Statistical models for data on recreational fishing

Danielle Hendricksen

School of Engineering, Mathematics and Physics

University of Kent

This dissertation is submitted for the degree of

Doctor of Philosophy

March 2025

Acknowledgements

I gratefully extend my appreciation to my supervisors, Eleni Matechou, David Maxwell, Oscar Rodriguez de Rivera Ortega, and Kieran Hyder. Their diverse expertise and genuine interest in this research, along with the collective time and effort they invested, have made this thesis possible. It has been a privilege to work with such a dedicated supervisory team.

I wish to express my deepest gratitude to Eleni for her guidance and support throughout my PhD journey. Her expertise and knowledge have been instrumental in the successful completion of this thesis. I am truly grateful for her commitment to this project. Her valuable suggestions, extensive feedback, and the time she has dedicated to supporting me in this research have greatly enhanced the quality of my work. Her contributions have been invaluable throughout this journey.

I wish to express my deepest gratitude to David for all his support and guidance through my PhD journey. His expertise in applied statistics, feedback and suggestions for refining my ideas and improving my work, and advice through the writing process have significantly contributed to the quality of my work. I truly appreciate the generous amount of time he has dedicated to supporting me on this journey. His patience, understanding, and encouragement through the challenges of this research have been exceptional, making this journey both successful and rewarding, and I will always be very grateful for this.

I wish to express my sincere gratitude to Oscar for his advice, support, and guidance. His extensive knowledge and expertise in spatiotemporal models were instrumental during the second year of my research, greatly enhancing my understanding and contributing significantly to the success of my research.

I wish to express my sincere gratitude to Kieran for his support during my research. His extensive fisheries expertise and comprehensive knowledge of recreational fishing provided a valuable external

perspective. His insightful research ideas, along with his valuable advice and suggestions were instrumental in shaping this research.

I am grateful for the opportunities and experiences provided by the University of Kent, which significantly enriched my PhD experience. I would like to express my appreciation to all the individuals across the University who contributed to providing these opportunities, as well as those who provided guidance and assistance. Their contributions have been valued throughout my academic journey.

I gratefully acknowledge the funding received from The University of Kent with a Vice Chancellor's Research Scholarship and also the School of Mathematics, Statistics and Actuarial Science. In addition, this work was supported by the Centre for Environment, Fisheries and Aquaculture Science (Cefas) via the Cefas internal Seedcorn funding programme.

Finally, I would like to express my gratitude to Fishbrain for providing the data used in this research.

Abstract

Recreational fishing is a globally significant activity, often involving angling but also encompassing various other methods in both marine and freshwater environments. While it provides notable health benefits, economic contributions, and conservation support, it can also exert substantial pressures on fish populations and ecosystems. To effectively manage and sustain this activity, it is crucial to quantify the scale, benefits, and impacts of recreational fishing. Traditional methods for monitoring, such as onsite surveys and recall surveys, face challenges in terms of cost, time, and data reliability, especially in accounting for the growing sector of angling tourism. Recent advancements in technology have introduced alternative data collection methods, particularly through smartphone applications. These apps, like Fishbrain, allow anglers to record their catches, offering a vast and continuous stream of data that could significantly enhance recreational fisheries management. However, to harness these data effectively, it is essential to develop robust statistical models to understand their strengths and limitations.

This thesis aims to advance the understanding and management of recreational fishing through the development and application of statistical models using data from traditional face-to-face surveys and app-based records.

Chapter 2 focuses on modelling marine recreational fishing data from a 2012-2013 UK survey, employing zero-inflated Poisson models with shrinkage methods to identify key predictors for catch rates. This chapter introduces a grid-based search algorithm for determining shrinkage penalties, and considers data on a number of key UK species.

Chapter 3 considers Fishbrain data from 2018 to 2021 to study the spatiotemporal patterns of recorded catches for four key marine species in the UK and Ireland. This analysis uses integrated

Laplace approximation methods to develop models that provide a framework for visualising large-scale catch data, comparing different error structures, and producing predictive maps for each species.

Chapter 4 extends the analysis to a global scale, examining angling tourism patterns using Fishbrain data. This chapter utilises network models and clustering methods to analyse over 100,000 catches, focusing on international travel and its changes in response to the COVID-19 pandemic. It identifies communities of countries with strong angling connections and explores the implications of these patterns for fisheries management.

Finally, Chapter 5 discusses the broader implications of the research and suggests future directions. The findings underscore the potential of app-based data to complement traditional survey methods, offering a rich resource for sustainable recreational fisheries management. This work contributes to the evidence base needed for informed marine policy and highlights the importance of integrating diverse data sources to support productive and sustainable fisheries.

Table of contents

List of figures **xi**

List of tables **xxvii**

1 Introduction **1**

2 Understanding factors driving recreational sea angling catch rates **5**

2.1 Introduction 7

2.2 Materials and methods 10

2.2.1 Survey 10

2.2.2 Covariates driving catch rates 11

2.2.3 Species 15

2.2.4 Zero-inflated Poisson models 21

2.2.5 Fitting the model 22

2.2.6 Checking goodness of fit with parametric bootstrapping 25

2.2.7 Simulations 25

2.3 Results 26

2.3.1 Simulation results 26

2.3.2 Case study results 27

2.4 Discussion 35

2.4.1 Modelling sea angling catches 35

2.4.2 Drivers of sea angling catches 39

2.4.3 Conclusion 39

3 Spatiotemporal models for mobile app records of fish catches 41

3.1 Introduction . . . 44

3.2 Materials and methods . . . 46

3.2.1 Data . . . 46

3.2.2 Modelling . . . 50

3.3 Results . . . 57

3.3.1 Models . . . 57

3.3.2 Predictions . . . 59

3.4 Discussion . . . 61

4 Global travel patterns in recreational fishing 65

4.1 Introduction . . . 67

4.2 Materials and methods . . . 72

4.2.1 Data preparation . . . 72

4.2.2 Networks and graphs . . . 99

4.3 Results . . . 107

4.3.1 Case study countries . . . 108

4.3.2 Europe . . . 122

4.3.3 World countries . . . 132

4.3.4 World regions . . . 142

4.3.5 Summary of results . . . 153

4.3.6 Comparing edge density in networks . . . 167

4.3.7 Comparing modularity in communities . . . 169

4.4 Discussion . . . 172

4.4.1 Findings . . . 172

4.4.2 Factors influencing travel and destination choices . . . 175

4.4.3 Comparing the United Kingdom and United States case studies . . . 183

4.4.4 Data and methods . . . 184

4.4.5 Limitations and biases . . . 186

4.4.6 Future work 190

5 Discussion **193**

5.1 Chapter 2 194

5.2 Chapter 3 196

5.3 Chapter 4 199

5.4 Summary of modelling approaches and future work 201

References **203**

Appendix A Understanding factors driving recreational sea angling catch rates – Supplementary material **221**

A.1 Data 222

A.1.1 Data cleaning 222

A.1.2 Covariates 222

A.2 Simulations 234

A.2.1 Setting up the simulations 234

A.2.2 Simulation results 238

A.3 Grid method 240

A.3.1 Fitting the models 240

A.3.2 Results 252

A.4 Default method 261

A.4.1 Fitting the models 261

A.4.2 Results 268

A.5 Comparing the grid and default methods 275

A.5.1 Loglikelihood values 275

A.5.2 Predicted catches 276

Appendix B Spatiotemporal models for mobile app records of fish catches – Supplementary material **279**

B.1 R code 280

B.2 Rate of catches reported 284
B.3 Computational time 289
B.4 Predictions 290

Appendix C Global travel patterns in angling – Supplementary material 295

C.1 Country codes 296
C.2 Regions manually allocated 298
C.3 General tourism trends 299
C.4 Travel patterns for individual species 303
C.5 Network graphs for the United Kingdom and United States with no minimum threshold 307
C.6 World regions community graphs 2017-2021 312
C.7 Number of catches recorded 314
C.7.1 Case study countries 315
C.7.2 Europe 319
C.7.3 World countries 323
C.7.4 World regions 327
C.7.5 Edge density 331

List of figures

2.1 Locations of catches, by species. 15

2.2 Proportion (%) of trips in each sea region in which cod, sea bass, mackerel and flounder are targeted (left column) and caught (right column). 17

2.3 Proportion (%) of trips in each sea region in which plaice, whiting, dab and skates/rays are targeted (left column) and caught (right column). 18

2.4 Proportion (%) of trips in each season of the year in which cod, sea bass, mackerel and flounder are targeted (left column) and caught (right column). 19

2.5 Proportion (%) of trips in each season of the year in which plaice, whiting, dab and skates/rays are targeted (left column) and caught (right column). 20

3.1 Rate of catches reported of sea bass, cod, mackerel and whiting for each quarter in 2021 at the locations in the Fishbrain app. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). . . . 49

3.2 INLA mesh used to fit the models. The mesh covers the UK and Ireland, and surrounding marine regions. 51

3.3 Lower bound of the number of users m_t, $t = 1, \ldots, 16$, by year and quarter. The dotted line corresponds to $t_1 - t_4$, the dotted and dashed line to $t_5 - t_8$, the dashed line to $t_9 - t_{12}$ and the solid line to $t_{13} - t_{16}$. 52

3.4 Rate of recorded catches, r_t, for each time period, per year and quarter, with $r_t = \sum_\ell y_{\ell t}/m_t$ where $y_{\ell t}$ is the number of recorded catches in time period t and location ℓ and m_t is the lower bound of the number of Fishbrain users at time period t. Note the different y-axis scales for the different species. 53

3.5 Locations at which predictions are generated. These points are within 1,000 metres inland to 22.22 km (12 nautical miles) offshore from the coast. 56

3.6 Predicted rates for 2021. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). The predicted rates are capped to the 90% quantile to optimise visualisation. 60

4.1 World regions. Data source: `countrycode` package in R (Arel-Bundock et al., 2018). 91

4.2 World regions: Number of trips from 2018 to 2021 inclusive. This shows the data given in Table 4.16. 95

4.3 World regions: Number of catches from 2018 to 2021 inclusive. This shows the data given in Table 4.17. 95

4.4 Number of trips and catches recorded for each origin-destination region pair, for the years 2018-2021. 97

4.5 Example graphs. 100

4.6 Network graphs of the trips to and from the United Kingdom. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than five trips are excluded. 112

4.7 Network graphs of the trips to and from the United States. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United States are purple. Edges representing trips to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than 30 trips are excluded. 116

4.8 Network graphs of the trips to and from Canada. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Canada are purple. Edges representing trips to Canada are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 118

4.9 Network graphs of the trips to and from Sweden. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Sweden are purple. Edges representing trips to Sweden are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 119

4.10 Network graphs of the trips to and from Australia. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Australia are purple. Edges representing trips to Australia are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 120

4.11 Network graphs of the trips to and from Mexico. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Mexico are purple. Edges representing trips to Mexico are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 121

4.12 The number of quarters during the period 1 January 2018 to 31 December 2021 in which each pair of countries is in the same community. The width of connections between countries is proportional to the number of times the connected countries are in the same community cluster. 127

4.13 Network graphs of Europe, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 129

4.14 Communities in the Europe networks based on the number of trips. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. 131

4.15 The number of quarters during the period 1 January 2018 to 31 December 2021 in which each pair of countries is in the same community. The width of connections between countries is proportional to the number of times the connected countries are in the same community cluster. 137

4.16 Network graphs of the world countries, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 139

4.17 Communities in the world countries networks based on the number of trips. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. 141

4.18 Network graphs of world regions, based on the number of trips between regions. The area of each vertex represents the total number of trips from that region to other regions. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The loops indicate the trips to a different country to the origin country, but still in the same region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 150

4.19 Communities in the world regions networks based on the number of trips (2018-2021). The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa 152

4.20 Trips between the United Kingdom and world regions. 154

4.21 Trips between the United States and world regions. 155

4.22 Number of trips from the top origin countries and to the top destination countries in the North America region. 159

4.23 Number of trips from the top origin countries and to the top destination countries in the Europe & Central Asia region. . . . 159
4.24 Number of trips from the top origin countries and to the top destination countries in the Latin America & Caribbean region. . . . 160
4.25 Number of trips from the top origin countries and to the top destination countries in the East Asia & Pacific region. . . . 160
4.26 Edge density (max vertices) for the networks based on the number of trips. . . . 168
4.27 modularity scores of the community clustering networks for Europe based on the number of trips. . . . 170
4.28 modularity scores of the community clustering networks for the world countries based on the number of trips. . . . 171
4.29 modularity scores of the community clustering networks for the world regions based on the number of trips. . . . 171
4.30 Mean international travel controls. These are calculated as the average of the restriction level each day in the corresponding quarter across the destination countries in the networks of the number of trips for the case. For example, the mean international travel control for Europe is based on only the countries in the Europe networks of trips, while for the United Kingdom and United States case studies, the level of restrictions to the case study country only is used to calculate the mean. Data source (Hale et al., 2021). . . . 181

A.1 Average catch rate, by species and type of fishing. . . . 226
A.2 Average catch rate, by species and fishing method. . . . 226
A.3 Average catch rate, by species and weekend. . . . 227
A.4 Average catch rate, by species and competitive fishing. . . . 227
A.5 Average catch rate, by species and angling club membership. . . . 228
A.6 Average catch rate, by species and sea. . . . 228
A.7 Average catch rate, by species and season. . . . 229
A.8 Average catch rate, by species and sea state. . . . 229

A.9 Average catch rate, by species and sediment type. 230
A.10 Correlations between numeric covariates. 230
A.11 Relationship between the number of fish of other species caught and catch size (of the modelled species). The plot shows the number of fish of the modelled species caught (y-axis) and the number of fish of other species caught (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting. 231
A.12 Relationship between angling group size and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of anglers in the angling group (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting. 231
A.13 Relationship between shore angling experience and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of days shore angling in the past three months (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting. 232
A.14 Relationship between boat angling experience and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of days boat angling in the past three months (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting. 232
A.15 Relationship between lunar illumination and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the lunar illumination on the day of the fishing trip (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting. 233
A.16 Loglikelihood values in simulations with 10 folds. 235
A.17 R code to generate the simulated data. 236

A.18 R code to run simulation models. 237

A.19 Loglikelihood values with 10 folds. 240

A.20 Loglikelihood values with different numbers of folds. 242

A.21 Loglikelihood values for pairs of λ values used in the grid method for mackerel. Zoom levels: 0 and 1. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. 243

A.22 Loglikelihood values for pairs of λ values used in the grid method for mackerel. Zoom levels: 2 and 3. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. 244

A.23 Grids using a wide range of λ values. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. Species shown: cod, sea bass, mackerel, and whiting. 246

A.24 Grids using a wide range of λ values. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. Species shown: flounder, plaice, dab and skates/rays. 247

A.25 Time taken (seconds) for each pair of λ values used in the grid method for mackerel. Zoom levels: 0 and 1. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. 250

A.26 Time taken (seconds) for each pair of λ values used in the grid method for mackerel. Zoom levels: 2 and 3. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. 251

A.27 Number of observed and predicted catches using the grid method for each interview. Species shown: cod, sea bass, mackerel and flounder. 252

A.28 Number of observed and predicted catches using the grid method for each interview. Species shown: plaice, whiting, dab and skates/rays. 253

A.29 Values of the λ pairs used in the Poisson and Binomial parts of the model using the default method with 100 λ pairs. Species shown: cod, sea bass, mackerel and flounder. Note the different axes scales for the different species. 262

A.30 Values of the λ pairs used in the Poisson and Binomial parts of the model using the default method with 100 λ pairs. Species shown: plaice, whiting, dab, and skates/rays. Note the different axes scales for the different species. 263
A.31 Loglikelihood value for each pair of λ values for the default method with n folds. . . 264
A.32 Number of observed and predicted catches using the default method for each interview. Species shown: cod, sea bass, mackerel and flounder. 273
A.33 Number of observed and predicted catches using the default method for each interview. Species shown: plaice, whiting, dab and skates/rays. 274
A.34 Grids with λ values used in the default method. The red diamond indicates the maximum loglikelihood found by the grid method. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. 275
A.35 Number of predicted catches for each interview using the grid and default methods. Species shown: cod, sea bass, mackerel and flounder. 276
A.36 Number of predicted catches for each interview using the grid and default methods. Species shown: plaice, whiting, dab and skates/rays. 277

B.1 R code to prepare the data. 281
B.2 R code to run the model. 282
B.3 R code to generate predictions from the model. 283
B.4 Rate of catches reported for sea bass in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.21. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). . . . 285
B.5 Rate of catches reported for cod in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.034. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). . . . 286

B.6 Rate of catches reported for mackerel in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.017. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). . . . 287

B.7 Rate of catches reported for whiting in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.054. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). . . . 288

B.8 Predicted number of catches of sea bass per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear. 291

B.9 Predicted number of catches of cod per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear. 292

B.10 Predicted number of catches of mackerel per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear. 293

B.11 Predicted number of catches of whiting per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear. 294

C.1 Number of tourists travelling from the top origin countries and to the top destination countries in the North America region. Outbound data are not available for the United States of America in 2021. Source: (World Tourism Organization, 2023) 299

C.2 Number of tourists (except United Kingdom departures, which consist of tourists and visitors) travelling from the top origin countries and to the top destination countries in the Europe & Central Asia region. Source: (World Tourism Organization, 2023) . . . 300

C.3 Number of tourists travelling from the top origin countries and to the top destination countries in the Latin America & Caribbean region. Outbound data are not available for Brazil. Number of outbound tourists from Puerto Rico (2018-2021): 731,000, 766,000, 688,000, and 659,000. Source: (World Tourism Organization, 2023) 301

C.4 Total departures and arrivals (tourists and visitors) from the top origin countries and to the top destination countries in the East Asia & Pacific region. Total arrivals to Australia (2018-2021): 9,246,000, 9,466,000, 1,828,000, and 246,000. Source: (World Tourism Organization, 2023) 302

C.5 Network graphs for Largemouth black bass, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 304

C.6 Network graphs for Northern pike, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 305

C.7 Network graphs for Smallmouth bass, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 306

C.8 Network graphs of the trips to and from the United Kingdom. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United Kingdom are included. 308

C.9 Network graphs based on catches recorded by users travelling from and to the United Kingdom. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United Kingdom are purple. Edges representing catches recorded by users travelling to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United Kingdom are included. 309

C.10 Network graphs of the trips to and from the United States. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United States are included. 310

C.11 Network graphs based on catches recorded by users travelling from and to the United States. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United States are purple. Edges representing catches recorded by users travelling to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United States are included. 311

C.12 Communities in the world regions networks based on the number of trips (2017-2021). The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa 313

C.13 Network graphs based on catches recorded by users travelling from and to the United Kingdom. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United Kingdom are purple. Edges representing catches recorded by users travelling to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than five trips are excluded. 316

C.14 Network graphs based on catches recorded by users travelling from and to the United States. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United States are purple. Edges representing catches recorded by users travelling to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than 30 trips are excluded. 318

C.15 Network graphs of Europe, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 320

C.16 Communities in the Europe networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. 322

C.17 Network graphs of world countries, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 324

C.18 Communities in the world countries networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. 326

C.19 Network graphs of world regions, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the region corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The loops indicate the number of catches recorded in a different country to the origin country, but within the same region. The area of the vertices and width of the edges are scaled consistently across all years and quarters. 328

C.20 Communities in the world regions networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa 330

C.21 Edge density (max vertices) for the networks based on the number of catches. 331

List of tables

2.1 Covariates relating to trip characteristics. 13

2.2 Covariates relating to angler characteristics. 14

2.3 Covariates relating to environment characteristics. 14

2.4 Number of fish caught per trip. 21

2.5 Number of fish caught per trip, by species. 21

2.6 Mean estimated values of coefficients in simulation results. 27

2.7 The λ values of the final model for the grid and default methods, using n folds. The λ values used in the grid method are displayed with the number of decimal places corresponding to the precision used in the calculations. The λ values used in the default method are displayed to five decimal places. 28

2.8 The log-likelihood values of the final model for the grid and default methods, using n folds. 28

2.9 Estimated coefficients in the Binomial part (whether or not fish were available) using the grid method. A positive value indicates a higher chance of a zero, and therefore a negative association between the covariate and the availability of fish. A negative value indicates a lower chance of a zero, and therefore a positive association between the covariate and the availability of fish. A zero indicates no association between the covariate and the availability of fish. Non-zero values are displayed to two decimal places. 29

2.10 Estimated coefficients in the Poisson part (how many fish of the species were caught given fish of that species were available) using the grid method. Non-zero values are displayed to two decimal places. 32
2.11 Proportion of observations within the interquartile range of the simulated values. . . 35

3.1 Frequency table of the number of catches recorded in each quarter at each location and for each species. The total number of location-quarter combinations is 21,584. . 48
3.2 Watanabe–Akaike information criterion (WAIC) and effective number of parameters for the iid and ar1 models for each species. 57
3.3 Posterior summaries of the intercept and of the fixed effects of quarters in the model for all species. The baseline is Q1. 58
3.4 Posterior summaries of the random effects in the best fitting model. 58

4.1 Users recording a catch, catches recorded and trips taken each year, as a result of travel to a different country. The number of users is the number of active users each year. Users may be active in multiple years. The number of trips per year is the sum of the number of trips in each quarter of that year. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. The proportions of users sum to greater than 100% due to some users recording catches in multiple years. The proportions of trips and catches do not sum to 100% due to rounding. . . 77
4.2 Users recording a catch, trips taken and catches recorded according to users' origin country, as a result of travel to a different country during the period 2018-2021. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. Countries are listed in order of number of users. 78

4.3 Users, trips and catches by destination, as a result of travel to a different country during the period 2018-2021. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. 'Users' is the number of users recording a catch as occurring in the country, 'trips' is the number of trips taken to the country, and 'catches' is the number of catches recorded as occurring in the country. Countries are listed in order of number of users. 79

4.4 Number of catches recorded by species 2018-2021, for the top 15 species recorded. The remaining 47% of species recorded each contributed less than 1% of the total number of catches recorded. The proportions are based on the total number of catches with the species recorded (92,489). 80

4.5 Number of catches recorded per year for the three most commonly recorded species. 80

4.6 Most common catch locations (destination country) for the three most commonly recorded species. 80

4.7 Method of fishing, for the catches recorded during the period 2018-2021. Note that only 38% of catch records have the species recorded. 81

4.8 Top destination countries from the United Kingdom and origin countries to the United Kingdom, according to the number of trips from and to the United Kingdom respectively, from 2018 to 2021 inclusive. The proportions are based on the total number of trips from the United Kingdom (2,153) and to the United Kingdom (2,400) respectively. 82

4.9 Top destination countries from the United Kingdom and origin countries to the United Kingdom, according to the number of catches reported by United Kingdom users and the number of catches reported as occurring in the United Kingdom respectively, from 2018 to 2021 inclusive. The proportions are based on a total of 5,491 catches reported by users from the United Kingdom and 5,636 catches reported in the United Kingdom respectively. 83

4.10 Top destination countries from the United States and top origin countries to the United States, according to the number of trips from and to the United States respectively, from 2018 to 2021 inclusive. The proportions are based on the total number of trips from the United States (31,658) and to the United States (7,098) respectively. 84

4.11 Top destination countries from the United States and origin countries to the United States, according to the number of catches reported by United States users and the number of catches reported as occurring in the United States respectively, from 2018 to 2021 inclusive. The proportions are based on a total of 52,847 catches reported by users from the United States and 26,133 catches reported in the United States. 85

4.12 Top 15 origin and destination countries in Europe according to the number of trips from and to these countries respectively, from 2018 to 2021 inclusive. The number of trips from the origin countries are the number of outbound trips from these countries to countries in Europe. The number of trips to the destination countries are the number of trips made to these countries from countries in Europe. Proportions are percentages of the total number of trips (6,454) and are given to 1 decimal place. Totals may not sum to 100% due to rounding. 87

4.13 Top 15 origin and destination countries in Europe according to the number of catches reported as occurring in Europe by users from the origin countries, and the number of catches reported as occurring in the destination countries by users from origin countries in Europe, respectively, from 2018 to 2021 inclusive. The proportions are based on 14,721 catches. 87

4.14 Top 15 world countries by number of trips from 2018 to 2021 inclusive. 90

4.15 Top 15 world countries by number of catches reported from 2018 to 2021 inclusive. . 90

4.16 World regions: Number of trips from 2018 to 2021 inclusive. Proportions are percentages of the total number of trips and are given to 1 decimal place. Totals may not sum to 100% due to rounding. These data are shown in Figure 4.2. 94

4.17 World regions: Number of catches reported from 2018 to 2021 inclusive. Proportions are percentages of the total number of catches and are given to 1 decimal place. Totals may not sum to 100% due to rounding. These data are shown in Figure 4.3. 94

4.18 Number of trips from each origin region and to each destination region, by year (including trips to other countries within the same region). 96
4.19 Number of trips for each origin-destination region pair 2018-2021. This information is shown in Figure 4.4 (a). 97
4.20 Number of catches recorded for each origin-destination region pair 2018-2021. This information is shown in Figure 4.4 (b). 97
4.21 Number of catches recorded by users from each country within each region 2018-2021. 98
4.22 Metrics for the United Kingdom network graphs based on the number of trips. In each year-quarter time period, the networks contain only the edges where there are at least five trips for that origin-destination country pair. 111
4.23 Metrics for the United States case study network graphs based on the number of trips. In each year-quarter time period, the networks contain only the edges where there are at least 30 trips for that origin-destination country pair. 115
4.24 Metrics for Europe network graphs based on the number of trips. 128
4.25 Number of quarters (maximum 16) in which each pair of countries in the Europe networks based on trips is in the same community. The countries are ordered geographically, approximately west to east, followed by Mediterranean countries. 130
4.26 Metrics for world countries network graphs based on number of trips. 138
4.27 Number of quarters (maximum 16) in which each pair of countries in the world countries networks based on trips is in the same community. The countries are ordered approximately geographically in a clockwise direction (the Americas, Europe, China, New Zealand, Australia, South Africa). 140
4.28 Metrics for world regions network graphs based on number of trips. 149
4.29 World regions: trips. Number of quarters (maximum 16) in which the corresponding pair of regions is in the same community from 2018-2021. 151
4.30 Comparing the communities in the networks of world regions based on the number of trips before and during the COVID-19 pandemic. Number of quarters in which the corresponding pair of regions is in the same community: (a) 2018 quarter 1 to 2020 quarter 1 inclusive (b) 2020 quarter 2 to 2021 quarter 4 inclusive. 151

A.1 Number (and percentage) of trips, by type of fishing 225

A.2 Mean of numeric covariates, by type of fishing. The means for shore fishing are based on 1,143 anglers and the means for boat fishing are based on 180 anglers. 225

A.3 Simulation results with constant parameters for the Binomial and Poisson distributions ($p = 0.5$ and $\lambda = 6$ respectively), using n folds. 238

A.4 Simulation results with constant parameters for the Binomial and Poisson distributions ($p = 0.5$ and $\lambda = 6$ respectively), using 10 folds. 238

A.5 Simulation results with parameters for the Binomial and Poisson distributions as a function of covariates, using n folds. 239

A.6 Simulation results with parameters for the Binomial and Poisson distributions as a function of covariates, using 10 folds. 239

A.7 Grid method durations. Number of λ pairs: zoom level 0 = 81, zoom level 1 = 400, zoom level 2 = 420, zoom level 3 = 420. Zoom level 3 was only used when at least one λ value at the maximum loglikelihood in zoom level 2 was 0.001 or less. There are no level 3 durations for skates/rays as the λ values corresponding to the highest loglikelihood at zoom level 2 were both greater than 0.001 for this species. 248

A.8 λ values used by the default method with 100 λ pairs and n folds and the resulting loglikelihood. Species shown: cod, sea bass, mackerel, and flounder. 265

A.9 λ values used by the default method with 100 λ pairs and n folds and the resulting loglikelihood. Species shown: plaice, whiting, dab, and skates/rays. 266

A.10 Default method durations (hours). 267

A.11 Estimated coefficients in the Binomial part (whether or not fish were available) using the default method. A positive value indicates a higher chance of a zero, and therefore a negative association between the covariate and the availability of fish. A negative value indicates a lower chance of a zero, and therefore a positive association between the covariate and the availability of fish. A zero indicates no association between the covariate and the availability of fish. Non-zero values are displayed to two decimal places. 268

A.12 Estimated coefficients in the Poisson part (number of fish caught) using the default method. A negative value indicates the covariate is negatively associated with the number of fish caught, a positive value indicates the covariate is positively associated with the number of fish caught and zero indicates there is no association between the covariate and the number of fish caught. Non-zero values are displayed to two decimal places. 270

A.13 Total number of observed catches and total number of predicted catches with the grid and default methods. 278

B.1 Computational time (minutes) to fit the iid and ar1 models for each species. 289

B.2 Computational time (minutes) to generate predictions from the ar1 model. Predictions for each quarter are based on data from the years 2018 to 2021 inclusive. Values are rounded to one decimal place. The total is the sum of quarters 1 to 4. Totals may not equal the sum of the quarterly values due to rounding. 289

C.1 Country codes and regions, listed alphabetically by country code (Arel-Bundock et al., 2018). 297

C.2 (a) Origin and (b) destination countries that did not have a matching region using the countrycode() function from the `countrycode` package Arel-Bundock et al. (2018) in R, the region to which these countries are allocated, and the number of catch records. 298

C.3 Mean edge density, mean distance and diameter for network graphs with number of catches, over the 16 quarters from 1 January 2018 to 31 December 2021. 314

C.4 Metrics for the United Kingdom network graphs based on the number of catches recorded. In each year-quarter time period, the networks contain only the edges where there are at least five trips for that origin-destination country pair. 315

C.5 Metrics for United States network graphs based on the number of catches recorded. In each year-quarter time period, the networks contain only the edges where there are at least 30 trips for that origin-destination country pair. 317

C.6 Metrics for Europe network graphs based on the number of catches recorded. 319

C.7 Number of quarters (maximum 16) in which each pair of countries in the Europe networks based on catches is in the same community. The countries are ordered geographically, approximately west to east, followed by Mediterranean countries. . . 321

C.8 Metrics for world countries network graphs based on the number of catches recorded. 323

C.9 Number of quarters (maximum 16) in which each pair of countries in the world countries networks based on catches is in the same community. The countries are ordered approximately geographically in a clockwise direction (the Americas, Europe, China, New Zealand, Australia, South Africa). 325

C.10 Metrics for world regions network graphs based on the number of catches recorded. . 327

C.11 Number of quarters (maximum 16) in which each pair of regions in the world regions networks based on catches is in the same community. 329

Chapter 1

Introduction

Recreational fishing is an important activity globally. It is defined as fishing when the fishers do not sell the fish they catch and the fishing activity is not predominantly for subsistence, cultural or heritage purposes (Pawson et al., 2008). It is often synonymous with angling (attempting to catch fish on hooks), but can include other methods, and encompasses both marine and river fishing. Recreational fishing can have substantial health benefits for individuals (Hyder et al., 2018), economic benefits for countries (Hyder et al., 2018) and local communities (Pita et al., 2021), and the ability to help drive conservation (Arlinghaus et al., 2021). However, it can also have considerable effects on fish populations and ecosystems (Arlinghaus and Cooke, 2009; Cooke and Cowx, 2004; Cooke et al., 2007; Dainys et al., 2022b; Hyder et al., 2014, 2018; Lewin et al., 2006, 2019; National Research Council, 2006). It is therefore important to be able to quantify the scale, benefits, and impact of recreational fishing activities.

Quantifying these measures can be challenging. Recreational fishing activity is monitored using standardised onsite surveys (such as creel and aerial surveys), recall surveys (such as post or phone surveys), and angler diaries. Onsite surveys are expected to provide unbiased and representative data. However, they are are often time-consuming and costly due to the challenges of surveying long coastlines and diverse fishing practices associated with this seasonally and spatially-variable activity (Skov et al., 2021). Meanwhile, recall surveys can have low response rates and potential bias (Dainys et al., 2022a).

Assessing fishing activity among individuals visiting a country presents significant challenges, as diary-based or telephone surveys typically do not include tourists. This segment of tourism, known as fishing tourism, is part of the ocean-related tourism economy (Hall, 2021; Mordue, 2009). Fishing tourism encompasses both domestic and international travel, with recreational fishing either as the primary reason for the travel or as a secondary activity during travel for other purposes (van den Heuvel et al., 2022). In some countries the economic contribution from international anglers can be substantial (Cisneros-Montemayor and Sumaila, 2010; Hannonen and Hoogendoorn, 2022). However, potential negative aspects of fishing tourism include ecological impacts (Cooke and Cowx, 2004; Dainys et al., 2022b) and natural capital impacts to local communities (Hall, 2021). Therefore, understanding fishing tourism can help balance the needs of fisheries, the environment, and local communities.

In recent years, new, alternative sources for collecting recreational fishing data have arisen. With the widespread use of smartphones, apps that collect recreational catch data have been developed (Fishbrain, 2023; Gundelund et al., 2021; Jiorle et al., 2016; Noleto-Filho et al., 2021; Papenfuss et al., 2015) and a large amount of information on recreational catches is uploaded to apps daily (Venturelli et al., 2017). Apps that enable anglers to record their catches could be a useful source of inexpensive catch information (Gundelund et al., 2020; Venturelli et al., 2017) to support fisheries management. For example, Fishbrain (Fishbrain, 2023) has millions of users worldwide logging thousands of catches each week. This provides an opportunity for a source of continuous, extensive citizen-science data. However, development and evaluation of modelling for this data source is required to understand its strengths and challenges.

This thesis uses data from face-to-face beach surveys of anglers[1] and anonymised Fishbrain data[2] obtained through a data-sharing agreement, building corresponding modelling approaches to generate new insights from the data. Chapter 2 starts with new modelling of marine recreational fishing data from a face-to-face survey in the United Kingdom. Chapter 3 then models Fishbrain data in this region to study spatial and temporal patterns in recorded catches of marine species, and Chapter 4 extends the scope to investigate international tourism patterns in marine and freshwater recreational fishing.

[1] *Sea Angling 2012 - Shore and private boat survey* dataset, available at https://doi.org/10.14466/CefasDataHub.27.
[2] Fishbrain data requests should be made directly to Fishbrain (fishbrain.com).

Specifically, Chapter 2 considers data from surveys of sea angling carried out around the coast of England in 2012 to 2013 (Armstrong et al., 2013), together with shrinkage methods in zero-inflated Poisson models, to identify important predictors for catch rates in angling trips. The chapter builds on existing approaches for fitting such models and identifying the shrinkage penalties, but proposes an alternative, grid-based search algorithm for identifying the penalty terms, which is shown to lead to models with a better predictive ability. The work demonstrates how shrinkage methods can be employed as an alternative to model selection criteria and provide a reliable approach for variable selection when modelling catch rates, especially in cases where there are a potentially large number of predictors to consider.

Chapter 3 considers Fishbrain reported catches for the years 2018 to 2021, for four core marine species: European sea bass (*Dicentrarchus labrax*), Atlantic cod (*Gadus morhua*), Atlantic mackerel (*Scomber scombrus*), and whiting (*Merlangius merlangus*), building spatiotemporal models to understand and predict reported catches around the coast of the United Kingdom and Ireland. The work, for the first time, considers this large unexploited source of marine recreational fishing data together with spatiotemporal models fitted using integrated Laplace approximation approaches (Rue et al., 2009), providing a framework for modelling and visualising large scale catch data. The chapter compares the fit of different error structures, and produces maps of predicted rates of reported catches for each species.

Chapter 4 considers global Fishbrain data to explore patterns of fishing tourism. This gives rise to a substantial data set, initially with over 7 million entries during the years 2018 to 2021, processed to 104,423 catches recorded by 37,531 users across 50,566 trips for the analysis of international travel. The chapter considers network models, and in particular directed graphs, for summarising and visualising data in terms of fishing trips between pairs of countries. It examines patterns using multiple subsets of the data covering world regions, world countries, and European countries, as well as the United Kingdom and United States as case studies to provide illustrative examples of travel related to individual countries. Clustering methods are used to identify communities of countries that have strong links in terms of travel of anglers between them. The spatial and temporal scale of the information allows changes in the travel patterns to be considered in relation to the COVID-19 pandemic and associated travel restrictions.

Finally, Chapter 5 provides a discussion of the work so far and of potential future research directions.

This thesis develops and applies statistical models to understand drivers and patterns of catch and fishing activity. From recent trends, it is likely that the scale and reach of app-based data will continue to grow. Therefore, methods to combine app data with face-to-face survey data will be a desirable future research direction. The work presented, and future developments from it, will add to the evidence-base for marine policy and recreational fisheries management, to support productive and sustainable seas.

Chapter 2

Understanding factors driving recreational sea angling catch rates

Abstract

Recreational sea angling catches can have considerable impact on fish stocks. At the same time, there is large variation in catches between anglers, times, and locations. This affects both our ability to reconstruct catches over time for use in stock assessment and the development of suitable management measures to restrict catches. However, the factors driving this variation are not well understood. In this study, we considered data from the onsite shore and private boat survey of recreational sea angling activity carried out around the entire coast of England in 2012-13. We fitted zero-inflated Poisson models to catches of eight species from 1,323 sea angling trips together with angler-specific, trip-related, and environmental variables. We used penalised regression, with the Least Absolute Selection and Shrinkage Operator (LASSO) penalty, to identify important predictors for catch rates of each species. Our findings suggest that duration of fishing and targeting a species are important predictors for the probability of fish availability, while duration of fishing and the number of other fish caught are important predictors for the catch rate, given availability, with fish availability and catch rates tending to increase as duration of fishing increases. Our work demonstrates that penalised regression offers an automated approach for identifying important predictors out of large pools of variables in this case, and hence provides the foundation for modelling and understanding the effects of angler-specific, trip-related, and environmental covariates on the catch rates of recreational angling.

2.1 Introduction

Recreational angling is an important activity that provides social and economic benefits, and can support rural communities (Arlinghaus and Cooke, 2009), but can also have substantial impacts on fish populations and ecosystems (Arlinghaus and Cooke, 2009; Cooke et al., 2007; Dainys et al., 2022b; Hyder et al., 2018; Lewin et al., 2006; National Research Council, 2006). However, it is

challenging to quantify this impact because it is difficult to estimate the number of fish caught by recreational anglers due to the diverse and dispersed nature of the activity. Hence, there is a general lack of data and methods for understanding the impact of recreational angling on fish stocks (Hyder et al., 2014, 2018, 2020b).

Estimating catch rates is traditionally done through estimating effort and catch per unit effort, with data obtained through surveys (Hyder et al., 2020b). There are many factors that could impact on recreational catch rates, including angling effort, anglers' expertise and knowledge, technology used, and fish abundance, size, feeding habits and quality of habitats (Mosley et al., 2022), and fish learning and experience (Louison et al., 2019). More recently, data mining and machine learning methods are being applied in fisheries, including for estimating catch rates (Gladju et al., 2022). However, the factors driving catch rates are still not well understood. One study, in Australia, which looked at the impact of sea surface temperature, current speed and fish abundance on catch rates of dolphinfish, found that only the sea surface temperature was significantly related to catch rates (Folpp and Lowry, 2006). Lunar phases were shown to impact catch rates for some species, but not others (Pulver, 2017). A study of the impacts on catch rates of the freshwater species Eurasian perch (*Perca fluviatilis*) in lakes in Germany showed that catch rates of this species were influenced by angler choices and characteristics and environmental factors, with angler experience, the type of bait/lure used, target species, water transparency and the seasonal availability of food all influencing catch rates (Heermann et al., 2013), while for marine species Hyder et al. (2021) and Hyder et al. (2024) identified avidity, age, and region as important predictors of catch by individual sea anglers in the United Kingdom (UK), with gender also recognised as a significant predictor in Hyder et al. (2024).

Identifying covariates that impact catch rates from a large number of potential covariates requires variable selection strategies. However, modelling recreational fishing catches is not straightforward as anglers frequently catch very few fish per trip, or none at all, so the distribution of fish caught is often highly skewed and zero-inflated. Zero-inflated data must be modelled by taking appropriate account of the excess zeros using corresponding zero-inflated models, as inappropriate models can lead to unreliable inference through biased parameter estimates and unreliable standard errors (Zuur et al., 2009). Additionally, when there is a large pool of available covariates to consider, employing variable selection approaches such as stepwise regression can be very time-consuming and can require

fitting of a considerably large number of models. Finally, with sparse data such as those considered in this chapter, overly complex models or different combinations of covariates can lead to optimisation issues, which often prohibits the use of automatic variable selection techniques, such as stepwise regression using AIC (Burnham and Anderson, 1998).

Penalised regression can offer a reliable and computationally feasible approach in such cases (Greenwood et al., 2020; Pavlou et al., 2016), and is useful when there is collinearity between covariates (Dormann et al., 2012). Penalised regression methods rely on estimation instead of hypothesis testing to select a model, and avoid multiple testing (Wu and Lange, 2008). Despite these advantages, to the best of our knowledge, penalised regression methods have not been used to identify important predictors of recreational fishing catches in the literature.

Sea angling is an important activity in England with around 2% of adults participating, spending £1.23 billion, and catching 10.1 million fish each year (Armstrong et al., 2013). Estimates of sea angling catches were generated in 2012-13 that included an onsite shore and private boat survey (Armstrong et al., 2013). However, no attempt was made to understand the factors driving catches by sea anglers on individual trips, which would provide useful information for both stock assessment and future surveys.

In this study, we propose a modelling framework for the number of catches per trip and associated covariates. We demonstrate the use of zero-inflated Poisson models, with the number of fish caught per trip for each species as the response, which is regressed against the available covariates. We illustrate the use of LASSO penalised regression for identifying the important predictors out of the available covariate pool for zero-inflated catch data and an extensive search algorithm for estimating the LASSO penalty terms in zero-inflated Poisson regression.

Penalised regression, such as the LASSO, relies on cross-validation approaches for selecting the most appropriate value of the penalty parameter. This can be computationally demanding, especially in the case of the zero-inflated models considered in this chapter, where a separate penalty term needs to be chosen for each part of the model (Binomial and Poisson). In this chapter, we develop an extensive search approach for selecting the two penalty terms, and compare its performance to default methods as implemented in an existing R package (Wang, 2020). Our results suggest that, for our data set, our approach can outperform the default in terms of yielding a model with lower cross

validation error, and hence better predictive ability. However, the time taken to run the models was considerably higher with our extensive search than using the method's default approach which uses an expectation-maximisation algorithm based on coordinate descent methods (Wang, 2020).

We apply our modelling framework to data on eight commonly-caught species: Atlantic cod (*Gadus morhua*), European sea bass (*Dicentrarchus labrax*), Atlantic mackerel (*Scomber scombrus*), European plaice (*Pleuronectes platessa*), European flounder (*Platichthys flesus*), whiting (*Merlangius merlangus*), common dab (*Limanda limanda*), and the skates and rays family (family Rajidae) (hereafter referred to as cod, sea bass, mackerel, plaice, flounder, whiting, dab, and skates/rays). Our results show that targeting a species is positively correlated with higher catch rates for that species, and that anglers tend to know which species may be available in the local area and target these species accordingly. For the majority of these eight species, almost all of the 18 covariates used were identified by the model as relevant to either the availability of the species or the catch rate given the species is available, however the effects varied between the different species. For most species, more covariates impacted the availability of fish than the catch rate. This is in contrast to the results obtained with the approach of Wang (2020), which indicated that across the species fewer covariates impacted the availability of fish than the catch rate.

The remainder of the chapter proceeds as follows. The methodology is presented in Section 2.2. Section 2.3 presents the results of a simulation study and of the analysis of the real data, and Section 2.4 contains a discussion of the findings.

2.2 Materials and methods

2.2.1 Survey

In 2012-13, surveys of sea angling were carried out around the coast of England (Armstrong et al., 2013). The surveys recorded angling activity, economic value and social aspects of anglers participating in sea angling in England. We used the data from the onsite survey of shore and private boat angling. Interviews were carried out over a 12-month period. The interviews took place at randomly selected shore angling and boat landing sites, with an increased known probability of selecting high activity sites to increase efficiency in use of surveyors. Where possible all anglers were interviewed,

otherwise anglers were selected randomly. Anglers were interviewed during or at the end of their fishing trips. A trip was defined as the fishing activity of a particular type (shore/boat) on a single day, which could be split across different locations or separate periods during the day. The data are available from the Cefas Data Hub (Armstrong et al., 2016), and contain angler, trip, catch and environment information, described in Section 2.2.2. A description of the survey methodology can be found in Armstrong et al. (2013) and Armstrong et al. (2016).

At each interview, the fishing activity of all anglers was recorded either for an individual angler or for a group of anglers according to whether or not the interviewed angler was pooling their catches with other anglers. Specifically, if interviewed shore anglers were pooling their catches, so that it was not possible to determine which fish was caught by which angler, then the angling activity of the group as a whole was recorded. If shore anglers were keeping their catches separate to other anglers, then the angling activity of the individual angler only was recorded, regardless of whether the angler was fishing alone or in a group with other anglers. When interviewing boat anglers, the angling activity of all anglers in the boat was recorded. Angling activity for the total fishing activity that had taken place on the day of the interview was recorded. This could be across any number of locations or separate periods, but needed to have been the same type (shore/boat) on the same day. Breaks and travel time were not included in the recorded time fished. After removing records with inconsistent or conflicting information, 1,323 interviews remained. Details on the data cleaning are given in Appendix A.

2.2.2 Covariates driving catch rates

The covariates available related to trip, angler and environment characteristics. When the interviewed angler was fishing within a group, the angler characteristics were only recorded for the 'primary respondent' and not for other group members. The trip and environment characteristics were the same for all anglers in an angling group. We calculated the number of hours fished in daylight, twilight and darkness from the time of interview and duration fished, both of which are recorded in the survey data. The duration fished was recorded as the actual number of hours (to the nearest half hour) spent fishing on the day of the interview at the time of the interview of the type of fishing (shore/boat) recorded at the interview, and does not include any planned further fishing time. The season of the

year and whether fishing was on a weekday or weekend were obtained from the date of the interview. We categorised sea state, available in the data as Beaufort scale values, as 'calm', 'moderate' and 'rough', representing Beaufort values 0-2, 3-4, and 5-7, respectively, due to the non-linear nature of the Beaufort scale. These conditions correspond to waves that do not break, wave crests that break, and waves that break with varying levels of spray, respectively (Met Office, 2010).

In addition to covariates collected as part of the survey, we included lunar illumination, as the movements of some fish species have been associated with the lunar cycle (Lowry et al., 2007). We calculated the lunar illumination for all trips in the data using the R package `lunar` (Lazaridis, 2014). We also included the type of seabed sediment in the area of the catch (Cooper et al., 2010), as fish distributions are well-known to be related to sediment type.

All covariates used in the model are given in Tables 2.1, 2.2 and 2.3, classified according to trip, angler and environment characteristics respectively. Further details on the covariates are given in Appendix A.

Table 2.1 Covariates relating to trip characteristics.

Covariate	**Description**	**Summary information**
Type	Whether the fishing was from shore or by boat. Of the 174 interviews with boat anglers, all were in boats with engines except 14 anglers (8% of boat anglers) who were in kayaks and 3 anglers (2% of boat anglers) who were in boats without engines.	Categorical *Shore* (87%) *Boat* (13%)
Method	Angling method used.	Categorical *Bait* (79%) *Lure/fly/jig* (11%) *Mix* of bait and lure/fly/jig methods (9%)
Total anglers	Number of anglers in the angling group and whose total catches were recorded.	Integer mean = 1.7, median = 1, minimum = 1, maximum = 11
Hours fished in daylight	Duration fished between sunrise and sunset.	Numerical mean = 2.3, median = 2, minimum = 0, maximum = 11.2
Hours fished in twilight	Number of hours fished during civil twilight (the period when the sun is below the horizon, but outdoor activities can usually continue without artificial illumination).	Numerical mean = 0.6, median = 0, minimum = 0, maximum = 1.5
Hours fished in darkness	Number of hours fished during nautical twilight, astronomical twilight and night. (During these periods artificial illumination is required for outdoor activities.)	Numerical mean = 0.1, median = 0, minimum = 0, maximum = 9.6
Target	Whether the species being modelled as the response variable was being targeted, either as the primary or secondary target.	Categorical *Yes* *No*
Competition	Whether or not the anglers were fishing in a competition at the time of the interview.	Categorical *Yes* (9%) *No* (91%)
Type of day	Whether the angling trip took place on a weekday or at the weekend.	Categorical *Weekday* (Monday to Friday) (45%) *Weekend* (Saturday and Sunday) (55%)
Other fish caught	Number of fish of other species caught (the total number of fish caught by the group excluding the species being modelled as the response variable).	Integer (Summary statistics vary according to species modelled.)

Table 2.2 Covariates relating to angler characteristics.

Covariate	Description	Summary information
Days shore angling	Number of days the interviewed angler participated in shore angling during the previous three months.	Numeric mean = 11.5, median = 5, minimum = 0, maximum = 92
Days boat angling	Number of days the interviewed angler participated in boat angling during the previous three months.	Numeric mean = 1.4, median = 0, minimum = 0, maximum = 90
Club membership	Whether the interviewed angler belonged to an angling club.	Categorical *Yes* = 21%, *No* = 79%

Table 2.3 Covariates relating to environment characteristics.

Covariate	Description	Summary information
Sea	Region of the coast, defined by the seas around the coast of England. As the Bristol Channel and Celtic Sea are adjacent they were combined into one category due to low counts in these categories.	Categorical *English Channel* *North Sea* *Bristol Channel/Celtic Sea* *Irish Sea*
Season	The quarter of the year in which the interview took place.	Categorical *Spring* (April to June) *Summer* (July to September) *Autumn* (October to December) *Winter* (January to March)
Sea state	The height of the waves, obtained from the Beaufort Scale value recorded at the interview	Categorical *Calm* (Beaufort Scale 0-2; waves up to 0.2m; waves do not break) *Moderate* (Beaufort Scale 3-4; waves 0.6m to 1m; wave crests break) *Rough* (Beaufort Scale 5-7; waves 2m to 4m; waves break with spray) There were no interviews with a sea state greater than Beaufort Scale 7 recorded.
Lunar illumination	The proportion of lunar illumination, calculated as the fraction of the Moon's disk that is illuminated when viewed from the Earth, on the day of the fishing trip. The lunar illumination ranges from 0 (new moon, no illumination) to 1 (full moon, full illumination), with 0.5 corresponding to half lit.	Numeric mean = 0.473, median = 0.433, minimum = 0.00005, maximum = 0.999997
Sediment	Type of seabed sediment off the coast at the location of the interview	Categorical *Sand* *Gravel* *Rock* *Mud*

2.2.3 Species

The locations of the catches of each species of interest are shown in Figure 2.1.

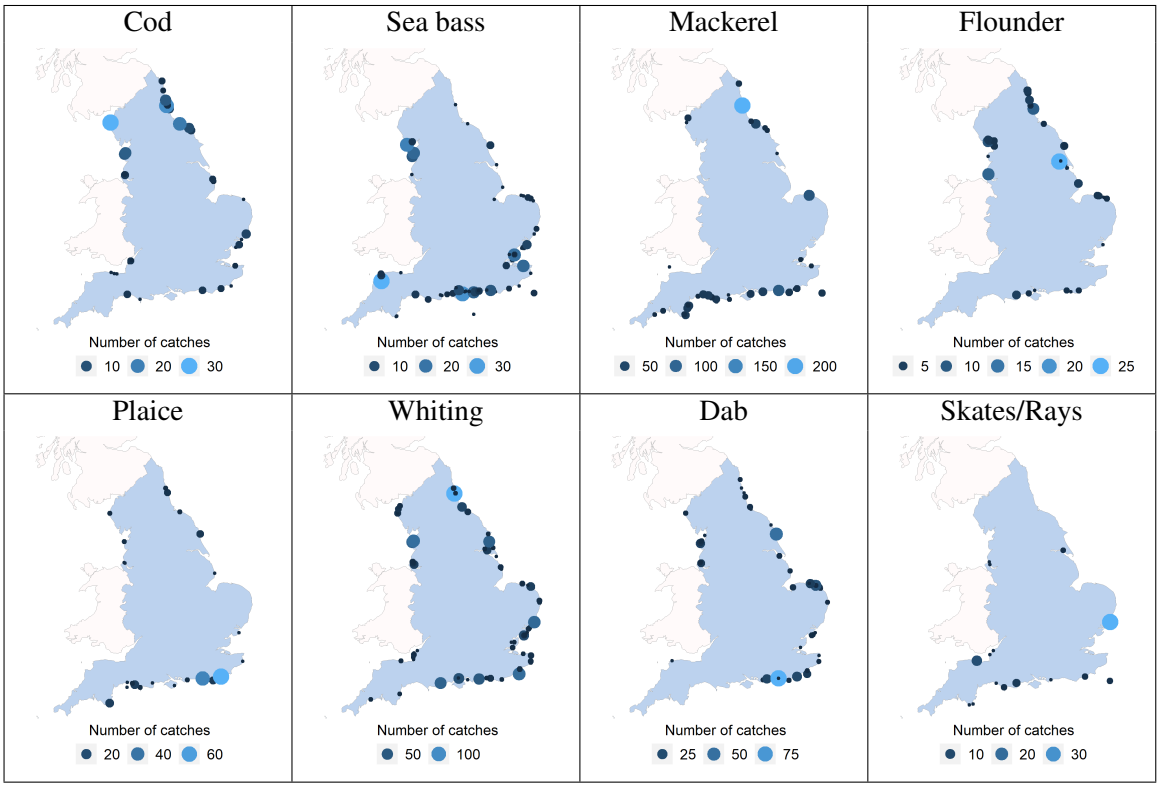


Fig. 2.1 Locations of catches, by species.

In around half of trips, anglers were targeting specific species, of which cod and sea bass were the most commonly targeted species, followed by mackerel. In the majority of the other trips, anglers were not targeting any species in particular, while in a small proportion of trips the target species was unknown or not recorded.

The proportion of trips in each sea region in which each species was targeted and caught is shown in Figures 2.2 and 2.3. A lower proportion of anglers fishing in the English Channel targeted cod compared to the proportion of anglers in other areas of the country who were targeting cod. In contrast, sea bass and mackerel were targeted by a higher proportion of anglers in the English Channel compared to the proportion in other regions. Targeting flounder and plaice was more common in the English Channel than other regions, targeting whiting was most common in the North Sea, and targeting skates/rays was most common in the Bristol Channel and Celtic Sea region. For most species,

the proportion of anglers catching the species in a region tended to follow the same patterns as the proportion of anglers targeting the species, although a lower proportion of anglers caught the species compared to the proportion targeting it. However, in some regions a higher proportion of anglers caught flounder, whiting and dab compared to the proportion of anglers targeting these species.

The proportion of trips in each season of the year in which each species was targeted and caught is shown in Figures 2.4 and 2.5. A higher proportion of anglers targeted cod and whiting in winter and autumn compared to in other seasons. In contrast, sea bass and mackerel were targeted by a higher proportion of anglers in the spring and summer. Flounder and dab were more commonly targeted in winter and autumn, plaice was more commonly targeted in winter and summer, and skates/rays were more commonly targeted in autumn, compared to in other seasons. The proportion of anglers catching each species across the seasons tended to follow the same pattern as the proportion of anglers targeting each species. A smaller proportion of anglers caught cod and sea bass, and a higher proportion caught whiting and dab, compared to the proportion targeting each of these species.

Fig. 2.2 Proportion (%) of trips in each sea region in which cod, sea bass, mackerel and flounder are targeted (left column) and caught (right column).

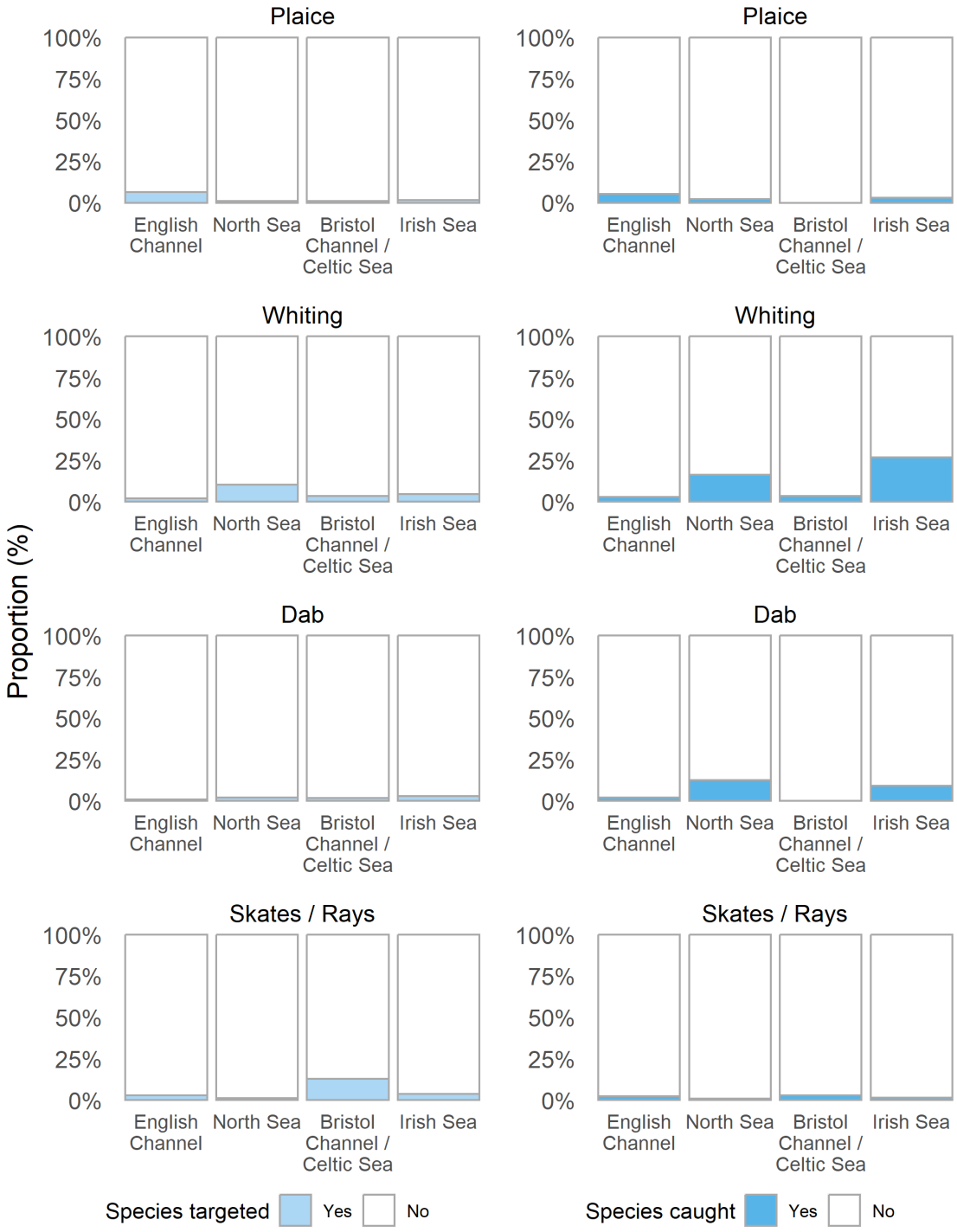


Fig. 2.3 Proportion (%) of trips in each sea region in which plaice, whiting, dab and skates/rays are targeted (left column) and caught (right column).

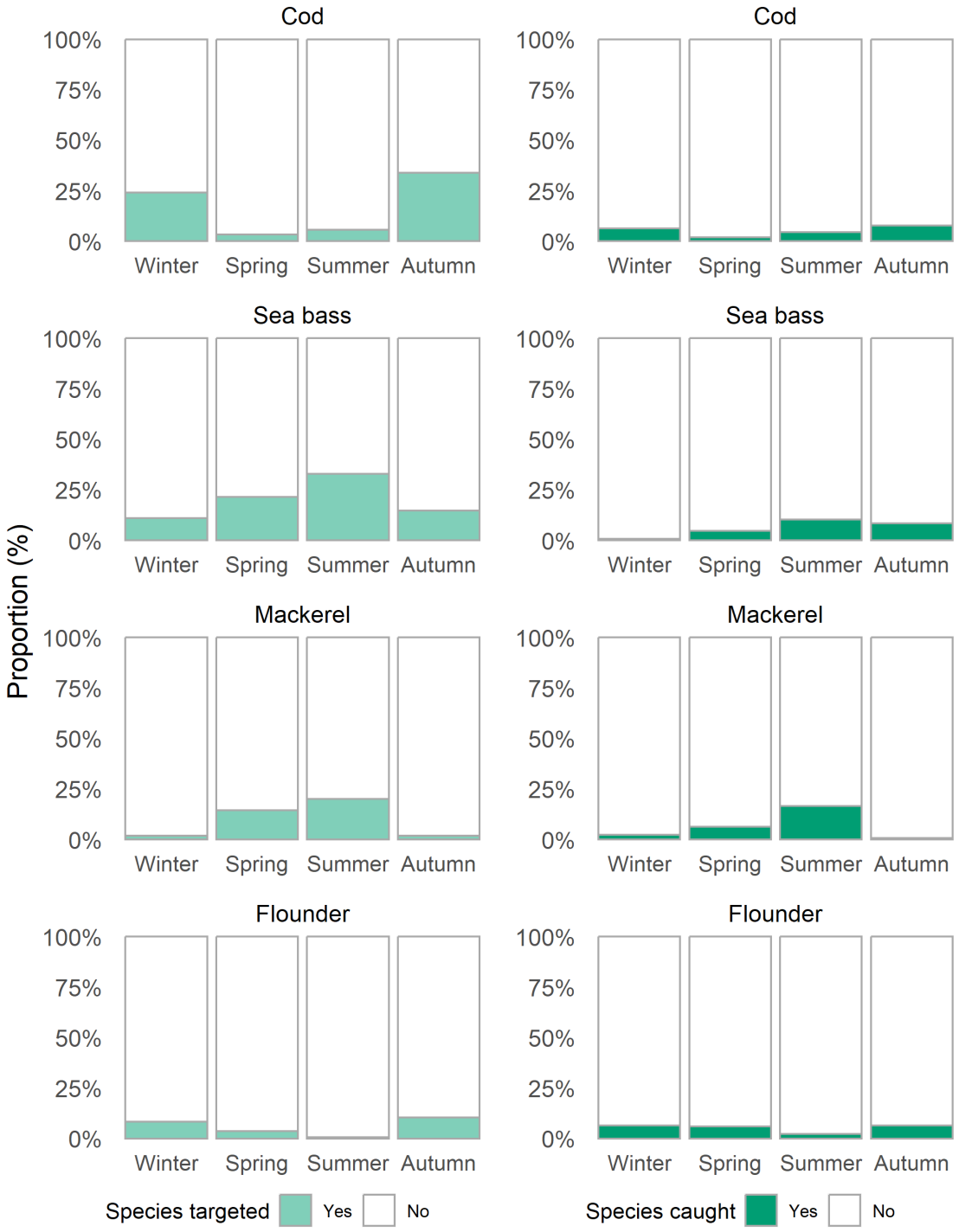


Fig. 2.4 Proportion (%) of trips in each season of the year in which cod, sea bass, mackerel and flounder are targeted (left column) and caught (right column).

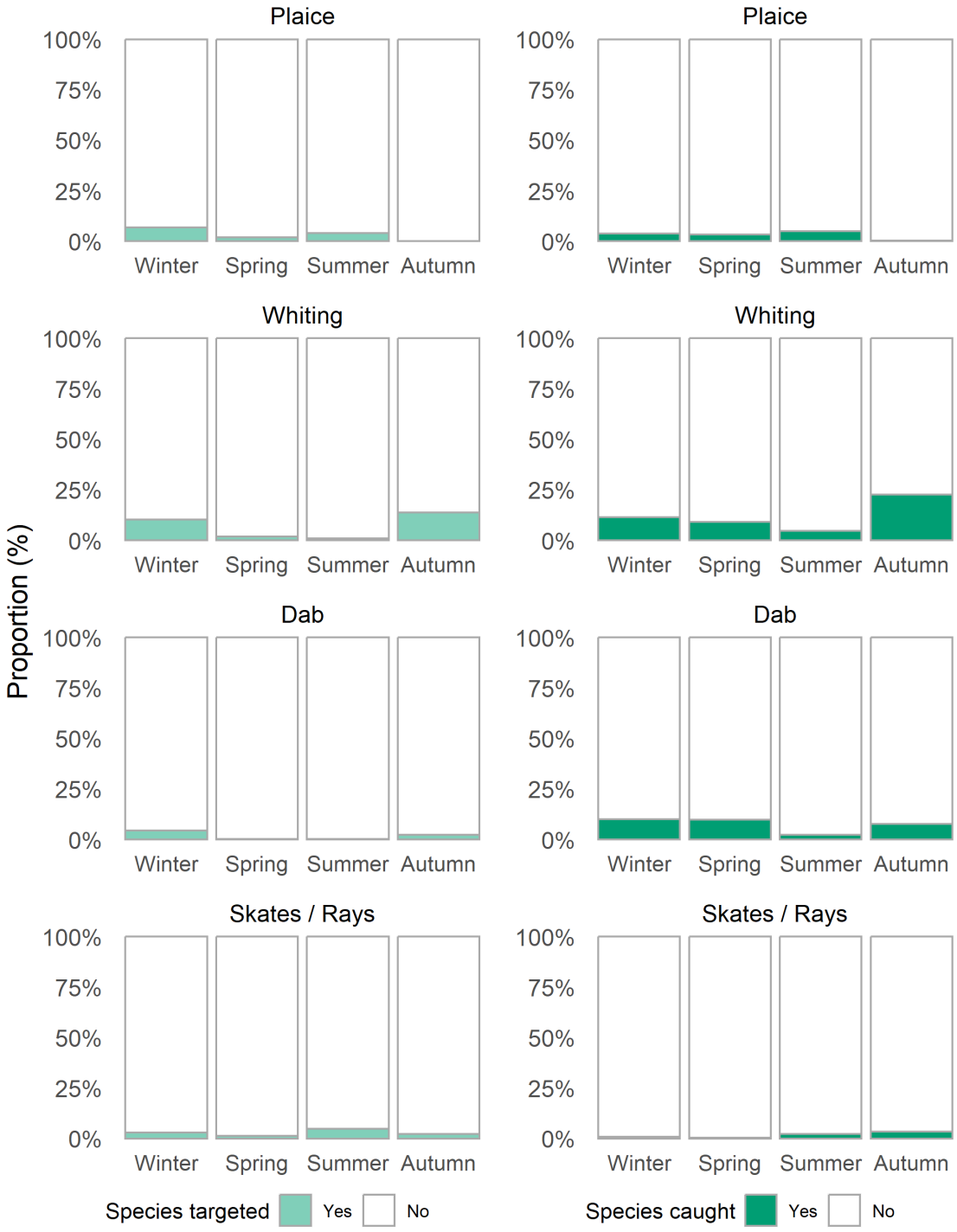


Fig. 2.5 Proportion (%) of trips in each season of the year in which plaice, whiting, dab and skates/rays are targeted (left column) and caught (right column).

2.2.4 Zero-inflated Poisson models

The distribution of the number of fish caught per trip (Table 2.4) is highly zero-inflated, with over half of fishing trips resulting in no catches. Even when fish are caught during a trip, usually the number is very low: the median number of fish caught per trip is 0, the mean is 3.3, and the upper quartile is 2. However, a very small number of trips resulted in a large number of fish caught, with the maximum number of fish recorded in one trip greater than 200.

Table 2.4 Number of fish caught per trip.

Number of fish	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15+
Frequency	760	164	95	64	39	30	23	8	13	12	9	9	5	5	4	83
Proportion	57%	12%	7%	5%	3%	2%	2%	1%	1%	1%	1%	1%	0.4%	0.4%	0.3%	6%

The distribution of the number of each species of interest caught per trip follows the same pattern as the total number of fish caught per trip (Table 2.5). Therefore, the data we model here, which correspond to the number of fish caught per trip for each species, are zero-inflated, and modelling needs to rely on models that account for this feature, such as the zero-inflated Poisson distribution, which we employ in this chapter.

Table 2.5 Number of fish caught per trip, by species.

	Number of fish																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15+	Maximum
Cod	1257	34	10	6	5	4	4	0	0	0	0	0	0	0	0	3	30
Sea bass	1238	51	13	2	3	5	3	1	1	1	0	0	1	0	1	3	19
Mackerel	1220	19	12	9	9	5	4	5	5	2	4	1	2	2	2	22	120
Flounder	1258	33	18	5	5	2	0	0	0	0	0	0	1	0	0	1	21
Plaice	1278	17	8	3	3	5	2	0	2	0	0	0	1	0	1	3	32
Whiting	1180	41	28	17	7	7	8	1	5	0	4	1	0	1	3	20	100
Dab	1232	42	19	4	4	7	1	1	2	1	1	0	3	0	1	5	79
Skates/Rays	1300	14	2	2	1	0	0	0	0	0	1	1	1	0	0	1	15

The zero-inflated Poisson model, a popular model for zero-inflated data, is a two-part process. The first part models the probability of obtaining a zero value for the response using a Binomial distribution. The second part models the value of the response, given that the first part did not yield a value of zero. The second part of the model assumes a response variable that is Poisson distributed, and therefore a zero count can be generated for the response variable in either part of the process. Hence, when modelling the number of catches per trip, zero-inflated models lead to a natural interpretation

of our data-generating process, with the first part (Binomial distribution) modelling the probability the modelled species was available to catch, and the second part (Poisson distribution) modelling the number of fish of that species caught, given that the species was available.

If Y_i denotes the random variable expressing the number of fish caught in trip i, then under the zero-inflated Poisson model, the probability of catching k fish in the i^{th} trip is

$$\begin{aligned} \mathscr{P}(Y_i = k) &= \pi_i + (1-\pi_i)e^{-\mu_i} \quad k = 0 \\ \mathscr{P}(Y_i = k) &= (1-\pi_i)\frac{{\mu_i}^k e^{-\mu_i}}{k!}, \quad k \geq 1 \end{aligned} \tag{2.2.1}$$

where π_i denotes the probability of sampling a 0 from the Binomial distribution for trip i and μ_i is the mean of the Poisson distribution for trip i. Both parameters, π and μ, can be functions of covariates, within a logistic and log-linear regression framework, respectively. These covariates do not need to be the same in both parts of the model and hence different subsets of covariates can be used in each part of the model, as different variables could influence each part of the model (Coxe et al., 2013).

We included all covariates in both parts of the model, as we believed they all have the potential to be associated with the number of fish available in angling areas and with the number of fish caught, given they are available.

2.2.5 Fitting the model

Given the large number of covariates available in this case, variable selection strategies such as stepwise regression would require a large number of models to be fitted for each species to identify the important predictors for each part of the process, i.e. the Binomial and the Poisson parts of the model. Furthermore, due to the sparseness of the data, some variable combinations may not be feasible as they can lead to convergence issues in the optimisation process. Other issues with stepwise regression include different models being selected according to the number of covariates available for selection and the order of introducing or removing covariates from the model, potential bias in the parameter estimates and potentially several different models with similar fits (Whittingham et al., 2006). In addition, standard errors can be too low, leading to over-confidence in the selected model (Smith, 2018). Although stepwise regression continues to be widely used, it has been the

focus of much criticism (Smith, 2018). Therefore, we propose instead the use of penalised regression, specifically the LASSO method, which allows us to consider all variables within a single model, with the penalty term automatically shrinking coefficients of predictors that are not important to zero, thereby simultaneously performing variable selection and coefficient estimation.

The LASSO method uses a *l*-1 penalty. This penalty is the sum of the absolute values of the coefficients scaled by the tuning parameter λ. The penalty shrinks the estimated coefficients towards zero by similar amounts, which allows some coefficients to become exactly zero. When λ is large enough, all coefficients become zero, leading to the null model with just the intercept term β_0. Therefore, the LASSO method can result in a "sparse" model containing a subset of the predictor variables, and so provides a method of variable selection, with the number of variables selected determined by the value of λ (James et al., 2017). In the zero-inflated models considered here, a separate λ parameter is needed in each of the Binomial (λ_{Bin}) and Poisson (λ_{Pois}) processes to control the shrinkage of each process independently. In general, the intercept term β_0 is not shrunk, which ensures that the mean fitted value equals the mean response. When using penalised regression, the covariates must have comparable scales, otherwise the amount by which the coefficients are shrunk would vary depending on the scales on which the variables are measured (James et al., 2017). Therefore, we standardised all numeric covariates by subtracting the mean and dividing by the standard deviation of the covariate prior to model-fitting. The data preparation and model fitting was carried out with the statistical software R (version 3.6.2) (R Core Team, 2019).

We fitted the penalised zero-inflated Poisson models using the zipath() function from the `mpath` package (Wang, 2020). This function fits zero-inflated regression models to count data using penalised maximum likelihood methods using an Expectation-Maximisation (EM) algorithm, which is based on coordinate descent methods (Wang et al., 2014).

Coordinate descent algorithms are simple, very fast, efficient and stable (Wu and Lange, 2008), and remain very fast for data with large numbers of observations and predictors (Friedman et al., 2010). The EM algorithm, which has been shown to work well with both small and large data sets, uses cross-validation to determine the values of the two LASSO λ parameters λ_{Bin} and λ_{Pois} that together give the best fitting model (Wang et al., 2014). Cross-validation splits the data set randomly into k groups, called folds, with approximately the same number of observations in each fold. Each

fold in turn is used as the validation set, with the other folds used to fit the model, and the fitted model is used to find the mean squared error for the validation set. This gives k estimates for the mean squared error, which are averaged to give the k-fold cross-validation estimate. Usually five or ten folds are used, as there is evidence that this generates estimates with a balance between bias and variance (James et al., 2017).

The zipath() function first uses all the data to determine pairs of values for λ_{Bin} and λ_{Pois}. For each pair of λ values considered, cross-validation is used to calculate the fit of the model with each of the k folds excluded in turn. The average cross-validated log-likelihood, described below, and corresponding standard deviation for this pair of λ values are then calculated as the mean from the k models. (R Package Documentation, 2021). Finally, the model is refitted with all the observed data using the pair of λ_{Bin} and λ_{Pois} values that was found to minimise the cross-validation error and maximise the cross validation log likelihood (Wang et al., 2014).

In the case of the leave-one-out cross validation that we employ here, the contribution of the ith observation to the cross validated log-likelihood is calculated by fitting the model to the other $n-1$ observations, i.e. excluding the ith observation, and calculating the log-likelihood at the ith observation using the estimated parameter for the model fitted to the $n-1$ observations. Therefore, the cross-validated log-likelihood assesses the out-of-sample fit of the model (Desmarais and Harden, 2014). If f is the probability density function for the employed model for data y then the ith contribution to the cross-validated log-likelihood is $\log(f(y_i|\widehat{\theta}_{-i}))$ where $\widehat{\theta}_{-i}$ are the estimated parameters from fitting the model to all observations except observation i.

In the case of zero-inflated Poisson models, the penalty parameters need to be chosen simultaneously. The default approach implemented in the `mpath` package (Wang, 2020) only considers a set number of pairs. The pairs of λ values used by the default method for each species are shown in Figures A.29 and A.30 in Appendix A, along with the corresponding loglikelihood values. However, in our data, we noted that even small changes in the chosen values of the penalty terms yielded considerably different estimated coefficients. Therefore, we developed an alternative approach for searching for the optimal values of the penalty terms, which is described below.

We did not rely on the default zipath() approach for identifying the optimal pair of penalty terms and instead we introduced our own extensive search approach examining the loglikelihood values

resulting from an extensive range of pairs of λ values. The process relies on setting up grids, with each cell corresponding to a pair of values for the two penalty terms. These grids get increasingly finer. In the first step, λ_{Bin} and λ_{Pois} take all possible pairs of values in $\{0.1, 0.2, ..., 0.9\}$. For each combination of these, we obtain the cross validation log-likelihood. In each subsequent step, the λ sequences "zoom in" on the values that have generated the highest cross-validated log-likelihood value in the previous step. Therefore, in the second step, the step size is 0.01, for example 0.20, 0.21, 0.22, ..., 0.30. This "zooming" in is repeated to find the pair of λ values to 3 decimal places that corresponded to the maximum cross-validated log-likelihood. We implement this by fitting multiple models with the zipath() function, using user-defined sequences for λ_{Bin} and λ_{Pois}, with each sequence containing just one value. We use leave-one-out cross-validation, in which there are the same number of folds (k) as observations in the entire data set (n), so that only a single observation is omitted in each fold, as this was found to be more stable in our simulation study (see Section 2.2.7).

2.2.6 Checking goodness of fit with parametric bootstrapping

We simulated 200 observations for each species, using the estimated Binomial probability and Poisson mean for each trip obtained from the corresponding fitted model. The proportion of observations from the modelled data that were in the interquartile range of the 200 simulated values for that observation was calculated, as an indication of how well the model fitted the data. This was done for each species modelled.

2.2.7 Simulations

We used simulations to investigate whether the model correctly identified the covariates that were related to the response variable in zero-inflated Poisson models using the grid-search approach described in the previous section.

In total, we performed 20 simulation sets, each with $n = 1000$ observations of two simulated numeric predictor variables, X1 and X2. In ten of the simulations, the values of the Binomial and Poisson distribution parameters were constant with $p = 0.5$ and $\lambda_{\text{mean}} = 6$ respectively ('scenario 1'). In the other ten simulations, the parameters of the Binomial and Poisson distributions were functions of the covariates X2 and X1 respectively ('scenario 2').

Both X1 and X2 were randomly generated by sampling with replacement from the sequence $[-3.00, -2.99, -2.98, ...2.98, 2.99, 3.00]$. Both simulated covariates were considered as potential predictors in the Binomial and Poisson parts of the model. The same set of X1 and X2 values was used for all 20 simulations.

The simulated response variable, y, was generated by sampling from the Binomial distribution with probability p. If the outcome was a success ($= 1$) then y was set to 0. If the outcome was a failure ($= 0$), then y was sampled from a Poisson distribution with mean λ_{mean}. The simulations were run with 10 folds and n folds, where n is the number of observations (leave-one-out cross validation) for comparison.

2.3 Results

2.3.1 Simulation results

Simulations showed that our approach, using leave-one-out cross validation and our extensive grid search, correctly identified the covariates that were related to the response variable in each case and for both parts of the zero-inflated model. The results are given in Table 2.6. In scenario 1, the estimated coefficients of both X1 and X2 were zero in eight of the ten simulations, and were close to zero in the remaining simulations (the maximum deviation from zero was 0.13). In scenario 2, the model correctly identified the covariate related to the simulated response y (X2 in the Binomial part and X1 in the Poisson part) in all simulations. The estimated coefficient of the covariate not related to the response in each part was either zero or close to zero in all simulations. Therefore, all covariates that impacted the response were identified by the model. Although unrelated variables were sometimes included in the final model, their coefficients tended to be close to zero and so had a low impact on the value of the response.

An explanation of how the simulations were carried out and the simulation results are given in Appendix A.

Table 2.6 Mean estimated values of coefficients in simulation results.

(a) Scenario 1 - simulating y values using constant parameters for the Binomial probability ($p = 0.5$) and Poisson mean ($\lambda_{\text{mean}} = 6$). The true values of both coefficients in the simulation are zero.

	Poisson		Binomial	
Run	X1	X2	X1	X2
1	-0.03	0	0	0.13
2	-0.03	0	0	0
3	0	-0.04	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0.10	0.12
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	-0.08	0

(b) Scenario 2 - simulating y values using the parameters of the Binomial and Poisson distributions as functions of the covariates $X2$ and $X1$ respectively. The true values for the coefficients of $X1$ and $X2$ are 0 and 2 for the Binomial part and 1 and 0 for the Poisson part respectively.

	Poisson		Binomial	
Run	X1	X2	X1	X2
1	0.99	0	-0.11	1.76
2	0.93	0	-0.42	1.63
3	0.96	-0.03	0.08	2.00
4	1.03	-0.03	0	1.84
5	0.99	-0.03	0.03	1.99
6	1.02	0	0	1.83
7	0.98	0.02	0.23	2.39
8	0.97	0	-0.17	1.72
9	0.98	-0.03	0.23	1.70
10	0.95	0	0	1.62

2.3.2 Case study results

We analysed data for the following eight species: cod, sea bass, mackerel, plaice, flounder, whiting, dab and skates/rays. For most species, the model retained almost all covariates in at least one part of the model. Seven of the 18 covariates were dropped for skates/rays, whereas for the remaining species, no more than two covariates were dropped. For most species, the environmental covariates impact the availability of fish more than they impact the catch rates.

Selecting a fitting approach

Examples of the loglikelihood grids obtained with the grid method are given in Appendix A. The λ penalty term values for the final models for each species, using the grid and default methods, are given in Table 2.7. A comparison of the loglikelihood values of the final models for each species, using the grid and default methods, is given in Table 2.8.

For all species, the grid method gave a model with a better out-of-sample fit than the default method, based on the cross-validated loglikelihood (Table 2.8). Table 2.7 demonstrates that the grid approach selects λ values that more heavily penalise the Poisson part of the model compared to the

Table 2.7 The λ values of the final model for the grid and default methods, using n folds. The λ values used in the grid method are displayed with the number of decimal places corresponding to the precision used in the calculations. The λ values used in the default method are displayed to five decimal places.

Method	Model part	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/ Rays
Grid	Poisson	0.001	0.017	0.054	0.042	0.0008	0.124	0.011	0.016
	Binomial	0.004	0.0007	0.0006	0.0004	0.0009	0.0008	0.0009	0.0012
Default	Poisson	0.00015	0.00043	0.05561	0.00591	0.00250	0.03461	0.00879	0.02219
	Binomial	0.00934	0.01535	0.02255	0.01397	0.01473	0.01861	0.01640	0.01848

Table 2.8 The log-likelihood values of the final model for the grid and default methods, using n folds.

Method	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Grid	-0.29832	-0.36292	-0.66406	-0.27356	-0.17340	-0.85278	-0.54873	-0.10151
Default	-0.30091	-0.36977	-0.71823	-0.30122	-0.21043	-0.91388	-0.58127	-0.12722

default approach, while the opposite is true for the penalty on the Binomial part. As a result, the grid method had a tendency to select more covariates in the Binomial part of the model than in the Poisson part for most species. In contrast, the default method tended to select more covariates in the Poisson part of the model than in the Binomial part. The covariates selected by the default method are given in Appendix A. The predictions for the number of fish caught obtained from the grid and default methods are shown in Figures A.27 and A.28 for the grid method, and Figures A.32 and A.33 for the default method, in Appendix A. Figures A.35 and A.36 in Appendix A compare the predictions from the two methods.

The grid search not only generated improved fits compared to the default method, but also identified different parameters as being important in both the probability of fish being available to catch (Binomial part) and the number of fish caught (Poisson part). This indicates that it can be valuable to seek a higher loglikelihood through a more extensive search of the λ parameter values, although this can take more computing time (see Appendix A for details on computational times).

The number of covariates that impacted the probability the species was available varied across the species. Most of the covariates impacted the probability of sea bass, mackerel, flounder, whiting and dab being available, while cod, plaice and skates/rays were impacted by fewer covariates. The number of covariates predicted to impact the number of fish caught, if the species was available, also varied across the species. Most of the covariates impacted the number of cod and plaice caught, while

the number of whiting and skates/rays caught was only predicted to be impacted by two covariates, and the number of flounder caught was not impacted by any of the covariates used in the model.

The estimated coefficients for all covariates considered are given in Tables 2.9 (Binomial part) and 2.10 (Poisson part). A value of zero indicates that the corresponding coefficient was shrunk to zero and therefore was not selected in that part of the model.

Probability of fish availability

The Binomial part models the probability the modelled species was available to catch in angling areas. There is variation in the factors that drive the availability of a species, which varies between species (Table 2.9).

Table 2.9 Estimated coefficients in the Binomial part (whether or not fish were available) using the grid method. A positive value indicates a higher chance of a zero, and therefore a negative association between the covariate and the availability of fish. A negative value indicates a lower chance of a zero, and therefore a positive association between the covariate and the availability of fish. A zero indicates no association between the covariate and the availability of fish. Non-zero values are displayed to two decimal places.

Grid method: Binomial part	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Intercept	2.63	4.85	4.26	3.46	2.84	3.36	3.46	5.43
Type: Boat	0	−1.09	−1.96	1.49	0	−1.02	−0.17	−1.14
Method: Lure/fly/jig	0	0.09	−1.14	2.01	1.06	0.16	1.19	0
Method: Mix	0	0	−1.55	0	0.48	1.37	0.36	0
Target: Yes	−1.32	−1.45	−2.00	−1.96	−3.00	−1.05	−2.49	−3.18
Hours fished in daylight	−0.47	−0.66	−0.34	−0.69	−0.60	−0.28	−0.46	−0.57
Hours fished in twilight	−0.14	0.35	−0.28	−0.20	0.28	0.22	0.08	0
Hours fished in darkness	−0.13	−0.26	0.14	0.07	0	−0.40	0	0.06
Number of other species caught	−0.08	0.20	−0.21	0.36	0	−0.18	−0.07	−0.05
Weekend: Yes	0	0.23	0.19	0.88	1.20	−0.02	0	0
Total anglers in group	0	0.10	−0.02	−0.19	−0.08	−0.07	−0.03	−0.11
Competition: Yes	0	0.22	0	0.28	0	0.22	0.23	0
Days shore angling	−0.11	0	0.02	−0.19	0	−0.04	0.02	0
Days boat angling	0	−0.02	0.13	−0.16	−0.07	−0.01	0.03	0
Belongs to angling club: Yes	0	−0.66	0.12	−0.26	0	−0.22	−0.35	−0.40
Sea: North Sea	0	−0.42	0.53	−1.56	0	−1.74	−1.64	0.11
Sea: Bristol Channel/Celtic Sea	0	0.83	1.16	1.91	0.94	0	0.86	0
Sea: Irish Sea	0	−0.69	0.08	−2.62	0	−2.07	−1.43	0
Season: Spring	0	−0.92	−0.65	0	−0.15	0.14	−0.39	0
Season: Summer	0	−1.21	−0.76	0.71	0	1.11	1.29	0
Season: Autumn	0	−1.45	1.22	0.21	0.94	−0.37	0.30	−0.76
Sea state: Moderate	0	0	0	0.35	−0.90	0	0	0.02
Sea state: Rough	0	−0.99	1.08	0.47	0.06	−0.49	0.20	0
Lunar illumination	0	−0.08	0.08	−0.12	0	0	0.21	0.33
Sediment: Gravel	0	0	−0.22	0	0	0.36	−0.51	0
Sediment: Rock	0	0	0.91	0.35	0	0.44	0	0
Sediment: Mud	0	0.37	0.94	0	0	0.85	0.82	0

Trip characteristics Sea bass, mackerel, whiting, dab, and skates/rays are more likely to be available in angling areas when fishing from a boat than when fishing from the shore. In contrast, flounder is more likely to be available when fishing from the shore than when fishing from a boat. Cod and plaice are equally likely to be available for both shore and boat anglers.

Mackerel is more likely, and sea bass, flounder, plaice, whiting, and dab less likely, to be available when anglers are using lure/fly/jig methods compared to when using bait. All species are more likely to be available when anglers are targeting them.

For all species, the probability of fish being available increases as the duration fished during daylight increases. As the duration fished during twilight increases, cod, mackerel, and flounder are more likely to be available, while sea bass, plaice, whiting, and dab are less likely to be available. As the duration fished during darkness increases, cod, sea bass, and whiting are more likely to be available, while mackerel, flounder, and skates/rays are less likely to be available.

As the number of fish of other species caught increases, cod, mackerel, whiting, dab, and skates/rays are more likely to be available, while sea bass and flounder are less likely to be available. The number of fish of other species caught is not predicted to be associated with whether plaice are available.

Whiting is more likely to be available in angling areas at weekends than on weekdays, while sea bass, mackerel, flounder, and plaice are less likely at weekends. Cod, dab, and skates/rays are equally likely to be available on weekdays and at weekends.

Angling group size is negatively correlated with sea bass availability. For all other species except cod, the probability that the species is available increases as the angling group size increases. Group size has no impact on the probability that cod is available.

The probability of sea bass, flounder, whiting, and dab being available is lower when fishing in a competition than when fishing is not part of a competition. There is no difference in the probability of cod, mackerel, plaice, and skates/rays being available when fishing in a competition compared to when fishing is not part of a competition.

Angler characteristics Anglers with more shore angling experience are more likely to have cod, flounder, and whiting available in the area when they are fishing, and less likely to have mackerel and

dab available. Anglers with more boat angling experience are more likely to have sea bass, flounder, plaice, and whiting available, and less likely to have mackerel and dab available.

Anglers who are members of an angling club are more likely to have sea bass, flounder, whiting, dab, and skates/rays available in areas where they fish, and less likely to have mackerel available, compared to anglers who are not members of an angling club.

Environment characteristics Sea bass, flounder, and dab are more likely to be available in angling areas in the North Sea and Irish Sea, and less likely in the Bristol Channel/Celtic Sea region compared to the English Channel. Mackerel is more likely to be available in the English Channel than in all other regions. Plaice is less likely to be available in angling areas in the Bristol Channel/Celtic Sea region and skates/rays are less likely in the North Sea, compared to in the English Channel, while whiting is more likely to be available in the North Sea and Irish Sea compared to the English Channel. There is no difference in the probability of cod being available across the different regions of the country.

Sea bass is less likely to be available in angling areas in winter than in other seasons. Mackerel is more likely in spring and summer and less likely in autumn, compared to winter. Flounder is less likely in summer and autumn compared to winter, while plaice is more likely in spring and less likely in autumn, compared to winter. Whiting is more likely to be available in autumn and winter compared to spring and summer. Dab is less likely in summer and autumn and more likely in spring, while skates/rays are more likely to be available in autumn, compared to winter.

Flounder and skates/rays are less likely, while plaice is more likely, to be available in moderate sea conditions compared to calm conditions. Sea bass and whiting are more likely, while mackerel, flounder, plaice, and dab are less likely, to be available during rough sea conditions than in calm conditions.

Sea bass and flounder are more likely to be available as lunar illumination increases. In contrast, mackerel, dab, and skates/rays are more likely to be available in lower lunar illumination conditions. Lunar illumination has no impact on the availability of cod, plaice, and whiting.

Sea bass, mackerel, whiting, and dab are less likely to be available where there is mud sediment, compared to sand sediment areas. Mackerel, flounder, and whiting are less likely to be available on

rock sediment compared to sand sediment. Whiting is less likely, while mackerel and dab are more likely, to be available in angling areas with gravel sediment compared to areas with sand sediment.

Number of fish caught

The Poisson distribution models the number of fish caught of the modelled species, given that the species is available. There is variation in the factors that impact catch rates, which varies between species (Table 2.10).

Table 2.10 Estimated coefficients in the Poisson part (how many fish of the species were caught given fish of that species were available) using the grid method. Non-zero values are displayed to two decimal places.

Grid method: Poisson part	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Intercept	−2.88	0.39	1.93	0.67	−0.84	1.74	0.77	0.51
Type: Boat	1.16	0.30	0	0	1.28	0	0.23	0
Method: Lure/fly/jig	0.14	0	0	0	0	0	0	0
Method: Mix	0.72	0	0	0	0.71	0	0	0
Target: Yes	0.74	0	0	0	0.30	0	0	0
Hours fished in daylight	0.27	0.20	0.20	0	0.57	0	0.24	0
Hours fished in twilight	0	0.03	0	0	0.19	0	0	0
Hours fished in darkness	0	−0.01	0	0	0	0	0	0
Number of other species caught	−0.06	−0.06	0.06	0	−0.03	0.16	0.24	0.31
Weekend: Yes	0.63	0	0	0	1.14	0	0	0
Total anglers in group	0	−0.33	0.21	0	0.04	0	0	−0.20
Competition: Yes	0.22	0	0	0	0	0	0	0
Days shore angling	−0.03	0	0	0	0.12	0	−0.13	0
Days boat angling	−0.27	0.16	0	0	−0.44	0.19	0.03	0
Belongs to angling club: Yes	−0.05	0	−0.04	0	0.91	0	0.74	0
Sea: North Sea	1.93	0	0	0	−0.65	0	0	0
Sea: Bristol Channel/Celtic Sea	1.59	0	0	0	0	0	0	0
Sea: Irish Sea	2.78	0	0	0	0	0	0	0
Season: Spring	−0.29	0	0	0	0.32	0	−0.21	0
Season: Summer	0.44	0	0	0	0.59	0	0	0
Season: Autumn	−0.17	0	0	0	0	0	0	0
Sea state: Moderate	0.04	0	0	0	−1.76	0	−0.92	0
Sea state: Rough	0	0	0	0	0	0	0	0
Lunar illumination	0	0	0.19	0	0.09	0	0.41	0
Sediment: Gravel	−0.81	0	0	0	0.43	0	0.49	0
Sediment: Rock	0	0	0	0	−1.06	0	0	0
Sediment: Mud	0.72	0	0	0	0	0	0	0

Trip characteristics When the species is available, the number of cod, sea bass, plaice and dab caught per angler is higher from a boat than from shore, but there is no difference for the other species. Using a lure or mix of bait and lures results in higher catches of cod, while a mix of methods increases catches of plaice. Targeting increases the number of cod and plaice caught, but has no influence on catches of other species. As the duration fished during daylight increases, more cod, sea bass,

mackerel, plaice and dab are caught, more sea bass and place in twilight, and fewer sea bass at night, if these species are available. Fewer cod, sea bass and plaice, and more mackerel, whiting, dab and skates/rays, are caught as the number of fish of other species caught increases. Higher numbers of cod and plaice are caught per angler at weekends than on weekdays. As group size increases, the number of mackerel and plaice caught per angler increases, while the number of sea bass and skates/rays decreases. Fishing in a competition results in a higher number of cod caught per angler, compared to not fishing in a competition.

Angler characteristics As the number of days shore angling experience increases, the number of plaice caught increases, while the number of cod and dab decrease, if these species are available. More sea bass, whiting and dab, and fewer cod and plaice, are caught by anglers with more boat angling experience, compared to anglers with less boat angling experience. Fewer cod and mackerel, and more plaice and dab, are caught by club members than by anglers who are not members of an angling club.

Environment characteristics More cod are caught in the North Sea, Bristol Channel/Celtic Sea and Irish Sea than the English Channel, while fewer plaice are caught in the North Sea than English Channel. More cod are caught in summer and winter than in spring and autumn, and more plaice are caught in spring and summer than winter. Fewer dab are caught in spring than in winter. More cod and fewer plaice and dab are caught during moderate than calm sea conditions. As lunar illumination increases, more mackerel, plaice and dab are predicted to be caught, if these species are available. Fewer cod and more plaice and dab are caught when angling over gravel than sand. Fewer plaice are caught in rocky areas, and more cod in mud sediment areas, than over sand.

Summary of case study results

For most species, the model retained almost all covariates in at least one part of the model. Seven of the 18 covariates were dropped for skates/rays, whereas for the remaining species, no more than two covariates were dropped. For most species, the environmental covariates impact the availability of fish more than they impact the catch rates.

Availability of fish All species are more likely to be available in boat angling areas, except flounder which is more likely by the shore, and cod and plaice which are equally likely in shore and boat angling areas. For all species, there is a high probability of the species being available if that species is being targeted, indicating that anglers tend to know which species are likely to be available and target them accordingly. As the number of hours fished during daylight increases, the probability of all species being available increases. Catching more fish of other species is associated with a higher probability of catching all species if they are available, except for sea bass, flounder, and plaice. Four of the species (sea bass, mackerel, flounder, plaice) are less likely to be available in angling areas at weekends compared to on weekdays, while one (whiting) is more likely to be available, with no difference for cod, dab, and skates/rays. Four of the species are less likely to be available during competitions, with no difference in the probability for the other species.

All species have at least one covariate relating to angler characteristics associated with the probability fish are available. Where multiple covariates relating to angler characteristics are included in the model for a species, these covariates are consistently associated with either a higher or lower chance of the species being in the angling area, for all species except dab.

The area of the country and season in which each species is more likely to be available varies by species. The sea conditions and sediment type also impact each species differently. Lunar illumination impacts five of the eight species, although not all in the same way.

Catch rates Catch rates of cod, sea bass, plaice, and dab are higher when fishing from boats than from shore, if these species are available in these areas. The fishing method has no impact on catch rates for most species, though cod is a notable exception with higher catch rates of cod associated with lure/fly/jig methods compared to bait. Targeting cod and plaice increases the number of those species caught, while for other species, targeting the species has no impact on the number of fish of that species caught. Catch rates of cod and plaice are higher at weekends compared to on weekdays. Fishing in a competition is not predicted to impact the number of fish caught per angler for any species except cod, with higher catch rates during competitive fishing than during non-competitive fishing.

Most species have at least one covariate relating to angler characteristics associated with catch rates. Environmental covariates do not impact catch rates of sea bass, flounder, whiting, and skates/rays.

The only environmental covariate impacting mackerel catch rates is lunar illumination, with higher illumination associated with higher catch rates. In contrast, most environmental covariates impact catch rates of cod, plaice, and dab, if they are available.

Parametric bootstrapping

For each species modelled, the proportion of observations that were in the interquartile range of the simulated values is shown in Table 2.11. The proportion of observations in the interquartile range was lowest for whiting (92%) and highest for skates/rays (99%). The proportions for the other species were between 95 and 97%.

Table 2.11 Proportion of observations within the interquartile range of the simulated values.

Species	Proportion (%) in IQR
Cod	96
Sea bass	95
Mackerel	95
Flounder	95
Plaice	97
Whiting	92
Dab	95
Skates/Rays	99

2.4 Discussion

Our aim was to identify the important predictors of the number of fish caught by marine recreational anglers. We used penalised regression, with the Least Absolute Selection and Shrinkage Operator (LASSO) penalty, to fit zero-inflated Poisson models for eight commonly-caught species.

2.4.1 Modelling sea angling catches

We have considered a large number of possible predictors for both parts of our zero-inflated Poisson model, some of which are correlated. However, as suggested by Hebiri and Lederer (2013) and by Dalalyan et al. (2017), the LASSO can still yield a reliable predictive model in this case. Where predictors are highly correlated, the LASSO will tend to select one at random in favour of the rest

(Friedman et al., 2010), although to overcome bias in the LASSO model, small λ values are required, which can lead to models containing "many false positive variables" (Fan and Lv, 2010, p. 8).

This method enabled fitting multiple models for a large number of possible combinations of variables, while allowing a realistic error distribution to describe the data. Simulation results showed that important effects were recovered. Our extensive approach to identify the λ values corresponding to the higher loglikelihoods increased run times, but gave a smaller cross-validation error than the default approach for all eight species modelled.

We used leave-one-out cross-validation (LOOCV), which is time consuming to fit as it requires the same number of models as data observations (n) (James et al., 2017). For our extensive search method, due to using only one pair of λ values to fit the model each time, it is necessary to use n folds as fewer than n folds results in substantially different loglikelihood values across the grids. Our extensive search was compared with the default method, and for this it was necessary to also used n folds, in order for the comparison between the loglikelihoods of the models to be valid. (James et al., 2017)

Although the loglikelihood values in the grids increased smoothly at the first and second zoom levels, this is not always the case at the third zoom level. It is therefore possible, or even likely, that the highest loglikelihood found by the grid method is not the best loglikelihood. It is also possible that higher loglikelihoods can be obtained by zooming in further, taking the λ values to a greater number of decimal places. Therefore, our method may not result in a better model than the default approach, especially if there are multiple regions with a high loglikelihood. For cod, the initial grid method process did not find a better model than obtained using the default method, and additional searching was required to improve on the default loglikelihood.

In our application, we found different coefficients were generated from the two approaches, despite similar loglikelihoods obtained from each method. This could be a consequence of the two parts of the zero-inflated model. Low counts can be generated by either a low probability of availability with a higher mean catch rate or a higher probability of availability with a low mean catch rate. Alternatively, the likelihood surface may be fairly flat, and hence different coefficient values can lead to similar likelihood values.

The grid method tended to select covariates in the Binomial part, while the default method tended to select covariates in the Poisson part. As a result, the two methods give different interpretations to how catches arise. The grid method tended to indicate that the covariates are associated with the probability that fish would be available, whereas in the default method results, the covariates tend to be associated with the number of fish caught, given that they were available.

Penalised regression in general is fast and efficient, but in this case we have found that the choice of the penalty terms is not straightforward, which increases the computational burden.

Neither the grid method nor the default method predicted the highest numbers of catches observed for cod, sea bass, mackerel, flounder, plaice and whiting. For dab and skates/rays, both methods predicted a substantially higher number of catches for one observation than the maximum number of catches observed of these species. The prediction plots in Appendix A illustrate these findings. The total number of catches predicted by the grid method was reasonably close to the total number of observed catches, and closer than the total predicted by the default method, for cod, sea bass, mackerel, flounder and whiting. The total number of dab and skates/rays predicted by the default method was over twice the observed catches, as a result of the extremely high prediction for one observation. Due to this, together with the tendency for the default method to underestimate the catches, for these two species the total catches predicted by the default method was closer to the total observed catches than the grid method predicted total. The total number of observed and predicted catches are in Table A.13 in Appendix A.

Limitations with the data include angler characteristics (number of days angling experience and club membership) only recorded for one angler in angling groups despite the catches of all anglers in the group being recorded. Public holidays falling on weekdays would have been better classed as 'weekend', rather than 'weekday' days. It is assumed that the catch rate is constant through each fishing trip.

Charter boat surveys were excluded from our research, due to differences in trip characteristics and catch rates, compared to shore and private boat angling (Armstrong et al., 2013). Consequently, our findings may not provide a complete picture of the variables affecting catch rates across all types of recreational fishing activities. Future research should consider including charter boat surveys to

provide a more comprehensive understanding of the factors influencing catch rates in recreational fishing.

Future work could compare the `mpath` package's (Wang, 2020) default method with different input parameters with results obtained using an extensive search similar to that used in this research, and different penalties could be considered instead of the LASSO. Ridge regression, like the LASSO, is a commonly used penalised regression method (James et al., 2017). In contrast to the l-1 penalty of the LASSO method, ridge regression uses a l-2 penalty, which is the sum of the squared coefficients scaled by the tuning parameter λ. This penalty scales all estimated coefficients towards zero by similar proportions, so they approach zero but do not become exactly zero. All covariates remain in the model, with the magnitude of their corresponding effect determined by the value of λ (James et al., 2017). Therefore, unlike the LASSO, ridge regression does not lead to sparse models, and so we did not consider it as it does not provide a method of variable selection. The elastic net is an additional method, based on ridge regression and LASSO (Zou and Hastie, 2005). Although it does enable variable selection, the elastic net method requires two penalty parameters for each part of the zero-inflated model. Given the complexity of our model, the sparseness of our data, and the fact that we need to specify two penalty terms, one for each part of the zero-inflated model, we decided in favour of the LASSO, as this was the simplest penalised regression method that enabled variable selection for both parts of the zero-inflated models considered in this research. Other potential penalty functions include the less-commonly used minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) (Wang et al., 2014). In addition, zero-inflated negative Binomial models, which are able to accommodate more overdispersion than zero-inflated Poisson models (Wang et al., 2014), could be investigated. Additional covariates such as water and weather data, including temperatures, tides and sunshine could be included in future studies. Due to the difficulties in obtaining this data in retrospect for all catch locations, this data should be collected at the time of the surveys. Furthermore, the data only include angler characteristics for the 'primary respondent' and not for other group members. In future surveys, this information should be collected for each angler in an angling group. Finally, models could be created for shore anglers and boat anglers separately, and be weighted based on the sampling probabilities for area and activity that were used within survey design.

2.4.2 Drivers of sea angling catches

Drivers of sea angling catches varied among species in both the Binomial and Poisson parts of the model. These varied in terms of characteristics related to the trip, angler, and environment. This variation is expected as, for example, species are associated with different habitats, are found different distances from the coast, and are found at different times of the year. There are few studies of factors driving sea angling (Folpp and Lowry, 2006; Heermann et al., 2013; Pulver, 2017), and only one (ongoing) study related to UK catch rates by anglers in the UK (Hyder et al., 2020a, 2021, 2024). Bayesian multilevel regression and post-stratification were used to model effort and catches by individual sea anglers in the UK and showed that avidity, age, gender, and region were important predictors of catch by anglers (Hyder et al., 2021, 2024). However, this research focused on per angler catches across all species rather than trip catches of individual species, so comparisons are difficult, although it showed that type of fishing is important.

The results can be compared with angler knowledge of where, when and how the case study species are commonly caught taken from BritishSeaFishing.co.uk. Generally, the spatial distributions and the seasonality of catches of the different species provided as advice for anglers agree with the predictions from the model, suggesting that the correct drivers are being identified. For example, the model predicts that the seasons do not impact the availability of cod, but anglers catch more cod in summer, if available. This may be because the number of anglers fishing from boats increases in the summer. Cod occur around UK coasts throughout the year, though the larger individuals migrate to cooler regions during warmer months, so cod are more common in autumn and winter. In summer, cod prefer regions further from the shore, so are more likely to be caught by boat anglers than shore anglers. However, it would be useful to compare this with additional data sets and trip level models based on sea angling diaries from the UK (Hyder et al., 2020a, 2021, 2024).

2.4.3 Conclusion

The aim of this study was to identify important predictors of catch rates for commonly-caught species by marine recreational anglers. We used LASSO penalised regression due to the large number of covariates considered. The models were fitted using the `mpath` package in R, with a grid search

space instead of the package's default option. Using a statistically designed survey conducted through face-to-face interviews with shore and boat anglers around the coast of England in 2012-13, we have demonstrated that marine recreational angling catch rates can be effectively modelled using LASSO penalised regression. These models enabled factors influencing catch rates to be identified. Our results identified covariates associated with a higher likelihood of fish being available and with higher catch rates. These showed that fish are more likely to be available if they are being targeted, indicating that anglers tend to know what species may be available and target these. Higher catch rates for each species were also associated with the species being targeted. Increased catch rates for targeted species is consistent with results of other studies (Pope et al., 2016).

Chapter 3

Spatiotemporal models for mobile app records of fish catches

Abstract

Marine recreational fisheries (MRF) have social and economic benefits, but can impact on fish populations and the environment. Therefore, data on these fisheries are vital to inform their monitoring and management. Traditional onsite surveys of MRF are often time-consuming and difficult to carry out over large areas, so novel sources of information are needed to improve understanding of spatiotemporal variation in catches. Angling apps provide a new source of this information, but the utility of these data to support management of MRF is not well understood. Fishbrain is a widely-used social networking app in which anglers can record their catches and associated information. In this chapter, we present a novel modelling approach for catches, fitting spatiotemporal models to catches recorded in Fishbrain, for four species commonly targeted by marine recreational anglers: European sea bass (*Dicentrarchus labrax*), Atlantic cod (*Gadus morhua*), Atlantic mackerel (*Scomber scombrus*) and whiting (*Merlangius merlangus*). The models are fitted using integrated nested Laplace approximations (INLA) via the `inlabru` package in R and incorporate seasonal effects and the estimated number of app users. The analysed data provide information on 5,893 catches across the four species from 1,349 coastal locations around the United Kingdom and Ireland, between the years 2018 to 2021. The model predictions show seasonal effects and clear spatial patterns of catches recorded by Fishbrain app users. These spatial patterns are consistent across the years studied and show regional and seasonal differences between species. The results support the case for development of methods to combine opportunistic app data and structured sampling, and further investigation of the propensity to use the app in order to convert recorded catches to overall catch rates.

3.1 Introduction

Marine recreational fishing can have a significant adverse impact on fish stocks and the sustainability of species (Cooke and Cowx, 2004, 2006; Hyder et al., 2018; Radford et al., 2018) and fishing activities can put marine habitats and ecosystems at risk (Lewin et al., 2006, 2019; Lloret et al., 2018). Data on marine recreational catches are vital to inform monitoring and management of fisheries, including on the number of people who engage in recreational fishing and their fishing patterns over space and time (Dainys et al., 2022a).

Traditionally, data are collected through onsite surveys (such as creel and aerial surveys), recall surveys (such as post or phone surveys) and angler diaries. Onsite surveys are often time-consuming, expensive, difficult to carry out over large spatial areas and can require specialist equipment and human resources (Skov et al., 2021), while recall surveys have low response rates and potential bias (Dainys et al., 2022a). Technology is enabling developments which can be a more cost effective approach (Dutterer et al., 2020; Hartill et al., 2020), including onsite cameras (Hartill et al., 2020), (Dutterer et al., 2020), drones (Dainys et al., 2022a) and smartphone applications (Venturelli et al., 2017) (hereafter referred to as 'apps').

An important aspect of data collection is that recreational fishing effort and locations fished can vary over short time scales due to socioeconomic factors, weather conditions and changes in the presence and abundance of fish (Hartill et al., 2020). These differences cannot be identified through surveys that are intermittent in space or time, but over time small differences can become significant, so knowledge of these is important for accurate monitoring purposes (Hartill et al., 2020). Citizen science projects, in which anyone can contribute records of their observations, enable data collection over large and continuous temporal and spatial scales, as participants can make observations at any time and from any location, as for example in the eBird project (Sullivan et al., 2014). These projects are becoming more popular and the cumulative number of citizen science projects in ecology and environmental topics has increased exponentially over the past 20 years (Pocock et al., 2017).

With the widespread use of smartphones, many apps that collect recreational catch data have been developed (Fishbrain, 2023; Gundelund et al., 2021; Jiorle et al., 2016; Noleto-Filho et al., 2021; Papenfuss et al., 2015), and a large amount of information on recreational catches is uploaded to apps

daily (Venturelli et al., 2017). Apps that provide an opportunity for anglers to record their catches could be a useful source of inexpensive catch information (Gundelund et al., 2020; Venturelli et al., 2017) to support fisheries management.

Fishing apps that collect catch records fall into two main categories: citizen science fishing apps created specifically for scientific research, such as Fangstjournalen (Gundelund et al., 2020, 2021) and iAngler (Jiorle et al., 2016), and social media apps that provide recreational anglers with tools such as social networking, maps, weather and water conditions, fishing tips, and personal catch diaries, such as MyCatch (Johnston et al., 2022) and Fishbrain (Fishbrain, 2023). However, the number of species covered by many fishing apps is limited (Noleto-Filho et al., 2021) as well as being restricted to particular regions, for example iAngler is focused on Florida, United States.

In contrast, Fishbrain users can record catches around the world. According to its website, in 2022, Fishbrain was the most popular fishing app worldwide, with 14 million users, who had recorded catches of 2,800 different species in 1.7 million locations (Fishbrain, 2022). In contrast to surveys, catches recorded in Fishbrain are not limited to a fixed time period or area as users can record catches made anytime and anywhere, potentially providing a large source of data unlimited in time and area, and valuable insights into recreational catches. However, users of fishing apps are unlikely to be representative of all fishers (Gundelund et al., 2020), and additionally can choose whether or not to record each catch they make. Despite the limitations and biases in opportunistic citizen science data, they can be valuable for improving understanding of ecological spatiotemporal process (Johnston et al., 2021), and data voluntarily recorded in fishing apps could be a useful and inexpensive supplement to data obtained from traditional methods for monitoring and management of fisheries (Papenfuss et al., 2015; Venturelli et al., 2017). Fishbrain users can choose to record any catch, and its associated information, including the location and date of the catch, the species caught and fishing method. The app also records weather and water conditions for the selected location and date, where these are available. This gives rise to a spatiotemporal data set of catch records for each species.

Spatiotemporal models (Blangiardo and Cameletti, 2015) provide the foundation for monitoring patterns and trends over space and time, and hence are appropriate for the Fishbrain app data that consist of catch records around the whole coastline and across several years. In recent years, fitting of such complex and large spatiotemporal models routinely relies on integrated nested Laplace

approximations (INLA), which provide an efficient approach for estimating the parameters in latent Gaussian models (Calculli et al., 2019; Sani et al., 2022; Zhang et al., 2023). INLA has now been established as the go-to tool for spatiotemporal modelling of large datasets (Lindgren and Rue, 2015; Niekerk et al., 2023), and in this chapter we employ this approach for fitting spatiotemporal models to catches recorded in Fishbrain of four marine species commonly-caught around the coast of the United Kingdom (UK) and Ireland: European sea bass (*Dicentrarchus labrax*), Atlantic cod (*Gadus morhua*), Atlantic mackerel (*Scomber scombrus*), and whiting (*Merlangius merlangus*) (hereafter referred to as sea bass, cod, mackerel, and whiting). Our analysis is the first to consider this type of modelling approach for data collected from a mobile app and aims to investigate whether spatiotemporal models fitted to catches recorded in an angling app can provide useful insights into marine recreational fishing. This approach provides a novel modelling framework for opportunistic records of fish catches, which cover a wider time period and finer spatial scale than data from traditional survey methods.

Section 3.2 introduces the data and describes the models, while Section 3.3 presents the results. This is followed by a Discussion in Section 3.4.

3.2 Materials and methods

3.2.1 Data

The data consist of catches recorded by users of the Fishbrain app available at www.fishbrain.com (Fishbrain, 2023), an app providing fishing information, social networking opportunities and the ability to record catches and associated information. The anonymised data were provided by Fishbrain for research purposes under a data-sharing agreement and contain all catches recorded up to 31 December 2021. Users of the app can choose to upload information about a specific catch, which corresponds to a single fish caught. Data associated with each catch record include location coordinates, date and time, species and anonymised user ID, along with autonomously collected data on weather and water conditions. As the app is user-focused, this information is only on the catches recorded by users. When recording catches, users of the Fishbrain app can select the location from a list of nearby locations and/or use the device's current location, which is determined by GPS coordinates (entering the location is optional). For this analysis, we used only those catches for which the location had

been selected by the user from the list of locations available in the app. We considered these to be more reliable than the device's current location at the time of recording the catch as users may delay recording catches until they are in a different location to where the catch occurred.

Separate models are fitted for each species of interest: sea bass, cod, mackerel and whiting. For each of these species, the data used are catches of the species recorded as occurring between 1 January 2018 and 31 December 2021, inclusive, at locations within 12 nautical miles of the coast ("territorial waters") of the UK and Ireland. Catch records for which the location or species caught were not recorded are excluded. The years prior to 2018 are excluded due to the low numbers of catches recorded during these years and the resulting sparse spatiotemporal data set before Fishbrain increased in popularity. Catches at locations over 12 nautical miles from the coast are excluded as recreational fishing generally occurs within 12 nautical miles of the coast. Catches recorded at inland locations more than 1,000 metres from the coast are excluded from the analysis, as in these cases it is assumed that either the species or location are incorrect. All data preparation was carried out in R (version 4.1.1) (R Core Team, 2021). Maps showing the locations of the reported catches for each species, by year and quarter, are given in Appendix B (Figures B.4 to B.7).

We model the total number of catches of the species of interest recorded at each location in each time period, which is defined as we explain below. We split the four years (2018-2021) into 16 time periods of three-month length, as is common practice in fisheries analyses. We also define quarter 1 (Q1) to correspond to January-March (winter) each year; quarter 2 (Q2) to April-June (spring); quarter 3 (Q3) to July-September (summer); and quarter 4 (Q4) to October-December (autumn), so that time periods $\{1, 5, 9, 13\}$ correspond to Q1, time periods $\{2, 6, 10, 14\}$ to Q2, time periods $\{3, 7, 11, 15\}$ to Q3 and time periods $\{4, 8, 12, 16\}$ to Q4. The data therefore consist of the number of catches, $y_{\ell t}$, recorded at location ℓ, $\ell = 1, \ldots, L$, and time period t, with t = 1 (Q1 2018) to 16 (Q4 2021), and the corresponding quarter. There are $L = 1,349$ locations, resulting in 21,584 unique location-quarter combinations. Locations with no catches of the species of interest recorded in any time period are also included in the model (as zeros). The data are highly zero-inflated, with no catches recorded in 95% of the location-time period combinations for sea bass, 97% for cod, 96% for mackerel and 98% for whiting (see Table 3.1).

Within each species, the spatial pattern in the recorded catches is similar across all years.

Table 3.1 Frequency table of the number of catches recorded in each quarter at each location and for each species. The total number of location-quarter combinations is 21,584.

Number recorded	Sea bass	Cod	Mackerel	Whiting
0	20424	21038	20800	21190
1	717	411	508	278
2	217	72	154	57
3	102	23	58	24
4	41	11	21	12
5	26	9	19	11
6	11	5	6	3
7	9	5	6	2
8	7	1	4	3
9	4	1	2	0
10	4	2	1	3
>10	22	6	5	1
Maximum	312	23	18	19

The rate of recorded catches for each species in each quarter in 2021 at each location is shown in Figure 3.1, where the rate of recorded catches at a location in a quarter is the number of catches recorded at that location in that quarter divided by the lower bound on the number of app users (which we define in Section 3.2.2). The corresponding plots for each species over the four-year period 2018-2021 are given in Figures B.4 to B.7 in Appendix B.

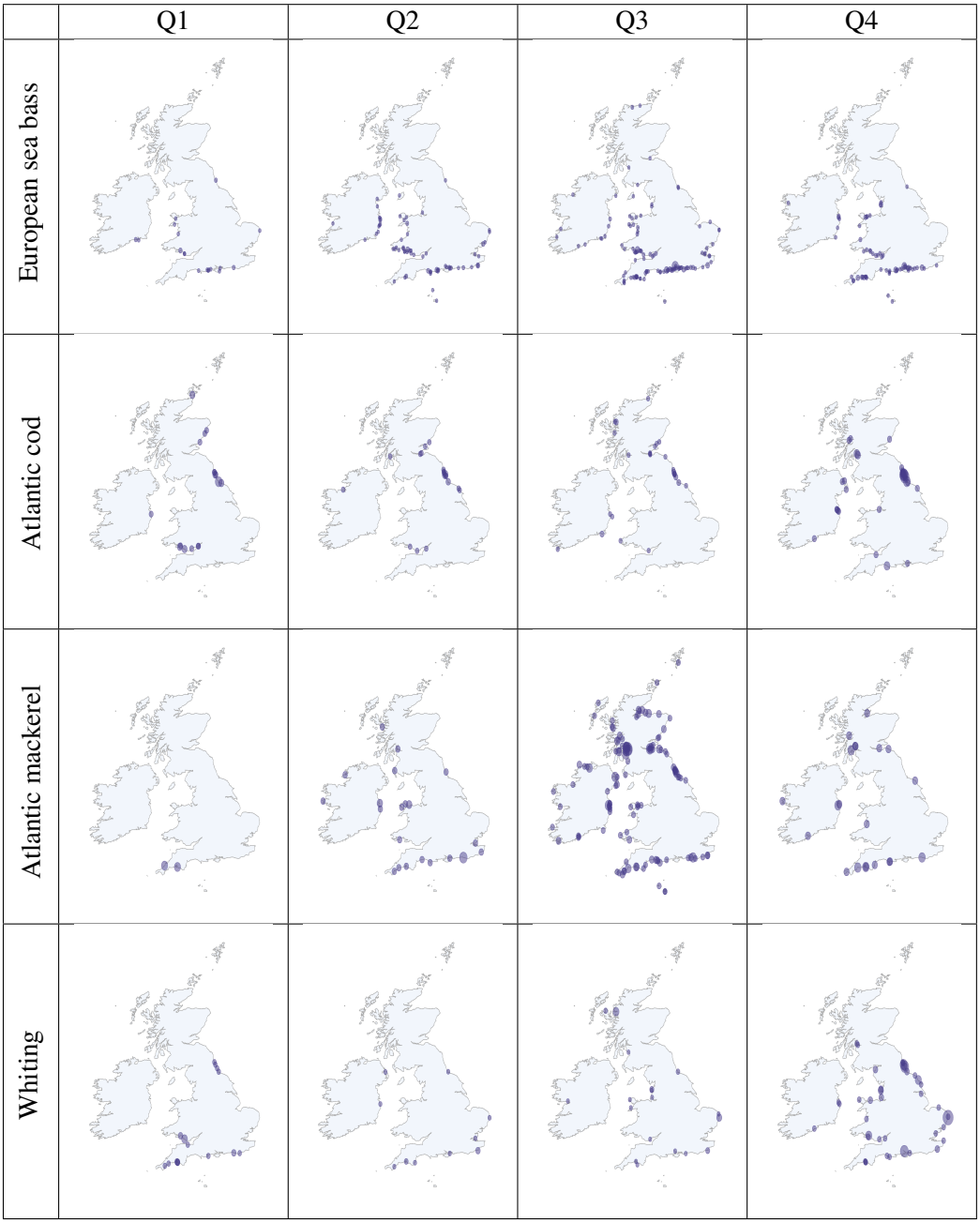


Fig. 3.1 Rate of catches reported of sea bass, cod, mackerel and whiting for each quarter in 2021 at the locations in the Fishbrain app. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December).

3.2.2 Modelling

Creating a mesh

We obtained world coastlines from the Natural Earth website (Natural Earth, 2021) and used the inla.mesh.2d() function from the R package `INLA` (Rue, Martino, and Chopin, 2009) to create a mesh covering the UK and Ireland, and surrounding marine areas. The extent of the marine areas covered by the mesh is determined automatically by the inla.mesh.2d() function. The mesh consists of triangular tiles and is used to approximate continuous space (Bachl et al., 2019). The minimum side length of the triangles in the mesh was set to 40 kilometres. This was a trade-off between the minimum values possible that we identified that could generate a mesh and generating a mesh that enabled the model to run successfully. The maximum side length was unrestricted, by using a high value (500 kilometres) for the 'max.edge' parameter in the inla.mesh.2d() function, to allow large triangles over the areas that are not adjacent to the coast as these areas are not of interest (see Figure 3.2).

Number of users

The number of Fishbrain users each time period, or a proxy for the number of users, needs to be considered as an offset in our model to account for increased exposure to catches with a larger number of users. The number of users is not known, but we can obtain a lower bound for the number of users in each time period, m_t, as the sum of the number of users reporting a catch at time t, n_t, and the number of users reporting a catch before and after time t, but not during time period t, z_t, so that $m_t = n_t + z_t$. This simply assumes that individual users keep the app, and hence are exposed to recording catches, during the time period without a catch recorded by these individual users. Values for $m_t, t = 1, \ldots, 16$ were obtained with the JS.counts() function from the R package `openCR` (Efford, 2022). The value of z_t for $t = 15$ and $t = 16$ given by the JS.counts() function, that is the number of known active users who did not report a catch at those times, is lower than expected, because of the truncation at $t = 16$, which means that users do not have the opportunity to report a catch after that time period, even if they are still active. Hence, we approximate z_{15} and z_{16} by $z_{15} = \frac{(z_3+z_7+z_{11})}{3}$ and $z_{16} = \frac{(z_4+z_8+z_{12})}{3}$, that is we set them equal to the mean values of corresponding zs in the same quarter

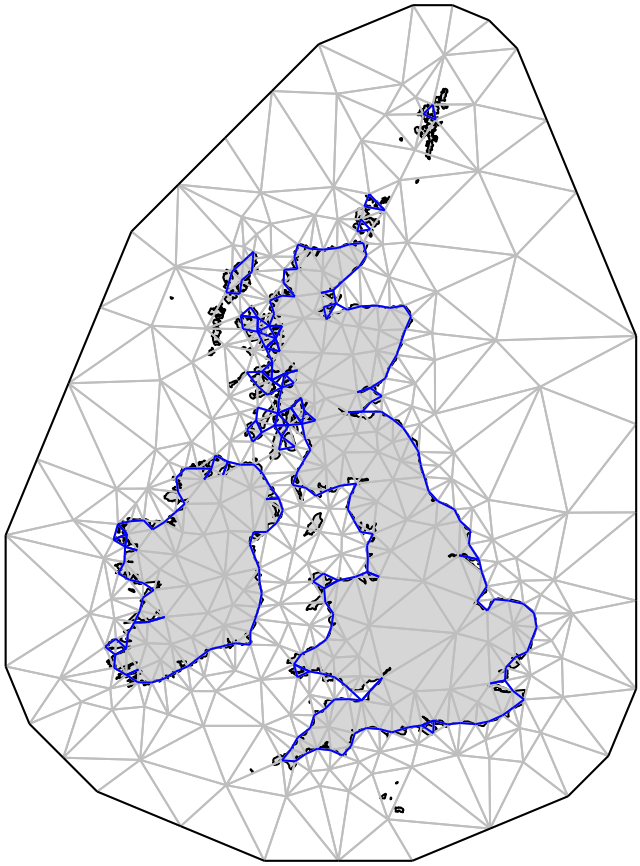


Fig. 3.2 INLA mesh used to fit the models. The mesh covers the UK and Ireland, and surrounding marine regions.

in each case, instead of z given by JS.counts(). The lower bound on the number of users for each time period, by year and quarter, is shown in Figure 3.3.

Figure 3.4 shows the rate of recorded catches, r_t, $t = 1,\ldots 16$, for each species, calculated as the ratio between number of reported catches across all locations in each time period, $\sum_\ell y_{\ell t}$ and the corresponding lower bound of users in that period, m_t, i.e. $r_t = \sum_\ell y_{\ell t}/m_t$.

The rate of recorded catches of sea bass and mackerel is highest in summer, while for cod it is highest in autumn, and for whiting in autumn and winter. For sea bass, the rate in summer 2019 was approximately three times higher than the corresponding rate in other years, and the rates in autumn 2020 and spring 2021 were lower compared to these quarters in other years. In contrast, the rate for mackerel in spring 2021 was higher compared to corresponding quarters in other years. For cod, the

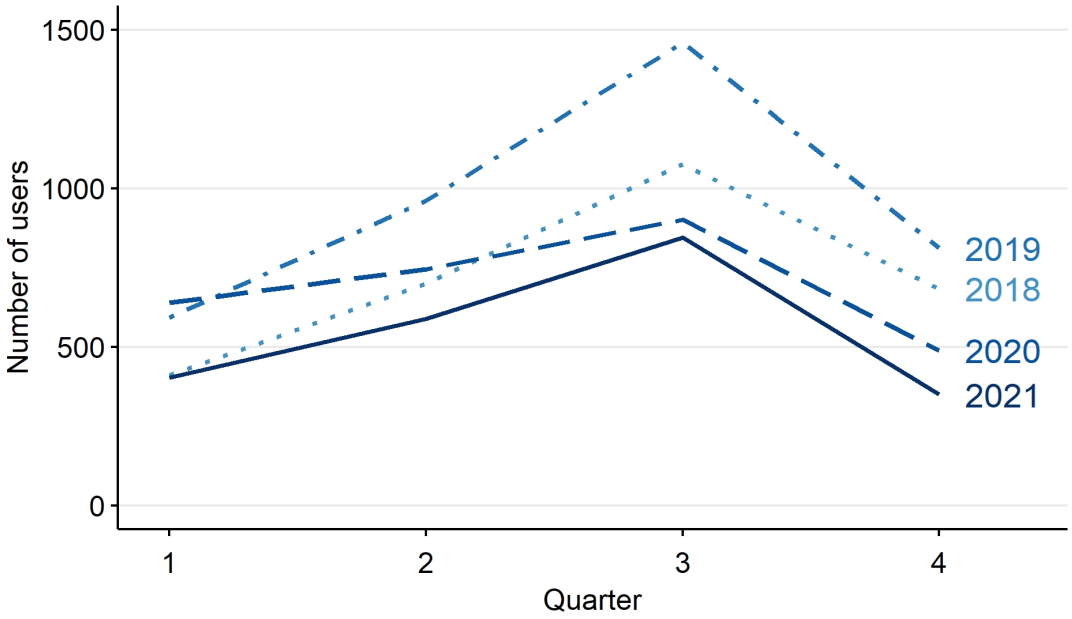


Fig. 3.3 Lower bound of the number of users m_t, $t = 1, \ldots, 16$, by year and quarter. The dotted line corresponds to $t_1 - t_4$, the dotted and dashed line to $t_5 - t_8$, the dashed line to $t_9 - t_{12}$ and the solid line to $t_{13} - t_{16}$.

rate was notably lower in summer 2021, but substantially higher the following quarter (autumn 2021), while for whiting the rate was higher in spring, summer and autumn 2021, compared to these quarters in other years.

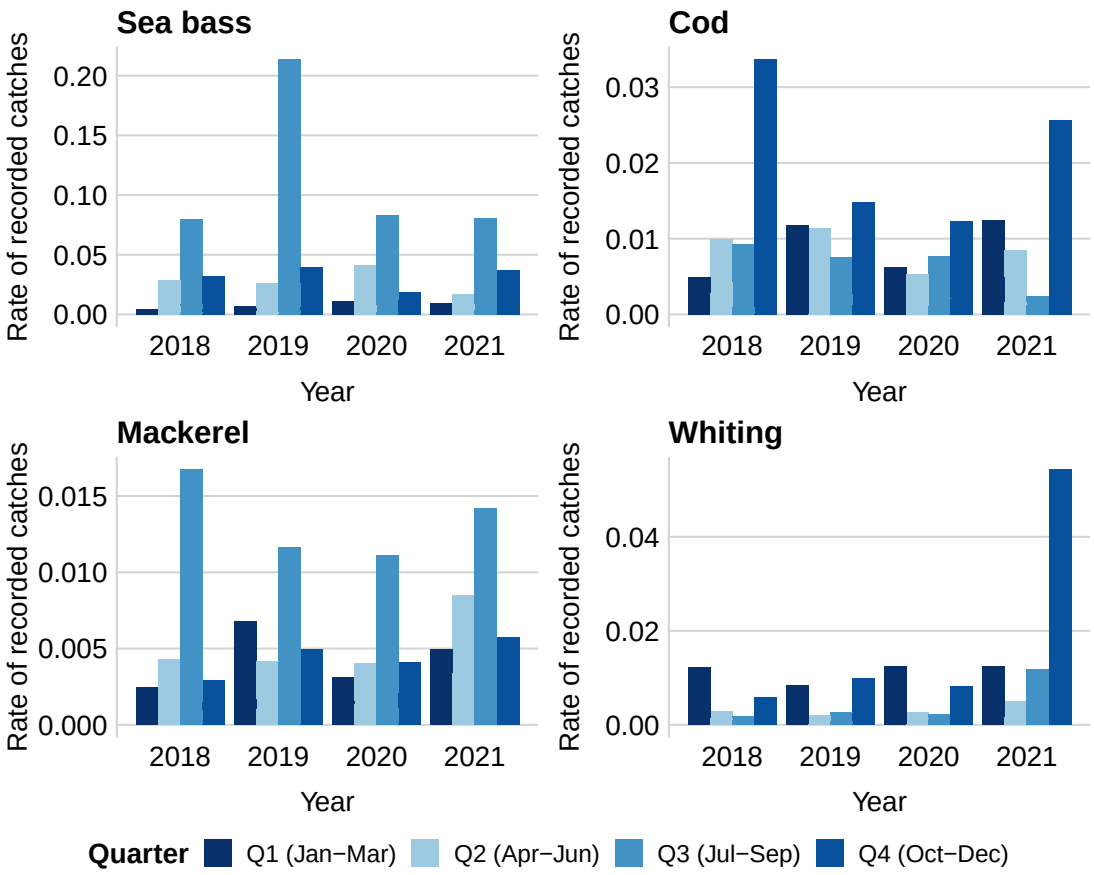


Fig. 3.4 Rate of recorded catches, r_t, for each time period, per year and quarter, with $r_t = \sum_\ell y_{\ell t}/m_t$ where $y_{\ell t}$ is the number of recorded catches in time period t and location ℓ and m_t is the lower bound of the number of Fishbrain users at time period t. Note the different y-axis scales for the different species.

Fitting the models

We model the number of recorded catches at location ℓ and time t, $y_{\ell t}$, for each species as the realisation of $Y_{\ell t}$, which is a zero-inflated Poisson variable. We model the mean of this Poisson distribution as a function of the fixed effects of quarters (Q1, Q2, Q3, Q4), and of a spatiotemporal random effect, as described below. We use the lower bound of the number of users at each time period, m_t, as an offset in the corresponding log-linear models.

$$Y_{\ell t} \sim \begin{cases} \text{Poi}(\lambda_{\ell t}) \text{ with probability } p \\ \\ 0 \text{ with probability } 1-p \end{cases} \tag{3.2.1}$$

where

$$\log(\lambda_{\ell t}) = \log(m_t) + \beta_0 + \beta_1 \mathscr{I}_{\text{quarter=Q2}} + \beta_2 \mathscr{I}_{\text{quarter=Q3}} + \beta_3 \mathscr{I}_{\text{quarter=Q4}} + \gamma_{\ell t},$$

with $\mathscr{I}_{\text{quarter=q}}$ equal to 1 if the corresponding observation belongs to quarter q, and 0 otherwise, and $\gamma_{\ell,t}$ the spatiotemporal random effect. We consider two cases for $\gamma_{\ell,t}$: an independent and identically distributed (iid) random effect and an autoregressive of order one (ar1) random effect, so that

$$\text{ar1} : \gamma_{\ell,t} = \alpha \gamma_{\ell,t-1} + \xi_{\ell t}$$

where $\xi_{\ell t}$ is a zero-mean Gaussian Field. $\xi_{\ell t}$ is assumed to be temporally independent and is characterised by the following spatiotemporal covariance function

$$\text{Cov}(\xi_{\ell t}, \xi_{j,u}) = \begin{cases} 0, & \text{if } t \neq u \\ \sigma^2_{\mathscr{C}} \mathscr{C}(\Delta_{ij}), & \text{if } t = u \end{cases}$$

for $i \neq j$, with $\mathscr{C}(\Delta_{ij})$ denoting the Matèrn spatial covariance function (see Blangiardo et al., 2013).

We fit the spatiotemporal models using the bru() function from the `inlabru` package in R (Bachl et al., 2019), which itself uses the R package `INLA` (Rue et al., 2009). In the models fitted here, we set $P(\rho < 500) = 0.5$ and $P(\sigma > 50) = 0.5$. This can be interpreted as a probability of 0.5 that the spatial range of the field is less than 500 metres, and a probability of 0.5 that the standard deviation

of the field is greater than 50 metres, where the spatial range is the distance at which the spatial autocorrelation is considered to be negligible. These probabilities were chosen to reflect our initial uncertainty and lack of prior knowledge regarding the spatial range and standard deviation of the field.

For each species, fitted models (iid and ar1) are compared using the Watanabe–Akaike information criterion (WAIC), which takes into consideration both model fit and complexity. The model with the lowest WAIC is considered the best fitting model, as it has the best trade-off in fit and complexity (Schrödle and Held, 2011).

Generating predictions

We generate predictions within an area around the coast, and predict rates, that is the estimated number of catches, $\widehat{y}_{\ell t}$ divided by the number of users m_t, in each case. A set of pixels is generated from the mesh, using the pixels() function from the R package `inlabru` (Bachl et al., 2019). These locations are shown in Figure 3.5. A mask covering the area from 1,000 metres inland to 12 nautical miles off the coast of the UK and Ireland was used to predict only in this region. The predictions of the catch rates at each pixel location were generated using the predict() function from the R package `inlabru` (Bachl et al., 2019) with this set of pixels and the number of samples set to 1,000. The model predicts the rate of catches at each pixel location during each time period, with the rate defined as the predicted number of catches at the location in the quarter divided by the lower bound for the number of users. This gives an upper bound for the rate for each location-quarter.

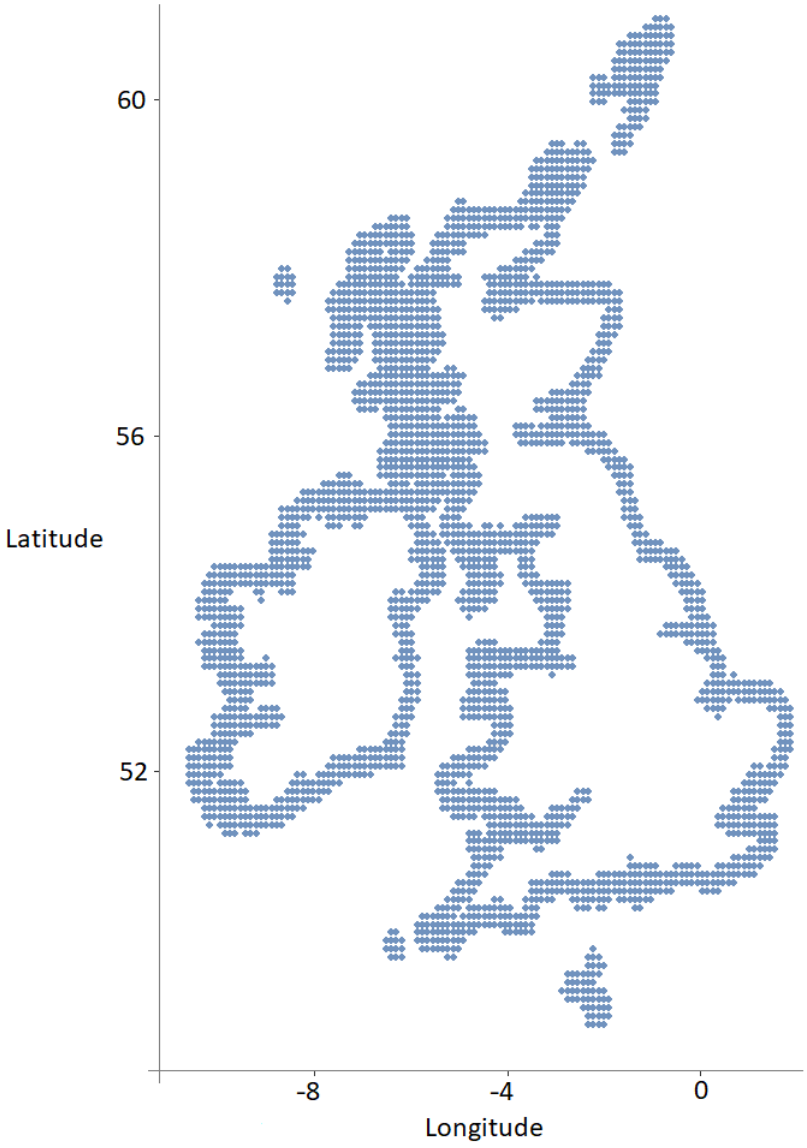


Fig. 3.5 Locations at which predictions are generated. These points are within 1,000 metres inland to 22.22 km (12 nautical miles) offshore from the coast.

3.3 Results

3.3.1 Models

For all four species, the best models (as determined by the lowest WAIC) were the models using ar1 random effects instead of iid (Table 3.2). For each species, the computational time required to fit the ar1 model was longer than the time required to fit the iid model (see Appendix B for details on computational times). However, for all species, the effective number of parameters was lower for the ar1 models, indicating that these models have a lower complexity than the iid models.

Table 3.2 Watanabe–Akaike information criterion (WAIC) and effective number of parameters for the iid and ar1 models for each species.

	Model			
	iid		**ar1**	
	WAIC	Parameters	WAIC	Parameters
Sea bass	12406	1446	11655	1334
Cod	5553	458	4875	214
Mackerel	7263	449	6938	272
Whiting	45739	14240	4102	206

Table 3.3 shows, for each species, the posterior summaries and 95% credible intervals of the estimated coefficients of the fixed effects of the model (quarters). The baseline is quarter 1 ('Intercept' row) and the values for quarters 2, 3 and 4 (rows 'Q2', 'Q3' and 'Q4' respectively) are the differences between the respective quarter and the baseline.

For sea bass and mackerel, all quarters have a strong positive effect on the response, which is the rates of catches reported, compared to quarter 1 (winter), with quarter 3 having the highest effect for both species. For cod, quarters 2 (spring) and 3 (summer) have a negative effect, albeit the latter is weaker, and quarter 4 (autumn) has a positive effect. Finally, for whiting, quarters 2 (spring) and 3 (summer) have a negative effect and quarter 4 (autumn) has a positive effect on the response.

Table 3.4 presents posterior summaries of the model random effects. For all species, there is a strong temporal correlation between quarters, indicated by the 'GroupRho' value being close to 1. The range of the spatial effect is 4.3 km for sea bass, 2.6 km for cod, 2.0 km for mackerel and 1.4 km for whiting.

Table 3.3 Posterior summaries of the intercept and of the fixed effects of quarters in the model for all species. The baseline is Q1.

	Mean	Stdev	Median	95% Credible Interval
Sea bass				
Intercept	−13.538	0.293	−13.537	(-14.116, -12.966)
Q2	1.822	0.151	1.820	(1.529, 2.123)
Q3	2.083	0.154	2.082	(1.786, 2.391)
Q4	1.499	0.159	1.498	(1.191, 1.815)
Cod				
Intercept	−10.690	0.172	−10.689	(-11.03, -10.354)
Q2	−0.468	0.134	−0.467	(-0.731, -0.204)
Q3	−0.212	0.125	−0.213	(-0.457, 0.034)
Q4	0.489	0.120	0.488	(0.256, 0.725)
Mackerel				
Intercept	−11.521	0.222	−11.516	(-11.972, -11.101)
Q2	1.452	0.213	1.447	(1.051, 1.886)
Q3	2.861	0.200	2.855	(2.487, 3.272)
Q4	0.943	0.228	0.938	(0.509, 1.403)
Whiting				
Intercept	−9.833	0.162	−9.832	(-10.153, -9.517)
Q2	−1.448	0.182	−1.446	(-1.811, -1.096)
Q3	−1.031	0.154	−1.030	(-1.333, -0.73)
Q4	0.454	0.125	0.454	(0.209, 0.701)

Table 3.4 Posterior summaries of the random effects in the best fitting model.

	Mean	Stdev	Median	95% Credible Interval
Sea bass				
Zero-probability parameter for ZIP1	0.51	0.02	0.51	(0.471, 0.55)
Range for field	4341.565	487.978	4308.976	(3462.31, 5421.399)
Stdev for field	51.506	5.633	51.259	(40.974, 63.484)
GroupRho for field	0.969	0.004	0.969	(0.959, 0.977)
Cod				
Zero-probability parameter for ZIP1	0.596	0.026	0.596	(0.544, 0.647)
Range for field	2606.632	608.685	2491.935	(1745.81, 4096.186)
Stdev for field	41.754	8.500	41.629	(25.877, 58.788)
GroupRho for field	0.981	0.004	0.981	(0.972, 0.988)
Mackerel				
Zero-probability parameter for ZIP1	0.652	0.018	0.652	(0.615, 0.686)
Range for field	2007.765	355.263	1994.7	(1360.26, 2745.918)
Stdev for field	41.265	9.838	39.629	(26.632, 64.799)
GroupRho for field	0.982	0.007	0.983	(0.964, 0.991)
Whiting				
Zero-probability parameter for ZIP1	0.788	0.001	0.788	(0.785, 0.79)
Range for field	1428.175	13.169	1429.836	(1398.743, 1449.225)
Stdev for field	66.565	0.435	66.548	(65.754, 67.461)
GroupRho for field	0.944	0.001	0.944	(0.941, 0.945)

3.3.2 Predictions

Figure 3.6 shows the spatial patterns in the predicted rates of catches reported by the best fitting model for each species in 2021. Due to the highest predicted rates being extreme values, we cap predictions at the 90% quantile on the maps to enable the spatial patterns in the predicted rates to be clearly visible. For all species, the pattern in the predicted rates was similar in the years 2018 to 2020 as in 2021. The prediction maps for each species, by year and quarter, are given in Appendix B (Figures B.8 to B.11), along with the computational times required to generate these predictions.

For each species, the spatial and seasonal patterns of the predicted rates of catches reported were consistent from one year to another, with only very minor fluctuations in the rate at any location from year to year.

For sea bass, the predicted rates of catches reported are higher in warmer months than cooler months, with noticeably lower rates in quarter 1 (January to March) than the rest of the year, and the highest predicted rates in quarter 3 (July to September). Higher rates are predicted along the south coast of England, and the south and west coasts of Wales, with local-scale variation between fishing sites in these areas. Higher rates are also predicted for the west of Scotland (although these predictions are from fewer sites with high reported catches; see Figure B.4), along with the south-east of Ireland and a few areas along the east coast of England and west of Ireland.

For cod, the predicted rates of catches reported are higher in cooler months than in warmer months, with predicted rates being highest in quarter 4 (October to December). Higher rates are spread along the north-east coast of England and east of Scotland, along with north-west Scotland, the Bristol Channel (between the south coast of Wales and south-west England) and some locations of reported catch on the south-east coast of Ireland.

For mackerel, the predicted rates of catches reported are highest in summer (quarter 3, July to September) in regions around most of the coast of the UK and Ireland, with the east of England, north coast of Scotland and southern Ireland tending to have lower rates. In contrast, predicted rates are lower during quarters 2 (April to June) and 4 (October to December) and concentrated at sites in southern England, north-east England, Scotland and north-west Wales, and low in quarter 1 (January to March) around the whole coastline.

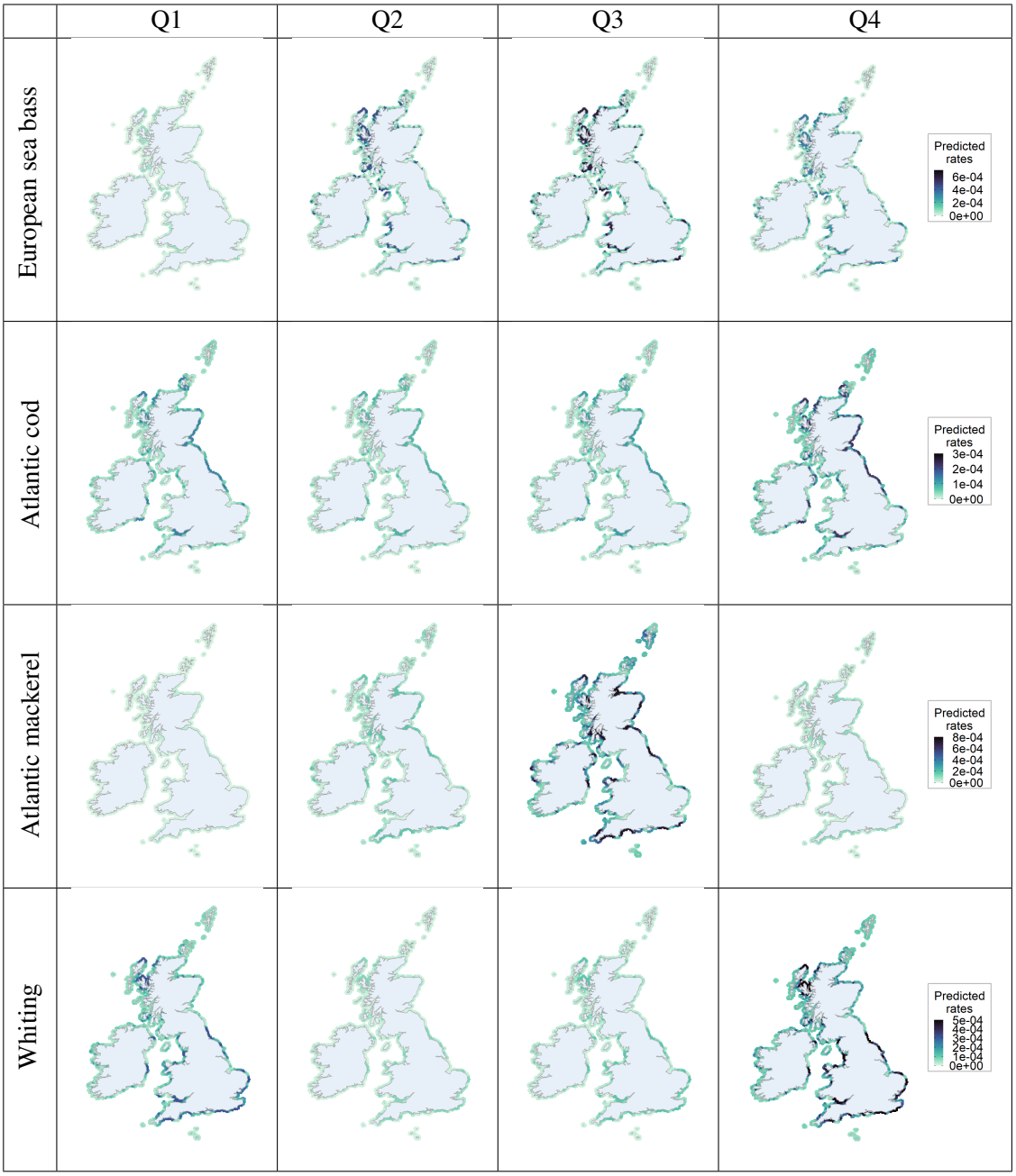


Fig. 3.6 Predicted rates for 2021. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). The predicted rates are capped to the 90% quantile to optimise visualisation.

Finally, for whiting, the predicted rates of catches reported are higher in cooler months (quarters 1 and 4) than in warmer months, with predicted rates being highest around the coast of England, and some areas of north-west Scotland.

3.4 Discussion

The aim of this study was to understand and predict reported catches around the coast of the UK and Ireland, and to investigate whether spatiotemporal models fitted to catches recorded in an angling app can provide useful insights into marine recreational fishing. Spatiotemporal zero-inflated Poisson models were fitted to catches of cod, sea bass, mackerel and whiting recorded by Fishbrain app users during the years 2018 to 2021 using the R package `inlabru` (Bachl et al., 2019), which fits models using the integrated nested Laplace approximation methods of the `INLA` package (Rue et al., 2009). Both iid and ar1 models were fitted for each species. Using the WAIC to compare the fit of the models, the ar1 model provided the better fit for all four species.

We found that the spatial pattern of reported catches is fairly consistent across quarters and years, and the spatial pattern of predicted catch rates is similar to the reported catch rate from the Fishbrain app. Higher catch rates of sea bass and mackerel are predicted in warmer months compared to cooler months, while the reverse pattern is seen for cod and whiting with higher predicted catch rates in cooler months. Predicted catch rates for sea bass tend to be higher in south and west regions, and for cod in north or east regions, while for mackerel and whiting predicted catches are more widespread, around most of the UK and Ireland for mackerel and around England and north-west Scotland for whiting. The overall predicted pattern of catches around the UK and across the seasons in general follows the expected spatiotemporal distribution for each species, although higher rates of sea bass and whiting are predicted around Scotland compared to rates obtained from UK sea angling diary information (Hyder et al. 2021, p. 43; Hyder et al. 2024, p. 71), suggesting this is an area where further data and model development are required. Furthermore, species distribution can be impacted by environmental conditions (Nikolioudakis et al., 2019) and future developments could include additional environmental covariates in the model.

In this study, the mesh used consisted of the highest density of triangles that enabled the model to be fitted to the data without fitting errors. In contrast to spatial methods that use regular grids, advantages of the mesh approach are that the density of triangles in the mesh is higher in regions with higher recorded catches with a 'smooth transition' to large triangles in regions with few recorded catches, generating better predictions in areas of interest whilst simultaneously reducing computation time and avoiding boundary effects (Lezama-Ochoa et al., 2020). Research has demonstrated that small variations in the mesh can substantially impact the predictions, with a trade-off between clear spatial effects and overfitting as the density of triangles in the mesh increases (Dambly et al., 2023). Further work could investigate the impact of different meshes on the predictions.

Marine recreational fishing provides socio-economic benefits, but can adversely impact fish stocks, therefore monitoring and management of marine recreational fishing is essential for sustainability of fish species and potential consequent wider impacts to the marine environment and ecosystems. Spatiotemporal models can provide valuable information on patterns of catches over space and time. However, obtaining data from probabilistic surveys is time-consuming and expensive. In contrast, user-generated data from apps such as Fishbrain do not require complex and resource-intensive data collection, and offer the advantage of being continuous across space and time. However, since Fishbrain app users are not a random sample from the population of anglers, the reported catches cannot be assumed to be a constant proportion of total catches made spatiotemporally. Anglers who use fishing apps have been found to have higher catch rates than anglers who do not engage with fishing apps (Gundelund et al., 2020) and biases are likely as catches are only reported by users who have chosen to use the app (Skov et al., 2021). Despite concerns that nonprobability data from apps is unable to provide the reliable evidence given by probability surveys (Brick et al., 2022), app data can nevertheless provide complementary insights into spatiotemporal patterns of catches (Skov et al., 2021), as well as invasive species, angler choices and changes to spatial distributions (Brick et al., 2022), and research into techniques for addressing biases and errors in citizen science data is ongoing (Johnston et al., 2023).

Furthermore, these opportunistic presence-only Fishbrain data are an incomplete picture of actual catches. Absences of catches in the data of any particular species at any given location on any given day may occur in four ways: users may not have gone fishing on that day at that location; users may

have gone fishing at that location but not caught the species; users may have gone fishing at that location and caught the species but chosen to not record the catch; or users may have gone fishing and caught the species, but recorded the catch incorrectly, for example recording the species or location incorrectly. Species distribution models using presence-only data are found to be less accurate for species with more versatility in their preferred habitats compared to species that have a more restricted habitat range, and using presence-absence data results in a more accurate model for species that inhabit a wide range of ecosystems (Brotons et al., 2004). Studies have shown that combining citizen science data and more robust data can improve accuracy over using single data sources alone (Adde et al., 2021; Matutini et al., 2021; Robinson et al., 2020). Therefore, a next step could be to investigate combining Fishbrain app data with other data sources.

Chapter 4

Global travel patterns in recreational fishing

Abstract

Recreational angling is a hugely important human activity around the world, with considerable benefits for individuals. Fishing tourism, defined as a specialised travel experience wherein individuals engage in fishing-related activities as the primary or secondary purpose of their trip, contributes substantially to local economies. Recent work has documented the effect of the COVID-19 pandemic on local angling activity, with results varying between countries and regions. In this paper, we demonstrate how data obtained from the mobile phone app Fishbrain can be used to study global fishing tourism. We employ network models to describe global travel patterns for recreational fishing and compare these patterns across years. Our findings suggest that international recreational fishing trips are more prevalent during summer months in the destination country, with travel to neighbouring countries often preferred over longer-distance travel, although a few countries appear to attract tourists from across the world. A considerable reduction in international travel for recreational fishing is evident from the start of the pandemic in 2020 quarter 2, continuing to various extents until the end of the study period in 2021 quarter 4, with variations between countries and regions in the impact on both inbound and outbound tourism.

4.1 Introduction

Recreational fishing is a popular leisure activity across the world (Arlinghaus et al., 2015; Cisneros-Montemayor and Sumaila, 2010), which connects individuals with nature, promotes well-being (Arlinghaus et al., 2002; Hook et al., 2022; Pita et al., 2022; Ryan et al., 2021) and provides opportunities for social engagement (Arlinghaus et al., 2002; McManus et al., 2011). It is important socially (Arlinghaus et al., 2021) and culturally (Arlinghaus et al., 2021; Hall, 2021; Mordue, 2009), and helps drive conservation (Arlinghaus et al., 2021). Furthermore, recreational fishing is economically

important (Hall, 2021), contributing to local economies through expenditure on fishing gear, travel and related services (Curtis et al., 2017; Pita et al., 2021).

International tourism contributes considerably to the global economy (Scott et al., 2012). Globally, China, India, and the United States generate the greatest number of tourists, including both domestic travellers and those travelling internationally (Adamiak and Szyda, 2022). In contrast, Europe has a higher proportion of international tourists (Adamiak and Szyda, 2022). The international tourism sector continues to experience global growth, with Asian countries in particular showing a significant upward trend in outbound international travel (McKercher and Mak, 2019). Recreational fishing is a substantial component of the tourism economy (Hall, 2021; Mordue, 2009). In some regions, international tourists comprise a significant portion of recreational fishing participants, contributing substantially to the local tourism economy (Pita et al., 2021). In Norway, the majority of participants in marine recreational angling are international tourists, while in northern Greece, approximately half of recreational fishers are international tourists (Pita et al., 2021). Similarly, Portugal's big game fishing sector relies heavily on international tourism (Pita et al., 2021). Across Australia, New Zealand, and the Pacific region, international tourists account for 6% of recorded catches (Britton et al., 2023).

Fishing tourism encompasses the travel (domestic and international) of individuals beyond their local area (United Nations, 2010), with recreational fishing either as the primary reason for the travel or as a secondary activity during travel for other purposes (van den Heuvel et al., 2022). Fishing tourists range from specialised enthusiasts (Mordue, 2009) to those who engage in recreational fishing as a secondary activity during their trip (Mordue, 2009), with tour operators offering specialised fishing holiday packages (Mordue, 2016). In this chapter we use the term 'trip' to refer to travel to a single different country, which may include multiple fishing locations within that country, with multiple country destinations treated as separate trips, although we note that for economic tourism statistics, the United Nations World Tourism Organization (UNWTO) define a trip as "the travel of a visitor from the time of leaving his/her usual residence until he/she returns" (United Nations, 2010, p. 13), which can therefore include destinations in multiple countries.

Fishing tourism patterns can vary significantly across spatial and temporal scales, as well as by target species (Camp et al., 2018). In some countries the economic contribution from international

anglers can be substantial (Cisneros-Montemayor and Sumaila, 2010; Hannonen and Hoogendoorn, 2022; Pita et al., 2021). However, potential negative aspects of fishing tourism include ecological impacts (Cooke and Cowx, 2004; Dainys et al., 2022b) including invasive species (Davis and Darling, 2017; Hall, 2021), and natural capital impacts to local communities (Hall, 2021). Knowledge of angler travel patterns can be valuable from an economic perspective and for effective fisheries and ecosystem management (Camp et al., 2018). Understanding and predicting anglers' travel patterns is valuable for maintaining sustainability of fish stocks and mitigating the spread of fish diseases or invasive species (Drake and Mandrak, 2010). Fishing tourists may be more experienced, fish more often, and be motivated by different factors compared to local anglers (Ditton et al., 2002) and fisheries with high levels of fishing tourism may require alternative management strategies (Camp et al., 2018). Therefore, understanding fishing tourism can help balance the needs of fisheries, the environment and local communities.

Despite the importance of the global fishing tourism industry (Hall, 2021; Hannonen and Hoogendoorn, 2022; Mordue, 2009), research on travel by anglers for recreational fishing remains limited (Camp et al., 2018; Hall, 2021; Hannonen and Hoogendoorn, 2022; Mordue, 2009, 2016), although research interest has increased in recent years (Hannonen and Hoogendoorn, 2022), on a variety of aspects including angler loyalty (Lam-González et al., 2021; van den Heuvel et al., 2022), motivation (Golden et al., 2019), distances travelled (Jalali et al., 2022), expenditure (Lam-González et al., 2021), and a social constructionist perspective on fishing tourism (Mordue, 2009). Around a quarter of studies on fishing tourism focus on a Nordic country (Hannonen and Hoogendoorn, 2022), often on providers of fishing opportunities (Fredman and Margaryan, 2021), with an issue of the *Scandinavian Journal of Hospitality and Tourism* dedicated to fishing tourism (Andersson, 2021; Hall, 2021; Lam-González et al., 2021; Merkel et al., 2021; Pokki et al., 2021; Stensland et al., 2021; Wheeller and Hall, 2021). Research also includes game fishing tourism (León et al., 2003; Mordue, 2016) and economic perspectives, for example the development of fishing tourism (Ditton et al., 2002; Stensland, 2013; Waldo and Paulrud, 2012). A review of research in angling tourism (Hannonen and Hoogendoorn, 2022) reveals three key research areas: management, impacts, and other aspects of angling tourism. Most studies focus on a single country (Hannonen and Hoogendoorn, 2022). In their analysis, Hannonen and Hoogendoorn (2022, p. 6) note that "tourism studies and journals on tourism

management / geography have, by and large, ignored angling tourism as a tourism activity across the globe" and recommend that "future research on angling tourism development requires broadening the geographical scope of studies" (Hannonen and Hoogendoorn, 2022, p. 6).

The COVID-19 pandemic has affected recreational fishing worldwide (Britton et al., 2023; Pita et al., 2021). Research on the impact of the COVID-19 pandemic on recreational fishing to date tends to focus primarily on the first few months of the pandemic. These studies indicate that the impact varied across different countries (Audzijonyte et al., 2023; Pita et al., 2021), likely due to the varying policies and restrictions implemented across different countries (Audzijonyte et al., 2023). For example, fishing effort increased in the United States and Canada, but mixed responses were observed in Europe, with fishing effort increasing in Denmark (Gundelund and Skov, 2021) but decreasing in the United Kingdom (Hook et al., 2022). As well as being impacted by travel restrictions (Hook et al., 2022; Ryan et al., 2021), individual circumstances impacted fishing effort, for example an increase in available time and desire for social connection was associated with an increase in fishing effort (Midway et al., 2021), but perceived risk of the pandemic reduced fishing activity (Hook et al., 2022). Existing research on the impact of the COVID-19 pandemic on recreational fishing includes impacts in the United Kingdom (Hook et al., 2022), Denmark (Gundelund and Skov, 2021), Canada (Howarth et al., 2021), United States (Midway et al., 2021; Paradis et al., 2020) and Australia (Ryan et al., 2021) during 2020, and comparisons between countries during 2020 and 2021 (Audzijonyte et al., 2023; Britton et al., 2023; Pita et al., 2021). To the best of our knowledge, the impact of the pandemic on global travel patterns for recreational fishing in the context of fishing tourism have not yet been investigated.

The widespread use of smartphones has generated vast quantities of high-resolution spatial data through mobile positioning data generated when mobile devices connect to the mobile operator's network (Kovács et al., 2021) and location data collected by smartphone applications (hereafter referred to as 'apps'). These data provide unprecedented opportunities to enhance our understanding of human behaviour across a wide range of fields, including tourism (Shoval and Ahas, 2016) and recreational fishing (Audzijonyte et al., 2023; Johnston et al., 2022; Papenfuss et al., 2015; Skov et al., 2021; Venturelli et al., 2017). Apps like Fishbrain (Fishbrain, 2023) allow anglers to record catches, share data, and contribute to citizen science initiatives. App data have enhanced our understanding of

spatial and temporal patterns of recreational angling travel in Canada (Papenfuss et al., 2015) and changes in fishing patterns during the COVID-19 pandemic (Audzijonyte et al., 2023), while mobile positioning data have provided insights into seasonal patterns of recreational fishing and popular fishing destinations of Latvian tourists in Estonia (Ahas et al., 2007) and general tourist travel at local (Kovács et al., 2021), regional (Kim et al., 2020) and international (Saluveer et al., 2020) scales.

In this chapter, we use data from Fishbrain (Fishbrain, 2023) from years 2018 to 2021 inclusive to study international fishing tourism, defined here as trips and catches occurring as a result of travel to a different country, identified by Fishbrain users reporting catch(es) in a different country to the country in which their device is registered. We investigate both seasonal patterns and changes during the COVID-19 pandemic. To achieve our goal, we employ network modelling of the catch records, where we define an observed interaction to be any instance of an angler reporting a catch in a country other than the country in which their device is registered. Network modelling is a widely used approach for studying interactions between individuals/places/products, and understanding and representing complex relationships. In the context of international tourism, which is the focus of this chapter, network modelling has been used by Liu et al. (2023) to study the links in international stock tourism and hospitality markets, with findings suggesting that "network interconnectedness can significantly explain the variations in tourism and hospitality market returns" (Liu et al., 2023, p. 2). However, to our knowledge, network modelling has not yet been used to study recreational fishing tourism.

The remainder of the chapter proceeds as follows. The methodology is presented in Section 4.2. Section 4.3 presents the results, which focus on networks of travel for world regions, world countries, and of countries within Europe. We also consider the United Kingdom and United States as illustrative examples of travel relating to individual countries. In each case, we construct networks, calculate network statistics, and identify communities within networks. We discuss our findings in general terms, and with respect to changes observed as a result of the COVID-19 pandemic. Finally, Section 4.4 contains a discussion of the findings and direction for future work.

4.2 Materials and methods

4.2.1 Data preparation

The dataset comprises catches documented by users of the Fishbrain app, which is accessible from www.fishbrain.com (Fishbrain, 2023). This app offers fishing insights, social networking features, and the option to record catches with related details, including the date and location of catch, fishing method, and species caught. For research purposes, Fishbrain provided anonymised data under a data-sharing agreement.

Complete data extracts of all catch records were obtained from Fishbrain (Fishbrain, 2023) on 18 August 2021 and 1 February 2022. Catches recorded as occurring between 1 January 2018 and 17 August 2021 were taken from the extract on 18 August 2021. Due to the amount of processing time that had already been carried out determining the country of the catch location, the catches dated as occurring from 18 August 2021 to 31 December 2021 in the second extract were appended to those from the first extract. Therefore, any catches occurring before 18 August 2021 that were recorded on or after 18 August 2021 are not included in this analysis.

When recording a catch, users can select the catch location from a pre-populated list of locations or use the GPS location of the phone (or both). The location selected by the user from the pre-populated list was considered to be more reliable than the GPS location of the phone at the time of recording the catch, and records without a location selected from the list were discarded, giving 7,771,485 catch records with the location recorded. Identifying the country of each catch record in turn was inefficient, and estimates using a small subset of data indicated that it would take months to complete. To improve efficiency, all unique locations and their coordinates were extracted from the data using the location ID, and the country of each unique location was determined using a spatial search with the over() function from the `sp` package (Pebesma and Bivand, 2005), using a map of the union of ESRI world country boundaries and the world Exclusive Economic Zones (Flanders Marine Institute, 2020). Since the majority of catches were in the United States, to improve efficiency further, the United States boundaries were extracted from the world map and each location was first assessed using the over() function from the `sp` package (Pebesma and Bivand, 2005) to determine whether it was within the boundaries of the United States. If the location was not in the United States, then the

corresponding country was identified using the over() function from the `sp` package (Pebesma and Bivand, 2005) using the entire world boundaries map. Finally, the country of each Fishbrain catch record was determined by matching its location ID to the list of unique locations and their countries, which was ordered numerically by location ID.

The country in which the subscriber identity module (SIM) card associated with the device is registered is available for every record, and is taken to be the country where the travel originates (hereafter referred to as the 'origin' country). All catches recorded as occurring in the same country as the origin country are excluded, leaving 104,749 catches recorded by 37,699 users in 50,821 trips to a different country (hereafter referred to as the 'destination' country) from the origin country. This represents 1.3% of the catch records for which the location was recorded.

The World Bank region of each country was found using the countrycode() function from the `countrycode` package (Arel-Bundock et al., 2018) in R. There were 11 countries with no matched region, which were manually assigned to the closest region using a map of World Bank regions (World Bank, 2018). As some islands in the French Southern Territories are closer to Sub-Saharan Africa, while others are closer to Antarctica, the latitude and longitude coordinates of the catch location of individual catches recorded as occurring in these islands was used to determine the region for catches in the French Southern Territories. Details of the countries without matching World Bank regions, their allocated regions, and the number of catches recorded are provided in Appendix C. A total of 326 records where either the origin or destination region is Antarctica are excluded from the analysis. This takes the number of catches recorded in the period 2018 to 2021 from 104,749 to 104,423, by 37,531 users, across 50,566 trips.

Separate sets of network models are fitted for the number of trips (which we define below) and the number of catches, recorded as occurring between 1 January 2018 and 31 December 2021, inclusive, as the result of travel to a different country. The years prior to 2018 are excluded due to lower numbers of catches recorded during these years, before the Fishbrain app increased in popularity. The data preparation and analysis was carried out in R (version 4.2.2) (R Core Team, 2022).

We split the four years (2018-2021) into 16 quarters. We define quarter 1 (Q1) to correspond to January-March each year; quarter 2 (Q2) to April-June; quarter 3 (Q3) to July-September; and quarter 4 (Q4) to October-December.

We define the number of trips per quarter as the number of unique origin-destination country pairs per user per quarter. For example, if a user from the United Kingdom recorded at least one catch in France and at least one catch in Sweden in the same quarter, this is defined as two trips. A user could make multiple trips to a destination within a quarter; however, it cannot be determined from the data available whether multiple catches recorded over a period of time were made during a single trip of a long duration, or during multiple short trips within that period. Therefore, for this analysis, it is assumed that each user makes a maximum of one trip to each country destination per quarter. It is also possible that catches recorded from a single trip occur in two (or more) quarters. In this case, these are treated in the analysis as multiple trips occurring in different quarters.

We create network models for the number of trips made by users of the Fishbrain app and the number of catches recorded, each quarter. We investigate travel within Europe, around the world, across world regions, and for two case study countries. To investigate the broad patterns, we include as many countries in the networks as possible whilst still being able to visualise the results and interpret the important links (from both economic and fisheries management perspectives), by excluding countries that contribute only a small number of trips or catches to the network and hence are of little consequence in terms of understanding important travel patterns. For the Europe and world countries networks, we select the top 15 countries by origin country and the top 15 countries by destination country, which results in around 20 countries in the networks due to overlap within the top 15 origin and destination countries. For the case study countries, we include in the networks only the edges with at least a threshold number of trips or catches, and the vertices associated with these edges. For the world regions, all trips or catches are included in the networks, as it is easy to visualise and investigate the patterns between the seven regions. We present the results based on the number of trips in Section 4.3. The corresponding results for the catches are in Appendix C.

Overview of data

We define the number of users in a quarter as the number of users actively using the app by recording at least one catch during the quarter.

The number and proportion of users, trips and catches recorded, along with the average number of trips per user, catches per user and catches per trip, are shown by year in Table 4.1 and by origin

and destination countries for the top countries by number of users in Tables 4.2 and 4.3, for the period 1 January 2018 to 31 December 2021. Note that users may visit multiple countries, so the proportions for the number of users sum to greater than 100%.

From 2018 to 2019, the number of users, trips and catches recorded increased (Table 4.1). Subsequently, in 2020, the number of users and trips decreased to less than in 2018, while the number of catches decreased to levels similar to 2018. From 2020 to 2021, the number of users, trips and catches recorded all increased, although the number of users remained less than in 2018 and the number of trips was approximately the same as in 2018, while the number of catches was higher than in 2018 but still lower than in 2019. The average number of trips and catches per user increased each year from 2018 to 2020, but then decreased in 2021. The number of catches per trip increased each year from 2018 to 2020, but remained the same from 2020 to 2021.

The United States has the greatest proportion of users, with around two-thirds of all app users, who accounted for half of all catches recorded during the period from 2018 to 2021 (Table 4.2). Canada, the United Kingdom, Sweden and Australia each have a moderate number of users.

Japan and Puerto Rico, and to a lesser extent Argentina, have a high average number of catches per user, trips per user and catches per trip, compared to other countries (Table 4.2). A number of users from Japan, Puerto Rico, and Argentina recorded an exceptionally high number of catches compared to users from other countries, resulting in a high mean catch rate per user for these countries. In total, 3,361 catches are recorded by a total of 70 users from Japan, resulting in a mean of 48.0 catches per user. Over two-thirds of these catches are recorded are by just four users, with all these catches in the United States. Similarly, users from Puerto Rico recorded 2,836 catches, with over one-third of these by four users, with an overall mean of 16.9 catches per user. The mean catch rate per user for Argentina is 11.4. In comparison, the mean catch rate for other countries is typically between 2 and 5 catches per user. In addition, Japan and Puerto Rico have a higher mean number of trips per user compared to other countries (3.1 and 2.7 respectively). For all other countries, the mean number of trips per user ranges from 1.3 to 2.2, and is typically around 1.3 to 1.6, with a mean of 1.7 across all countries. The mean number of catches per trip is also higher for Japan, Puerto Rico and Argentina (15.4, 6.2 and 7.7 respectively) in comparison with other countries, all of which have a mean of between 1.5 and 3.6 catches per trip.

The countries with the highest number of Fishbrain users visiting the country and recording a catch as occurring there are Canada (21%), the United States (12%), and Mexico (12%) (Table 4.3). These countries also have the highest number of inbound trips, and collectively these destinations received almost half of all trips. Although the United States did not have the highest number of inbound visitors, it had the highest number of catches recorded by visitors, accounting for 25% of all catches recorded. Additionally, it was the country with the highest mean number of catches per visitor (5.9), driven by a high mean number of trips per user to this destination (1.6) and a high mean number of catches per trip (3.7).

Of the 104,423 catches recorded during the period 1 January 2018 to 31 December 2021, 89% have the species recorded. Of the catches with the species recorded, the most commonly recorded species are Largemouth black bass (16%), Northern pike (8%) and Smallmouth bass (5%). In total, 1,567 different species are recorded during the period. Notably, there were only 14 species where each accounted for more than 1% of the total catches. These 14 species comprised 52% of the catches with the species recorded. The remaining 48% of catches were of species with less than 1% of the total number of records with the species recorded.

Table 4.4 shows the number of catches recorded by species, for the top 15 species. Tables 4.5 and 4.6 show the number of catches recorded by year and by location (destination country), respectively, for the three most commonly recorded species. The number of catches recorded per year varies by species, with higher numbers of Largemouth black bass recorded in 2019 and 2020, Northern pike in 2019 and 2021, and Smallmouth bass in 2018 and 2019, compared to other years. The most common catch locations were the United States and Canada for Largemouth black bass and Smallmouth bass, and Canada and Sweden for Northern pike.

The fishing method was recorded for 38% of catches. The most commonly recorded fishing methods were casting (39%) and bottom fishing (21%). The number of catches by each fishing method recorded is shown in Table 4.7.

Table 4.1 Users recording a catch, catches recorded and trips taken each year, as a result of travel to a different country. The number of users is the number of active users each year. Users may be active in multiple years. The number of trips per year is the sum of the number of trips in each quarter of that year. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. The proportions of users sum to greater than 100% due to some users recording catches in multiple years. The proportions of trips and catches do not sum to 100% due to rounding.

	Counts			Proportions (%)			Averages		
Year	Users	Trips	Catches	Users	Trips	Catches	Trips per user	Catches per user	Catches per trip
2018	11 079	12 908	24 299	30	25	23	1.17	2.19	1.88
2019	12 064	14 450	29 756	32	29	28	1.20	2.47	2.06
2020	9108	11 154	24 233	24	22	23	1.22	2.66	2.17
2021	10 118	12 054	26 135	27	24	25	1.19	2.58	2.17
Total	42 369	50 566	104 423	113	101	99			

Table 4.2 Users recording a catch, trips taken and catches recorded according to users' origin country, as a result of travel to a different country during the period 2018-2021. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. Countries are listed in order of number of users.

	Counts			Proportions (%)			Averages		
Origin country	Users	Trips	Catches	Users	Trips	Catches	Trips per user	Catches per user	Catches per trip
United States	25 269	31 658	52 847	67.3	62.6	50.6	1.3	2.1	1.7
Canada	2785	3918	7997	7.4	7.7	7.7	1.4	2.9	2.0
United Kingdom	1503	2153	5491	4.0	4.3	5.3	1.4	3.7	2.6
Sweden	1122	1578	3305	3.0	3.1	3.2	1.4	2.9	2.1
Australia	897	1175	1976	2.4	2.3	1.9	1.3	2.2	1.7
Germany	382	620	1468	1.0	1.2	1.4	1.6	3.8	2.4
Brazil	366	544	1399	1.0	1.1	1.3	1.5	3.8	2.6
Mexico	345	607	1674	0.9	1.2	1.6	1.8	4.9	2.8
France	303	419	1054	0.8	0.8	1.0	1.4	3.5	2.5
Ireland	296	490	1218	0.8	1.0	1.2	1.7	4.1	2.5
Denmark	287	430	1431	0.8	0.9	1.4	1.5	5.0	3.3
Poland	232	469	1259	0.6	0.9	1.2	2.0	5.4	2.7
Finland	220	298	491	0.6	0.6	0.5	1.4	2.2	1.6
South Africa	216	339	906	0.6	0.7	0.9	1.6	4.2	2.7
Netherlands	209	273	751	0.6	0.5	0.7	1.3	3.6	2.8
Norway	188	278	570	0.5	0.5	0.5	1.5	3.0	2.1
Puerto Rico	168	459	2836	0.4	0.9	2.7	2.7	16.9	6.2
Switzerland	135	199	344	0.4	0.4	0.3	1.5	2.5	1.7
Spain	132	184	588	0.4	0.4	0.6	1.4	4.5	3.2
New Zealand	121	176	272	0.3	0.3	0.3	1.5	2.2	1.5
Belgium	111	220	534	0.3	0.4	0.5	2.0	4.8	2.4
United States Minor Outlying Islands	102	225	812	0.3	0.4	0.8	2.2	8.0	3.6
Singapore	85	125	241	0.2	0.2	0.2	1.5	2.8	1.9
Lithuania	82	115	218	0.2	0.2	0.2	1.4	2.7	1.9
Italy	81	158	502	0.2	0.3	0.5	2.0	6.2	3.2
China	76	110	252	0.2	0.2	0.2	1.4	3.3	2.3
Japan	70	218	3361	0.2	0.4	3.2	3.1	48.0	15.4
Romania	68	116	249	0.2	0.2	0.2	1.7	3.7	2.1
Hong Kong	66	138	333	0.2	0.3	0.3	2.1	5.0	2.4
Latvia	65	107	194	0.2	0.2	0.2	1.6	3.0	1.8
Argentina	60	89	686	0.2	0.2	0.7	1.5	11.4	7.7

Table 4.3 Users, trips and catches by destination, as a result of travel to a different country during the period 2018-2021. Proportions (as percentages) are calculated based on the total number of users recording a catch (37,531), trips taken (50,566) and catches recorded (104,423) during the four-year period. 'Users' is the number of users recording a catch as occurring in the country, 'trips' is the number of trips taken to the country, and 'catches' is the number of catches recorded as occurring in the country. Countries are listed in order of number of users.

	Counts			Proportions (%)			Averages		
Destination country	Users	Trips	Catches	Users	Trips	Catches	Trips per user	Catches per user	Catches per trip
Canada	7809	9496	17 973	20.8	18.8	17.2	1.2	2.3	1.9
United States	4440	7098	26 133	11.8	14.0	25.0	1.6	5.9	3.7
Mexico	4345	5082	7785	11.6	10.1	7.5	1.2	1.8	1.5
Australia	2019	2535	4409	5.4	5.0	4.2	1.3	2.2	1.7
United Kingdom	1785	2400	5636	4.8	4.7	5.4	1.3	3.2	2.3
Sweden	1642	2230	4866	4.4	4.4	4.7	1.4	3.0	2.2
Norway	1079	1322	2850	2.9	2.6	2.7	1.2	2.6	2.2
Bahamas	908	985	1477	2.4	1.9	1.4	1.1	1.6	1.5
Brazil	852	914	1644	2.3	1.8	1.6	1.1	1.9	1.8
China	810	930	1380	2.2	1.8	1.3	1.1	1.7	1.5
Puerto Rico	668	1028	2172	1.8	2.0	2.1	1.5	3.3	2.1
Costa Rica	597	644	1244	1.6	1.3	1.2	1.1	2.1	1.9
New Zealand	494	608	1053	1.3	1.2	1.0	1.2	2.1	1.7
Finland	489	638	1330	1.3	1.3	1.3	1.3	2.7	2.1
France	482	603	1405	1.3	1.2	1.3	1.3	2.9	2.3
Spain	444	525	883	1.2	1.0	0.8	1.2	2.0	1.7
Ireland	426	531	936	1.1	1.1	0.9	1.2	2.2	1.8
Denmark	391	523	885	1.0	1.0	0.8	1.3	2.3	1.7
Netherlands	339	503	1248	0.9	1.0	1.2	1.5	3.7	2.5
Cuba	298	341	521	0.8	0.7	0.5	1.1	1.7	1.5
Dominican Republic	250	269	383	0.7	0.5	0.4	1.1	1.5	1.4
Thailand	248	275	536	0.7	0.5	0.5	1.1	2.2	1.9
Russian Federation	243	257	313	0.6	0.5	0.3	1.1	1.3	1.2
Japan	232	297	530	0.6	0.6	0.5	1.3	2.3	1.8
South Africa	228	277	413	0.6	0.5	0.4	1.2	1.8	1.5
Belize	227	258	560	0.6	0.5	0.5	1.1	2.5	2.2
Nigeria	216	217	267	0.6	0.4	0.3	1.0	1.2	1.2
India	211	233	313	0.6	0.5	0.3	1.1	1.5	1.3
Argentina	205	221	311	0.5	0.4	0.3	1.1	1.5	1.4
Germany	203	255	382	0.5	0.5	0.4	1.3	1.9	1.5

Table 4.4 Number of catches recorded by species 2018-2021, for the top 15 species recorded. The remaining 47% of species recorded each contributed less than 1% of the total number of catches recorded. The proportions are based on the total number of catches with the species recorded (92,489).

Species	Species common name	Number of catches	Proportion (%)
Micropterus salmoides	Largemouth black bass	14941	16
Esox lucius	Northern pike	7188	8
Micropterus dolomieu	Smallmouth bass	4657	5
Oncorhynchus mykiss	Rainbow trout	3340	4
Perca fluviatilis	European perch	3335	4
Sander vitreus	Walleye	3199	3
Cyprinus carpio	Common carp	2642	3
Lepomis macrochirus	Bluegill	1536	2
Coryphaena hippurus	Common dolphinfish	1469	2
Salmo trutta trutta	Sea trout	1363	1
Ictalurus punctatus	Channel catfish	1311	1
Perca flavescens	American yellow perch	1062	1
Sphyraena barracuda	Great barracuda	1000	1
Caranx hippos	Crevalle jack	964	1
Gadus morhua	Atlantic cod	909	0.98

Table 4.5 Number of catches recorded per year for the three most commonly recorded species.

Year	Largemouth black bass	Northern pike	Smallmouth bass
2018	3376	1449	1246
2019	4326	2069	1220
2020	4084	1536	1044
2021	3155	2134	1147
Total	14941	7188	4657

Table 4.6 Most common catch locations (destination country) for the three most commonly recorded species.

Destination	Largemouth black bass	Destination	Northern pike	Destination	Smallmouth bass
United States	8647	Canada	2102	Canada	2984
Canada	2446	Sweden	1685	United States	1361
Mexico	844	United Kingdom	819	Mexico	47
Republic of Korea	351	Finland	615	Australia	23
China	302	United States	518	Puerto Rico	12
Australia	201	Netherlands	404	United Kingdom	12
United Kingdom	148	Ireland	164	China	11
South Africa	139	Denmark	142	Nigeria	9
Spain	80	Norway	104	Kyrgyzstan	8
Italy	78	France	93	Cuba	7

Table 4.7 Method of fishing, for the catches recorded during the period 2018-2021. Note that only 38% of catch records have the species recorded.

Method	Number of catches	Proportion (%)
Casting	15631	39
Bottom fishing	8539	21
Trolling	3686	9
Jig fishing	3349	8
Fly fishing	2133	5
Free line	1463	4
Pole fishing	1414	4
Sea angling	1033	3
Jerk fishing	932	2
Vertical jigging	618	2
Surfcasting	485	1
Ice fishing	463	1
Spear fishing	229	1
Hand lining	114	0.3
Total	40172	100

Case study countries

The case studies of the individual countries explore the travel to and from the United Kingdom and the United States, two countries in different world regions with some of the highest numbers of travel catches recorded (see Tables 4.14 and 4.15).

In these case studies, we consider the most common interactions, representing the most popular destinations to and from the case study country, by including only the edges that meet a minimum threshold number of trips from the origin to destination vertex each quarter, along with the associated vertices. For the United Kingdom, the threshold is five. For the United States, due to the larger number of trips, the threshold is 30. The thresholds for the networks based on the number of catches are the same as for the number of trips. This provides a focus on the most common interactions and graphs that are easily interpretable.

Table 4.8 Top destination countries from the United Kingdom and origin countries to the United Kingdom, according to the number of trips from and to the United Kingdom respectively, from 2018 to 2021 inclusive. The proportions are based on the total number of trips from the United Kingdom (2,153) and to the United Kingdom (2,400) respectively.

Origin country			**Destination country**		
Country	Total trips	Proportion	Country	Total trips	Proportion
United States	419	19.5	United States	1659	69.1
France	271	12.6	Ireland	296	12.3
Ireland	215	10.0	Poland	75	3.1
Australia	147	6.8	Australia	49	2.0
Spain	99	4.6	Bulgaria	35	1.5
Canada	94	4.4	Canada	34	1.4
Poland	65	3.0	Denmark	27	1.1
Norway	57	2.6	South Africa	25	1.0
Sweden	55	2.6	Isle of Man	21	0.9
Thailand	43	2.0	Romania	17	0.7
Greece	41	1.9	Spain	15	0.6
Jersey	37	1.7	France	14	0.6
Romania	30	1.4	Italy	12	0.5
New Zealand	26	1.2	Sweden	8	0.3
Netherlands	25	1.2	Belgium	7	0.3
Lithuania	21	1.0			
Denmark	20	0.9			
Guernsey	20	0.9			
Portugal	19	0.9			

Table 4.9 Top destination countries from the United Kingdom and origin countries to the United Kingdom, according to the number of catches reported by United Kingdom users and the number of catches reported as occurring in the United Kingdom respectively, from 2018 to 2021 inclusive. The proportions are based on a total of 5,491 catches reported by users from the United Kingdom and 5,636 catches reported in the United Kingdom respectively.

Origin country			**Destination country**		
Country	Total catches	Proportion	Country	Total catches	Proportion
United States	1673	30.5	United States	3246	57.6
France	669	12.2	Ireland	775	13.8
Australia	388	7.1	Denmark	477	8.5
Ireland	380	6.9	Poland	198	3.5
Korea (the Republic of)	245	4.5	Italy	148	2.6
Jersey	162	3.0	Bulgaria	132	2.3
Sweden	159	2.9	Isle of Man	123	2.2
Spain	158	2.9	Spain	98	1.7
Canada	153	2.8	Australia	65	1.2
Norway	131	2.4	South Africa	52	0.9
Poland	121	2.2	Canada	48	0.9
Thailand	91	1.7	Uruguay	44	0.8
Greece	90	1.6	France	27	0.5
Netherlands	58	1.1	Romania	21	0.4
New Zealand	55	1.0	Netherlands	14	0.2
Guernsey	51	0.9			
Denmark	47	0.9			
Latvia	42	0.8			
Romania	36	0.7			
Lithuania	32	0.6			
Portugal	30	0.5			
Brazil	27	0.5			
Mexico	26	0.5			
Italy	25	0.5			
Croatia	24	0.4			
Iceland	24	0.4			
Turkey	24	0.4			
Bulgaria	23	0.4			
Finland	22	0.4			
Singapore	21	0.4			
South Africa	21	0.4			
Cabo Verde	20	0.4			
Belgium	19	0.3			
Cuba	19	0.3			
Maldives	19	0.3			
Costa Rica	18	0.3			
Hungary	18	0.3			
Russian Federation	16	0.3			
Bermuda	15	0.3			
Cyprus	15	0.3			
Barbados	14	0.3			

Table 4.10 Top destination countries from the United States and top origin countries to the United States, according to the number of trips from and to the United States respectively, from 2018 to 2021 inclusive. The proportions are based on the total number of trips from the United States (31,658) and to the United States (7,098) respectively.

Origin country			**Destination country**		
Country	Total trips	Proportion	Country	Total trips	Proportion
Canada	9063	28.6	Canada	2993	42.2
Mexico	4792	15.1	Mexico	567	8.0
Australia	2017	6.4	Puerto Rico	455	6.4
United Kingdom	1659	5.2	United Kingdom	419	5.9
Sweden	1035	3.3	United States Minor Outlying Islands	215	3.0
Puerto Rico	1008	3.2	Australia	202	2.8
Bahamas	939	3.0	Brazil	192	2.7
Brazil	792	2.5			
China	758	2.4			
Costa Rica	586	1.9			
New Zealand	258	0.8			
Norway	251	0.8			
Ireland	246	0.8			
Japan	234	0.7			
Finland	233	0.7			

Table 4.11 Top destination countries from the United States and origin countries to the United States, according to the number of catches reported by United States users and the number of catches reported as occurring in the United States respectively, from 2018 to 2021 inclusive. The proportions are based on a total of 52,847 catches reported by users from the United States and 26,133 catches reported in the United States.

Origin country			**Destination country**		
Country	Total catches	Proportion	Country	Total catches	Proportion
Canada	16936	32.0	Canada	6525	25.0
Mexico	7322	13.9	Japan	3172	12.1
United Kingdom	3246	6.1	Puerto Rico	2832	10.8
Australia	3233	6.1	United Kingdom	1673	6.4
Puerto Rico	2142	4.1	Mexico	1615	6.2
Sweden	1906	3.6	Brazil	828	3.2
Bahamas	1403	2.7	United States Minor Outlying Islands	801	3.1
Brazil	1112	2.1	Argentina	591	2.3
Costa Rica	1108	2.1	Korea (the Republic of)	588	2.3
China	1020	1.9	Philippines	481	1.8
Belize	488	0.9	Albania	456	1.7
Korea (the Republic of)	465	0.9	Australia	436	1.7
Norway	453	0.9	South Africa	409	1.6
Virgin Islands (U.S.)	448	0.8	Peru	343	1.3
Ireland	447	0.8	France	311	1.2
Japan	423	0.8	Uruguay	276	1.1
New Zealand	414	0.8	Germany	273	1.0
Finland	351	0.7	Bahamas	227	0.9
Cuba	327	0.6	Holy See	223	0.9
Dominican Republic	302	0.6			
South Africa	291	0.6			
Spain	271	0.5			
India	266	0.5			
Denmark	256	0.5			
Nigeria	243	0.5			

Europe

To create network models for Europe, the data were filtered to include only records where both the origin and destination country were in Europe. This resulted in 14,721 catches recorded across 6,454 trips, by 4,417 users.

The top 15 origin countries and top 15 destination countries according to the total number of trips (number of trips in each quarter summed over the 16 quarters of the period 1 January 2018 to 31 December 2021) were selected to be included in the network models. All trips from the top 15 origin countries to the top 15 destination countries were included in the networks. The majority of the top 15 origin countries were also in the top 15 destination countries, resulting in a maximum of 19 unique countries in the network models of the number of trips in Europe each quarter. There were 4,868 trips between the top 15 origin and top 15 destination countries in the period 2018 to 2021, representing 75% of all Europe trips during this period. We used the same approach to determine the countries to create the network models for the number of catches in Europe, which resulted in a maximum of 18 unique countries each quarter in the networks for Europe based on the number of catches recorded. There were 11,641 catches recorded by users from the top 15 origin countries as occurring in the top 15 destination countries in the period 2018 to 2021, representing 79% of all Europe catches during this period.

Table 4.12 shows the top 15 origin countries according to the number of outbound trips from countries in Europe (to destination countries in Europe), and the top 15 destination countries according to the number of inbound trips to countries in Europe (from countries in Europe). Sweden and United Kingdom users made the highest number of outbound trips, while the destination countries receiving the highest number of inbound trips were Sweden and Norway. Table 4.13 shows similar information based on the number of catches recorded. Users from Sweden and the United Kingdom recorded the highest number of catches, while the countries with the highest number of catches recorded as occurring in the country are Sweden, Norway, and the United Kingdom.

Table 4.12 Top 15 origin and destination countries in Europe according to the number of trips from and to these countries respectively, from 2018 to 2021 inclusive. The number of trips from the origin countries are the number of outbound trips from these countries to countries in Europe. The number of trips to the destination countries are the number of trips made to these countries from countries in Europe. Proportions are percentages of the total number of trips (6,454) and are given to 1 decimal place. Totals may not sum to 100% due to rounding.

Origin country			**Destination country**		
Country	Total trips	Proportion	Country	Total trips	Proportion
Sweden	1221	18.9	Sweden	1126	17.4
United Kingdom	1156	17.9	Norway	1042	16.1
Germany	494	7.7	United Kingdom	575	8.9
Poland	429	6.6	France	455	7.0
Denmark	382	5.9	Finland	387	6.0
Ireland	376	5.8	Denmark	353	5.5
Norway	236	3.7	Netherlands	349	5.4
Finland	234	3.6	Spain	296	4.6
France	227	3.5	Ireland	261	4.0
Belgium	201	3.1	Germany	170	2.6
Netherlands	188	2.9	Greece	137	2.1
Switzerland	146	2.3	Croatia	122	1.9
Spain	111	1.7	Poland	118	1.8
Lithuania	109	1.7	Italy	106	1.6
Latvia	107	1.7	Romania	65	1.0
Romania	107	1.7			
Rest of Europe	730	11.3	Rest of Europe	892	13.8
Total	6454	100.0	Total	6454	100.0

Table 4.13 Top 15 origin and destination countries in Europe according to the number of catches reported as occurring in Europe by users from the origin countries, and the number of catches reported as occurring in the destination countries by users from origin countries in Europe, respectively, from 2018 to 2021 inclusive. The proportions are based on 14,721 catches.

Origin country			**Destination country**		
Country	Total catches	Proportion	Country	Total catches	Proportion
Sweden	2477	16.8	Sweden	2658	18.1
United Kingdom	2434	16.5	Norway	2283	15.5
Denmark	1307	8.9	United Kingdom	2100	14.3
Poland	1161	7.9	France	1161	7.9
Germany	1119	7.6	Netherlands	937	6.4
Ireland	984	6.7	Finland	925	6.3
Netherlands	502	3.4	Denmark	603	4.1
Belgium	492	3.3	Spain	547	3.7
France	445	3.0	Ireland	448	3.0
Spain	410	2.8	Greece	402	2.7
Norway	393	2.7	Germany	259	1.8
Finland	390	2.6	Croatia	256	1.7
Italy	349	2.4	Poland	206	1.4
Switzerland	239	1.6	Jersey	163	1.1
Bulgaria	201	1.4	Italy	161	1.1

World countries

We followed a similar approach to select the countries to include in the world country models as for selecting countries for the Europe models. The top 15 origin countries and top 15 destination countries according to the total number of trips (number of trips in each quarter summed over the 16 quarters of the period 1 January 2018 to 31 December 2021) were selected to be included in the network models. All trips from the top 15 origin countries to the top 15 destination countries were included in the networks. The majority of the top 15 origin countries were also in the top 15 destination countries, resulting in a maximum of 20 unique countries in the network models of the number of trips worldwide each quarter. There were 32,854 trips between the top 15 origin and top 15 destination countries in the period 2018 to 2021, representing 65% of the 50,566 worldwide trips during this period. We used the same approach to determine the countries to create the network models for the number of catches worldwide, resulting in a maximum of 21 unique countries each quarter in the worldwide networks based on the number of catches recorded. There were 68,631 catches recorded by users from the top 15 origin countries as occurring in the top 15 destination countries in the period 2018 to 2021, representing 66% of the 104,423 worldwide catches during this period.

Table 4.14 shows the top 15 origin countries according to the number of outbound trips from countries around the world, and the top 15 destination countries according to the number of inbound trips. Users from the United States made the greatest number of outbound trips, followed by users from Canada, the United Kingdom, Sweden, and Australia. The destination countries receiving the highest number of inbound trips were Canada, the United States, and Mexico, followed by Australia, the United Kingdom, and Sweden. In this subset of the data, the vast majority of trips originate from the United States (84%), while a much lower proportion (19%) of trips are to the United States. The reverse pattern is seen for Canada and Mexico, which receive a higher proportion of trips (25% and 14% respectively) compared to the proportion of trips originating from these countries (10% and 2% respectively). In contrast, each of the other countries has a similar proportion of outbound and inbound trips.

Table 4.15 shows the top 15 origin countries according to the number of catches recorded by users from these countries, and the top 15 destination countries according to the number of number of

catches recorded as occurring in these destinations. The highest proportion of catches is recorded by users from the United States (51%), with smaller proportions by users from Canada, the United Kingdom, Japan, Sweden, and Puerto Rico (from 8% to 3%). The destination countries with the highest number of catches recorded as occurring in the country are the United States (25%) and Canada (17%), followed by Mexico, the United Kingdom, Sweden, and Australia (from 7% to 4%).

Comparing Tables 4.14 and 4.15 shows that Japan and Puerto Rico are among the origin countries whose users recorded the highest number of catches (Table 4.15), but are not placed similarly high according to the number of trips made by users from these countries (Table 4.14). In contrast, other countries are generally clustered in similar positions in both tables. This is due to the high number of catches recorded by a minority of users from these countries, resulting in a higher mean number of catches per user compared to other countries (as seen in Table 4.2 and discussed in Section 4.2.1). In contrast, destination countries are in similar rank order in each table, reflecting a more consistent mean number of catches per trip across destinations (as seen in Table 4.3 in Section 4.2.1).

Table 4.14 Top 15 world countries by number of trips from 2018 to 2021 inclusive.

Origin country			Destination country		
Country	Total trips	Proportion	Country	Total trips	Proportion
United States	31 867	84.5	Canada	9497	25.2
Canada	3923	10.4	United States	7109	18.9
United Kingdom	2156	5.7	Mexico	5082	13.5
Sweden	1578	4.2	Australia	2537	6.7
Australia	1190	3.2	United Kingdom	2400	6.4
Germany	620	1.6	Sweden	2230	5.9
Mexico	607	1.6	Norway	1322	3.5
Brazil	547	1.5	Puerto Rico	1028	2.7
Ireland	490	1.3	Bahamas	985	2.6
Poland	469	1.2	China	930	2.5
Puerto Rico	459	1.2	Brazil	915	2.4
Denmark	430	1.1	Costa Rica	644	1.7
France	419	1.1	Finland	638	1.7
South Africa	340	0.9	New Zealand	608	1.6
Finland	299	0.8	France	603	1.6

Table 4.15 Top 15 world countries by number of catches reported from 2018 to 2021 inclusive.

Origin country			Destination country		
Country	Total catches	Proportion	Country	Total catches	Proportion
United States	53 075	50.7	United States	26 187	25.0
Canada	8002	7.6	Canada	17 975	17.2
United Kingdom	5496	5.2	Mexico	7785	7.4
Japan	3361	3.2	United Kingdom	5636	5.4
Sweden	3305	3.2	Sweden	4866	4.6
Puerto Rico	2836	2.7	Australia	4414	4.2
Australia	1993	1.9	Norway	2850	2.7
Mexico	1674	1.6	Puerto Rico	2172	2.1
Germany	1468	1.4	Brazil	1645	1.6
Denmark	1431	1.4	Bahamas	1477	1.4
Brazil	1403	1.3	France	1405	1.3
Poland	1259	1.2	China	1380	1.3
Ireland	1218	1.2	Finland	1330	1.3
France	1054	1.0	Netherlands	1248	1.2
South Africa	907	0.9	Costa Rica	1244	1.2

World regions

The number of trips from an origin region to a destination region per quarter is the sum of the number of trips from countries in the origin region to countries in the destination region per quarter, where the number of trips is the number of unique origin-destination country pairs per user per quarter (as defined in Section 4.2). As the origin and destination countries can be in the same region, the networks based on world regions can have self-loops (edges from a vertex to the same vertex), and the corresponding graphs are therefore multi-graphs (Kolaczyk and Csárdi, 2020). The world regions are shown in Figure 4.1.

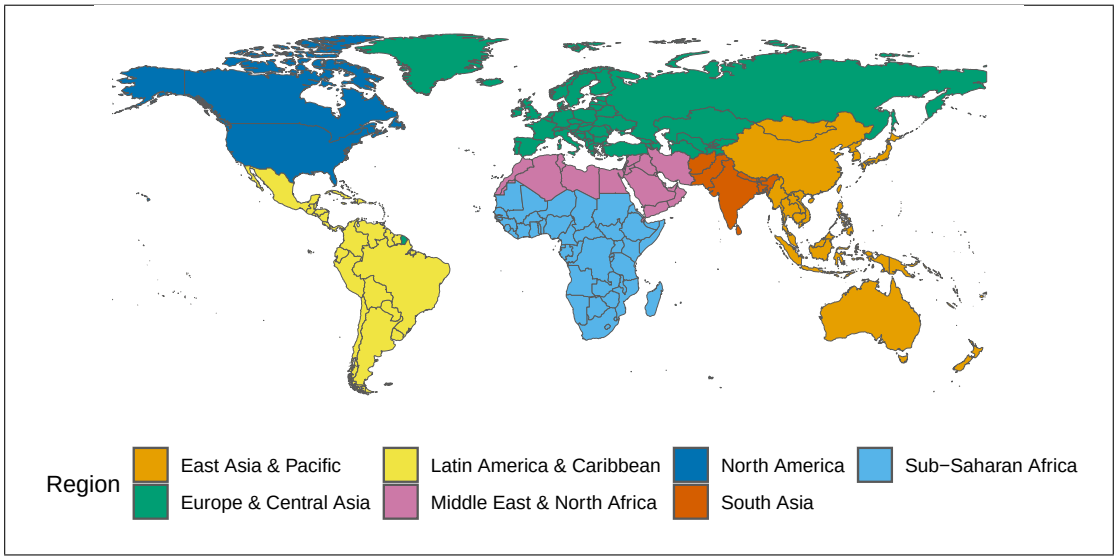


Fig. 4.1 World regions. Data source: `countrycode` package in R (Arel-Bundock et al., 2018).

Table 4.16 shows the proportion of trips to and from each region for the four-year period from 2018 to 2021. These data are illustrated in Figure 4.2. North America has the highest proportion of outbound and inbound trips, although the proportion of outbound trips (70%) is considerably higher than the proportion of inbound trips (33%). In contrast, Europe & Central Asia has a lower proportion of outbound trips (18%) compared to inbound trips (25%). Similarly, outbound trips from the Latin America & Caribbean and East Asia & Pacific regions each comprise 5% of the total trips recorded over the four-year period, although these regions receive 24% and 12% of inbound trips, respectively.

Outbound trips from other regions each comprise 1% or fewer of the total outbound trips, while inbound trips to other regions each comprise 4% or fewer of total inbound trips.

Table 4.17 shows the proportion of catches reported by users from each origin region, and the proportion of catches reported in each destination region, for the four-year period from 2018 to 2021. These data are illustrated in Figure 4.3. The proportion of catches reported by users from North America (58%) is substantially lower than the proportion of outbound trips (70%). This is offset by higher proportions of catches reported than outbound trips from most of the other regions. In contrast, the proportion of catches reported by visitors to the North America region is higher (43%) than the proportion of inbound trips to this region (33%). This is offset by lower proportions of catches reported than inbound trips to all other regions except Europe & Central Asia, which has similar proportions of catches reported by visitors to this region (24%) as inbound trips (25%).

Table 4.18a shows the number of trips from each origin region in each year. The number of trips from the North America region is fairly consistent across the years 2018 to 2020, with a notable increase in 2021. In contrast, the number of trips from the Europe & Central Asia and East Asia & Pacific regions declines in 2020 compared to the previous two years, with a further decline in 2021. Trips from the Latin America & Caribbean region decline gradually over the four-year period.

Table 4.18b shows the number of trips to each destination region in each year. The number of trips to the North America and Latin America & Caribbean regions is fairly consistent in 2018 and 2019, and then decreases in 2020. In 2021, trips to the North America region decline further, while trips to the Latin America & Caribbean region increase slightly. There is a notable increase in the number of trips to Europe & Central Asia from 2018 to 2019, which is not observed in any other region. Trips to this region then decrease substantially in 2020, but recover in 2021. Trips to the East Asia & Pacific region are fairly consistent across the years 2018 to 2020, with a slight increase in 2021.

Table 4.19 shows the number of trips recorded for each origin-destination region pair. These data are illustrated in Figure 4.4a. Top destinations for trips from North America are the North America and Latin America & Caribbean regions, followed by the Europe & Central Asia and East Asia & Pacific regions. Most trips from Europe & Central Asia are to other countries in this region, with North America being the second most popular destination. Top destinations from the East Asia & Pacific region are North America and other countries in the East Asia & Pacific region. A substantial

proportion of trips from the Latin America & Caribbean region are to North America, with relatively few trips to destinations within the Latin America & Caribbean region itself.

Table 4.20 shows the number of catches recorded for each origin-destination region pair. These data are illustrated in Figure 4.4b. The proportion of catches recorded in each region by users from North America is broadly similar to the proportion of trips. In contrast, from the Europe & Central Asia, Latin America & Caribbean, and East Asia & Pacific regions, the proportion of catches recorded in North America is higher than the corresponding proportion of trips. For each of these regions, this is offset by a lower proportion of catches recorded in countries within the region itself.

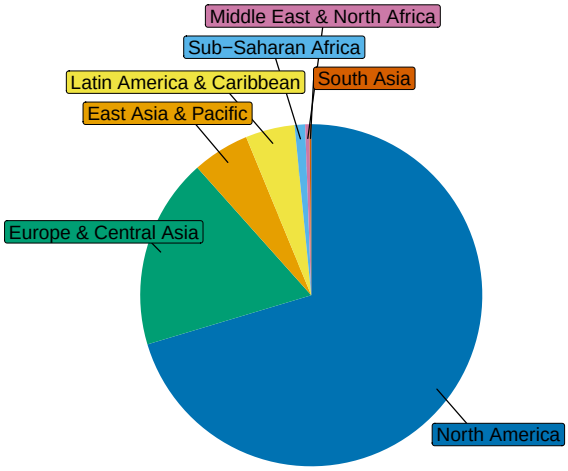
Table 4.21 shows the top countries in each region, by number of catches recorded, from origin regions and to destination regions. The United States and Canada are the dominant countries in the North America region, while Europe & Central Asia has a higher number of countries with more equal proportions of catches. In the other regions, around two or three countries typically account for the highest proportions, with other countries comprising the remainder.

Table 4.16 World regions: Number of trips from 2018 to 2021 inclusive. Proportions are percentages of the total number of trips and are given to 1 decimal place. Totals may not sum to 100% due to rounding. These data are shown in Figure 4.2.

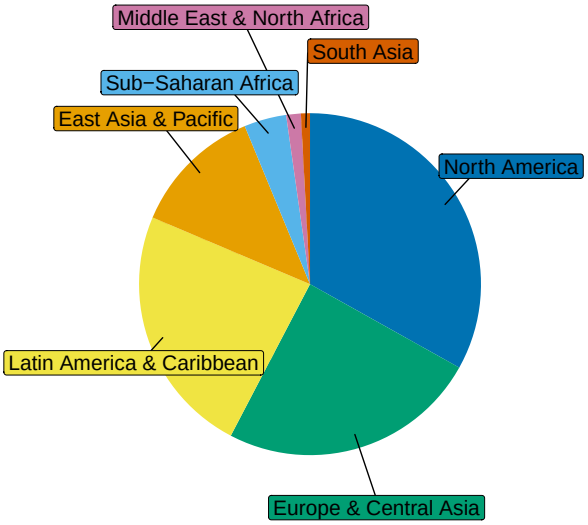
Origin region (outbound trips)			**Destination region (inbound trips)**		
Region	Total trips	Proportion	Region	Total trips	Proportion
North America	35580	70.4	North America	16745	33.1
Europe & Central Asia	9120	18.0	Europe & Central Asia	12415	24.6
East Asia & Pacific	2724	5.4	Latin America & Caribbean	11989	23.7
Latin America & Caribbean	2352	4.7	East Asia & Pacific	6264	12.4
Sub-Saharan Africa	501	1.0	Sub-Saharan Africa	2022	4.0
Middle East & North Africa	204	0.4	Middle East & North Africa	701	1.4
South Asia	85	0.2	South Asia	430	0.9
Total	50566	100.0	Total	50566	100.0

Table 4.17 World regions: Number of catches reported from 2018 to 2021 inclusive. Proportions are percentages of the total number of catches and are given to 1 decimal place. Totals may not sum to 100% due to rounding. These data are shown in Figure 4.3.

Origin region (outbound trips)			**Destination region (inbound trips)**		
Region	Total catches	Proportion	Region	Total catches	Proportion
North America	60849	58.3	North America	44342	42.5
Europe & Central Asia	22962	22.0	Europe & Central Asia	25246	24.2
East Asia & Pacific	9578	9.2	Latin America & Caribbean	19549	18.7
Latin America & Caribbean	9245	8.9	East Asia & Pacific	10968	10.5
Sub-Saharan Africa	1209	1.2	Sub-Saharan Africa	2670	2.6
Middle East & North Africa	346	0.3	Middle East & North Africa	943	0.9
South Asia	234	0.2	South Asia	705	0.7
Total	104426	100.0	Total	104426	100.0

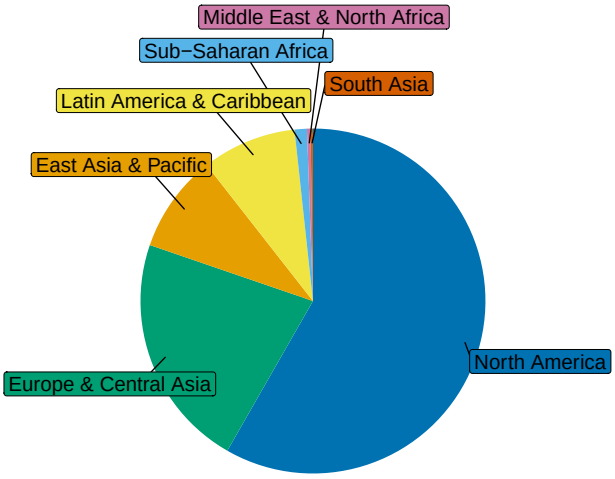


(a) Origin region (outbound trips).

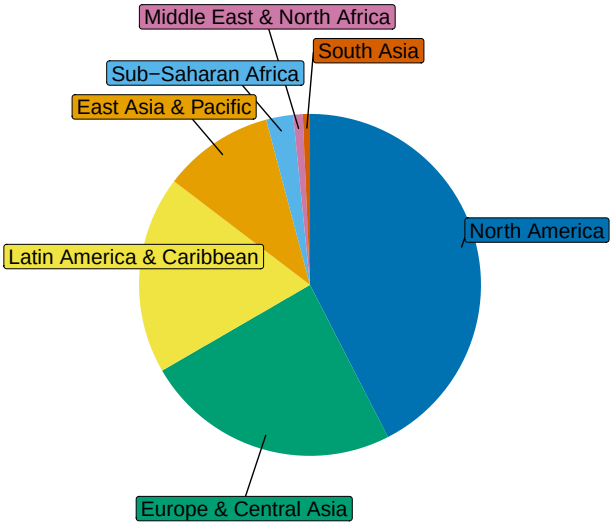


(b) Destination region (inbound trips).

Fig. 4.2 World regions: Number of trips from 2018 to 2021 inclusive. This shows the data given in Table 4.16.



(a) Origin region (catches recorded by users from these regions).



(b) Destination region (catches recorded in these regions).

Fig. 4.3 World regions: Number of catches from 2018 to 2021 inclusive. This shows the data given in Table 4.17.

Table 4.18 Number of trips from each origin region and to each destination region, by year (including trips to other countries within the same region).

(a) Number of trips by origin region.

Origin region	2018	2019	2020	2021	Total
North America	8570	8881	8091	10038	35580
Europe & Central Asia	2432	3695	1754	1239	9120
East Asia & Pacific	895	963	577	289	2724
Latin America & Caribbean	771	649	551	381	2352
Sub-Saharan Africa	153	178	113	57	501
Middle East & North Africa	62	60	50	32	204
South Asia	25	24	18	18	85
Total	12908	14450	11154	12054	50566

(b) Number of trips by destination region.

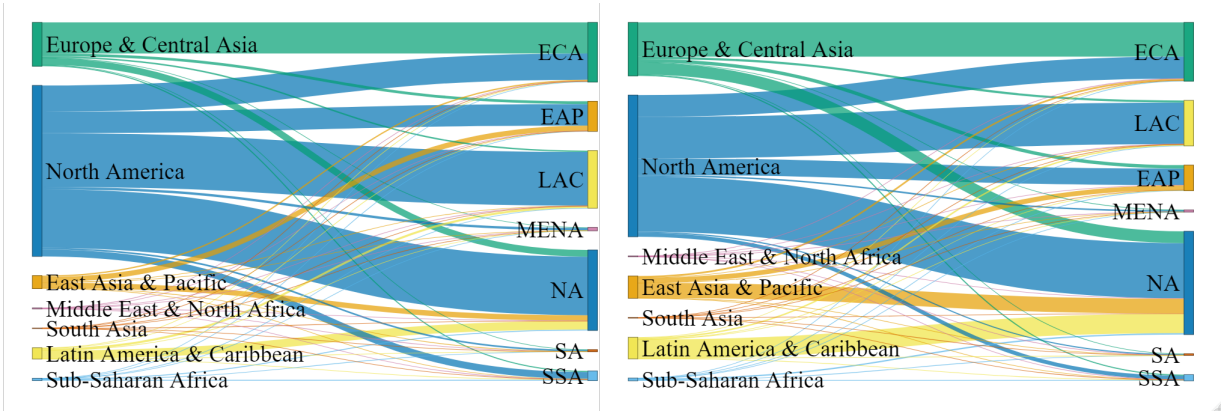
Destination region	2018	2019	2020	2021	Total
North America	4790	4894	3645	3416	16745
Europe & Central Asia	2395	3559	2804	3657	12415
Latin America & Caribbean	3374	3668	2353	2594	11989
East Asia & Pacific	1560	1541	1476	1687	6264
Sub-Saharan Africa	515	512	554	441	2022
Middle East & North Africa	138	152	229	182	701
South Asia	136	124	93	77	430
Total	12908	14450	11154	12054	50566

Table 4.19 Number of trips for each origin-destination region pair 2018-2021. This information is shown in Figure 4.4 (a).

	Destination							
Origin	EAP	ECA	LAC	MENA	NA	SA	SSA	Total
East Asia & Pacific (EAP)	1089	283	74	30	1120	21	107	2724
Europe & Central Asia (ECA)	578	6454	300	114	1392	54	228	9120
Latin America & Caribbean (LAC)	51	87	434	2	1739	1	38	2352
Middle East & North Africa (MENA)	21	47	5	46	53	12	20	204
North America (NA)	4478	5468	11165	500	12204	335	1430	35580
South Asia (SA)	16	13	0	6	47	3	0	85
Sub-Saharan Africa (SSA)	31	63	11	3	190	4	199	501
Total	6264	12415	11989	701	16745	430	2022	50566

Table 4.20 Number of catches recorded for each origin-destination region pair 2018-2021. This information is shown in Figure 4.4 (b).

	Destination							
Origin	EAP	ECA	LAC	MENA	NA	SA	SSA	Total
East Asia & Pacific (EAP)	2086	746	137	46	6408	27	128	9578
Europe & Central Asia (ECA)	1469	14721	926	154	5091	137	464	22962
Latin America & Caribbean (LAC)	114	180	620	3	8284	1	43	9245
Middle East & North Africa (MENA)	24	85	5	55	140	17	20	346
North America (NA)	7100	9385	17848	671	23681	502	1662	60849
South Asia (SA)	55	19	0	11	145	4	0	234
Sub-Saharan Africa (SSA)	120	110	13	3	593	17	353	1209
Total	10968	25246	19549	943	44342	705	2670	104423



(a) Number of trips. The data is in Table 4.19. The total number of trips is 50,566.

(b) Number of catches recorded. The data is in Table 4.20. The total number of catches is 104,423.

Fig. 4.4 Number of trips and catches recorded for each origin-destination region pair, for the years 2018-2021.

Table 4.21 Number of catches recorded by users from each country within each region 2018-2021.

Region	Origin country	Catches	Destination country	Catches
North America	United States	52 847	United States	26 133
	Canada	7997	Canada	17 973
	Other	5	Other	236
	Total	**60849**	**Total**	**44342**
Europe & Central Asia	United Kingdom	5491	United Kingdom	5636
	Sweden	3305	Sweden	4866
	Germany	1468	Norway	2850
	Denmark	1431	France	1405
	Poland	1259	Finland	1330
	Other	10 008	Other	9159
	Total	**22962**	**Total**	**25246**
East Asia & Pacific	Japan	3361	Australia	4409
	Australia	1976	China	1380
	United States Minor Outlying Islands	812	New Zealand	1053
	Republic of Korea	677	Republic of Korea	721
	Philippines	525	Thailand	536
	Other	2227	Other	2869
	Total	**9578**	**Total**	**10968**
Latin America & Caribbean	Puerto Rico	2836	Mexico	7785
	Mexico	1674	Puerto Rico	2172
	Brazil	1399	Brazil	1644
	Argentina	686	Bahamas	1477
	Peru	357	Costa Rica	1244
	Other	2293	Other	5227
	Total	**9245**	**Total**	**19549**
Sub-Saharan Africa	South Africa	906	South Africa	413
	Nigeria	72	Nigeria	267
	Zimbabwe	58	Mozambique	157
	Kenya	48	Ghana	116
	Mozambique	30	Mali	115
	Other	95	Other	1602
	Total	**1209**	**Total**	**2670**
Middle East & North Africa	United Arab Emirates	94	United Arab Emirates	190
	Israel	50	Egypt	147
	Syrian Arab Republic	44	Saudi Arabia	112
	Palestine, State of	39	Algeria	90
	Egypt	23	Libya	62
	Other	96	Other	342
	Total	**346**	**Total**	**943**
South Asia	Afghanistan	129	India	313
	Nepal	39	Maldives	142
	India	35	Sri Lanka	125
	Bangladesh	11	Nepal	45
	British Indian Ocean Territory	9	Pakistan	32
	Other	11	Other	48
	Total	**234**	**Total**	**705**

4.2.2 Networks and graphs

Network models

Network models provide a valuable framework for modelling interactions between entities, in this case countries (or regions), in complex systems (Kolaczyk and Csárdi, 2020). For our cases, countries (or regions) are represented by vertices (nodes) and travel between vertices is represented by edges. Edge weights (or strengths) quantify the intensity or frequency of interactions between connected vertices (Newman, 2010), in this case, of travel for fishing purposes between vertices.

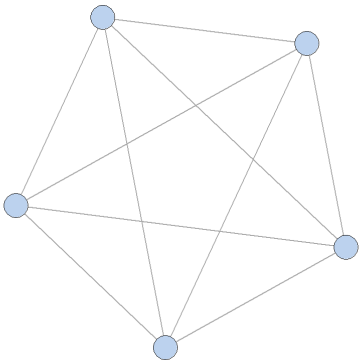
Directed graphs

Networks are typically represented by graphs, with summary statistics describing the network or individual entities and connections obtained from the graph (Kolaczyk and Csárdi, 2020). Graphs are structures that consist of vertices to represent entities and edges to represent connections or associations between these entities, where edges can be 'unweighted' or 'weighted' (Newman, 2010). We summarise connections between vertices using directed graphs (digraphs), which are graph structures consisting of vertices and directed edges (arcs) that indicate a one-way relationship between vertices (Kolaczyk and Csárdi, 2020; Newman, 2010). Therefore, directed graphs incorporate the notion of flow or directionality between connected vertices, which in this case indicates the direction of travel of anglers whose device is registered in the 'from' vertex (origin country or region) and have recorded catches in the 'to' vertex (destination country or region) (Newman, 2010).

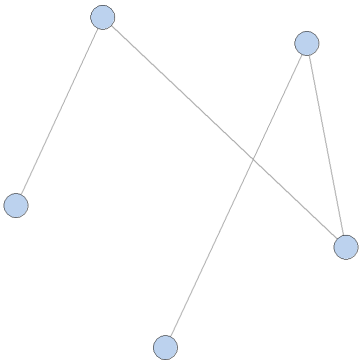
Directed graphs were created using the graph_from_data_frame() function from the R package `igraph` (Csardi and Nepusz, 2006) with the interactions between vertices as weighted edges, where the weights are equal to the number of observed interactions between each pair of vertices. For this research, we used the number of trips and the number of catches recorded as interactions between the vertices.

Network metrics

There are many ways to compare networks (Tantardini et al., 2019; Wills and Meyer, 2020). We are interested in comparing travel over time, and since we are representing travel as connections between



(a) Graph with edge density = 1, mean distance = 1 and diameter = 1.



(b) Graph with edge density = 0.2, mean distance = 2 and diameter = 4.

Fig. 4.5 Example graphs.

vertices, the 'edge density', a measure of the density of connections within a network (Newman, 2010), provides a comparison of the number of routes travelled in different time periods. The edge density is the ratio of the number of edges in the network and the maximum number of possible edges between all vertex pairs (Schaeffer, 2007). In a directed network, the maximum number of edges is $n(n-1)$, where n is the number of vertices in the network. In an undirected network, the maximum number of edges is $\frac{n(n-1)}{2}$. In networks that allow self-loops from a vertex to the same vertex, the maximum number of vertices is $n(n-1)+n$ for directed networks and $\frac{n(n-1)}{2}+n$ for undirected networks.

Since the edge density describes the density of connections in the graph, lower values indicate fewer connections (a sparser network). This is demonstrated in Figure 4.5, where the more densely connected graph, graph (a), has an edge density of 1, and the less densely connected graph, graph (b), has an edge density of 0.2. For our fishing tourism cases, a lower edge density indicates fewer 'connected' vertices (where an edge between the 'from' vertex and the 'to' vertex exists, indicating travel from the country (or region) represented by the origin vertex to the country (or region) represented by the destination vertex). For a given case (for example, travel within Europe), time periods with lower edge densities have fewer origin-destination pairs, meaning fewer routes were travelled, compared to time periods with higher edge densities.

As the edge density is dependent on the number of vertices in the network, it does not provide a comparable measure of the density of networks across different time periods when the number of vertices in the networks are different. In our cases, networks across the time periods have different numbers of vertices, as countries (or regions) are excluded from networks when there are no edges to and from them, as a result of the absence of travel to and from the country (or region) in that time period. However, these edges and vertices are possible edges within the network and not accounting for them within the edge density metric results in a higher edge density value than if these are considered as potential edges that are not present. Networks with fewer edges can be more dense than networks with more edges, if the network with more edges also has more vertices.

Therefore, we introduce a modified edge density based on the maximum number of vertices in a network, hereafter referred to as 'edge density (max vertices)'. We define the edge density (max vertices) as the ratio of the number of edges in the network and the maximum number of possible edges if all vertices are present in the network. For example, in a network with a potential maximum of $n = 12$ vertices, the number of possible edges when all vertices are present is $n(n-1) = 132$. The 'edge density (max vertices)' is calculated as the actual number of edges in the network divided by 132, even if the network contains fewer than 12 vertices.

The 'diameter' and 'mean distance' of a network can also give an indication of the network density. Both these metrics are based on distances within the network. The 'distance' between a pair of vertices is the length of the shortest path (geodesic) between them, where a path is a route between two distinct vertices that contains no repeated edges or vertices, the length of a path is the number of edges in the path, and the shortest path between two vertices is the minimum length path between them (Kolaczyk and Csárdi, 2020). The diameter is the greatest distance between any pair of vertices (Newman, 2010). Since its value is determined by just two vertices, the diameter may be impacted by small changes within the network (Newman, 2010). Furthermore, since it is based on the maximum shortest distance in the network, it may not provide a good representation of typical shortest path distances in the network (Newman, 2010). The mean distance (or 'average distance') is the mean of the shortest path between every pair of connected vertices (Scardoni and Laudanna, 2012), and therefore based on information from the entire network rather than only two values. In

directed graphs, the shortest path in both directions between each pair of vertices is used to calculate the mean distance.

Both the diameter and mean distance can provide an indication of the density of a network, with low values associated with dense networks and higher values with sparser networks (Scardoni and Laudanna, 2012), since shorter paths between vertices tend to arise from a higher density of connections within the network. This is demonstrated in Figure 4.5, where the more densely connected graph, graph (a), has a lower diameter and mean distance (both 1), while the less densely connected graph, graph (b), has a higher diameter (4) and mean distance (2). However, this is not always the case, as high values can arise in dense networks or, in the case of the diameter, in networks containing dense sub-networks (Scardoni and Laudanna, 2012). Furthermore, meaningful comparisons require consideration of the number of vertices in the networks being compared, as networks with a small number of vertices will naturally have a low diameter and mean distance (Scardoni and Laudanna, 2012).

Since paths through the networks do not have an interpretable meaning in the context of this analysis of travel for fishing purposes, the diameter and average distance provide an abstract comparison of the density of connections in networks with a similar numbers of vertices. Therefore, in our cases, where the number of countries (or regions) in networks are comparable, a network with a lower diameter or mean distance suggests more connected countries (or regions) compared to a network with a higher diameter or mean distance.

The diameter and mean distance can be calculated as 'unweighted' where each edge has an equal weight (usually 1), or 'weighted' which is calculated by summing the edge weights (Csardi and Nepusz, 2006). For our analysis, the amount of travel between vertices (edge weights) is informative only in the context of directly connecting two vertices. The sum of edge weights on paths containing more than two vertices is meaningless, as users can travel directly from one vertex to another. Therefore, the weighted diameter and mean distance are not relevant in our cases. However, the unweighted diameter and mean distance may provide an indication of the density of connections within our networks.

Communities

Real-world networks often contain groups of vertices that are more densely interconnected with one another than with the rest of the network (Fortunato, 2010). These dense subgroups are known as communities or clusters (Schaeffer, 2007). In contrast, random graphs typically lack such community structure, displaying a more uniform distribution of connections (Fortunato, 2010). Vertices within clusters often share similar characteristics or behaviours, and identifying clusters in real-world networks can reveal insights into patterns of interaction and influence within the network (Fortunato, 2010).

There are many ways of defining clusters (Fortunato, 2010; Schaeffer, 2007), though vertices within a cluster should be connected by at least one path entirely within the cluster (Schaeffer, 2007). Clustering algorithms aim to reveal community structure in networks by identifying dense subgraphs. Effective clustering results in clusters with a higher edge density within each cluster compared to the overall graph density, and a lower density of edges between clusters compared to the overall graph density (Schaeffer, 2007). In weighted graphs, the sum of edge weights can be considered instead of the number of edges, so that higher density clusters contain numerous or strongly-weighted (potentially influential) edges while lower density clusters contain few or lower-weighted edges (Schaeffer, 2007).

Graph clustering methods encompass a diverse range of approaches (Fortunato, 2010; Lancichinetti and Fortunato, 2009; Schaeffer, 2007). These include hierarchical techniques that iteratively merge or split clusters to create a hierarchy of communities at different levels of granularity (Fortunato, 2010; Schaeffer, 2007), partitioning approaches that divide the data into a predetermined number of clusters (Fortunato, 2010), methods based on statistical inference (Fortunato, 2010), spectral techniques utilising eigenvalues of graph matrices (Fortunato, 2010; Newman, 2013; Schaeffer, 2007), and optimisation-based methods that maximise measures of cluster quality such as modularity (Fortunato, 2010) and betweenness (Newman and Girvan, 2004). Overlap between these categories is common. For instance, some hierarchical methods also employ optimisation techniques, as used in the greedy optimisation of modularity (Clauset et al., 2004), while modularity can also be optimised through spectral methods (Fortunato, 2010). Furthermore, the edge betweenness method (Newman and Girvan,

2004) can be categorised as both a partitioning method and an optimisation-based method. Vertices can be allocated to clusters based on their properties, such as similarity or distance in attributes relevant to the network context or the graph structure, or according to the degree of connectivity between vertices (Schaeffer, 2007). Finally, while cluster quality measures are utilised to evaluate clusters in optimisation methods, they can also be used to compare different clustering outcomes.

The `igraph` package (Csardi and Nepusz, 2006) in R contains several community detection algorithms. We initially assessed four algorithms, which use different approaches: 'edge betweenness' as implemented in the R function cluster_edge_betweenness() (Newman and Girvan, 2004), 'propagating labels' as implemented in the R function cluster_label_prop() (Raghavan et al., 2007), 'greedy optimisation of modularity' as implemented in the R function cluster_fast_greedy() (Clauset et al., 2004), and the 'Louvain method' as implemented in the R function cluster_louvain() (Blondel et al., 2008).

Betweenness algorithms iteratively identify and remove edges with the highest 'edge betweenness'. This measure gives more weight to edges between clusters and less weight to edges within clusters, effectively distinguishing between edges that link communities and edges within communities (Newman and Girvan, 2004). Consequently, edges between clusters have higher betweenness, while those within clusters have lower betweenness (Newman and Girvan, 2004). There are many measures of betweenness, including measures based on shortest paths, random walks and current flow (Newman and Girvan, 2004). Shortest path measures are straightforward and efficient to implement and have been shown to be equally effective as some more computationally complex methods (Newman and Girvan, 2004). The edge betweenness algorithm in the `igraph` package (Csardi and Nepusz, 2006) uses the number of shortest paths to calculate betweenness.

The propagating labels algorithm, as proposed by Raghavan et al. (2007), offers a computationally efficient approach to community detection. This method focuses solely on the network vertices, without assessing community strengths or implementing an optimisation process (Raghavan et al., 2007). The algorithm, as implemented in the `igraph` package (Csardi and Nepusz, 2006), initially assigns labels to vertices according to their current community. Each iteration re-labels each vertex with the most common community among its neighbours, until labels stabilise (Raghavan et al., 2007).

Optimising the modularity is potentially the most popular clustering method (Newman, 2013). Modularity is a measure of the quality of clusters in a network, based on the network structure and resulting communities (Clauset et al., 2004; Newman, 2004). It quantifies the degree to which a network division results in a high proportion of edges within clusters compared to edges between clusters (Clauset et al., 2004; Schaeffer, 2007). Higher modularity scores indicate more edges within communities and fewer edges between communities (Clauset et al., 2004). The modularity is zero if the proportion of internal edges within clusters is no greater than in a randomised network (Clauset et al., 2004).

Different ways of representing modularity have been proposed (Newman, 2006). One common representation of modularity is the fraction of edges within communities compared to the expected fraction in a random model with the same number of vertices (Clauset et al., 2004; Newman, 2006; Newman and Girvan, 2004). Due to the computational complexity of evaluating all possible cluster configurations, many good approximation approaches to maximise the modularity have been developed (Newman, 2006). These include greedy approaches and methods based on eigenvectors of the modularity matrix, which is derived from the network structure and is not affected by clustering outcomes (Newman, 2006). The greedy optimisation of modularity algorithm, introduced by Newman (2004) and computationally enhanced by Clauset et al. (2004), offers a fast and widely applicable method for identifying communities within a network (Clauset et al., 2004). This agglomerative algorithm initially treats each vertex as a separate community and then iteratively merges communities. It employs a greedy approach, selecting the best option to maximise the modularity measure at each step, based solely on the current state of the network at that iteration (Clauset et al., 2004; Newman, 2004). Modularity scores above 0.3 typically indicate significant community structure (Clauset et al., 2004; Newman, 2004). It is a commonly-used (Kolaczyk and Csárdi, 2020) and effective (Newman, 2006) algorithm for identifying communities in networks.

The Louvain method, introduced by Blondel et al. (2008), is a multi-level optimisation approach. It is an iterative process, with the implementation in the `igraph` package (Csardi and Nepusz, 2006) based on optimising modularity. Each iteration consists of two stages. The first stage optimises modularity by moving individual nodes between communities to obtain a local maxima for the

modularity, with these first-stage communities then used as nodes in a new network in the second stage. The two stages are repeated until maximum modularity is achieved (Blondel et al., 2008).

We assessed all four algorithms using both weighted and unweighted graphs. The propagating labels, greedy optimisation of modularity, and Louvain algorithms require undirected networks. We constructed undirected networks using the as.undirected() function from the `igraph` (Csardi and Nepusz, 2006) package with argument 'mode = "collapse" ' to create exactly one undirected edge between each pair of vertices that have at least one directed edge in the directed graph. This approach avoids multiple edges in the undirected graph, with the weight of the undirected edge being the sum of the weights of the directed edge(s) it replaced. Using undirected graphs to identify communities assumes that entities which are connected in any direction are similar (Clauset et al., 2004). The edge betweenness algorithm accepts directed and undirected networks and was assessed with all combinations of directed/undirected and weighted/unweighted networks. For this algorithm, the directed graph was converted to an undirected graph using the same method as for the other algorithms.

The greedy optimisation of modularity and Louvain algorithms resulted in several clusters, most of which contained multiple vertices. In contrast, the edge betweenness and propagating labels algorithms tended to produce either a single cluster containing all (or most) vertices or many clusters, each containing a single vertex. The greedy optimisation of modularity and Louvain algorithms with weighted graphs generally produced fewer, more geographically-defined clusters. In contrast, the unweighted graphs exhibited more complex patterns characterised by a slightly greater number of clusters and less regional clustering, resulting in greater geographical overlap. The modularity scores of the weighted networks were higher overall than those of the unweighted networks.

Since paths through the network are not meaningful in our context, the betweenness approach may be less suitable for identifying clusters in our networks. Additionally, the propagating labels algorithm focuses on vertices, whereas our interest is in the travel between vertices, represented by edges. Therefore, the betweenness and propagating labels algorithms are less relevant to our analysis compared to the greedy optimisation of modularity algorithm and Louvain method, which may contribute to the weaker cluster patterns they generated.

We chose to identify communities using weighted networks with the greedy optimisation of modularity algorithm proposed by Clauset et al. (2004), as implemented in the R function clus-

ter_fast_greedy() from the `igraph` package (Clauset et al., 2004; Csardi and Nepusz, 2006), due to its simplicity, applicability to our context, and informative results it generates.

4.3 Results

In this section, we present and describe the results obtained based on the number of trips. We show the results based on the number of catches recorded in Appendix C. These show that the identified patterns based on the number of trips and number of catches are similar.

First we present the results of the individual country case studies, then look at networks of countries within Europe and around the world, and finally look at networks across world regions. For all these cases we present directed graphs, by quarter, for the years 2018 to 2021, summarising the asymmetric interactions between vertices, along with the corresponding network metrics described in Section 4.2.2. For the networks within Europe, around the world, and across world regions we also look at the community structure within the networks. We provide community graphs showing the structure of the communities in the weighted undirected graphs identified using the greedy optimisation of modularity algorithm (Clauset et al., 2004) implemented with the R function cluster_fast_greedy() in the `igraph` package (Clauset et al., 2004; Csardi and Nepusz, 2006), as described in Section 4.2.2, along with a frequency table of the number of quarters in which each pair of countries has been allocated to the same community.

In the network graphs, the area of each vertex represents the total number of trips from that origin to all destinations worldwide (except Antarctica). This includes destinations removed when filtering to the top countries (in the case of the Europe and world countries networks) or number of trips or catches (for the case study countries). The vertex colour represents the origin region. As the network graphs are based on directed networks, each edge represents travel from the origin vertex to the destination vertex, with the direction of travel indicated by an arrow from the origin to the destination. The width of the edges represents the number of trips from the origin vertex to the destination vertex. For the networks based on the number of catches, the edge widths represent the number of catches recorded as occurring in the destination vertex by users whose device is registered in the origin vertex. Edge loops in the graphs of world regions represent travel to a destination country in the same region

as the origin country. In the network graphs for Europe, world countries, and world regions, the edge colour represents the world region of the origin vertex. In the network graphs for the individual case study countries, the edge colour represents travel to or from the case study country. Within the network graphs for each case, the area of the vertices and width of the edges are scaled consistently across all 16 quarters.

As the community graphs are based on the undirected networks, the edges represent travel in either or both directions between the pair of vertices they link. Red edges in the community graphs represent travel between vertices in different clusters, and black edges represent travel between vertices in the same cluster. Vertex colours indicate the cluster each country (or region) is allocated to, and are assigned automatically by the cluster_fast_greedy() function (Clauset et al., 2004).

The position of the vertices representing the countries (or regions) in all network and community graphs approximately matches their geographical location, to provide a visual representation of travel patterns across the world. Vertices representing countries are labelled with country codes, with the corresponding country names and regions listed in Table C.1 in Appendix C. Vertices representing regions are labelled with region codes, with the corresponding region names in the figure caption.

For each case, we describe the seasonal patterns and trends over time, in terms of the number of connections and the strength of connections, where the number of connections is the number of origin-destination pairs connected by travel, and the strength of connections is represented by the number of trips between origin-destination pairs.

In all cases, due to the small network sizes, the computational time required to fit the network model and identify communities for each quarter was negligible (i.e. less than one second). Model fitting and community identification were performed using R version 4.2.2 (R Core Team, 2022) with the `igraph` (version 1.5.0) (Csardi and Nepusz, 2006) package on a personal computer running Windows 10.

4.3.1 Case study countries

For each case study country, we present the network metrics and directed graphs summarising the asymmetric interactions between vertices, where the width of the edges represents the number of trips.

For the case studies, the mean distance and diameter of the networks are not useful, as in each case the network is based only on travel to and from the case study country, so there is no possibility for connections between pairs of countries where the case study country is not either an origin or destination country within the pair. Therefore, the diameter of the network is always 2. The mean distance is linked to the number of countries (vertices) in the network since each country (except the case study country itself) will have a path of length 1 to the case study country, and path of length 2 to every other country in the network (via the case study country). Community networks are also not relevant, as these models represent only connections to and from a single country.

United Kingdom

We created network models representing the most popular travel routes to and from the United Kingdom, by including only the edges representing at least five trips from the origin vertex to the destination vertex each quarter, along with the associated vertices, where the United Kingdom is either the origin or destination vertex.

Table 4.22 presents the network metrics and Figure 4.6 presents the directed graphs summarising the number of trips by users from the United Kingdom to other countries and by users from other countries to the United Kingdom. Since edges representing fewer than five trips are excluded, the graphs only show the connections where five or more users from the origin country recorded at least one catch in the destination country in that quarter.

Seasonal patterns

Number of connections There are seasonal patterns in the number of connected countries in 2018 and 2019. The network graphs (Figure 4.6) show that in these years quarter 3 has the most densely connected networks (highest number of edges), while quarter 1 has the least densely connected networks. This is reflected in the network metrics (Table 4.22), with quarter 3 having the highest edge density (max vertices) and quarter 1 having the lowest edge density (max vertices) compared to the other quarters. This shows that there are more connections between the United Kingdom and other

countries during the United Kingdom summer months and fewer during the United Kingdom winter months, compared to the rest of the year.

Number of trips There are also seasonal patterns in the number of trips. Edge widths in 2018 and 2019 from the United Kingdom to European countries tend to be wider in quarter 3 compared to other quarters, while edge widths to Australia are wider in quarters 1 and 4, indicating a preference to travel during the warmest months of the destination country.

Trends over time

Number of connections There is a decrease in the number of connected countries from around 2020 quarter 3 onwards (Figure 4.6). In 2020 quarter 2, the number of connections is lower compared to 2019 quarter 2, and the same as 2018 quarter 2. From 2020 quarter 3 onwards, there is a strong reduction in the number of vertices and edges in the networks, and the edge density (max vertices) is lower in all quarters during this period, compared to corresponding quarters in previous years (Table 4.22). This shows a sharp drop in the number of connections between countries in all quarters from 2020 quarter 3 onwards, compared to the corresponding quarters previously.

Number of trips The seasonal pattern in the number of trips is evident across the four-year period, with an increasing trend in the number of trips over the four-year period. The number of trips drops in 2020 quarters 2 and 3 compared to previous years, but then increases above 2018 and 2019 levels in all quarters from 2020 quarter 4 onwards (Table 4.22).

Prior to 2020 quarter 2, the top outbound destinations in terms of number of trips (widest edges) from the United Kingdom are the United States and France, followed by Ireland, Australia and Spain. All these countries, except Spain, remained top destinations from 2020 quarter 2 onwards, although the number of trips reduced to around half of previous levels. Prior to 2020 quarter 2, the highest number of inbound trips to the United Kingdom were from the United States and Ireland. In contrast to the decrease in the number of outbound trips from the United Kingdom, the number of inbound trips from the United States and Ireland increased in the period from 2020 quarter 2 onwards, compared to the previous period.

Table 4.22 Metrics for the United Kingdom network graphs based on the number of trips. In each year-quarter time period, the networks contain only the edges where there are at least five trips for that origin-destination country pair.

Year	Quarter	Number of vertices	Number of edges	Edge density (max vertices)	Number of trips
2018	1	5	6	0.022	72
2018	2	8	9	0.033	173
2018	3	17	21	0.077	337
2018	4	10	11	0.040	106
2019	1	5	6	0.022	75
2019	2	16	17	0.062	283
2019	3	17	19	0.070	390
2019	4	8	11	0.040	114
2020	1	6	8	0.029	96
2020	2	7	9	0.033	158
2020	3	7	8	0.029	270
2020	4	4	4	0.015	126
2021	1	4	4	0.015	86
2021	2	3	4	0.015	364
2021	3	7	8	0.029	574
2021	4	5	6	0.022	259

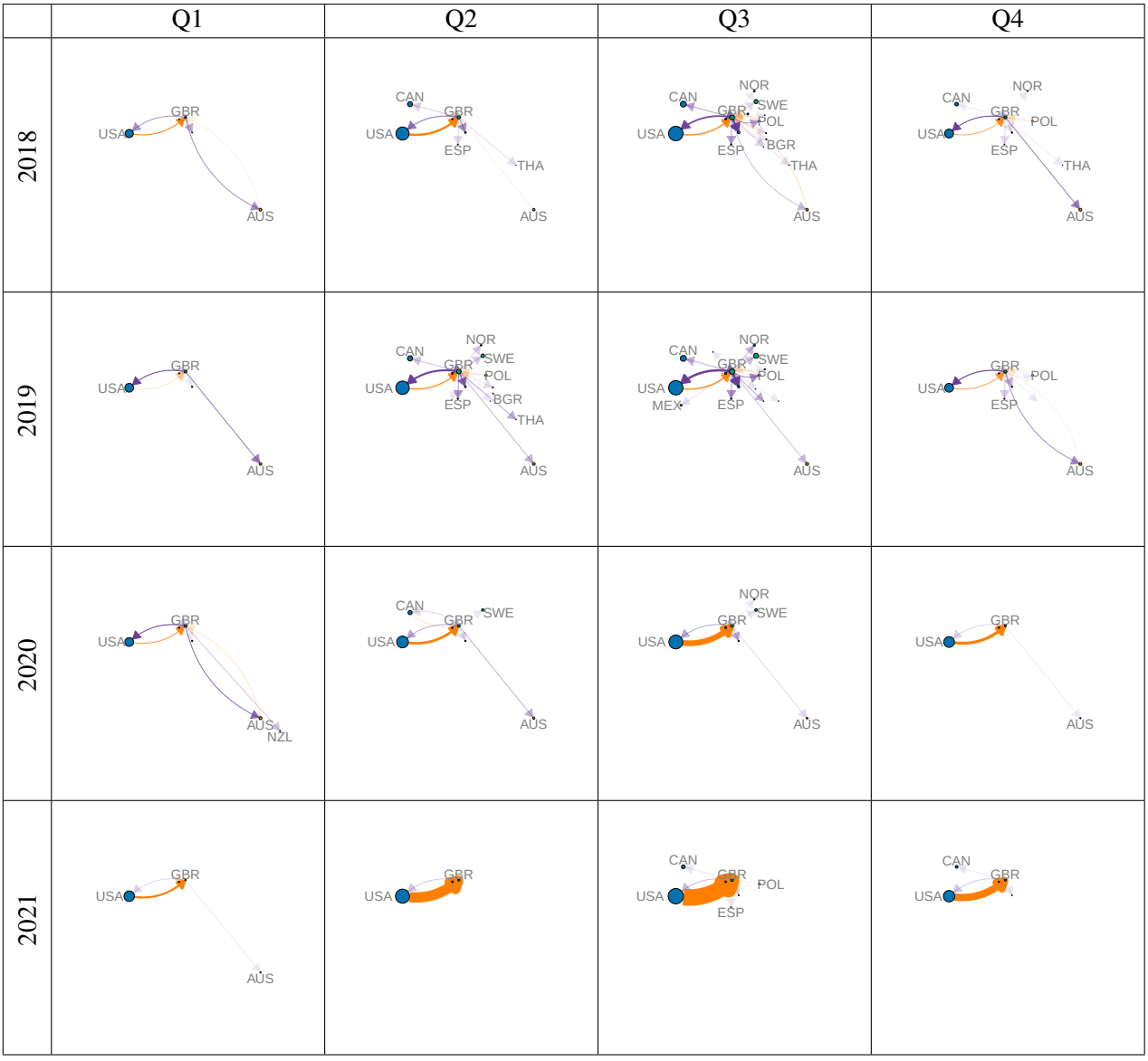


Fig. 4.6 Network graphs of the trips to and from the United Kingdom. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than five trips are excluded.

United States

We created network models representing the most popular travel to and from the United States, by including only the edges representing at least 30 trips from the origin to destination vertex each quarter, along with the associated vertices, where the United States is either the origin or destination vertex.

Table 4.23 presents the network metrics and Figure 4.7 presents the directed graphs summarising the number of trips by United States users to other countries and by users from other countries to the United States. Edges representing fewer than 30 trips are excluded. Therefore, the graphs only show the connections where 30 or more users from the origin country recorded at least one catch in the destination country in that quarter.

Seasonal patterns

Number of connections. There are seasonal patterns in the number of connected countries throughout the four-year period. The network graphs (Figure 4.7) and edge density (max vertices) (Table 4.23) show that quarters 2 and 3 have the highest number of edges and most densely connected networks, while quarters 1 and 4 have the least densely connected networks. This indicates that there are more connections between the United States and other countries during the spring and summer months of the United States, and fewer during their autumn and winter. Connections to countries in the Latin America & Caribbean and East Asia & Pacific (Australia and China) regions and to Canada are evident throughout the year, while connections to Europe & Central Asia are predominantly in quarters 2 and 3.

Number of trips. There are also seasonal patterns in the number of trips between connected countries. The number of trips (Table 4.23) is highest in quarter 3, followed by quarter 2. Quarters 1 and 4 have a much lower number of trips, with the lowest in quarter 4. Seasonal patterns are only apparent for Canada and Mexico. Edge widths on the network graphs (Figure 4.7) from the United States to Canada and Mexico are wider in quarters 2 and 3 than in quarters 1 and 4. In contrast, edge widths to the United States from Canada are fairly consistent across the year. This indicates that there

are more trips from the United States to Canada and Mexico in spring and summer than in autumn and winter, while in contrast there is no clear seasonal pattern in the number of trips to the United States from Canada. There are no clear seasonal patterns in the number of trips to other destinations, though the number of trips to other destinations is relatively low compared to the number to Canada and Mexico.

Trends over time

Number of connections The number of connections (number of edges) remains fairly consistent over time, with the seasonal pattern evident across the four-year period. There is a slight increase in the number of edges, vertices and edge density in 2021 quarters 1, 2 and 3, compared to previous years, showing an increase in the number of popular countries and connections between countries during these quarters (Table 4.23).

In contrast, there is a change in the pattern of connections from 2020 quarter 2 onwards (Figure 4.7). The number of connections to Scandinavian countries increases, while connections to China cease. Popular destinations from the United States to Latin America & Caribbean countries also change from 2020 quarter 2 onwards. Costa Rica and Brazil are rarely both in the network in the same quarter, with Costa Rica tending to be in the networks prior to 2020 quarter 2, and Brazil tending to be in the networks from 2020 quarter 2 onwards. Prior to 2020 quarter 2, there was a connection to Costa Rica every quarter (at least 30 trips per quarter), while Brazil was generally absent from the networks (fewer than 30 trips per quarter). From 2020 quarter 2 until 2021 quarter 1 inclusive, this pattern reversed, with a connection to Brazil in every quarter during this period, while Costa Rica was absent from the networks. Interestingly, there is a large number of trips to Brazil in 2021 quarter 4. Both Costa Rica and Puerto Rico are absent from the networks in this quarter, which is the only quarter without a connection from the United States to Puerto Rico.

Number of trips The number of trips remains consistent from 2018 to 2019, but has an overall increasing trend during 2020 and 2021, apart from a decrease in the number of trips in 2020 quarters 2 and 3 compared to other years (Table 4.23). The seasonal pattern in the number of trips is evident across the four-year period.

Prior to 2020 quarter 2, the top outbound destinations (widest edges in Figure 4.7) from the United States are Canada and Mexico. Other destinations with a moderate number of trips are Puerto Rico, the Bahamas, China, Australia, and Costa Rica. From 2020 quarter 2 onwards, Canada and Mexico remained top destinations, with the number of trips reducing by about 20%. Trips to the Latin America & Caribbean countries of Puerto Rico, the Bahamas, and Costa Rica decreased, while trips to China ceased completely from 2020 quarter 3 onwards. In contrast, trips to Australia, the United Kingdom, and Sweden increased substantially during this period. The number of trips to Australia is particularly high in quarters 1 and 4, while trips to the United Kingdom are across all quarters, and to Sweden in quarters 2, 3 and 4. The number of trips from the United States to Ireland becomes at least 30 in 2021 quarters 2 and 3.

Table 4.23 Metrics for the United States case study network graphs based on the number of trips. In each year-quarter time period, the networks contain only the edges where there are at least 30 trips for that origin-destination country pair.

Year	Quarter	Number of vertices	Number of edges	Edge density (max vertices)	Number of trips
2018	1	7	8	0.044	715
2018	2	11	13	0.071	2202
2018	3	11	14	0.077	2765
2018	4	8	9	0.049	871
2019	1	7	7	0.038	791
2019	2	9	12	0.066	2209
2019	3	10	13	0.071	2729
2019	4	8	9	0.049	835
2020	1	7	8	0.044	822
2020	2	10	12	0.066	1539
2020	3	11	13	0.071	2205
2020	4	9	9	0.049	1156
2021	1	9	9	0.049	1164
2021	2	12	14	0.077	2323
2021	3	14	16	0.088	2944
2021	4	8	8	0.044	1400

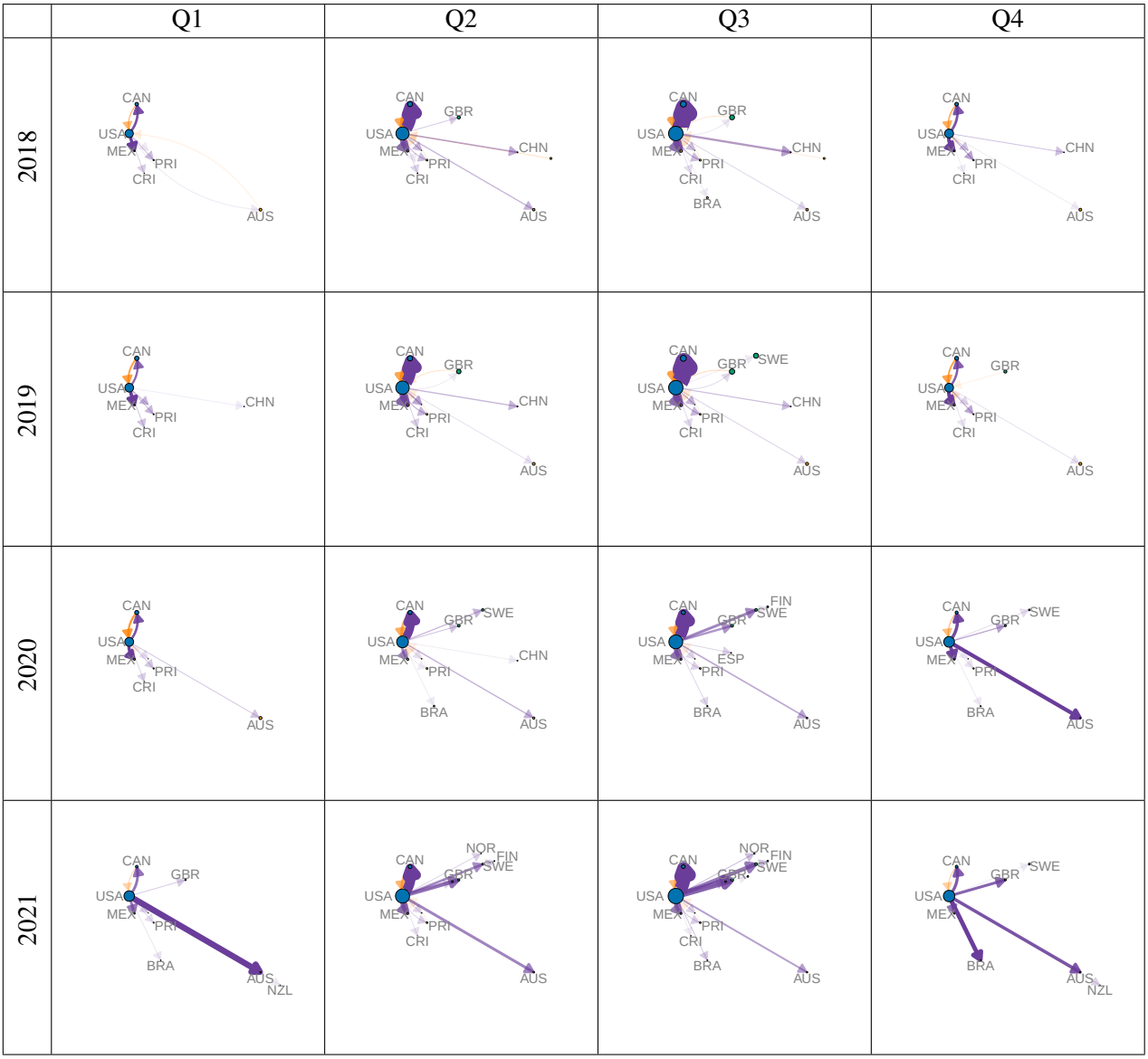


Fig. 4.7 Network graphs of the trips to and from the United States. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United States are purple. Edges representing trips to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than 30 trips are excluded.

Overview of other countries

Figures 4.8 to 4.11 show the networks based on the number of trips for other top countries: Canada, Sweden, Australia, and Mexico, by origin and destination, with all trips to and from the county included (i.e. a minimum threshold of 1). These show differences in the number of connections and in the dominance of trips from or to the country.

The majority of trips to Canada are from the United States, with a strong seasonal pattern of more trips in spring and summer. Trips from Canada are to destinations scattered worldwide, although there are relatively few to Europe, while in contrast, trips to Canada are often from northern Europe countries.

There are more connections from Sweden and Australia than to these countries, with more connections to geographically closer countries than for Canada. In particular, there are many connections from Sweden to southern Europe, and from Australia to destinations in the East Asia & Pacific region which are not commonly visited by users from other countries.

Mexico shows some differences to the other countries, with fewer connections overall. Most of these are inbound to Mexico from a few select countries, mainly Australia, New Zealand, and a few countries in Europe.

The decrease in the number of connections from 2020 quarter 2 onwards is clear for all countries. The increase in trips from the United States to Australia and Sweden after 2020 quarter 2, seen in the United States case study (Section 4.3.1), is also clearly evident in the networks for Australia and Sweden. Additionally, the consistent seasonal pattern of trips between the United States and Canada is evident in both the United States and Canada networks.

Fig. 4.8 Network graphs of the trips to and from Canada. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Canada are purple. Edges representing trips to Canada are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

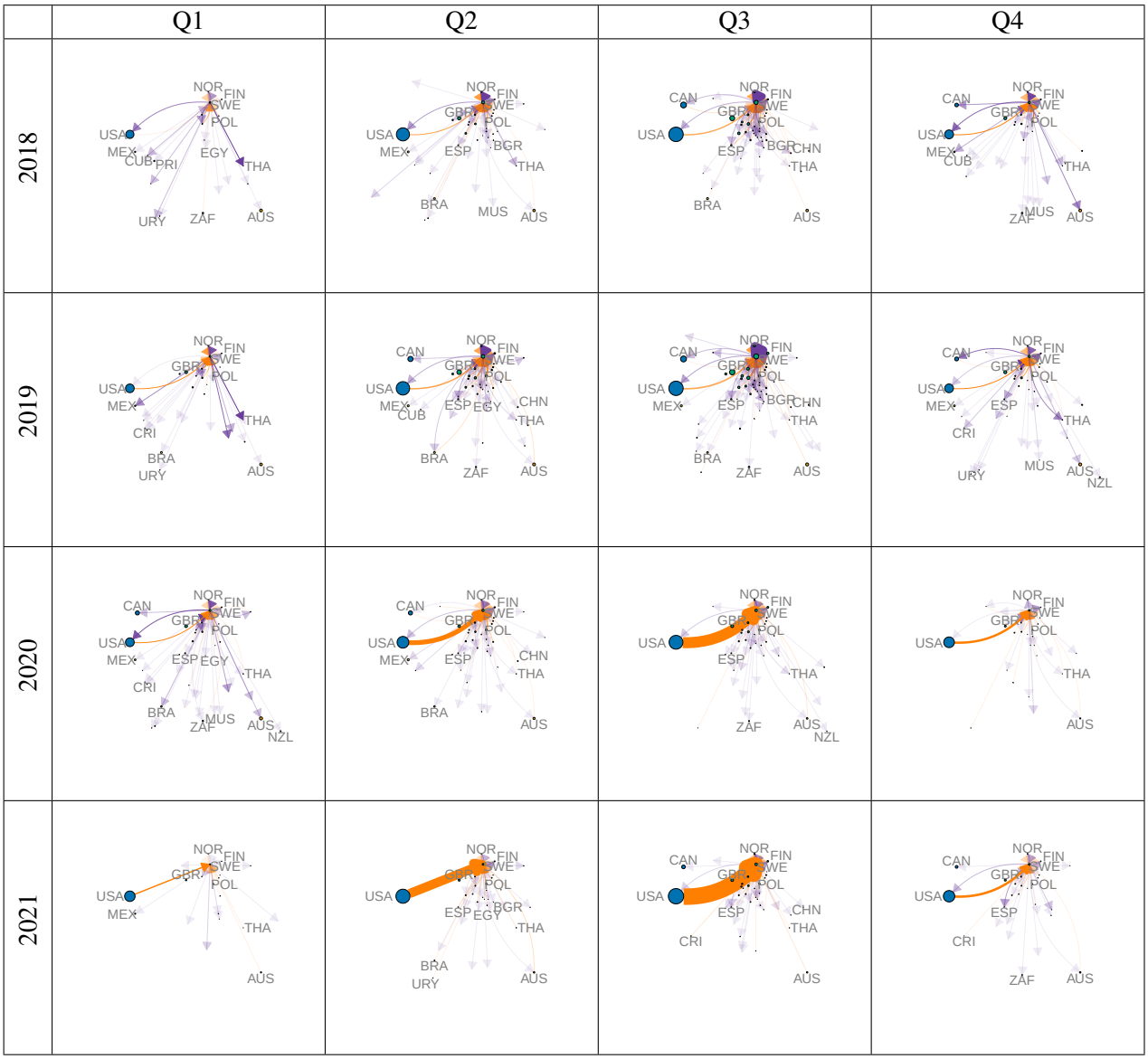


Fig. 4.9 Network graphs of the trips to and from Sweden. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Sweden are purple. Edges representing trips to Sweden are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

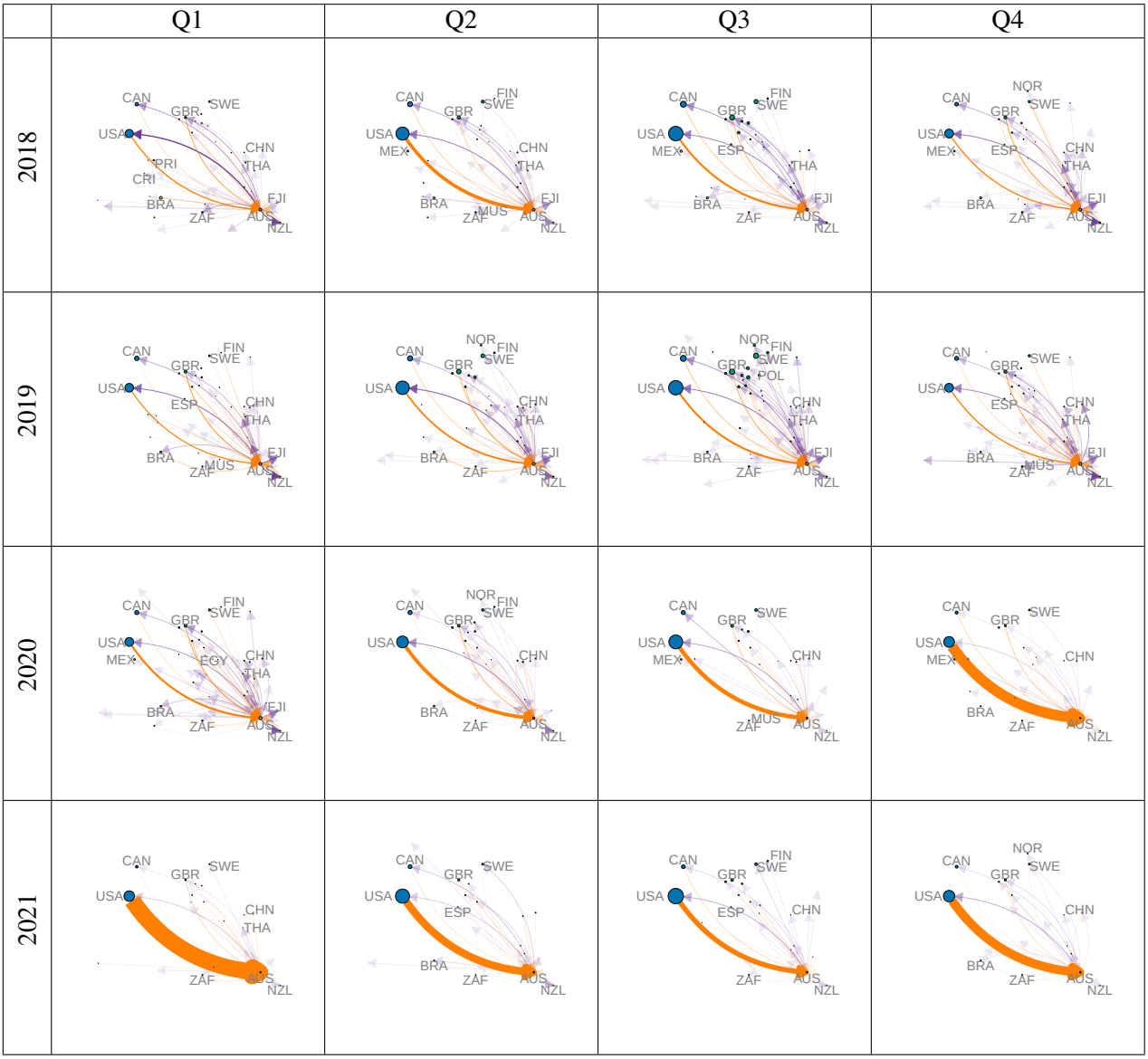


Fig. 4.10 Network graphs of the trips to and from Australia. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Australia are purple. Edges representing trips to Australia are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Fig. 4.11 Network graphs of the trips to and from Mexico. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from Mexico are purple. Edges representing trips to Mexico are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

4.3.2 Europe

We created network models representing the most popular travel between countries in Europe, by including only the top 15 European countries by the total number of outbound trips over the four-year period 2018-2021, and the top 15 European countries by the total number of inbound trips, as vertices, along with all trips between these countries as edges. We then identified the structure of communities within the network.

Table 4.24 presents the network metrics and Figure 4.13 presents the directed graphs summarising the number of trips between European countries. Table 4.25 presents a frequency table of the number of quarters in which each pair of countries has been allocated to the same community and Figure 4.14 presents community graphs showing the structure of the communities each quarter.

Networks

Seasonal patterns

Number of connections There are seasonal patterns in the number of connected countries throughout the four-year period. The network graphs (Figure 4.13) show that quarter 3 has the most densely connected networks (highest number of edges), while quarter 1 has the least densely connected networks. This is reflected in the network metrics (Table 4.24), with quarter 3 having the highest edge density (max vertices), and lowest mean distance and diameter compared to the other quarters, in all years. The reverse pattern is seen for quarter 1, which has low edge density values and high mean distance and diameter values compared to the other quarters. These indicate that more routes are travelled during the summer months and fewer during the winter months, compared to the rest of the year.

Number of trips There are also seasonal patterns in the number of trips (edge widths in Figure 4.13), which has a similar seasonal pattern to the number of connections. Edge widths tend to be narrower in quarter 1 compared to other quarters, indicating fewer trips between connected countries in the winter months compared to the rest of the year, while wider edge widths in quarter 3 indicate more people travelling between connected countries in summer months.

Trends over time

Number of connections After an initial increase in the number of connections from 2018 to 2019, there is a strong decrease in the number of connected countries from 2020 quarter 2 onwards. The density of connections is higher throughout 2019 compared to corresponding quarters in 2018, shown by a higher edge density (max vertices) and indicated by a lower mean distance and diameter in 2019 compared to corresponding quarters in 2018 (Table 4.24). However, from 2020 quarter 2 onwards, the edge density decreases and the mean distance and diameter increase compared to corresponding quarters in previous years, indicating a decrease in the number of connections between countries during this period, which continues until 2021 quarter 3. The networks are especially sparse from 2020 quarter 4 to 2021 quarter 2. Connections that remain are often between countries that are geographically close, especially in quarters 1 and 4 (Figure 4.13).

Number of trips The number of trips increases from 2018 to 2019, but then decreases substantially from 2020 quarter 2 onwards (Table 4.24). The highest number of outbound trips prior to 2020 quarter 2 are from Sweden and the United Kingdom, with a substantial number from Germany. From 2020 quarter 2 onwards, Sweden and the United Kingdom continue to have the highest number of outbound trips, followed by Ireland. The highest number of inbound trips prior to 2020 quarter 2 are to Norway and Sweden, with a substantial number to France. From 2020 quarter 2 onwards, Sweden and Norway have the highest number of inbound trips, followed by the United Kingdom.

The highest proportions of trips from the United Kingdom throughout the four-year period are to Ireland and France. Tourists from Sweden tended to visit Norway, Finland, and Denmark, although there is a substantial decrease in the proportion visiting Norway from 2020 quarter 2 onwards. Sweden is a popular destination for tourists from Poland. The majority of trips from Ireland were to the United Kingdom.

Ireland is the only country with an increase in outbound trips during the period from 2020 quarter 2 onwards, compared to the previous period. The other countries had a reduction in the number of outbound trips by 30-70%, with a total reduction in trips within Europe of around 50%.

Communities

Community graphs In general, within the community clusters (Figure 4.14) there is evidence of north-south grouping patterns, and less geographical overlap between clusters in the north-south direction than in the west-east direction. In contrast, there is less evidence of west-east grouping patterns, with clusters tending to cover a wide west-east area, and spanning the entire west to east of Europe in many quarters. This suggests a tendency to travel further across latitudes than across longitudes.

The community graphs in all quarters except one have three community clusters. The Nordic countries (Denmark, Norway, Sweden, Finland) and Baltic countries (Latvia, Lithuania) are in a community together in all quarters, and the countries in the centre of the geographic region tend to form a second cluster, with a third cluster of countries spanning west to east across southern Europe. Quarter 1 in 2021 is the only quarter when there were four communities, showing a more disconnected network compared to all other quarters, indicating travel was more regional during this quarter compared to during other quarters.

There tends to be slightly less geographical overlap in community clusters from 2020 quarter 3 to 2021 quarter 3 inclusive, suggesting Fishbrain app users preferred to travel less far and chose destinations slightly closer to home during this period.

Community pairs Table 4.25 shows the number of times each pair of countries is in the same community cluster, out of a maximum of 16 (four quarters per year over four years). The countries are ordered approximately geographically from west to east, followed by Mediterranean countries. Figure 4.12 shows this information on a map, with the width of connections between countries proportional to the number of times the connected countries are in the same community cluster.

Considering strong connections between countries to be when they are frequently in the same community group, Table 4.25 shows that strong connections often occur between countries in similar geographical areas, while weak connections tend to occur between countries in different geographical areas.

Strongly connected countries There are three distinct geographical areas containing groups of neighbouring countries with strong connections within the group. In size order, the first group is the

Nordic (Denmark, Finland, Norway, and Sweden) and Baltic (Latvia and Lithuania) countries. The second group is Switzerland, Belgium, the Netherlands, and Germany, and the third is the United Kingdom, Ireland, and France.

Within the Nordic-Baltic group, there are very strong connections between all Nordic countries and between the Baltic and Nordic countries, while the connection within the Baltic countries between Latvia and Lithuania is only fairly strong. All countries in the Nordic-Baltic group (except Norway) are situated around the Baltic Sea. Within the Switzerland-Belgium-Netherlands-Germany group, the connections are strongest between the countries that are geographically closer. The strongest connections are along the North Sea coastline, between the Netherlands and its immediate neighbours, Belgium to the west and Germany to the east. Germany is fairly strongly connected with Switzerland and Belgium, countries with which it shares a border. Switzerland is moderately connected with the Netherlands and Belgium, which are in the same geographic area but are not immediate neighbours. Within the United Kingdom-Ireland-France group, the connections are strongest between the United Kingdom and Ireland.

The Central and Eastern Europe (CEE) countries and Mediterranean countries are two further regional groups, spanning a wider geographical range than the three regional groups already discussed, with varying levels of connections.

Central and Eastern Europe countries The Central and Eastern Europe (CEE) countries (Croatia, Latvia, Lithuania, Poland, and Romania) are geographically dispersed, situated from the Baltic Sea in northern Europe to the Adriatic Sea and Black Sea in southern Europe. The northern CEE countries (Latvia, Lithuania, and Poland), which are close neighbours and all situated on the coast of the Baltic Sea, have slightly stronger connections with the Nordic countries than with each other. The southern CEE countries, Croatia with coastline on the Adriatic Sea and Romania with coastline on the Black Sea, are not immediate neighbours and are only very weakly connected with each other, being in the same community only once. The connections between the northern and southern CEE countries are weak. All northern CEE countries have very weak or no connections with the United Kingdom-Ireland-France and Switzerland-Belgium-Netherlands-Germany groups, except for Poland which has slightly stronger connections with the United Kingdom-Ireland-France group. The southern CEE countries have different levels of connection with other countries. Croatia has moderate connections

with the Switzerland-Belgium-Netherlands-Germany group and weak connections with the United Kingdom-Ireland-France group, with the opposite for Romania.

Mediterranean countries None of the Mediterranean countries (Spain, Italy, and Greece) are immediate land neighbours, and connection strengths between them vary. Connections between Italy, which is geographically central within this group of countries, and the other Mediterranean countries are weak, while there are fairly strong connections between the two most geographically distant Mediterranean countries, Spain and Greece.

Similarly, the strength of connections between the Mediterranean countries and the United Kingdom-Ireland-France, Switzerland-Belgium-Netherlands-Germany and CEE groups varies. Overall, connections with the United Kingdom-Ireland-France group are stronger for Spain than for Greece, with both having slightly stronger connections with France than with the United Kingdom and Ireland. Overall, these connections are moderate in strength, except for a fairly strong connection between Spain and neighbouring France, which is Spain's strongest connection. In contrast, Spain and Greece are weakly connected with the countries in the Switzerland-Belgium-Netherlands-Germany group and CEE countries. The strongest connection for both Spain and Greece with countries in these two groups is with Romania, while the weakest is with Croatia. The reverse is seen for Italy, which has very weak connections with the United Kingdom-Ireland-France group, but stronger connections with the countries in the Switzerland-Belgium-Netherlands-Germany group to the north, and with Croatia to the east across the Adriatic Sea.

Italy and the two southern CEE countries, Croatia and Romania, are the only countries that do not have strong connections with any other country. The strongest connections for these countries are fairly weak connections between Italy and the Switzerland-Belgium-Netherlands-Germany group, fairly weak connections between Croatia and the Switzerland-Belgium-Netherlands-Germany group, rising to moderate connections to Germany, and moderate connections between Romania and the United Kingdom-Ireland-France group.

Weak or no connections Countries which are weakly connected tend to be in different geographical areas. Connections between the three groups of strongly connected countries are weak or non-existent.

The Nordic countries have no connections with the United Kingdom-Ireland-France group, and very weak or no connections with the countries in the Switzerland-Belgium-Netherlands-Germany group. Connections with countries further south (the southern CEE and Mediterranean countries) are slightly stronger, although still weak.

Connections between the United Kingdom-Ireland-France and Switzerland-Belgium-Netherlands-Germany groups are weak, particularly with the United Kingdom and Ireland with which there are no or very weak connections. Connections with mainland Europe neighbour France are slightly stronger but still in the range of very to fairly weak.

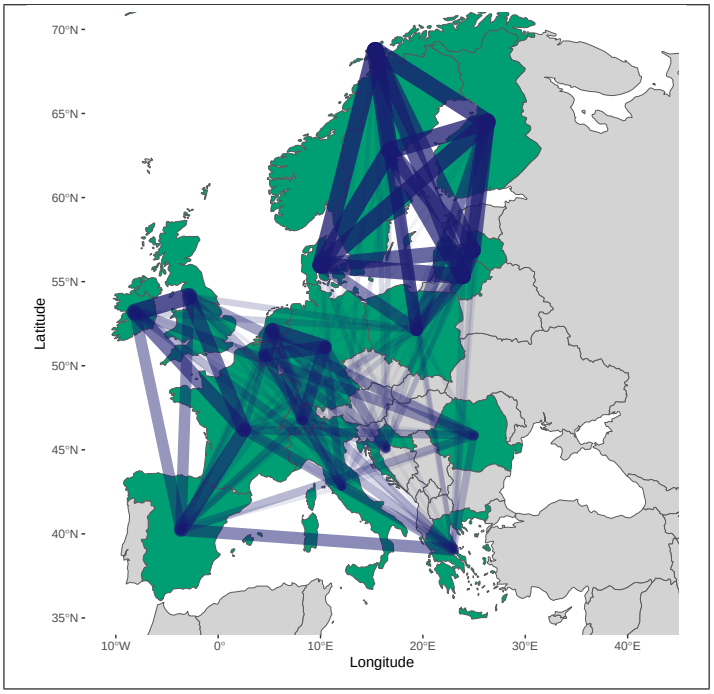


Fig. 4.12 The number of quarters during the period 1 January 2018 to 31 December 2021 in which each pair of countries is in the same community. The width of connections between countries is proportional to the number of times the connected countries are in the same community cluster.

Table 4.24 Metrics for Europe network graphs based on the number of trips.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)	Number of trips
2018	1	15	30	0.143	0.088	2.40	5	60
2018	2	19	65	0.190	0.190	1.86	3	304
2018	3	19	116	0.339	0.339	1.46	2	678
2018	4	18	64	0.209	0.187	1.90	4	179
2019	1	16	50	0.208	0.146	1.94	4	115
2019	2	19	111	0.325	0.325	1.50	3	623
2019	3	19	133	0.389	0.389	1.38	2	1119
2019	4	19	81	0.237	0.237	1.71	3	234
2020	1	17	52	0.191	0.152	2.03	4	130
2020	2	19	72	0.211	0.211	1.80	4	271
2020	3	19	91	0.266	0.266	1.62	3	372
2020	4	18	44	0.144	0.129	2.11	4	115
2021	1	17	25	0.092	0.073	2.79	6	51
2021	2	18	53	0.173	0.155	2.02	4	163
2021	3	19	86	0.251	0.251	1.69	4	346
2021	4	18	48	0.157	0.140	2.09	5	108

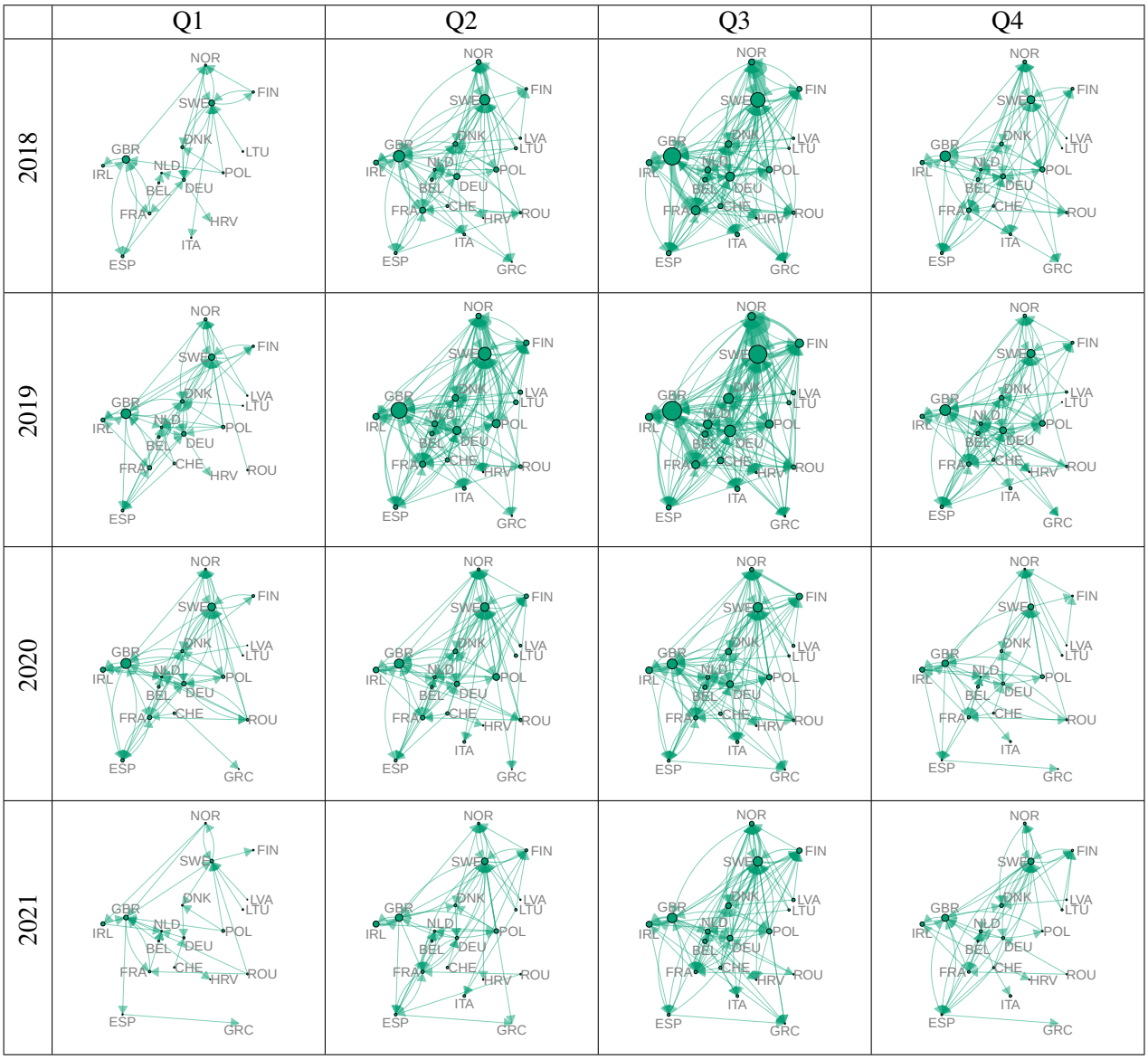


Fig. 4.13 Network graphs of Europe, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table 4.25 Number of quarters (maximum 16) in which each pair of countries in the Europe networks based on trips is in the same community. The countries are ordered geographically, approximately west to east, followed by Mediterranean countries.

	IRL	GBR	FRA	CHE	BEL	NLD	DEU	DNK	NOR	SWE	FIN	POL	LTU	LVA	ROU	HRV	GRC	ITA
United Kingdom (GBR)	16																	
France (FRA)	13	13																
Switzerland (CHE)	3	3	5															
Belgium (BEL)	2	2	5	8														
Netherlands (NLD)	0	0	3	10	14													
Germany (DEU)	0	0	3	11	12	14												
Denmark (DNK)	0	0	0	2	0	0	2											
Norway (NOR)	0	0	0	2	0	0	2	16										
Sweden (SWE)	0	0	0	2	0	0	2	16	16									
Finland (FIN)	0	0	0	2	0	0	2	16	16	16								
Poland (POL)	5	5	5	1	1	0	0	11	11	11	11							
Lithuania (LTU)	1	1	1	2	0	0	2	15	15	15	15	10						
Latvia (LVA)	1	1	1	3	0	0	1	14	14	14	14	10	13					
Romania (ROU)	9	9	7	3	1	2	2	0	0	0	0	3	1	0				
Croatia (HRV)	2	2	2	7	7	7	9	4	4	4	4	2	4	3	1			
Greece (GRC)	7	7	9	4	3	3	2	3	3	3	3	4	4	4	6	0		
Italy (ITA)	3	3	3	7	6	7	7	3	3	3	3	2	4	2	3	5	4	
Spain (ESP)	10	10	13	6	5	4	4	1	1	1	1	4	2	2	7	1	11	3

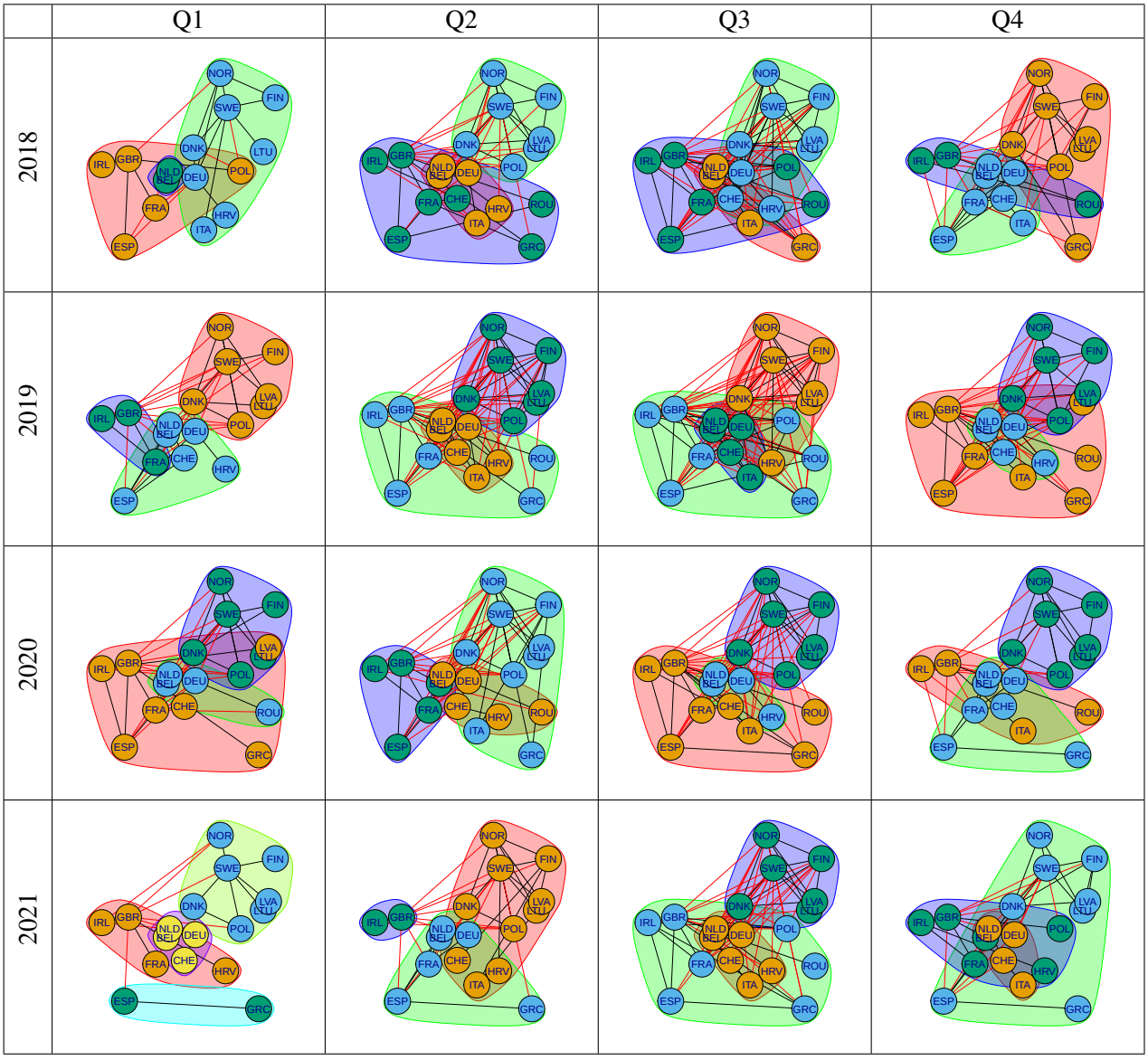


Fig. 4.14 Communities in the Europe networks based on the number of trips. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities.

4.3.3 World countries

We created network models representing the most popular travel between countries globally, by including only the top 15 countries by the total number of outbound trips over the four-year period 2018-2021, and the top 15 countries by the total number of inbound trips, as vertices, along with all trips between these countries as edges. We then identified the structure of communities within the network.

Table 4.26 presents the network metrics and Figure 4.16 presents the directed graphs summarising the number of trips between countries globally. Table 4.27 presents a frequency table of the number of quarters in which each pair of countries has been allocated to the same community and Figure 4.17 presents community graphs showing the structure of the communities each quarter.

Networks

Seasonal patterns

Number of connections There are seasonal patterns in the number of connected countries in 2018 and 2019, although not as strongly evident as in the Europe networks. In these years, quarter 3 has the highest number of edges and quarters 1 and 4 have a lower number of edges, with variation in quarter 2 each year (Figure 4.16). This is reflected in the network metrics (Table 4.26), with quarter 3 having the highest edge density (max vertices) and lowest mean distance and diameter, and quarters 1 and 4 tending to have a low edge density (max vertices) and higher mean distance and diameter, compared to the other quarters in this period. These indicate that there are more connections during the northern hemisphere summer and fewer during their autumn and winter, compared to the rest of the year, with the number of connections during the northern hemisphere spring less consistent. The network graphs (Figure 4.16) show that the number of connections from northern hemisphere countries to Australia is greater in quarters 1 and 4 (Australian summer) than in quarters 2 and 3, while the reverse is seen for travel in the opposite direction, with the number of connections from Australia to northern hemisphere countries greater in quarters 2 and 3 (northern hemisphere summer) than in quarters 1 and 4. The number of connections from Europe to countries in the North America and Latin America & Caribbean regions are also higher during quarters 2 and 3, compared to quarters

1 and 4. This shows a tendency towards a greater number of connections during the warmer months of the destination country, as seen in the Europe networks.

Number of trips There is a clear seasonal pattern in the number of trips (Table 4.26) throughout the four-year period. Each year, quarter 3 has the most trips, followed by quarter 2, with the fewest trips in quarter 1. This shows more trips are recorded by Fishbrain users during the northern hemisphere spring and summer, compared to other times of the year.

Trends over time

Number of connections After an initial increase in the number of connections from 2018 to 2019, there is a decrease in the number of connected countries in 2020. This aligns with the pattern observed in the Europe networks. The density of connections is higher throughout 2019 compared to corresponding quarters in 2018, shown by a higher edge density (max vertices) and indicated by a lower mean distance in 2019, compared to corresponding quarters in 2018 (Table 4.26). In 2020 quarter 2, the edge density decreases compared to 2019, but remains higher than in 2018. From 2020 quarter 3 onwards, the edge density decreases, the mean distance increases, and the diameter increases or remains the same, compared to corresponding quarters in previous years, indicating that the number of connections between countries is decreasing during this period, which continues until 2021 quarter 3 when the trend starts to reverse. The number of connected countries is lowest in 2021 quarter 1. The decrease in density of connections is less visible in the graphs of global networks (Figure 4.16) than in the graphs of the Europe networks (Figure 4.13), possibly as a result of more connections in the network obscuring the pattern. The reduction in number of connections is more apparent in the north-south direction than in the west-east direction, especially the connections of Australia, New Zealand, and Brazil with Europe.

Number of trips The number of trips increases from 2018 to 2019, then decreases in 2020 quarters 2 and 3 by just over 25% compared to corresponding quarters in previous years. Following this, the total number of trips increases in 2020 quarter 4 and 2021 quarter 1 to more than in

corresponding quarters in 2018 and 2019. In 2021, quarters 2 and 3 have similar levels of trips to 2018, while the number of trips in quarter 4 rise above previous quarter 4 levels.

The highest number of outbound trips prior to 2020 quarter 2 is from the United States and Canada, with a moderate number from the United Kingdom and Sweden. From 2020 quarter 2 onwards, the United States and Canada continue to have the highest number of outbound trips, with a moderate number from the United Kingdom. The number of outbound trips from most countries decreases by around 10-70% in the period from 2020 quarter 2 onwards compared to the period 2018 to 2019, although there is an increase in the number of trips from the United States and Ireland by 13% and 14% respectively. The largest percentage decrease in the number of outbound trips is from Sweden. Overall, the number of outbound trips decreases by 7% from the period 2018-2019 to 2020-2021.

Popular destinations from the United States throughout the four-year period are Canada and Mexico, although from 2020 quarter 2 onwards there is a notable increase in the proportion of trips to Australia, the United Kingdom and Sweden, with the number of trips to Australia increasing substantially from 2020 quarter 4 onwards. During the period from 2020 quarter 2 onwards, 93% of trips to Australia, 85% of trips to the United Kingdom, and 78% of trips to Sweden originate from the United States, compared to 67%, 46% and 25% respectively during the period 2018 to 2019. There is also a substantial increase in trips from the United States to Brazil in 2021 quarter 4.

The highest number of inbound trips prior to 2020 quarter 2 are to Canada, the United States, and Mexico. From 2020 quarter 2 onwards, these countries continue to have the highest number of inbound trips, along with a substantial number to Australia and the United Kingdom. The largest decrease (75%) in the number of inbound trips is to China, while the number of trips to Australia, the United Kingdom, and Brazil approximately triples, and the number to Sweden and Finland approximately doubles.

The majority of inbound trips to Canada and Mexico throughout the four-year period are from the United States (97% and 95% respectively). This is reciprocated, as the majority of outbound trips from Canada and Mexico are to the United States (88% and 95% respectively). The majority of inbound trips to the United States are from Canada, with a moderate number from Mexico, Puerto Rico, and the United Kingdom, all of which decrease in 2020 and 2021 compared to 2018 and 2019, with greater decreases from Canada and the United Kingdom than from Mexico and Puerto Rico.

As seen in the United Kingdom case study, the largest proportions of inbound trips to the United Kingdom are from the United States and Ireland, with the number of trips from both these countries increasing from 2020 quarter 2 onwards, compared to the period from 2018 to 2019. Most trips to Australia are from the United States, the majority of which were in 2020 and 2021.

Communities

Community graphs The community cluster patterns in the world countries (Figure 4.17) are similar to those in Europe, with clusters tending to occur in regional groups and in the north-south direction, rather than west-east. There is less geographical overlap between clusters in the north-south direction than in the west-east direction. These patterns suggest a tendency to travel further across latitudes than across longitudes, which is also seen in the Europe networks. The community graphs in all quarters have either two or three community clusters, with seasonal patterns in the clusters.

Until 2020 quarter 2 inclusive, quarters 1 and 4 have two clusters, while quarter 3 has three clusters, indicating a tendency towards longer distance travel in quarters 1 and 4, with shorter distance travel in quarters 2 and 3. Where networks consist of two clusters, these tend to be split into a cluster containing the Americas and China, and a cluster containing Australia, New Zealand, South Africa, and the countries in Europe. In quarter 3, the networks consist of three clusters, which also have a cluster containing the Americas and China, with the remaining countries split into a cluster containing Australia, New Zealand, South Africa, the United Kingdom, Ireland, and France, and a cluster containing countries from the rest of Europe. Cluster patterns in quarter 2 are more variable, ranging from the two-cluster pattern seen in quarters 1 and 4 to the three-cluster pattern seen in quarter 3, and contain exceptions to the typical patterns. These exceptions include Brazil joining the cluster containing European countries (in 2019) and a cluster consisting of just Australia and New Zealand (in 2018).

From 2020 quarter 3 onwards, the cluster pattern changes. In 2020 quarter 3, the cluster pattern is similar to quarter 3 in previous years, except Australia is in the cluster containing the Americas and China, and New Zealand joins the mainland Europe cluster. In the next two quarters (2020 quarter 4 and 2021 quarter 1), Europe forms one cluster and the rest of the world forms a second cluster, although New Zealand and Germany break into an additional third small cluster in 2021 quarter 1.

From 2021 quarter 2 onwards, the cluster pattern is consistent, with mainland Europe countries in one cluster and the rest of the world (the Americas, China, Australia, New Zealand, South Africa, the United Kingdom, and Ireland) in a second cluster.

The red edges indicate travel between clusters, while black edges indicate travel within clusters. Prior to 2020 quarter 4, the community graphs have many red edges, indicating many routes connecting countries in different clusters. From 2020 quarter 4 onwards, there are fewer red edges, indicating less travel between community clusters, suggesting that communities are more highly separated during this period.

Community pairs Table 4.27 shows the number of times each pair of countries is in the same community cluster, out of a maximum of 16 (four quarters over four years). The countries are ordered in regional groups, approximately geographically from west to east. Figure 4.15 shows this information on a map, with the width of connections between countries proportional to the number of times the connected countries are in the same community cluster.

Very strong connections Considering strong connections between countries to be when they are frequently in the same community group, Table 4.27 shows groups of countries in the same geographical area which have very strong connections: countries in the Americas; the United Kingdom and Ireland; the Nordic countries (Denmark, Finland, Norway, and Sweden), Germany and Poland; Australia and New Zealand. There are also very strong connections between China and the Americas, which are not in the same geographical area.

Fairly strong connections There are fairly strong connections between countries in the same geographical area and across the world: South Africa with Australia and New Zealand; the United Kingdom and Ireland with Australia, New Zealand, and South Africa; France with all European countries.

Moderate connections France has moderate connections with Australia, New Zealand, and South Africa, while the United Kingdom and Ireland have moderate connections with countries in mainland Europe (except France, with which there is a fairly strong connection).

Fairly weak connections Australia, New Zealand and South Africa are all fairly weakly connected with mainland Europe, the Americas, and China.

Very weak connections The United Kingdom and Ireland are very weakly connected with the Americas and China, and are only in the same community in 2021 quarters 2, 3 and 4.

No connections Mainland Europe countries are never in the same community as the Americas and China, although Brazil joins the cluster containing mainland Europe countries in one quarter (2019 quarter 2).

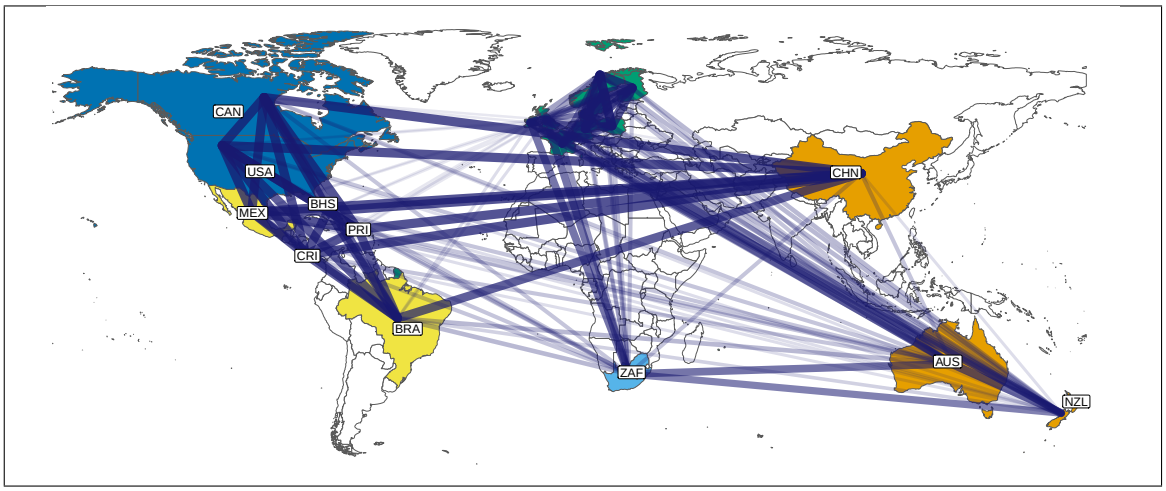


Fig. 4.15 The number of quarters during the period 1 January 2018 to 31 December 2021 in which each pair of countries is in the same community. The width of connections between countries is proportional to the number of times the connected countries are in the same community cluster.

Table 4.26 Metrics for world countries network graphs based on number of trips.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)	Number of trips
2018	1	20	79	0.208	0.208	1.67	3	1039
2018	2	20	75	0.197	0.197	1.66	3	2525
2018	3	20	99	0.261	0.261	1.54	2	3368
2018	4	20	82	0.216	0.216	1.68	3	1239
2019	1	20	83	0.218	0.218	1.63	3	1175
2019	2	20	108	0.284	0.284	1.50	2	2808
2019	3	20	112	0.295	0.295	1.48	2	3654
2019	4	20	91	0.239	0.239	1.59	3	1208
2020	1	20	82	0.216	0.216	1.62	2	1190
2020	2	20	89	0.234	0.234	1.60	3	1927
2020	3	20	79	0.208	0.208	1.64	3	2577
2020	4	20	63	0.166	0.166	1.73	3	1430
2021	1	19	52	0.152	0.137	1.83	3	1353
2021	2	20	72	0.189	0.189	1.74	4	2541
2021	3	20	79	0.208	0.208	1.66	3	3137
2021	4	20	68	0.179	0.179	1.76	3	1683

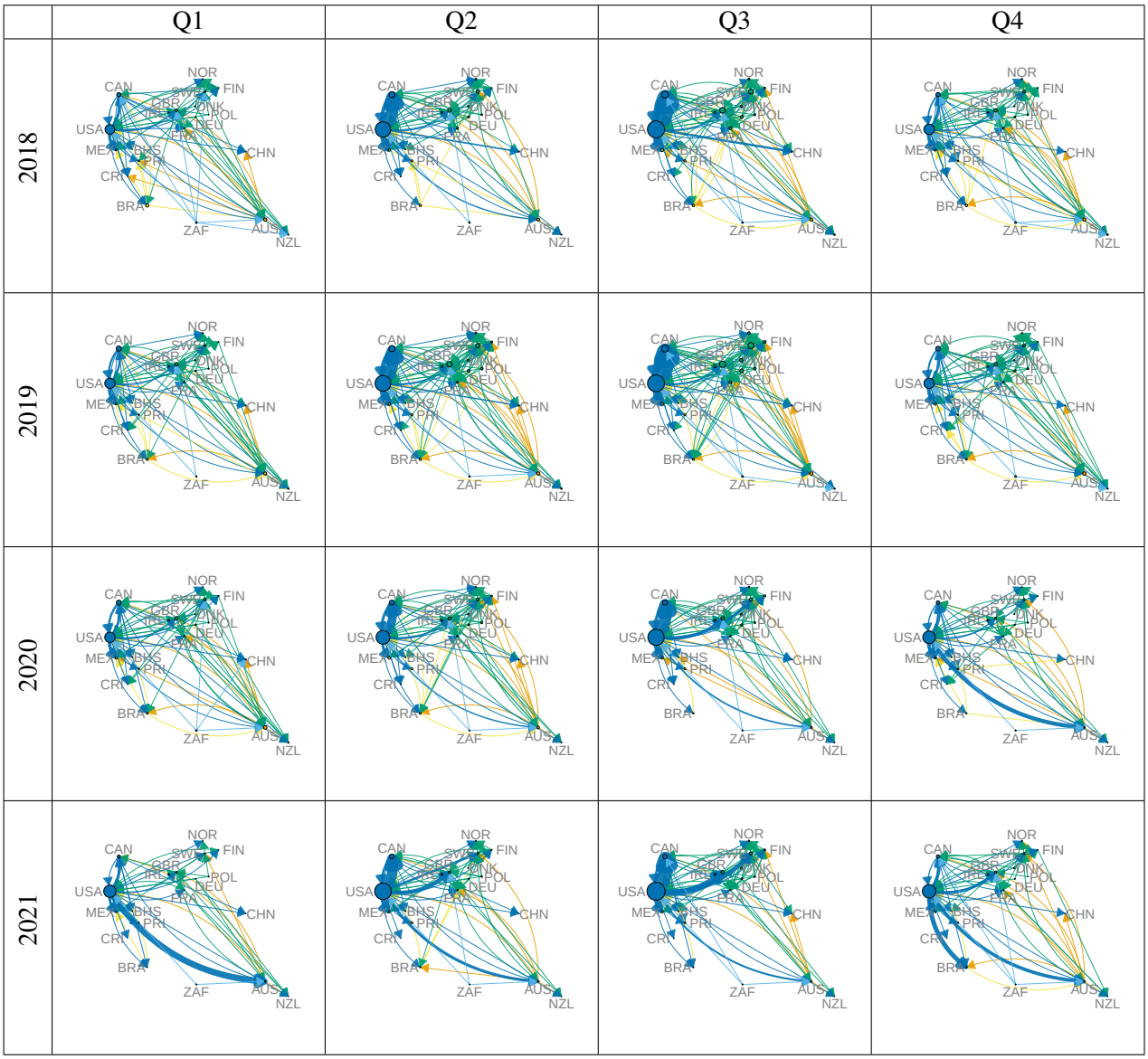


Fig. 4.16 Network graphs of the world countries, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table 4.27 Number of quarters (maximum 16) in which each pair of countries in the world countries networks based on trips is in the same community. The countries are ordered approximately geographically in a clockwise direction (the Americas, Europe, China, New Zealand, Australia, South Africa).

		CAN	USA	MEX	BHS	CRI	PRI	BRA	IRL	GBR	FRA	DEU	DNK	NOR	SWE	FIN	POL	CHN	NZL	AUS
United States	USA	16																		
Mexico	MEX	16	16																	
Bahamas	BHS	16	16	16																
Costa Rica	CRI	16	16	16	16															
Puerto Rico	PRI	16	16	16	16	16														
Brazil	BRA	15	15	15	15	15	15													
Ireland	IRL	3	3	3	3	3	3	4												
United Kingdom	GBR	3	3	3	3	3	3	4	16											
France	FRA	0	0	0	0	0	0	1	13	13										
Germany	DEU	0	0	0	0	0	0	1	8	8	11									
Denmark	DNK	0	0	0	0	0	0	1	8	8	11	15								
Norway	NOR	0	0	0	0	0	0	1	9	9	12	15	15							
Sweden	SWE	0	0	0	0	0	0	1	9	9	12	15	15	16						
Finland	FIN	0	0	0	0	0	0	1	9	9	12	15	15	16	16					
Poland	POL	0	0	0	0	0	0	1	9	9	12	15	15	16	16	16				
China	CHN	16	16	16	16	16	16	15	3	3	0	0	0	0	0	0	0			
New Zealand	NZL	4	4	4	4	4	4	5	12	12	9	8	7	7	7	7	7	4		
Australia	AUS	6	6	6	6	6	6	7	12	12	9	6	6	6	6	6	6	6	14	
South Africa	ZAF	6	6	6	6	6	6	7	13	13	10	6	6	6	6	6	6	6	12	13

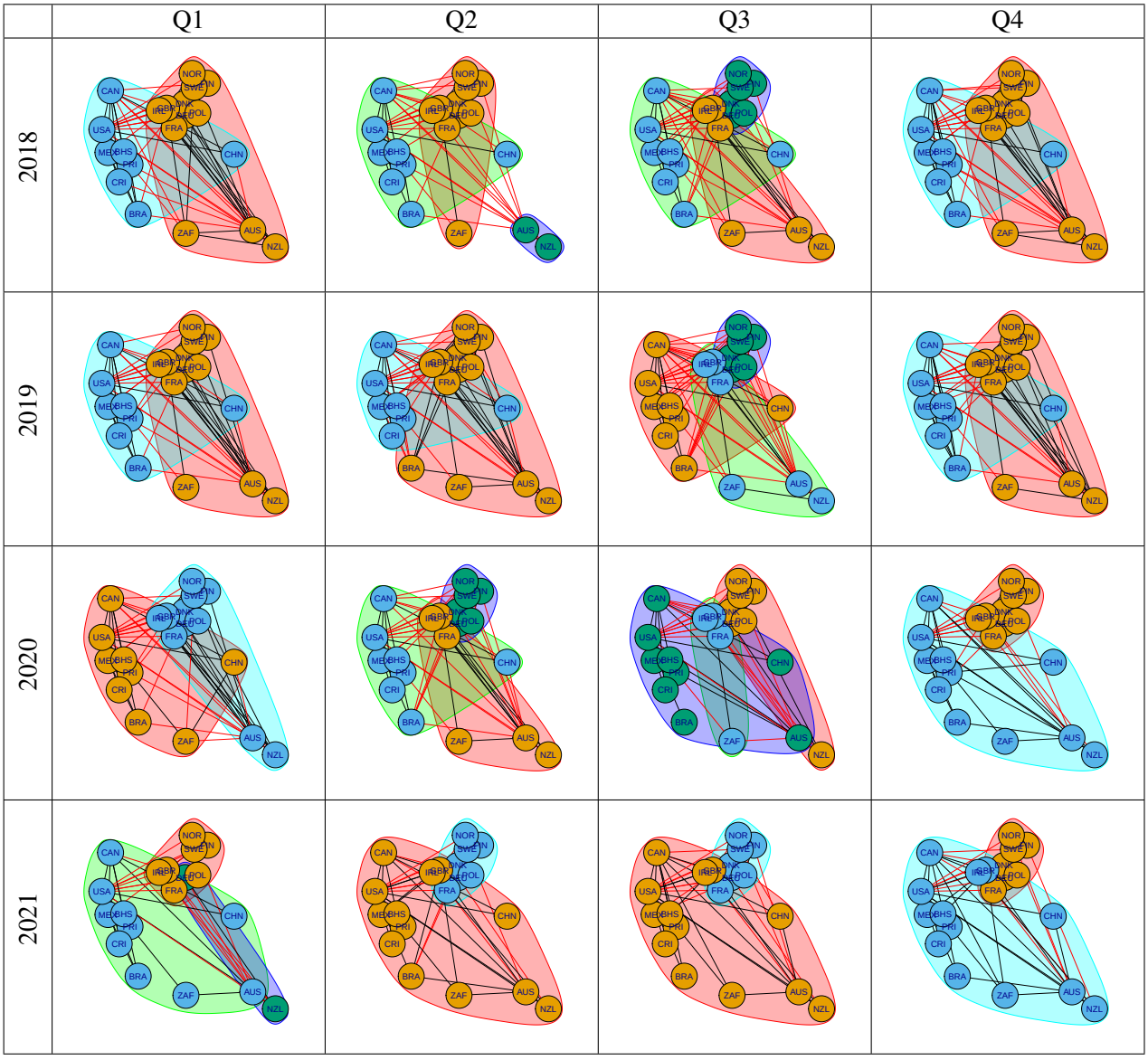


Fig. 4.17 Communities in the world countries networks based on the number of trips. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities.

4.3.4 World regions

We created network models representing the most popular travel between world regions, by including all trips over the four-year period 2018-2021 as edges, along with the associated vertices representing regions. We then identified the structure of communities within the network.

Table 4.28 presents the network metrics and Figure 4.18 presents the directed graphs summarising the number of trips between world regions. Tables 4.29 and 4.30 present a frequency table of the number of quarters in which each pair of regions has been allocated to the same community and Figure 4.19 presents community graphs showing the structure of the communities each quarter.

The metrics for the networks based on the number of trips discussed in this section and the number of catches recorded (in Appendix C) are identical. This is because these metrics are based on the number of edges within the networks. The world regions networks based on trips and catches have identical edges as they use all the data with no minimum threshold for the number of trips or catches. The size of the vertices and the edges weights are not identical on the trips and catches networks, as different users record different numbers of catches. However, this does not impact the unweighted metrics, which we use here. The loops on the network and community graphs show the number of catches recorded in a different country to the origin country, but within the same region.

Networks

Seasonal patterns

Number of connections There are no clear seasonal patterns in the number of connected regions. In contrast to the networks for Europe and world countries, the edge density (max vertices), mean distance and diameter (Table 4.28) are fairly consistent across quarters. This is reflected in the network graphs (Figure 4.18), with most regions connected to all other regions throughout the four-year period.

Number of trips However, there are seasonal patterns in the number of trips throughout the four-year period. Across the networks as a whole, each year, quarter 3 has the most trips, followed by quarter 2, with the fewest trips in quarter 1 (Table 4.28). This seasonal pattern is the same as in the networks of the top world countries (Section 4.3.3).

There are also seasonal patterns in the number of trips within and between regions. With respect to trips within the same region, the number of trips within the North America and Europe & Central Asia regions is higher (wider edge widths on the loops) in quarters 2 and 3 compared to quarters 1 and 4, indicating that there are more trips between countries within these regions during their spring and summer, compared to autumn and winter (Figure 4.18). In contrast, the number of trips within the East Asia & Pacific region is around 15-30% higher in quarter 4, and within the Latin America & Caribbean region around 25-40% higher in quarter 1, compared to the rest of the year. Both these regions are predominantly in tropical regions and the southern hemisphere, again indicating a preference for travel in warmer months, though to a lesser extent than seen in the regions in the northern latitudes, possibly due to substantial parts of these regions experiencing warmer temperatures in general throughout the year in comparison.

In terms of trips between regions, the number of trips from the North America region to the Latin America & Caribbean region is higher in quarters 2 and 3 than in quarters 1 and 4, particularly during 2018 and 2019 (Figure 4.18), despite less of a contrast in temperatures across the year in many of these destination countries compared to countries in other regions.

Trends over time

Number of connections There is a strong decrease in the number of connections within the world regions networks from 2020 quarter 2 onwards (Table 4.28). Prior to 2020 quarter 2, the edge density (max vertices), mean distance and diameter are fairly consistent, indicating a consistent number of connections between regions during this period. In contrast, during the period 2020 quarter 2 to 2021 quarter 1, the edge density decreases each quarter, then remains consistently low, and increases slightly in 2021 quarter 4, indicating a reduction in the number of connections within the network during this period compared to previously. The mean distance indicates a similar trend, increasing from 2020 quarter 2, and remaining high during 2021 quarters 1 to 3. The mean number of edges in the networks prior to 2020 quarter 2 is 40 (range 39 to 41), and decreases to 35 (range 34 to 37) during the period from 2020 quarter 2 onwards, showing a complete absence of trips between

some regions in some quarters. The density of connections in the networks is lowest in 2021 quarters 1, 2 and 3.

The decrease in the number of connections is mainly focused around connections from the Latin America & Caribbean region (Figure 4.18). In all quarters from 2020 quarter 3 onwards, there is a reduction in the number of connections from the Latin America & Caribbean region compared to the previous period, with the greatest reduction in 2021 quarter 1, when the only connection is to the North America region. There are no connections from the Latin America & Caribbean region to Sub-Saharan Africa from 2020 quarter 3 onwards, and to the East Asia & Pacific region in 2020 quarter 3 and 2021 quarter 1. Furthermore, there are no connections from Sub-Saharan Africa to the Latin America & Caribbean region for over a year, from 2020 quarter 2 to 2021 quarter 3 inclusive, though this connection is only present in two-thirds of the networks prior to 2020 quarter 2. There is a notable absence of connections in 2020 quarter 1 to Europe & Central Asia from the Latin America & Caribbean and Sub-Saharan Africa regions. In contrast, there is no change in the number of connections from the North America, Europe & Central Asia, and East Asia & Pacific regions to all other regions throughout the four-year period.

Number of trips The total number of trips increases each quarter until 2020 quarter 1 compared to the corresponding quarter in the previous year, but then decreases substantially from 2020 quarter 2 onwards (Table 4.28). The highest number of outbound trips each year throughout the four-year period are from the North America and Europe & Central Asia regions, although the number of trips from Europe & Central Asia decreases by over half from 2020 quarter 2 onwards compared to corresponding quarters in the period 2018 to 2019. In contrast, while the number of trips from the North America region shows an initial decrease in 2020 quarters 2 and 3 (by 15-25%), this is followed by an increase in 2020 quarter 4 and 2021 quarter 1 of 34%, and a substantial increase of 50% in 2021 quarter 4 compared to corresponding quarters in the period 2018 to 2019. There is a moderate number of outbound trips from the East Asia & Pacific and Latin America & Caribbean regions. North America is the only region with an increase in outbound trips during the period from 2020 quarter 2 onwards compared to the previous period. Outbound trips from the other regions decrease by 30-50% during this period. This reduction is offset by the increase in trips from North America, resulting in a

global decrease of around 15% in the number of trips between these two periods. The regions with the highest number of inbound trips prior to 2020 quarter 2 are the North America, Latin America & Caribbean, and Europe & Central Asia regions. From 2020 quarter 2 onwards, the number of trips to the North America and Latin America & Caribbean regions decreases by around 30%, while trips to Europe & Central Asia increase by around 10%.

The number of trips (edge widths) between regions varies by region and over time (Figure 4.18). The main destinations from North America throughout the four-year period are other countries in the North America region (34%) and the Latin America & Caribbean region (31%). During the period 2018 to 2019, 39% of trips are to other countries in the North America region, 37% are to the Latin America & Caribbean region, and 8% are to Europe & Central Asia. From 2020 quarter 2 onwards, the number of trips from North America to the Latin America & Caribbean region and within the North America region decreases (by around 25% each), but increases to the East Asia & Pacific and Europe & Central Asia regions (by approximately 50% and 200% respectively), resulting in a more equal distribution of trips (approximately 25 to 30%) to each of the North America, Latin America & Caribbean, and Europe & Central Asia regions. The increase in number of trips from North America to the East Asia & Pacific region in quarters 1 and 4 from 2020 quarter 4 onwards reflects the same pattern seen in the world country networks, where there is an increase in the number of trips from the United States (North America region) to Australia (East Asia & Pacific region) during the same quarters. The main destinations from Europe & Central Asia throughout the four-year period are other countries in Europe & Central Asia (70%) and North America (15%), with the proportion of trips to each region remaining fairly consistent through the four-year period. Tourists from the Latin America & Caribbean region tended to visit North America (67%) or other countries in the Latin America & Caribbean region (24%) prior to 2020 quarter 2, although there is a notable decrease in the proportion visiting other Latin America & Caribbean countries (to 9%) from 2020 quarter 2 onwards. This reduction is almost matched by an increase in the proportion of trips to North America. Similarly, tourists from the East Asia & Pacific region tended to visit North America or other East Asia & Pacific countries (approximately 40% each), with the number of trips to both regions decreasing by around half from 2020 quarter 2 onwards, compared to the previous period.

The seasonal patterns in the number of trips each quarter across the four-year period vary for the different regions (Table 4.28). The seasonal pattern of more trips within the North America and Europe & Central Asia regions (wider edge widths on the loops) in quarters 2 and 3 than in quarters 1 and 4 is evident throughout the four-year period (Figure 4.18). Therefore, although the number of trips within both these regions decreases in all quarters from 2020 quarter 2 onwards, anglers continued to prefer to take trips within the region in the warmer months compared to during colder months. In contrast, the seasonal pattern of a higher number of trips within the Latin America & Caribbean region in quarter 1 and the East Asia & Pacific region in quarter 4, compared to other quarters, is only evident until 2020 quarter 1. In subsequent quarters, the number of trips within these regions is low and more consistent across all quarters, especially from 2020 quarter 3 onwards. The seasonal pattern of a higher number of trips from the North America region to the Latin America & Caribbean region in quarters 2 and 3, compared to quarters 1 and 4, is stronger prior to 2020 quarter 2. From this period onwards, the seasonal pattern is more subtle, characterised by fewer trips and less variation in the number of trips between quarters, although there is a sharp decrease in the number of trips in 2020 quarter 2 and a sharp rise in 2021 quarter 4.

Communities

Community graphs The clearest finding in the communities graphs for the world regions is that there are different cluster patterns before 2020 quarter 2 compared to the period from this quarter onwards (Figure 4.19). Prior to 2020 quarter 2, the cluster patterns tend to consist of several, fairly evenly sized clusters, containing neighbouring regions. From 2020 quarter 2 onwards, there is one large cluster containing most regions and spanning a wide geographical area, together with one or two small clusters.

Prior to 2020 quarter 2 Prior to 2020 quarter 2, the networks contain three or four community clusters, with the largest cluster consisting of three or four regions, and one or two clusters consisting of just one region (Figure 4.19). An exception is 2019 quarter 4, which has five clusters, each consisting of one or two regions. Clusters tend to occur in regional groups, with nearby regions clustered together. The only exception is that South Asia joins the cluster containing the North

America and Latin America & Caribbean regions in around half the networks during this period, forming a cluster that spans a wide geographical range, and creating the only geographical overlap of clusters. However, trips from and to the South Asia region represent 1% of the total trips during the period 2018 to 2021, with approximately 75% either to or from the North America region (Table 4.19).

The North America and Latin America & Caribbean regions are always together in a cluster (Figure 4.19). The two other regions with substantial proportions of overall trips (inbound and outbound combined), Europe & Central Asia and East Asia & Pacific, are only together in the same cluster in one quarter (2018 quarter 1). Of the regions with a small proportion of the overall trips, Sub-Saharan Africa tends to be in a single-vertex cluster, while in contrast, South Asia and the Middle East & North Africa are never in single-vertex clusters. South Asia tends to join the North America and Latin America & Caribbean cluster in quarters 2, 3 and 4, and joins the cluster containing Europe & Central Asia in quarter 1. The Middle East & North Africa tends to join Europe & Central Asia.

In quarter 1, the North America and Latin America & Caribbean regions form a cluster, with Europe & Central Asia and the two regions with the fewest trips, the Middle East & North Africa and South Asia, in another cluster. The East Asia & Pacific region is in a different cluster each year. Sub-Saharan Africa, with relatively few trips, is usually in a single-vertex cluster.

In quarter 2, the North America and Latin America & Caribbean cluster also contains South Asia. Europe & Central Asia forms a single-vertex cluster. The East Asia & Pacific, Middle East & North Africa and Sub-Saharan Africa regions form two clusters, split differently each year.

In quarter 3, the North America and Latin America & Caribbean cluster continues to contain South Asia, along with Sub-Saharan Africa in 2018. Europe & Central Asia forms a cluster with the Middle East & North Africa. The East Asia & Pacific region is in a single-vertex cluster. Sub-Saharan Africa is in a single-vertex cluster in 2019.

In quarter 4, there is less consistency across the two years, although the North America and Latin America & Caribbean regions remain together in a cluster in both years, and the East Asia & Pacific and Sub-Saharan Africa regions are both in single-vertex clusters. Europe & Central Asia, the Middle East & North Africa and South Asia are in different clusters in different years.

From 2020 quarter 2 onwards From 2020 quarter 2 onwards, there is a distinct change in the clustering pattern (Figure 4.19). The number of clusters within the networks decreases and the pattern of countries grouped together changes. During this period, the networks have two or three community clusters, with one large cluster containing five or six of the seven regions, and either one or two clusters containing a single region or one cluster containing two regions. The large cluster is spread over a wide north-south and west-east geographical region. An exception is 2021 quarter 1 which has three clusters in a similar pattern to quarter 1 in previous years (2018 to 2020), showing a more disconnected network compared to the other quarters in this period, indicating travel was more regional during this quarter compared to during other quarters. Europe & Central Asia is either in a single-vertex cluster or forms a cluster with a region with a low proportion of trips (apart from in the 2021 quarter 1 anomaly), with the remaining regions forming the large cluster, showing that Europe is fairly isolated from the rest of the world during this period.

Community pairs Table 4.29 shows the number of times each region is in the same community cluster, out of a maximum of 16 (four quarters over four years). The regions are ordered approximately geographically from west to east. Table 4.30 splits this data into the periods prior to and from 2020 quarter 2 onwards, as the cluster patterns change significantly at this point.

Very strong connections Considering strong connections between regions to be when they are frequently in the same community group, Table 4.29 shows there are very strong connections between the neighbouring regions of North America and Latin America & Caribbean, which are in the same community in all quarters throughout the period 2018 to 2021.

Fairly strong connections There are fairly strong connections across the world for South Asia with the North America and Latin America & Caribbean regions throughout the four-year period, although the number of trips from and to South Asia represents only 1% of the total trips.

Moderate connections The Middle East & North Africa region is moderately connected with neighbouring regions South Asia and Europe & Central Asia over the four-year period, although it also has fewer than 1% of the total inbound and outbound trips. From 2020 quarter 2 onwards, its connections are stronger with South Asia and weaker with Europe & Central Asia, compared to the prior period.

Fairly weak connections Regions that are fairly weakly connected are geographically distant: East Asia & Pacific and Sub-Saharan Africa with all other regions (except Europe & Central Asia, where the connections are weaker); and the Middle East & North Africa region with the North America and Latin America & Caribbean regions. All these connections are weak or non-existent prior to 2020 quarter 2, and become fairly strong or moderate from this quarter onwards. Europe & Central Asia and South Asia are also fairly weakly connected.

Very weak or no connections Europe & Central Asia has very weak connections with the East Asia & Pacific and Sub-Saharan Africa regions, and no connections with the North America and Latin America & Caribbean regions throughout the four-year period.

In summary, prior to 2020 quarter 2, there are very strong connections between the North America and Latin America & Caribbean regions, with weak or no connections between other region pairs. An exception is that South Asia and the Middle East & North Africa have moderate connections with some other regions, though they each have around 1% of the total inbound and outbound trips. From 2020 quarter 2 onwards, Europe & Central Asia becomes more isolated, while the rest of the regions become more connected with each other.

Table 4.28 Metrics for world regions network graphs based on number of trips.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)	Number of trips
2018	1	7	41	0.837	0.837	1.17	2	1702
2018	2	7	39	0.796	0.796	1.21	2	3844
2018	3	7	40	0.816	0.816	1.21	2	5228
2018	4	7	39	0.796	0.796	1.21	2	2134
2019	1	7	40	0.816	0.816	1.21	2	1990
2019	2	7	40	0.816	0.816	1.19	2	4522
2019	3	7	40	0.816	0.816	1.19	2	5725
2019	4	7	40	0.816	0.816	1.21	2	2213
2020	1	7	40	0.816	0.816	1.19	2	2080
2020	2	7	37	0.755	0.755	1.26	2	3070
2020	3	7	36	0.735	0.735	1.29	2	3851
2020	4	7	35	0.714	0.714	1.31	2	2153
2021	1	7	34	0.694	0.694	1.33	2	1910
2021	2	7	34	0.694	0.694	1.33	2	3536
2021	3	7	34	0.694	0.694	1.33	2	4271
2021	4	7	35	0.714	0.714	1.29	2	2337

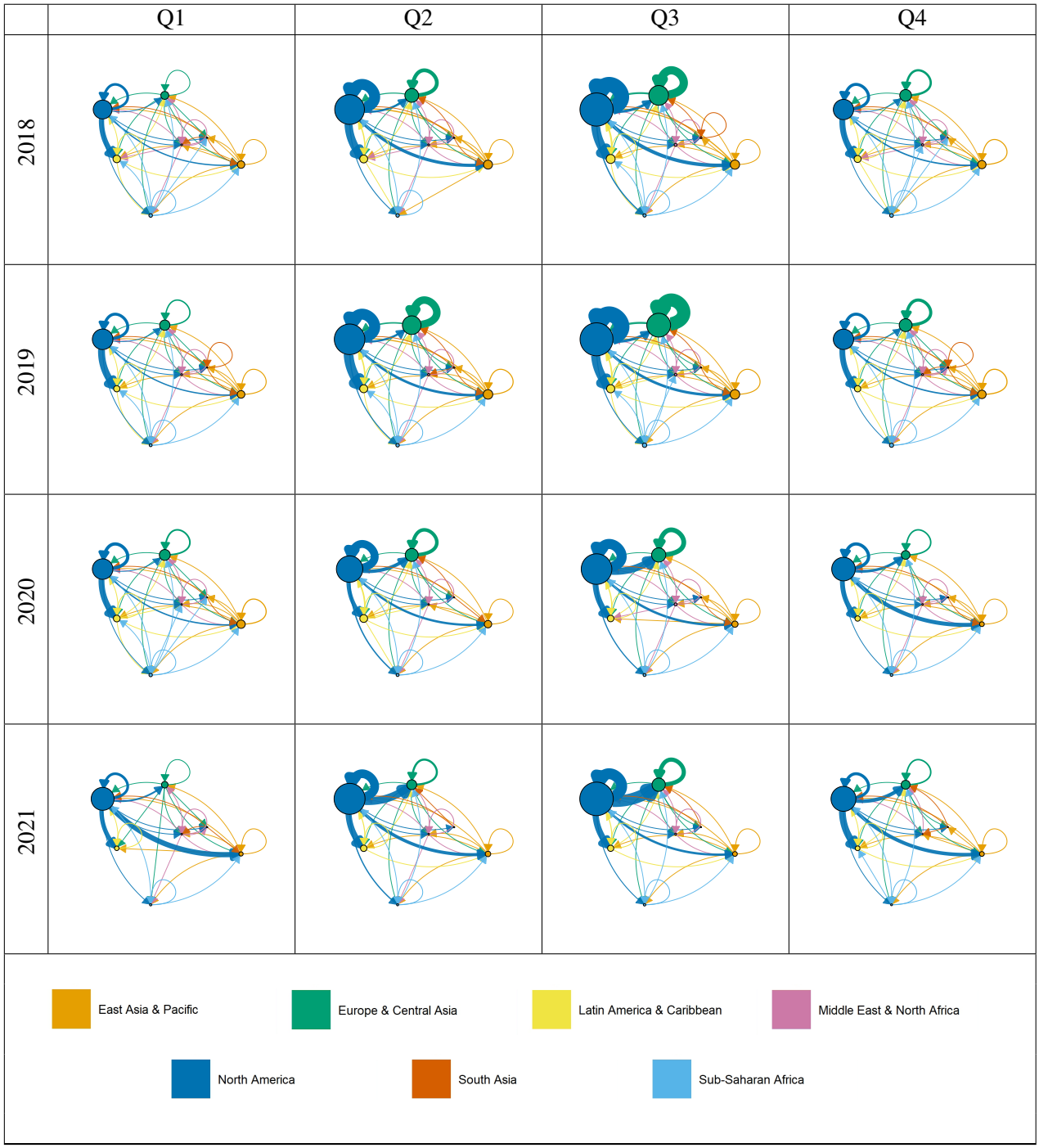


Fig. 4.18 Network graphs of world regions, based on the number of trips between regions. The area of each vertex represents the total number of trips from that region to other regions. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The loops indicate the trips to a different country to the origin country, but still in the same region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table 4.29 World regions: trips. Number of quarters (maximum 16) in which the corresponding pair of regions is in the same community from 2018-2021.

	North America	Latin America & Caribbean	Europe & Central Asia	Middle East & North Africa	South Asia	East Asia & Pacific
Latin America & Caribbean	16					
Europe & Central Asia	0	0				
Middle East & North Africa	5	5	8			
South Asia	11	11	4	10		
East Asia & Pacific	6	6	1	6	6	
Sub-Saharan Africa	5	5	1	4	5	4

Table 4.30 Comparing the communities in the networks of world regions based on the number of trips before and during the COVID-19 pandemic. Number of quarters in which the corresponding pair of regions is in the same community: (a) 2018 quarter 1 to 2020 quarter 1 inclusive (b) 2020 quarter 2 to 2021 quarter 4 inclusive.

(a) Number of quarters (maximum 9) in which the corresponding pair of regions is in the same community, from 2018 quarter 1 to 2020 quarter 1 inclusive.

	North America	Latin America & Caribbean	Europe & Central Asia	Middle East & North Africa	South Asia	East Asia & Pacific
Latin America & Caribbean	9					
Europe & Central Asia	0	0				
Middle East & North Africa	0	0	6			
South Asia	5	5	3	4		
East Asia & Pacific	0	0	1	2	1	
Sub-Saharan Africa	1	1	0	1	1	1

(b) Number of quarters (maximum 7) in which the corresponding pair of regions is in the same community, from 2020 quarter 2 to 2021 quarter 4 inclusive.

	North America	Latin America & Caribbean	Europe & Central Asia	Middle East & North Africa	South Asia	East Asia & Pacific
Latin America & Caribbean	7					
Europe & Central Asia	0	0				
Middle East & North Africa	5	5	2			
South Asia	6	6	1	6		
East Asia & Pacific	6	6	0	4	5	
Sub-Saharan Africa	4	4	1	3	4	3

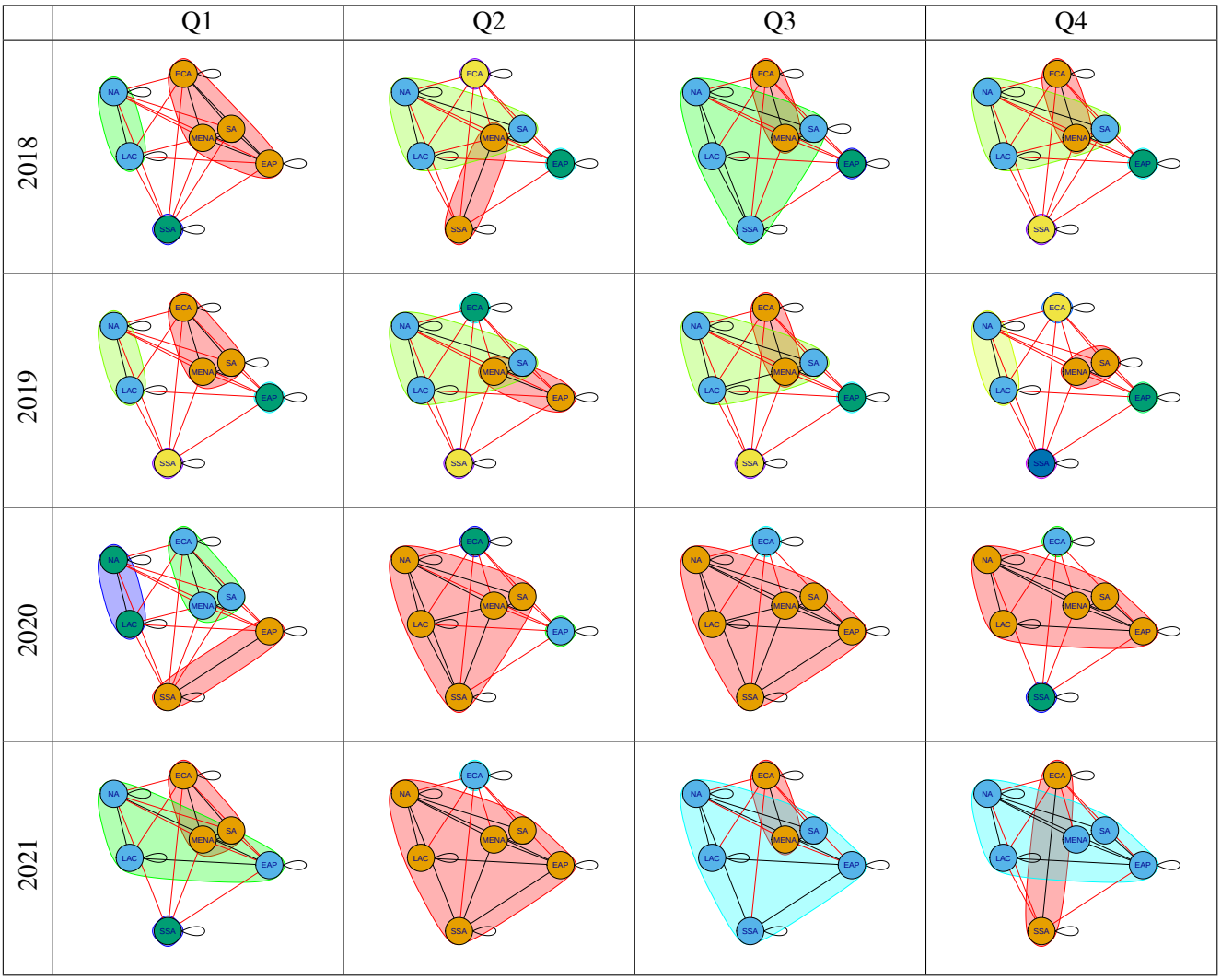


Fig. 4.19 Communities in the world regions networks based on the number of trips (2018-2021). The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa

4.3.5 Summary of results

United Kingdom

Seasonal patterns Seasonal patterns show that there are more connections and more trips from the United Kingdom to countries in the Europe & Central Asia and North America regions during the northern hemisphere warmer months and fewer during the colder months, compared to the rest of the year. There are more trips to Australia during the United Kingdom colder months, which correspond with the warmer Australian months, compared to the rest of the year. These indicate a preference to travel during the warmest months of the destination country.

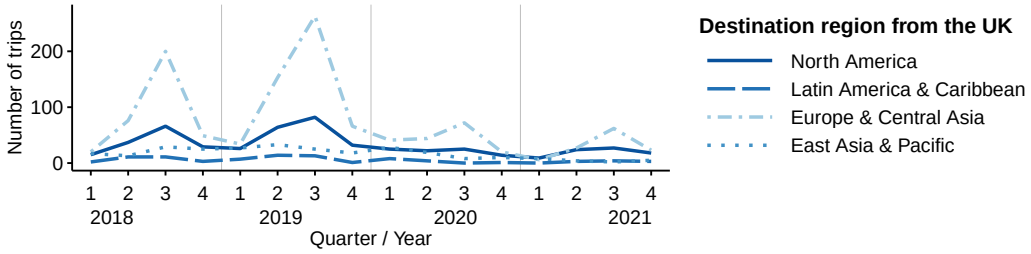
Trends over time From 2020 quarter 3 onwards, there is a large reduction in the number of connections to and from the United Kingdom compared to previously. There is also a large reduction in the number of trips during this period, although trips to the United Kingdom from the United States increased.

Figure 4.20 shows (a) the number of trips to each region by United Kingdom users, and (b) the number of trips to the United Kingdom by users from other regions, across the four-year period from 2018 to 2021. There are seasonal patterns in the number of trips from the United Kingdom to all regions in 2018 and 2019, and to the Europe & Central Asia region in 2020 and 2021, although the number of trips decreases sharply in 2020 and 2021 compared to 2018 and 2019. In contrast, the number of trips to the United Kingdom from the North America region increases in 2020 and 2021 compared to 2018 to 2019, due to the increase from the United States.

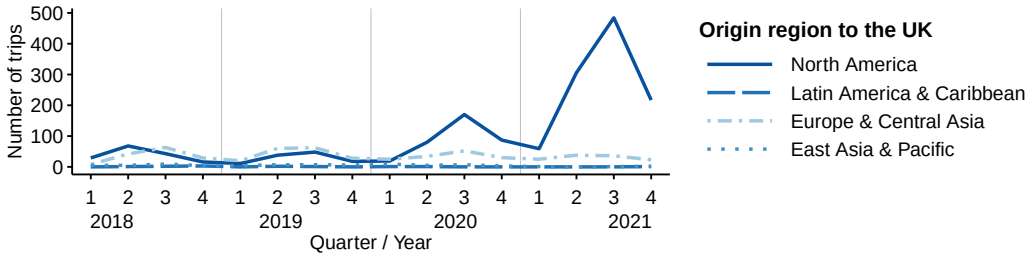
United States

Seasonal patterns Seasonal patterns show that there are more connections and more trips between the United States and other countries during the United States spring and summer (quarters 2 and 3) than in the autumn and winter (quarters 1 and 4).

Trends over time The number of connections to popular destinations (at least 30 trips recorded in a quarter) decreases from 2018 to 2019, remains fairly constant from 2019 to 2020, and increases from 2020 to 2021. Costa Rica is a popular destination from the United States (at least 30 trips recorded in



(a) Number of outbound trips from the United Kingdom.



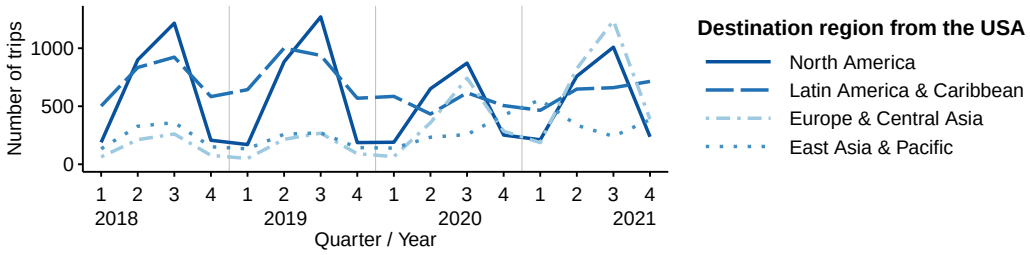
(b) Number of inbound trips to the United Kingdom.

Fig. 4.20 Trips between the United Kingdom and world regions.

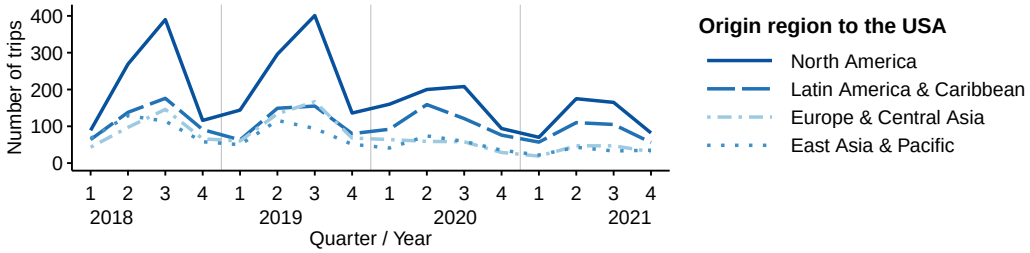
a quarter) prior to 2021 quarter 2, while Brazil is generally absent from the networks (fewer than 30 trips per quarter). From 2021 quarter 2 onwards this pattern reverses, and Costa Rica is less popular while Brazil becomes more popular with at least 30 trips recorded in most quarters. The United Kingdom, Norway, Sweden and Finland increase in popularity (number of connections and number of trips) from 2020 quarter 2 onwards, and Ireland increases in popularity in 2021 quarters 2 and 3. There is an increase in the number of trips recorded from the United States to Australia from 2020 quarter 3 onwards, with a substantial increase in quarters 1 and 4 during this period. Connections between the United States and Canada and from the United States to Mexico are lower in 2021 quarter 1 compared to other quarters. There is a large number of trips from the United States to Brazil in 2021 quarter 4 compared to other quarters, which coincides with fewer trips to Costa Rica and Puerto Rico in this quarter.

Figure 4.21 shows (a) the number of trips to each region by users from the United States, and (b) the number of trips to the United States by users from other regions, across the four-year period from 2018 to 2021. There are seasonal patterns in the number of trips from the United States to all regions in 2018 and 2019, and to Europe & Central Asia and the North America regions in 2020 and 2021

(where the North America region here is dominated by trips to Canada). In 2020 and 2021, there is a decrease in the number of trips to the Latin America & Caribbean and North America regions, and an increase to Europe & Central Asia and the East Asia & Pacific region, compared to 2018 and 2019. In contrast, the seasonal pattern in the number of trips to the United States from all regions occurs throughout the four-year period, despite trips from all regions decreasing in 2020 and 2021 compared to 2018 to 2019 (although the decrease from the Latin America & Caribbean region is much weaker compared to from the other regions).



(a) Number of outbound trips from the United States.



(b) Number of inbound trips to the United States.

Fig. 4.21 Trips between the United States and world regions.

Europe

Networks

Seasonal patterns Seasonal patterns show more connections and more trips between connected countries in summer and fewer in winter, compared to the rest of the year.

Trends over time There is a reduction in travel in 2020, with fewer connections in all quarters from 2020 quarter 3 onwards, and around 50% fewer trips overall from 2020 quarter 2 onwards,

compared to 2018 and 2019. Ireland is the only country with an increase in outbound trips during this period. The top origin countries are Sweden and the United Kingdom, with Sweden and Norway being the top destination countries.

Communities

Community graphs Community clusters tend to have a cluster containing the Nordic and Baltic countries, a cluster containing countries across the entire west-east range, including the United Kingdom, Ireland and much of mainland Europe, and a smaller cluster of countries in the middle of the geographical region. However, there is variation in this pattern, especially in quarters 1 and 4 and from 2020 quarter 2 onwards.

Community pairs Neighbouring countries by land or sea tend to be connected, with strong connections between Ireland, United Kingdom and France, and also between Switzerland, Belgium, the Netherlands and Germany, with the strongest connections in this group between closest neighbours. There are also strong connections between the Nordic (Denmark, Finland, Norway and Sweden) and northern CEE (Latvia, Lithuania and Poland) countries, all of which (except Norway) have coastline on the Baltic Sea. Italy is moderately connected with Croatia across the Adriatic Sea, though only weakly connected with Mediterranean neighbours Spain and Greece.

There are also some notable connections between countries that are not in the local geographical area. Ireland, the United Kingdom and France are strongly connected with Spain and moderately connected with Greece and Romania, while Switzerland, Belgium, Netherlands and Germany are moderately connected with Italy and Croatia. Across the Mediterranean, there are strong connections between Spain and Greece.

World countries

Networks

Seasonal patterns The total number of connections and trips is highest in quarter 3, followed by quarter 2, with the fewest in quarter 1. This shows that travel increases during northern hemisphere

warmer months and is less during its cooler months. This may be due to a high proportion of countries where catches are recorded being in the northern hemisphere as travel patterns to individual countries show that there is more travel in the warmer months of the destination country than in the colder months. This reflects the pattern of more trips in warmer months and fewer in cooler months seen in the Europe networks.

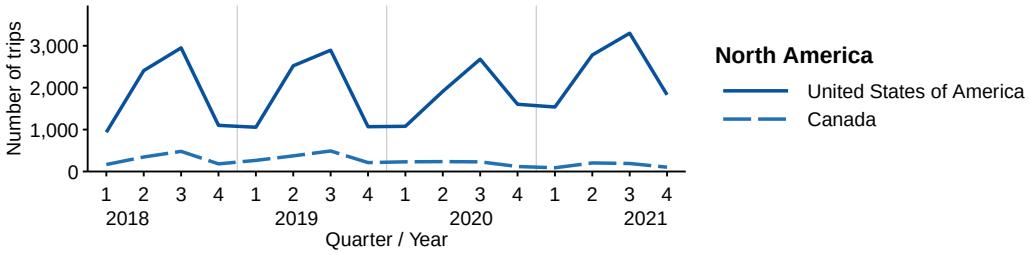
Trends over time The number of connections increases from 2018 to 2019, then decreases from 2020 quarter 2 onwards, and is especially low in 2021 quarter 1, however the overall reduction is less strong than seen in the Europe networks. The reduction in connections is greater in the north-south direction than in the west-east direction, especially the connections of Australia, New Zealand and Brazil with Europe.

The overall number of trips increases from 2018 to 2019, decreases in 2020 quarters 2 and 3 (by just over 25%), and then starts to increase again from 2020 quarter 4 onwards, compared to corresponding quarters in the previous year, with the number of trips in 2021 quarter 4 higher than in this quarter previously. While the number of trips from most countries decreases from 2020 quarter 2 onwards, there is an increase in the number of trips from the United States and Ireland, with an substantial increase from both these countries to the United Kingdom. Trips from the United States to Australia increase in quarters 1 and 4 (Australian warmer months) from 2020 quarter 4 onwards, and to Brazil in 2021 quarter 4, which is also seen in the United States case study. The United States and Canada have the highest numbers of outbound and inbound trips recorded. The United States is a popular destination from Canada and Mexico, and Canada and Mexico are popular destinations for United States tourists.

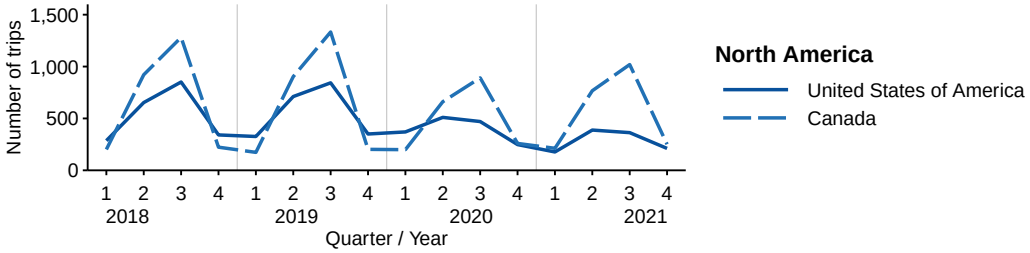
Figures 4.22 to 4.25 show the number of recreational fishing trips from the top origin countries and to the top destination countries in the four regions with the highest number of trips. Figures C.1 to C.4 in Appendix C illustrate the outbound and inbound tourism trends for these countries from 2018 to 2021. These demonstrate the number of tourists (individuals staying overnight for a minimum of one night) or, where data on the number of tourists is not available, the total number of tourists and visitors (individuals on same-day excursions, with no overnight stays) (World Tourism Organization, 2023). The number of outbound fishing trips from the United States remained relatively

stable across the four-year period (Figure 4.22a), despite a sharp decline in general tourism in 2020, with no data available for 2021 (Figure C.1a). In contrast, declines in outbound fishing trips from other countries broadly reflected their general outbound tourism trends during 2020 and 2021, with substantial declines observed in Australia and the Europe & Central Asia countries. More moderate declines occurred in Mexico and Brazil, although the overall number of outbound fishing trips from Latin America & Caribbean countries is relatively low across the four-year period, and general outbound tourism data are unavailable for Brazil. Fishing trips from Puerto Rico remained consistent, mirroring the general outbound tourism trend here (Figures 4.24a and C.3a).

Most countries experienced declines in inbound fishing trips during 2020 and 2021, which were more moderate than the trends in inbound general tourism to these countries (World Tourism Organization, 2023). In contrast, fishing tourism to the United Kingdom increased in 2020 compared to 2018-2019, followed by a considerable further increase in 2021 (Figure 4.23b). Conversely, general inbound tourism to the United Kingdom sharply declined in 2020, then declined slightly further in 2021 (Figure C.2b). Inbound fishing trips to Sweden remained consistent between 2019 and 2020, despite a decline in general tourism, but both fishing and general tourism rose in 2021. The notable increases in fishing trips to Australia in 2020 and 2021 (Figure 4.25b) and to Brazil at the end of 2021 (Figure 4.24b) are not seen in general inbound tourism trends to these countries (Figures C.4b and C.3b).

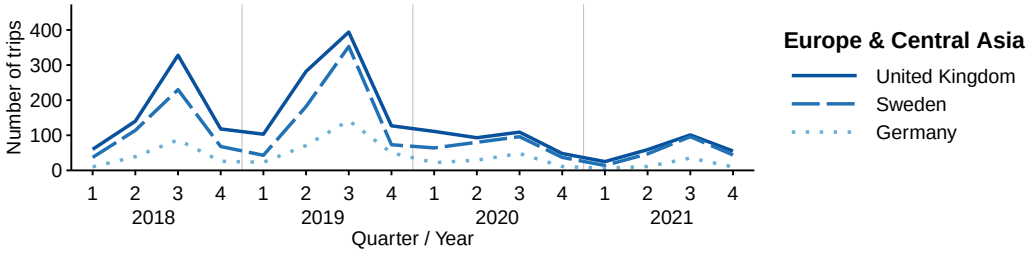


(a) Origin countries (outbound trips).

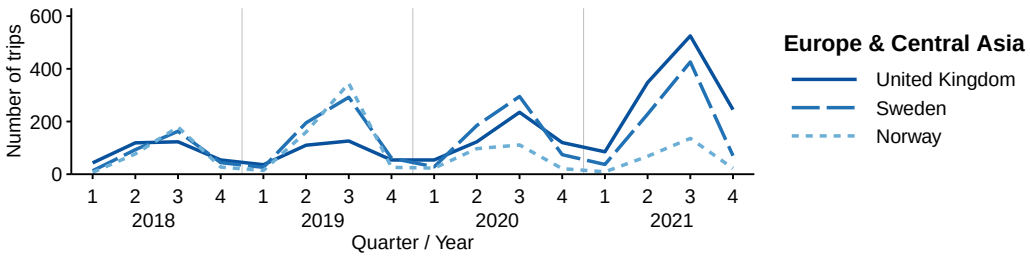


(b) Destination countries (inbound trips).

Fig. 4.22 Number of trips from the top origin countries and to the top destination countries in the North America region.

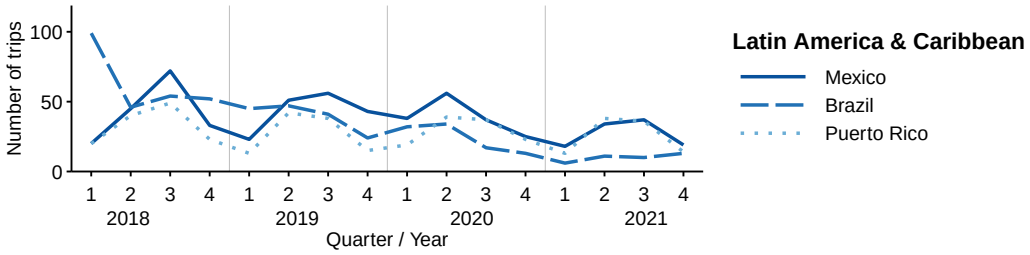


(a) Origin countries (outbound trips).

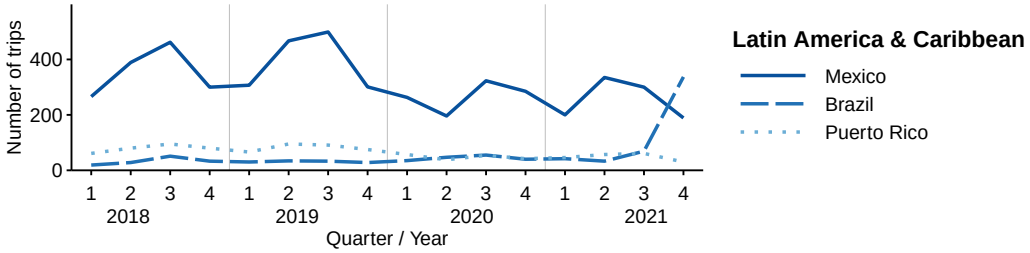


(b) Destination countries (inbound trips).

Fig. 4.23 Number of trips from the top origin countries and to the top destination countries in the Europe & Central Asia region.

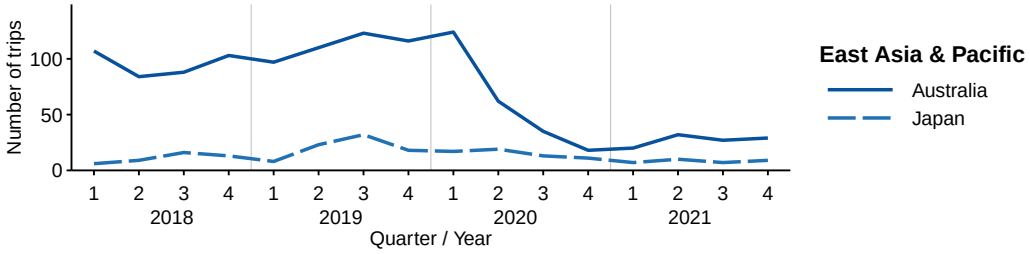


(a) Origin countries (outbound trips).

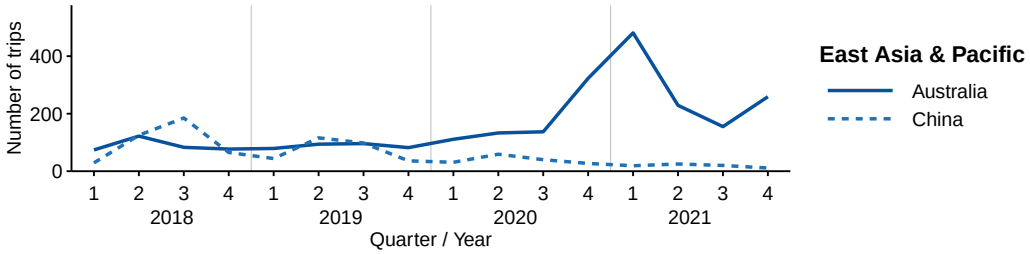


(b) Destination countries (inbound trips).

Fig. 4.24 Number of trips from the top origin countries and to the top destination countries in the Latin America & Caribbean region.



(a) Origin countries (outbound trips).



(b) Destination countries (inbound trips).

Fig. 4.25 Number of trips from the top origin countries and to the top destination countries in the East Asia & Pacific region.

Communities

Community graphs Clusters tend to occur in regional groups and in the north-south direction. Prior to 2020 quarter 3 there are seasonal patterns in the clusters, with the Americas and China in one cluster, and the rest of the world forming one cluster in quarters 1 and 4 and two clusters in quarter 3 (with the United Kingdom, Ireland, France, Australia, New Zealand, South Africa in one cluster and the rest of Europe in the other). The patterns across quarter 2 have more inconsistencies than other quarters, which is also the case in the world regions community patterns. Over the period from 2020 quarter 3 to 2021 quarter 1, there is a progression from the split between the Americas and China with the rest of the world to a split between mainland Europe and the rest of the world. From 2021 quarter 2, mainland Europe is in one cluster and the rest of the world is in another cluster.

Community pairs Countries in the same geographical area are often connected, with strong connections between countries in the Americas, between the United Kingdom and Ireland, and between Australia and New Zealand. The strong connections between the Nordic countries (Denmark, Finland, Norway and Sweden) seen in the Europe networks are seen in the world networks, where they are also strongly connected with Germany and Poland. There are also strong connections between China and the Americas, and fairly strong connections between South Africa with Australia and New Zealand; and the United Kingdom and Ireland with Australia, New Zealand and South Africa, which are not in the same geographical area.

The impact of geographical distance is further seen in France having fairly strong connections with all European countries, while the United Kingdom and Ireland only have moderate connections with countries in mainland Europe.

Australia, New Zealand and South Africa are all fairly weakly connected with the Americas, China and mainland Europe (an exception is France, with which they have moderate connections). The weakest connections are between Europe with the Americas and China. The United Kingdom and Ireland are rarely in a community with the Americas and China, and mainland Europe countries are never in a community with the Americas and China.

World regions

Networks

Seasonal patterns The number of connections is fairly consistent across quarters, with no clear seasonal patterns. However, there are seasonal patterns in the number of trips, with a preference for trips taken in the destination's warmer months, compared to its cooler months, reflecting the pattern seen in the world countries networks (Section 4.3.3). Overall, each year, users take the most trips in quarter 3, followed by a substantial number in quarter 2, and the fewest trips in quarter 1. This reflects the seasonal trip patterns of the regions with the largest proportion of Fishbrain users (North America and Europe & Central Asia).

In terms of trips within regions, there are more trips within regions during warmer months compared to colder months. Trip frequency is higher in quarters 2 and 3 in the northern hemisphere regions of North America and Europe & Central Asia, and in quarters 4 and 1 in the Latin America & Caribbean and East Asia & Pacific regions, which are both mainly situated in the tropics and southern hemisphere. The seasonal variation in the trips within these regions is less pronounced than for trips within the northern hemisphere regions. However, it is important to note the lower number of trips recorded in these regions (434 and 1,089 respectively across the four-year period) compared to the northern hemisphere regions (12,204 within North America and 6,454 within Europe & Central Asia).

In terms of seasonal patterns in the number of trips between regions, there is a higher preference to travel south from North America to the Latin America & Caribbean region during the North American spring and summer, than in autumn and winter.

Trends over time Prior to 2020 quarter 2, the number of connections is relatively stable. From 2020 quarter 2 onwards, there is a decrease in the number of connections, primarily to and from the Latin America & Caribbean region, with the number of connections between the world regions lowest in 2021 quarter 1. There is no change in the number of connections from North America, Europe & Central Asia and East Asia & Pacific over the four-year period.

The number of trips from all regions, except North America, decreases from 2020 quarter 2 onwards. North America experiences an initial decline in the first two quarters of this period, but

the number of outbound trips is subsequently higher than in the corresponding quarters of previous years. Overall, the total number of trips recorded from all regions decreases by approximately 15% compared to 2018-2019. Throughout the four-year period, the North America and Europe & Central Asia regions consistently have the highest number of outbound trips each year.

The regions with highest number of inbound trips are North America, Latin America & Caribbean and Europe & Central Asia. From 2020 quarter 2 onwards, the number of trips to the North America and Latin America & Caribbean regions declines, while trips to Europe & Central Asia increase, compared to the previous period. The most popular destinations from North America prior to 2020 quarter 2 are within the North America region and to the Latin America & Caribbean region, but subsequently are more evenly distributed between these regions and Europe & Central Asia. This trend aligns with the increase in trips from the United States to the United Kingdom (Figure 4.20). The main destinations from Europe & Central Asia throughout the four-year period are countries within the same region and the North America region. A decline in trips within Europe & Central Asia from 2020 quarter 2 mirrors the decrease in trips seen in the Europe networks. For tourists from the Latin America & Caribbean and East Asia & Pacific regions, the top destinations during the four-year period are the North America region and countries within their own region, although from 2020 quarter 2 onwards the proportion of trips within the Latin America & Caribbean region decreases substantially, while the proportion to North America increases.

The seasonal pattern in the number of trips within North America and Europe & Central Asia is evident throughout the four-year period. Despite the decline in trips within Europe & Central Asia from 2020 quarter 2 onwards, the seasonal pattern remains. In contrast, the seasonal pattern within the Latin America & Caribbean and East Asia & Pacific regions is only apparent in 2018 and 2019. From 2020 quarter 2 onwards, the number of trips within these regions decreases to a low and consistent level throughout the year.

An increase in trips from North America to the East Asia & Pacific region is evident in quarters 1 and 4 from 2020 quarter 4 onwards. This trend aligns with the substantial increase in trips from the United States to Australia, as observed in the United States case study and world countries networks.

Communities

Community graphs The community cluster analysis reveals a distinct change in cluster patterns from 2020 quarter 2 onwards. Prior to this period, the networks comprise fairly evenly sized clusters containing neighbouring regions, and one or two regions in single-vertex clusters. From 2020 quarter 2 onwards, the networks typically feature one large cluster consisting of five or six regions across the world, and either one or two regions in single-vertex clusters or a small cluster containing two regions. An exception occurs in 2021 quarter 1, which has a pattern more aligned with those prior to 2020 quarter 2.

Prior to 2020 quarter 2, networks typically contain three or four clusters, with each cluster consisting of up to three neighbouring regions. The only exception occurs when South Asia, which accounts for fewer than 1% of the total inbound and outbound trips, is included in the cluster with the geographically distant North America and Latin America & Caribbean regions. This situation arises in around half of the networks, forming a cluster that spans a wide geographical range, and creating the only geographical overlap of clusters during this period.

During this period, the North America and Latin America & Caribbean regions, which are in close proximity geographically, are always in the same cluster. In contrast, Europe & Central Asia, is often in a single-vertex cluster, despite its close proximity to other regions. When Europe & Central Asia is not in a single-vertex cluster, the neighbouring Middle East & North Africa region is always included. Additionally, South Asia primarily joins Europe & Central Asia when it is not with the Americas.

The East Asia & Pacific and Sub-Saharan Africa regions are more frequently in single-vertex clusters than in clusters with other regions. This tendency may be due to the countries with the highest numbers of inbound or outbound trips in these regions (Australia, New Zealand, Japan, China, United States Minor Outlying Islands, and Republic of Korea in the East Asia & Pacific region; and South Africa in Sub-Saharan Africa) being geographically distant from other regions. When in a cluster with other regions, they are typically with geographically closer regions.

The Middle East & North Africa and South Asia, the two regions with the lowest inbound and outbound trips, are never in single-vertex clusters. This may result from their proximity to other regions or relatively low number of trips. South Asia clusters with Europe & Central Asia in quarter 1,

and typically clusters with the geographically distant North America and Latin America & Caribbean regions in quarters 2, 3, and 4. Notably, South Asia is the only region that clusters with every other region at some point. The Middle East & North Africa is often with neighbouring Europe & Central Asia and South Asia, and sometimes with the East Asia & Pacific region to the east or Sub-Saharan Africa to the south. Unlike South Asia, the Middle East & North Africa is never in a cluster with the Americas across the Atlantic Ocean.

Regions which are never in a cluster together during this period are often geographically distant. The North America and Latin America & Caribbean regions are never in a cluster with Europe & Central Asia or the Middle East & North Africa across the Atlantic Ocean, nor with the East Asia & Pacific region on the opposite side of the world. Similarly, Europe & Central Asia is never clustered with geographically distant Sub-Saharan Africa, which lies beyond the Middle East & North Africa to the south.

From 2020 quarter 2 onwards, the clustering pattern changes markedly. The number of clusters decreases, with one large cluster emerging that contains most regions, while one or two regions form either a small cluster or remain as single-vertex clusters. This contrasts sharply with the higher number and fairly evenly sized clusters in the previous period. Notably, 2021 quarter 1 presents an exception, reverting to a pattern similar to the previous period. Europe & Central Asia consistently forms a single-vertex cluster from 2020 quarter 2 to 2021 quarter 2, with 2021 quarter 1 being the only exception. This suggests reduced connectivity between Europe & Central Asia and other regions compared to the period prior to 2020 quarter 2. Subsequently, Europe & Central Asia is in a cluster with another region, indicating an increase in connectivity with other parts of the world in 2021 quarters 3 and 4.

Throughout both periods, the North America and Latin America & Caribbean regions are consistently clustered together. During the period prior to 2020 quarter 2, this cluster primarily includes South Asia in quarters 2, 3 and 4, with Sub-Saharan Africa joining once. Notably, the Europe & Central Asia, East Asia & Pacific and Middle East & North Africa regions never join this cluster during this period. From 2020 quarter 2 onwards, most regions are in the same cluster as the Americas, except for Europe & Central Asia, which is never part of this cluster throughout the four-year period. Interestingly, although the East Asia & Pacific region is never in the same cluster as the Americas

prior to 2020 quarter 3, from this quarter onwards it consistently joins the Americas cluster. This includes 2021 quarter 1, which exhibits a cluster pattern similar to the period prior to 2020 quarter 2, in contrast to the patterns observed during subsequent quarters.

Additionally, the Europe & Central Asia, East Asia & Pacific and Sub-Saharan Africa regions are in single-vertex clusters in some quarters during both periods. However, Europe & Central Asia forms a single-vertex cluster more frequently from 2020 quarter 2 onwards, while Sub-Saharan Africa and East Asia & Pacific are less frequently in single-vertex clusters during this period. This shift reflects the increased isolation of Europe & Central Asia from other the regions. The Middle East & North Africa and South Asia, both of which have fewer than 1% of inbound and outbound trips, are never in single-vertex clusters throughout the four-year period.

In summary, the analysis reveals distinct clustering patterns. Specifically, throughout the period prior to 2020 quarter 2, geographical proximity often correlates with clustering patterns, although there are exceptions. From 2020 quarter 2 onwards, these patterns shift significantly, with increased connectivity among most regions and a notable isolation of Europe & Central Asia. Overall, these observations highlight the dynamic nature of regional connectivity and the links between geography and cluster patterns.

Community pairs The strength of connections between each pair of world regions describes the number of times the regions are in the same community over the four-year period. The pattern of connections between world regions is substantially different in the period prior to 2020 quarter 2 onwards, compared to this quarter onwards.

The neighbouring North America and Latin America & Caribbean regions consistently exhibit strong connections throughout the four-year period. Prior to 2020 quarter 2, connections among most other regions are very weak. The East Asia & Pacific region, which is geographically relatively isolated from other regions, has only very weak connections with them. Similarly, Sub-Saharan Africa, primarily represented by South Africa in terms of trip numbers, displays weak connections with other regions, possibly due to South Africa's geographical remoteness from other regions. Despite its central location and proximity to other regions, Europe & Central Asia is also weakly connected with most regions. The two regions with the lowest trip numbers exhibit stronger connectivity patterns: the

Middle East & North Africa is fairly strongly connected with Europe & Central Asia, while South Asia has moderate connections with most regions and is the only region to have notable connections with the Americas during this period.

From 2020 quarter 2 onwards, the pattern of connections changes markedly. The strength of connections between Europe & Central Asia and other regions declines. Conversely, the strength of connections among the other regions increases. Particularly noteworthy is the development of strong connections between the Americas and the East Asia & Pacific, Middle East & North Africa, and Sub-Saharan Africa regions, with which there were previously very weak or no connections.

In summary, the cluster analysis reveals a significant shift in global connectivity patterns from 2020 quarter 2, characterised by a marked increase in inter-regional connectivity among most regions, particularly involving the Americas, contrasted by the isolation of Europe & Central Asia.

4.3.6 Comparing edge density in networks

The patterns in the previous section can be summarised by considering the changes in edge density over time (Figure 4.26). The edge density in the world regions networks decreases from 2020 quarter 2 onwards, which reflects fewer connections between the regions, with a slight upturn towards the end of 2021. For world countries and Europe, the decrease is seen slightly later, in 2020 quarter 3, and a clear seasonal cycle is also visible with more connections in quarter 3 (northern hemisphere summer) and fewer in quarter 1 (northern hemisphere winter) compared to the rest of the year. For the case study countries, there is a seasonal pattern prior to the start of the COVID-19 pandemic, but from 2020 quarter 2 onwards, the edge density in the United Kingdom decreases, while for the United States it increases, showing a decrease in the number of connections to/from the United Kingdom and an increase to/from the United States.

The absolute values of edge density differ among the world regions, world countries, Europe, United Kingdom, and United States networks due to the data selection method impacting the number of vertices. There are only seven world regions, with no minimum number of trips between regions, so the graph is densely connected. The world countries and Europe networks are based on the top 15 origin and top 15 destination countries, and therefore have a maximum of around 20 vertices, while the United Kingdom and United States focus on trips to and from a single country.

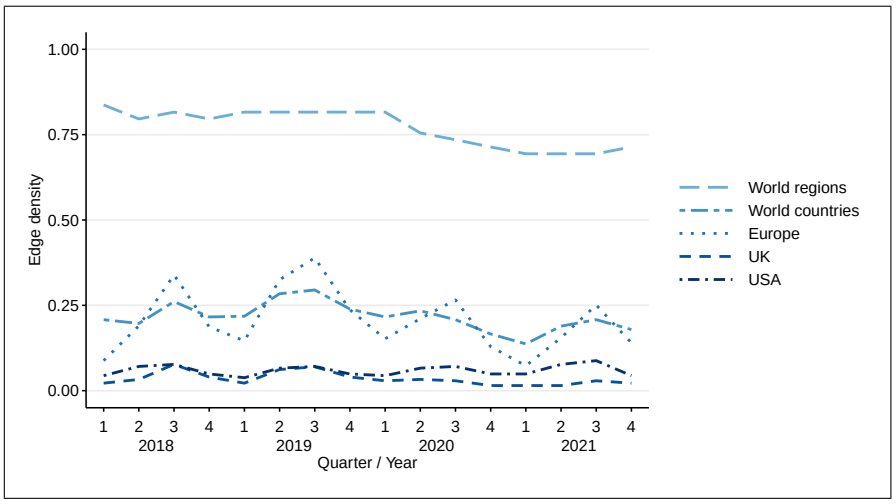


Fig. 4.26 Edge density (max vertices) for the networks based on the number of trips.

4.3.7 Comparing modularity in communities

The modularity scores of the clustering results using the greedy optimisation of modularity algorithm for each quarter are shown in Figures 4.27, 4.28, and 4.29, based on the weighted and unweighted graphs.

According to the modularity scores, the weighted graphs give better clustering for the Europe networks than the unweighted graphs, throughout the four-year period. The modularity scores of the cluster patterns in the weighted graphs are typically around 0.3 or above, indicating possible significant community structure in the networks (Clauset et al., 2004; Newman, 2004), while the modularity scores for the unweighted graphs are almost consistently below 0.3 throughout the four-year period.

In contrast, for the world countries and world regions, the weighted graphs tend to have higher modularity scores only prior to 2020 quarters 2 and 3, respectively. From these periods onwards, the unweighted graphs have higher modularity scores, indicating these generate better clustering patterns than the weighted graphs during these periods. However, for both the world countries and world regions, the modularity scores typically do not exceed 0.3 with either the weighted or unweighted graphs, throughout the four-year period.

The modularity scores for clusters created using the greedy optimisation of modularity and Louvain algorithms are almost identical for all cases (i.e. for Europe, world countries and world regions, using weighted and unweighted networks, in all quarters). Modularity scores for the communities created using the betweenness and propagating labels algorithms are lower than for the communities created with the greedy optimisation of modularity and Louvain algorithms in all quarters for all cases.

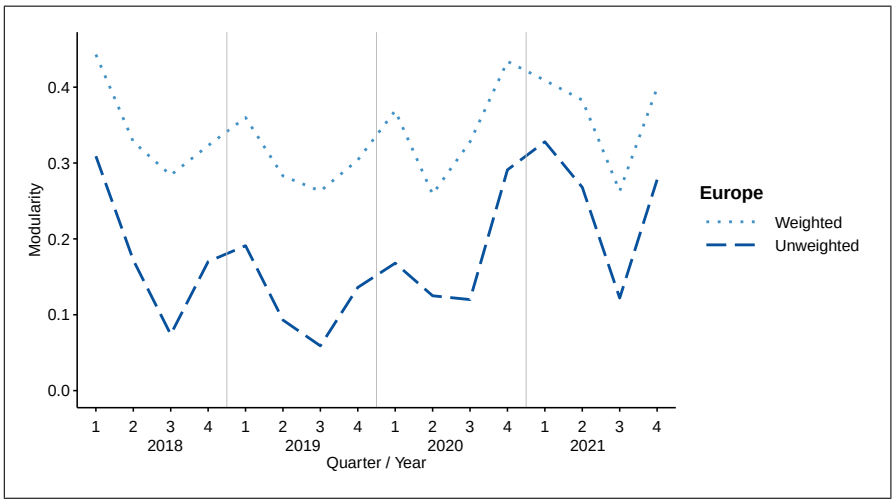


Fig. 4.27 modularity scores of the community clustering networks for Europe based on the number of trips.

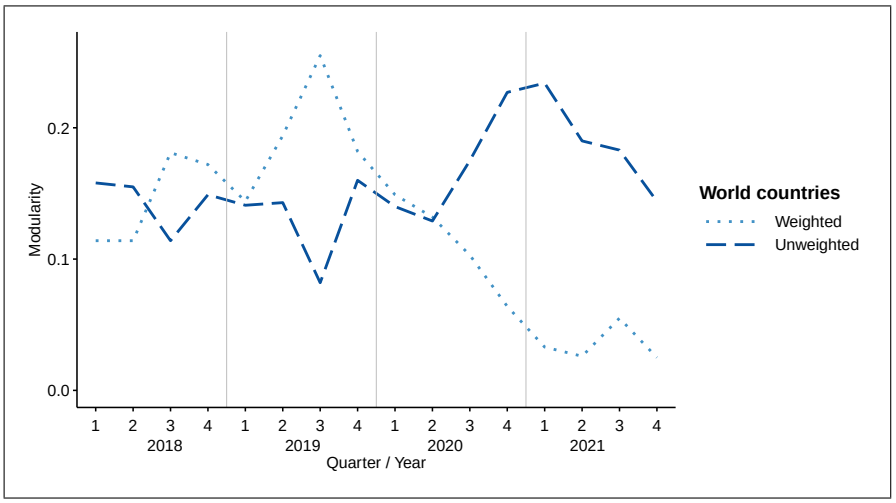


Fig. 4.28 modularity scores of the community clustering networks for the world countries based on the number of trips.

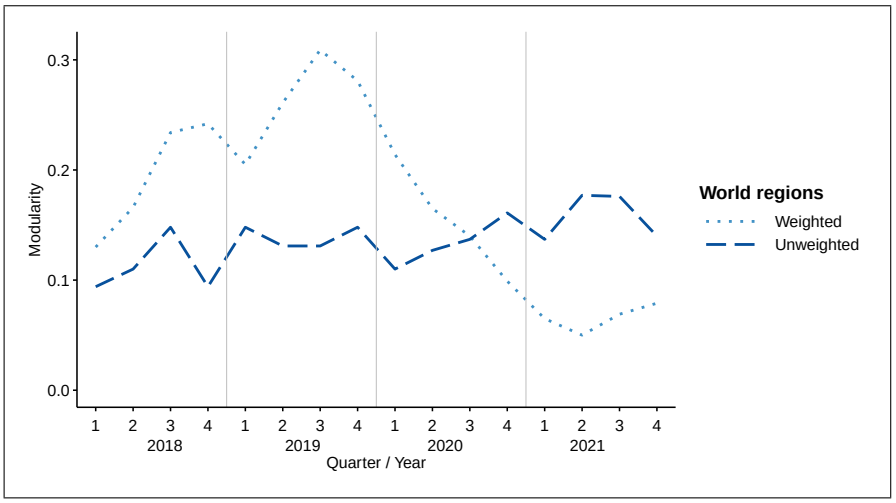


Fig. 4.29 modularity scores of the community clustering networks for the world regions based on the number of trips.

4.4 Discussion

4.4.1 Findings

This chapter explored seasonal patterns and the impact of the COVID-19 pandemic on international fishing tourism, giving new insights into travel patterns and trends. Utilising network models, we analysed travel patterns to and from the United Kingdom and United States, within Europe, and globally, focusing on both trip numbers and catches recorded over a four-year period from 2018 to 2021. The findings highlighted preferences among anglers for warmer destinations during cooler months. As expected, there was a considerable reduction in international travel for recreational fishing from 2020 quarter 2 onwards, particularly evident in Europe, indicating the pandemic's disruptive influence on international travel networks. However, intriguing variations in the pandemic's impact emerge between Europe and global networks, suggesting differential responses to travel restrictions and containment measures across regions. Community clustering analysis revealed a tendency for countries and regions geographically closer to cluster together, indicating that travel is more frequently regional than long-distance.

Comparison with general tourism

Data from the World Tourism Organization (UNWTO) (UN Tourism, 2021) (now known as UN Tourism[1]) show that, overall, the number of international tourists worldwide decreased in 2020 by 74% compared to 2019 (UN Tourism, 2021). There was little improvement in 2021, with the number of international tourists worldwide 72% lower overall compared to 2019. The change in the overall number of fishing trips recorded worldwide by Fishbrain users was more moderate, decreasing by 23% in 2020 compared to 2019, with an increase of 8% in 2021 compared to 2020, although trips in 2021 remained 17% below 2019 levels.

According to the UNWTO, the decreases in international tourism in 2020 compared to 2019 varied by region, with a 69% decrease in travel to the Americas, a 70% decrease to European countries, and the greatest decrease of 84% in travel to Asia and the Pacific (UN Tourism, 2021). The decline in the number of trips by Fishbrain users also varied by region, with more substantial differences

[1]The World Tourism Organization (UNWTO) changed its name to UN Tourism on 23 January 2024.

between regions than international tourism in general, although the regions used in our analysis do not exactly match the UNWTO regions. The number of trips by Fishbrain users decreased by 26% to North America, 21% to Europe & Central Asia, 36% to the Latin America & Caribbean region, and 4% to the East Asia & Pacific region.

In 2021, increases on 2020 levels of international tourism were reported in most regions, with different rates of increase in different regions (UN Tourism, 2022). The highest reported increases in inbound tourists were in Europe and the Americas (19% and 17% increase on 2020 respectively), with the countries with the highest increases in these regions being in the Mediterranean and Central Eastern European regions, and in the Caribbean, Central and North America (UN Tourism, 2022). There were 30% more trips by Fishbrain users to countries in Europe & Central Asia in 2021 compared to 2020, which broadly aligns with the increase in general tourism to this region. However, in sharp contrast to the increase in general tourism to North America, Fishbrain users made 6% fewer trips to this region during this period.

There was a small reported increase of 12% in international tourist travel to Africa in 2021 compared to 2020 (UN Tourism, 2022). In contrast, the number of trips by Fishbrain users to both the Middle East & North Africa and Sub-Saharan Africa regions decreased by 20% over this period. A similar decrease (24%) in international tourism to the Middle East was reported over this period (UN Tourism, 2022). In Asia and the Pacific, international tourist arrivals in 2021 were 65% lower than in 2020, and 94% below levels before the pandemic (UN Tourism, 2022). In contrast, trips by Fishbrain users to the East Asia & Pacific region increased by 14% from 2020 to 2021, with levels in 2021 10% higher than in 2019. This increase is mainly due to the high number of trips by United States users to Australia during 2020 and 2021.

In summary, both general and fishing tourism experienced declines in 2020 and 2021, compared to 2019. However, the decline in fishing tourism was to a lesser extent than that of general tourism. In 2020, all world regions experienced a smaller decline in fishing tourism than in general tourism, compared to 2019. However, in 2021, while the Americas experienced an increase in general tourism compared to 2020, a decline in fishing tourism was observed. In contrast, while general tourism to Asia and the Pacific remained relatively low, fishing tourism increased.

Comparison with fishing trends

Changes in recreational fishing patterns during the COVID-19 pandemic varied spatially and were linked to levels of restrictions nationally and, in some cases, regionally (Britton et al., 2023; Pita et al., 2021). While interest in recreational fishing increased worldwide during 2020, as evidenced by a rise in internet searches (Britton et al., 2023), overall, opportunities to engage in fishing activities decreased globally (Pita et al., 2021).

During the early stages of the pandemic, many countries prohibited recreational fishing (Pita et al., 2021). However, some countries, including Denmark, Latvia, the Netherlands, Norway (Pita et al., 2021), and the USA (Paradis et al., 2020), allowed and, in some cases, encouraged it. After initial restrictions were lifted, several countries noted an increase in participation (Britton et al., 2023; Pita et al., 2021). Through 2020, varying levels of restrictions on travel and activities led to differing impacts on fishing effort (Britton et al., 2023). For example, Denmark and Portugal reported overall increased participation, while Germany and Spain experienced declines (Pita et al., 2021). Impacts may also have varied between freshwater and sea fishing. For instance, in the United Kingdom, participation among established sea anglers and fishing effort increased in summer 2020, but overall sea angling participation, effort and catches declined in 2020 compared to previous years (Hook et al., 2022), while increases in freshwater licence sales suggested increased participation in freshwater fishing (Pita et al., 2021).

Britton et al. (2023) found that overall and individual fishing effort among Fishbrain users were lower in both 2020 and 2021 compared to 2019 in Europe and Oceania[2], despite a previously increasing trend. In contrast, participation in the North America region increased substantially, although these increases diminished partially or entirely in 2021 (Britton et al., 2023).

In terms of fishing tourism, the impacts on domestic tourist numbers during 2020 varied by country. For instance, domestic recreational fishing tourism increased in Australia after national travel restrictions eased (Ryan et al., 2021), while Germany experienced a decline (Britton et al., 2023). Conversely, declines in international fishing tourism were reported worldwide, including in Canada (Britton et al., 2023), Denmark (Britton et al., 2023; Gundelund and Skov, 2021; Pita et al., 2021),

[2]Oceania generally refers to islands of the Pacific Ocean, including Australia, New Zealand, and various island nations. The East Asia & Pacific region comprises Oceania, along with East Asian countries such as China and Japan.

Norway, Greece, Portugal, and Spain (Pita et al., 2021), while border closures suspended international fishing tourism in Australia (Britton et al., 2023). Although there were some seasonal increases during the summer of 2020, coinciding with a seasonal rise in international general tourism in Europe (Marques Santos et al., 2020), overall international tourism levels remained below those of 2018 and 2019, likely due, at least in part, to ongoing international travel restrictions.

Specifically, our findings show that fishing activity reported by Fishbrain users travelling from and to the Europe & Central Asia and East Asia & Pacific regions decreased from 2019 to 2020, and remained low in 2021, broadly aligning with the trends within these regions identified by Britton et al. (2023). However, in contrast to the rise in fishing activity by Fishbrain users in the North America region from 2019 to 2020, and the subsequent decline in 2021 (Britton et al., 2023), our findings indicate that international trips by Fishbrain users from North America decreased from 2019 to 2020, and subsequently increased in 2021. Trips to North America also declined from 2019 to 2020, then declined further in 2021.

4.4.2 Factors influencing travel and destination choices

Impact of distance on travel

In our analysis, cluster patterns show that for international fishing tourism, strongly connected countries – defined as those frequently in the same community cluster – are often in close geographical proximity. This observation aligns with broader patterns in general international tourism, where more than 50% of trips are to immediate neighbouring countries (McKercher and Mak, 2019; McKercher et al., 2008) and 93% are to destinations within 2,000 kilometres (McKercher and Mak, 2019)[3]. However, the geographical area of strongly connected countries is more extensive for fishing tourism. For example, strongly connected countries in the same geographical area can be situated across an area broader than 1,000 kilometers (for example the Americas) or are not necessarily immediate land neighbours (for example the Nordic and Baltic countries, around the Baltic Sea).

'Distance decay theory', which suggests that demand decreases exponentially with increased distance, has been shown to apply to destination choice for international tourism (McKercher and Mak,

[3]For context, the average distance between European capital cities is approximately 1,000 kilometers. The greatest distance between any two European capital cities is approximately 4,000 kilometres.

2019; McKercher et al., 2008). While distance does not have a direct impact, it is linked to multiple factors determining destination choice (McKercher and Mak, 2019). However, the relationship between some origin-destination country pairs exhibits anomalies, with a higher proportion of travel to a long-distance destination compared to typical distance-decay patterns. For instance, travel from Japan to the United States (McKercher et al., 2008), as well as from Australia and New Zealand to Europe (McKercher and Mak, 2019; McKercher et al., 2008), is greater than typically expected for these distances. Our findings indicate that the United States is also a popular destination from Japan for Fishbrain users, although a substantial proportion of catches reported in the United States by Japanese users is by a relatively small proportion of users. However, in contrast to general tourism, for Australian anglers, the United States is a more popular fishing destination than Europe.

In terms of general tourism, only the United States and Spain are sufficiently appealing destinations to attract tourists from multiple countries further than typical distances, with the United States being unparalleled in attracting tourists from worldwide, while Spain's effect is on European tourists (McKercher et al., 2008). With fishing tourism, the United States is popular, despite its distance, with tourists from Australia, the United Kingdom, France, and South Africa, as well as Japan. The proportion of fishing trips from Australia to the United States (32%) almost equals the proportion to close neighbour New Zealand (36%), while the highest proportion of fishing trips from the United Kingdom is to the United States (37%), considerably higher than to neighbouring France (24%) despite its proximity and a land-link through the Channel Tunnel. Similarly, the highest proportion of fishing trips from France is also to the United States (30%), more than double the proportion to its most popular European destination, despite the travel from France to European destinations being quicker, more cost effective, and often accessible by land. Almost two-thirds of fishing trips from South Africa are to the United States (65%), although this could also be due to an 'Effective Tourism Exclusion Zone (ETEZ)' (McKercher et al., 2008) imposed by the surrounding oceans. It is not known from our analysis whether the United States has a particular fishing appeal or whether trips would have occurred anyway with fishing as a secondary activity rather than the main trip purpose. Since the United States is a more popular destination for fishing tourism from Australia, while Europe is more popular for general tourism (McKercher and Mak, 2019), this suggests that the United States may have a particular appeal for fishing enthusiasts.

Sweden and Australia are more popular destinations for fishing tourists from the United States, compared to other fishing destinations in the same geographic area. Trips from the United States to Australia are almost ten times greater than trips to New Zealand, and four times greater to Sweden than to Norway and Finland, suggesting that Australia and Sweden have a particular appeal to United States anglers over other countries in these geographical areas. Sweden is also preferred over Norway and Finland for anglers from Poland, Germany, and Denmark, with over three times as many trips from each of these countries to Sweden compared to Norway, and a comparatively low number of trips to Finland.

Influences on destination choice

Factors influencing the choice of destination for general tourism trips vary according to the duration of the trip (Carril-Caccia et al., 2024; McKercher et al., 2008). For short duration trips (up to three nights), shorter distances (Carril-Caccia et al., 2024; McKercher et al., 2008), neighbouring countries by land (Carril-Caccia et al., 2024), cost and time factors (McKercher et al., 2008), and a similar culture (such as language and religion) in the destination country to the origin country (Carril-Caccia et al., 2024; McKercher et al., 2008) are preferred. The strong connections in our findings between fishing destination countries within close geographical proximity suggests these factors may also be important for fishing destination choice. However, this is not always the case. From a cultural and geographical perspective, France is a member of the Western European countries, along with Switzerland, Belgium, the Netherlands, and Germany, but only has weak fishing tourism connections with them. In terms of recreational fishing trips, it has stronger connections with Spain, the United Kingdom and Ireland. Affordable air travel between northern and southern Europe increases the distance in these directions for which general tourism demand remains high (McKercher and Mak, 2019), and may also be a factor in the north-south grouping patterns evident in the Europe fishing tourism community clusters. Furthermore, cultural similarities, such as a common language, may extend the influence of fishing tourism over longer distances compared to general tourism. Despite large geographical distances, Australia, New Zealand, and South Africa are strongly connected with the United Kingdom and Ireland for fishing tourism, and cultural similarities and a shared language

may contribute to this connection; however, none of these five countries exhibit equally strong connections with the United States, despite also having a similar culture and common language.

For longer duration general tourism trips (four nights or more), destination choice is impacted by distance (Carril-Caccia et al., 2024) and destination attributes (McKercher et al., 2008). The popularity of the United States, Australia, and Sweden as long distance fishing tourism destinations indicates that destination appeal may also be highly influential in destination choice for fishing trips. Carril-Caccia et al. (2024) also found a positive correlation between long duration trips and the proportion of individuals from other countries residing in the origin country, which may have been particularly relevant during pandemic restrictions.

Species availability also influences fishing destination choices. Research has shown that the distance anglers travel to recreational fishing locations depends on their target species (Camp et al., 2018), highlighting the importance of understanding species preferences when analysing fishing tourism trends. The United States is a popular 'trophy fishing' destination (Boon et al., 2024), attracting anglers seeking to catch large and noteworthy species. This may contribute to its appeal as a destination for Fishbrain users. Interestingly, other renowned trophy fishing destinations such as Japan and New Zealand (Boon et al., 2024) do not emerge as popular international destinations for Fishbrain users in our findings. However, it's worth noting that the global number of trophy fishing participants remains unknown (Boon et al., 2024). If the proportion of Fishbrain users participating in trophy fishing is relatively low, analysis at a species level may be required to identify any international tourism trends in trophy fishing.

Largemouth black bass, the most-recorded species on the Fishbrain app by international tourists, is a sought-after trophy species (Dotson et al., 2013). Recent studies indicate a reduction over time in the size of trophy fish caught (Bellquist and Semmens, 2016; Boon et al., 2024) and increased pressure on trophy species in previously less-exploited areas (Boon et al., 2024). Trophy fishing typically involves substantial financial expenditure (Boon et al., 2024), and understanding trophy fishing tourism can inform both tourism strategies and conservation efforts, balancing economic benefits with environmental protection.

Network graphs showing a preliminary analysis for the top three species recorded in Fishbrain, Largemouth black bass, Northern pike, and Smallmouth bass are in Appendix C. Interestingly, the

number of connections and trips for these individual species remain more consistent over the four-year period than for all fishing tourism trips, particularly for Largemouth black bass and Northern pike. This highlights the value of species-specific analyses in uncovering nuanced trends within international fishing tourism.

Key connections

Key general tourism flows worldwide are between China and Hong Kong, and between the United States, Mexico and Canada (Llano et al., 2023a). Our results show that the high general tourism flows between the United States, Mexico, and Canada are also seen in fishing tourism. Hong Kong does not feature in the world country networks due to low catch numbers, but 76% of the 138 trips from Hong Kong are to China. However, there are no catches recorded in Hong Kong by users from China. Therefore, the popularity of Hong Kong for general tourism from China does not appear to extend to fishing tourism.

Key general tourism flows within Europe are from Germany to Poland, France, and Italy, from the United Kingdom to Spain, and from France to Spain and Italy (Llano et al., 2023a). Of these, the strongest fishing connections occur between Germany and Italy, and for Spain with the United Kingdom and France, although our results do not indicate the direction of the connection. Despite strong general tourism connections, fishing tourism connections for Germany with Poland and France, and between France and Italy, are very weak. Furthermore, our findings show a strong fishing tourism connection for Spain with Greece and Ireland. In contrast, Spain and Greece are not strongly connected in general tourism. However, while general tourism from Spain to Ireland is weak, Spain is a key general tourism destination from Ireland (interactive data and maps by Llano et al. (2023a), available at: Llano et al., 2023b).

Our findings also show strong fishing tourism connections between the Nordic countries of Denmark, Finland, Norway, and Sweden. In contrast, in general tourism, the highest numbers of tourists are primarily among Denmark, Norway, and Sweden (interactive data and maps by Llano et al. (2023a), available at: Llano et al., 2023b). Despite Finland's strong fishing tourism connections with other Nordic countries in our community analysis, it has substantially fewer fishing tourists from

the countries in this group compared to the other Nordic countries. This aligns with the pattern of general tourism from other Nordic countries to Finland.

Key origin and destination countries

Origin countries Recreational fishing participation is influenced by culture and socio-economic factors within countries (Arlinghaus et al., 2021). Countries with high recreational fishing participation rates by their citizens include the United States and Canada in the North America region; Ecuador, Colombia, Venezuela, Brazil, and Argentina in the Latin America & Caribbean region; the Nordic (Denmark, Finland, Iceland, Norway, Sweden) and Baltic (Estonia, Latvia, Lithuania) countries in Europe & Central Asia; and Australia, New Zealand, and Japan in the East Asia & Pacific region, although data is not available from many countries across the world (Arlinghaus et al., 2021). Although participation rates are lower in France, Germany, the Netherlands, Switzerland, and the United Kingdom compared to the Nordic and Baltic countries (Arlinghaus et al., 2021), comparatively higher population densities (Arlinghaus et al., 2021) result in a substantial number of individuals participating in recreational fishing in these countries. Although the data used in this research covers only a small proportion of countries worldwide, these countries tend to correspond to those with higher participation in recreational fishing. However, there is limited recreational fishing data available for Africa, substantial areas of Central and South America, and Asia (Arlinghaus et al., 2021).

Destination countries Many countries across the world have developed their fishing tourism opportunities, including Canada, Cuba, Chile, France, Spain, Slovenia, Russia, the United Kingdom, and United States (Mordue, 2009). Notable fishing tourism destinations include Argentina, Chile, Mexico, Belize, Botswana, Mongolia, Gabon, and Slovenia (Hannonen and Hoogendoorn, 2022). Of these, the countries in the North America and Latin America & Caribbean regions feature strongly as destinations among Fishbrain travellers, along with France, Spain, and the United Kingdom from Europe & Central Asia. Botswana, Mongolia, and Gabon did not feature in any networks, while Russia and Slovenia were in only a few. Many top destination countries for Fishbrain users are also top origin countries (Table 4.21), which may suggest that international recreational fishing tourism may be more popular with people who reside in countries where recreational fishing is popular.

Impacts of the COVID-19 pandemic

The COVID-19 pandemic significantly disrupted global travel (Gössling et al., 2021) and had a major impact on tourism (Gössling et al., 2021; Liu et al., 2023). The most notable decreases in international tourists and trip durations occurred from April to June 2020 and in December 2020, when many travel restrictions were in place (Roman et al., 2022). However, the decline in international recreational fishing was less severe than the overall decline in international travel and tourism (see Section 4.4.1).

Travel restrictions, measured on a scale from 0 (no restrictions) to 4 (total border closure) (Hale et al., 2023), varied significantly by country and over time. For example, restrictions in the United Kingdom were implemented more slowly compared to the United States and Europe (Figure 4.30). This may have contributed to the increase in trips to the United Kingdom from North America early in the pandemic (Figure 4.20). However, even as restrictions intensified, trips to the United Kingdom continued to rise in 2021. In addition, trips to Sweden (Figure 4.23b) remained consistent from 2019 to 2020 despite high travel controls (Hale et al., 2021).

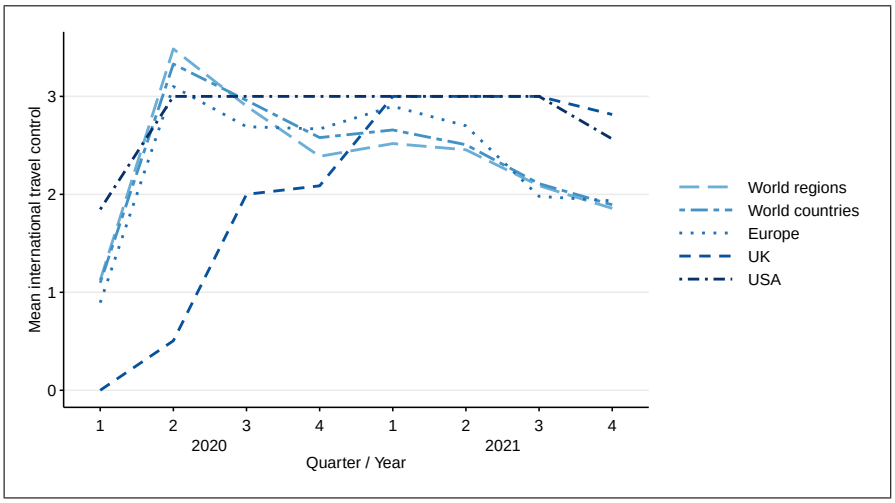


Fig. 4.30 Mean international travel controls. These are calculated as the average of the restriction level each day in the corresponding quarter across the destination countries in the networks of the number of trips for the case. For example, the mean international travel control for Europe is based on only the countries in the Europe networks of trips, while for the United Kingdom and United States case studies, the level of restrictions to the case study country only is used to calculate the mean. Data source (Hale et al., 2021).

Several factors could explain why international recreational fishing did not decline as sharply as other forms of tourism despite the international travel controls. One possibility is that individuals

residing in international second homes, or those already living abroad, started or continued to engage in fishing activities. Second homes, which are often located in leisure areas (Hall, 2014), became temporary residences for many during the initial stages of the pandemic (Alonsopérez et al., 2022; Järv et al., 2021; Willberg et al., 2021). This facilitated extended stays and may have increased the number of fishing trips recorded, even as overall international travel decreased, especially where individuals anticipated the impending travel restrictions and relocated prior to the introduction of restrictions.

Furthermore, recreational fishing effort increased by around 20% during the initial COVID-19 lockdown periods across several European countries (Audzijonyte et al., 2023), with elevated levels persisting into late 2021. It is not reported what proportions of these increases were driven by local residents, domestic tourists, or international tourists. However, in Denmark, the proportion of new international users on the Fangstjournalen angling platform was approximately 25-33% of pre-pandemic levels (Gundelund and Skov, 2021), suggesting some ongoing participation despite the pandemic.

Perceptions of travel risk varied during the pandemic, influenced by factors including age and nationality (Köchling et al., 2022; Terzić et al., 2022). Individuals comfortable travelling during a crisis tend to enjoy connecting with nature (Calderón et al., 2022; Marques Santos et al., 2020) and participating in tourism activities that are active, outdoor, or enhance well-being (Hajibaba et al., 2015). In Europe, an increase in tourist participation in outdoor activities was evident during summer 2020 (Marques Santos et al., 2020). Recreational fishing aligns well with these preferences (Hook et al., 2022), potentially contributing to a higher proportion of tourists engaging in this activity compared to pre-pandemic levels.

Perceptions of travel risk were lower among younger individuals (Hajibaba et al., 2015; Köchling et al., 2022; Terzić et al., 2022). In addition, a temporary increase in younger anglers was seen in 2020 across multiple countries (Britton et al., 2023). This suggests that there may have been an increase in younger individuals participating in recreational fishing during international travel in this period.

It is important to note that catches recorded on platforms like Fishbrain depend on anglers' willingness to log their catches. This reporting behaviour can vary based on angler characteristics, which may differ between established anglers and those who started fishing during the pandemic

(Britton et al., 2023; Gundelund and Skov, 2021). Additionally, it is possible that existing Fishbrain users may have become more inclined to record their catches during the pandemic compared to the pre-pandemic period.

Increases in reported catches could be due to increases in catch rates. For example, reduced commercial fishing activity resulted in higher fish stocks (Pita et al., 2021), which may have led to higher recreational fishing catch rates. Increases in reported catches could also be due to increased effort or increases in reporting rates.

In summary, while international tourism saw a significant decline during the pandemic, international recreational fishing was less affected. This resilience may be due to factors such as an increased interest in fishing, shifts in residency patterns, and the specific characteristics of travellers who continued to engage in recreational fishing.

4.4.3 Comparing the United Kingdom and United States case studies

The impact of the pandemic is highly evident on travel from the United Kingdom, while in contrast travel from the United States shows less impact. This is possibly due to different levels of restrictions in the United Kingdom (Hook et al., 2022) compared to the United States, or across different states in the United States (Paradis et al., 2020).

In the United States, the restrictions on recreational fishing varied across different states, but overall there was a positive attitude from authorities towards recreational fishing (Paradis et al., 2020). In the early weeks of the pandemic, 92% of states allowed recreational fishing (although some restrictions to protect public health were implemented, such as banning competitions), with some states actively promoting recreational fishing, opening fisheries earlier than usual in the season, or removing the requirement for fishing licences during this period (Paradis et al., 2020). As a result, recreational fishing in the United States increased during the initial months of the pandemic (Midway et al., 2021). In contrast, in the United Kingdom, recreational fishing was prohibited for several weeks between March and May 2020 (Environment Agency, 2020; Hook et al., 2022) although Hook et al. (2022, p. 5) report that in April 2020, “57% of individuals who would have typically have gone sea angling did not do so”. It is possible that the differences in the responses by authorities within these countries may have influenced both international travel and leisure activity choices during trips,

contributing to a smaller decline in international recreational fishing among Fishbrain users from the United States compared to those from the United Kingdom.

The network graphs for the United States case study have substantially more connections compared to those of the United Kingdom, despite a higher minimum threshold for the number of trips between country pairs applied to the United States networks. Furthermore, the same pattern is evident when no minimum threshold is applied (see Appendix C). The higher number of connections in the United States networks may be due to the higher popularity of the Fishbrain app in the United States, resulting in a larger user base compared to the United Kingdom.

4.4.4 Data and methods

For all cases except the world regions, we limited the networks to represent only the strongest travel connections, to enable the important patterns to be clearly identified, since when no filtering is applied, high numbers of edges representing comparatively small numbers of trips or catches result in the graphs being cluttered and harder to interpret. There are choices in the way the data are filtered. Filtering on vertices, as with the Europe and world countries networks, allows a consistent set of countries across each quarter, while filtering based on edges, as with the individual case study countries, allows differences between the networks to be seen more clearly. Networks based on the number of trips and catches for the United Kingdom and United States with all trips and catches included (Figures C.8 to C.11 in Appendix C) show the same patterns as those filtered with a minimum threshold (in Section 4.3.1). This indicates that the conclusions are robust to the use of the threshold in these cases.

In the data, the metrics are correlated with the number of trips. When no correlation exists, the number of edges increases proportionally with the number of trips. Conversely, if travel patterns are highly fixed, the number of edges remains fairly consistent regardless of the number of trips. In our data, the number of edges (origin-destination pairs) increases as the number of trips increases, indicating that the connectivity between countries (or regions) is related to the total number of trips across the network. Furthermore, as described in the definitions of the metrics, we observe that edge density (max vertices) tends to be correlated with the mean distance and diameter, with an increase in edge density (max vertices) linked to a decrease in both mean distance and diameter.

Due to lower numbers of users prior to 2018 while the Fishbrain app was still gaining popularity and the resulting sparseness of international catches recorded, we have only used four years of data to detect patterns, two years prior to the pandemic and the first two years of the pandemic's impact. Additional analysis shows that, prior to the pandemic, the community patterns in 2017 are similar to those in 2018 and 2019 quarters 1 and 4, with more variation between the community patterns in quarters 2 and 3, especially quarter 3 (see Figure C.12 in Appendix C). This consistency in community patterns across 2017-2019, particularly in quarters 1 and 4, reinforces the validity of our two-year pre-pandemic baseline and strengthens the reliability of our comparative analysis with the pandemic period.

There are many ways of identifying communities within networks, and some options available in the `igraph` (Csardi and Nepusz, 2006) package are considered in Section 4.2.2. The computational complexity and running time of many community detection algorithms increase exponentially as the number of vertices in the network increases (Schaeffer, 2007). Additionally, the efficiency of clustering algorithms for large networks depends to an extent on the relative density of clusters compared to the overall network density (Emmons et al., 2016). Given the small number of vertices in the networks in this analysis, we did not consider computational complexity when selecting a community detection algorithm. To identify communities, we used the greedy optimisation of modularity algorithm by Clauset et al. (2004), which aims to achieve a high number of edges within clusters and a low number of edges between clusters. This algorithm requires undirected graphs and therefore assumes that countries (or regions) which are connected in any direction are similar. However, it is important to note that for some countries (or regions) the number of trips in each direction is substantially different.

Comparing the communities based on weighted networks (given in the Results section for the networks based on the number of trips and in Appendix C for the networks based on the number of catches) with the communities in unweighted networks based on the number of trips and catches, respectively (not included in this thesis), showed that the observed patterns in the community clusters based on weighted networks also exist in the communities based on unweighted networks, but to a slightly lesser extent. When using weighted networks, connections between countries (or regions) become stronger (based on the number of times each country (or region) is in a community cluster

together), indicating that the community cluster patterns are not driven by the network weighting, but that the weighting strengthens the existing connections.

4.4.5 Limitations and biases

Limitations

While the findings of this research provide valuable insights, there are several limitations to consider.

We assumed that each user makes a maximum of one trip to each destination country per quarter. However, multiple catches recorded in a quarter may have resulted from multiple trips. Furthermore, catches from a single trip recorded as occurring in more than one quarter are treated as multiple trips occurring in different quarters. Likewise, a trip (as defined according the UNWTO definition (United Nations, 2010)) to multiple countries is treated in our data as multiple trips to each destination country.

We have also assumed that all catches reported as occurring in a different country to the country in which the device is registered are made by tourists. However, catches may be recorded by individuals employed in the destination country or who are visiting for longer than one year and so are not tourists according to the UNWTO definition (United Nations, 2010). Catches recorded by such individuals may be incorrectly identified as tourism catches if these individuals use a device registered in their home country rather than the country they are visiting.

Furthermore, we assume that the country in which the device is registered is the user's country of residence. In reality, users may use a device registered in a country other than their country of residence. This would result in an incorrect classification of the origin country, and in some cases may even result in a catch being classified as occurring in the country of residence, rather than from international travel.

The data were filtered to include only the dominant travel. Two different filtering methods were used: by vertex size (countries with the highest numbers of outbound and inbound trips) for Europe and world countries, and by edge weight (connections with the highest travel volumes) for the individual country case studies. These filtering decisions were made to ease interpretation of the corresponding results, but they limit opportunities for comparison between different sets of results.

When filtering by vertex size, we considered only travel from the top 15 origin countries to the top 15 destination countries, which means that for some country pairs only travel in one direction is included. For example, in the Europe networks based on the number of trips, Belgium is only in the top 15 origin countries and Greece is only in the top 15 destination countries. Therefore, trips from Belgium to Greece are included, but trips from Greece to Belgium are not included. Similarly, trips from Croatia and Italy to Switzerland, Lithuania and Latvia are not included (while trips in the opposite direction are included). This may impact the community clusters, as some connections are based on two-way flows, while others are based on one-way flows. For the worldwide networks based on the number of trips, trips from Denmark, Germany, Ireland, Poland and South Africa to the Bahamas, China, Costa Rica, Norway and New Zealand are not included. If all countries from the top 15 origin countries and top 15 destination countries were included as origin and destination countries in the analysis, this would have included an additional 383 Europe trips (6%) and 2,007 worldwide trips (4%) during the period 2018 to 2021.

Although community detection algorithms that optimise modularity are effective on networks with similar-sized clusters and vertex degrees, they are less successful at identifying clusters in networks with different sized clusters and a wide range of vertex degrees (Lancichinetti and Fortunato, 2009). This limitation is particularly relevant for real-world networks, which often exhibit such heterogeneous structures (Fortunato and Castellano, 2012; Schaeffer, 2007). Furthermore, the greedy optimisation of modularity algorithm is biased towards the creation of large clusters over small ones (Wakita and Tsurumi, 2007). However, as our networks have a small number of vertices, the difference in size between clusters is likely to be smaller, which may make these limitations less of an issue in these networks than in larger networks. Nevertheless, it remains a consideration that could be worth further investigation in future research.

Additionally, the greedy optimisation of the modularity algorithm employs an agglomerative approach, making it susceptible to issues associated with such methods. These include failing to detect communities, and the propensity to identify only the central vertices within a community while overlooking more marginal ones. This could also be considered in further research.

We rely on the records reported by Fishbrain users, which may contain errors, for example relating to location of catch or species identification. For instance, Smallmouth bass occurs naturally only

in North America, but has been released in some countries for sport fishing (Loppnow et al., 2013). Therefore, the small numbers of Smallmouth bass recorded in countries in other regions may result from misidentification of species, or be genuine catches of Smallmouth bass.

The analysis used only four years of data, covering the two years prior to the pandemic and two years from the beginning of the pandemic. Data covering a longer time period is needed to confidently establish seasonal patterns and trends over time, particularly the recovery from the pandemic.

Although the Fishbrain app has a global user base, the data analysed in this research predominantly consist of catches recorded by users from a limited number of countries, primarily high-income countries and, to a lesser extent, upper-middle-income countries (World Bank, 2018). This geographic concentration could introduce biases, as fishing practices and user demographics may vary significantly across different regions and income levels. Nevertheless, there is substantial overlap between the countries included in this study and those with the highest rates of recreational fishing (Arlinghaus et al., 2021). However, it is important to note that data availability is limited in some parts of the world (Arlinghaus et al., 2021), which may further affect the representativeness of the findings.

The data consist of the catches recorded by users of a single app, which may limit the generalisability of the findings. Furthermore, anglers who use apps may not be representative of the general angling population. Users of the Fangstjournalen angling citizen science platform have been shown to not be representative of the overall angling population, as they are more specialised and have higher catch rates (Gundelund et al., 2020). It is possible that this is also the case for users of the Fishbrain app. Despite this, there were similarities between catches recorded on the Fangstjournalen platform and traditional survey data, although this was limited to sea trout catches over a three-month period in a specific location and may not be the case more generally (Gundelund et al., 2021). Additionally, differences may exist between Fishbrain users and users of the Fangstjournalen platform, which could mean that these findings may not be applicable to Fishbrain users. Nevertheless, this indicates the potential of data recorded by angling app users to provide useful insights into recreational fishing.

Finally, the study focused on catch rates, and it did not consider qualitative factors such as users' motivations and attitudes towards fishing and data reporting. Incorporating these aspects could provide a more comprehensive understanding of biases and their underlying causes.

Potential biases in the data

Recreational fishing data collected through apps enable use of larger datasets, spanning a wider geographical area and longer time periods compared to traditional surveys. Furthermore, apps can collect extensive data and provide a higher spatial resolution than traditional surveys (Jiorle et al., 2016), for example the Fishbrain app collects a wide variety of app-generated data including location coordinates and weather and sea conditions in addition to data recorded by users, such as the fishing method and species caught. However, there are many potential biases in recreational fishing data obtained from apps (Venturelli et al., 2017), with key biases including avidity bias, angler characteristics, and reporting bias (Jiorle et al., 2016; Johnston et al., 2022; Papenfuss et al., 2015), as well as retention bias (Jiorle et al., 2016; Papenfuss et al., 2015). These biases are not necessarily limited to app data and also occur in other methods of collecting recreational fishing data (Jiorle et al., 2016), for example there is the potential for reporting bias in data collected through angler diaries (Jiorle et al., 2016) and avidity bias in surveys with self-selecting participants (Bethlehem, 2010).

Avidity bias and biases from angler demographics and behaviour arise from the self-selection nature of app use (Gundelund et al., 2020), from choosing whether or not to use the app, and once using the app, choices in which, if any, catches are recorded. Avidity bias can lead to a high proportion of the data recorded by a low proportion of users (Papenfuss et al., 2015), while biases in angler demographics and fishing behaviour can arise from a variety of characteristics, such as potentially higher app use by urban users (Johnston et al., 2022). Reporting bias occurs when incorrect data are recorded, either unintentionally or intentionally, for example, recording an incorrect location or species, or understating or overstating catches. In citizen science projects, which similarly rely on user-reported data, volunteers often report fewer common species and record unusual species more frequently (Dickinson et al., 2010). In the case of Fishbrain, opportunities for overstating catches are low, although anglers may choose to only record some catches, or particular species or catches in particular locations ('affinity' bias (Johnston et al., 2022)), which could introduce spatial bias (Johnston et al., 2022) into the data. Although incorrect identification of species can introduce bias (Johnston et al., 2022), the species caught is not relevant for the analysis in this research, and therefore this is not an issue here. Retention bias may be introduced through short term use of the app

(Papenfuss et al., 2015) if short-term and long-term users have systematic differences, and it is not known whether reduced or discontinued reporting of catches by individual users is due to reduced or discontinued fishing activity or app use (Jiorle et al., 2016). In contrast, a benefit of app data is that recall bias is likely to be lower than in retrospective surveys, due to the ability to record catches soon after they occur (Jiorle et al., 2016; Johnston et al., 2022; Venturelli et al., 2017). Despite potential biases, there is increasing evidence that the quality of self-reported data can be similar to traditional surveys in some situations (Johnston et al., 2022), though further research is required to understand whether this applies in other situations (Johnston et al., 2022) and whether app data can provide similar conclusions to more traditional data collection methods at some spatial scales (Jiorle et al., 2016). Additionally, there are many approaches for addressing various biases in app data (Venturelli et al., 2017), including for retention and reporting biases (Venturelli et al., 2017) and avidity bias (Jiorle et al., 2016).

Despite these limitations and potential biases, this research provides an initial insight into global travel patterns for recreational fishing. Further studies could investigate the limitations of the methods used and biases in the data, the impacts of these, and ways to address them.

4.4.6 Future work

In this chapter we presented an initial investigation of seasonal patterns and trends over time in international fishing tourism. Future research could investigate this further, using formal graph comparison measures that consider networks with different numbers of vertices, and using statistical testing to identify significant differences eg (Tantardini et al., 2019; Wills and Meyer, 2020). As an alternative to the multiple static networks used in this research, dynamic network analysis (Kolaczyk and Csárdi, 2020) could be used to investigate changes in international fishing tourism over time. Any future research could also use a longer time span of data to increase the clarity and reliability of the findings.

Identifying clusters or communities in directed graphs is not a straightforward problem (Fortunato, 2010), and there is much less research on clustering in directed graphs than in undirected graphs (Malliaros and Vazirgiannis, 2013). Finding clusters in directed graphs is often done by removing the edge directions to obtain an undirected graph and identifying clusters in this, which is the method

used in this research. However, this method does not accurately consider the asymmetric structure of a directed graph (Malliaros and Vazirgiannis, 2013) and therefore disregards potentially valuable information. Future work could investigate finding communities using the directed networks instead of converting them to undirected networks.

Additionally, future work could explore alternative optimisation functions, for example the Constant Potts Model, which addresses some of the issues with modularity (Traag et al., 2019). Future work could also explore alternative community detection approaches, for example spectral techniques (Fortunato, 2010; Newman, 2013; Schaeffer, 2007), statistical inference methods (Fortunato, 2010), or density-based methods such as the ST-DBSCAN algorithm for spatiotemporal data (Birant and Kut, 2007).

We investigated patterns in international fishing tourism using simple measures describing an overview of the networks. However, relationships may exist at finer scales, and future research could investigate whether nested travel patterns exist within clusters (Fortunato, 2010). Future studies could also explore the possibility of an overlapping community structure, where vertices may belong to multiple communities (Fortunato, 2010). Additionally, further work could aim to identify whether there are different layers of relationships, for example for different target species, or differentiate travel patterns by Fishbrain users for whom fishing is primary reason for travel and those for whom it is a secondary activity on travel for other or general purposes. Economic and other country indicators could be used to identify the role of vertex attributes in global travel patterns. An alternative approach beyond the straightforward comparisons of simple network metrics used in our study, which could be considered in future research for a more detailed investigation of interactions between countries and factors influencing travel patterns, would be to compare the number of communities in the fishing tourism networks with the number of communities in random graphs of the same order (number of vertices) and size (number of edges), and with graphs of the same degree sequence (vertices with the same number of incident edges) (Kolaczyk and Csárdi, 2020).

We filtered the data to include only the main travel in each case. Future work could explore using the same subset of countries for both the origin and destination. Future work could also include all interactions, which would enable the number of connections for each country to be identified and compared. There are fewer methods for investigating interactions (edges) in networks compared to

entities (vertices), and to further investigate travel between countries the network could be converted to a line graph by switching edges and vertices, to enable vertex comparison methods to investigate the travel between countries (or regions) to be used (Kolaczyk and Csárdi, 2020).

Fishbrain users tend to record certain species more than other species (Table 4.4) and future research could focus on international fishing tourism patterns and distances travelled for specific target species, as research has shown that distances travelled within local regions for recreational fishing vary by species of interest (Camp et al., 2018).

Future work could also explore the impact of restrictions in specific modes of transport, such as air travel, on international fishing tourism. International tourist volumes are highly correlated with air and train passenger volumes (Llano et al., 2023a). The European air transport network (EATN) experienced a considerable decrease in 2020 quarter 2, with the number of daily flights initially up to 90% below typical levels, and around 50% of typical levels during the second half of the period, with the average journey time tripling (Mueller, 2022). In comparison, the number of international trips within Europe by Fishbrain users decreased overall by 53% in this quarter, compared to the same quarter in the previous year. Travel volumes within community clusters in the air travel network, particularly the Nordic airport cluster, reduced less than travel between clusters (Mueller, 2022). Conversely, our findings show the reverse pattern occurred for the Nordic cluster in fishing tourism, with around a 60% decrease in trips within the Nordic cluster and a 40% reduction in trips between the Nordic countries and other clusters. Further research could investigate the effects of different types of travel disruptions and changes in passenger volumes on international fishing tourism.

International recreational fishing tourism patterns and trends are likely to continue changing over time, as tourism spatiotemporal patterns shift in response to societal changes and economic conditions (Adamiak and Szyda, 2022; McKercher and Mak, 2019). Additionally, societal changes are occurring across many communities, with impacts on recreational fishing participation (Arlinghaus et al., 2015). The potential impacts of future societal changes on recreational fishing remain uncertain (Arlinghaus et al., 2021). Given the dynamic nature of both tourism and societal transformations, ongoing research will be essential to understand and adapt to the evolving landscape of international recreational fishing tourism, ensuring sustainable practices that can mitigate the impacts on fish populations and allow the continued quality of the fishing experience.

Chapter 5

Discussion

The thesis proposed a number of modelling approaches for data on recreational fishing, ranging from structured survey data to opportunistic data from a mobile phone app, and at different spatial and temporal scales, from regional to global, and from quarterly to annual data. The methods considered include approaches for efficient variable selection, for flexible spatiotemporal modelling, and for uncovering network structures.

All of the analysis considered in the thesis dealt with complex, multidimensional and large datasets, and the corresponding data cleaning and modelling processes were challenging and computationally demanding. The thesis considered appropriate techniques and, where appropriate, reliable approximations to model the corresponding data. The grid-based approach introduced in Chapter 2 was needed to reliably search the parameter space of the two penalty terms of the corresponding model. However, this search was considerably time-demanding, typically requiring fitting the model at least 900 times for each species. Similarly, Chapter 3 considered complex and large spatiotemporal data, and fitting the corresponding models, despite the employed integrated Laplace approximations, was computationally demanding. Finally, Chapter 4 required looping through over 7,000,000 records to assign them to an origin and destination country, which took around 30 days. All of the aforementioned computational challenges were overcome using, where possible, efficient code, parallel computing and high-performance clusters.

5.1 Chapter 2

Chapter 2 demonstrated the use of LASSO penalised regression as a viable option for performing variable selection in complex models with several variables and the results demonstrated the flexibility of the approach and ease with which a large number of predictors can be considered in different parts of the model. The chapter introduced a grid-based approach for selecting the two penalty terms in zero-inflated Poisson models and compared the results to the default approach of the R package `mpath` (Wang, 2020). The results suggested that different coefficients were generated from the two approaches, despite similar maximised log-likelihood values obtained from each method, which is possibly a feature of the sparseness of the data and the complexity of the model. The grid method tended to select covariates in the binomial part, so that the covariates are associated with the probability that fish would be available in angling areas, while the default method tended to select covariates in the Poisson part, so that the covariates are associated with the number of fish caught, given that they were available, giving different interpretations to how catches arise.

We applied our modelling framework to data on eight commonly-caught species: Atlantic cod (*Gadus morhua*), European sea bass (*Dicentrarchus labrax*), Atlantic mackerel (*Scomber scombrus*), European plaice (*Pleuronectes platessa*), European flounder (*Platichthys flesus*), whiting (*Merlangius merlangus*), common dab (*Limanda limanda*), and the skates and rays family (family Rajidae). The covariates in our models related to three main categories:

- Trip characteristics: fishing location (from shore or by boat), fishing method, target species, hours fished during daylight, twilight and darkness, number of fish of other species caught, type of day (weekday/weekend), total anglers in the group, fishing as part of a competition.
- Angler characteristics: number of days shore and boat angling within the previous three months, angling club membership.
- Environment characteristics: sea, season, sea state, lunar illumination, seabed sediment type.

The results of the grid method suggest that for all species there is a high probability of the species being available in angling areas if the species is being targeted, indicating that anglers tend to know which species are likely to be available and target them accordingly. All species are more likely to

be available in boat angling areas than shore angling areas, except flounder which is more likely by the shore, and cod and plaice which are equally likely in both areas. Mackerel are more likely to be available when lure/fly/jig methods are used, while all other species except cod and skates/rays are more likely to be available when bait methods are used. As the number of hours fished during daylight increases, the probability of all species being available increases. Other trip covariates (number of hours fished in twilight or darkness, number of fish of other species caught, weekday/weekend, fishing as part of a competition, number of anglers in a group) have different impacts for different species on the probability the species are available in angling areas (including positive, negative, and no significant impacts). All species have at least one covariate relating to angler characteristics associated with the probability fish of that species are available. The environmental covariates have varying impacts on the probability of each species being available. The probability that cod are available is not impacted by any environmental covariates. For all other species, all or most environmental covariates impact the probability that the species are available in angling areas.

Regarding the number of fish caught, catch rates of cod, sea bass, plaice, and dab are higher when fishing from boats than from shore, if these species are available, with no significant difference for the other species. The fishing method has no significant impact on catch rates for most species, though cod is a notable exception, with higher catch rates of cod associated with lure/fly/jig methods compared to bait. Targeting cod and plaice increases catch rates of these species, while for other species targeting the species has no association with the number of fish of that species caught. Other trip covariates (number of hours fished in daylight, twilight or darkness, number of fish of other species caught, weekday/weekend, number of anglers in a group, fishing as part of a competition) have different impacts on catch rates for different species. Most species have at least one covariate relating to angler characteristics associated with catch rates. Environmental covariates do not impact catch rates of sea bass, flounder, whiting, and skates/rays, while the only environmental covariate impacting mackerel catch rates is lunar illumination. In contrast, most environmental covariates impact catch rates of cod, plaice, and dab, if they are available.

Identifying important predictors for catch rates for different species is important for understanding and predicting catches from recreational anglers, and can inform sustainable angling practices and effective fisheries management strategies. The grid approach enabled us to identify the optimal

penalty values but it is time-consuming, as discussed above. Considering a fine grid was important in this case, as even small changes to the penalty terms led to considerable differences in covariate coefficients, probably due to data sparseness, but for the approach to be more widely adopted more work is required to develop less computationally demanding ways for identifying the most suitable penalty values.

5.2 Chapter 3

Chapter 3 proposed the use of integrated Laplace approximations (INLA) for modelling large spatiotemporal data. The models allowed us to quantify changes in catches across space and time for four important species in the United Kingdom and Ireland. The results suggested that despite using non-probability app data, the overall predicted spatiotemporal patterns of the rate of catches recorded across the seasons around the United Kingdom and Ireland broadly follow the expected spatiotemporal distribution for each species. Spatiotemporal zero-inflated Poisson models were fitted to catches of European sea bass (*Dicentrarchus labrax*), Atlantic cod (*Gadus morhua*), Atlantic mackerel (*Scomber scombrus*), and whiting (*Merlangius merlangus*), comparing two contrasting random effects distributions, iid (independent and identically distributed Gaussian random effects) and ar1 (autoregressive model of order 1), with the results suggesting that the ar1 models, which account for correlation between successive time periods, provide a better fit for all species. The results show a seasonal impact on rates of catches reported for all four species, with a strong temporal correlation between quarters. For each species, the spatial pattern of the predicted rates of catches reported show strong consistency from year to year, with only minimal variations in the rate at any location between years.

The rates of catches reported for sea bass and mackerel are highest in quarter 3 (summer) and lowest in quarter 1 (winter). For cod and whiting, the rates of catches reported are highest in quarter 4 (autumn) and lowest in quarters 2 (spring) and 3 (summer). For sea bass, higher rates are predicted along the south coasts of England and Wales and in south-east Ireland, on the west coasts of Wales, Scotland and Ireland, and a few areas along the east coast of England, with some local-scale variation between fishing sites in these areas. For mackerel, the high predicted rates in quarter 3 (summer) occur around the majority of the United Kingdom and Ireland, although areas on the east coast of

England and north coast of Scotland have lower rates. During quarters 2 (spring) and 4 (autumn), higher predicted rates are generally concentrated at sites in southern England and Scotland. The low rates in quarter 1 (winter) occur around the entire United Kingdom and Ireland coastline. For cod, the high predicted rates in quarter 4 (autumn) are spread along the north-east coast of England and north-west and east coasts of Scotland to the north, and along the Bristol Channel and south-east Ireland to the south, while for whiting, predicted rates are highest around the coast of England and some areas of north-west Scotland.

Monitoring and management of marine recreational fishing is essential for sustainability of fish species and mitigating potential associated impacts on marine environments and ecosystems. The spatiotemporal models enabled us to identify patterns of catches over space and time using Fishbrain app data. These data do not require the resource-intensive collection associated with probabilistic surveys, and have the advantage of being continuous across space and time. However, the reported catches cannot be assumed to be a constant proportion of total catches made spatiotemporally as Fishbrain app users are not a random sample from the population of anglers. Furthermore, these opportunistic presence-only Fishbrain data are an incomplete picture of actual catches. Nevertheless, app data can provide complementary insights into spatiotemporal patterns of catches and changes to spatial distributions, as well as angler choices.

It is important to note that the use of INLA in our analysis required several decisions, such as mesh specifications and prior distributions for parameters related to the spatiotemporal correlation. Since these choices may influence results, further work could explore their effects more thoroughly. In our analysis, we built spatiotemporal models assuming separable space and time effects. Future research could consider alternatives, such as shared component modelling to link the probabilities and abundance in semi-continuous datasets (Paradinas et al., 2017), or the approach by Lindgren et al. (2024), which models spatiotemporal Gaussian Matérn fields using stochastic partial differential equations, allowing for nonseparable spatiotemporal covariance structures that provide greater flexibility and interpretability in modelling spatial and temporal smoothness and range.

A next step could be to investigate combining app data with other, more robust, data sources for a more accurate and higher resolution insight into recreational fishing catch rates. Integrating multiple data sources enables unstructured data (such as presence-only data), which often cover larger areas

or longer timescales, to supplement higher quality datasets, which tend to have fewer observations (Robinson et al., 2018, 2020). A simple and common (Fletcher Jr. et al., 2019, p. 3) approach to integrating data involves pooling data (merging multiple datasets into a single dataset before analysis), which often requires transforming higher-quality datasets to match lower-quality datasets, for example removing absences from presence-absence data to combine it with presence-only data (Isaac et al., 2020). However, pooling assumes similar observation processes across datasets, which is frequently unrealistic and can lead to inaccurate results (Fletcher Jr. et al., 2019). In contrast to data pooling and early approaches that combined model outputs[1], recently-developed integrated models offer powerful frameworks for combining diverse data sources within a single framework without losing the characteristics of the data, enabling more robust inferences (Isaac et al., 2020; Miller et al., 2019; Zipkin and Saunders, 2018). In a fisheries context, Stock Assessment Models (SAMs) combine data, historically through the separate application of multiple datasets or, more recently, using integrated model methods (Maunder and Punt, 2013). Integrated Population Models (IPMs), which integrate multiple datasets using classical or Bayesian approaches to model population dynamics (Schaub and Abadi, 2011; Zipkin and Saunders, 2018), are similar to SAMs, but IPMs have traditionally been applied to land-based ecosystems, while SAMS are used for fisheries (Schaub et al., 2024). However, IPMs have significant potential for use in fish population studies (Schaub et al., 2024). Species Distribution Models (SDMs) traditionally use a single source of species occurrence or abundance data together with environmental variables to model species distributions (Elith and Leathwick, 2009; Karp et al., 2025), but recently, Integrated Distribution Models (IDMs) have been developed to enable the integration of multiple data sources within SDMs, resulting in improved model robustness and greater reliability of ecological inference (Miller et al., 2019; Paradinas et al., 2023). In addition, other data-integrated approaches have been developed, including Integrated Movement Models (IMMs) (Buderman et al., 2025) and models integrating data sources for estimating angler effort (Tucker et al., 2024), reflecting the growing use of integrated approaches in ecology. While these approaches focus on single species, ecological systems are typically composed of multiple interacting species. Integrated Community Models (ICMs) (Zipkin et al., 2023) extend IPM concepts to multi-species

[1] Early approaches to using multiple data sources typically relied on separate models for each source and did not share information between sources (Fletcher Jr. et al., 2019; Maunder and Punt, 2013).

systems. In fisheries, Multi-Species Models (MSMs) extend SAMs to incorporate various levels of ecological complexity, and are increasingly used to support Ecosystem-Based Fisheries Management (EBFM) (Couve et al., 2024; Karp et al., 2023). Although SDMs have been extended to Joint Species Distribution Models (JSDMs) (Poggiato et al., 2021; Roberts et al., 2022; Wilkinson et al., 2021), IDMs for communities (or JSDMs with integrated data) are not yet established. However, Integrated Community Occupancy Models (ICOMs) have recently emerged (Ardiantiono et al., 2025; Doser et al., 2022b), extending Single Species Occupancy Models (a type of SDM when used for spatial prediction (Doser et al., 2022a; Rushing et al., 2019)) to jointly model multiple species using integrated data sources. Using integrated modelling approaches, the presence-only Fishbrain data could be combined with presence-absence or other data to enhance mapping of catch rates and species distributions, in a similar way to modelling distribution and relative abundance more commonly used for terrestrial species (Grattarola et al., 2023; Morera-Pujol et al., 2023; Thorson et al., 2016). Furthermore, ICMs (or other community approaches) could be used to model multiple species simultaneously instead of in separate models as in our research. The Fishbrain data could also enhance the structured survey data used in Chapter 2 to further understand drivers and spatiotemporal patterns of recreational fishing catches. Currently, integrated modelling concepts are an important future research direction in ecological monitoring and an open challenge (Isaac et al., 2020; Johnston et al., 2023; Zipkin et al., 2019).

5.3 Chapter 4

Chapter 4 considered network models for the Fishbrain data, enabling us to understand global recreational fishing tourism patterns and quantify the effect of the pandemic on these. The analysis examined travel patterns to and from the United Kingdom and United States, within Europe, and globally, focusing on trips and catches recorded. In each case, the seasonal patterns and trends over time were explored, in terms of the number of connections and number of trips or catches between pairs of countries. The chapter presented directed graphs, by quarter, for the four-year period from 2018 to 2021, summarising global travel for recreational fishing, and community graphs showing the structure of the communities within these networks. The results highlight seasonal and spatial patterns

in international recreational fishing travel, and the trends over time show a considerable reduction in international trip numbers and catches recorded from 2020 quarter 2 onwards in response to the COVID-19 pandemic, although this decrease is less substantial than reported decreases in international travel and tourism in general, especially in the initial stages of the pandemic. The observed patterns and trends based on trips and catches recorded demonstrate consistent similarities in each case.

The findings demonstrate a preference among anglers for warmer destinations during cooler months, in terms of both the number of connections and number of trips. Globally, the highest number of trips is seen in quarter 3, followed by quarter 2, with the lowest in quarter 1, a pattern which is likely linked to the highest proportions of Fishbrain users being in North America and Europe. The impact of the pandemic on angling tourism was not uniform across regions. While trips originating from Europe and the East Asia & Pacific region decreased substantially, those from the United States showed minimal reduction. Trips to countries in the Americas decreased, while trips to the United Kingdom, Sweden, and Australia increased (mainly due to tourists from the United States), contrary to the general trend. Community clustering analysis revealed strong regional connections in angling tourism, with notable clusters including the Americas, the Baltic and Nordic countries, Western European nations (Switzerland, Belgium, Netherlands, Germany), and the United Kingdom, Ireland, and France. This regional clustering aligns with general tourism patterns, where factors such as proximity and cultural similarities influence destination choices (Carril-Caccia et al., 2024; McKercher et al., 2008), although strongly connected regions in angling tourism are often found over broader geographical areas than general short-duration tourism trips. The study also identified certain countries, such as the United States, Australia, and Sweden, that may have particular appeal for long-distance angling tourists. This suggests that for longer duration trips, destination attributes may outweigh proximity factors in angling tourism, echoing findings in general tourism literature (McKercher et al., 2008). These results contribute to our understanding of recreational fishing as a subset of international tourism, highlighting both similarities and distinctions from general tourism patterns. The findings suggest a complex relationship between geographical, cultural, and destination-specific factors in shaping international angling tourism flows, while also demonstrating the sector's vulnerability to global disruptions such as the COVID-19 pandemic.

Recreational fishing is a substantial component of the global tourism economy (Hall, 2021; Mordue, 2009), and understanding and predicting anglers' travel patterns can be valuable for effective fisheries and ecosystem management (Camp et al., 2018), including sustainability and mitigating the spread of invasive species or disease, and from an economic perspective (Camp et al., 2018), to help balance conflicting needs of fisheries, the environment and local communities. Network models and community detection algorithms offered the opportunity to investigate both seasonal patterns and the impact of the pandemic on global recreational fishing tourism. This initial investigation used simple measures and a subset of countries to focus on the key travel, and identified communities using undirected graphs, therefore disregarding potentially valuable information. Future research could explore more detailed or specific analyses, including travel for various target species or travel motivations, and incorporate economic and other country indicators, while identifying communities using directed networks could reveal relationships not evident in communities based on undirected networks. Expanding the study to more countries and including more recent data would enhance comprehensiveness, and employing formal graph comparison measures and statistical testing would strengthen the robustness and validity of findings.

5.4 Summary of modelling approaches and future work

Penalised regression methods, such as those applied to the Cefas structured survey data in Chapter 2, provide a way to identify covariates that significantly influence catch rates from a large number of potential covariates, some of which may be correlated. The INLA approach provides a flexible and computationally efficient way for modelling large spatiotemporal data, such as the Fishbrain data considered in Chapter 3. The models can be readily extended to account for other appropriate spatial or temporal covariates to improve our understanding of reported catches. However, these data only contain reported catches, and hence results should be interpreted appropriately. Future work could consider joint modelling of these presence-only data with structured survey data, by combining the zero-inflated penalised regression modelling work of Chapter 2 and the spatiotemporal models in Chapter 3, to distinguish actual catch rates from reporting patterns.

Network modelling offers unique insights into angling tourism and opens up avenues for future research that can consider economic, demographic, social and other country indicators, such as those produced by the World Bank or United Nations, to provide better understanding of the barriers and opportunities for supporting this important socio-economic activity and managing its impacts. These indicators could encompass a wide range of factors relevant to recreational fishing and tourism. Economic indicators may include gross domestic product (GDP), unemployment rates, and the specific GDP contribution from recreational fishing. Demographic indicators could encompass population size and age distribution, while social indicators might feature education levels and poverty rates. Additional relevant metrics could incorporate employment generated by recreational fishing, fishing activity indicators, tourism statistics, regulatory factors (such as permit requirements), conservation measures, immigration and mobility data (including visa regulations and international travel statistics), and even internet search volumes for fishing-related terms as a proxy for interest in recreational fishing across different countries. A natural extension of the work is to consider more recent data and to model all years jointly within a network framework, accounting for the effect of time within a single model. Combining the local spatial information from Chapter 3 with the global travel data from Chapter 4 has the potential to provide a more detailed picture of anglers' local and international travel patterns and trends as a whole. This could reveal insights, such as whether anglers who fish extensively locally are less likely to take international fishing trips, or if high levels of local fishing within an angling community are associated with regular international travel.

The thesis developed and applied statistical models to understand drivers and patterns of catch rates and fishing activity, regionally and globally. Considering both traditional surveys and app-based data collection provided opportunities for studying recreational fishing, offering valuable insights into both local and international fishing activities. The scale and reach of app-based data will continue to grow, making it essential to develop integrated methods that consider data from multiple sources to overcome limitations of separate data sources. Future research should focus on further developing these methods to support sustainable fishing practices and the conservation of aquatic ecosystems, ensuring that the social, environmental, and economic benefits of recreational fishing can be maximised.

References

Czesław Adamiak and Barbara Szyda. Combining conventional statistics and big data to map global tourism destinations before Covid-19. Journal of Travel Research, 61(8):1848–1871, 2022. doi: https://doi.org/10.1177/00472875211051418.

Antoine Adde, Clara Casabona i Amat, Marc J. Mazerolle, Marcel Darveau, Steven G. Cumming, and Robert B. O'Hara. Integrated modeling of waterfowl distribution in western Canada using aerial survey and citizen science (eBird) data. Ecosphere, 12(10), 2021. doi: https://doi.org/10.1002/ecs2.3790.

Rein Ahas, Anto Aasa, Ülar Mark, Taavi Pae, and Ain Kull. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. Tourism Management, 28(3):898–910, 2007. doi: https://doi.org/10.1016/j.tourman.2006.05.010.

Maria José Alonsopérez, Juan Gabriel Brida, and Mara Leticia Rojas. Second homes: A bibliometric analysis and systematic literature review. Journal of Tourism, Heritage & Services Marketing (JTHSM), 8(1):16–26, 2022. doi: https://doi.org/10.5281/zenodo.6581498.

Malin Andersson. From pluri-activity to entrepreneurship: Swedish inshore commercial fisheries navigating in the service-oriented economy. Scandinavian Journal of Hospitality and Tourism, 21 (4):374–390, 2021. doi: https://doi.org/10.1080/15022250.2021.1906744.

Ardiantiono, Nicolas J. Deere, David J. I. Seaman, U. Mamat Rahmat, Eka Ramadiyanta, Muhammad I. Lubis, Ahtu Trihangga, Ahmad Yasin, Gunawan Alza, Dessy P. Sari, Muhammad Daud, Ridha Abdullah, Rina Mutia, Dewi Melvern, Tarmizi, Jatna Supriatna, and Matthew J. Struebig. Improved cost-effectiveness of species monitoring programs through data integration. Current Biology, 35 (2):391–397, 2025. doi: https://doi.org/10.1016/j.cub.2024.11.051.

Vincent Arel-Bundock, Nils Enevoldsen, and CJ Yetman. countrycode: An R package to convert country names and country codes. Journal of Open Source Software, 3(28):848, 2018. doi: https://doi.org/10.21105/joss.00848.

Robert Arlinghaus and Steven J. Cooke. Recreational fisheries: socioeconomic importance, conservation issues and management challenges. In B. Dickson, J. Hutton, and W. M. Adams, editors, Recreational hunting, conservation and rural livelihoods: science and practice. Wiley-Blackwell, Chichester, 2009. doi: https://doi.org/10.1002/9781444303179.ch3.

Robert Arlinghaus, Thomas Mehner, and Ian G. Cowx. Reconciling traditional inland fisheries management and sustainability in industrialized countries, with emphasis on Europe. Fish and Fisheries, 3(4):261–316, 2002. doi: https://doi.org/10.1046/j.1467-2979.2002.00102.x.

Robert Arlinghaus, Robert Tillner, and M. Bork. Explaining participation rates in recreational fishing across industrialised countries. Fisheries Management and Ecology, 22(1):45–55, 2015. doi: https://doi.org/10.1111/fme.12075.

Robert Arlinghaus, Øystein Aas, Josep Alós, Ivan Arismendi, Shannon Bower, Steven Carle, Tomasz Czarkowski, Kátia M. F. Freire, John Hu, Len M. Hunt, Roman Lyach, Andrzej Kapusta, Pekka Salmi, Alexander Schwab, Jun ichi Tsuboi, Marek Trella, Daryl McPhee, Warren Potts, Arkadiusz Wołos, and Zi-Jiang Yang. Global Participation in and Public Attitudes Toward Recreational Fishing: International Perspectives and Developments. Reviews in Fisheries Science & Aquaculture, 29(1): 58–95, 2021. doi: https://doi.org/10.1080/23308249.2020.1782340.

Mike Armstrong, Adam Brown, Jodie Hargreaves, Kieran Hyder, Sarah Pilgrim-Morrison, Max Munday, Steven Proctor, Annette Roberts, and Kevin Williamson. Sea Angling 2012 – a survey of recreational sea angling activity and economic value in England. Department for Environment, Food and Rural Affairs. London, UK, 2013.

Mike Armstrong, Adam Brown, Jodie Hargreaves, Kieran Hyder, Sarah Pilgrim-Morrison, Max Munday, Stephen Proctor, Annette Roberts, and Kevin Willamson. Sea Angling 2012 - Shore and private boat survey. V1. Cefas, UK, 2016. Available at: https://doi.org/10.14466/CefasDataHub.27 (Accessed: 20 November 2019).

Asta Audzijonyte, Fernando Mateos-González, Justas Dainys, Casper Gundelund, Christian Skov, J. Tyrell DeWeber, Paul Venturelli, Vincentas Vienožinskis, and Carl Smith. High-resolution app data reveal sustained increases in recreational fishing effort in Europe during and after COVID-19 lockdowns. Royal Society Open Science, 10, 2023. doi: https://doi.org/10.1098/rsos.230408.

Fabian E. Bachl, Finn Lindgren, David L. Borchers, and Janine B. Illian. inlabru: an R package for Bayesian spatial modelling from ecological survey data. Methods in Ecology and Evolution, 10: 760–766, 2019. doi: doi:10.1111/2041-210X.13168.

Lyall Bellquist and Brice X. Semmens. Temporal and spatial dynamics of 'trophy'-sized demersal fishes off the california (USA) coast, 1966 to 2013. Marine Ecology Progress Series, 547:1–18, 2016. doi: https://doi.org/10.3354/meps11667.

Jelke Bethlehem. Selection Bias in Web Surveys. International Statistical Review, 78(2):161–188, 2010. doi: https://doi.org/10.1111/j.1751-5823.2010.00112.x.

Derya Birant and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering, 60(1):208–221, 2007. doi: https://doi.org/10.1016/j.datak.2006.01.013.

Marta Blangiardo and Michela Cameletti. Spatial and spatio-temporal Bayesian models with R-INLA. John Wiley & Sons, 2015.

Marta Blangiardo, Michela Cameletti, Gianluca Baio, and Håvard Rue. Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology, 7:39–55, 2013. doi: https://doi.org/10.1016/j.sste.2013.07.003.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008 (10):P10008, 2008. doi: https://doi.org/10.48550/arXiv.0803.0476.

James S. Boon, Grace Vaudin, Hannah Millward-Hopkins, Bethan C. O'Leary, Colin J. McClean, and Bryce D. Stewart. Shifts in the size and distribution of marine trophy fishing world records. Aquatic Conservation: Marine and Freshwater Ecosystems, 34(1):e4051, 2024. doi: https://doi.org/10.1002/aqc.4051.

J. Michael Brick, William R. Andrews, and John Foster. A Review of Nonprobability Sampling Using Mobile Apps for Fishing Effort and Catch Surveys. Transactions of the American Fisheries Society, 151(1):42–49, 2022. doi: https://doi.org/10.1002/tafs.10342.

J. Robert Britton, Adrian C. Pinder, Josep Alós, Robert Arlinghaus, Andy J. Danylchuk, Wendy Edwards, Kátia M. F. Freire, Casper Gundelund, Kieran Hyder, Ivan Jarić, Robert Lennox, Wolf-Christian Lewin, Abigail J. Lynch, Stephen R. Midway, Warren M. Potts, Karina L. Ryan, Christian Skov, Harry V. Strehlow, Sean R. Tracey, Jun ichi Tsuboi, Paul A. Venturelli, Jessica L. Weir, and Marc Simon Weltersbach & Steven J. Cooke. Global responses to the COVID-19 pandemic by recreational anglers: considerations for developing more resilient and sustainable fisheries. Reviews in Fish Biology and Fisheries, 33(4):1095–1111, 2023. doi: https://doi.org/10.1007/s11160-023-09784-5.

Lluís Brotons, Wilfried Thuiller, Miguel B. Araújo, and Alexandre H. Hirzel. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 27(4): 437–448, 2004. doi: https://doi.org/10.1111/j.0906-7590.2004.03764.x.

Frances E. Buderman, Ephraim M. Hanks, Viviana Ruiz-Gutierrez, Michael Shull, Robert K. Murphy, and David A. W. Miller. Integrated movement models for individual tracking and species distribution data. Methods in Ecology and Evolution, 16(2):345–361, 2025. doi: https://doi.org/10.1111/2041-210X.14482.

Kenneth P. Burnham and David R. Anderson. Practical use of the information-theoretic approach. Springer, 1998.

C. Calculli, A. Pollice, I. Paradinas, L. Sion, and P. Maiorano. An INLA spatio-temporal model for zero-inflated marine plastic litter abundance. GRASPA WORKING PAPERS, pages 62–65, 2019.

Michael Moya Calderón, Kevin Chavarría Esquivel, María Margarita Arrieta García, and Carlos Barriocanal Lozano. Tourist behaviour and dynamics of domestic tourism in times of COVID-19. Current Issues in Tourism, 25(14):2207–2211, 2022. doi: https://doi.org/10.1080/13683500.2021.1947993.

Edward V. Camp, Robert N.M. Ahrens, Chelsey Crandall, and Kai Lorenzen. Angler travel distances: Implications for spatial approaches to marine recreational fisheries governance. Marine Policy, 87: 263–274, 2018. doi: https://doi.org/10.1016/j.marpol.2017.10.003.

Federico Carril-Caccia, José María Martín Martín, and Francisco Javier Sáez-Fernández. How important are borders for tourism? The case of Europe. Tourism Economics, 30(1):27–43, 2024. doi: https://doi.org/10.1177/13548166221132452.

Andrés M. Cisneros-Montemayor and U. Rashid Sumaila. A global estimate of benefits from ecosystem-based marine recreation: potential impacts and implications for management. Journal of Bioeconomics, 12:245–268, 2010. doi: https://doi.org/10.1007/s10818-010-9092-7.

Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. Physical Review E, 70(6):066111, 2004. doi: https://doi.org/10.1103/PhysRevE.70.066111.

Steven J. Cooke and Ian G. Cowx. The Role of Recreational Fishing in Global Fish Crises. BioScience, 54(9):857–859, 2004. doi: https://doi.org/10.1641/0006-3568(2004)054[0857:TRORFI]2.0.CO;2.

Steven J. Cooke and Ian G. Cowx. Contrasting recreational and commercial fishing: Searching for common issues to promote unified conservation of fisheries resources and aquatic environments. Biological Conservation, 128(1):93–108, 2006. doi: https://doi.org/10.1016/j.biocon.2005.09.019.

Steven J. Cooke, Cory D. Suski, Kenneth G. Ostrand, David H. Wahl, and David P. Philipp. Physiological and Behavioral Consequences of Long-term Artificial Selection for Vulnerability to Recreational Angling in a Teleost Fish. Physiological and Biochemical Zoology, 80(5), 2007. doi: https://doi.org/10.1086/520618.

R. Cooper, S. Green, and D. Long. User Guide for the British Geological Survey DiGSBS250K Dataset. British Geological Survey Internal Report IR/11/026, 2010.

Pablo Couve, Nixon Bahamon, Cristian M. Canales, and Joan B. Company. Systematic Review of Multi-Species Models in Fisheries: Key Features and Current Trends. Fishes, 9(10):372, 2024. doi: https://doi.org/10.3390/fishes9100372.

Stefany Coxe, Stephen G. West, and Leona S. Aiken. Generalized linear models. In The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis, chapter 3. Oxford University Press, 2013. ISBN 9780199934898. doi: https://doi.org/10.1093/oxfordhb/9780199934898.013.0003.

Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. Complex Systems, 1695(5):1–9, 2006.

John Curtis, Stephen Hynes, Paul O'Reilly, and Benjamin Breen. Recreational angling tournaments: participants' expenditures. Journal of Sport & Tourism, 21(3):201–221, 2017. doi: https://doi.org/10.1080/14775085.2017.1322998.

Justas Dainys, Harry Gorfine, Fernando Mateos-González, Christian Skov, Robertas Urbanavičius, and Asta Audzijonyte. Angling counts: Harnessing the power of technological advances for recreational fishing surveys. Fisheries Research, 254, 2022a. doi: https://doi.org/10.1016/j.fishres.2022.106410.

Justas Dainys, Eglė Jakubavičiūtė, Harry Gorfine, Mindaugas Kirka, Alina Raklevičiūtė, Augustas Morkvėnas, Žilvinas Pūtys, Linas Ložys, and Asta Audzijonyte. Impacts of Recreational Angling on Fish Population Recovery after a Commercial Fishing Ban. Fishes, 7(5):232, 2022b. doi: https://doi.org/10.3390/fishes7050232.

Arnak S. Dalalyan, Mohamed Hebiri, and Johannes Lederer. On the prediction performance of the Lasso. Bernoulli, 23(1):552–581, 2017. doi: https://doi.org/10.3150/15-BEJ756.

Lea I. Dambly, Nick J. B. Isaac, Kate E. Jones, Katherine L. Boughey, and Robert B. O'Hara. Integrated species distribution models fitted in INLA are sensitive to mesh parameterisation. Ecography, 2023(7), 2023. doi: https://doi.org/10.1111/ecog.06391.

A. J. S. Davis and J. A. Darling. Recreational freshwater fishing drives non-native aquatic species richness patterns at a continental scale. Diversity and Distributions, 23(6):692–702, 2017. doi: https://doi.org/10.1111/ddi.12557.

Bruce A. Desmarais and Jeffrey J. Harden. An unbiased model comparison test using cross-validation. Quality & Quantity, 48:2155–2173, 2014.

Janis L. Dickinson, Benjamin Zuckerberg, and David N. Bonter. Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annual Review of Ecology, Evolution, and Systematics, 41(1):149–172, 2010.

Robert B. Ditton, Stephen M. Holland, and David K. Anderson. Recreational Fishing as Tourism. Fisheries, 27(3):17–24, 2002. doi: https://doi.org/10.1577/1548-8446(2002)027⟨0017:RFAT⟩2.0.CO;2.

Carsten F. Dormann, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, Bernd Gruber, Bruno Lafourcade, Pedro J. Leitão, Tamara Münkemüller, Colin McClean, Patrick E. Osborne, Björn Reineking, Boris Schröder, Andrew K. Skidmore, Damaris Zurell, and Sven Lautenbach. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1):27–46, 2012. doi: https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Jeffrey W. Doser, Andrew O. Finley, Marc Kéry, and Elise F. Zipkin. spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. Methods in Ecology and Evolution, 13(8):1670–1678, 2022a. doi: https://doi.org/10.1111/2041-210X.13897.

Jeffrey W. Doser, Wendy Leuenberger, T. Scott Sillett, Michael T. Hallworth, and Elise F. Zipkin. Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. Methods in Ecology and Evolution, 13(4):919–932, 2022b. doi: https://doi.org/10.1111/2041-210X.13811.

Jason R. Dotson, Micheal S. Allen, Janice A. Kerns, and William F. Pouder. Utility of restrictive harvest regulations for trophy Largemouth Bass management. North American Journal of Fisheries Management, 33(3):499–507, 2013. doi: https://doi.org/10.1080/02755947.2013.769921.

D. Andrew R. Drake and Nicholas E. Mandrak. Least-cost transportation networks predict spatial interaction of invasion vectors. Ecological Applications, 20(8):2286–2299, 2010. doi: https://doi.org/10.1890/09-2005.1.

Andrew C. Dutterer, Jason R. Dotson, Brandon C. Thompson, Christopher J. Paxton, and William F. Pouder. Estimating Recreational Fishing Effort Using Autonomous Cameras at Boat Ramps versus Creel Surveys. North American Journal of Fisheries Management, 40(6):1367–1378, 2020. doi: https://doi.org/10.1002/nafm.10490.

M. G. Efford. openCR: Open population capture-recapture models. R package version 2.2.2, 2022. URL https://CRAN.R-project.org/package=openCR/.

Jane Elith and John R. Leathwick. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics, 40:677–697, 2009. doi: https://doi.org/10.1146/annurev.ecolsys.110308.120159.

Scott Emmons, Stephen Kobourov, Mike Gallant, and Katy Börner. Analysis of network clustering algorithms and cluster quality metrics at scale. PLOS ONE, 11(7):e0159161, 2016. doi: https://doi.org/10.1371/journal.pone.0159161.

Environment Agency. Environment Agency and Angling Trust welcome lifting of restrictions on recreational fishing, 2020. Available at: https://www.gov.uk/government/news/environment-agency-and-angling-trust-welcome-lifting-of-restrictions-on-recreational-fishing (Accessed: 26 April 2024).

Jianqing Fan and Jinchi Lv. A selective overview of variable selection in high dimensional feature space. Statistica Sinica, 20(1):101–148, 2010.

Fishbrain. Fishbrain, 2022. URL https://fishbrain.com/.

Fishbrain. Fishbrain, 2023. URL https://fishbrain.com/.

Flanders Marine Institute. Union of the ESRI Country shapefile and the Exclusive Economic Zones (version 3), 2020. Available at: https://doi.org/10.14284/403 (Accessed: 10 November 2021).

Robert J. Fletcher Jr., Trevor J. Hefley, Ellen P. Robertson, Benjamin Zuckerberg, Robert A. McCleery, and Robert M. Dorazio. A practical guide for combining data to model species distributions. Ecology, 100(6):e02710, 2019. doi: https://doi.org/10.1002/ecy.2710.

Heath Folpp and Michael Lowry. Factors affecting recreational catch rates associated with a Fish Aggregating Device (FAD) off the NSW coast, Australia. Bulletin of Marine Science, 78(1):185–193, 2006.

Santo Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75–174, 2010. doi: https://doi.org/10.1016/j.physrep.2009.11.002.

Santo Fortunato and Claudio Castellano. Community structure in graphs. In R. Meyers, editor, Computational Complexity. Springer, New York, NY, 2012. doi: https://doi.org/10.1007/978-1-4614-1800-9_33.

Peter Fredman and Lusine Margaryan. 20 years of Nordic nature-based tourism research: a review and future research agenda. Scandinavian Journal of Hospitality and Tourism, 21(1):14–25, 2021. doi: https://doi.org/10.1080/15022250.2020.1823247.

Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software, 33(1):1–22, 2010.

J. Gladju, Biju Sam Kamalam, and A. Kanagaraj. Applications of data mining and machine learning framework in aquaculture and fisheries: A review. Smart Agricultural Technology, 2, 2022. doi: https://doi.org/10.1016/j.atech.2022.100061.

Abigail S. Golden, Christopher M. Free, and Olaf P. Jensen. Angler preferences and satisfaction in a high-threshold bucket-list recreational fishery. Fisheries Research, 220:105364, 2019. doi: https://doi.org/10.1016/j.fishres.2019.105364.

Florencia Grattarola, Diana E. Bowler, and Petr Keil. Integrating presence-only and presence–absence data to model changes in species geographic ranges: An example in the Neotropics. Journal of Biogeography, 50(9):1561–1575, 2023. doi: https://doi.org/10.1111/jbi.14622.

Christopher J. Greenwood, George J. Youssef, Primrose Letcher, Jacqui A. Macdonald, Lauryn J. Hagg, Ann Sanson, Jenn Mcintosh, Delyse M. Hutchinson, John W. Toumbourou, Matthew Fuller-Tyszkiewicz, and Craig A. Olsson. A comparison of penalised regression methods for informing the selection of predictive markers. PLOS ONE, 15:1–14, 2020. doi: https://doi.org/10.1371/journal.pone.0242730.

Casper Gundelund and Christian Skov. Changes in angler demography and angling patterns during the Covid-19 lockdown in spring 2020 measured through a citizen science platform. Marine Policy, 131, 2021. doi: https://doi.org/10.1016/j.marpol.2021.104602.

Casper Gundelund, Robert Arlinghaus, Henrik Baktoft, Kieran Hyder, Paul Venturelli, and Christian Skov. Insights into the users of a citizen science platform for collecting recreational fisheries data. Fisheries Research, 229, 2020. doi: https://doi.org/10.1016/j.fishres.2020.105597.

Casper Gundelund, Paul Venturelli, Bruce W. Hartill, Kieran Hyder, Hans Jakob Olesen, and Christian Skov. Evaluation of a citizen science platform for collecting fisheries data from coastal sea trout anglers. Canadian Journal of Fisheries and Aquatic Sciences, 78(11):1576–1585, 2021. doi: https://doi.org/10.1139/cjfas-2020-0364.

Stefan Gössling, Daniel Scott, and C. Michael Hall. Pandemics, tourism and global change: a rapid assessment of COVID-19. Journal of Sustainable Tourism, 29(1):1–20, 2021. doi: https://doi.org/10.1080/09669582.2020.1758708.

Homa Hajibaba, Ulrike Gretzel, Friedrich Leisch, and Sara Dolnicar. Crisis-resistant tourists. Annals of Tourism Research, 53:46–60, 2015. doi: https://doi.org/10.1016/j.annals.2015.04.001.

Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour, 2021. doi: https://doi.org/10.1038/s41562-021-01079-8.

Thomas Hale, Anna Petherick, Toby Phillips, Jessica Anania, Bernardo Andretti de Mello, Noam Angrist, Roy Barnes, Thomas Boby, Emily Cameron-Blake, Alice Cavalieri, Martina Di Folco, Benjamin Edwards, Lucy Ellen, Jodie Elms, Rodrigo Furst, Liz Gomes Ribeiro, Kaitlyn Green, Rafael Goldszmidt, Laura Hallas, Nadezhda Kamenkovich, Beatriz Kira, Sandhya Laping, Maria Luciano, Saptarshi Majumdar, Thayslene Marques Oliveira, Radhika Nagesh, Annalena Pott, Luyao Ren, Julia Sampaio, Helen Tatlow, Will Torness, Adam Wade, Samuel Webster, Andrew Wood, Hao Zha, Yuxi Zhang, and Andrea Vaccaro. Variation in Government Responses to COVID-19. Version 15. Blavatnik School of Government Working Paper, 2023. Available at: www.bsg.ox.ac.uk/covidtracker (Accessed: 17 April 2024).

C. Michael Hall. Second home tourism: An international review. Tourism Review International, 18 (3):115–135, 2014. doi: https://doi.org/10.3727/154427214X14101901317039.

C. Michael Hall. Tourism and fishing. Scandinavian Journal of Hospitality and Tourism, 21(4): 361–373, 2021. doi: https://doi.org/10.1080/15022250.2021.1955739.

Olga Hannonen and Gijsbert Hoogendoorn. Angling tourism: A state-of-the-art review. Matkailututkimus, 18(2):6–30, 2022. doi: https://doi.org/10.33351/mt.116555.

Bruce W. Hartill, Stephen M. Taylor, Krystle Keller, and Marc Simon Weltersbach. Digital camera monitoring of recreational fishing effort: Applications and challenges. Fish and Fisheries, 21(1): 204–215, 2020. doi: https://doi.org/10.1111/faf.12413.

M. Hebiri and J. Lederer. How correlations influence lasso prediction. IEEE Transactions On Information Theory, 59(3), 2013.

L. Heermann, M. Emmrich, M. Heynen, M. Dorow, U. König, J. Borcherding, and R. Arlinghaus. Explaining recreational angling catch rates of Eurasian perch, Perca fluviatilis: the role of natural and fishing-related environmental factors. Fisheries Management and Ecology, 20(2-3):187–200, 2013. doi: https://doi.org/10.1111/fme.12000.

Samantha A. Hook, Adam Brown, Brigid Bell, Jo Kroese, Zachary Radford, and Kieran Hyder. The Impact of COVID-19 on Participation, Effort, Physical Activity, and Well-Being of Sea Anglers in the UK. Frontiers in Marine Science, 9, 2022. doi: https://doi.org/10.3389/fmars.2022.815617.

A. Howarth, A.L. Jeanson, A.E.I. Abrams, C. Beaudoin, I. Mistry, A. Berberi, N. Young, V.M. Nguyen, S.J. Landsman, A.N. Kadykalo, A.J. Danylchuk, and S.J. Cooke. COVID-19 restrictions and recreational fisheries in Ontario, Canada: preliminary insights from an online angler survey. Fisheries Research, 240:105961, 2021. doi: https://doi.org/10.1016/j.fishres.2021.105961.

Kieran Hyder, Mike Armstrong, Keno Ferter, and Harry V. Strehlow. Recreational sea fishing – the high value forgotten catch. ICES Insight, 51:8–15, 2014.

Kieran Hyder, Marc Simon Weltersbach, Mike Armstrong, Keno Ferter, Bryony Townhill, Anssi Ahvonen, Robert Arlinghaus, Andrei Baikov, Manuel Bellanger, Janis Birzaks, Trude Borch, Giulia Cambie, Martin de Graaf, Hugo M. C. Diogo, Łukasz Dziemian, Ana Gordoa, Ryszard Grzebielec, Bruce Hartill, Anders Kagervall, Kostas Kapiris, Martin Karlsson, Alf Ring Kleiven, Adam M. Lejk, Harold Levrel, Sabrina Lovell, Jeremy Lyle, Pentti Moilanen, Graham Monkman, Beatriz Morales-Nin, Estanis Mugerza, Roi Martinez, Paul O'Reilly, Hans Jakob Olesen, Anastasios Papadopoulos, Pablo Pita, Zachary Radford, Krzysztof Radtke, William Roche, Delphine Rocklin, Jon Ruiz, Callum Scougal, Roberto Silvestri, Christian Skov, Scott Steinback, Andreas Sundelöf, Arvydas Svagzdys, David Turnbull, Tessa van der Hammen, David van Voorhees, Frankwin van Winsen, Thomas Verleye, Pedro Veiga, Jon-Helge Vølstad, Lucia Zarauz, Tomas Zolubas, and Harry V. Strehlow. Recreational sea fishing in Europe in a global context—Participation rates,

fishing effort, expenditure, and implications for monitoring and assessment. Fish and Fisheries, 19 (2):225–243, 2018. doi: https://doi.org/10.1111/faf.12251.

Kieran Hyder, Adam Brown, Mike Armstrong, Brigid Bell, Kirsty Bradley, Elena Couce, Iain Gibson, Francesca Hardman, James Harrison, Vanessa Haves, Samantha Hook, Jo Kroese, Gavin Mellor, Eleanor MacLeod, Angela Muench, Zachary Radford, and Bryony Townhill. Participation, catches and economic impact of sea anglers resident in the UK in 2016 & 2017. Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK, 2020a.

Kieran Hyder, Christos D Maravelias, Marloes Kraan, Zachary Radford, and Raul Prellezo. Marine recreational fisheries — current state and future opportunities. ICES Journal of Marine Science, 77(6):2171–2180, 2020b. doi: https://doi.org/10.1093/icesjms/fsaa147.

Kieran Hyder, Adam Brown, Mike Armstrong, Brigid Bell, Samantha Alison Hook, Jo Kroese, and Zachary Radford. Participation, effort, and catches of sea anglers resident in the UK in 2018 & 2019. Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK, 2021.

Kieran Hyder, Adam Brown, Brigid Bell, Kirsty Bradley, Wendy Edwards, Samantha A. Hook, Rebecca Mills, Jo Kroese, and Zachary Radford. Participation, effort, catches, and impact of COVID-19 of sea anglers resident in the UK in 2016-21. Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, UK, 2024.

Nick J.B. Isaac, Marta A. Jarzyna, Petr Keil, Lea I. Dambly, Philipp H. Boersch-Supan, Ella Browning, Stephen N. Freeman, Nick Golding, Gurutzeta Guillera-Arroita, Peter A. Henrys, Susan Jarvis, José Lahoz-Monfort, Jörn Pagel, Oliver L. Pescott, Reto Schmucki, Emily G. Simmonds, and Robert B. O'Hara. Data Integration for Large-Scale Models of Species Distributions. Trends in Ecology & Evolution, 35(1):56–67, 2020. doi: https://doi.org/10.1016/j.tree.2019.08.006.

Ali Jalali, Justin D. Bell, Harry K. Gorfine, Simon Conron, and Khageswor Giri. Angling to reach a destination to fish—Exploring the land and water travel dynamics of recreational fishers in Port Phillip Bay, Australia. Frontiers in Marine Science, 8:793074, 2022. doi: https://doi.org/10.3389/fmars.2021.793074.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning. Springer, New York, 2017. doi: https://doi.org/10.1007/978-1-4614-7138-7.

Ryan P. Jiorle, Robert N. M. Ahrens, and Micheal S. Allen. Assessing the Utility of a Smartphone App for Recreational Fishery Catch Data. Fisheries, 41(12):758–766, 2016. doi: https://doi.org/10.1080/03632415.2016.1249709.

Alison Johnston, Wesley M. Hochachka, Matthew E. Strimas-Mackey, Viviana Ruiz Gutierrez, Orin J. Robinson, Eliot T. Miller, Tom Auer, Steve T. Kelling, and Daniel Fink. Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. Diversity and Distributions, 27:1265– 1277, 2021. doi: https://doi.org/10.1111/ddi.13271.

Alison Johnston, Eleni Matechou, and Emily B. Dennis. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. Methods in Ecology and Evolution, 14(1):103–116, 2023. doi: https://doi.org/10.1111/2041-210X.13834.

Fiona D. Johnston, Sean Simmons, Brett van Poorten, and Paul Venturelli. Comparative analyses with conventional surveys reveal the potential for an angler app to contribute to recreational fisheries monitoring. Canadian Journal of Fisheries and Aquatic Sciences, 79(1):31–46, 2022.

Olle Järv, Ago Tominga, Kerli Müürisepp, and Siiri Silm. The impact of COVID-19 on daily lives of transnational people based on smartphone data: Estonians in Finland. Journal of Location Based Services, 15(3):169–197, 2021. doi: https://doi.org/10.1080/17489725.2021.1887526.

Melissa A. Karp, Jason S. Link, Max Grezlik, Steve Cadrin, Gavin Fay, Patrick Lynch, Howard Townsend, Richard D. Methot, Grant D. Adams, Kristan Blackhart, Caren Barceló, Andre Buchheister, Matthew Cieri, David Chagaris, Villy Christensen, J. Kevin Craig, Jonathan Cummings, Matthew D. Damiano, Mark Dickey-Collas, Bjarki Thór Elvarsson, Sarah Gaichas, Melissa A. Haltuch, Janne B. Haugen, Daniel Howell, Isaac C. Kaplan, Willem Klajbor, Scott I. Large, Michelle Masi, Jason McNamee, Brandon Muffley, Sarah Murray, Éva Plagányi, David Reid, Anna Rindorf, Sagarese Skyler R., Amy M. Schueller, Robert Thorpe, James T. Thorson, Maciej T. Tomczak, Vanessa Trijoulet, and Rudi Voss. Increasing the uptake of multispecies models in fisheries management. ICES Journal of Marine Science, 80(2):243–257, 2023. doi: https://doi.org/10.1093/icesjms/fsad001.

Melissa A. Karp, Megan Cimino, J. Kevin Craig, Daniel P. Crear, Christopher Haak, Elliott L. Hazen, Isaac Kaplan, Donald R. Kobayashi, Hassan Moustahfid, Barbara Muhling, Malin L. Pinsky, Laurel A. Smith, James T. Thorson, and Phoebe A. Woodworth-Jefcoats. Applications of species distribution modeling and future needs to support marine resource management. ICES Journal of Marine Science, 82(3):fsaf024, 2025. doi: https://doi.org/10.1093/icesjms/fsaf024.

Yoon Jung Kim, Dong Kun Lee, and Choong Ki Kim. Spatial tradeoff between biodiversity and nature-based tourism: Considering Mobile Phone-Driven Visitation Pattern. Global Ecology and Conservation, 21:e00899, 2020.

Eric D. Kolaczyk and Gábor Csárdi. Statistical Analysis of Network Data with R. Springer International Publishing AG, Reading, Massachusetts, 2020.

Zoltán Kovács, György Vida, Ábel Elekes, and Tamás Kovalcsik. Combining social media and mobile positioning data in the analysis of tourist flows: A case study from Szeged, Hungary. Sustainability, 13(5):2926, 2021.

Anne Köchling, Marit Gundersen Engeset, Julian Reif, Nadine Yarar, Jarmo Ritalahti, Eva Holmberg, and Jan Velvin. Between fearful homebodies and carefree travel lovers: identifying tourist segments during the Covid-19 pandemic in Finland, Germany, and Norway. Current Issues in Tourism, 25 (7):1074–1087, 2022. doi: https://doi.org/10.1080/13683500.2022.2026304.

Yen E. Lam-González, Carmelo J. Leon, Matias M. Gonzalez Hernandez, and Javier de Leon. The structural relationships of destination image, satisfaction, expenditure and loyalty in the context of fishing tourism. Scandinavian Journal of Hospitality and Tourism, 21(4):422–441, 2021. doi: https://doi.org/10.1080/15022250.2021.1884596.

Andrea Lancichinetti and Santo Fortunato. Community detection algorithms: a comparative analysis. Physical Review E, 80(5):056117, 2009. doi: https://doi.org/10.1103/PhysRevE.80.056117.

Emmanuel Lazaridis. lunar: Lunar Phase & Distance, Seasons and Other Environmental Factors (version 0.1-04), 2014. URL http://statistics.lazaridis.eu.

Wolf-Christian Lewin, Robert Arlinghaus, and Thomas Mehner. Documented and potential biological impacts of recreational fishing: Insights for management and conservation. Reviews in Fisheries Science, 14(4):305–367, 2006. doi: https://doi.org/10.1080/10641260600886455.

Wolf-Christian Lewin, Marc Simon Weltersbach, Keno Ferter, Kieran Hyder, Estanis Mugerza, Raúl Prellezo, Zachary Radford, Lucia Zarauz, and Harry Vincent Strehlow. Potential Environmental Impacts of Recreational Fishing on Marine Fish Stocks and Ecosystems. Reviews in Fisheries Science & Aquaculture, 27(3):287–330, 2019. doi: https://doi.org/10.1080/23308249.2019.1586829.

Nerea Lezama-Ochoa, Maria Grazia Pennino, Martin A. Hall, Jon Lopez, and Hilario Murua. Using a Bayesian modelling approach (INLA-SPDE) to predict the occurrence of the Spinetail Devil Ray (Mobular mobular). Scientific Reports, 10, 2020. doi: https://doi.org/10.1038/s41598-020-73879-3.

Carmelo J. León, Jorge E. Araña, and Arturo Melián. Tourist use and preservation benefits from big-game fishing in the Canary Islands. Tourism Economics, 9(1):53–65, 2003. doi: https://doi.org/10.5367/000000003101298268.

Finn Lindgren and Håvard Rue. Bayesian Spatial Modelling with R-INLA. Journal of Statistical Software, 63(19):1–25, 2015. doi: https://10.18637/jss.v063.i19.

Finn Lindgren, Haakon Bakka, David Bolin, Elias Krainski, and Håvard Rue. A diffusion-based spatio-temporal extension of Gaussian Matérn fields. SORT (Statistics and Operations Research Transactions), 48(1):3–66, 2024. doi: https://doi.org/10.57645/20.8080.02.13.

Yan Liu, Xian Cheng, Stephen Shaoyi Liao, and Feng Yang. The impact of COVID-19 on the tourism and hospitality Industry: Evidence from international stock markets. The North American Journal of Economics and Finance, 64, 2023. doi: https://doi.org/10.1016/j.najef.2022.101875.

Carlos Llano, Juan Pardo, Santiago Pérez-Balsalobre, and Julián Pérez. Estimating multicountry tourism flows by transport mode. Annals of Tourism Research, 103:103672, 2023a. doi: https://doi.org/10.1016/j.annals.2023.103672.

Carlos Llano, Juan Pardo, Santiago Pérez-Balsalobre, and Julián Pérez. Dataset on international tourists flows, 2023b. Available at: https://public.tableau.com/app/profile/author6764/viz/tourism_internacional_test_16511462531930/Dashboard1 (Accessed: 21 August 2024).

Josep Lloret, Ian G. Cowx, Henrique Cabral, Margarida Castro, Toni Font, Jorge M.S. Gonçalves, Ana Gordoa, Ellen Hoefnagel, Sanja Matić-Skoko, Eirik Mikkelsen, Beatriz Morales-Nin, Dimitrios K. Moutopoulos, Marta Muñoz, Miguel Neves dos Santos, Pedro Pintassilgo, Cristina Pita, Konstantinos I. Stergiou, Vahdet Ünal, Pedro Veiga, and Karim Erzini. Small-scale coastal fisheries in European Seas are not what they were: Ecological, social and economic changes. Marine Policy, 98:176–186, 2018. doi: https://doi.org/10.1016/j.marpol.2016.11.007.

Grace L. Loppnow, Kris Vascotto, and Paul A. Venturelli. Invasive smallmouth bass (Micropterus dolomieu): history, impacts, and control. Management of Biological Invasions, 4(3):191–206, 2013. doi: http://dx.doi.org/10.3391/mbi.2013.4.3.02.

Michael J. Louison, Cory D. Suski, and Jeffrey A. Stein. Largemouth bass use prior experience, but not information from experienced conspecifics, to avoid capture by anglers. Fisheries Management and Ecology, 26(6):600–610, 2019. doi: https://doi.org/10.1111/fme.12372.

M. Lowry, D. Williams, and Y. Metti. Lunar landings—relationship between lunar phase and catch rates for an australian gamefish-tournament fishery. Fisheries Research, 88(1–3):15–23, 2007. doi: https://doi.org/10.1016/j.fishres.2007.07.011.

Fragkiskos D. Malliaros and Michalis Vazirgiannis. Clustering and community detection in directed networks: A survey. Physics Reports, 533(4):95–142, 2013. doi: https://doi.org/10.1016/j.physrep.2013.08.002.

Anabela Marques Santos, Carmen Madrid, Karel Haegeman, and Alessandro Rainoldi. Behavioural changes in tourism in times of Covid-19. JRC121262, 22(3):121–147, 2020. doi: https://data.europa.eu/doi/10.2760/00411.

Florence Matutini, Jacques Baudry, Guillaume Pain, Morgane Sineau, and Joséphine Pithon. How citizen science could improve species distribution models and their independent assessment. Ecology and Evolution, 11(7):3028–3039, 2021. doi: https://doi.org/10.1002/ece3.7210.

Mark N. Maunder and André E. Punt. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142:61–74, 2013. doi: https://doi.org/10.1016/j.fishres.2012.07.025.

Bob McKercher and Barry Mak. The impact of distance on international tourism demand. Tourism Management Perspectives, 31:340–347, 2019. doi: https://doi.org/10.1016/j.tmp.2019.07.004.

Bob McKercher, Andrew Chan, and Celia Lam. The impact of distance on international tourist movements. Journal of Travel Research, 47(2):208–224, 2008. doi: https://doi.org/10.1177/0047287508321191.

Alexandra McManus, Wendy Hunt, Jessica Storey, and James White. Identifying the health and well-being benefits of recreational fishing. FRDC Project, 2011.

Annabell Merkel, Filippa Säwe, and Cecilia Fredriksson. The seaweed experience: exploring the potential and value of a marine resource. Scandinavian Journal of Hospitality and Tourism, 21(4):391–406, 2021. doi: https://doi.org/10.1080/15022250.2021.1879671.

Met Office. National Meteorological Library and Archive Fact sheet 6 - The Beaufort Scale. Exeter, UK, 10/0425 edition, 2010. URL https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/library/publications/factsheets/factsheet_6-the-beaufort-scale.pdf. (Accessed: 28 September 2021).

Stephen R. Midway, Abigail J. Lynch, Brandon K. Peoples, Michael Dance, and Rex Caffey. COVID-19 influences on US recreational angler behavior. PLOS ONE, 16(8):1–16, 2021. doi: https://doi.org/10.1371/journal.pone.0254652.

David A. W. Miller, Krishna Pacifici, Jamie S. Sanderlin, and Brian J. Reich. The recent past and promising future for data integration methods to estimate species' distributions. Methods in Ecology and Evolution, 10(1):22–37, 2019. doi: https://doi.org/10.1111/2041-210X.13110.

Tom Mordue. Angling in modernity: a tour through society, nature and embodied passion. Current Issues in Tourism, 12(5–6):529–552, 2009. doi: https://doi.org/10.1080/13683500903043244.

Tom Mordue. Game-angling Tourism: Connecting People, Places and Natures. International Journal of Tourism Research, 18(3):269–276, 2016. doi: https://doi.org/10.1002/jtr.2003.

Virginia Morera-Pujol, Philip S. Mostert, Kilian J. Murphy, Tim Burkitt, Barry Coad, Barry J. McMahon, Maarten Nieuwenhuis, Kevin Morelle, Alastair I. Ward, and Simone Ciuti. Bayesian species distribution models integrate presence-only and presence–absence data to predict deer distribution and relative abundance. Ecography, 2023(2):06451, 2023. doi: https://doi.org/10.1111/ecog.06451.

Camille L. Mosley, Colin J. Dassow, John Caffarelli, Alexander J. Ross, Greg G. Sass, Stephanie L. Shaw, Christopher T. Solomon, and Stuart E. Jones. Species differences, but not habitat, influence catch rate hyperstability across a recreational fishery landscape. Fisheries Research, 255:106438, 2022. doi: https://doi.org/10.1016/j.fishres.2022.106438.

Falko Mueller. Examining COVID-19-triggered changes in spatial connectivity patterns in the European air transport network up to June 2021. Research in Transportation Economics, 94: 101127, 2022. doi: https://doi.org/10.1016/j.retrec.2021.101127.

National Research Council. Review of Recreational Fisheries Survey Methods. The National Academies Press, Washington, DC, 2006. doi: https://doi.org/10.17226/11616. (Accessed: 28 July 2021).

Natural Earth. 1:10m Physical Vectors - Coastlines. Version 4.1.0, 2021. Available at: https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-coastline/ (Accessed: 22 July 2021).

M. E. J. Newman. Fast algorithm for detecting community structure in networks. Physical Review E, 69(6):066133, 2004. doi: https://doi.org/10.1103/PhysRevE.69.066133.

M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. Physical Review E, 74(3):036104, 2006. doi: https://doi.org/10.1103/PhysRevE.74.036104.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Physical Review E, 69(2):026113, 2004. doi: https://doi.org/10.1103/PhysRevE.69.026113.

Mark E. J. Newman. Networks: an introduction. Oxford University Press, 2010. ISBN 9780191637766.

Mark E. J. Newman. Spectral methods for community detection and graph partitioning. Physical Review E, 88(4):042822, 2013. doi: https://doi.org/10.1103/PhysRevE.88.042822.

Janet Van Niekerk, Elias Krainski, Denis Rustand, and Håvard Rue. A new avenue for Bayesian inference with INLA. Computational Statistics & Data Analysis, 181, 2023. doi: https://doi.org/10.1016/j.csda.2023.107692.

N Nikolioudakis, H J Skaug, A H Olafsdottir, T Jansen, J A Jacobsen, and K Enberg. Drivers of the summer-distribution of Northeast Atlantic mackerel (Scomber scombrus) in the Nordic Seas from 2011 to 2017; a Bayesian hierarchical modelling approach. ICES Journal of Marine Science, 76 (2):530–548, 2019. doi: https://doi.org/10.1093/icesjms/fsy085.

Eurico Mesquita Noleto-Filho, Ronaldo Angelini, Jeroen Steenbeek, and Adriana Rosa Carvalho. New, flexible and open-source fisheries self-reporting app: The Shiny4SelfReport. SoftwareX, 16, 2021. doi: https://doi.org/10.1016/j.softx.2021.100843.

Jason T. Papenfuss, Nicholas Phelps, David Fulton, and Paul A. Venturelli. Smartphones reveal angler behavior: a case study of a popular mobile fishing application in Alberta, Canada. Fisheries, 40(7): 318–327, 2015. doi: https://doi.org/10.1080/03632415.2015.1049693.

Iosu Paradinas, David Conesa, Antonio López-Quílez, and José María Bellido. Spatio-Temporal model structures with shared components for semi-continuous species distribution modelling. Spatial Statistics, 22(2):434–450, 2017. doi: https://doi.org/10.1016/j.spasta.2017.08.001.

Iosu Paradinas, Janine B. Illian, Alexandre Alonso-Fernández, Maria Grazia Pennino, and Sophie Smout. Combining fishery data through integrated species distribution models. ICES Journal of Marine Science, 80(10):2579–2590, 2023. doi: https://doi.org/10.1093/icesjms/fsad069.

Yves Paradis, Simon Bernatchez, Dominique Lapointe, and Steven J. Cooke. Can You Fish in a Pandemic? An Overview of Recreational Fishing Management Policies in North America During the COVID-19 Crisis. Fisheries, 46(2):81–85, 2020. doi: https://doi.org/10.1002/fsh.10544.

Menelaos Pavlou, Gareth Ambler, Shaun Seaman, Maria De Iorio, and Rumana Z Omar. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Statistics in Medicine, 35(7):1159–1177, 2016. doi: https://doi.org/10.1002/sim.6782.

M. G. Pawson, H. Glenn, and G. Padda. The definition of marine recreational fishing in Europe. Marine Policy, 32(3):339–350, 2008. doi: https://doi.org/10.1016/j.marpol.2007.07.001.

Edzer Pebesma and Roger S. Bivand. S classes and methods for spatial data: the sp package. R News, 5(2):9–13, 2005. URL https://CRAN.R-project.org/doc/Rnews/.

Pablo Pita, Gillian B. Ainsworth, Bernardino Alba, Antônio B. Anderson, Manel Antelo, Josep Alós, Iñaki Artetxe, Jérôme Baudrier, José J. Castro, Belén Chicharro, Karim Erzini, Keno Ferter, Mafalda Freitas, Laura García-de-la Fuente, José A. García-Charton, María Giménez-Casalduero, Antoni M. Grau, Hugo Diogo, Ana Gordoa, Filipe Henriques, Kieran Hyder, David Jiménez-Alvarado, Paraskevi K. Karachle, Josep Lloret, Martin Laporta, Adam M. Lejk, Arnau L. Dedeu, Pablo Martín-Sosa, Lllibori Martínez, Antoni M. Mira, Beatriz Morales-Nin, Estanis Mugerza, Hans J. Olesen, Anastasios Papadopoulos, João Pontes, José J. Pascual-Fernández, Ariadna Purroy, Milena Ramires, Mafalda Rangel, José Amorim Reis-Filho, Jose L. Sánchez-Lizaso, Virginia Sandoval, Valerio Sbragaglia, Luis Silva, Christian Skov, Iván Sola, Harry V. Strehlow, María A. Torres, Didzis Ustups, Tessa van der Hammen, Pedro Veiga, Leonardo A. Venerus, Thomas Verleye, Sebastián Villasante, Marc Simon Weltersbach, and Lucía Zarauz. First Assessment of the Impacts of the COVID-19 Pandemic on Global Marine Recreational Fisheries. Frontiers in Marine Science, 8, 2021. doi: https://doi.org/10.3389/fmars.2021.735741.

Pablo Pita, Matthew O. Gribble, Manel Antelo, Gillian Ainsworth, Kieran Hyder, Matilda van den Bosch, and Sebastián Villasante. Recreational Fishing, Health and Well-Being: Findings from a Cross-Sectional Survey. Ecosystems and People, 18(1):530–546, 2022. doi: https://doi.org/10.1080/26395916.2022.2112291.

Michael J. O. Pocock, John C. Tweddle, Joanna Savage, Lucy D. Robinson, and Helen E. Roy. The diversity and evolution of ecological and environmental citizen science. PLOS ONE, 12(4), 2017. doi: https://doi.org/10.1371/journal.pone.0172579.

Giovanni Poggiato, Tamara Münkemüller, Daria Bystrova, Julyan Arbel, James S. Clark, and Wilfried Thuiller. On the Interpretations of Joint Modeling in Community Ecology. Trends in Ecology & Evolution, 36(5):391–401, 2021. doi: https://doi.org/10.1016/j.tree.2021.01.002.

Heidi Pokki, Jani Pellikka, Päivi Eskelinen, and Pentti Moilanen. Regional fishing site preferences of subgroups of Finnish recreational fishers. Scandinavian Journal of Hospitality and Tourism, 21(4):442–457, 2021. doi: https://doi.org/10.1080/15022250.2020.1860814.

K. L. Pope, C. J. Chizinski, C. L. Wiley, and D. R. Martin. Influence of anglers' specializations on catch, harvest, and bycatch of targeted taxa. Fisheries Research, 183:128–137, 2016.

J. R. Pulver. Does the Lunar Cycle Affect Reef Fish Catch Rates? North American Journal of Fisheries Management, 37(3):536–549, 2017. doi: https://doi.org/10.1080/02755947.2017.1293574.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL https://www.R-project.org/.

R Package Documentation. cv.zipath: Cross-validation for zipath., 2021. URL https://rdrr.io/cran/mpath/man/cv.zipath.html. (Accessed: 02-10-2021).

Zachary Radford, Kieran Hyder, Lucía Zarauz, Estanis Mugerza, Keno Ferter, Raul Prellezo, Harry Vincent Strehlow, Bryony Townhill, Wolf-Christian Lewin, and Marc Simon Weltersbach. The impact of marine recreational fishing on key fish stocks in European waters. PLOS ONE, 13 (9):e0201666, 2018. doi: https://doi.org/10.1371/journal.pone.0201666.

Usha Nandini Raghavan, R. Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E, 76(3):036106, 2007. doi: https://doi.org/10.1103/PhysRevE.76.036106.

Sarah M. Roberts, Patrick N. Halpin, and James S. Clark. Jointly modeling marine species to inform the effects of environmental change on an ecological community in the Northwest Atlantic. Scientific Reports, 12(1):132, 2022. doi: https://doi.org/10.1038/s41598-021-04110-0.

Orin J. Robinson, Viviana Ruiz-Gutierrez, Daniel Fink, Robert J. Meese, Marcel Holyoak, and Evan G. Cooch. Using citizen science data in integrated population models to inform conservation. Biological Conservation, 227:361–368, 2018. doi: https://doi.org/10.1016/j.biocon.2018.10.002.

Orin J. Robinson, Viviana Ruiz-Gutierrez, Mark D. Reynolds, Gregory H. Golet, Matthew Strimas-Mackey, and Daniel Fink. Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. Diversity and Distributions, 26(8):976–986, 2020. doi: https://doi.org/10.1111/ddi.13068.

Michał Roman, Monika Roman, Emilia Grzegorzewska, Piotr Pietrzak, and Kamil Roman. Influence of the COVID-19 Pandemic on Tourism in European Countries: Cluster Analysis Findings. Sustainability, 14(3), 2022. doi: https://doi.org/10.3390/su14031602.

Håvard Rue, Sara Martino, and Nicholas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71:319–392, 2009. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Clark S. Rushing, J. Andrew Royle, David J. Ziolkowski, and Keith L. Pardieck. Modeling spatially and temporally complex range dynamics when detection is imperfect. Scientific Reports, 9:12805, 2019. doi: https://doi.org/10.1038/s41598-019-48851-5.

Karina L. Ryan, Cameron J. Desfosses, Ainslie M. Denham, Stephen M. Taylor, and Gary Jackson. Initial insights on the impact of COVID-19 on boat-based recreational fishing in Western Australia. Marine Policy, 132:104646, 2021. doi: https://doi.org/10.1016/j.marpol.2021.104646.

Erki Saluveer, Janika Raun, Margus Tiru, Laura Altin, Jaanus Kroon, Tarass Snitsarenko, Anto Aasa, and Siiri Silm. Methodological framework for producing national tourism statistics from mobile positioning data. Annals of Tourism Research, 81:102895, 2020. doi: https://doi.org/10.1016/j.annals.2020.102895.

Asrul Sani, Bahriddin Abapihi, Mukhsar, Ramadhan Tosepu, Ida Usman, and Gusti Arviani Rahman. Bayesian temporal, spatial and spatio-temporal models of dengue in a small area with INLA. International Journal of Modelling and Simulation, pages 1–13, 2022.

Giovanni Scardoni and Carlo Laudanna. Centralities Based Analysis of Complex Networks. New Frontiers in Graph Theory, 323, 2012.

Satu Elisa Schaeffer. Graph clustering. Computer Science Review, 1(1):27–64, 2007. doi: https://doi.org/10.1016/j.cosrev.2007.05.001.

Michael Schaub and Fitsum Abadi. Integrated population models: a novel analysis framework for deeper insights into population dynamics. Journal of Ornithology, 152(Suppl 1):227–237, 2011. doi: https://doi.org/10.1007/s10336-010-0632-7.

Michael Schaub, Mark N. Maunder, Marc Kéry, James T. Thorson, Eiren K. Jacobson, and André E. Punt. Lessons to be learned by comparing integrated fisheries stock assessment models (SAMs) with integrated population models (IPMs). Fisheries Research, 272:106925, 2024. doi: https://doi.org/10.1016/j.fishres.2023.106925.

B. Schrödle and L. Held. Spatio-temporal disease mapping using INLA. Environmetrics, 22:725–734, 2011. doi: https://doi.org/10.1002/env.1065.

Daniel Scott, Stefan Gössling, and C. Michael Hall. International tourism and climate change. Wiley Interdisciplinary Reviews: Climate Change, 3(3):213–232, 2012. doi: https://doi.org/10.1002/wcc.165.

Noam Shoval and Rein Ahas. The use of tracking technologies in tourism research: The first decade. Tourism Geographies, 18(5):587–606, 2016.

Christian Skov, Kieran Hyder, Casper Gundelund, Anssi Ahvonen, Jérôme Baudrier, Trude Borch, Sara deCarvalho, Karim Erzini, Keno Ferter, Fabio Grati, Tessa van derHammen, Jan Hinriksson, Rob Houtman, Anders Kagervall, Kostas Kapiris, Martin Karlsson, Adam M Lejk, Jeremy M Lyle, Roi Martinez-Escauriaza, Pentti Moilanen, Estanis Mugerza, Hans Jakob Olesen, Anastasios Papadopoulos, Pablo Pita, João Pontes, Zachary Radford, Krzysztof Radtke, Mafalda Rangel, Oscar Sagué, Hege A Sande, Harry V Strehlow, Rūdolfs Tutiņš, Pedro Veiga, Thomas Verleye, Jon Helge Vølstad, Joseph W Watson, Marc Simon Weltersbach, Didzis Ustups, and Paul A Venturelli. Expert opinion on using angler Smartphone apps to inform marine fisheries management: status, prospects, and needs. ICES Journal of Marine Science, 78(3):967–978, 2021. doi: https://doi.org/10.1093/icesjms/fsaa243.

G. Smith. Step away from stepwise. Journal of Big Data, 5(32), 2018. doi: https://doi.org/10.1186/s40537-018-0143-6.

Stian Stensland. Landowners' perception of risk sources and risk management strategies in Norwegian salmon angling tourism. Scandinavian Journal of Hospitality and Tourism, 13(3):208–227, 2013.

Stian Stensland, Mehmet Mehmetoglu, Åste Sætre Liberg, and Øystein Aas. Angling destination loyalty – A structural model approach of freshwater anglers in Trysil, Norway. Scandinavian Journal of Hospitality and Tourism, 21(4):407–421, 2021. doi: https://doi.org/10.1080/15022250.2021.1921022.

Brian L. Sullivan, Jocelyn L. Aycrigg, Jessie H. Barry, Rick E. Bonney, Nicholas Bruns, Caren B. Cooper, Theo Damoulas, André A. Dhondt, Tom Dietterich, Andrew Farnsworth, Daniel Fink, John W. Fitzpatrick, Thomas Fredericks, Jeff Gerbracht, Carla Gomes, Wesley M. Hochachka, Marshall J. Iliff, Carl Lagoze, Frank A. La Sorte, Matthew Merrifield, Will Morris, Tina B. Phillips, Mark Reynolds, Amanda D. Rodewald, Kenneth V. Rosenberg, Nancy M. Trautmann, Andrea Wiggins, David W. Winkler, Weng-Keen Wong, Christopher L. Wood, Jun Yu, and Steve Kelling. The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation, 169:31–40, 2014. doi: https://doi.org/10.1016/j.biocon.2013.11.003.

Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. Comparing methods for comparing networks. Scientific Reports, 9:17557, 2019. doi: https://doi.org/10.1038/s41598-019-53708-y.

Aleksandra Terzić, Biljana Petrevska, and Dunja Demirović Bajrami. Personalities shaping travel behaviors: post-COVID scenario. Journal of Tourism Futures, 2022. doi: https://doi.org/10.1108/JTF-02-2022-0043.

James T. Thorson, James N. Ianelli, Elise A. Larsen, Leslie Ries, Mark D. Scheuerell, Cody Szuwalski, and Elise F. Zipkin. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Global Ecology and Biogeography, 25(9):1144–1158, 2016. doi: https://doi.org/10.1111/geb.12464.

V. A. Traag, L. Waltman, and N. J. Van Eck. From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports, 9:5233, 2019. doi: https://doi.org/10.1038/s41598-019-41695-z.

Caroline M. Tucker, Simone Collier, Geoffrey Legault, George E. Morgan, and Derrick K. de Kerckhove. Estimating angler effort and catch from a winter recreational fishery using a novel Bayesian methodology to integrate multiple sources of creel survey data. Fisheries Research, 272:106932, 2024. doi: https://doi.org/10.1016/j.fishres.2023.106932.

UN Tourism. 2020: Worst Year in Tourism History with 1 Billion Fewer International Arrivals. UNWTO News Release, January 2021. URL https://www.unwto.org/news/2020-worst-year-in-tourism-history-with-1-billion-fewer-international-arrivals.

UN Tourism. Tourism Grows 4% in 2021 but Remains Far Below Pre-Pandemic Levels. UNWTO News Release, January 2022. URL https://www.unwto.org/news/tourism-grows-4-in-2021-but-remains-far-below-pre-pandemic-levels.

United Nations. International Recommendations for Tourism Statistics 2008. ST/ESA/STAT/SER.M/83/Rev.1. New York, 2010. Available at: https://www.unwto.org/tourism-statistics/on-basic-tourism-statistics-irts-2008 (Accessed: 17 May 2024).

Lotte van den Heuvel, Malgorzata Blicharska, Stian Stensland, and Patrik Rönnbäck. Been there, done that? Effects of centrality-to-lifestyle and experience use history on angling tourists' loyalty to a Swedish salmon fishery. Journal of Outdoor Recreation and Tourism, 39, 2022. doi: https://doi.org/10.1016/j.jort.2022.100549.

Paul A Venturelli, Kieran Hyder, and Christian Skov. Angler apps as a source of recreational fisheries data: opportunities, challenges and proposed standards. Fish and Fisheries, 18(3):578–595, 2017. doi: https://doi.org/10.1111/faf.12189.

Ken Wakita and Toshiyuki Tsurumi. Finding community structure in mega-scale social networks. In Proceedings of the 16th International Conference on World Wide Web, WWW '07, pages 1275–1276, New York, NY, USA, 2007. Association for Computing Machinery. doi: https://doi.org/10.1145/1242572.1242805.

Staffan Waldo and Anton Paulrud. Obstacles to developing recreational fishing enterprises in Sweden. Scandinavian Journal of Hospitality and Tourism, 12(2):121–139, 2012.

Zhu Wang. mpath: Regularized Linear Models, R package version 0.4-2.16, 2020. URL https://CRAN.R-project.org/package=mpath.

Zhu Wang, Shuangge Ma, Ching-Yun Wang, Michael Zappitelli, Prasad Devarajan, and Chirag Parikh. EM for regularized zero-inflated regression models with applications to postoperative morbidity after cardiac surgery in children. Statistics in Medicine, 33(29):5192–5208, 2014. doi: https://doi.org/10.1002/sim.6314.

Brian Wheeller and C. Michael Hall. Fish tales, red herrings: (and gaffes?). Scandinavian Journal of Hospitality and Tourism, 21(4):458–469, 2021. doi: https://doi.org/10.1080/15022250.2021.1887759.

Mark J. Whittingham, Philip A. Stephens, Richard B. Bradbury, and Robert P. Freckleton. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology, 75:1182–1189, 2006. doi: https://doi.org/10.1111/j.1365-2656.2006.01141.x.

David P. Wilkinson, Nick Golding, Gurutzeta Guillera-Arroita, Reid Tingley, and Michael A. McCarthy. Defining and evaluating predictions of joint species distribution models. Methods in Ecology and Evolution, 12(3):394–404, 2021. doi: https://doi.org/10.1111/2041-210X.13518.

Elias Willberg, Olle Järv, Tuomas Väisänen, and Tuuli Toivonen. Escaping from Cities during the COVID-19 Crisis: Using Mobile Phone Data to Trace Mobility in Finland. ISPRS International Journal of Geo-Information, 10(2):103, 2021. doi: https://doi.org/10.3390/ijgi10020103.

Peter Wills and François G. Meyer. Metrics for graph comparison: A practitioner's guide. PLOS ONE, 15(2), 2020. doi: https://doi.org/10.1371/journal.pone.0228728.

World Bank. Atlas of Sustainable Development Goals 2018: World Development Indicators (English). Washington, D.C., 2018. Available at: https://doi.org/10.1596/978-1-4648-1250-7 (Accessed: 17 May 2024).

World Tourism Organization. Methodological Notes to the Tourism Statistics Database, 2023 Edition. UNWTO, Madrid, 2023. Available at: https://doi.org/10.18111/9789284424160 (Accessed: 18 August 2024).

Tong Tong Wu and Kenneth Lange. Coordinate descent algorithms for lasso penalized regression. Annals of Applied Statistics, 2(1):224 – 244, 2008. doi: https://doi.org/10.1214/07-AOAS147.

Zhongwei Zhang, Elias Krainski, Peng Zhong, Harvard Rue, and Raphaël Huser. Joint modeling and prediction of massive spatio-temporal wildfire count and burnt area data with the INLA-SPDE approach. Extremes, 26(2):339–351, 2023.

Elise F. Zipkin and Sarah P. Saunders. Synthesizing multiple data types for biological conservation using integrated population models. Biological Conservation, 217:240–250, 2018. doi: https://doi.org/10.1016/j.biocon.2017.10.017.

Elise F. Zipkin, Brian D. Inouye, and Steven R. Beissinger. Innovations in data integration for modeling populations. Ecology, 100(6):1–3, 2019. URL https://www.jstor.org/stable/26675541.

Elise F. Zipkin, Jeffrey W. Doser, Courtney L. Davis, Wendy Leuenberger, Samuel Ayebare, and Kayla L. Davis. Integrated community models: A framework combining multispecies data sources to estimate the status, trends and dynamics of biodiversity. Journal of Animal Ecology, 92(12):2248–2262, 2023. doi: https://doi.org/10.1111/1365-2656.14012.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2):301–320, 2005. doi: https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Alain F. Zuur, Elena N. Ieno, Neil Walker, Anatoly A. Saveliev, and Graham M. Smith. Mixed Effects Models and Extensions in Ecology with R. In M. Gail, K. Krickeberg, J. M. Samet, A. Tsiatis, and W. Wong, editors, Statistics for Biology and Health, chapter 11, pages 261–294. Springer, New York, 2009. doi: https://doi.org/10.1007/978-0-387-87458-6.

Appendix A

Understanding factors driving recreational sea angling catch rates – Supplementary material

A.1 Data

A.1.1 Data cleaning

Interviews removed during the data cleaning process due to missing data or inconsistencies were:

- The 113 interviews with status 'refused' were removed.
- Interviews with missing data for any of the following variables were removed: fishing method (14 interviews), hours fished (23 interviews), days shore angling (14 interviews), sea state (65 interviews).
- There were 25 interviews where there was no record of any catches, but the observed or unobserved catch field indicated there had been a catch. As a result of this conflict, these 25 interviews were removed.
- Two multiday trips of '70' and '48' in hours fished and hours still to fish respectively were removed as these variables represent the duration of a trip and, as a trip represents the fishing activities in one day, the maximum possible value is 24.

Other changes made to the data were:

- At some sites, sea state was recorded as a range of two Beaufort scale readings (for example '1-2' or '5-6'). There were 23 interviews at these sites, and the sea state for these interviews has been set as the lower value in the range.

The species in the Skates/Rays group are: blonde ray, common skate, ray, skate/ray (unknown), smalleyed (painted) ray, spotted ray (homelyn), thornback ray (roker), and white skate.

A.1.2 Covariates

Tables A.1 and A.2 show descriptive statistics for the trips and environment covariates, split by shore anglers and boat anglers.

Figures A.1 to A.9 illustrate the average catch rate for each of the eight modelled species, by covariate, for each of the categorical covariates. Figure A.10 illustrates the correlation between the

numeric covariates. Figures A.11 to A.15 illustrate the relationship between each numeric covariate and catch size of the eight modelled species. The number of hours fished in daylight, twilight and darkness are not included, as the number of fish caught is not categorised by these time periods. To enhance visual clarity, the catch size is capped at 50. This results in the exclusion of four data points that exceed this threshold. These represent trips where 120 mackerel, 100 whiting, 79 dab, and 58 whiting were caught.

Overall, catch rates for the categorical covariates are generally higher for mackerel and whiting than for the other species. Catch rates are higher for most species when fishing from a boat than from shore. Only mackerel has a notable difference in catch rates between different fishing methods. Catch rates between weekdays and the weekend vary by species. Cod, mackerel, and whiting ave higher catch rates at weekends, while flounder and dab have higher catch rates on weekdays. There is a notable difference in the catch rates of mackerel between competitive and non-competitive fishing, with differences also apparent for most other species. Catch rates for bass, plaice, whiting, and dab are substantially higher among anglers who are members of an angling club, compared to anglers who are not angling club members. In contrast, catch rates for mackerel are lower among angling club members.

Cod, bass and flounder have higher catch rates in the Irish Sea, while plaice has higher catch rates in the English Channel, compared to other regions. Mackerel has higher catch rates in the English Channel and North Sea, compared to other regions. Of all species, whiting has the greatest difference in catch rates between regions, with its highest rates in the Irish Sea and North Sea. Catch rates for dab are lowest in the Bristol Channel/Irish Sea, and fairly similar across the other regions. Skates/rays have similar catch rates across all regions.

The most notable differences in catch rates across seasons are for mackerel, for which catch rates are highest in summer, and whiting, for which catch rates are highest in autumn and winter. For cod and bass, catch rates are higher in summer and autumn, than in winter and spring. For Flounder and skates/rays, catch rates are higher in autumn than in other seasons. For plaice, catch rates are higher in summer than in other season, while dab has fairly consistent catch rates across the year.

The calmer the sea conditions, the higher the catch rates of mackerel. In contrast, the highest catch rates for whiting are in rough sea conditions. Catch rates for dab are lower in moderate sea

conditions, compared to calm or rough sea conditions. Catch rates for the other species do not vary substantially under different sea conditions.

Catch rates for mackerel are high in areas with sand or rock sediment, moderate in areas with gravel sediment, and low in mud sediment areas. Catch rates for whiting and dab are higher in sand and gravel sediment areas, than in rock and mud sediment areas. In contrast, catch rates for cod are higher in mud sediment areas than in areas with other sediment types. Catch rates for the other species vary less across areas with different sediment types.

There is moderate correlation between the number of hours fished in twilight and the number of hours fished in darkness. Correlations between the other numeric covariates are relatively low.

For all species, there are no clear trends in the relationship between the catch size of the modelled species and any of the numeric covariates.

Table A.1 Number (and percentage) of trips, by type of fishing

		Shore	Boat	Total
Trip				
Method	Bait	958 (92%)	86 (8%)	1044 (100%)
	Lure/fly/jig	107 (71%)	44 (29%)	151 (100%)
	Mix	78 (61%)	50 (39%)	128 (100%)
Weekend	No	556 (92%)	50 (8%)	606 (100%)
	Yes	587 (82%)	130 (18%)	717 (100%)
Competition	No	1045 (86%)	164 (14%)	1209 (100%)
	Yes	98 (86%)	16 (14%)	114 (100%)
Club	No	923 (89%)	115 (11%)	1038 (100%)
	Yes	220 (77%)	65 (23%)	285 (100%)
Environment				
Season	Autumn	232 (89%)	28 (11%)	260 (100%)
	Spring	291 (97%)	8 (3%)	299 (100%)
	Summer	324 (72%)	128 (28%)	452 (100%)
	Winter	296 (95%)	16 (5%)	312 (100%)
Sea	Bristol Channel/Celtic Sea	107 (98%)	2 (2%)	109 (100%)
	English Channel	440 (80%)	111 (20%)	551 (100%)
	Irish Sea	114 (88%)	16 (12%)	130 (100%)
	North Sea	482 (90%)	51 (10%)	533 (100%)
Sea state	Calm	594 (87%)	90 (13%)	684 (100%)
	Moderate	437 (85%)	77 (15%)	514 (100%)
	Rough	112 (90%)	13 (10%)	125 (100%)
Lunar phase	Full	277 (91%)	28 (9%)	305 (100%)
	New	304 (82%)	68 (18%)	372 (100%)
	Waning	223 (86%)	36 (14%)	259 (100%)
	Waxing	339 (88%)	48 (12%)	387 (100%)
Sediment	Gravel	241 (90%)	26 (10%)	267 (100%)
	Mud	115 (94%)	7 (6%)	122 (100%)
	Rock	29 (62%)	18 (38%)	47 (100%)
	Sand	758 (85%)	129 (15%)	887 (100%)
Total		1143 (86%)	180 (14%)	1323 (100%)

Table A.2 Mean of numeric covariates, by type of fishing. The means for shore fishing are based on 1,143 anglers and the means for boat fishing are based on 180 anglers.

Mean	Shore	Boat
Angler characteristics		
Days shore angling	12.8	2.3
Days boat angling	0.4	8.2
Trip characteristics		
Hours fished in daylight	2.0	4.2
Hours fished in twilight	0.1	0.03
Hours fished in darkness	0.2	0.03
Total anglers in group	1.6	2.2

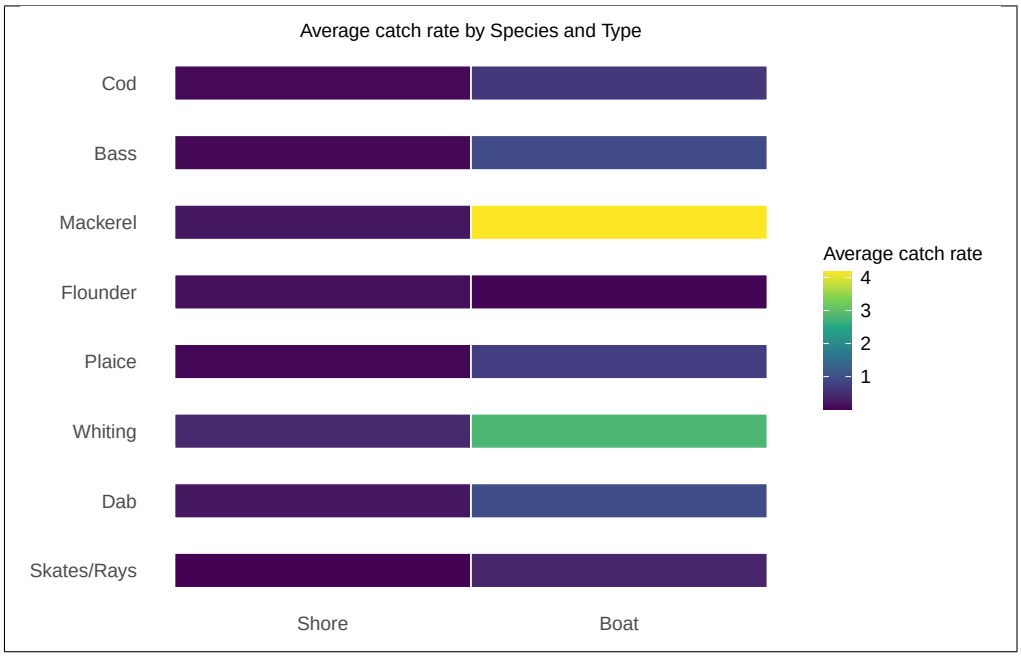


Fig. A.1 Average catch rate, by species and type of fishing.

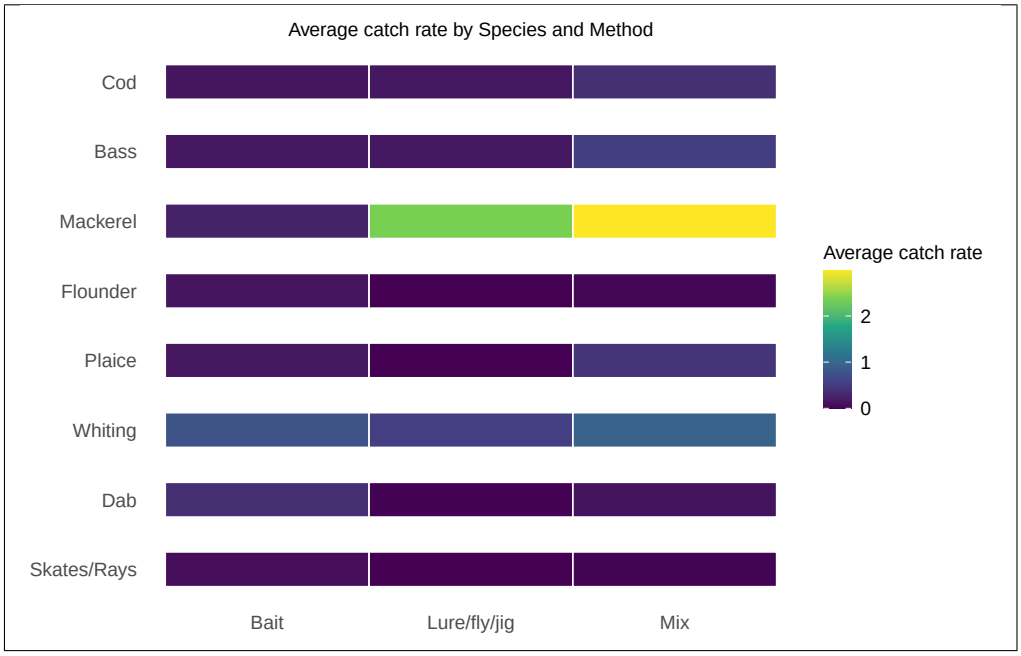


Fig. A.2 Average catch rate, by species and fishing method.

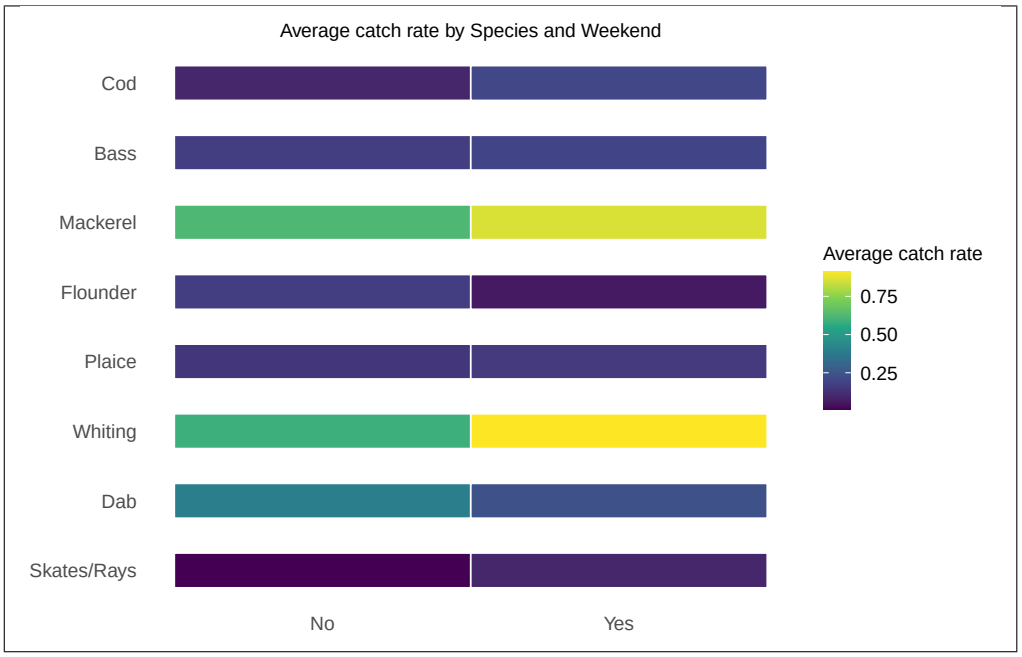


Fig. A.3 Average catch rate, by species and weekend.

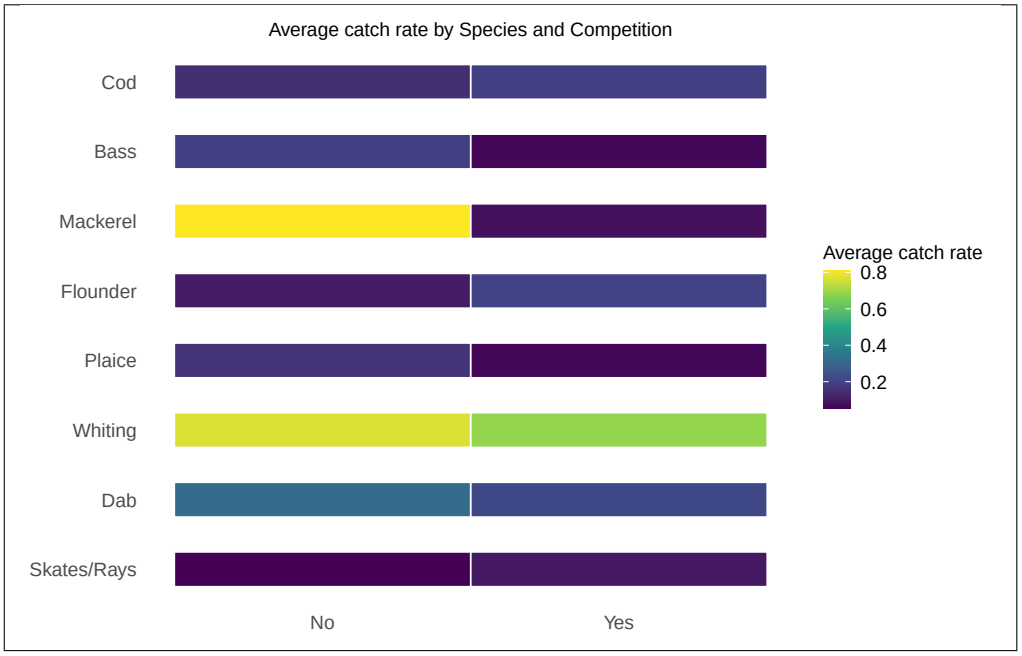


Fig. A.4 Average catch rate, by species and competitive fishing.

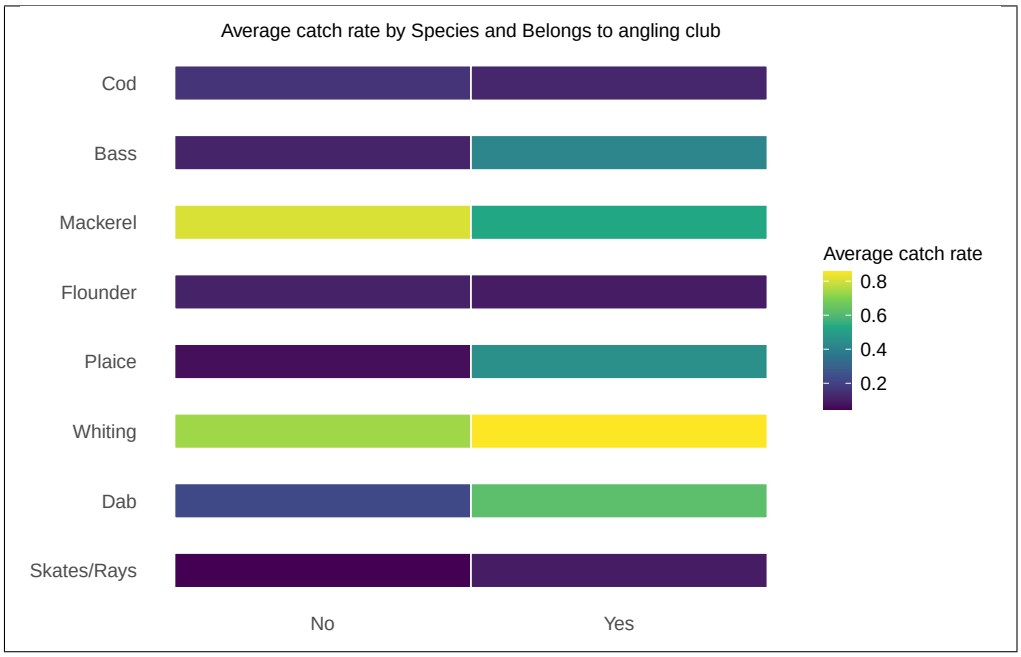


Fig. A.5 Average catch rate, by species and angling club membership.

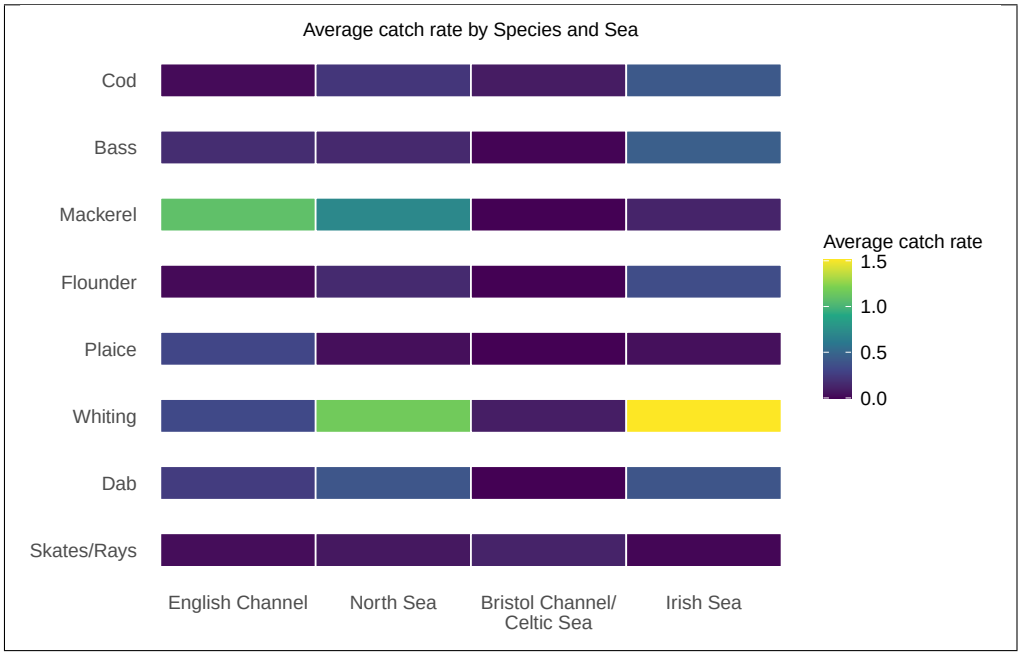


Fig. A.6 Average catch rate, by species and sea.

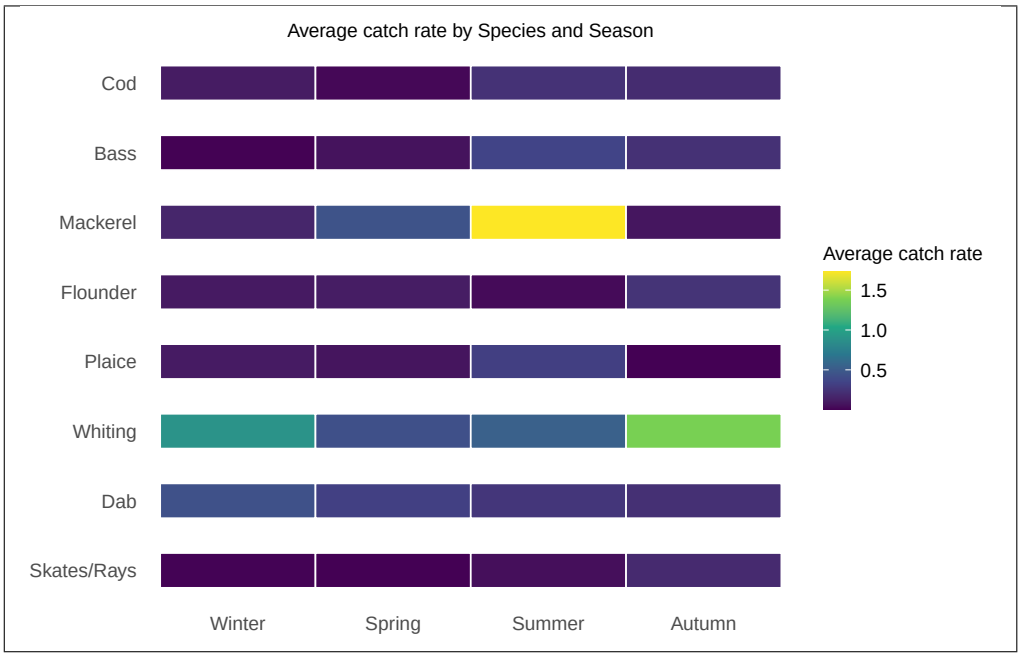


Fig. A.7 Average catch rate, by species and season.

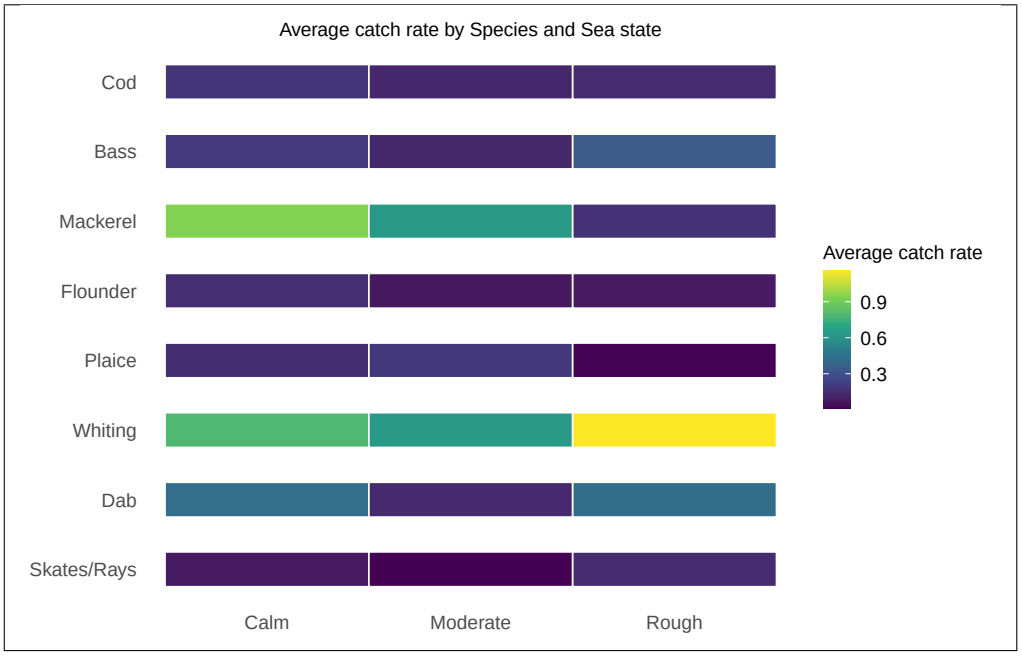


Fig. A.8 Average catch rate, by species and sea state.

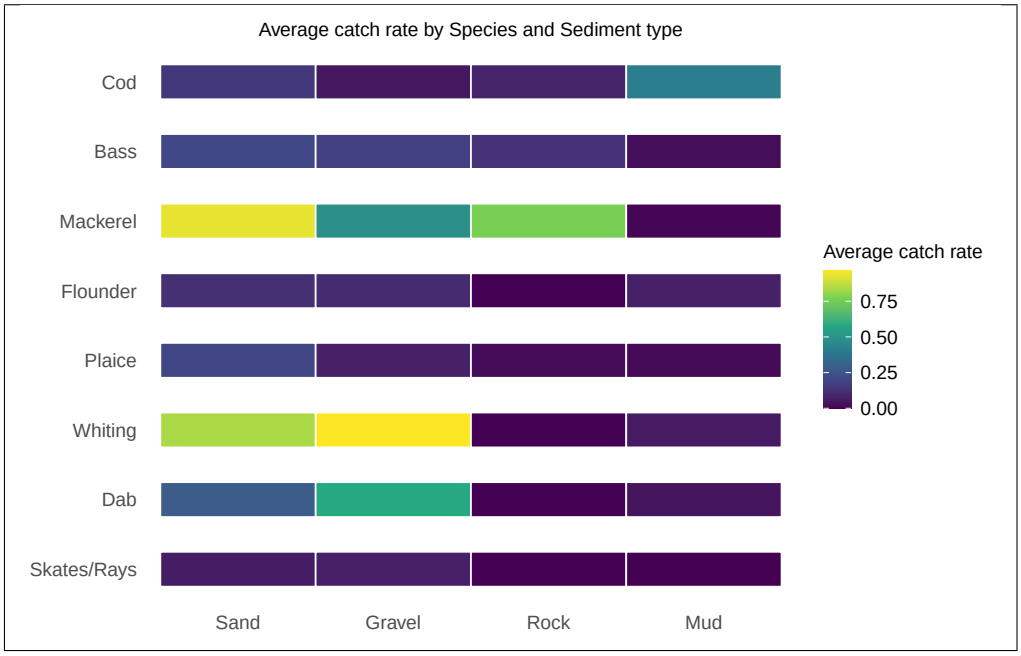


Fig. A.9 Average catch rate, by species and sediment type.

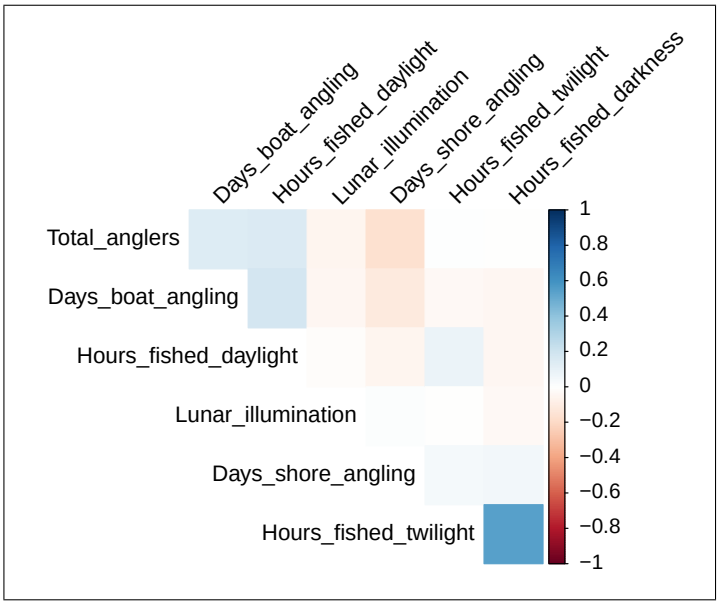


Fig. A.10 Correlations between numeric covariates.

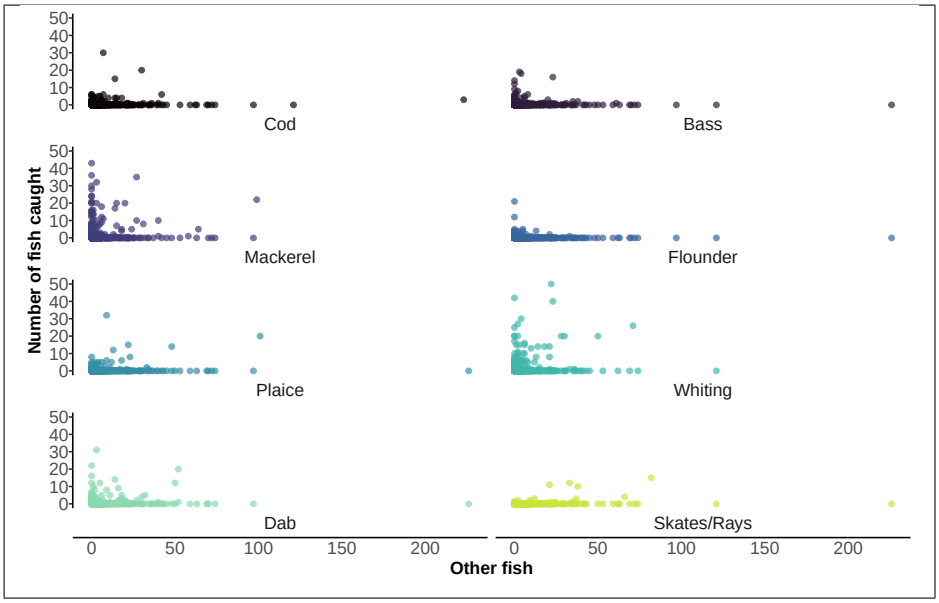


Fig. A.11 Relationship between the number of fish of other species caught and catch size (of the modelled species). The plot shows the number of fish of the modelled species caught (y-axis) and the number of fish of other species caught (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting.

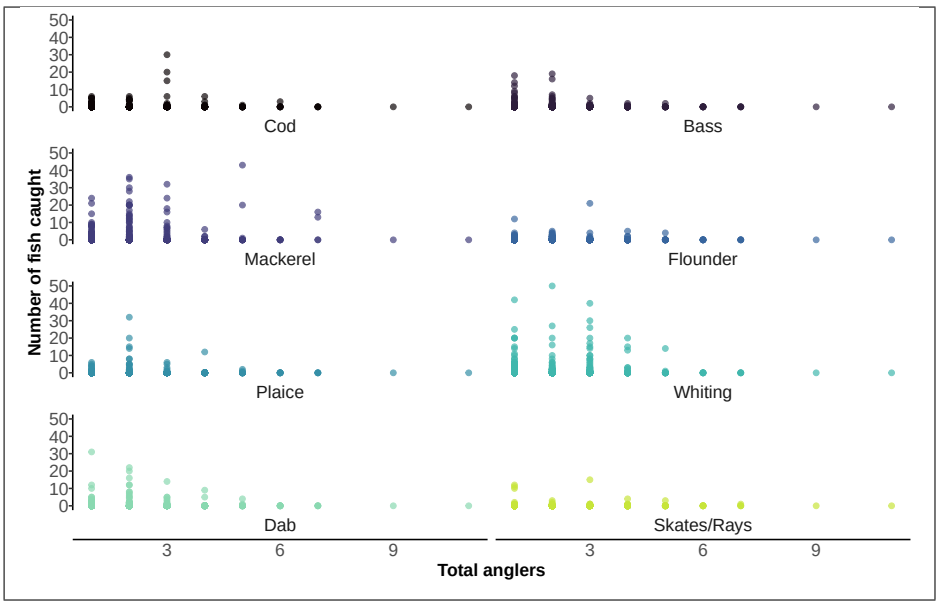


Fig. A.12 Relationship between angling group size and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of anglers in the angling group (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting.

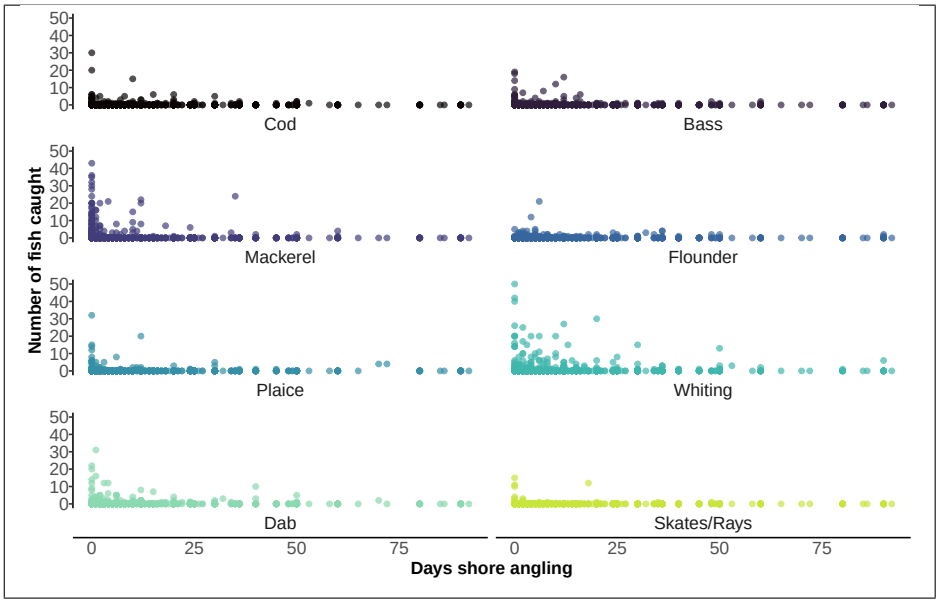


Fig. A.13 Relationship between shore angling experience and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of days shore angling in the past three months (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting.

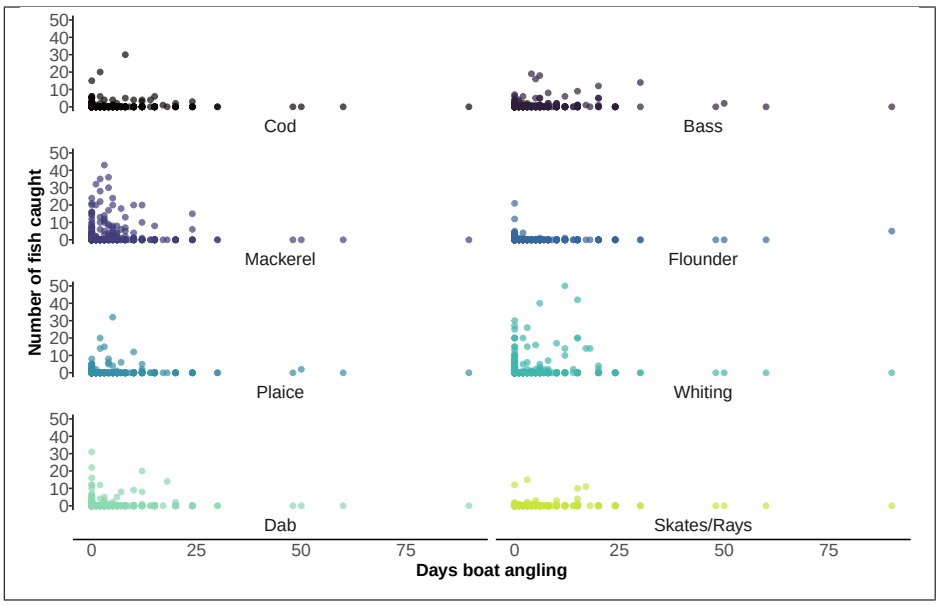


Fig. A.14 Relationship between boat angling experience and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the number of days boat angling in the past three months (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting.

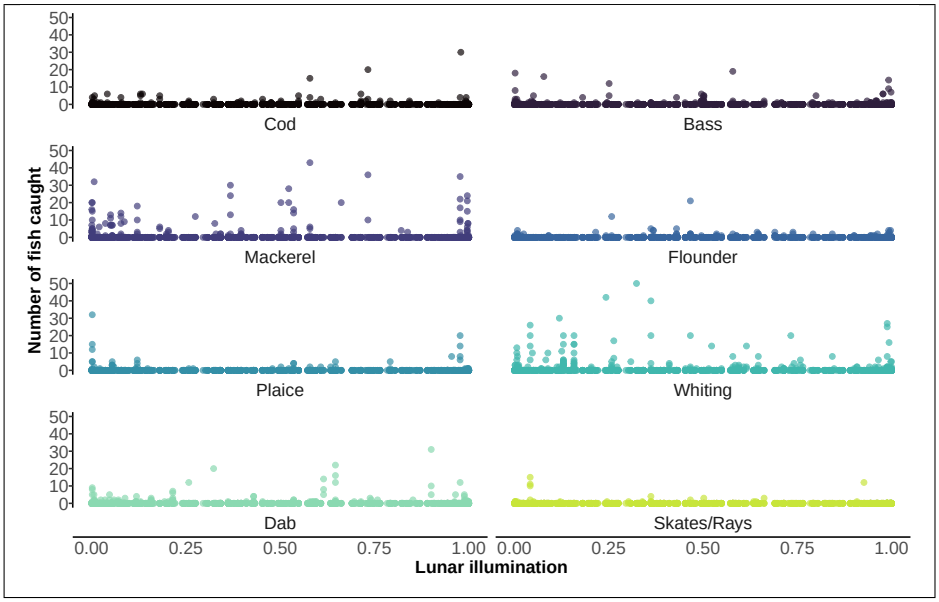


Fig. A.15 Relationship between lunar illumination and catch size. The plot shows the number of fish of the modelled species caught (y-axis) and the lunar illumination on the day of the fishing trip (x-axis). Each point represents a single fishing trip. The y-axis is capped at 50, resulting in the exclusion of four data points: one for mackerel, one for dab, and two for whiting.

A.2 Simulations

A.2.1 Setting up the simulations

Simulation grids

Initially, we tried 10 folds as this is more computationally efficient. However, substantially different values of the cross-validated log-likelihood were obtained across different runs of 10-fold cross-validation (see Figure A.16), which made it impossible to identify the optimal pair of λ values.

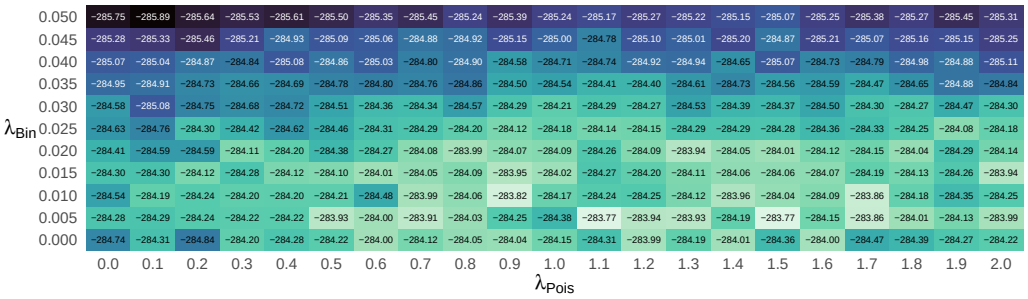
Therefore, we use n folds ('leave-one-out cross validation'). The advantage of n folds in this case is that there is no randomness in the cross-validation process and hence there is no risk for the noise to dominate over the signal.

The Binomial probability is given by

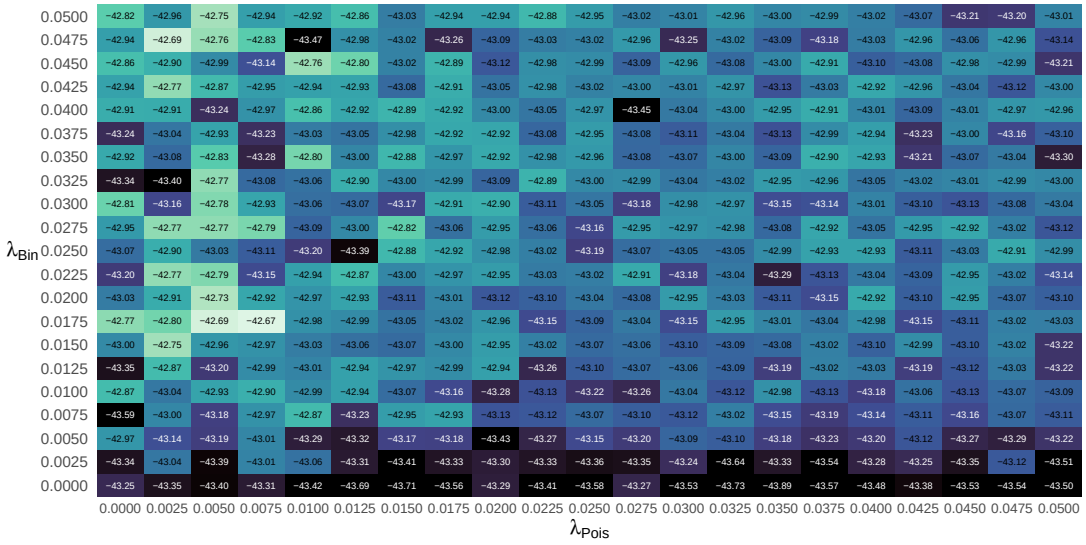
$$p = \frac{e^{(-1+2\cdot \mathrm{X2})}}{1+e^{(-1+2\cdot \mathrm{X2})}}$$

and the Poisson mean given by

$$\lambda = e^{(0.4+1\cdot \mathrm{X1})}$$



(a) Y values simulated using p and λ as a function of covariates.



(b) Y values simulated using constant p and λ.

Fig. A.16 Loglikelihood values in simulations with 10 folds.

R code

The R code used to generate the simulated data is provided in Figure A.17.

```
n = 1000
X1 <- sample(seq(-3, 3, by = 0.01), n, replace = TRUE)
X2 <- sample(seq(-3, 3, by = 0.01), n, replace = TRUE)

# when p and L are constant
p <- 0.5
lambda <- 6

# when p and L are a function of covariates
p <- exp(-1 + 2 * X2)/(1 + exp(-1 + 2 * X2))
lambda <- exp(0.4 + 1 * X1)

y <- ifelse(rbinom(n, size = 1, prob = p) == 1,
   0,
   rpois(n, lambda = lambda))
dataForSimulation <- data.frame(y, X1, X2)
colnames(dataForSimulation) <- c("y", "X1", "X2")
```

Fig. A.17 R code to generate the simulated data.

We fitted zero-inflated Poisson models with X1 and X2 included in both the Binomial and Poisson parts of the model to the simulated *y* values using LASSO penalised regression, using the zipath() function from the *mpath* package (Wang, 2020). The R code used to fit the models with the simulated data is provided in Figure A.18.

```
mod.zip <- zipath(y ~ X1 + X2 | X1 + X2,
   data = dataForSimulation,
   family = "poisson",
   nlambda = 1,
   lambda.count = seq(from = lambdaCountValues[i],
   to = lambdaCountValues[i],
   length.out = SeqLength.Poisson),
   lambda.zero = seq(from = lambdaZeroValues[j],
   to = lambdaZeroValues[j],
   length.out = SeqLength.Binomial))

mod.cvzip <- cv.zipath(y ~ X1 + X2 | X1 + X2,
   data = dataForSimulation,
   family = "poisson",
   nlambda = 1,
   nfolds = n,
   lambda.count = seq(from = lambdaCountValues[i],
   to = lambdaCountValues[i],
   length.out = SeqLength.Poisson),
   lambda.zero = seq(from = lambdaZeroValues[j],
   to = lambdaZeroValues[j],
   length.out = SeqLength.Binomial))

# 'from' = 'to' as each sequence contained one value only
```

Fig. A.18 R code to run simulation models.

A.2.2 Simulation results

The results of the simulations with constant parameters in both the Binomial and Poisson parts, for n and 10 folds are given in Tables A.3 and A.4 respectively.

Table A.3 Simulation results with constant parameters for the Binomial and Poisson distributions ($p = 0.5$ and $\lambda = 6$ respectively), using n folds.

	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
lambda.count	0.030842	0.032150	0.061046	0.009599	0.000154	0.009599	0.000154	0.011674	0.061956	0.030651
lambda.zero	0.024681	0.021364	0.016939	0.009209	0.006714	0.009209	0.006714	0.007703	0.022520	0.011047
(Intercept) (Count)	1.779681	1.778982	1.799319	1.806936	1.766910	1.806936	1.766910	1.813755	1.785502	1.803405
X1 (Count)	0	0	0	0.017742	−0.013060	0.017742	−0.013060	0	0	0
X2 (Count)	0	0	0	−0.012410	0.015223	−0.012410	0.015223	0	0	0
(Intercept) (Zero)	−0.069540	0.070869	−0.008760	0.123361	0.044157	0.123361	0.044157	−0.048450	−0.001150	0.003387
X1 (Zero)	0	0	0	−0.015710	0	−0.015710	0	0	0	0
X2 (Zero)	0	0	0	0	−0.056650	0	−0.056650	0	0	0

Table A.4 Simulation results with constant parameters for the Binomial and Poisson distributions ($p = 0.5$ and $\lambda = 6$ respectively), using 10 folds.

	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
lambda.count	0.030842	0.032150	0.061046	0.015285	0.000017	0.089523	0.001436	0.011674	0.061956	0.030651
lambda.zero	0.024681	0.021364	0.016939	0.010344	0.003842	0.016092	0.011734	0.007703	0.022520	0.011047
(Intercept) (Count)	1.779681	1.778982	1.799319	1.807088	1.766914	1.808212	1.766958	1.813755	1.785502	1.803405
X1 (Count)	0	0	0	0.016450	−0.013070	0	−0.012760	0	0	0
X2 (Count)	0	0	0	−0.011160	0.015232	0	0.014979	0	0	0
(Intercept) (Zero)	−0.069540	0.070869	−0.008760	0.123452	0.044078	0.123948	0.044580	−0.048450	−0.001150	0.003387
X1 (Zero)	0	0	0	−0.013080	0.005112	0	0	0	0	0
X2 (Zero)	0	0	0	0	−0.063650	0	−0.044900	0	0	0

The results of the simulations with the parameters in the Binomial and Poisson distributions as a function of the covariates for n and 10 folds are given in Tables A.5 and A.6 respectively.

Table A.5 Simulation results with parameters for the Binomial and Poisson distributions as a function of covariates, using n folds.

	Run1	Run2	Run3	Run4	Run5
lambda.count	0.00016	0.00018	0.00016	0.00017	0.00017
lambda.zero	0.02367	0.02226	0.02404	0.02304	0.02406
(Intercept) (Count)	0.41521	0.38874	0.40499	0.44183	0.46796
X1 (Count)	0.99938	1.01324	0.98098	0.98735	1.00323
X2 (Count)	0.02131	0.00002	−0.01583	0.02783	0.01519
(Intercept) (Zero)	−0.65962	−0.57045	−0.57803	−0.47553	−0.63337
X1 (Zero)	0	0	0	0	0
X2 (Zero)	1.33486	1.18199	1.28603	1.17975	1.38654

Table A.6 Simulation results with parameters for the Binomial and Poisson distributions as a function of covariates, using 10 folds.

	Run1	Run2	Run3	Run4	Run5
lambda.count	0.00016	0.00018	0.00016	0.00017	0.00017
lambda.zero	0.02367	0.02226	0.02404	0.02304	0.02406
(Intercept) (Count)	0.41521	0.38874	0.40499	0.44183	0.46796
X1 (Count)	0.99938	1.01324	0.98098	0.98735	1.00323
X2 (Count)	0.02131	0.00002	−0.01583	0.02783	0.01519
(Intercept) (Zero)	−0.65960	−0.57045	−0.57803	−0.47553	−0.63337
X1 (Zero)	0	0	0	0	0
X2 (Zero)	1.33486	1.18199	1.28603	1.17975	1.38654

A.3 Grid method

A.3.1 Fitting the models

Grids with 10 folds When using 10 folds, the cross-validated log-likelihood values varied across the grids, which made it impossible to identify where to zoom in to approach the optimal pair of λ values. Examples of two grids obtained using 10 folds with the same data are presented in Figure A.19.

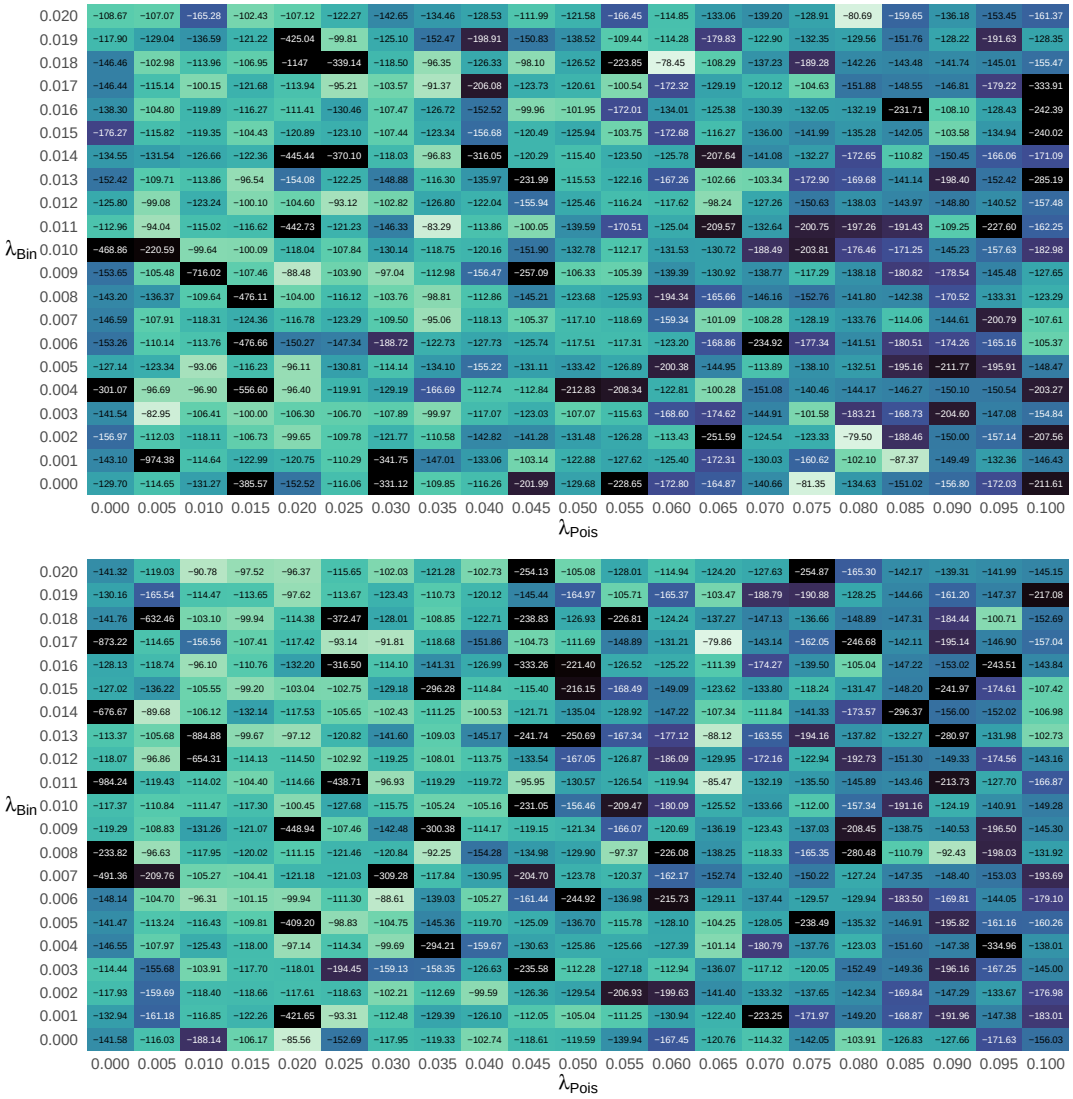
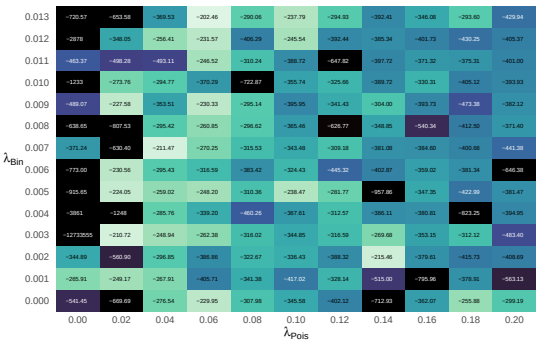
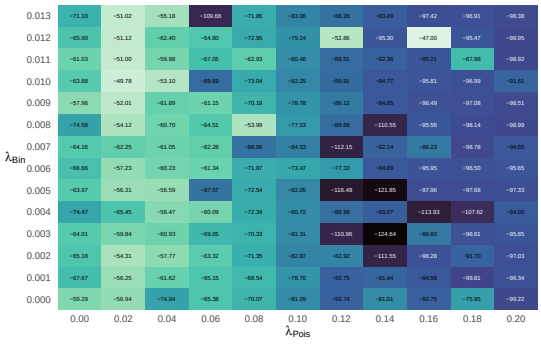


Fig. A.19 Loglikelihood values with 10 folds.

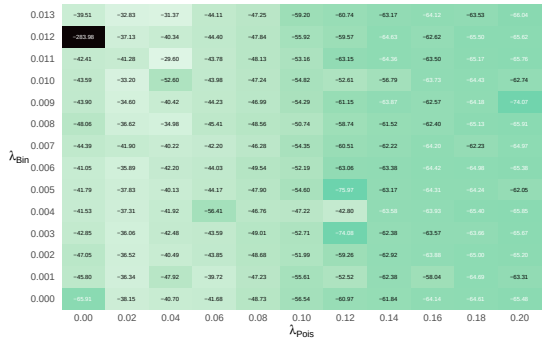
Grids with other numbers of folds As the number of folds increases, the region of highest loglikelihood becomes more distinct, although it remains difficult to identify an area for the next zoom level. Therefore, using n folds is preferable due to its stability. Examples of the grids obtained using different numbers of folds demonstrating this are presented in Figure A.20.



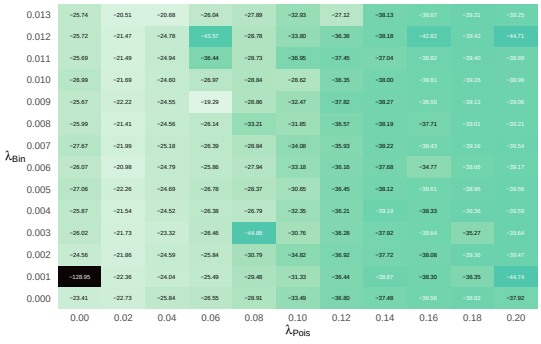
(a) 5 folds.



(b) 20 folds.



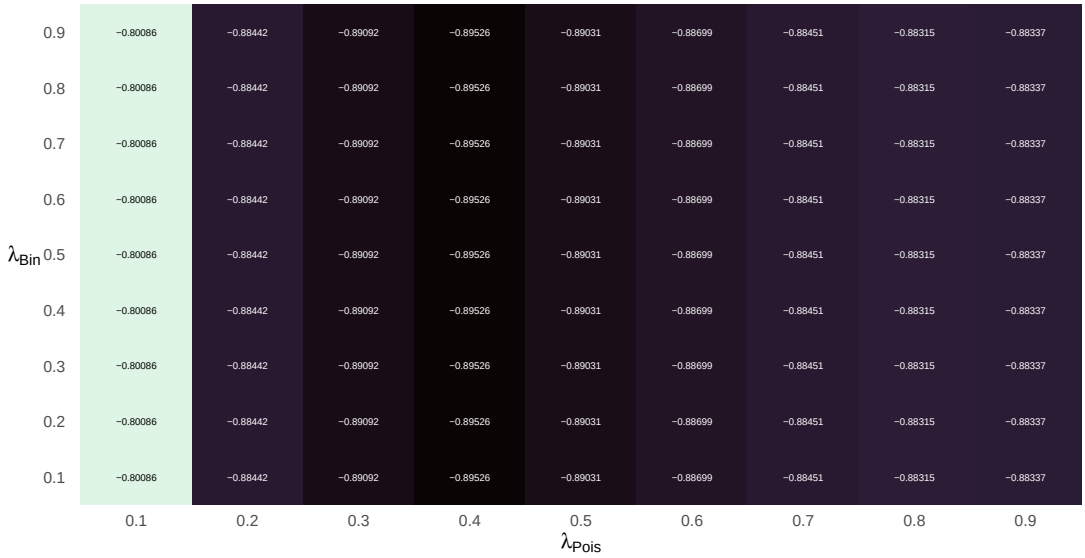
(c) 30 folds.



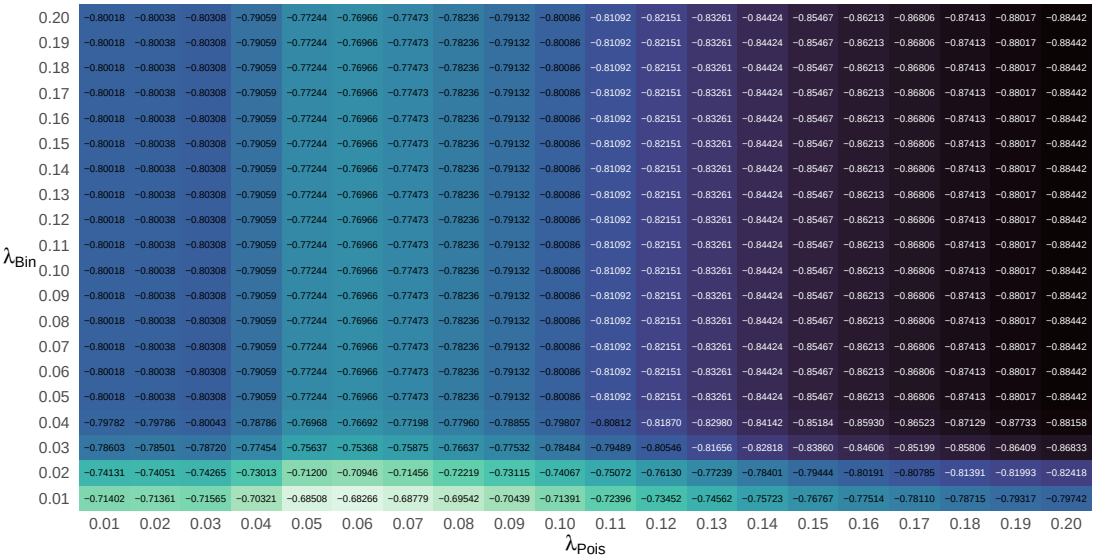
(d) 50 folds.

Fig. A.20 Loglikelihood values with different numbers of folds.

Grids with n folds Figures A.21 and A.22 demonstrate examples of the grids of loglikelihood values for each pair of λ values at each zoom level, for mackerel. Within each grid, a colour scale of white for the highest (best) loglikelihood values through to black for the lowest loglikelihood values is used.

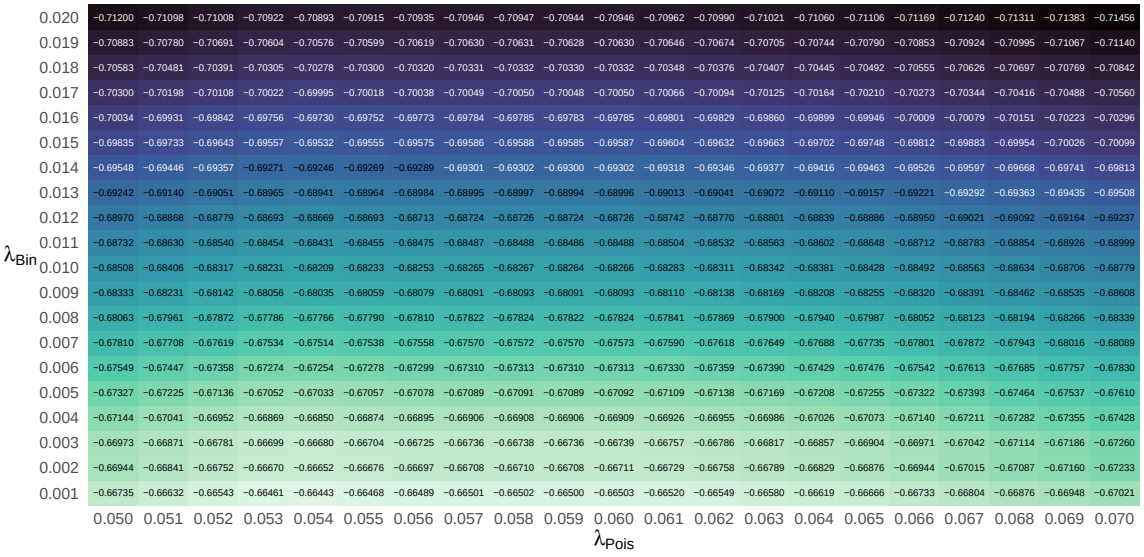


(a) Zoom level 0.

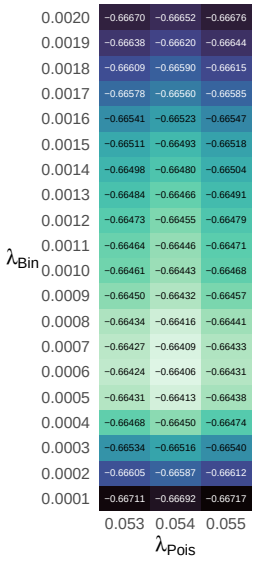


(b) Zoom level 1.

Fig. A.21 Loglikelihood values for pairs of λ values used in the grid method for mackerel. Zoom levels: 0 and 1. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis.



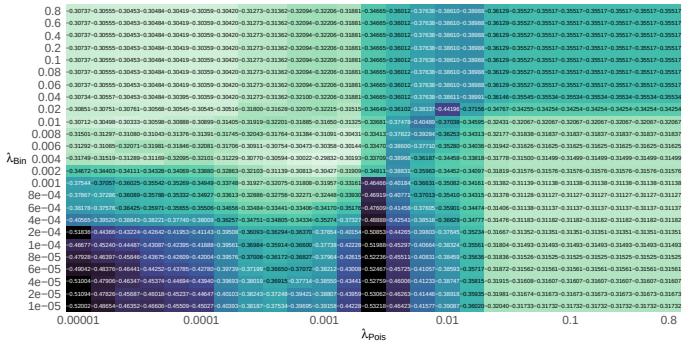
(a) Zoom level 2.



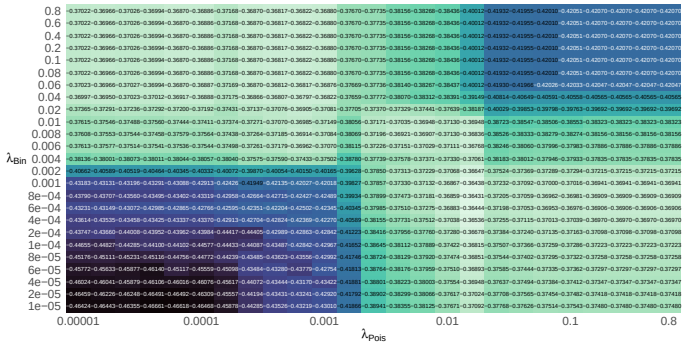
(b) Zoom level 3.

Fig. A.22 Loglikelihood values for pairs of λ values used in the grid method for mackerel. Zoom levels: 2 and 3. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis.

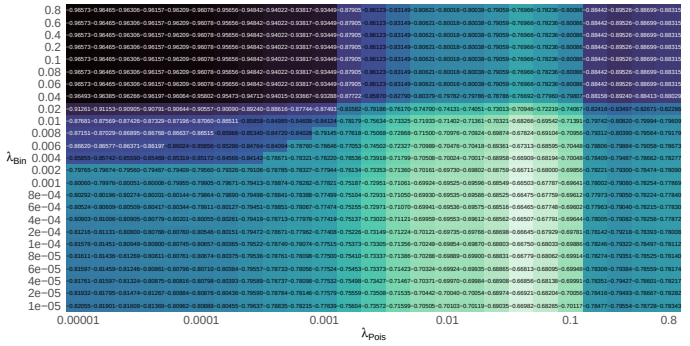
Checks To check that the maximum loglikelihood found by zooming in appeared to be in the best region, loglikelihood values for a wide range of λ values were calculated for each species. Figures A.23 (cod, sea bass, mackerel, flounder) and A.24 (plaice, whiting, dab, skates/rays) show the resulting grids. Within each grid, a colour scale of white for the highest (best) loglikelihood values through to black for the lowest loglikelihood values is used.



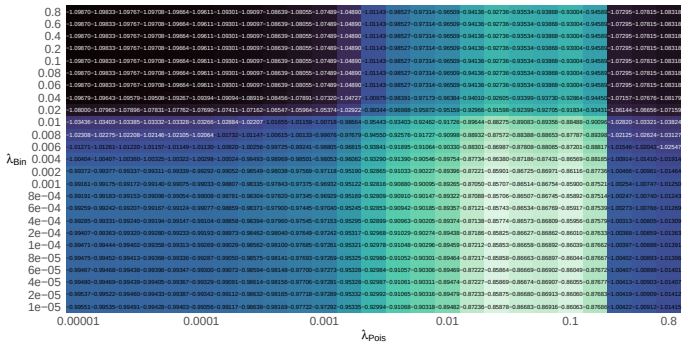
(a) Cod.



(b) Sea bass.

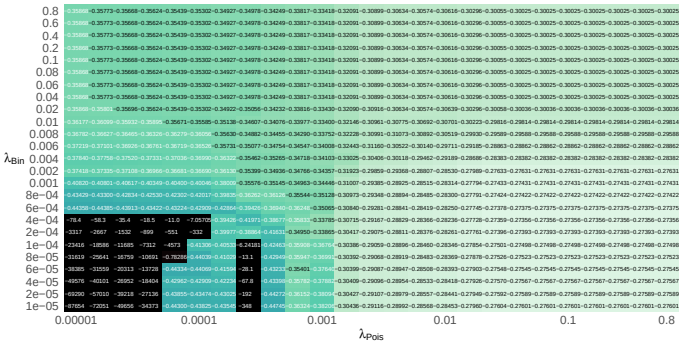


(c) Mackerel.

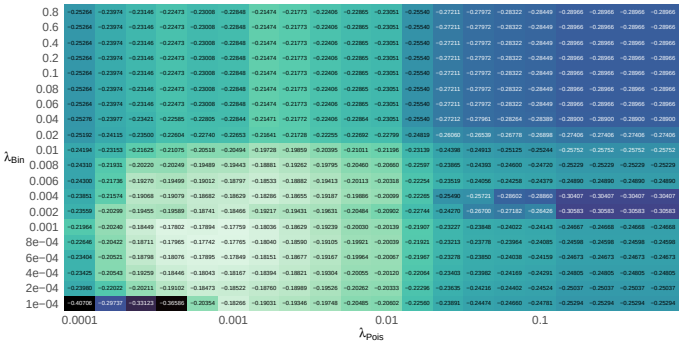


(d) Whiting.

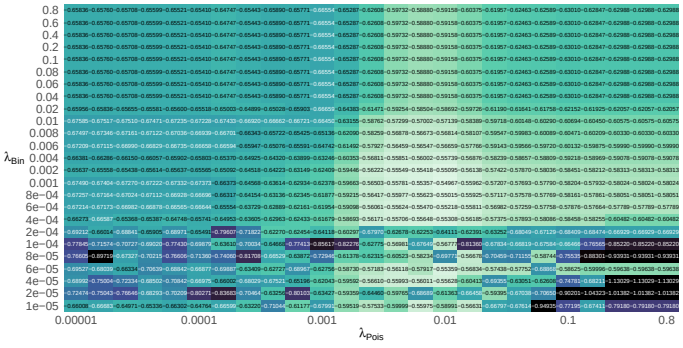
Fig. A.23 Grids using a wide range of λ values. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. Species shown: cod, sea bass, mackerel, and whiting.



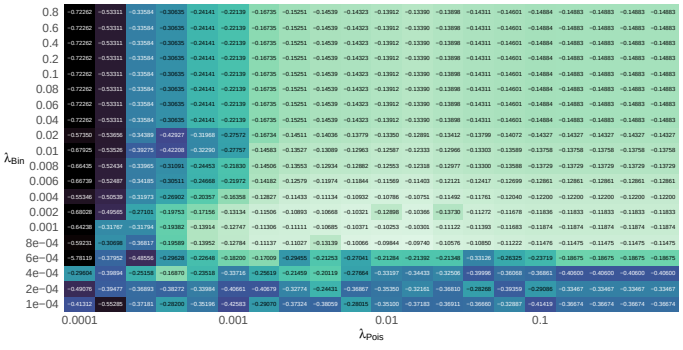
(a) Flounder.



(b) Plaice.



(c) Dab.



(d) Skates/rays.

Fig. A.24 Grids using a wide range of λ values. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis. Species shown: flounder, plaice, dab and skates/rays.

Computational time

The average time taken (in seconds) for each λ pair at each zoom level for each species using the grid method is shown in Table A.7a. The total time taken (in hours) for each zoom level is shown in Table A.7b.

Table A.7 Grid method durations. Number of λ pairs: zoom level 0 = 81, zoom level 1 = 400, zoom level 2 = 420, zoom level 3 = 420. Zoom level 3 was only used when at least one λ value at the maximum loglikelihood in zoom level 2 was 0.001 or less. There are no level 3 durations for skates/rays as the λ values corresponding to the highest loglikelihood at zoom level 2 were both greater than 0.001 for this species.

(a) Average λ pair duration (seconds).

	Zoom Level			
	0	1	2	3
Cod	27	40	42	106
Sea bass	27	45	91	195
Mackerel	32	47	55	103
Flounder	25	28	40	103
Plaice	26	39	263	1731
Whiting	32	47	50	85
Dab	27	30	79	1586
Skates/Rays	25	26	52	-

(b) Total duration (hours).

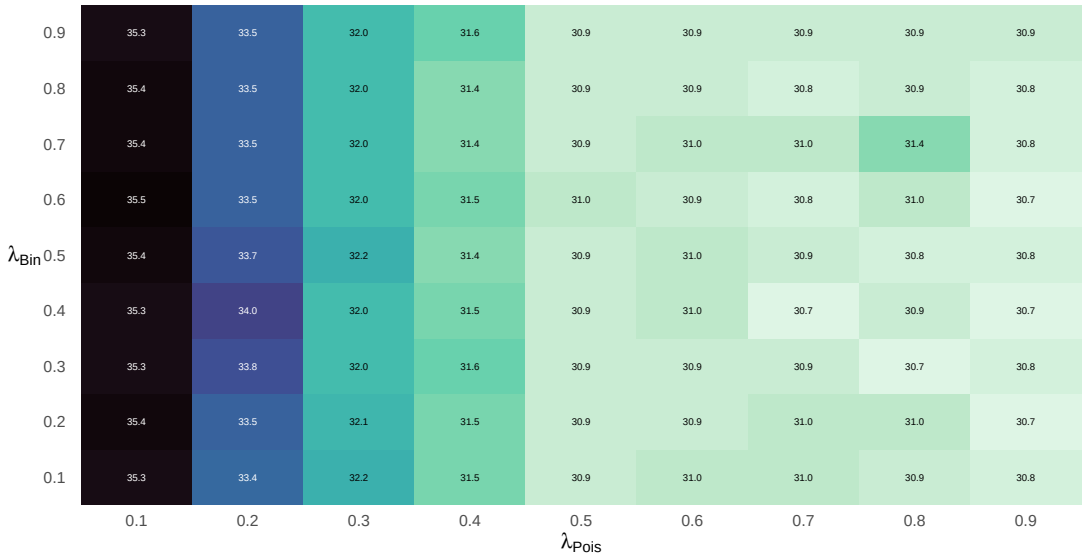
	Zoom Level				
	0	1	2	3	Total
Cod	0.6	4.5	5.0	12.3	22.4
Sea bass	0.6	5.0	10.7	22.8	39.1
Mackerel	0.7	5.2	6.4	12.0	24.3
Flounder	0.6	3.1	4.7	12.0	20.4
Plaice	0.6	4.3	30.6	202.0	237.5
Whiting	0.7	5.2	5.8	9.9	21.6
Dab	0.6	3.3	9.3	185.0	198.2
Skates/Rays	0.6	2.9	6.0	-	9.5

The average time taken to calculate the loglikelihood for each λ pair varied by species and increased as the zoom level increased and the λ values became closer to those producing higher loglikelihoods. In addition, differences in average calculation times between species widened as the zoom level increased. At zoom level 0, average durations for each λ pair were similar across the different species, ranging from 25 to 32 seconds. At zoom level 3, typical average durations were around 100 seconds, although the average duration for plaice and dab was over 1,000 seconds. The relative calculation time for a species at one zoom level does not necessarily predict its relative duration at other zoom levels. For example, whiting was among the species with the longest average calculation times at zoom level 0, yet had the shortest average duration at zoom level 3.

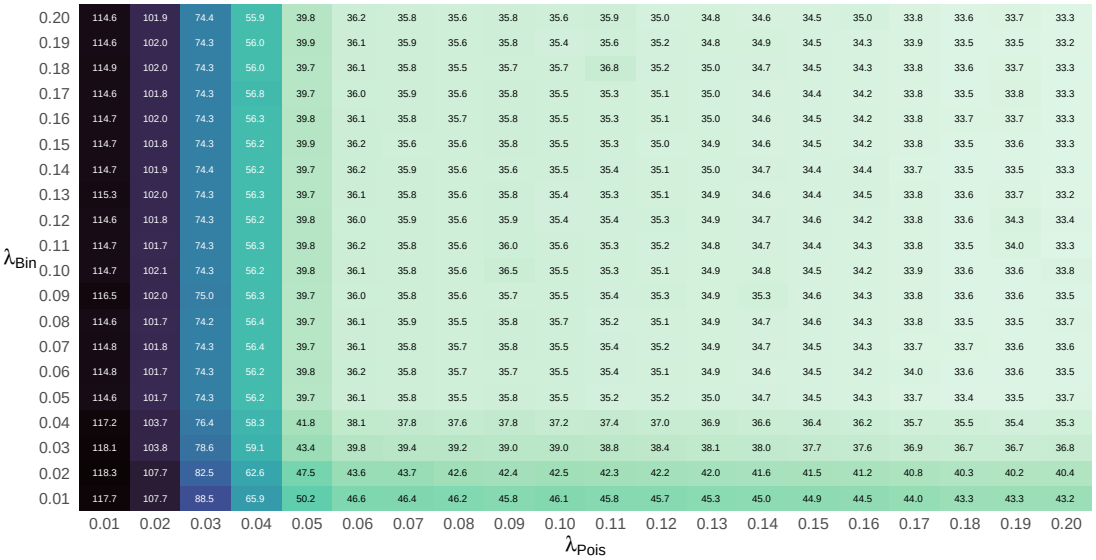
The total computational time depends on the number of λ pairs in the grids, which can be adjusted according to computational resources or specific research objectives. For example, zoom levels 0 to 2 were completed for all species, with further zooming in to level 3 implemented only when λ_{Bin} or λ_{Pois} corresponding to the highest loglikelihood at level 2 were 0.001 or less. The values presented in

Table A.7 are from a complete re-run of the fitting using a full grid of λ values at each zoom level. The re-run was performed using R version 4.4.1 (R Core Team, 2024) on Ubuntu 22.04.4 LTS in a shared computing environment. With this approach, typical total durations were around 20 hours, but ranged from 10 to over 200 hours. For the results given in Chapter 2, to reduce computational time, only three λ_{Pois} values were used at zoom level 3. However, this required re-running level 3 with an extended range of λ values if the highest loglikelihood was at the boundary of the grid, which occurred for bass, plaice, dab, and skates/rays. In addition, for cod, it was necessary to carry out further searching at level 3 to achieve a higher loglikelihood than obtained by the default method.

Figures A.25 and A.26 demonstrate the computational time (in seconds) for each pair of λ values at each zoom level in the re-run with full grids, for mackerel. Within each grid, a colour scale of white for the fastest durations through to black for the slowest durations is used.

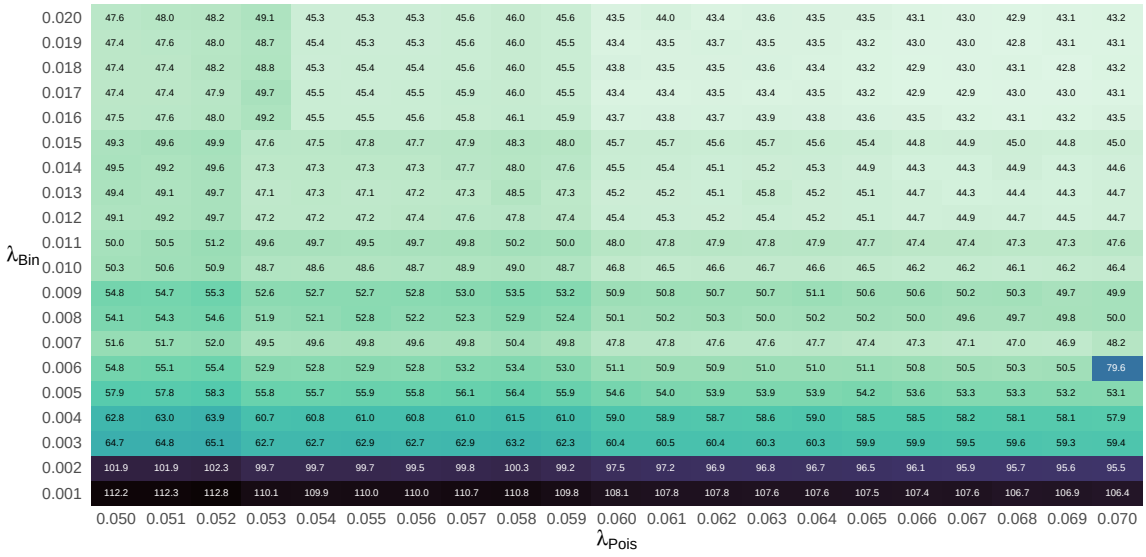


(a) Zoom level 0.

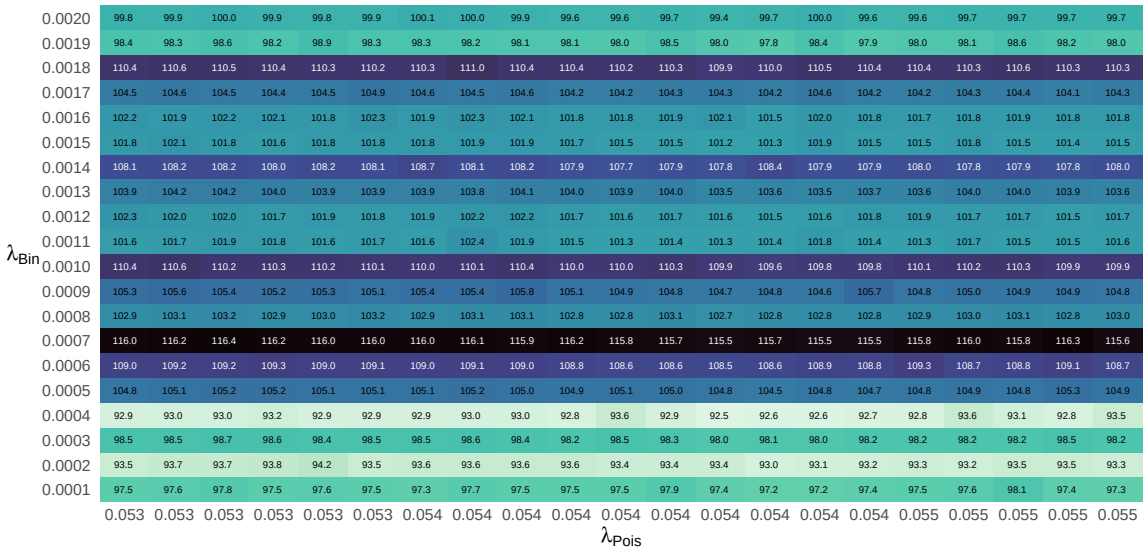


(b) Zoom level 1.

Fig. A.25 Time taken (seconds) for each pair of λ values used in the grid method for mackerel. Zoom levels: 0 and 1. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis.



(a) Zoom level 2.



(b) Zoom level 3.

Fig. A.26 Time taken (seconds) for each pair of λ values used in the grid method for mackerel. Zoom levels: 2 and 3. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis.

A.3.2 Results

Comparing observed and predicted catches

Figures A.27 (cod, sea bass, mackerel, flounder) and A.28 (plaice, whiting, dab, skates/rays) compare the recorded and predicted number of catches of the target species for each interview, using the grid method.

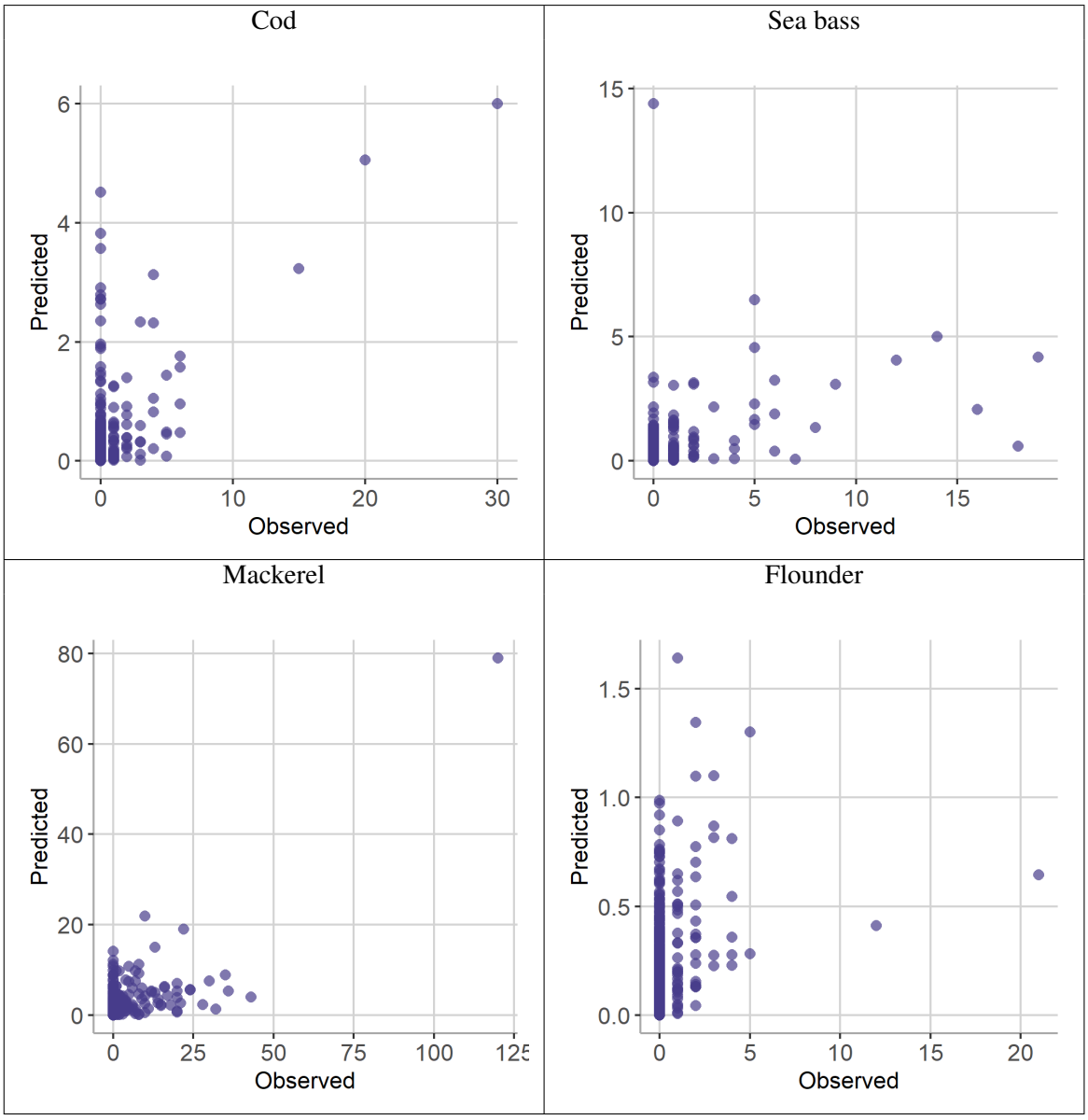


Fig. A.27 Number of observed and predicted catches using the grid method for each interview. Species shown: cod, sea bass, mackerel and flounder.

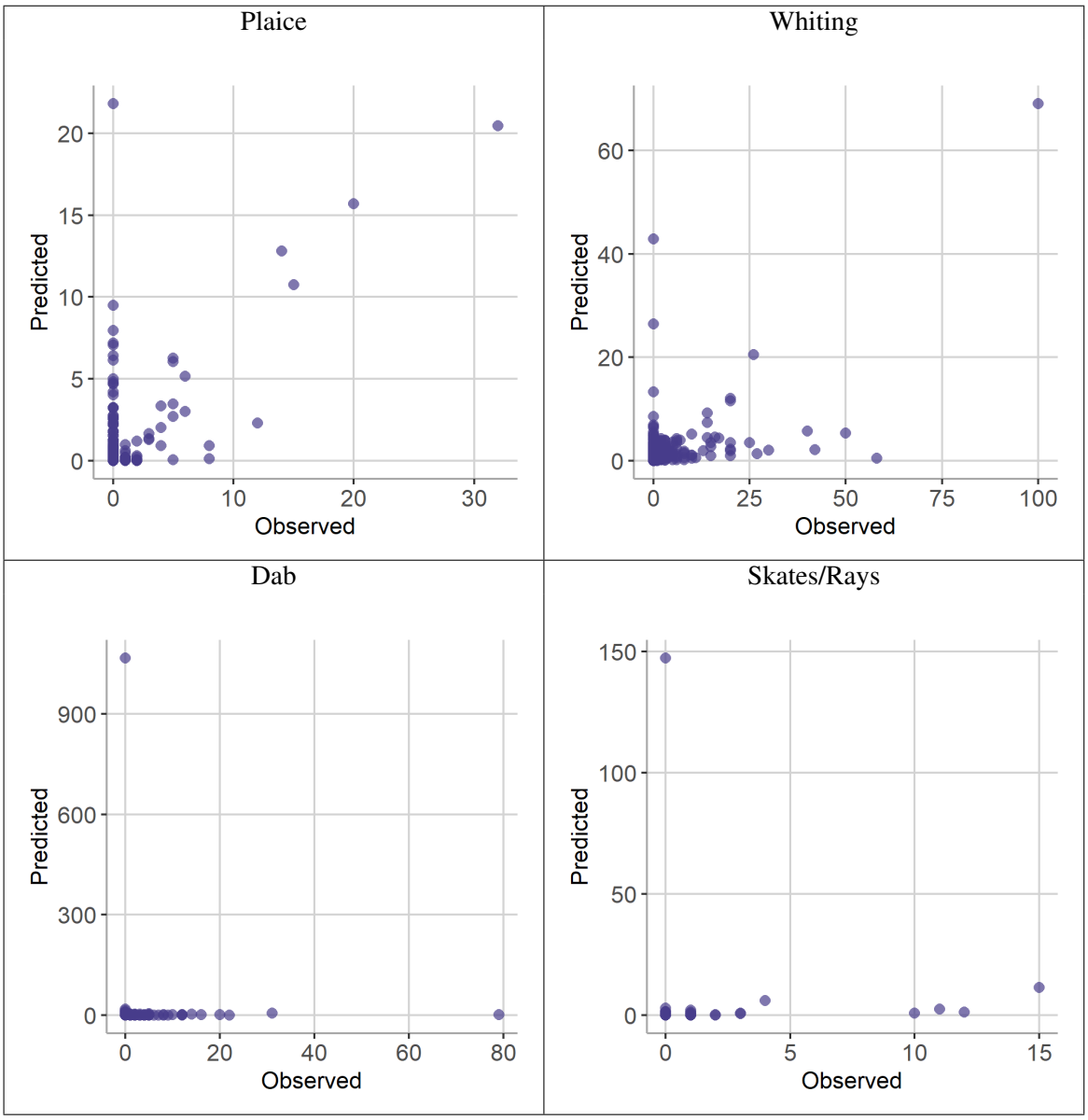


Fig. A.28 Number of observed and predicted catches using the grid method for each interview. Species shown: plaice, whiting, dab and skates/rays.

Results by species

This section contains the results summarised by species. These results are also described in Sections 2.3.2 and 2.3.2 by covariate.

Cod The probability of cod being available is higher when it is being targeted than when it is not being targeted. The probability increases with the duration fished during daylight, twilight and darkness, and with increasing numbers of fish of other species caught. The probability of cod availability also increases as the number of days shore angling experience increases. None of the environment covariates are associated with the probability of cod being available in the area.

If cod is available, most covariates influence the number caught per angler. Catch rates are higher from a boat than from shore, when using lure/fly/jig or a mix of methods compared to bait, and when cod is being targeted. More cod are caught per angler as hours fished during daylight increases, while fewer cod are caught if more fish of other species are caught. Catch rates are also higher at weekends compared to on weekdays, and when fishing in a competition compared to when fishing is not part of a competition.

The number of cod caught per angler, if they are available, decreases as the number of days shore angling and boat angling in the previous three months increases. The number of cod caught by members of an angling club is lower than by non-members. Therefore, angler characteristics that indicate higher angling experience is associated with a lower number of cod caught, if cod are available in the area.

Given cod are available, more are caught per angler in the North Sea, Bristol Channel/Celtic Sea and Irish Sea than in the English Channel. More cod are caught per angler in summer, and fewer in spring and autumn, compared to in winter. Catch rates are higher in moderate compared to calm sea conditions. More cod per angler are caught in areas with mud sediment, and fewer in areas with gravel sediment, compared to in areas with sand sediment.

Sea bass Sea bass are more likely to be available when fishing from a boat than from shore, for anglers using bait compared to lure/fly/jig methods, and when anglers are targeting them. The probability of sea bass being available increases with the duration fished during daylight and darkness,

and decreases with the duration fished during twilight. As the number of fish of other species caught increases, sea bass are less likely to be available. They are also less likely to be available as angling group size increases. Sea bass are less likely to be available at weekends than on weekdays, and when fishing in a competition compared to when fishing is not part of a competition.

The probability of sea bass being available is not associated with the number of days shore angling during the previous three months, but increases as the number of days boat angling in the previous three months increases. Sea bass are more likely to be available when anglers are members of an angling club compared to when anglers are not angling club members.

Sea bass are more likely to be available in angling areas in the North Sea and Irish Sea, and less likely to be available in angling areas in the Bristol Channel/Celtic Sea, compared to angling areas in the English Channel. They are less likely to be available during winter compared to other seasons. They are more likely to be available in rough sea conditions compared to calm sea conditions. The probability of sea bass being available in angling areas increases as lunar illumination increases. There is no difference in the probability of sea bass being available in gravel or rock sediment areas compared to sand areas, but they are less likely when the sediment is mud.

If sea bass are available, the number of sea bass caught per angler is higher when fishing from a boat than from shore. More sea bass are caught per angler as hours fished during daylight and twilight increase, and fewer are caught as hours fished during darkness increases. Fewer sea bass are caught as more fish of other species are caught and as the angling group size increases. The catch rate increases as the number of days boat angling in the previous three months increases. No environmental covariates are associated with sea bass catch rates.

Mackerel Mackerel are more likely to be available in the area when fishing from a boat than from shore, and when using lure/fly/jig or a mix of methods compared to bait. Mackerel are also more likely to be available when they are being targeted than when they are not targeted. The probability of availability increases with the duration fished during daylight and twilight, but decreases with the duration fished during darkness. Catching more fish of other species increases the probability of mackerel being available. The probability of mackerel being available also increases as the angling group size increases. Mackerel are less likely to be available at weekends than on weekdays. There is

no difference in the probability of mackerel being available when fishing in a competition compared to when fishing is not part of a competition.

The probability of mackerel being available decreases as the number of days shore and boat angling in the previous three months increases. The probability is also lower if an angler is a member of an angling club. Therefore, angler characteristics that indicate higher angling avidity are associated with a lower probability of mackerel being available in the area.

Mackerel are more likely to be available in angling areas in the English Channel than in all other regions. They are more likely to be available in spring and summer, and less likely during autumn, compared to winter. They are less likely to be available in rough sea conditions compared to calm sea conditions. The probability of mackerel being available in angling areas decreases as lunar illumination increases. Mackerel are more likely to be available in gravel sediment areas, and less likely in rock and mud sediment areas, compared to sand sediment areas.

If mackerel are available, catch rates increase as the duration fished during daylight increases. Catch rates also increase as more fish of other species are caught and as the angling group size increases. Fewer mackerel are caught by anglers who are members of an angling club compared to non-members. The only environmental covariate associated with mackerel catch rates is lunar illumination, with catch rates increasing as lunar illumination increases.

Flounder All covariates are associated with the probability that flounder are available. In contrast, no covariates are associated with the number of flounder caught, given they are available.

Flounder are more likely to be available when fishing from shore than from a boat, when using bait compared to lure/fly/jig methods, and when they are being targeted. The probability that flounder are available increases with the duration fished during daylight and twilight, and decreases with the duration fished during darkness. Catching more fish of other species is associated with a lower probability of flounder being available. Flounder are more likely to be available as the angling group size increases. Flounder are less likely to be available at weekends than on weekdays, and when fishing in a competition compared to when fishing is not part of a competition.

The probability of flounder being available increases as the number of days shore angling and boat angling in the previous three months increases. The probability is also higher if an angler is a

member of an angling club. Therefore, angler characteristics indicating higher angling experience are associated with a higher probability of flounder being available in the area.

Flounder are more likely to be available in angling areas in the North Sea and Irish Sea, and less likely in the Bristol Channel/Celtic Sea, compared to the English Channel. They are less likely to be available in summer and autumn, compared to winter. They are less likely to be available in moderate and rough sea conditions compared to calm conditions. The probability of flounder being available in angling areas increases as lunar illumination increases. Flounder are less likely to be available in rock sediment areas, and equally likely in gravel, mud and sand sediment areas.

Plaice Plaice are equally likely to be available when fishing from the shore or by boat. However, they are more likely to be available when anglers use bait compared to lure/fly/jig methods or a mix of methods. The probability of plaice being available is higher when they are being targeted than when they are not targeted. The probability of plaice being available increases with the duration fished during daylight and decreases with the duration fished during twilight. The number of fish of other species caught and fishing in a competition do not affect the probability of plaice being available. The probability of plaice being available increases as the angling group size increases. Plaice are more likely to be available on weekdays than at weekends.

The probability of plaice being available is not influenced by the number of days spent shore angling during the previous three months, but increases as the number of days boat angling in the previous three months increases. The probability is not associated with whether the interviewed angler is a member of an angling club.

There is no difference in the probability of plaice being available in North Sea and Irish Sea angling areas compared to angling areas in the English Channel, but they are less likely to be available in angling areas in the Bristol Channel/Celtic Sea compared to the English Channel. They are more likely to be available in spring, and less likely in autumn, compared to in winter. They are more likely to be available in moderate sea conditions, and less likely in rough sea conditions, compared to during calm sea conditions. Lunar illumination and the seabed sediment type are not associated with the probability of plaice being available in angling areas.

When plaice are available, catch rates are higher for boat anglers than for shore anglers, and when using a mix of lure/fly/jig and bait methods compared to only bait. More plaice are caught per angler when plaice are being targeted, compared to when they are not being targeted. The number of plaice caught increases as the duration fished during daylight and twilight increases, if they are available. The duration fished in darkness does not impact the number of plaice caught. Fewer plaice are caught per angler as more fish of other species are caught, but the number caught increases as the angling group size increases. More plaice are caught per angler at weekends compared to on weekdays. Fishing in a competition does not impact the number of plaice caught compared to not fishing in a competition.

The number of plaice caught, if they are available, increases as the number of days shore angling in the previous three months increases, but decreases as the number of days boat angling increases. The number of plaice caught by members of an angling club is higher than the number caught by non-members.

Given plaice are available, fewer are caught by anglers fishing in the North Sea than by anglers fishing in the English Channel. There is no difference in the number of plaice caught in the Bristol Channel/Celtic Sea and Irish Sea regions, compared to the English Channel. More plaice are caught in spring and summer compared to in winter, given that plaice are available in the area. Given plaice are available, more are caught in calm sea conditions compared to in moderate sea conditions, with no difference in the number caught in calm and rough sea conditions. If plaice are available, more are caught as the lunar illumination increases. Compared to areas where the seabed sediment is sand, more plaice are caught in areas with gravel sediment, fewer are caught in rock sediment areas, with no difference in the number caught when the sediment type is mud, if plaice are available in the angling area.

Whiting Whiting are more likely to be available when fishing from a boat than from shore, and when using bait compared to lure/fly/jig or a mix of methods. The probability of whiting being available is higher when they are being targeted than when they are not targeted. The probability increases with the duration fished during daylight and darkness, and decreases with the duration fished during twilight. The probability of whiting being available increases as more fish of other species are

caught and as the angling group size increases. Whiting are more likely to be available at weekends than on weekdays. Whiting are less likely to be available when fishing in a competition compared to when fishing is not part of a competition.

The probability of whiting being available increases as the number of days shore angling and boat angling in the previous three months increases. The probability is also higher if an angler is a member of an angling club. Therefore, angler characteristics indicating higher angling avidity are associated with a higher probability of whiting being available in the area.

Whiting are more likely to be available in angling areas in the North Sea and Irish Sea, compared to angling areas in the English Channel. There is no difference in the probability of whiting being available in the Bristol Channel/Celtic Sea angling areas compared to English Channel angling areas. They are less likely to be available in spring and summer, and more likely during autumn, compared to winter. They are more likely to be available in rough sea conditions compared to calm conditions. Lunar illumination is not associated with the probability of whiting being available in angling areas. Whiting are less likely to be available in gravel, rock and mud sediment areas compared to sand.

More whiting are caught as more fish of other species are caught, and as the number of days boat angling experience increases, if whiting are available in the area.

Dab The probability of dab being available in the area is higher when fishing from a boat than from shore, and when using bait compared to lure/fly/jig or a mix of methods. The probability of dab being available is higher when they are being targeted than when they are not targeted. The probability increases with the duration fished during daylight, and decreases with the duration fished during twilight. The probability of dab being available increases as more fish of other species are caught and as the angling group size increases. There is no difference in probability of dab being available at weekends compared to on weekdays. Dab are less likely to be available when fishing in a competition compared to when fishing is not part of a competition.

The probability of dab being available decreases as the number of days shore and boat angling in the previous three months increases. However, dab are more likely to be available when an angler is a member of an angling club than for anglers who are not members of an angling club.

Dab are more likely to be available in angling areas in the North Sea and Irish Sea, and less likely to be available in the Bristol Channel/Celtic Sea angling areas, compared to angling areas in the English Channel. They are more likely to be available in spring, and less likely during summer and autumn, compared to winter. They are less likely to be available in rough sea conditions compared to calm conditions. The probability of dab being available in angling areas decreases as lunar illumination increases. Dab are more likely to be available in gravel sediment areas, and less likely in mud sediment areas, compared to sand sediment areas.

If dab are available, the number of dab caught per angler is higher when fishing from a boat than from shore and as the duration fished in daylight increases. More dab are caught as more fish of other species are caught. As the number of days shore angling experience increases, the number of dab caught decreases, while more dab are caught by anglers as boat angling experience increases. More dab are caught by angling club members than by anglers who are not members of an angling club.

Skates/Rays The probability of skates/rays being available in the area is higher when fishing from a boat than from shore, and when they are being targeted compared to when they are not targeted. There is no association between the fishing method used and the probability of skates/rays being available. The probability that skates/rays are available increases as the duration fished during daylight increases, and decreases as the duration fished during darkness increases. The probability increases as more fish of other species are caught and as the angling group size increases. There is no difference in the probability of skates/rays being available on weekdays and at weekends, or when fishing in a competition compared to when fishing is not part of a competition.

The probability of skates/rays being available is not associated with the number of days shore angling and boat angling in the previous three months. However, the probability is higher if an angler is a member of an angling club, compared to non-members.

Skates/rays are less likely to be available in angling areas in the North Sea, compared to angling areas in the English Channel. There is no difference in the probability of skates/rays being available in the Irish Sea and Bristol Channel/Celtic Sea angling areas compared to English Channel angling areas. They are more likely to be available in autumn, and equally likely in spring and summer, compared to in winter. They are less likely to be available in moderate sea conditions compared to

calm sea conditions. The probability of skates/rays being available in angling areas decreases as lunar illumination increases. There is no difference in the probability of skates/rays being available in different types of seabed sediment within angling areas.

More skates/rays are caught per angler as more fish of other species are caught. Increasing angling group size results in fewer skates/rays caught, if skates/rays are available.

A.4 Default method

A.4.1 Fitting the models

Method

Using the default method from the `mpath` package, we fitted models using *n* folds, with the method's default setting of 100 pairs of λ values[1]. Figures A.29 (cod, sea bass, mackerel, flounder) and A.30 (plaice, whiting, dab, skates/rays) show the λ sequences selected by the default method for each species and Figure A.31 shows the loglikelihood value for each λ pair, for all species. The values of the λ pairs and the corresponding loglikelihood are given in Tables A.8 and A.9. The results are given in Tables A.11 (cod, sea bass, mackerel, flounder) and A.12 (plaice, whiting, dab, skates/rays).

[1]Testing with 500 pairs to examine scaling effects showed that the λ range remained similar, with additional pairs primarily increasing density along the existing path rather than extending it.

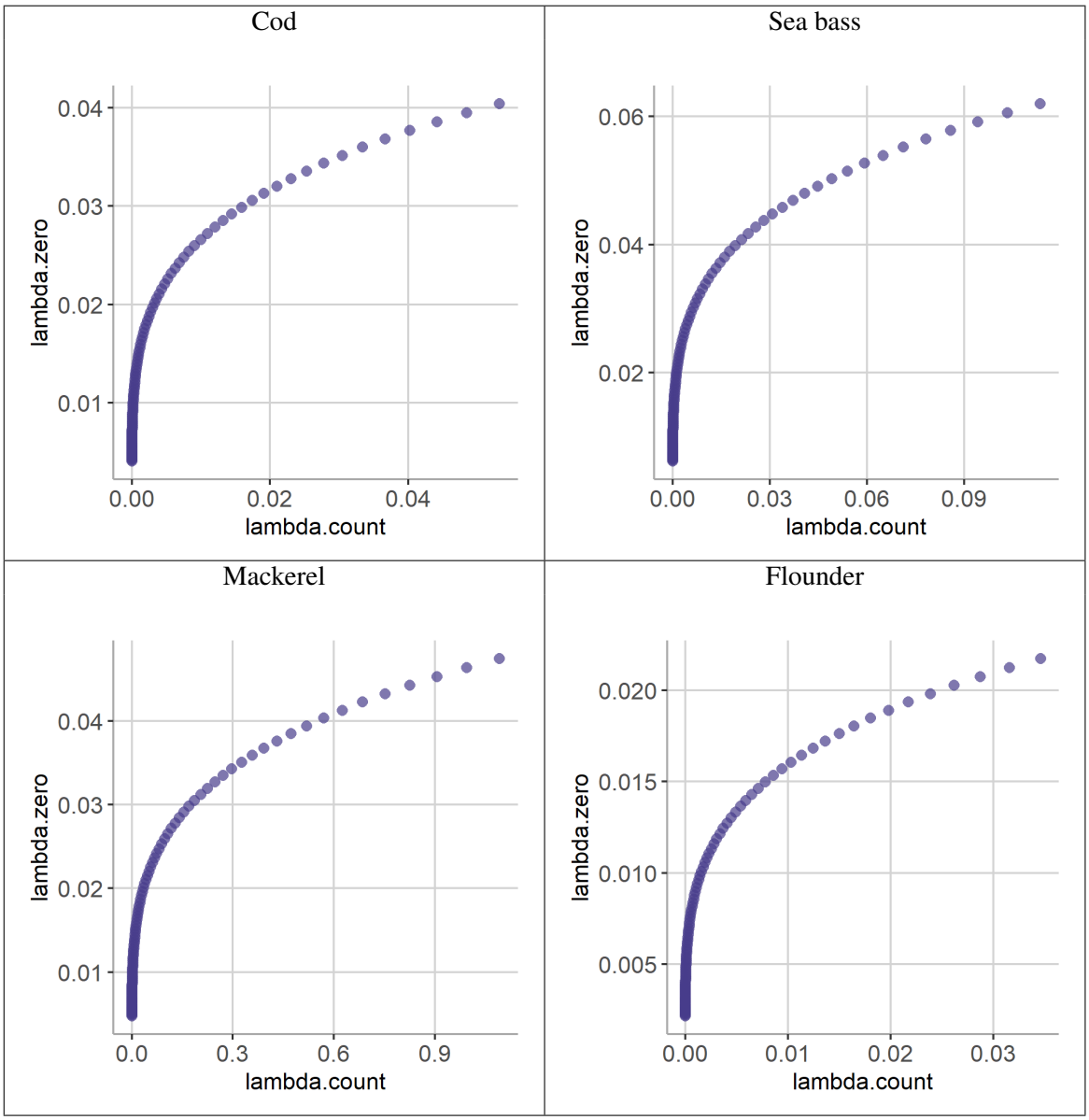


Fig. A.29 Values of the λ pairs used in the Poisson and Binomial parts of the model using the default method with 100 λ pairs. Species shown: cod, sea bass, mackerel and flounder. Note the different axes scales for the different species.

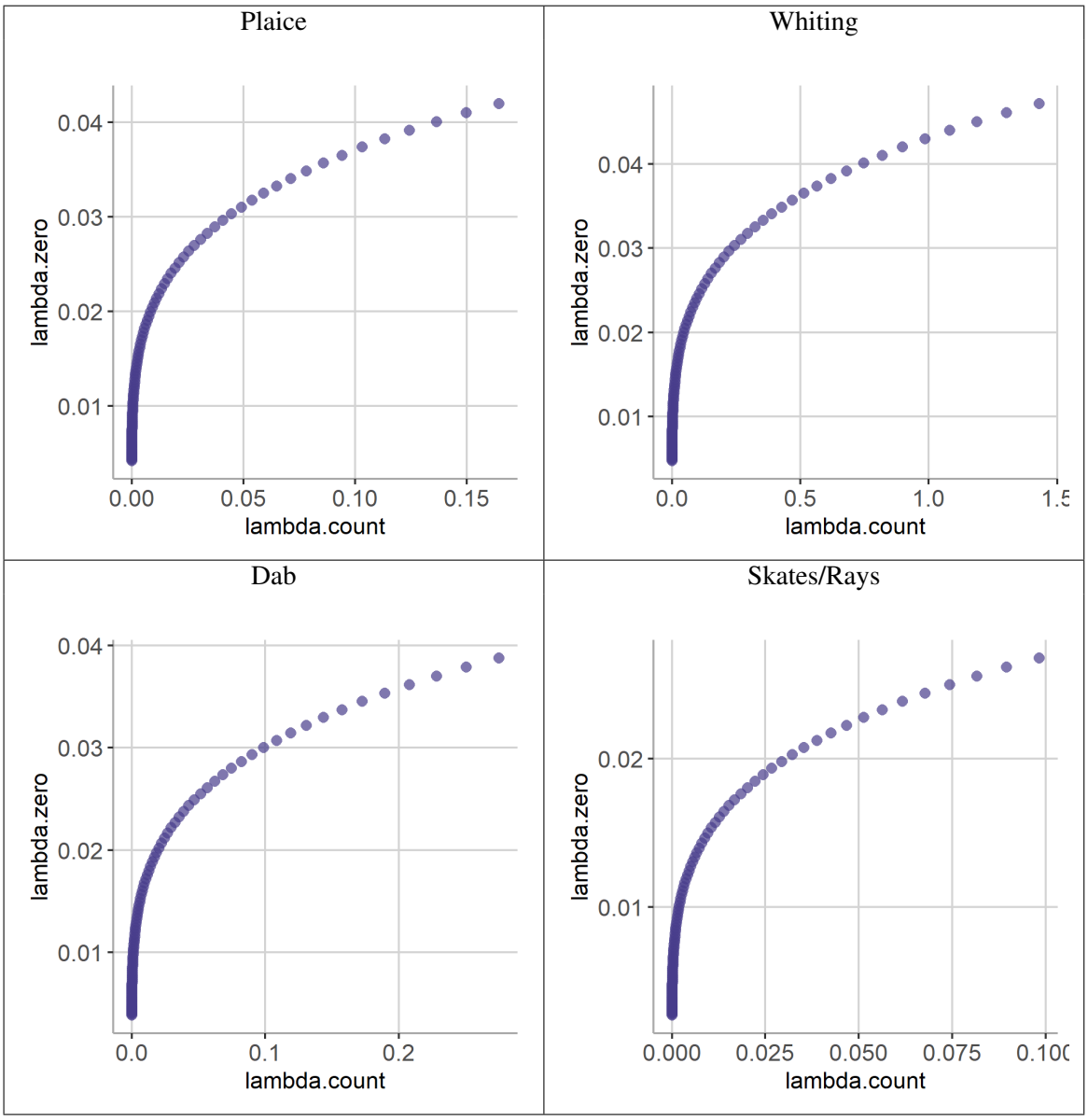


Fig. A.30 Values of the λ pairs used in the Poisson and Binomial parts of the model using the default method with 100 λ pairs. Species shown: plaice, whiting, dab, and skates/rays. Note the different axes scales for the different species.

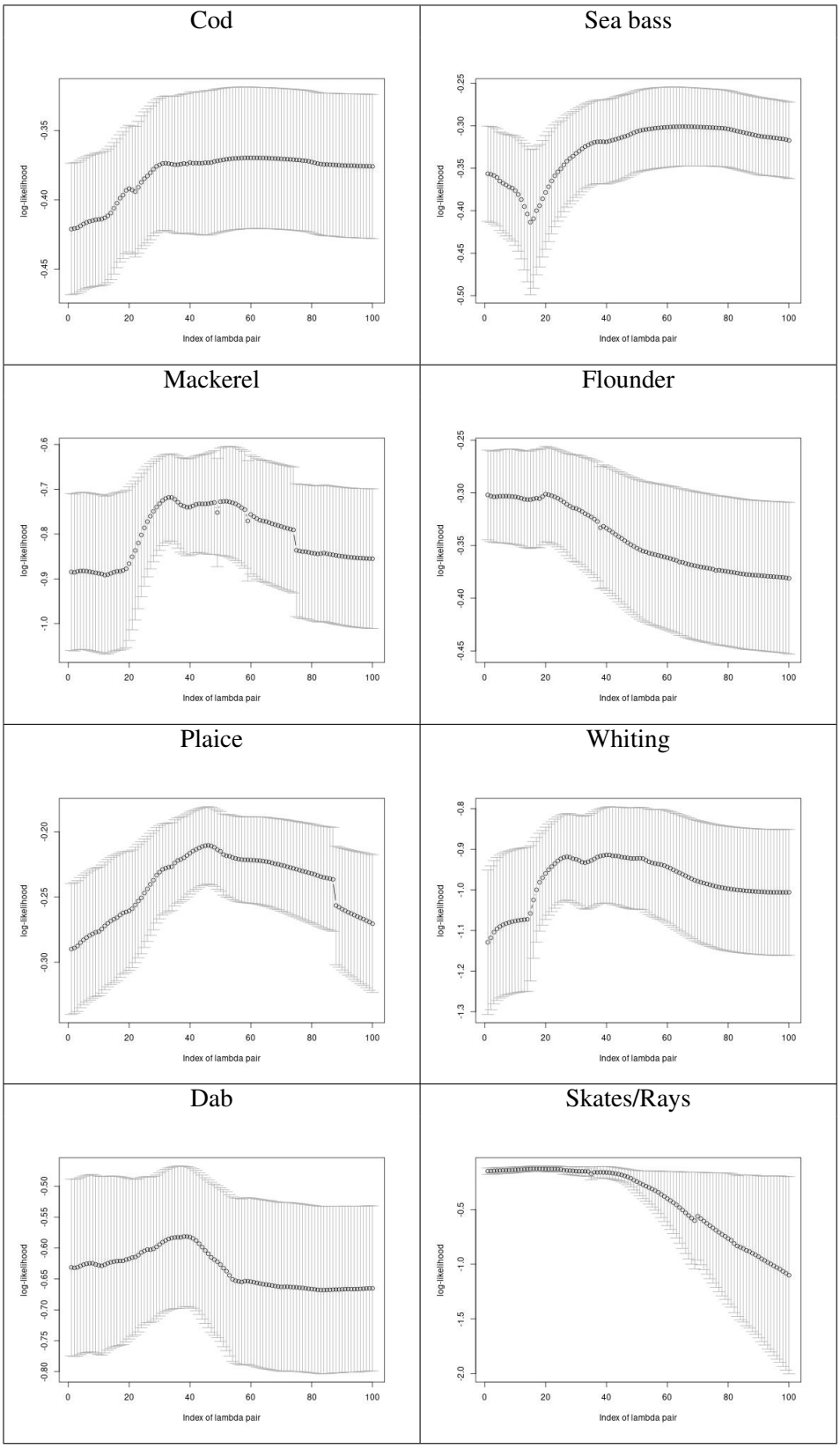


Fig. A.31 Loglikelihood value for each pair of λ values for the default method with n folds.

Table A.8 λ values used by the default method with 100 λ pairs and n folds and the resulting loglikelihood. Species shown: cod, sea bass, mackerel, and flounder.

Cod			Sea bass			Mackerel			Flounder		
lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood
0.05319	0.04043	−459.77698	0.11354	0.06197	−549.00374	1.09168	0.04746	−1145.12612	0.03461	0.02173	−390.10751
0.04846	0.03950	−456.84456	0.10345	0.06054	−543.71965	0.99470	0.04637	−1107.61235	0.03153	0.02123	−389.02250
0.04416	0.03859	−454.00107	0.09426	0.05915	−539.46562	0.90633	0.04531	−1081.56502	0.02873	0.02074	−386.58736
0.04024	0.03770	−450.70290	0.08589	0.05779	−533.91153	0.82582	0.04427	−1062.67012	0.02618	0.02026	−383.45630
0.03666	0.03684	−443.93952	0.07826	0.05646	−529.13376	0.75245	0.04325	−1048.53380	0.02385	0.01980	−381.21158
0.03340	0.03599	−436.06719	0.07130	0.05516	−525.03690	0.68561	0.04225	−1037.71143	0.02173	0.01934	−379.27405
0.03044	0.03516	−429.84583	0.06497	0.05389	−521.56961	0.62470	0.04128	−1029.23548	0.01980	0.01890	−377.71192
0.02773	0.03435	−424.63868	0.05920	0.05266	−518.62054	0.56920	0.04033	−1022.30844	0.01804	0.01846	−376.44989
0.02527	0.03356	−420.28941	0.05394	0.05145	−516.09778	0.51864	0.03941	−1016.74730	0.01644	0.01804	−375.27947
0.02302	0.03279	−416.63700	0.04915	0.05026	−513.92689	0.47256	0.03850	−1012.23564	0.01498	0.01763	−373.29706
0.02098	0.03204	−413.24145	0.04478	0.04911	−512.05843	0.43058	0.03761	−1008.54122	0.01365	0.01722	−371.37739
0.01912	0.03130	−408.50438	0.04080	0.04798	−506.34978	0.39233	0.03675	−1002.90948	0.01244	0.01682	−369.43330
0.01742	0.03058	−403.03318	0.03718	0.04687	−501.06801	0.35748	0.03591	−989.20587	0.01133	0.01644	−366.48487
0.01587	0.02988	−397.96940	0.03388	0.04580	−496.58995	0.32572	0.03508	−978.32782	0.01033	0.01606	−362.43661
0.01446	0.02919	−391.17080	0.03087	0.04474	−492.91802	0.29678	0.03427	−969.61886	0.00941	0.01569	−356.96257
0.01318	0.02852	−384.39238	0.02812	0.04372	−489.27348	0.27042	0.03349	−962.59462	0.00857	0.01533	−352.68753
0.01201	0.02787	−378.97633	0.02563	0.04271	−484.51842	0.24639	0.03272	−954.45793	0.00781	0.01498	−349.81205
0.01094	0.02722	−372.70742	0.02335	0.04173	−479.52137	0.22451	0.03196	−944.88791	0.00712	0.01463	−347.22100
0.00997	0.02660	−363.05166	0.02127	0.04077	−475.00462	0.20456	0.03123	−932.54157	0.00648	0.01430	−343.09079
0.00908	0.02599	−356.13748	0.01938	0.03983	−470.98042	0.18639	0.03051	−921.17390	0.00591	0.01397	−338.30346
0.00827	0.02539	−350.57252	0.01766	0.03892	−466.95186	0.16983	0.02981	−911.26729	0.00538	0.01365	−334.45045
0.00754	0.02481	−346.22459	0.01609	0.03802	−458.89213	0.15474	0.02912	−896.73412	0.00491	0.01333	−331.57238
0.00687	0.02424	−342.28514	0.01466	0.03715	−450.54966	0.14100	0.02845	−884.04654	0.00447	0.01303	−328.91666
0.00626	0.02368	−339.38276	0.01336	0.03629	−444.11753	0.12847	0.02780	−873.07100	0.00407	0.01273	−327.00116
0.00570	0.02313	−336.06409	0.01217	0.03546	−438.35190	0.11706	0.02716	−863.53179	0.00371	0.01243	−325.29861
0.00520	0.02260	−332.59995	0.01109	0.03464	−433.60835	0.10666	0.02654	−855.20170	0.00338	0.01215	−323.84643
0.00474	0.02208	−329.94955	0.01011	0.03385	−428.66921	0.09718	0.02593	−847.89287	0.00308	0.01187	−322.33619
0.00431	0.02158	−327.57654	0.00921	0.03307	−424.47723	0.08855	0.02533	−841.44939	0.00281	0.01160	−320.94297
0.00393	0.02108	−324.81316	0.00839	0.03231	−420.79406	0.08068	0.02475	−835.74148	0.00256	0.01133	−319.30760
0.00358	0.02059	−320.94964	0.00765	0.03157	−417.76717	0.07352	0.02418	−830.66077	0.00233	0.01107	−317.78142
0.00326	0.02012	−317.48528	0.00697	0.03084	−415.05869	0.06698	0.02362	−826.08676	0.00212	0.01081	−316.40985
0.00297	0.01966	−314.78624	0.00635	0.03013	−412.70255	0.06103	0.02308	−821.87836	0.00193	0.01057	−315.12383
0.00271	0.01921	−312.49841	0.00578	0.02944	−410.78725	0.05561	0.02255	−816.31544	0.00176	0.01032	−313.99597
0.00247	0.01877	−310.37991	0.00527	0.02876	−408.82469	0.05067	0.02203	−810.00834	0.00161	0.01009	−312.80321
0.00225	0.01833	−308.41067	0.00480	0.02810	−406.45988	0.04617	0.02152	−804.11219	0.00146	0.00985	−311.71994
0.00205	0.01791	−306.71938	0.00438	0.02745	−403.99183	0.04207	0.02103	−797.85675	0.00133	0.00963	−310.76664
0.00187	0.01750	−305.25134	0.00399	0.02682	−401.74790	0.03833	0.02055	−787.35171	0.00122	0.00941	−309.88076
0.00170	0.01710	−303.95519	0.00363	0.02621	−399.94727	0.03493	0.02007	−776.53142	0.00111	0.00919	−309.16280
0.00155	0.01670	−302.66237	0.00331	0.02560	−398.65975	0.03182	0.01961	−766.31089	0.00101	0.00898	−308.32495
0.00141	0.01632	−301.48161	0.00302	0.02502	−396.92917	0.02900	0.01916	−756.93746	0.00092	0.00877	−307.58759
0.00129	0.01595	−300.35685	0.00275	0.02444	−395.39930	0.02642	0.01872	−748.75293	0.00084	0.00857	−307.01073
0.00117	0.01558	−299.41406	0.00250	0.02388	−393.36631	0.02407	0.01829	−741.40424	0.00076	0.00837	−306.21023
0.00107	0.01522	−298.54210	0.00228	0.02333	−391.08760	0.02193	0.01787	−734.77106	0.00070	0.00818	−305.63644
0.00097	0.01487	−297.78507	0.00208	0.02279	−389.55653	0.01999	0.01746	−726.11766	0.00063	0.00799	−304.99149
0.00089	0.01453	−297.12920	0.00189	0.02227	−387.52430	0.01821	0.01706	−717.69459	0.00058	0.00781	−304.22912
0.00081	0.01420	−296.60437	0.00173	0.02176	−385.89520	0.01659	0.01667	−709.87283	0.00053	0.00763	−303.66531
0.00074	0.01387	−296.10654	0.00157	0.02126	−384.51559	0.01512	0.01628	−702.43107	0.00048	0.00745	−303.29024
0.00067	0.01355	−295.66480	0.00143	0.02077	−383.52021	0.01378	0.01591	−695.90810	0.00044	0.00728	−302.72764
0.00061	0.01324	−295.27431	0.00131	0.02029	−382.10256	0.01255	0.01554	−690.19128	0.00040	0.00712	−302.20104
0.00056	0.01293	−294.92812	0.00119	0.01982	−381.15452	0.01144	0.01519	−685.09250	0.00036	0.00695	−301.51368
0.00051	0.01264	−294.61972	0.00108	0.01937	−380.26928	0.01042	0.01484	−680.43612	0.00033	0.00679	−301.15117
0.00046	0.01235	−294.34343	0.00099	0.01892	−379.48172	0.00949	0.01450	−676.31875	0.00030	0.00664	−300.54457
0.00042	0.01206	−294.09446	0.00090	0.01849	−378.78265	0.00865	0.01416	−672.50436	0.00027	0.00648	−300.26047
0.00038	0.01178	−293.86877	0.00082	0.01806	−378.16109	0.00788	0.01384	−668.92064	0.00025	0.00633	−299.82783
0.00035	0.01151	−293.66305	0.00075	0.01765	−377.60695	0.00718	0.01352	−665.72867	0.00023	0.00619	−299.49306
0.00032	0.01125	−293.47448	0.00068	0.01724	−377.11147	0.00654	0.01321	−662.89216	0.00021	0.00605	−299.16418
0.00029	0.01099	−293.30074	0.00062	0.01685	−376.66704	0.00596	0.01290	−659.59077	0.00019	0.00591	−298.85416
0.00026	0.01074	−293.13988	0.00057	0.01646	−376.26710	0.00543	0.01261	−656.15531	0.00017	0.00577	−298.56515
0.00024	0.01049	−292.99029	0.00051	0.01608	−375.81902	0.00495	0.01232	−653.22402	0.00016	0.00564	−298.29616
0.00022	0.01025	−292.85059	0.00047	0.01571	−375.55165	0.00451	0.01203	−650.34809	0.00014	0.00551	−298.04538
0.00020	0.01001	−292.71966	0.00043	0.01535	−375.25323	0.00411	0.01176	−647.21286	0.00013	0.00538	−297.81100
0.00018	0.00978	−292.59655	0.00039	0.01500	−374.97520	0.00374	0.01149	−644.43846	0.00012	0.00526	−297.59141
0.00017	0.00956	−292.48044	0.00035	0.01465	−374.71905	0.00341	0.01122	−641.65261	0.00011	0.00514	−297.25325
0.00015	0.00934	−292.37065	0.00032	0.01431	−374.48311	0.00311	0.01096	−639.24395	0.00010	0.00502	−297.07940
0.00014	0.00912	−292.26661	0.00029	0.01399	−374.26530	0.00283	0.01071	−637.14479	0.00009	0.00490	−296.81868
0.00013	0.00891	−292.10686	0.00027	0.01366	−374.06368	0.00258	0.01047	−635.49342	0.00008	0.00479	−296.54803
0.00011	0.00871	−292.06440	0.00024	0.01335	−373.87657	0.00235	0.01023	−633.41984	0.00007	0.00468	−296.11845
0.00010	0.00851	−291.94363	0.00022	0.01304	−373.70247	0.00214	0.00999	−631.75632	0.00007	0.00457	−295.94214
0.00010	0.00831	−291.83379	0.00020	0.01274	−373.54010	0.00195	0.00976	−630.42894	0.00006	0.00447	−295.67091
0.00009	0.00812	−291.72742	0.00019	0.01245	−373.38833	0.00178	0.00954	−628.99895	0.00006	0.00437	−295.39433
0.00008	0.00794	−291.62375	0.00017	0.01216	−373.24620	0.00162	0.00932	−627.86859	0.00005	0.00427	−295.12265
0.00007	0.00775	−291.52274	0.00015	0.01188	−373.11283	0.00148	0.00910	−626.73637	0.00005	0.00417	−294.85991
0.00007	0.00758	−291.42449	0.00014	0.01161	−372.98748	0.00135	0.00889	−625.67963	0.00004	0.00407	−294.60764
0.00006	0.00740	−291.32909	0.00013	0.01134	−372.86949	0.00123	0.00869	−624.69784	0.00004	0.00398	−294.36626
0.00005	0.00723	−291.23662	0.00012	0.01108	−372.75827	0.00112	0.00849	−623.68840	0.00004	0.00389	−294.13568
0.00005	0.00706	−291.14713	0.00011	0.01083	−372.65331	0.00102	0.00829	−622.74566	0.00003	0.00380	−293.77555
0.00005	0.00690	−291.06063	0.00010	0.01058	−372.55413	0.00093	0.00810	−621.80841	0.00003	0.00371	−293.49317
0.00004	0.00674	−290.97713	0.00009	0.01034	−372.46032	0.00085	0.00792	−620.88402	0.00003	0.00362	−293.40034
0.00004	0.00659	−290.89660	0.00008	0.01010	−372.37151	0.00077	0.00774	−620.09375	0.00002	0.00354	−292.99626
0.00003	0.00644	−290.69456	0.00007	0.00987	−372.20405	0.00070	0.00756	−619.32211	0.00002	0.00346	−292.75598
0.00003	0.00629	−290.61028	0.00007	0.00964	−372.07669	0.00064	0.00738	−618.59854	0.00002	0.00338	−292.67033
0.00003	0.00614	−290.31266	0.00006	0.00942	−371.67260	0.00058	0.00721	−616.90152	0.00002	0.00330	−292.42259
0.00003	0.00600	−289.56064	0.00006	0.00920	−371.19299	0.00053	0.00705	−616.21440	0.00002	0.00323	−292.00508
0.00002	0.00587	−288.86918	0.00005	0.00899	−370.75654	0.00048	0.00689	−615.49421	0.00002	0.00315	−291.78841
0.00002	0.00573	−288.26126	0.00005	0.00878	−370.33485	0.00044	0.00673	−614.82498	0.00001	0.00308	−291.44341
0.00002	0.00560	−287.69908	0.00004	0.00858	−369.92648	0.00040	0.00657	−614.24969	0.00001	0.00301	−291.08850
0.00002	0.00547	−287.18433	0.00004	0.00838	−369.33970	0.00037	0.00642	−613.56863	0.00001	0.00294	−290.73794
0.00002	0.00534	−286.71630	0.00003	0.00819	−369.14102	0.00033	0.00627	−613.05488	0.00001	0.00287	−290.32985
0.00001	0.00522	−286.16019	0.00003	0.00800	−368.57718	0.00030	0.00613	−612.44590	0.00001	0.00281	−290.03901
0.00001	0.00510	−285.76883	0.00003	0.00782	−368.21678	0.00028	0.00599	−611.98305	0.00001	0.00274	−289.73555
0.00001	0.00498	−285.35365	0.00003	0.00764	−368.04124	0.00025	0.00585	−611.43299	0.00001	0.00268	−289.43749
0.00001	0.00487	−284.94565	0.00002	0.00746	−367.75890	0.00023	0.00572	−611.01346	0.00001	0.00262	−289.14956
0.00001	0.00476	−284.61432	0.00002	0.00729	−367.47325	0.00021	0.00559	−610.36996	0.00001	0.00256	−288.87310
0.00001	0.00465	−284.25193	0.00002	0.00712	−367.19275	0.00019	0.00546	−609.81995	0.00001	0.00250	−288.60836
0.00001	0.00454	−283.96719	0.00002	0.00696	−366.92057	0.00017	0.00533	−609.27282	0.00001	0.00244	−288.35519
0.00001	0.00444	−283.65089	0.00002	0.00680	−366.65813	0.00016	0.00521	−608.74650	0.00001	0.00238	−288.11328
0.00001	0.00433	−283.37897	0.00002	0.00664	−366.40610	0.00014	0.00509	−608.23992	0.00000	0.00233	−287.88223
0.00001	0.00424	−283.14412	0.00001	0.00649	−366.16473	0.00013	0.00497	−607.75209	0.00000	0.00228	−287.66161
0.00001	0.00414	−282.90980	0.00001	0.00634	−365.93408	0.00012	0.00486	−607.28212	0.00000	0.00222	−287.45097
0.00001	0.00404	−282.68011	0.00001	0.00620	−365.71401	0.00011	0.00475	−606.82916	0.00000	0.00217	−287.24987

Table A.9 λ values used by the default method with 100 λ pairs and n folds and the resulting loglikelihood. Species shown: plaice, whiting, dab, and skates/rays.

Plaice			Whiting			Dab			Skates/Rays		
lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood	lambda.count	lambda.zero	loglikelihood
0.16442	0.04196	−373.22432	1.43024	0.04718	−1469.48923	0.27487	0.03877	−809.46335	0.09830	0.02681	−188.87908
0.14982	0.04100	−364.80943	1.30318	0.04610	−1384.38076	0.25045	0.03788	−790.55945	0.08957	0.02620	−187.33311
0.13651	0.04006	−358.58737	1.18741	0.04504	−1336.23607	0.22821	0.03700	−776.63048	0.08161	0.02559	−177.48884
0.12438	0.03914	−353.78947	1.08193	0.04400	−1305.37698	0.20793	0.03615	−766.04298	0.07436	0.02501	−173.38892
0.11333	0.03824	−349.97405	0.98581	0.04299	−1284.14001	0.18946	0.03532	−757.81983	0.06775	0.02443	−170.02611
0.10326	0.03736	−346.86686	0.89823	0.04200	−1268.84637	0.17263	0.03451	−751.32861	0.06174	0.02387	−167.24092
0.09409	0.03650	−344.28754	0.81844	0.04104	−1257.48231	0.15729	0.03372	−746.13779	0.05625	0.02332	−164.90907
0.08573	0.03566	−342.11221	0.74573	0.04009	−1248.84257	0.14332	0.03294	−741.94208	0.05125	0.02278	−162.93879
0.07812	0.03484	−340.25276	0.67948	0.03917	−1242.15807	0.13059	0.03218	−737.50609	0.04670	0.02226	−161.26060
0.07118	0.03404	−336.80750	0.61912	0.03827	−1236.91405	0.11899	0.03144	−728.52921	0.04255	0.02175	−159.82087
0.06485	0.03326	−331.53738	0.56412	0.03739	−1232.75305	0.10842	0.03072	−718.08666	0.03877	0.02125	−158.57745
0.05909	0.03249	−326.84199	0.51400	0.03653	−1229.41963	0.09878	0.03002	−709.34980	0.03533	0.02076	−157.49678
0.05384	0.03174	−322.75747	0.46834	0.03569	−1226.67423	0.09001	0.02933	−702.10809	0.03219	0.02028	−156.55188
0.04906	0.03101	−319.13326	0.42673	0.03487	−1224.07653	0.08201	0.02865	−696.09711	0.02933	0.01982	−155.72080
0.04470	0.03030	−316.08545	0.38882	0.03407	−1221.91138	0.07473	0.02799	−691.10137	0.02672	0.01936	−155.00309
0.04073	0.02960	−313.48012	0.35428	0.03329	−1220.09333	0.06809	0.02735	−686.94449	0.02435	0.01892	−154.34766
0.03711	0.02892	−311.26560	0.32281	0.03252	−1218.55605	0.06204	0.02672	−683.48141	0.02219	0.01848	−152.60325
0.03381	0.02826	−309.39300	0.29413	0.03177	−1217.24737	0.05653	0.02611	−680.59250	0.02022	0.01806	−150.99695
0.03081	0.02761	−307.81591	0.26800	0.03104	−1212.66492	0.05151	0.02551	−678.17898	0.01842	0.01764	−149.52413
0.02807	0.02697	−302.98224	0.24419	0.03033	−1201.61859	0.04693	0.02492	−671.79294	0.01678	0.01724	−148.20501
0.02558	0.02635	−297.73767	0.22250	0.02963	−1192.89469	0.04276	0.02435	−666.44098	0.01529	0.01684	−146.59055
0.02331	0.02575	−293.19140	0.20273	0.02895	−1185.92345	0.03896	0.02379	−657.70781	0.01393	0.01645	−145.04784
0.02124	0.02516	−288.23094	0.18472	0.02829	−1180.19144	0.03550	0.02324	−648.82343	0.01270	0.01607	−143.70073
0.01935	0.02458	−283.74510	0.16831	0.02764	−1175.52624	0.03235	0.02271	−641.32497	0.01157	0.01570	−142.54692
0.01763	0.02401	−279.12504	0.15336	0.02700	−1171.70478	0.02947	0.02218	−634.52397	0.01054	0.01534	−141.55780
0.01606	0.02346	−275.28074	0.13974	0.02638	−1168.55571	0.02686	0.02167	−628.10775	0.00960	0.01499	−140.70701
0.01464	0.02292	−272.14468	0.12732	0.02577	−1165.94621	0.02447	0.02118	−621.26787	0.00875	0.01465	−138.80288
0.01334	0.02239	−269.53882	0.11601	0.02518	−1163.65314	0.02230	0.02069	−612.06001	0.00797	0.01431	−136.99551
0.01215	0.02188	−267.37527	0.10571	0.02460	−1159.88195	0.02032	0.02021	−603.87658	0.00727	0.01398	−135.42131
0.01107	0.02138	−265.57161	0.09631	0.02404	−1155.88652	0.01851	0.01975	−596.73767	0.00662	0.01366	−134.05642
0.01009	0.02089	−264.04694	0.08776	0.02348	−1152.50342	0.01687	0.01929	−590.51231	0.00603	0.01334	−129.27266
0.00919	0.02041	−262.75588	0.07996	0.02294	−1141.02927	0.01537	0.01885	−585.08583	0.00550	0.01304	−126.30548
0.00838	0.01994	−260.10239	0.07286	0.02242	−1130.01249	0.01400	0.01842	−580.35747	0.00501	0.01274	−123.80007
0.00763	0.01948	−257.71408	0.06639	0.02190	−1119.63956	0.01276	0.01799	−576.23824	0.00456	0.01245	−121.89183
0.00695	0.01903	−255.65380	0.06049	0.02140	−1109.35337	0.01163	0.01758	−572.64943	0.00416	0.01216	−120.48308
0.00634	0.01859	−254.00359	0.05511	0.02091	−1097.74638	0.01059	0.01718	−569.52150	0.00379	0.01188	−119.30142
0.00577	0.01816	−251.90254	0.05022	0.02042	−1087.98767	0.00965	0.01678	−566.44376	0.00345	0.01161	−117.84119
0.00526	0.01775	−250.01229	0.04576	0.01996	−1079.75730	0.00879	0.01640	−563.90931	0.00314	0.01134	−116.79644
0.00479	0.01734	−248.40213	0.04169	0.01950	−1071.82257	0.00801	0.01602	−561.22083	0.00287	0.01108	−115.80364
0.00437	0.01694	−247.05786	0.03799	0.01905	−1064.95153	0.00730	0.01565	−559.11607	0.00261	0.01082	−114.53796
0.00398	0.01655	−245.92709	0.03461	0.01861	−1059.03600	0.00665	0.01529	−556.73378	0.00238	0.01058	−113.37976
0.00363	0.01617	−245.16555	0.03154	0.01818	−1052.53602	0.00606	0.01494	−554.20823	0.00217	0.01033	−112.36704
0.00330	0.01580	−244.34994	0.02874	0.01776	−1046.10456	0.00552	0.01460	−551.78018	0.00198	0.01009	−111.44766
0.00301	0.01544	−243.15754	0.02618	0.01736	−1040.54962	0.00503	0.01426	−549.11468	0.00180	0.00986	−110.57685
0.00274	0.01508	−241.80523	0.02386	0.01696	−1035.23102	0.00459	0.01393	−546.68551	0.00164	0.00964	−109.85235
0.00250	0.01473	−240.37146	0.02174	0.01657	−1030.28974	0.00418	0.01361	−544.31724	0.00149	0.00941	−109.07442
0.00228	0.01440	−239.16353	0.01981	0.01619	−1025.98668	0.00381	0.01330	−542.14307	0.00136	0.00920	−108.30567
0.00207	0.01406	−238.26165	0.01805	0.01581	−1019.59701	0.00347	0.01299	−540.30037	0.00124	0.00899	−107.40952
0.00189	0.01374	−236.76586	0.01644	0.01545	−1013.16136	0.00316	0.01269	−538.63950	0.00113	0.00878	−106.28737
0.00172	0.01343	−235.29254	0.01498	0.01510	−1007.20137	0.00288	0.01240	−537.12802	0.00103	0.00858	−105.00340
0.00157	0.01312	−233.04374	0.01365	0.01475	−1001.28856	0.00262	0.01212	−535.78578	0.00094	0.00838	−103.87152
0.00143	0.01281	−230.91094	0.01244	0.01441	−996.35828	0.00239	0.01184	−534.45574	0.00085	0.00819	−102.93216
0.00130	0.01252	−228.90116	0.01133	0.01408	−990.95580	0.00218	0.01157	−533.07010	0.00078	0.00800	−102.15076
0.00119	0.01223	−227.41718	0.01033	0.01375	−985.64252	0.00198	0.01130	−530.52757	0.00071	0.00782	−101.43430
0.00108	0.01195	−225.78344	0.00941	0.01344	−980.48980	0.00181	0.01104	−528.57839	0.00065	0.00764	−100.80435
0.00099	0.01168	−223.37579	0.00857	0.01313	−976.30686	0.00165	0.01079	−527.08235	0.00059	0.00746	−100.26020
0.00090	0.01141	−221.22158	0.00781	0.01283	−972.78175	0.00150	0.01054	−524.73819	0.00054	0.00729	−99.79681
0.00082	0.01115	−219.26775	0.00712	0.01253	−968.99390	0.00137	0.01030	−523.34156	0.00049	0.00712	−99.38642
0.00075	0.01089	−217.48175	0.00649	0.01224	−965.91779	0.00125	0.01006	−522.16709	0.00045	0.00696	−98.98870
0.00068	0.01064	−215.75391	0.00591	0.01196	−962.96708	0.00114	0.00983	−521.11184	0.00041	0.00680	−98.62976
0.00062	0.01039	−214.26404	0.00538	0.01169	−960.01297	0.00103	0.00960	−520.18926	0.00037	0.00664	−98.31837
0.00056	0.01016	−212.85595	0.00491	0.01142	−957.23311	0.00094	0.00938	−519.40270	0.00034	0.00649	−98.04180
0.00051	0.00992	−211.55534	0.00447	0.01116	−954.60164	0.00086	0.00917	−518.62579	0.00031	0.00634	−97.79375
0.00047	0.00969	−210.34778	0.00407	0.01090	−952.10401	0.00078	0.00896	−517.91032	0.00028	0.00619	−97.56375
0.00043	0.00947	−209.11553	0.00371	0.01065	−949.81829	0.00071	0.00875	−517.24554	0.00026	0.00605	−97.33781
0.00039	0.00925	−208.11658	0.00338	0.01040	−947.74616	0.00065	0.00855	−516.62254	0.00023	0.00591	−97.14404
0.00035	0.00904	−207.01556	0.00308	0.01017	−945.84225	0.00059	0.00835	−515.88596	0.00021	0.00578	−96.98403
0.00032	0.00883	−206.13727	0.00281	0.00993	−944.04023	0.00054	0.00816	−515.15394	0.00019	0.00564	−96.84788
0.00029	0.00863	−205.25664	0.00256	0.00970	−942.36660	0.00049	0.00797	−514.52118	0.00018	0.00551	−96.72920
0.00027	0.00843	−204.42813	0.00233	0.00948	−940.77784	0.00045	0.00779	−513.62148	0.00016	0.00539	−96.62581
0.00024	0.00824	−203.55825	0.00212	0.00926	−939.31489	0.00041	0.00761	−513.17424	0.00015	0.00526	−96.52885
0.00022	0.00805	−202.95731	0.00194	0.00905	−937.95942	0.00037	0.00744	−512.69787	0.00013	0.00514	−96.45123
0.00020	0.00786	−202.20159	0.00176	0.00884	−936.70441	0.00034	0.00726	−512.24534	0.00012	0.00502	−96.37860
0.00018	0.00768	−201.66894	0.00161	0.00864	−935.53765	0.00031	0.00710	−511.80782	0.00011	0.00491	−96.31316
0.00017	0.00751	−201.09150	0.00146	0.00844	−934.44891	0.00028	0.00693	−511.38303	0.00010	0.00480	−96.25449
0.00015	0.00733	−200.49344	0.00133	0.00825	−933.41292	0.00026	0.00677	−510.80587	0.00009	0.00469	−96.20346
0.00014	0.00716	−199.99118	0.00122	0.00806	−932.43468	0.00023	0.00662	−510.43259	0.00008	0.00458	−96.15695
0.00013	0.00700	−199.49505	0.00111	0.00787	−931.51293	0.00021	0.00647	−509.95735	0.00008	0.00447	−96.11507
0.00012	0.00684	−199.02219	0.00101	0.00769	−930.69323	0.00019	0.00632	−509.32110	0.00007	0.00437	−96.07733
0.00011	0.00668	−198.57272	0.00092	0.00751	−929.82674	0.00018	0.00617	−508.22123	0.00006	0.00427	−96.04326
0.00010	0.00653	−197.94354	0.00084	0.00734	−929.01839	0.00016	0.00603	−507.28800	0.00006	0.00417	−96.01247
0.00009	0.00638	−197.41309	0.00076	0.00717	−928.24300	0.00015	0.00589	−506.23864	0.00005	0.00408	−95.98459
0.00008	0.00623	−196.72996	0.00070	0.00701	−927.54398	0.00013	0.00576	−505.23521	0.00005	0.00398	−95.95931
0.00007	0.00609	−196.15087	0.00063	0.00685	−926.79229	0.00012	0.00562	−504.28409	0.00004	0.00389	−95.93636
0.00007	0.00595	−195.58152	0.00058	0.00669	−926.16022	0.00011	0.00550	−503.38449	0.00004	0.00380	−95.91549
0.00006	0.00581	−195.04092	0.00053	0.00653	−925.47365	0.00010	0.00537	−502.30126	0.00004	0.00371	−95.89649
0.00006	0.00568	−194.53320	0.00048	0.00638	−924.89937	0.00009	0.00525	−501.65590	0.00003	0.00363	−95.87916
0.00005	0.00555	−194.11386	0.00044	0.00624	−924.26993	0.00008	0.00512	−500.89368	0.00003	0.00354	−95.86334
0.00005	0.00542	−193.68470	0.00040	0.00609	−923.74623	0.00008	0.00501	−500.16034	0.00003	0.00346	−95.84888
0.00004	0.00530	−193.27155	0.00036	0.00595	−923.15756	0.00007	0.00489	−499.46198	0.00002	0.00338	−95.83564
0.00004	0.00517	−192.73486	0.00033	0.00582	−922.68093	0.00006	0.00478	−498.57856	0.00002	0.00331	−95.82350
0.00003	0.00505	−192.44757	0.00030	0.00568	−922.07282	0.00006	0.00467	−497.89879	0.00002	0.00323	−95.81236
0.00003	0.00494	−191.95753	0.00027	0.00555	−921.53882	0.00005	0.00456	−496.74739	0.00002	0.00316	−95.80212
0.00003	0.00482	−191.69918	0.00025	0.00543	−920.84291	0.00005	0.00446	−496.11778	0.00002	0.00308	−95.74450
0.00003	0.00471	−191.37531	0.00023	0.00530	−920.27171	0.00004	0.00435	−495.30881	0.00002	0.00301	−95.66914
0.00002	0.00461	−190.95284	0.00021	0.00518	−919.62205	0.00004	0.00425	−494.51600	0.00001	0.00294	−95.55364
0.00002	0.00450	−190.72502	0.00019	0.00506	−919.08504	0.00004	0.00416	−493.75409	0.00001	0.00288	−95.43176
0.00002	0.00440	−190.43305	0.00017	0.00494	−918.40286	0.00003	0.00406	−493.02528	0.00001	0.00281	−95.24697
0.00002	0.00430	−190.14502	0.00016	0.00483	−917.77319	0.00003	0.00397	−492.07591	0.00001	0.00274	−95.17621
0.00002	0.00420	−189.73133	0.00014	0.00472	−917.02372	0.00003	0.00388	−491.55578	0.00001	0.00268	−95.00239

Computational time

The total time taken by the default method for each species is given in Table A.10.

Table A.10 Default method durations (hours).

Species	Total duration
Cod	4.5
Sea bass	4.4
Mackerel	1.9
Flounder	2.5
Plaice	3.5
Whiting	1.1
Dab	2.4
Skates/Rays	11.8

A.4.2 Results

Probability of each species being available in the fishing area

This section presents the results of the Binomial part of the default method, which models the probability the modelled species is available to catch.

Table A.11 Estimated coefficients in the Binomial part (whether or not fish were available) using the default method. A positive value indicates a higher chance of a zero, and therefore a negative association between the covariate and the availability of fish. A negative value indicates a lower chance of a zero, and therefore a positive association between the covariate and the availability of fish. A zero indicates no association between the covariate and the availability of fish. Non-zero values are displayed to two decimal places.

Default method: Binomial part	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Intercept	1.09	1.04	2.80	2.25	2.55	2.22	2.48	3.86
Type: Boat	0	0	−0.72	0	0	0	0	0
Method: Lure/fly/jig	0	0	0	0	0	0	0	0
Method: Mix	0	0	0	0	0	0	0	0
Target: Yes	0	0	−0.95	0	0	0	0	0
Hours fished in daylight	0	−0.25	−0.18	0	−0.25	−0.07	−0.19	−0.31
Hours fished in twilight	−0.12	0	0	0	0	0	0	0
Hours fished in darkness	0	0	0	0	0	−0.19	0	0
Number of other species caught	−0.08	0	−0.07	0	0	−0.11	0	−0.07
Weekend: Yes	0	0	0	0	0	0	0	0
Total anglers in group	0	0	0	0	0	0	0	0
Competition: Yes	0	0	0	0	0	0	0	0
Days shore angling	0	0	0	0	0	0	0	0
Days boat angling	0	0	0	0	0	0	0	0
Belongs to angling club: Yes	0	0	0	0	0	0	0	0
Sea: North Sea	0	0	0	0	0	−0.15	−0.36	0
Sea: Bristol Channel/Celtic Sea	0	0	0	0	0	0	0	0
Sea: Irish Sea	0	0	0	0	0	0	0	0
Season: Spring	0	0	0	0	0	0	0	0
Season: Summer	0	0	−0.03	0	0	0.02	0	0
Season: Autumn	0	0	0	0	0	−0.22	0	0
Sea state: Moderate	0	0	0	0	0	0	0	0
Sea state: Rough	0	0	0	0	0	0	0	0
Lunar illumination	0	0	0	0	0	0	0	0
Sediment: Gravel	0	0	0	0	0	0	0	0
Sediment: Rock	0	0	0	0	0	0	0	0
Sediment: Mud	0	0	0	0	0	0	0	0

For all species, very few covariates are selected by the model in the first (Binomial) part. The trip-related covariates not selected for any species are: fishing method, weekend, number of anglers in the angling group, and whether the fishing is part of a competition. None of the angler-specific covariates are selected in this part of the model for any species. The environmental covariates not selected for any species are: sea state, lunar phase, and seabed sediment. These covariates are not associated with the availability of fish of any of the modelled species.

No covariates are significant for flounder, suggesting that none impact the probability that flounder are available.

Omitting the observation with a high number of catches did not have a notable impact on the results.

Trip characteristics Mackerel are more likely to be available when fishing from a boat than when fishing from the shore. The other species are equally likely to be available for both shore and boat anglers. The choice of fishing method used (bait, lure/fly/jig or a mix of both methods) has no impact on whether each species is likely to be available in the area.

Mackerel are more likely to be available when they are being targeted than when they are not targeted. The choice of target did not impact the probability of the other species being available.

For all species except cod and flounder, the probability of the species being available increases as the duration fished during daylight increases. As the duration fished in twilight increases, cod are more likely to be available. As the duration fished during darkness increases, whiting are more likely to be available.

As the number of fish of other species caught increases, cod, mackerel, whiting and skates/rays are more likely to be available. The number of fish of other species caught has no impact on the probability of sea bass, flounder, plaice and dab being available.

None of the species are more or less likely to be available at weekends than on weekdays.

Angler characteristics Angling experience and club membership are not linked to the probability of any species being available.

Environment characteristics Whiting and dab are more likely to be available in angling areas in the North Sea than in the English Channel. None of the other species are more or less likely to be available in different regions.

Mackerel are more likely to be available in summer than in winter, and whiting are less likely in summer and more likely in autumn than in winter. The seasons are not linked to a higher or lower probability of the other species being available.

The sea state, lunar phase and sediment type are not linked to an increased or decreased likelihood of any of the species being available.

Number of fish caught, given they are available

This section presents the results of the Poisson part of the default method, which models the number of fish caught of the modelled species, given that the species is available.

Table A.12 Estimated coefficients in the Poisson part (number of fish caught) using the default method. A negative value indicates the covariate is negatively associated with the number of fish caught, a positive value indicates the covariate is positively associated with the number of fish caught and zero indicates there is no association between the covariate and the number of fish caught. Non-zero values are displayed to two decimal places.

Default method: Poisson part	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Intercept	−4.85	−3.82	1.93	−0.43	−0.83	1.43	0.62	0.43
Type: Boat	1.24	0.57	0	0	0.81	0.34	0.25	0
Method: Lure/fly/jig	0.50	0.35	0	0	0	0	0	0
Method: Mix	1.18	1.05	0	0	0.65	0	0	0
Target: Yes	1.85	0.62	0	0	0.63	0.49	0	0
Hours fished in daylight	0.72	0.50	0.19	0.53	0.62	0.06	0.25	0
Hours fished in twilight	−0.13	−0.02	0	0.27	0	0	0	0
Hours fished in darkness	0.22	−0.06	0	0	0	0.05	0	0
Number of other species caught	−0.09	−0.14	0.06	−0.45	−0.02	0.12	0.25	0.29
Weekend: Yes	0.69	0.02	0	−0.37	0	0	−0.04	0
Total anglers in group	0.03	−0.53	0.21	0.30	0.07	0.11	0	−0.03
Competition: Yes	1.01	−0.95	0	0	0	0	0	0
Days shore angling	0.15	−0.14	0	−0.06	−0.16	−0.10	−0.20	0
Days boat angling	−0.26	0.15	0	−0.03	−0.29	0.13	0.02	0
Belongs to angling club: Yes	−0.11	0.80	−0.03	0	1.02	0	0.90	0
Sea: North Sea	2.32	0.15	0	0.35	−0.20	0	0	0
Sea: Bristol Channel/Celtic Sea	2.25	−1.29	0	0	0	0	0	0
Sea: Irish Sea	3.55	0.96	0	0.83	0	0	0	0
Season: Spring	−0.59	1.79	0	0	0	0	−0.19	0
Season: Summer	0.21	2.18	0	0	0.07	0	0	0
Season: Autumn	−0.58	2.11	0	0.31	−0.59	0	0	0
Sea state: Moderate	0.11	−0.18	0	0	−0.39	0	−1.06	0
Sea state: Rough	−0.19	0.84	0	0	0	0	0	0
Lunar illumination	−0.15	0.10	0.19	−0.07	0.26	−0.05	0.39	0
Sediment: Gravel	−0.52	−0.58	0	0	0	0.07	0.67	0
Sediment: Rock	−1.47	−0.70	0	0	0	0	0	0
Sediment: Mud	1.02	−0.80	0	0	0	0	0	0

Trip characteristics The model predicts that boat anglers catch more cod, sea bass, plaice, whiting and dab per angler than shore anglers, if the species if available. For mackerel, flounder and skates/rays there is no difference in the predicted number of fish caught between shore and boat anglers. Anglers using lure/fly/jig methods are predicted to catch more cod and sea bass than anglers using bait, and anglers using both bait and lure/fly/jig methods are predicted to catch more cod, sea bass and plaice than anglers using bait only.

Anglers targeting cod, sea bass, plaice and whiting are predicted to catch more fish of the species targeted than anglers not targeting the species, if the species is available. Anglers targeting mackerel, flounder, dab and skates/rays are not predicted to catch any more or less fish of the targeted species than anglers not targeting the species.

The number of fish caught per angler increases with the duration fished during daylight for all species except skates/rays. The number of flounder caught increases and the number of cod and sea bass caught decreases as the duration fished during twilight increases. The number of cod and whiting increases and the number of sea bass decreases as the duration fished during darkness decreases.

When available, more mackerel, whiting, dab and skates/rays, and fewer cod, sea bass, flounder and plaice are predicted to be caught as the number of fish of other species caught increases. Anglers fishing at weekends are predicted to catch more cod and sea bass, and fewer flounder and dab than anglers fishing on weekdays, if the species is available. As the number of anglers in the group increases, the number of sea bass and skates/rays caught per angler decreases, while the catch rates of the other species increase, except for dab, for which the group size has no impact on the number of fish caught. Anglers fishing in a competition are predicted to catch more cod and fewer sea bass than anglers not fishing in a competition.

Angler characteristics Anglers with more shore angling experience are predicted to catch more cod and fewer sea bass, flounder, plaice, whiting, and dab than anglers with less shore angling experience. Anglers with more boat angling experience are predicted to catch more sea bass, whiting, and dab and fewer cod, flounder and plaice than anglers with less boat angling experience. Anglers belonging to an angling club are predicted to catch more sea bass, plaice and dab and fewer cod and mackerel than anglers not belonging to an angling club.

Environment characteristics Anglers fishing in the North Sea are predicted to catch more cod, sea bass and flounder than anglers fishing in the English Channel, if these species are available in the fishing area. Anglers fishing in the Bristol Channel/Celtic Sea area are predicted to catch more cod and fewer sea bass, while anglers fishing in the Irish Sea are predicted to catch more cod, sea bass, and flounder than anglers fishing in the English Channel, given these species are available in the area.

Anglers are predicted to catch more sea bass in spring, summer and autumn than in winter, given sea bass are available. It is predicted that anglers will catch more cod in summer and fewer in spring and autumn compared to in winter, given cod are available. Anglers are predicted to catch more plaice in summer and fewer in autumn, more flounder in autumn, and fewer dab in spring, compared to in winter, given these species are available.

Sea bass, plaice, and dab are less likely, and cod more likely, to be caught in moderate sea conditions compared to in calm sea conditions. Sea bass are more likely, and cod less likely, to be caught in rough than in calm sea conditions, given that the species is available at the time of fishing.

More sea bass, mackerel, plaice and dab, and fewer cod, flounder, and whiting are predicted to be caught as lunar illumination increases, given the species is available. Catch rates for skates/rays are not impacted by lunar illumination.

The number of cod and sea bass caught per angler is lower in gravel and rock sediment areas, compared to areas with sand sediment. The number of cod caught is higher, and the number of sea bass is lower, in mud sediment areas than in areas with sand sediment. The number of whiting and dab caught is higher in gravel sediment areas than in sand sediment areas.

Comparing observed and predicted catches

Figures A.32 (cod, sea bass, mackerel and flounder) and A.33 (plaice, whiting, dab and skates/rays) compare the recorded and predicted number of catches of the target species for each interview, using the default method.

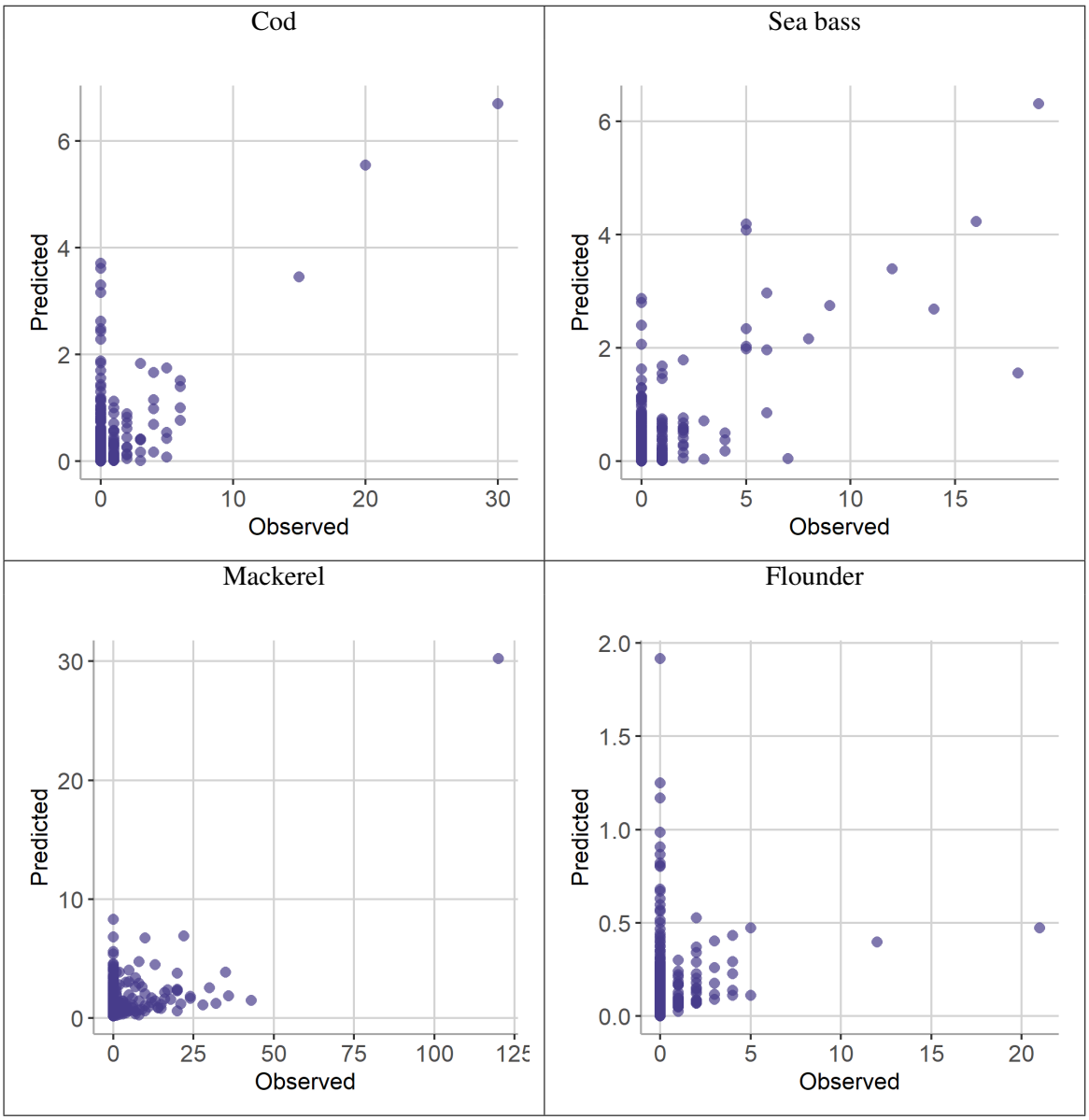


Fig. A.32 Number of observed and predicted catches using the default method for each interview. Species shown: cod, sea bass, mackerel and flounder.

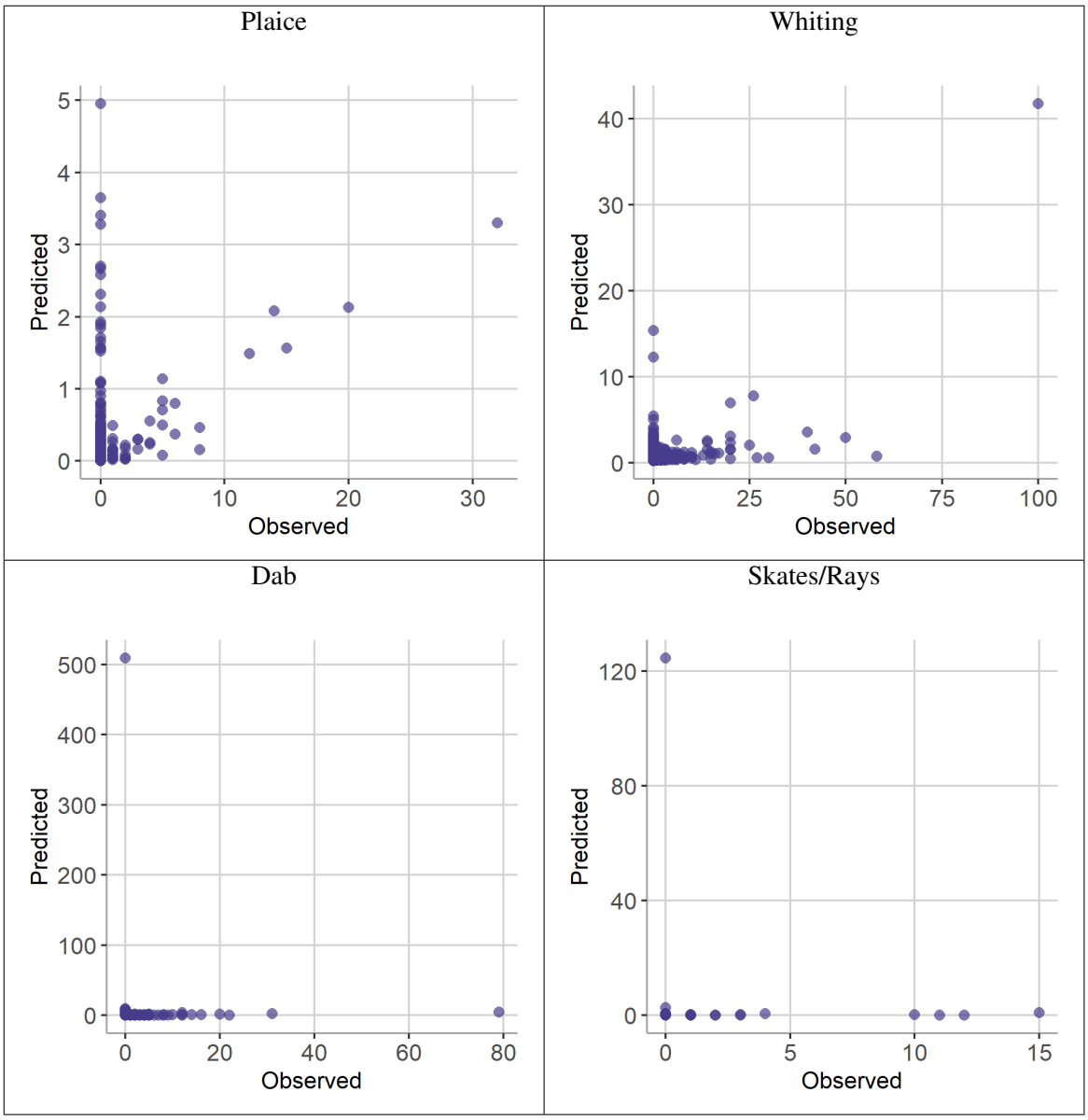
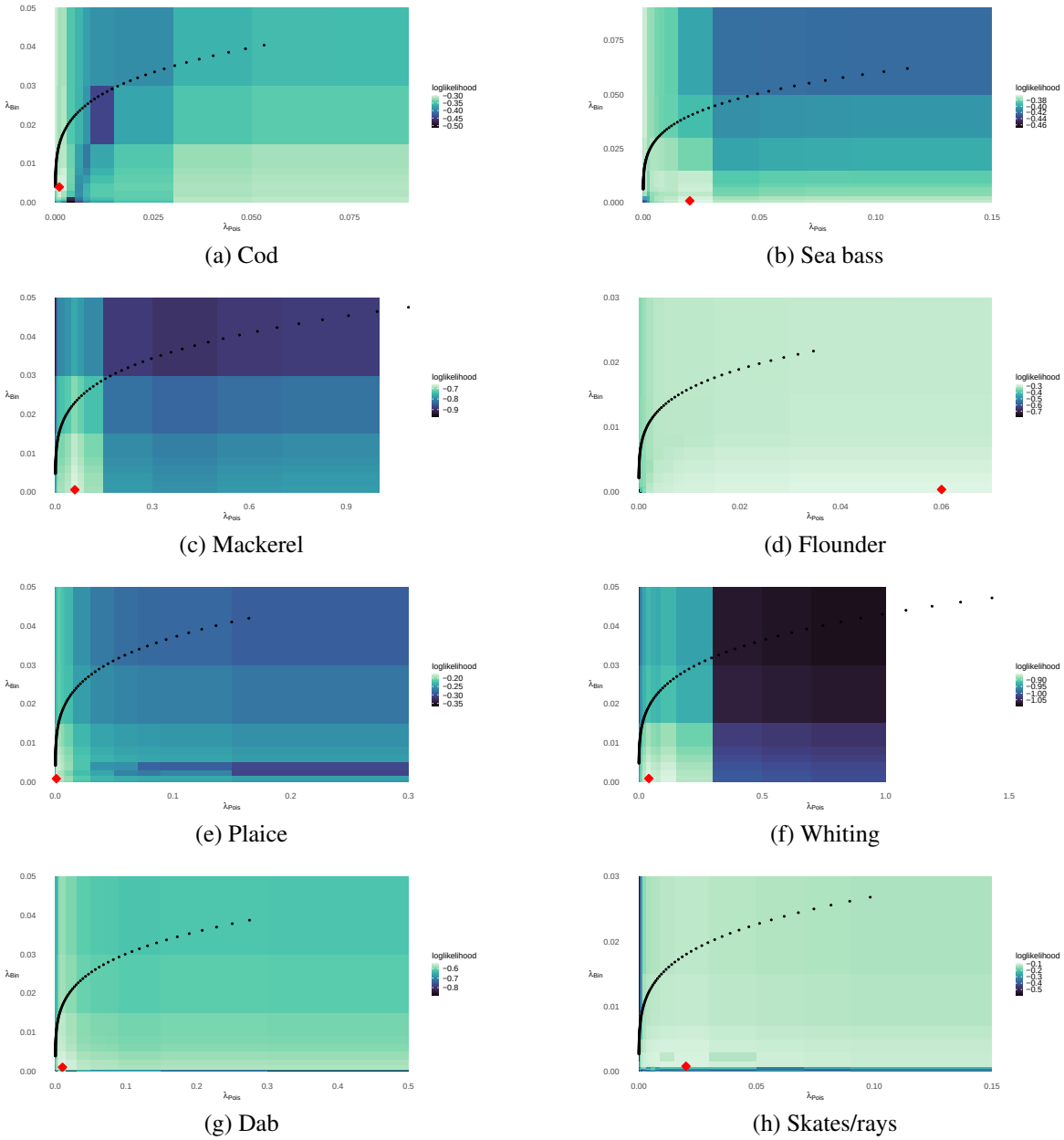


Fig. A.33 Number of observed and predicted catches using the default method for each interview. Species shown: plaice, whiting, dab and skates/rays.

A.5 Comparing the grid and default methods

A.5.1 Loglikelihood values

Figure A.34 compares the lambda values used by the default method with the lambda values and corresponding loglikelihood obtained with the grid method, for each species.



(a) Cod

(b) Sea bass

(c) Mackerel

(d) Flounder

(e) Plaice

(f) Whiting

(g) Dab

(h) Skates/rays

Fig. A.34 Grids with λ values used in the default method. The red diamond indicates the maximum loglikelihood found by the grid method. λ values for the Poisson (count) process are on the x-axis. λ values for the Binomial (zero) process are on the y-axis.

A.5.2 Predicted catches

Figures A.35 (cod, sea bass, mackerel, flounder) and A.36 (plaice, whiting, dab, skates/rays) compare the number of catches of the target species predicted by the grid and default methods, for each interview.

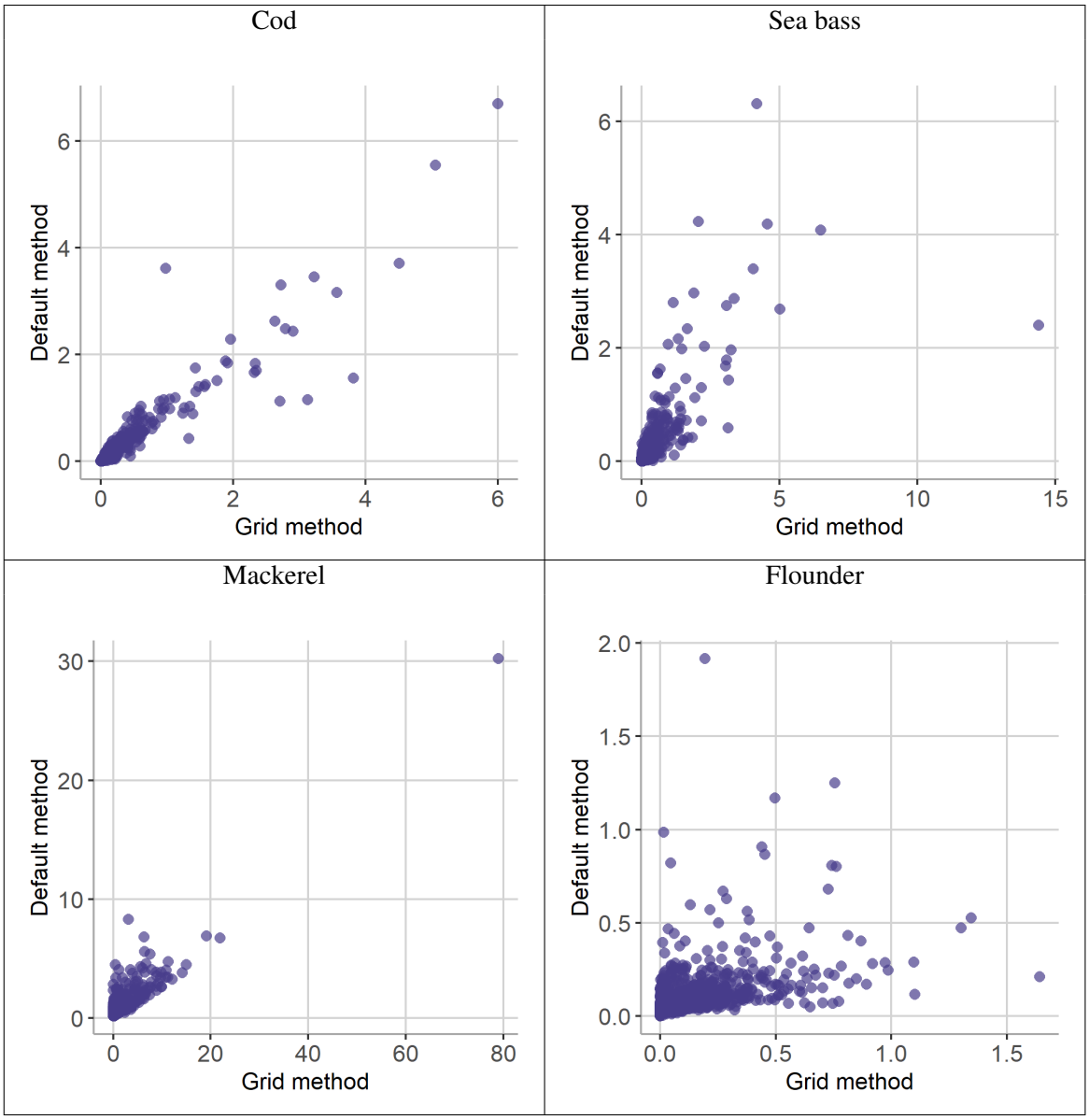


Fig. A.35 Number of predicted catches for each interview using the grid and default methods. Species shown: cod, sea bass, mackerel and flounder.

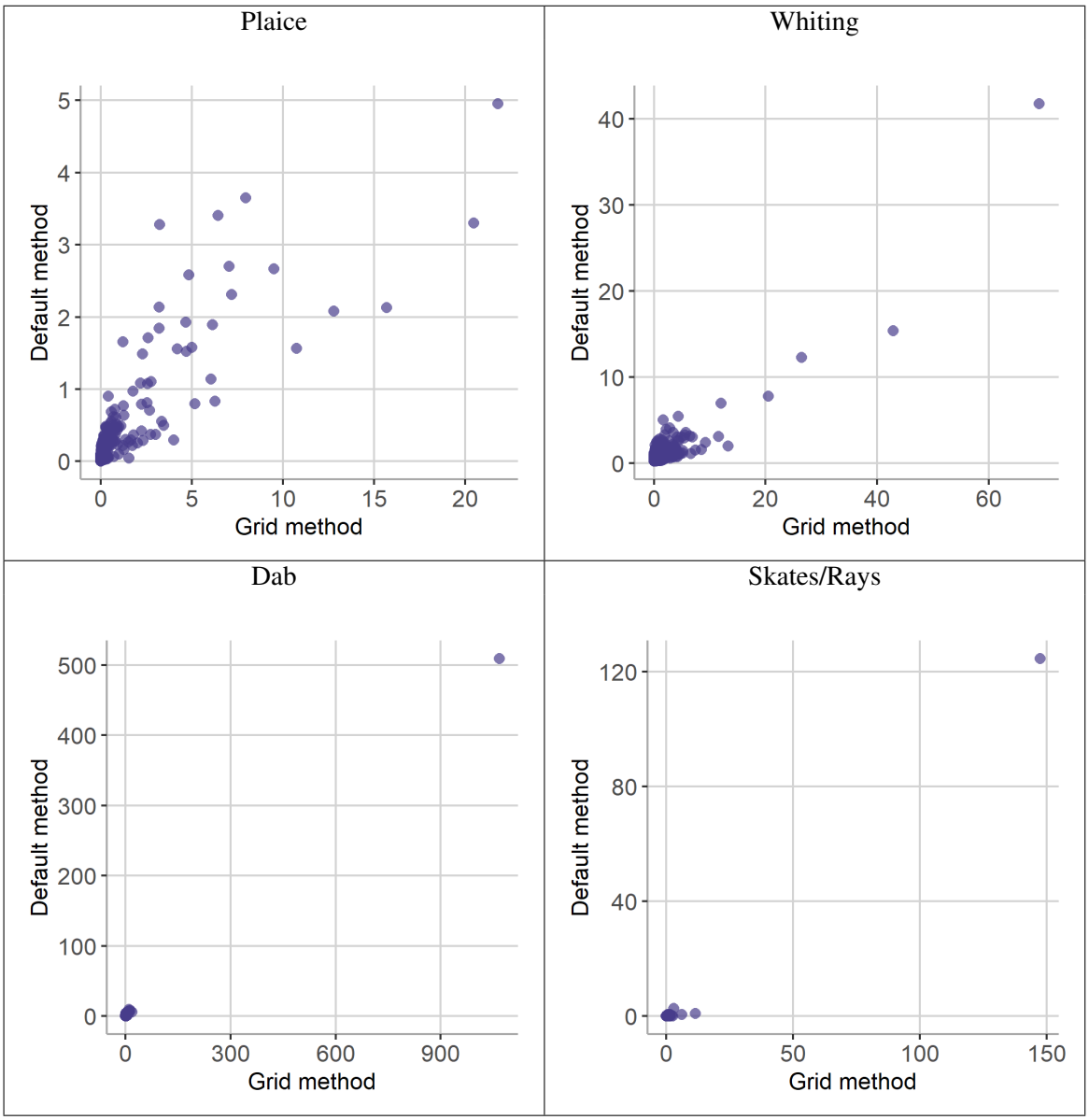


Fig. A.36 Number of predicted catches for each interview using the grid and default methods. Species shown: plaice, whiting, dab and skates/rays.

The total number of observed and predicted catches for each species is given in Table A.13.

Table A.13 Total number of observed catches and total number of predicted catches with the grid and default methods.

	Cod	Sea bass	Mackerel	Flounder	Plaice	Whiting	Dab	Skates/Rays
Observed	201	241	989	147	200	1008	409	76
Predicted - Grid	176	247	997	147	309	1047	1489	219
Predicted - Default	165	183	881	124	141	833	904	177

For cod, sea bass, mackerel, flounder and whiting, the total number of predicted catches is closer to the observed total catches with the grid method results compared to the default method results. For plaice, dab and skates/rays, the total number of predicted catches is closer to the observed number of catches by the default method.

Appendix B

Spatiotemporal models for mobile app records of fish catches – Supplementary material

B.1 R code

The R code used to prepare the data, fit the models, and generate predictions is provided in Figures B.1, B.2, and B.3.

```
# Set up

library(tidyverse)
library(raster)
library(INLA)
library(inlabru)

bru_options_set(bru_verbose = TRUE,
                control.compute = list(dic = TRUE, waic = TRUE, cpo=TRUE),
                control.inla = list(int.strategy="eb"))

colnames(fbdata)
# "location_lat" "location_lng"
# "quarter" "Q2" "Q3" "Q4"
# "timePeriod"
# "users"   # lower bound for the number of users
# "Dicentrarchus.labrax" "Gadus.morhua"
# "Scomber.scombrus" "Merlangius.merlangus"

fbdata$timePeriod <- fbdata$timePeriod - min(fbdata$timePeriod) + 1

timePeriod <- fbdata$timePeriod
ntimePeriod <- length(unique(fbdata$timePeriod))

data_sp <-
  sp::SpatialPointsDataFrame(
    coords = fbdata %>% dplyr::select(location_lng, location_lat),
    data = fbdata %>% dplyr::select(timePeriod,
                                    year,
                                    quarter, Q2, Q3, Q4,
                                    count = "Gadus.morhua",
                                    users),
    proj4string = CRS(SRS_string = "EPSG:4326")
  )
data_sp <- sp::spTransform(data_sp, CRS(SRS_string = "EPSG:3857"))

mesh <-
  inla.mesh.2d(
    boundary = inla.sp2segment(coastline),
    max.edge = c(100000, 500000),   # largest triangle edge length
    cutoff = 30000,   # min distance between points
    crs = CRS(SRS_string = "EPSG:3857")
  )
```

Fig. B.1 R code to prepare the data.

```
# Model

# model type
model_type <- "iid"   # "iid" or "ar1"

matern <-
  inla.spde2.pcmatern(mesh,
                      prior.sigma = c(50, 0.5),
                      prior.range = c(500, 0.5))

cmp1 <- count ~
  field(main = coordinates,
        model = matern,
        group = timePeriod,
        group_mapper = bru_mapper_index(ntimePeriod),
        control.group = list(model = model_type)
  ) +
  Intercept(1) +
  Q2 + Q3 + Q4 +
  myoffset(log(data_sp$users), model = "offset")

bru_model1 = bru(cmp1,
                 family = "zeroinflatedpoisson1",
                 data = data_sp,
                 options = list(verbose = TRUE))
```

Fig. B.2 R code to run the model.

```
# Predictions

# CRS of 'coastline' shapefile is EPSG:3857
# library(raster) for this operation
ring <-
  rgeos::gBuffer(coastline,
                 width = NISTunits::NISTmileNauticalTOmeter(12)) -
  rgeos::gBuffer(coastline, width = -1000)

ring <- sp::spTransform(ring, CRS(SRS_string = "EPSG:3857"))

n_samples <- 1000

pxl <- pixels(mesh1, nx = 300, ny = 300, mask = ringM)
pxl <- cprod(data.frame(timePeriod = seq_len(ntimePeriod)), pxl)

predictions_Q1 <-
  predict(bru_model1,
          pxl,
          formula = ~ exp(Intercept + field),
          n.samples = n_samples)

predictions_Q2 <-
  predict(bru_model1,
          pxl,
          formula = ~ exp(Intercept + field + Q2),
          n.samples = n_samples)

predictions_Q3 <-
  predict(bru_model1,
          pxl,
          formula = ~ exp(Intercept + field + Q3),
          n.samples = n_samples)

predictions_Q4 <-
  predict(bru_model1,
          pxl,
          formula = ~ exp(Intercept + field + Q4),
          n.samples = n_samples)
```

Fig. B.3 R code to generate predictions from the model.

B.2 Rate of catches reported

The rate of recorded catches for each species in each quarter over the four-year period 2018-2021 at each location is shown in Figures B.4 to B.7, where the rate of recorded catches at a location in a quarter is the number of catches recorded at that location in that quarter divided by the lower bound on the number of app users (which we define in Section 3.2.2).

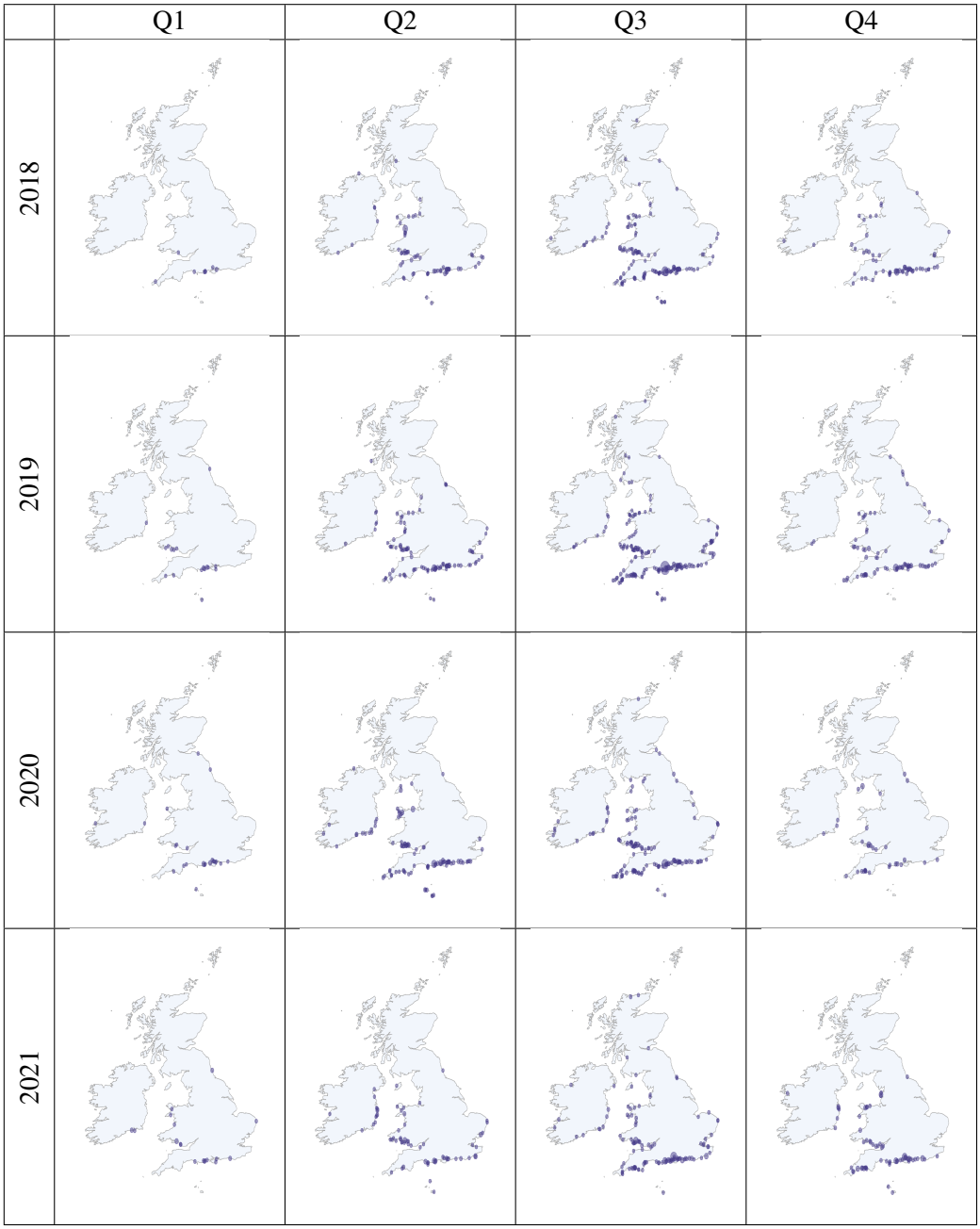


Fig. B.4 Rate of catches reported for sea bass in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.21. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December).

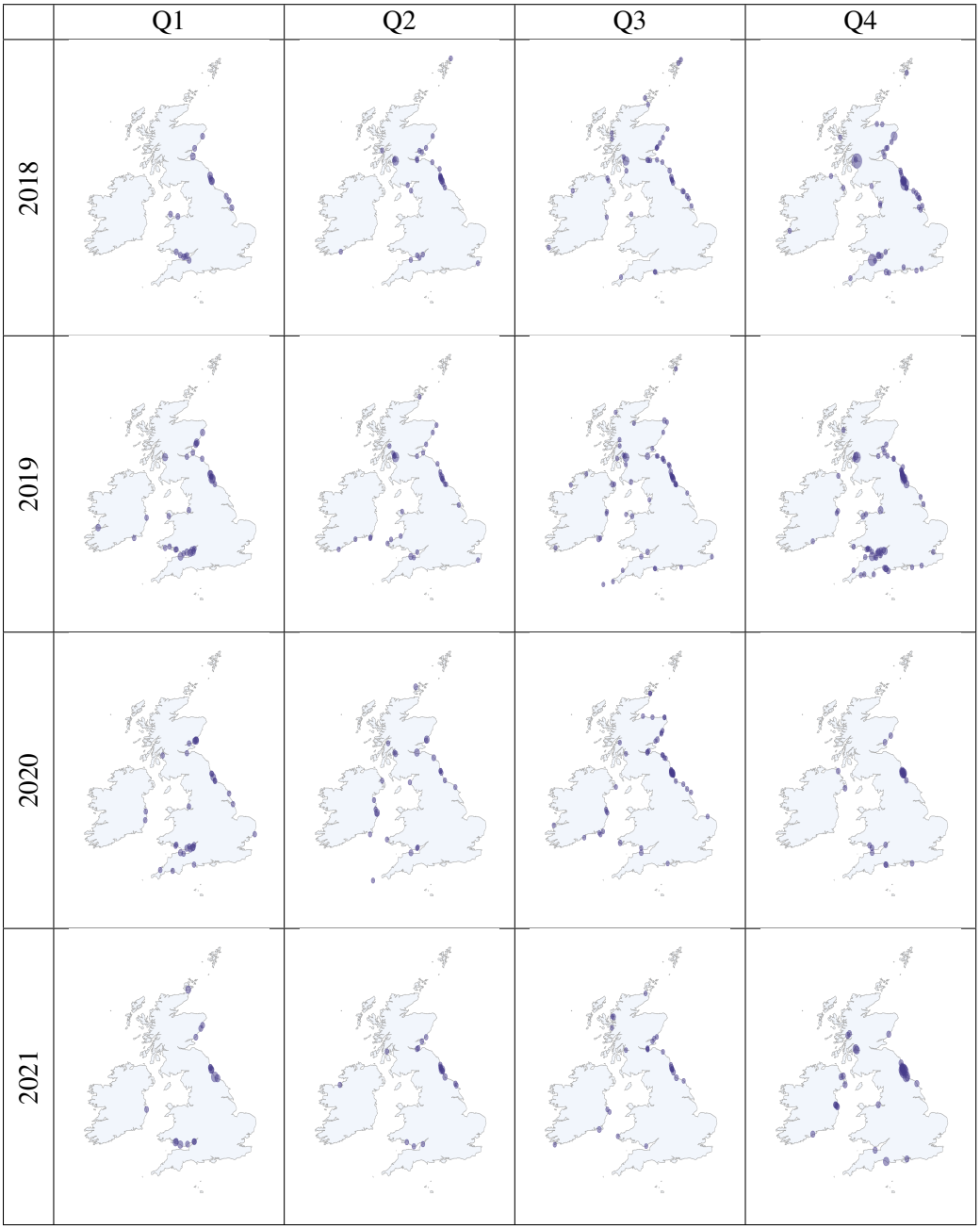


Fig. B.5 Rate of catches reported for cod in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.034. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December).

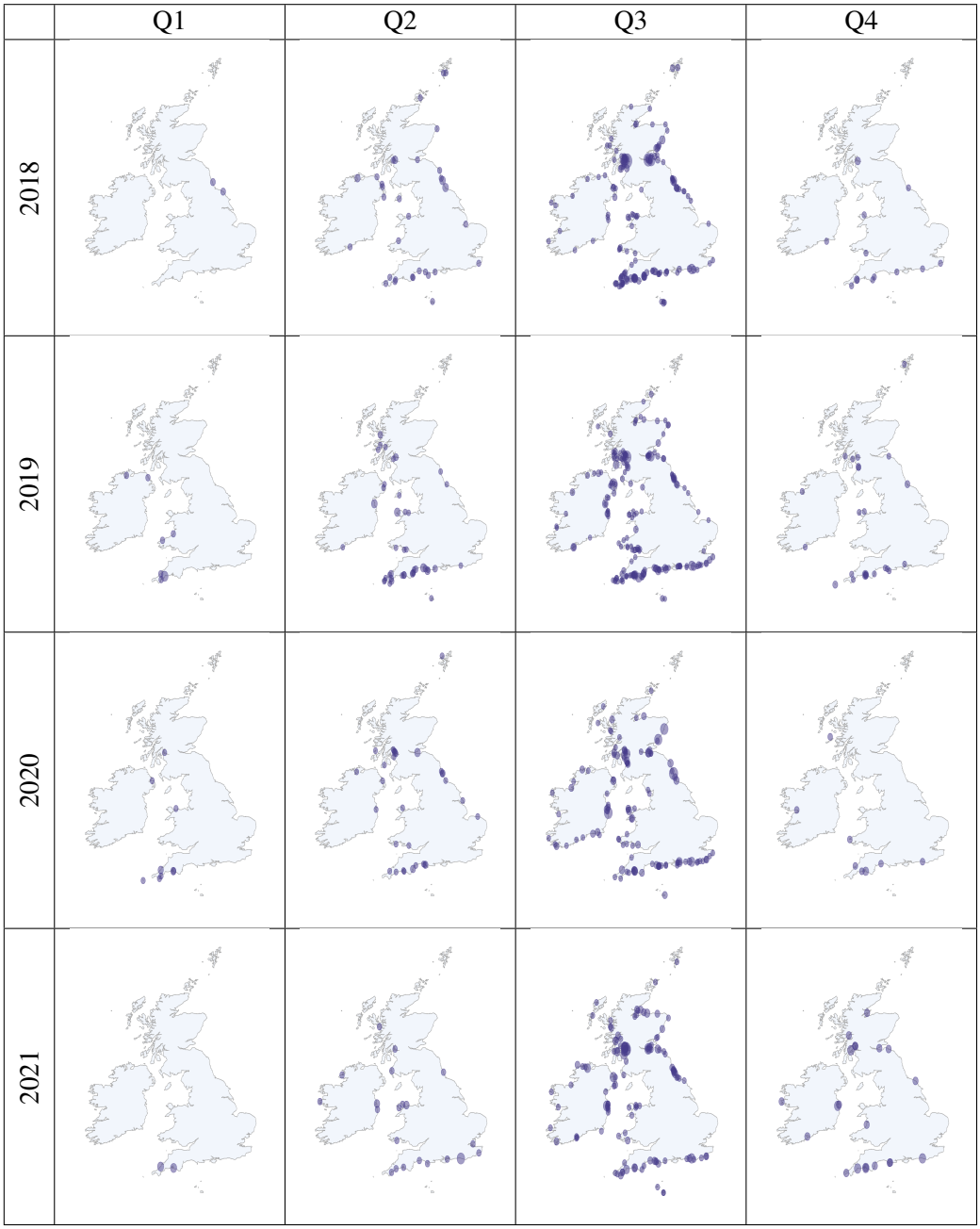


Fig. B.6 Rate of catches reported for mackerel in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.017. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December).

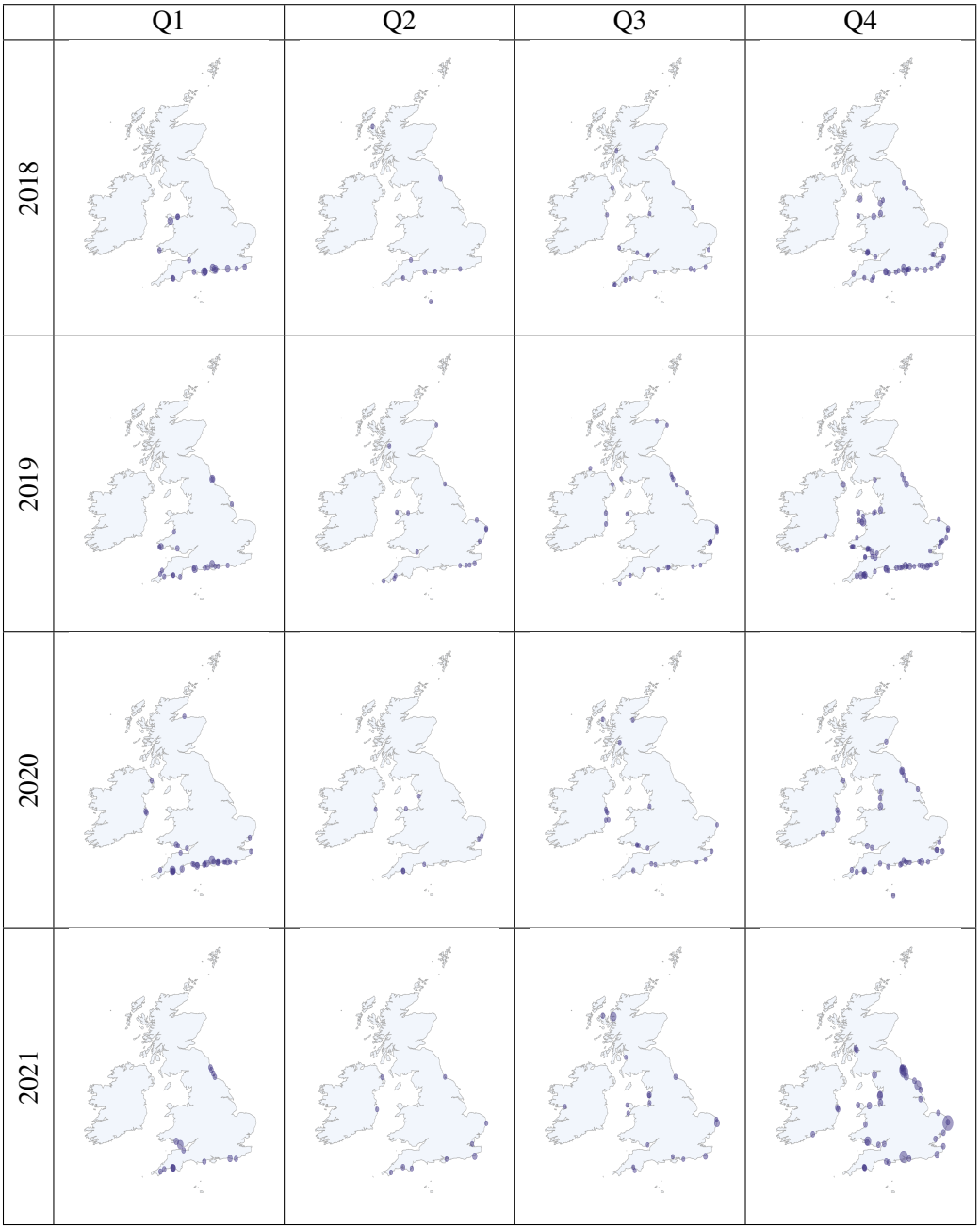


Fig. B.7 Rate of catches reported for whiting in the years 2018 to 2021 at the locations in the Fishbrain app. The highest rate is 0.054. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December).

B.3 Computational time

The computational time required to fit the models is influenced by the density of triangles in the mesh and the model complexity. Using the mesh shown in Figure 3.2 for both the iid and ar1 models, the time taken to fit the iid models ranged from one to four minutes depending on the species, while the ar1 models took approximately ten minutes for sea bass, cod, and mackerel, and nearly three hours for whiting. The mesh consisted of 921 triangles, providing a balance between computational efficiency and spatial resolution. The relative computational time required to fit the iid model for each species does not necessarily predict the relative time required to fit the ar1 model for that species. For example, for cod, the time taken to fit the ar1 model was approximately three times longer than the time taken to fit the iid model, while for whiting, it was over 40 times longer. Table B.1 presents the computational times (in minutes) required to fit the iid and ar1 models for each species.

Table B.1 Computational time (minutes) to fit the iid and ar1 models for each species.

Species	iid	ar1
Sea bass	4.3	11.2
Cod	1.0	8.4
Mackerel	0.9	13.6
Whiting	3.8	160.5

The computational time required to generate the predictions was consistent across quarters and species, ranging from four to five minutes per quarter. Table B.2 shows the time taken (in minutes) to generate predictions for each species using the ar1 model.

Table B.2 Computational time (minutes) to generate predictions from the ar1 model. Predictions for each quarter are based on data from the years 2018 to 2021 inclusive. Values are rounded to one decimal place. The total is the sum of quarters 1 to 4. Totals may not equal the sum of the quarterly values due to rounding.

	Quarter				
Species	1	2	3	4	Total
Sea bass	4.4	4.3	4.2	4.3	17.3
Cod	4.7	4.6	4.6	4.6	18.5
Mackerel	4.6	4.7	4.6	4.6	18.4
Whiting	5.0	4.7	4.7	4.6	19.1

Model fitting and prediction generation were performed using R version 4.2.2 (R Core Team, 2022) with the `INLA` (version 22.05.07) (Rue et al., 2009) and `inlabru` (version 2.7.0) (Bachl et al., 2019) packages on a personal computer running Windows 10.

B.4 Predictions

The predicted rates for each species, by quarter, for the years 2018 to 2021 are shown in Figures B.8 to B.11.

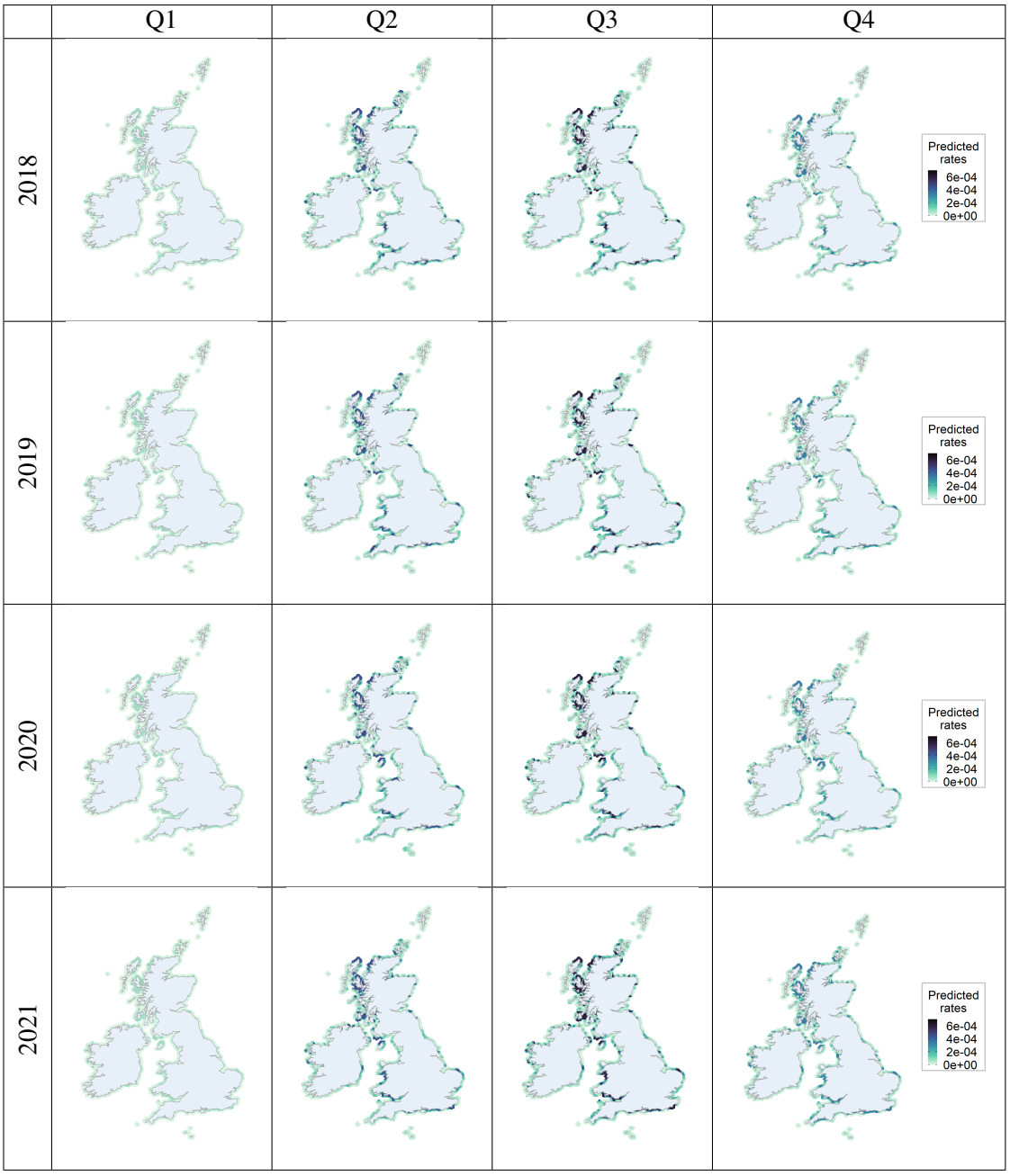


Fig. B.8 Predicted number of catches of sea bass per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear.

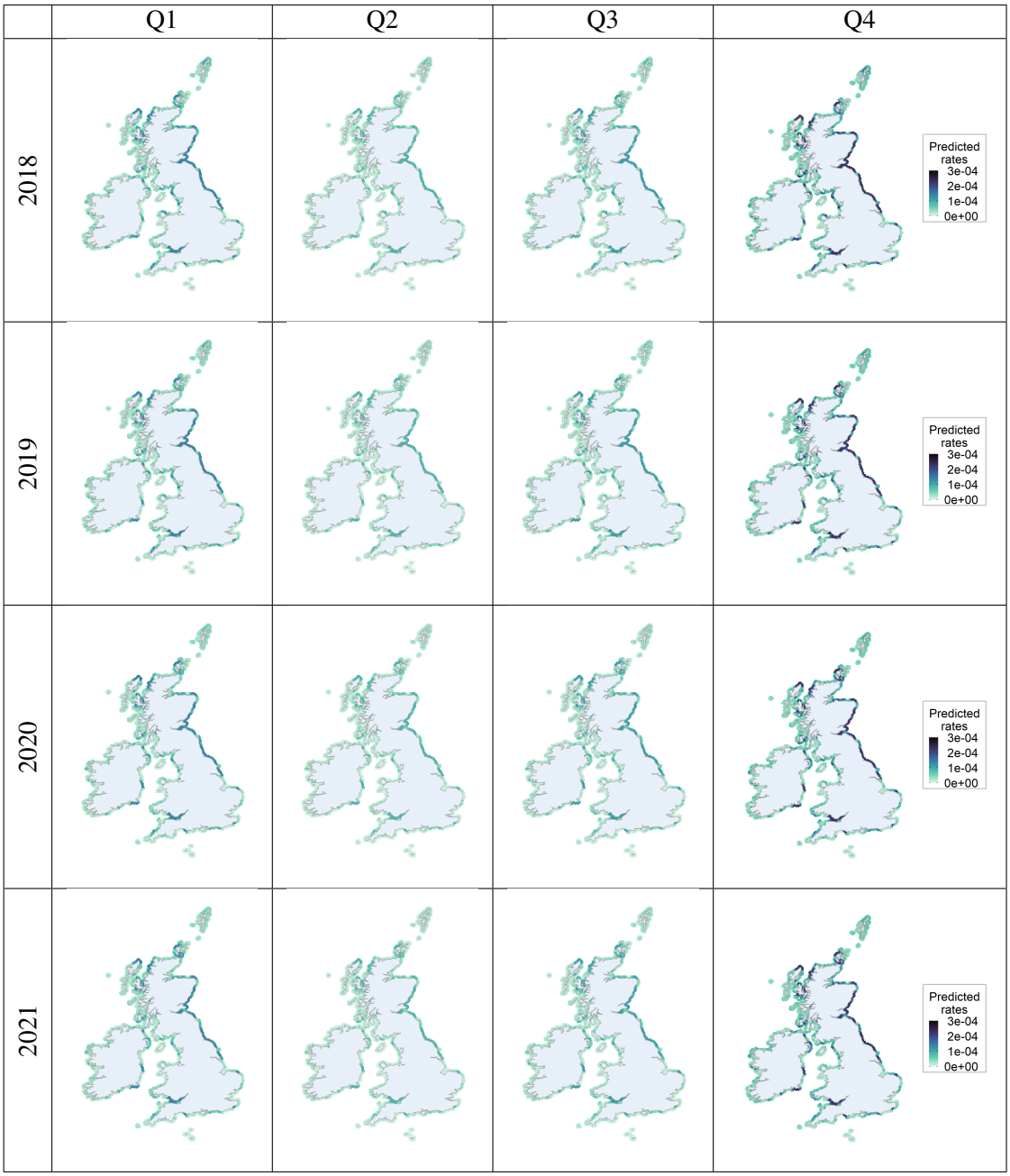


Fig. B.9 Predicted number of catches of cod per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear.

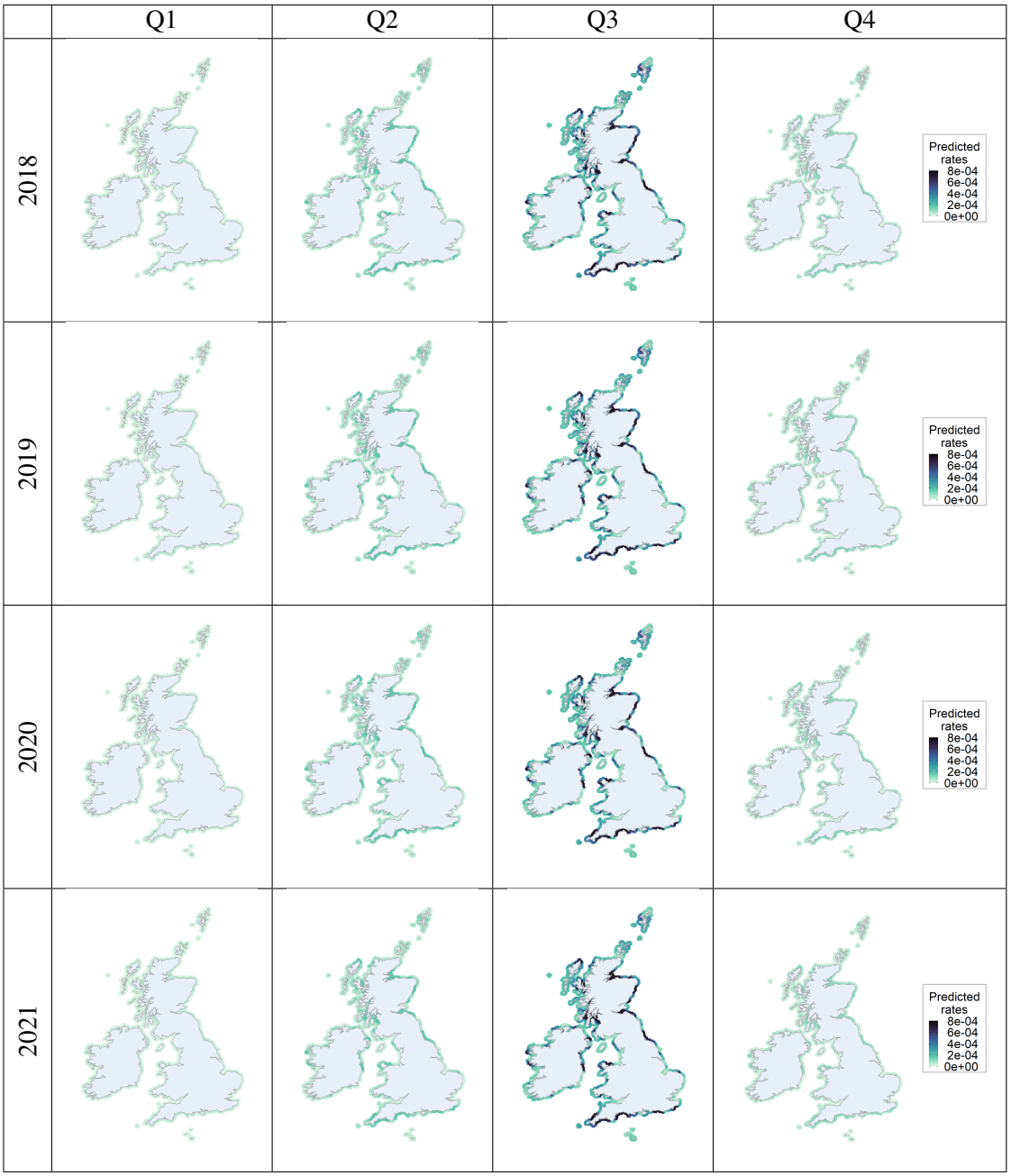


Fig. B.10 Predicted number of catches of mackerel per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear.

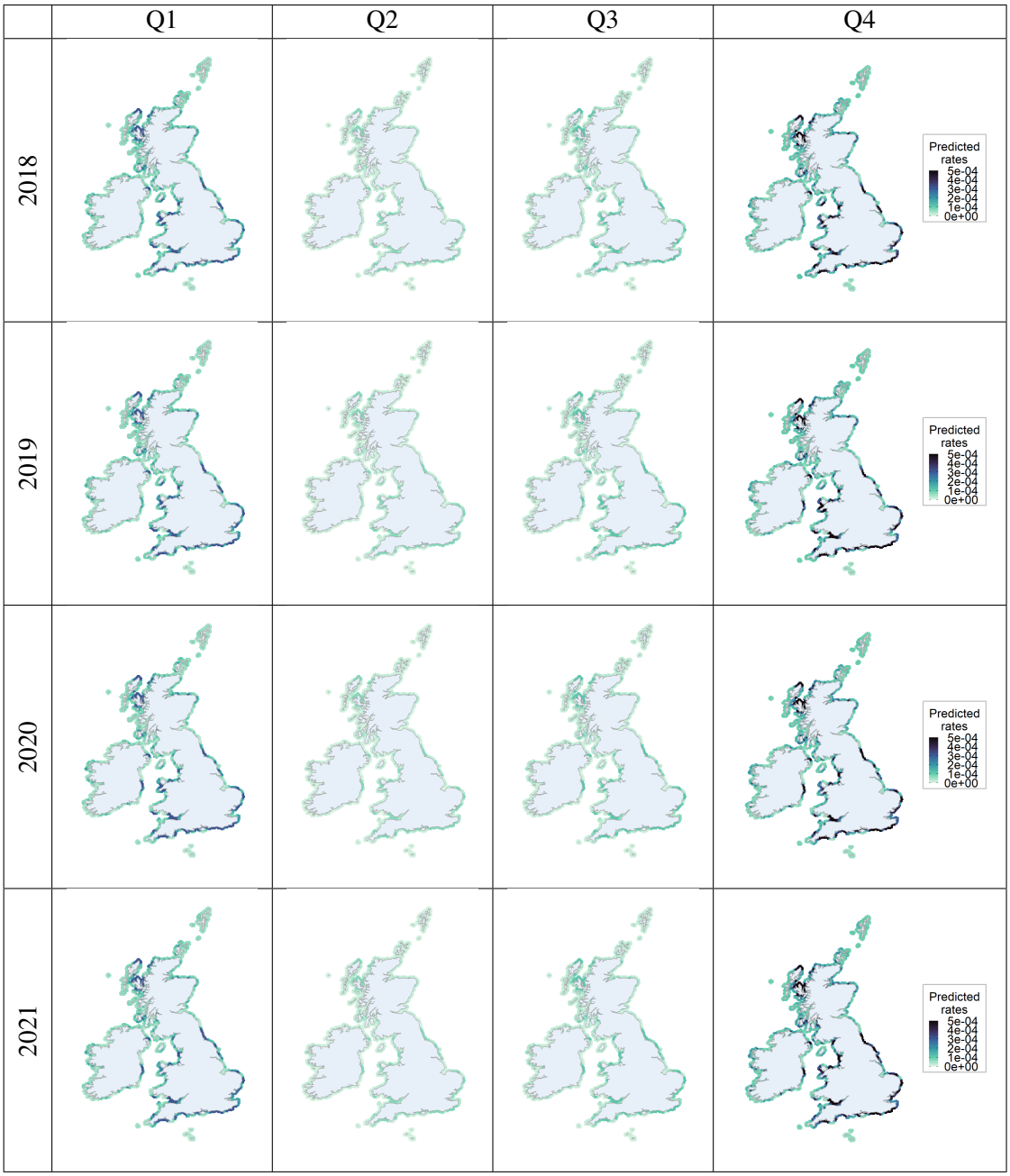


Fig. B.11 Predicted number of catches of whiting per app user in the years 2018 to 2021 around the coast of the UK and Ireland. The first column is quarter 1 (January to March), the second column is quarter 2 (April to June), the third column is quarter 3 (July to September) and the fourth column is quarter 4 (October to December). Due to the highest predicted rates being extreme values, the top 10% of predicted rates are capped at the 90% quantile to enable the spatial patterns in the predicted rates to be clear.

Appendix C

Global travel patterns in angling – Supplementary material

C.1 Country codes

This section contains a reference table of the International Organization for Standardization (ISO) alpha-3 country codes (ISO 3166) used in this research, listed alphabetically by country code. The table includes corresponding country names and World Bank regions (World Bank, 2018), obtained using the R package `countrycode` (Arel-Bundock et al., 2018).

Table C.1 Country codes and regions, listed alphabetically by country code (Arel-Bundock et al., 2018).

Country code	Country name	Region	Country code	Country name	Region
AFG	Afghanistan	South Asia	ISL	Iceland	Europe & Central Asia
ALA	Aland Islands	Europe & Central Asia	ITA	Italy	Europe & Central Asia
ALB	Albania	Europe & Central Asia	JEY	Jersey	Europe & Central Asia
AND	Andorra	Europe & Central Asia	JPN	Japan	East Asia & Pacific
ARG	Argentina	Latin America & Caribbean	KAZ	Kazakhstan	Europe & Central Asia
ATG	Antigua and Barbuda	Latin America & Caribbean	KGZ	Kyrgyzstan	Europe & Central Asia
AUS	Australia	East Asia & Pacific	KOR	Republic of Korea	East Asia & Pacific
AUT	Austria	Europe & Central Asia	LIE	Liechtenstein	Europe & Central Asia
BEL	Belgium	Europe & Central Asia	LKA	Sri Lanka	South Asia
BGR	Bulgaria	Europe & Central Asia	LTU	Lithuania	Europe & Central Asia
BHS	Bahamas	Latin America & Caribbean	LUX	Luxembourg	Europe & Central Asia
BIH	Bosnia and Herzegovina	Europe & Central Asia	LVA	Latvia	Europe & Central Asia
BLR	Belarus	Europe & Central Asia	MCO	Monaco	Europe & Central Asia
BLZ	Belize	Latin America & Caribbean	MDA	Moldova	Europe & Central Asia
BMU	Bermuda	North America	MDV	Maldives	South Asia
BRA	Brazil	Latin America & Caribbean	MEX	Mexico	Latin America & Caribbean
BRB	Barbados	Latin America & Caribbean	MKD	Republic of North Macedonia	Europe & Central Asia
CAN	Canada	North America	MNE	Montenegro	Europe & Central Asia
CHE	Switzerland	Europe & Central Asia	NGA	Nigeria	Sub-Saharan Africa
CHN	China	East Asia & Pacific	NLD	Netherlands	Europe & Central Asia
COL	Colombia	Latin America & Caribbean	NOR	Norway	Europe & Central Asia
CPV	Cabo Verde	Sub-Saharan Africa	NZL	New Zealand	East Asia & Pacific
CRI	Costa Rica	Latin America & Caribbean	PAN	Panama	Latin America & Caribbean
CUB	Cuba	Latin America & Caribbean	PER	Peru	Latin America & Caribbean
CYP	Cyprus	Europe & Central Asia	PHL	Philippines	East Asia & Pacific
CZE	Czechia	Europe & Central Asia	POL	Poland	Europe & Central Asia
DEU	Germany	Europe & Central Asia	PRI	Puerto Rico	Latin America & Caribbean
DNK	Denmark	Europe & Central Asia	PRT	Portugal	Europe & Central Asia
DOM	Dominican Republic	Latin America & Caribbean	ROU	Romania	Europe & Central Asia
EGY	Egypt	Middle East & North Africa	RUS	Russian Federation	Europe & Central Asia
ESH	Western Sahara	Sub-Saharan Africa	SGP	Singapore	East Asia & Pacific
ESP	Spain	Europe & Central Asia	SJM	Svalbard and Jan Mayen	Europe & Central Asia
EST	Estonia	Europe & Central Asia	SRB	Serbia	Europe & Central Asia
FIN	Finland	Europe & Central Asia	SVK	Slovakia	Europe & Central Asia
FLK	Falkland Islands	Latin America & Caribbean	SVN	Slovenia	Europe & Central Asia
FRA	France	Europe & Central Asia	SWE	Sweden	Europe & Central Asia
FRO	Faroe Islands	Europe & Central Asia	SYR	Syrian Arab Republic	Middle East & North Africa
GBR	United Kingdom	Europe & Central Asia	THA	Thailand	East Asia & Pacific
GEO	Georgia	Europe & Central Asia	TUR	Turkey	Europe & Central Asia
GGY	Guernsey	Europe & Central Asia	UKR	Ukraine	Europe & Central Asia
GIB	Gibraltar	Europe & Central Asia	UMI	United States Minor Outlying Islands	East Asia & Pacific
GRC	Greece	Europe & Central Asia	URY	Uruguay	Latin America & Caribbean
GRL	Greenland	Europe & Central Asia	USA	United States of America	North America
GUM	Guam	East Asia & Pacific	UZB	Uzbekistan	Europe & Central Asia
HRV	Croatia	Europe & Central Asia	VAT	Holy See	Europe & Central Asia
HUN	Hungary	Europe & Central Asia	VIR	Virgin Islands (U.S.)	Latin America & Caribbean
IDN	Indonesia	East Asia & Pacific	VNM	Viet Nam	East Asia & Pacific
IMN	Isle of Man	Europe & Central Asia	ZAF	South Africa	Sub-Saharan Africa
IOT	British Indian Ocean Territory	South Asia	ZWE	Zimbabwe	Sub-Saharan Africa
IRL	Ireland	Europe & Central Asia			

C.2 Regions manually allocated

This section presents information about the countries that could not be automatically assigned to a World Bank region using the countrycode() function from the `countrycode` package Arel-Bundock et al. (2018) in R. It contains two tables: one for origin countries and another for destination countries. Each table lists the countries without an automatic region match, the region they were manually assigned to, and the number of catches recorded for each country.

These countries were manually allocated to the closest region, by reference to a World Bank regions map (World Bank, 2018). Due to the geographic spread of the French Southern Territories, the latitude and longitude coordinates of each catch location were used to determine the appropriate region of individual catches, as some islands are closer to Sub-Saharan Africa while others are nearer to Antarctica.

Table C.2 (a) Origin and (b) destination countries that did not have a matching region using the countrycode() function from the `countrycode` package Arel-Bundock et al. (2018) in R, the region to which these countries are allocated, and the number of catch records.

(a) Origin country

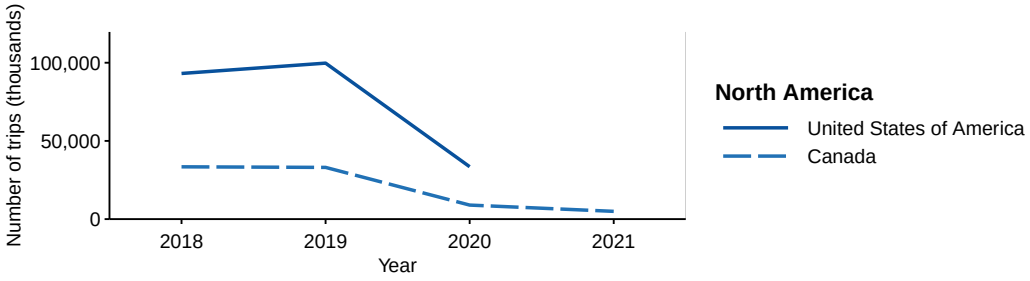
Origin country	Country code	Allocated region	Number of catches recorded
Antarctica	ATA	Antarctica	61
Bouvet Island	BVT	Antarctica	1
British Indian Ocean Territory	IOT	South Asia	9
Netherlands Antilles	ANT	Latin America & Caribbean	22
Reunion	REU	Sub-Saharan Africa	1
United States Minor Outlying Islands	UMI	East Asia & Pacific	812

(b) Destination country

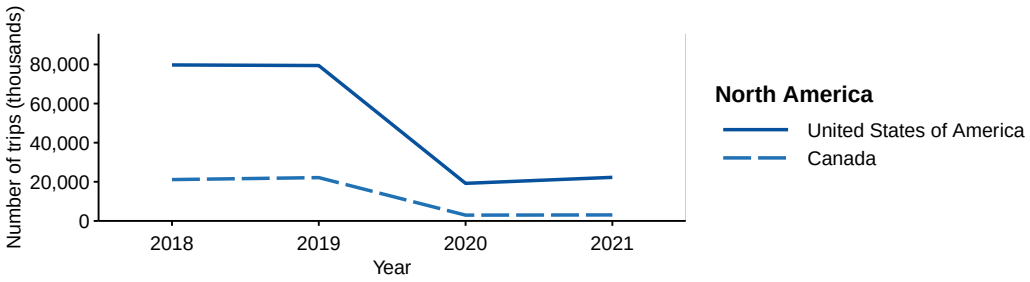
Destination country	Country code	Allocated region	Number of catches recorded
Antarctica	ATA	Antarctica	257
French Southern Territories	ATF	Antarctica	7
French Southern Territories	ATF	Sub-Saharan Africa	4
Reunion	REU	Sub-Saharan Africa	4
Saint Helena, Ascension and Tristan da Cunha	SHN	Sub-Saharan Africa	27
Svalbard and Jan Mayen	SJM	Europe & Central Asia	32
United States Minor Outlying Islands	UMI	East Asia & Pacific	22
Wallis and Futuna	WLF	East Asia & Pacific	1
Western Sahara	ESH	Sub-Saharan Africa	29

C.3 General tourism trends

This section contains Figures C.1 to C.4, which illustrate the outbound and inbound general tourism trends (as given by UNWTO tourism data (World Tourism Organization, 2023)) for the countries with the highest numbers of trips by Fishbrain users in the North America, Europe & Central Asia, Latin America & Caribbean, and East Asia & Pacific regions, from 2018 to 2021. These figures show tourism volumes, specifically the number of tourists (individuals staying overnight for a minimum of one night) or, where data on the number of tourists is not available, the total number of tourists and visitors (individuals on same-day excursions, with no overnight stays) (World Tourism Organization, 2023).

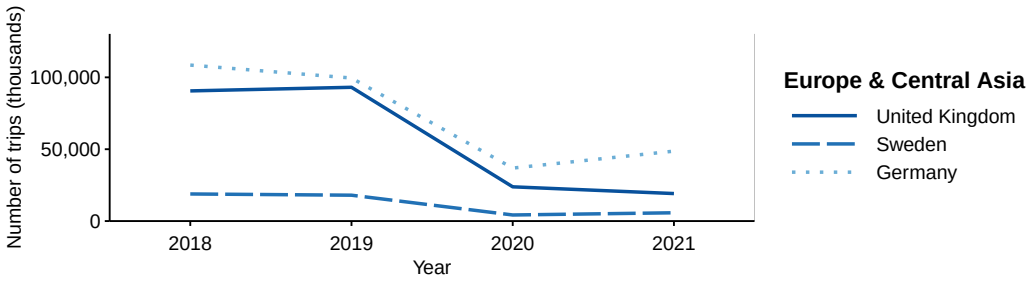


(a) Origin countries (outbound trips).

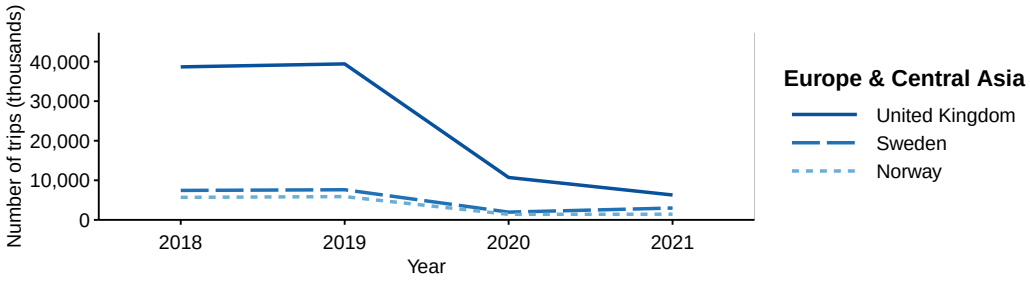


(b) Destination countries (inbound trips).

Fig. C.1 Number of tourists travelling from the top origin countries and to the top destination countries in the North America region. Outbound data are not available for the United States of America in 2021. Source: (World Tourism Organization, 2023)

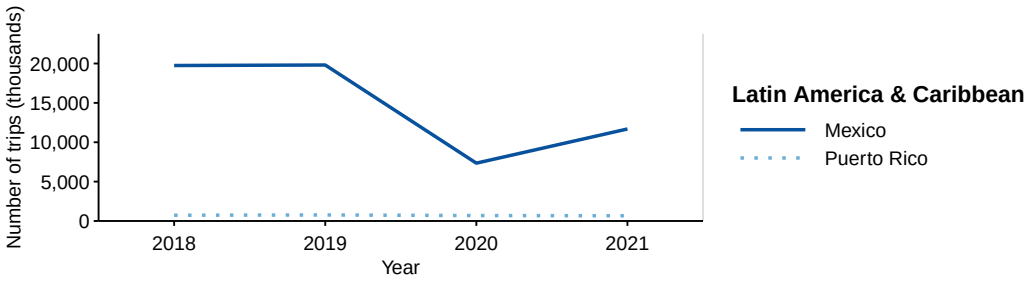


(a) Origin countries (outbound trips).

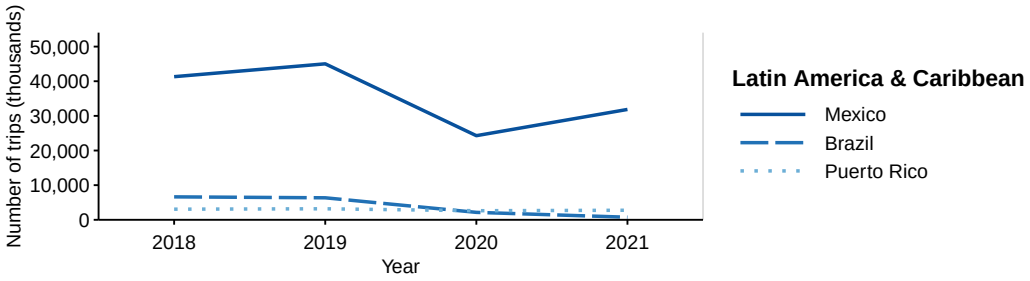


(b) Destination countries (inbound trips).

Fig. C.2 Number of tourists (except United Kingdom departures, which consist of tourists and visitors) travelling from the top origin countries and to the top destination countries in the Europe & Central Asia region. Source: (World Tourism Organization, 2023)

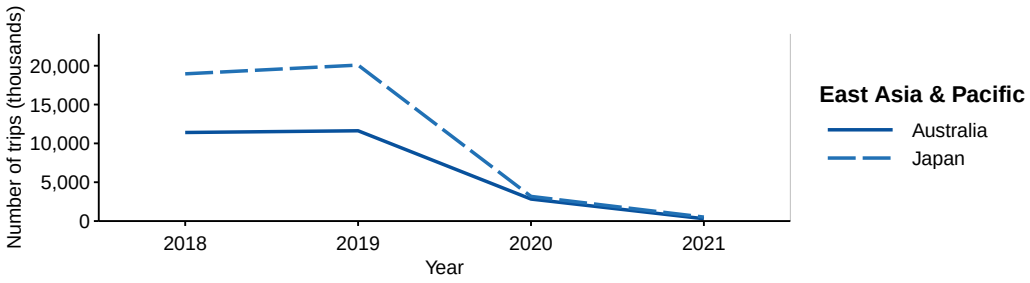


(a) Origin countries (outbound trips).

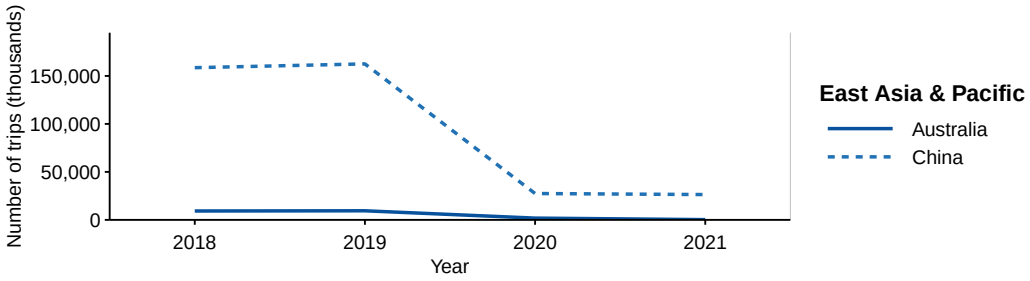


(b) Destination countries (inbound trips).

Fig. C.3 Number of tourists travelling from the top origin countries and to the top destination countries in the Latin America & Caribbean region. Outbound data are not available for Brazil. Number of outbound tourists from Puerto Rico (2018-2021): 731,000, 766,000, 688,000, and 659,000. Source: (World Tourism Organization, 2023)



(a) Origin countries (outbound trips).



(b) Destination countries (inbound trips).

Fig. C.4 Total departures and arrivals (tourists and visitors) from the top origin countries and to the top destination countries in the East Asia & Pacific region. Total arrivals to Australia (2018-2021): 9,246,000, 9,466,000, 1,828,000, and 246,000. Source: (World Tourism Organization, 2023)

C.4 Travel patterns for individual species

Figures C.5, C.6 and C.7 show networks based on the number of trips for the top three species caught (Largemouth black bass, Northern pike, and Smallmouth bass). These networks show only the most popular travel routes for each species, by including only the edges that meet a minimum threshold number of trips from the origin to destination vertex each quarter, along with the associated vertices, as described in Section 4.2.1. The minimum threshold is five for Largemouth black bass, and two for Northern pike and Smallmouth bass.

Trips for all three species predominantly occur during the northern hemisphere spring and summer. Interestingly, the number of connections and trips for these individual species remain more consistent over the four-year period than for all fishing tourism trips, particularly for Largemouth black bass and Northern pike.

The majority of trips for Largemouth black bass are between the United States and Canada, and from the North America region to countries in the Europe & Central Asia and East Asia & Pacific regions. Trips for Northern pike tend to be between the United States and Canada, and within Europe & Central Asia, with a few connections between countries in the North America and Europe & Central Asia regions. Most trips for Smallmouth bass are from the United States to Canada. There is a notably higher number of connections for Smallmouth bass from Europe & Central Asia to the United States and Canada in 2018 quarter 3, and from the East Asia & Pacific region to the United States in 2019 quarter 2, compared to other quarters during the four-year period. There are also more connections from the United States to countries in the east of the world in 2018 quarter 2, than in any other quarter.

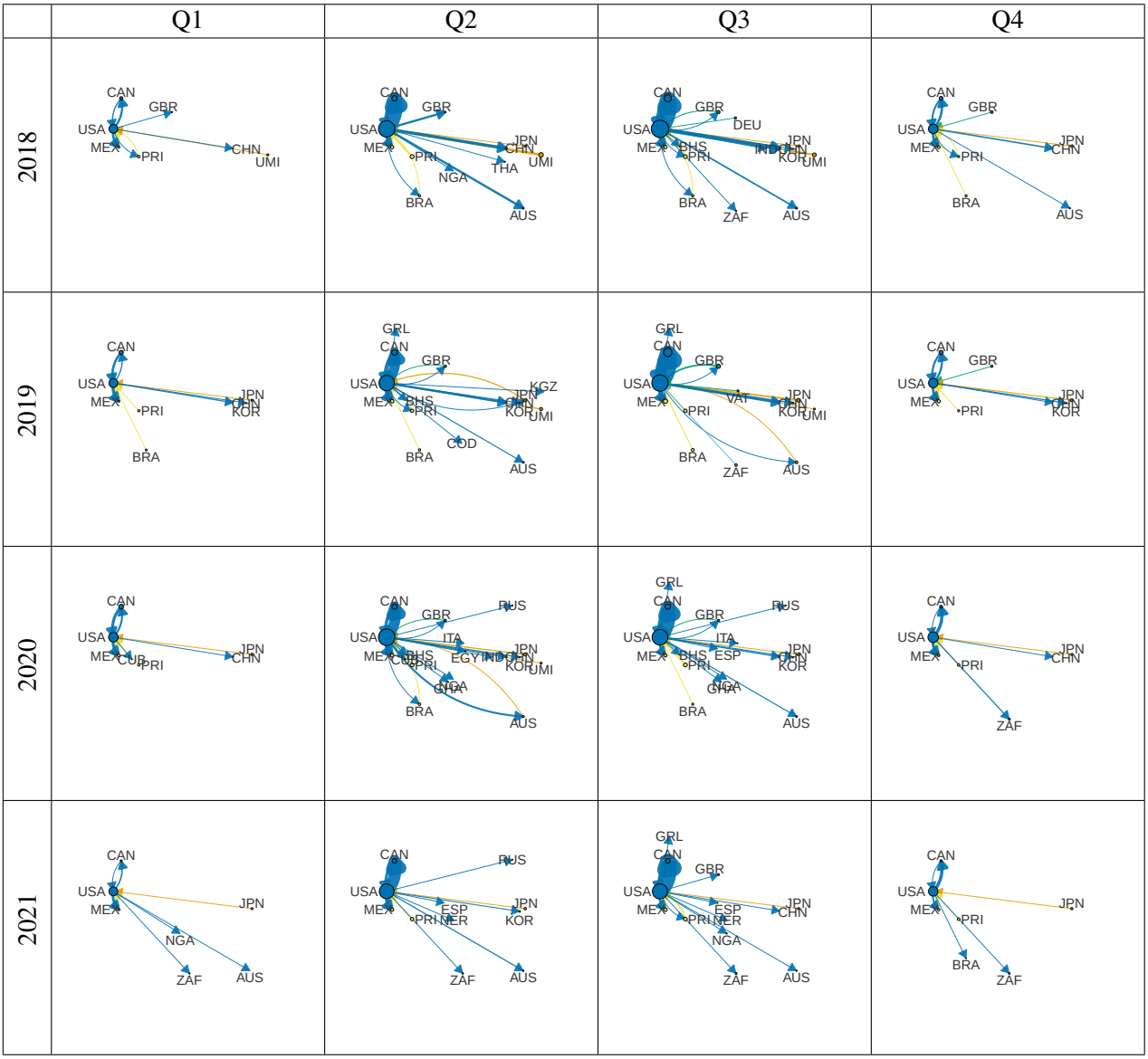


Fig. C.5 Network graphs for Largemouth black bass, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Fig. C.6 Network graphs for Northern pike, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

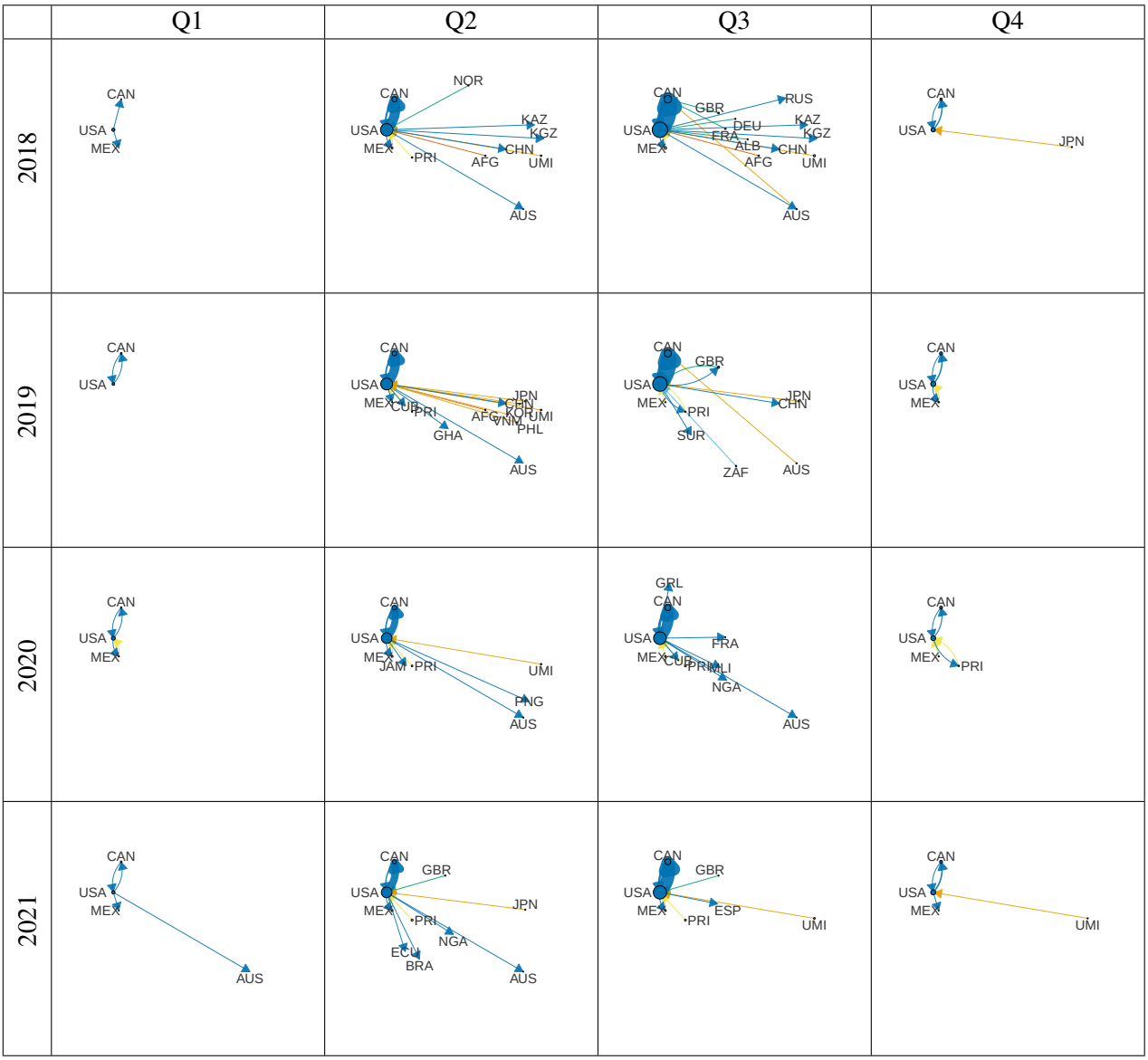


Fig. C.7 Network graphs for Smallmouth bass, based on the number of trips between countries. The area of each vertex represents the total number of trips from that country to other countries. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

C.5 Network graphs for the United Kingdom and United States with no minimum threshold

This section contains network graphs for the individual case study countries (the United Kingdom and the United Sates) based on all trips and catches with no minimum threshold on the number of trips or catches included between each origin-destination country pair in each quarter.

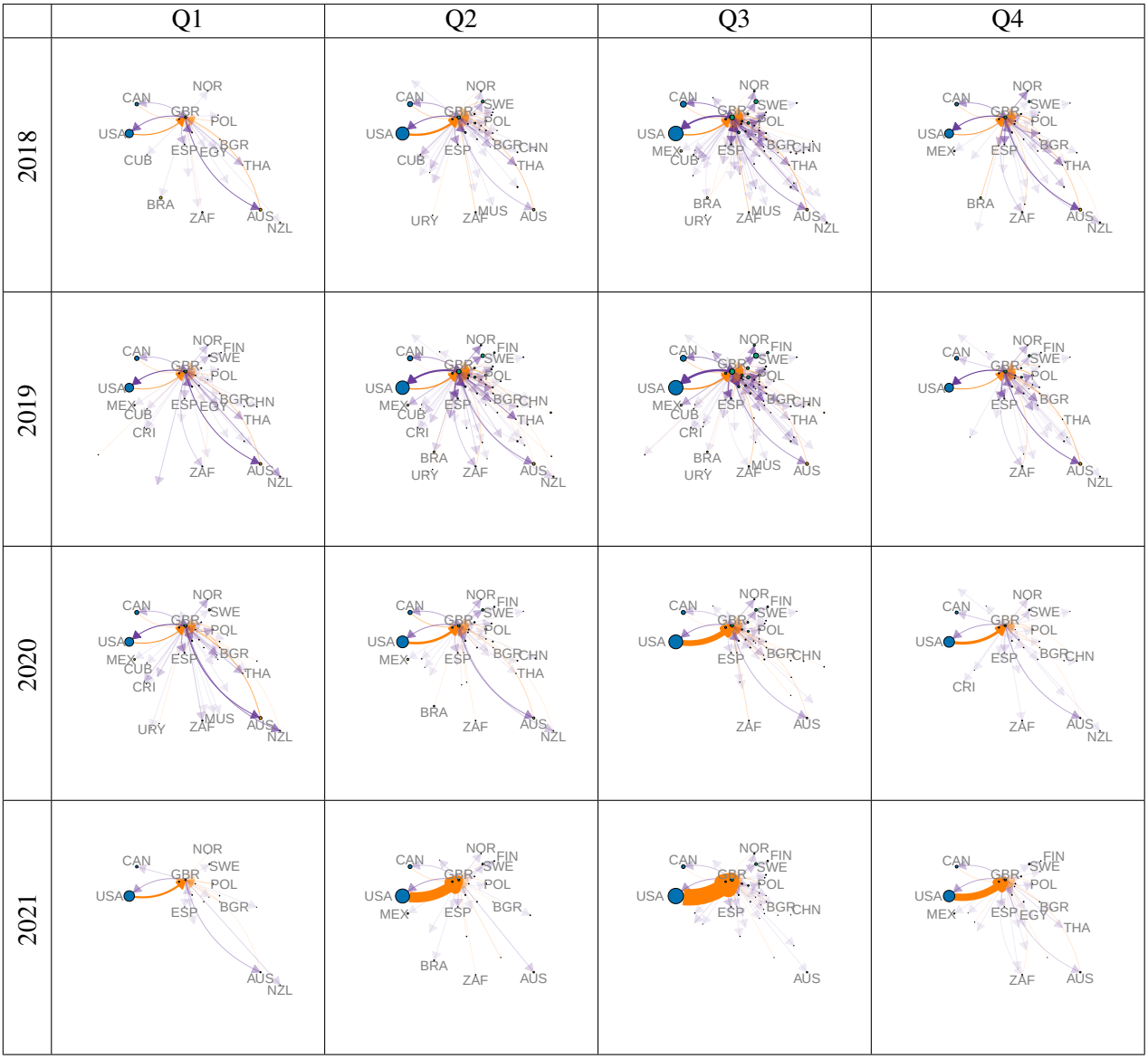


Fig. C.8 Network graphs of the trips to and from the United Kingdom. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United Kingdom are included.

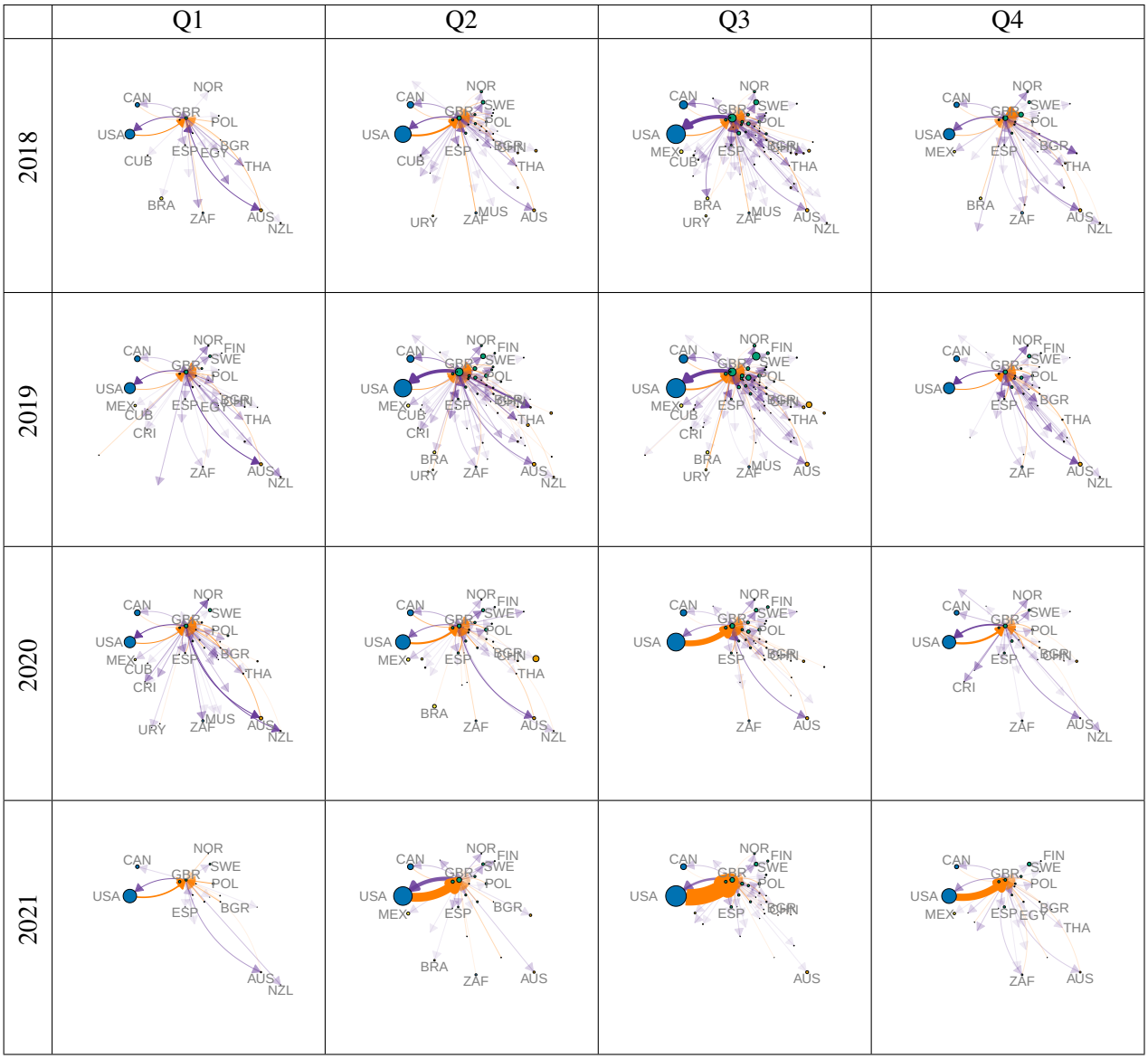


Fig. C.9 Network graphs based on catches recorded by users travelling from and to the United Kingdom. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United Kingdom are purple. Edges representing catches recorded by users travelling to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United Kingdom are included.

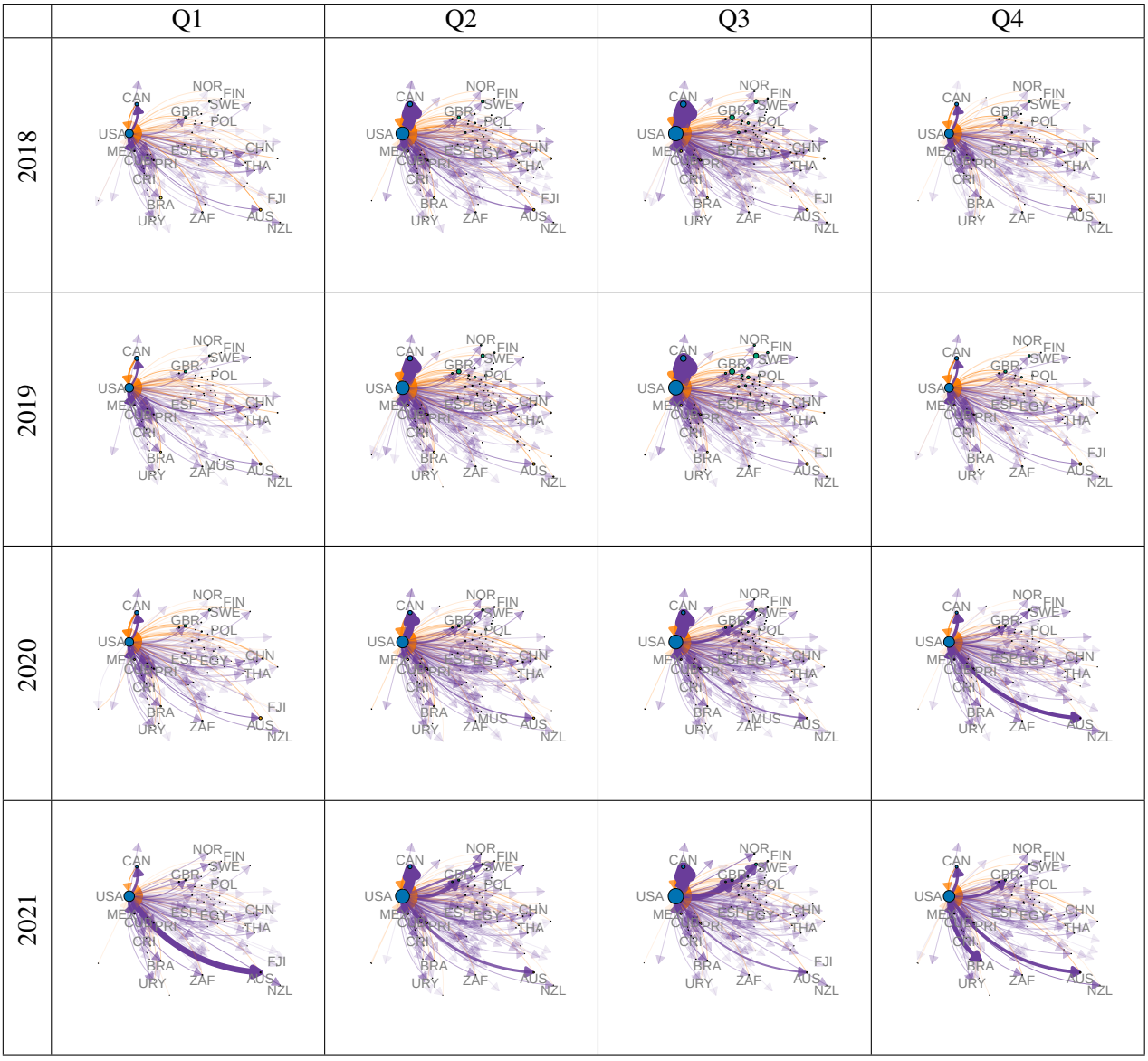


Fig. C.10 Network graphs of the trips to and from the United States. The area of each vertex represents the total number of trips from that country to other countries. The vertex colours represent the origin region. The width of the edges represents the number of trips from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing trips from the United Kingdom are purple. Edges representing trips to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United States are included.

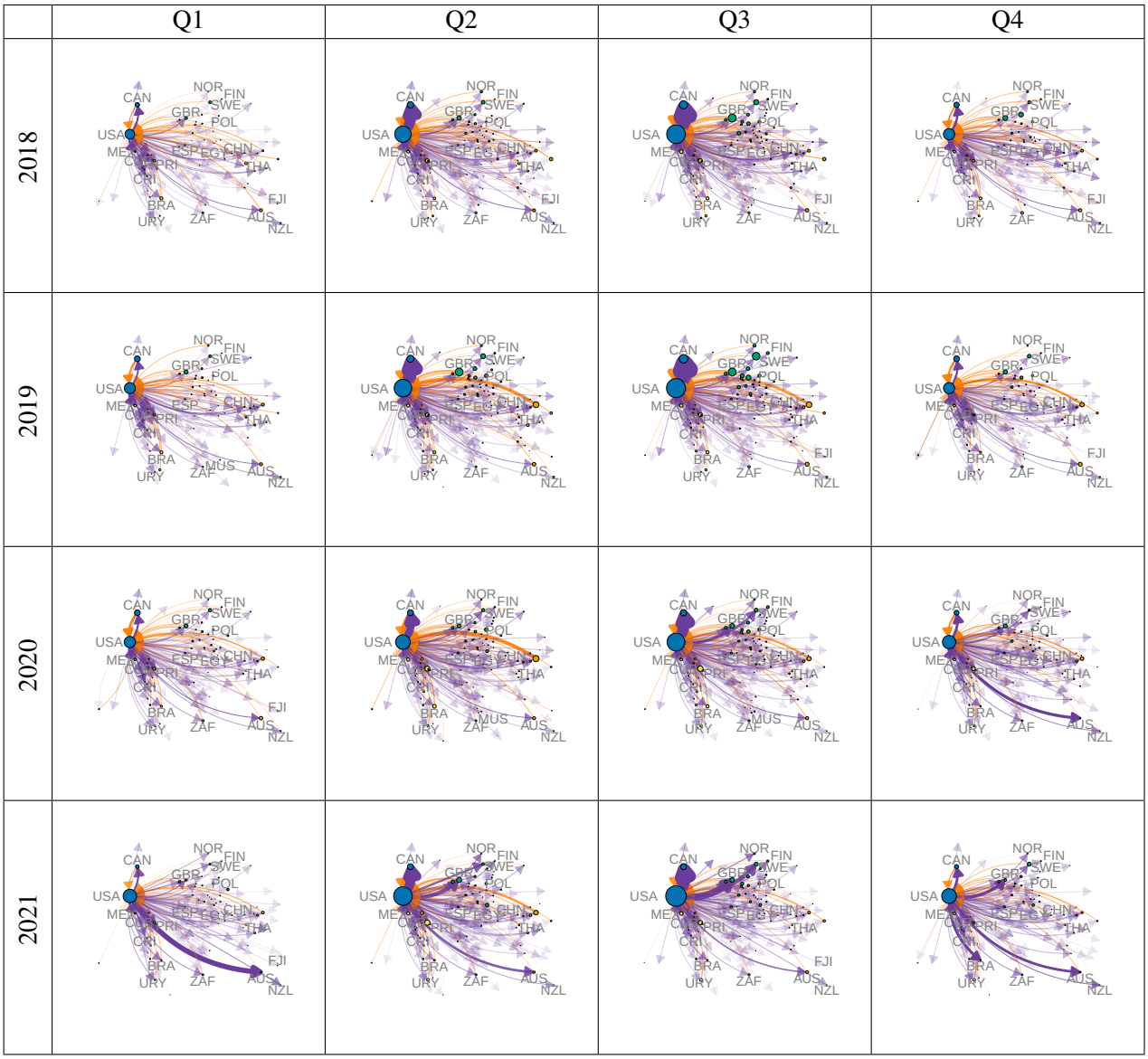


Fig. C.11 Network graphs based on catches recorded by users travelling from and to the United States. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United States are purple. Edges representing catches recorded by users travelling to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: All trips to and from the United States are included.

C.6 World regions community graphs 2017-2021

This section contains community graphs for the world regions, by quarter, for 2017 to 2021.

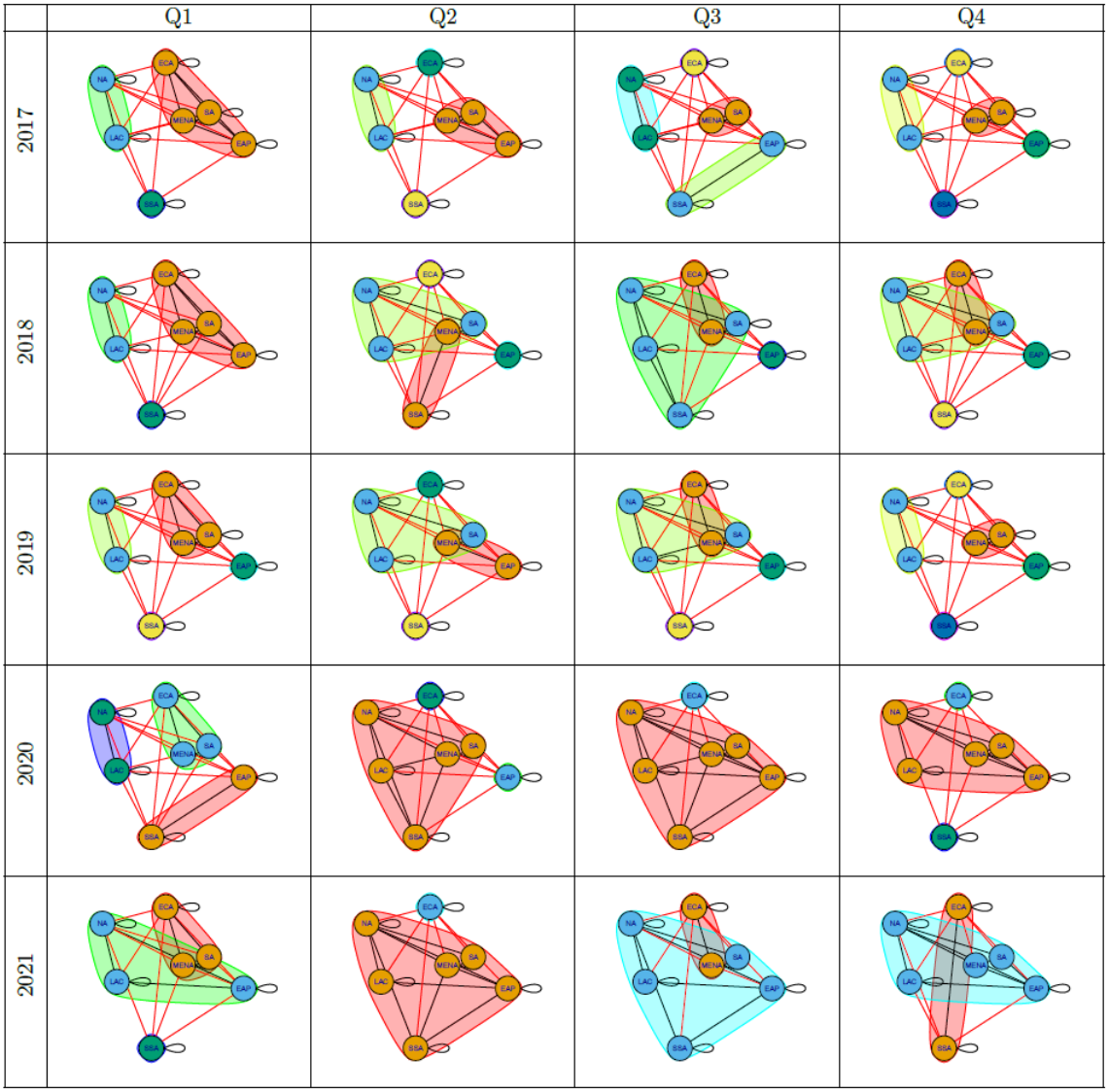


Fig. C.12 Communities in the world regions networks based on the number of trips (2017-2021). The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa

C.7 Number of catches recorded

This section contains the results for the networks summarising the number of catches recorded, for the individual country case studies, countries within Europe, countries around the world, and across world regions.

In all cases we provide network maps, by quarter, for the years 2018 to 2021. For the networks within Europe, around the world and across world regions we also provide a frequency table of the quarters in which the corresponding pair of countries has been allocated to the same community, a table of community metrics and maps of the communities.

On all network and community graphs, the location of the vertices representing the countries approximately matches their geographical location. The width of the edges on the network graphs represents the number of catches recorded by users from the origin country made in the destination country.

Table C.3 Mean edge density, mean distance and diameter for network graphs with number of catches, over the 16 quarters from 1 January 2018 to 31 December 2021.

	Edge density	Mean distance (unweighted)	Diameter (unweighted)
World Regions	0.77	1.25	2.00
World Countries	0.20	1.65	2.88
Europe	0.24	1.89	3.75
United Kingdom	0.10	1.77	2.00
United States	0.07	1.84	2.00

C.7.1 Case study countries

United Kingdom

Table C.4 Metrics for the United Kingdom network graphs based on the number of catches recorded. In each year-quarter time period, the networks contain only the edges where there are at least five trips for that origin-destination country pair.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)
2018	1	7	8	0.190	0.009	1.56
2018	2	17	21	0.077	0.023	1.82
2018	3	26	33	0.051	0.035	1.87
2018	4	18	21	0.069	0.023	1.83
2019	1	13	14	0.090	0.015	1.75
2019	2	28	33	0.044	0.035	1.85
2019	3	31	39	0.042	0.042	1.89
2019	4	14	18	0.099	0.019	1.80
2020	1	14	17	0.093	0.018	1.80
2020	2	14	17	0.093	0.018	1.80
2020	3	17	19	0.070	0.020	1.82
2020	4	13	13	0.083	0.014	1.75
2021	1	7	7	0.167	0.008	1.56
2021	2	14	16	0.088	0.017	1.78
2021	3	12	15	0.114	0.016	1.77
2021	4	8	9	0.161	0.010	1.67

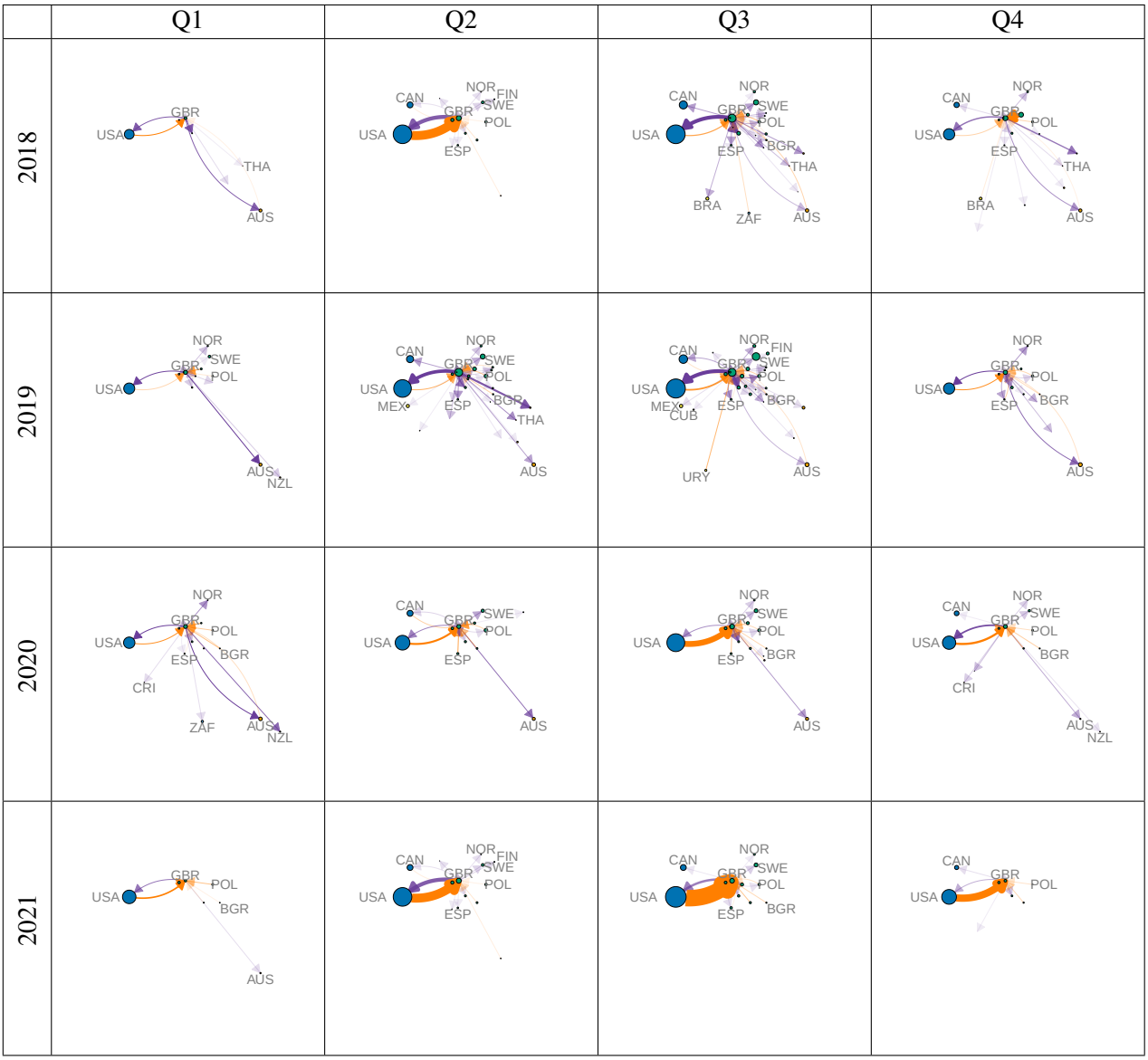


Fig. C.13 Network graphs based on catches recorded by users travelling from and to the United Kingdom. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United Kingdom are purple. Edges representing catches recorded by users travelling to the United Kingdom are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than five trips are excluded.

United States

Table C.5 Metrics for United States network graphs based on the number of catches recorded. In each year-quarter time period, the networks contain only the edges where there are at least 30 trips for that origin-destination country pair.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)
2018	1	11	14	0.127	0.022	1.76
2018	2	20	26	0.068	0.040	1.86
2018	3	18	24	0.078	0.037	1.85
2018	4	16	18	0.075	0.028	1.80
2019	1	16	19	0.079	0.029	1.82
2019	2	23	28	0.055	0.043	1.87
2019	3	24	30	0.054	0.046	1.88
2019	4	14	19	0.104	0.029	1.82
2020	1	13	15	0.096	0.023	1.77
2020	2	20	28	0.074	0.043	1.87
2020	3	25	32	0.053	0.049	1.89
2020	4	17	21	0.077	0.032	1.82
2021	1	17	20	0.074	0.031	1.81
2021	2	26	31	0.048	0.048	1.88
2021	3	24	28	0.051	0.043	1.87
2021	4	18	20	0.065	0.031	1.82

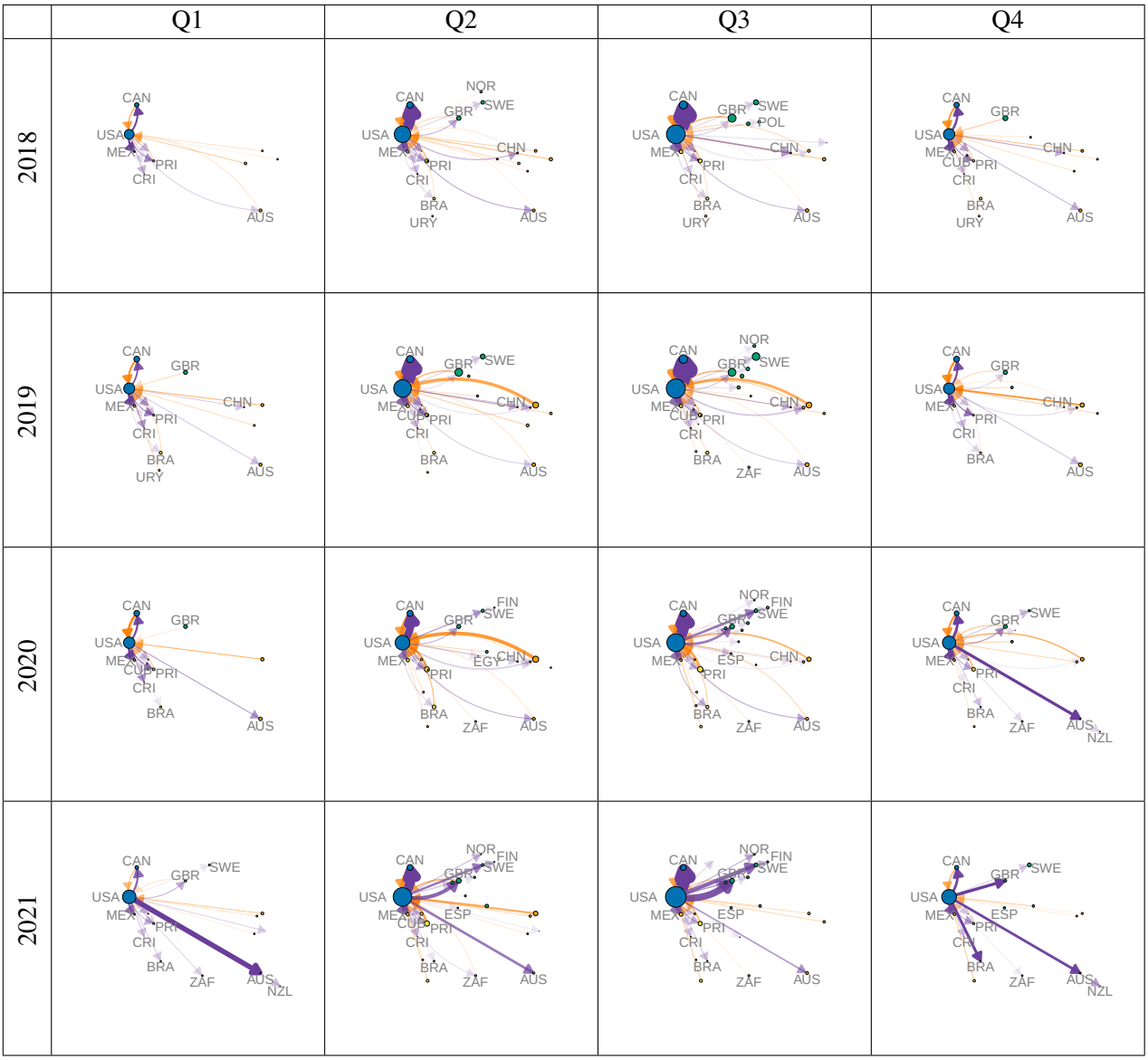


Fig. C.14 Network graphs based on catches recorded by users travelling from and to the United States. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The vertex colours represent the origin region. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. Edges representing catches recorded by users from the United States are purple. Edges representing catches recorded by users travelling to the United States are orange. The area of the vertices and width of the edges are scaled consistently across all years and quarters. Note: edges representing fewer than 30 trips are excluded.

C.7.2 Europe

Table C.6 Metrics for Europe network graphs based on the number of catches recorded.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)
2018	1	15	30	0.143	0.098	2.40	5
2018	2	18	63	0.206	0.206	1.87	3
2018	3	18	116	0.379	0.379	1.47	3
2018	4	18	65	0.212	0.212	1.92	4
2019	1	17	51	0.188	0.167	1.99	4
2019	2	18	110	0.359	0.359	1.49	3
2019	3	18	131	0.428	0.428	1.40	3
2019	4	18	81	0.265	0.265	1.71	3
2020	1	17	53	0.195	0.173	2.02	4
2020	2	18	74	0.242	0.242	1.75	4
2020	3	18	93	0.304	0.304	1.61	3
2020	4	16	44	0.183	0.144	2.07	4
2021	1	18	27	0.088	0.088	2.73	6
2021	2	18	55	0.180	0.180	2.03	4
2021	3	18	85	0.278	0.278	1.67	3
2021	4	18	52	0.170	0.170	2.07	4

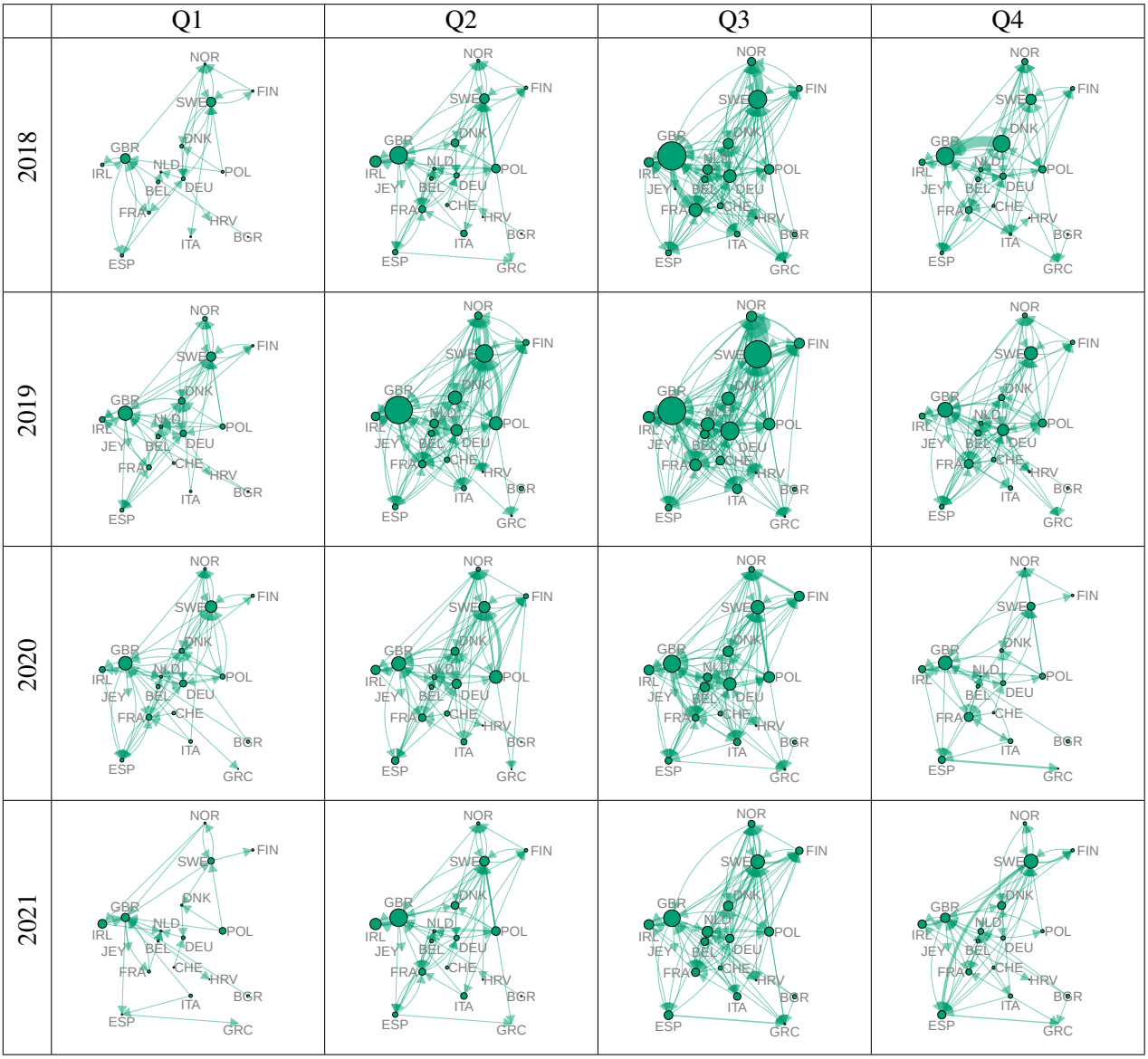


Fig. C.15 Network graphs of Europe, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table C.7 Number of quarters (maximum 16) in which each pair of countries in the Europe networks based on catches is in the same community. The countries are ordered geographically, approximately west to east, followed by Mediterranean countries.

	IRL	GBR	JEY	FRA	CHE	BEL	NLD	DEU	DNK	NOR	SWE	FIN	POL	LTU	HRV	GRC	ITA
United Kingdom (GBR)	16																
Jersey (JEY)	14	14															
France (FRA)	10	10	9														
Switzerland (CHE)	5	5	5	8													
Belgium (BEL)	2	2	2	8	7												
Netherlands (NLD)	0	0	0	6	7	14											
Germany (DEU)	0	0	0	5	8	11	13										
Denmark (DNK)	2	2	2	1	2	0	1	4									
Norway (NOR)	0	0	0	0	3	0	1	4	14								
Sweden (SWE)	0	0	0	0	3	0	1	4	14	16							
Finland (FIN)	0	0	0	1	5	1	2	5	12	14	14						
Poland (POL)	3	3	2	3	4	0	1	3	11	13	13	11					
Lithuania (LTU)	2	2	2	2	4	1	2	3	13	13	13	11	10				
Croatia (HRV)	3	3	3	4	6	7	7	6	4	5	5	5	4	6			
Greece (GRC)	7	7	7	10	6	5	3	3	1	1	1	2	1	2	2		
Italy (ITA)	9	9	8	5	7	2	2	4	4	6	6	5	7	5	8	4	
Spain (ESP)	8	8	8	9	6	6	4	3	1	1	1	1	1	2	2	13	5

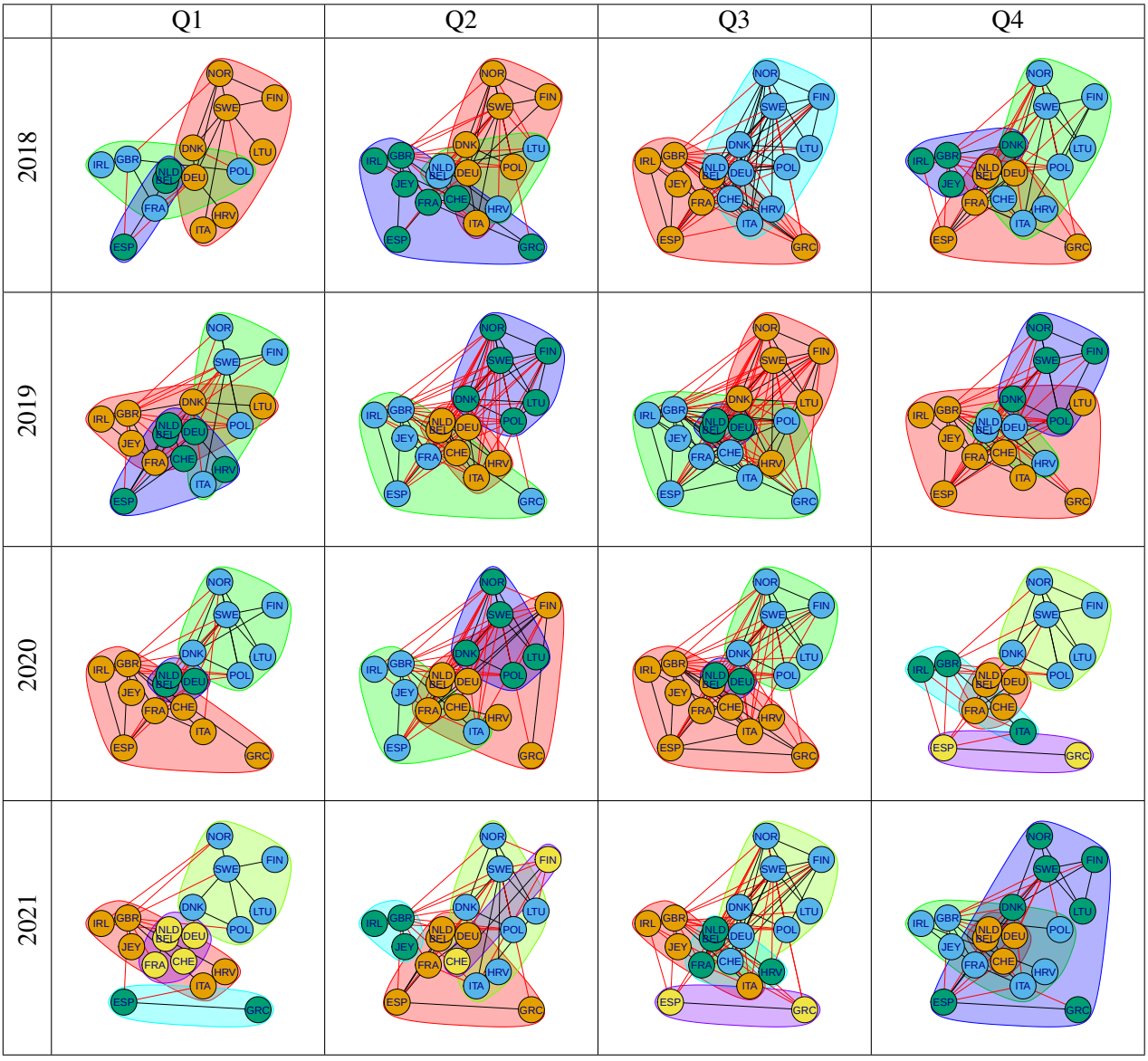


Fig. C.16 Communities in the Europe networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities.

C.7.3 World countries

Table C.8 Metrics for world countries network graphs based on the number of catches recorded.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)
2018	1	21	72	0.171	0.171	1.72	3
2018	2	21	76	0.181	0.181	1.66	3
2018	3	21	101	0.240	0.240	1.53	2
2018	4	21	80	0.190	0.190	1.69	3
2019	1	21	79	0.188	0.188	1.66	3
2019	2	21	106	0.252	0.252	1.52	3
2019	3	21	119	0.283	0.283	1.45	2
2019	4	21	89	0.212	0.212	1.60	3
2020	1	21	78	0.186	0.186	1.64	2
2020	2	21	91	0.217	0.217	1.59	3
2020	3	21	80	0.190	0.190	1.64	3
2020	4	21	65	0.155	0.155	1.74	3
2021	1	20	51	0.134	0.121	1.84	3
2021	2	21	73	0.174	0.174	1.74	4
2021	3	21	79	0.188	0.188	1.66	3
2021	4	21	70	0.167	0.167	1.71	3

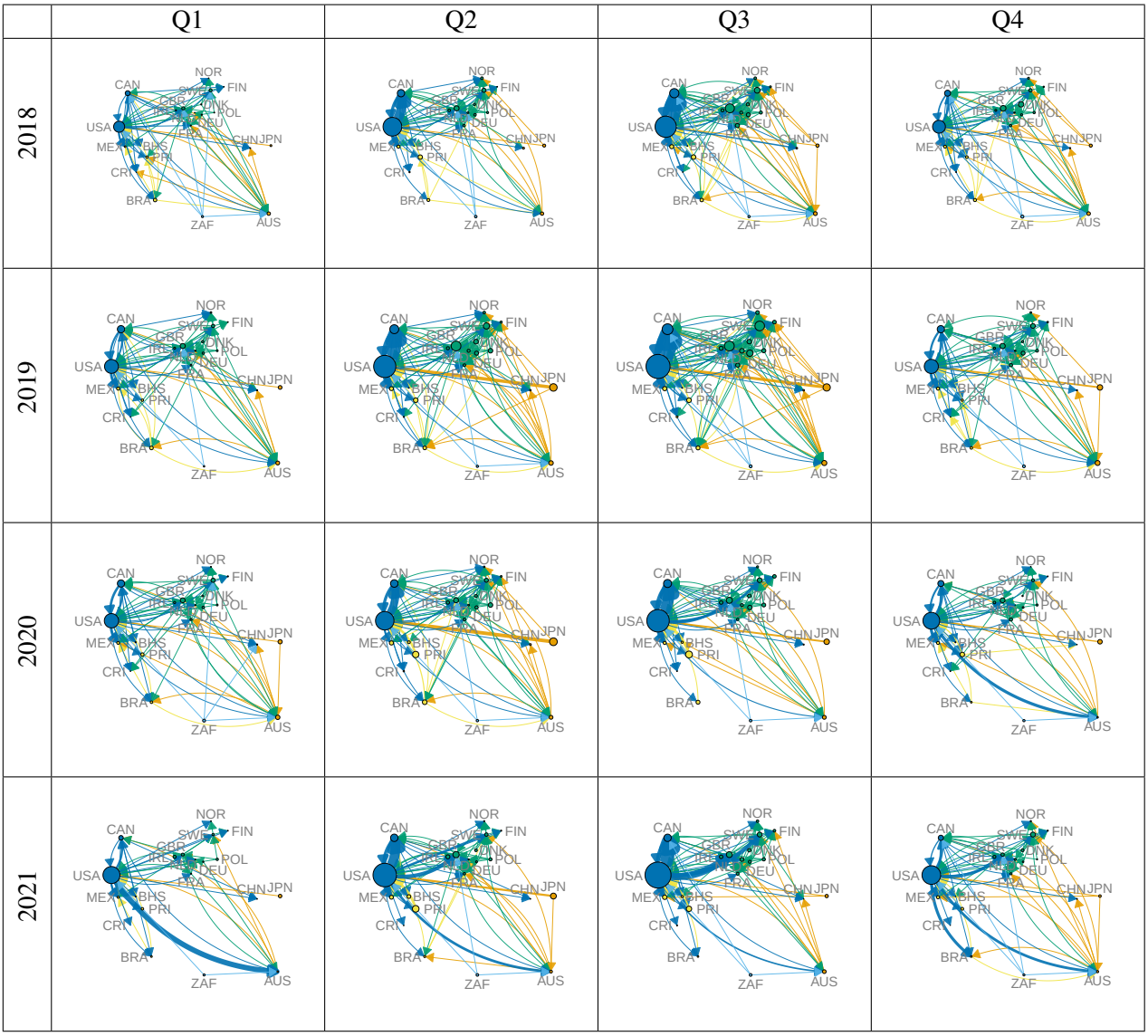


Fig. C.17 Network graphs of world countries, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the country corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table C.9 Number of quarters (maximum 16) in which each pair of countries in the world countries networks based on catches is in the same community. The countries are ordered approximately geographically in a clockwise direction (the Americas, Europe, China, New Zealand, Australia, South Africa).

		CAN	USA	MEX	BHS	CRI	PRI	BRA	IRL	GBR	FRA	NLD	DEU	DNK	NOR	SWE	FIN	POL	CHN	JPN	AUS
United States	USA	16																			
Mexico	MEX	16	16																		
Bahamas	BHS	16	16	16																	
Costa Rica	CRI	15	15	15	15																
Puerto Rico	PRI	16	16	16	16	15															
Brazil	BRA	12	12	12	12	11	12														
Ireland	IRL	2	2	2	2	3	2	4													
United Kingdom	GBR	2	2	2	2	3	2	4	16												
France	FRA	0	0	0	0	0	0	4	13	13											
Netherlands	NLD	0	0	0	0	0	0	4	11	11	14										
Germany	DEU	0	0	0	0	0	0	4	11	11	14	16									
Denmark	DNK	0	0	0	0	1	0	4	11	11	12	14	14								
Norway	NOR	0	0	0	0	1	0	4	12	12	13	15	15	15							
Sweden	SWE	0	0	0	0	1	0	4	12	12	13	15	15	15	16						
Finland	FIN	0	0	0	0	1	0	4	12	12	13	15	15	15	16	16					
Poland	POL	0	0	0	0	1	0	4	13	13	14	14	14	14	15	15	15				
China	CHN	16	16	16	16	15	16	12	2	2	0	0	0	0	0	0	0	0			
Japan	JPN	15	15	15	15	14	15	11	3	3	1	1	1	1	1	1	1	1	15		
Australia	AUS	4	4	4	4	3	4	6	12	12	12	11	11	11	11	11	11	11	4	5	
South Africa	ZAF	9	9	9	9	8	9	9	7	7	7	6	6	6	6	6	6	6	9	10	11

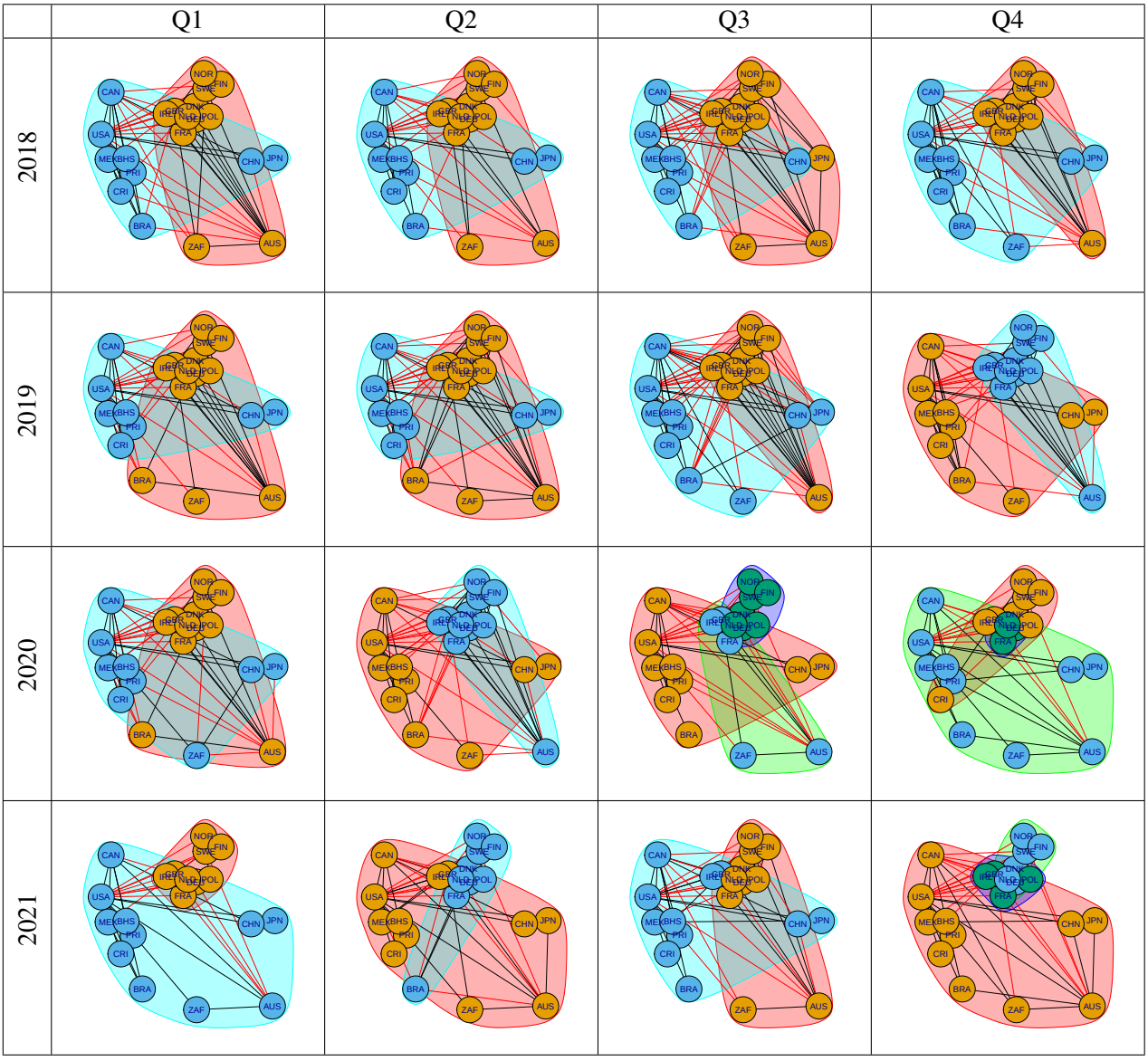


Fig. C.18 Communities in the world countries networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities.

C.7.4 World regions

The number of catches reported grouped by world region for 2018 to 2021 is shown in Table C.10 and Figure C.19.

Table C.10 Metrics for world regions network graphs based on the number of catches recorded.

Year	Quarter	Number of vertices	Number of edges	Edge density	Edge density (max vertices)	Mean distance (unweighted)	Diameter (unweighted)
2018	1	7	41	0.837	0.837	1.17	2
2018	2	7	39	0.796	0.796	1.21	2
2018	3	7	40	0.816	0.816	1.21	2
2018	4	7	39	0.796	0.796	1.21	2
2019	1	7	40	0.816	0.816	1.21	2
2019	2	7	40	0.816	0.816	1.19	2
2019	3	7	40	0.816	0.816	1.19	2
2019	4	7	40	0.816	0.816	1.21	2
2020	1	7	40	0.816	0.816	1.19	2
2020	2	7	37	0.755	0.755	1.26	2
2020	3	7	36	0.735	0.735	1.29	2
2020	4	7	35	0.714	0.714	1.31	2
2021	1	7	34	0.694	0.694	1.33	2
2021	2	7	34	0.694	0.694	1.33	2
2021	3	7	34	0.694	0.694	1.33	2
2021	4	7	35	0.714	0.714	1.29	2

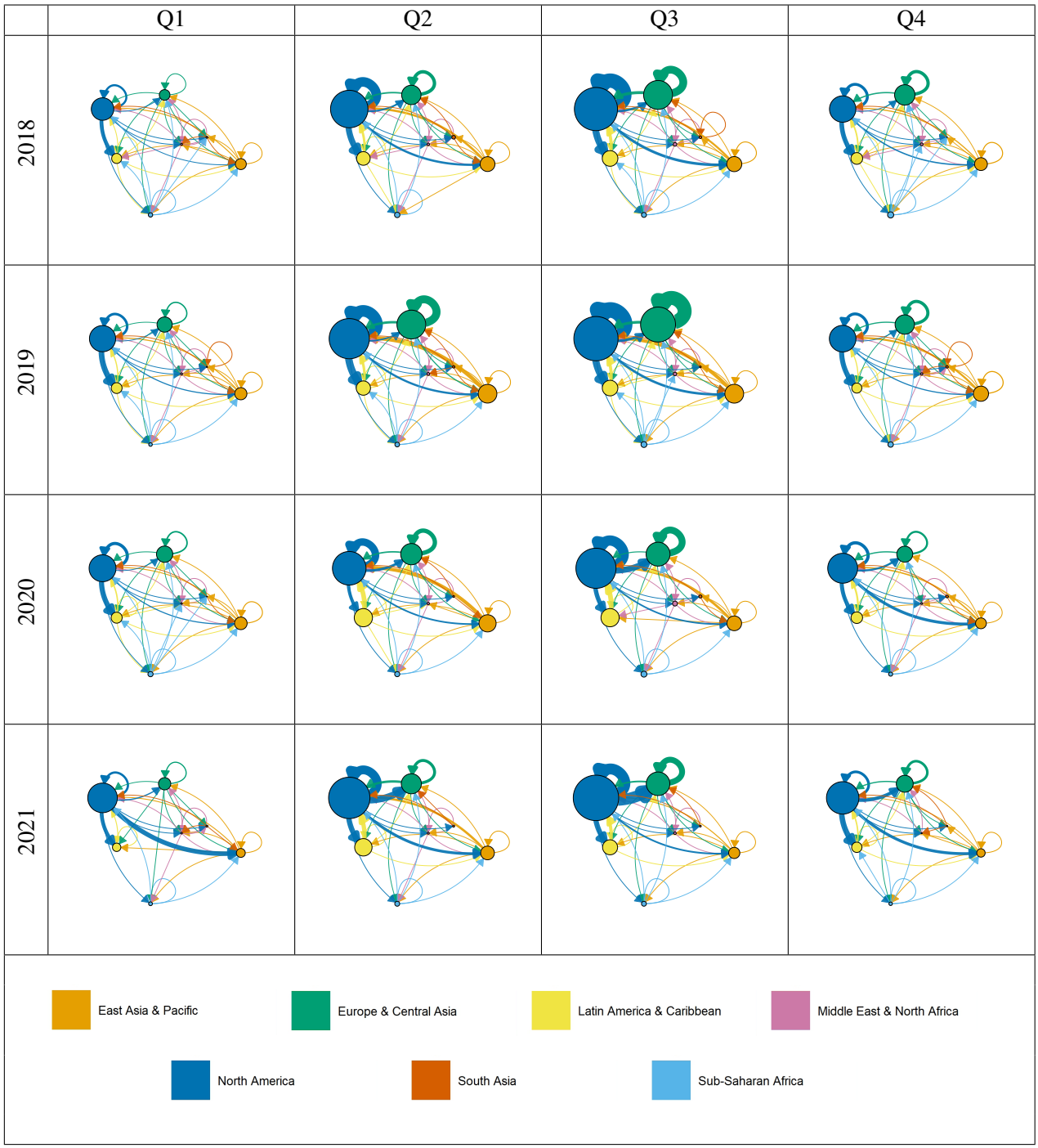


Fig. C.19 Network graphs of world regions, based on catches recorded. The area of each vertex represents the total number of catches recorded by users whose device SIM card was registered in the region corresponding to that vertex. The width of the edges represents the number of catches recorded by users travelling from the origin vertex to the destination vertex, indicated by the direction of the arrow. The vertex and edge colours represent the origin region. The loops indicate the number of catches recorded in a different country to the origin country, but within the same region. The area of the vertices and width of the edges are scaled consistently across all years and quarters.

Table C.11 Number of quarters (maximum 16) in which each pair of regions in the world regions networks based on catches is in the same community.

	North America	Latin America & Caribbean	Europe & Central Asia	Middle East & North Africa	South Asia	East Asia & Pacific
Latin America & Caribbean	16					
Europe & Central Asia	0	0				
Middle East & North Africa	6	6	4			
South Asia	11	11	3	9		
East Asia & Pacific	9	9	1	7	10	
Sub-Saharan Africa	6	6	3	7	7	7

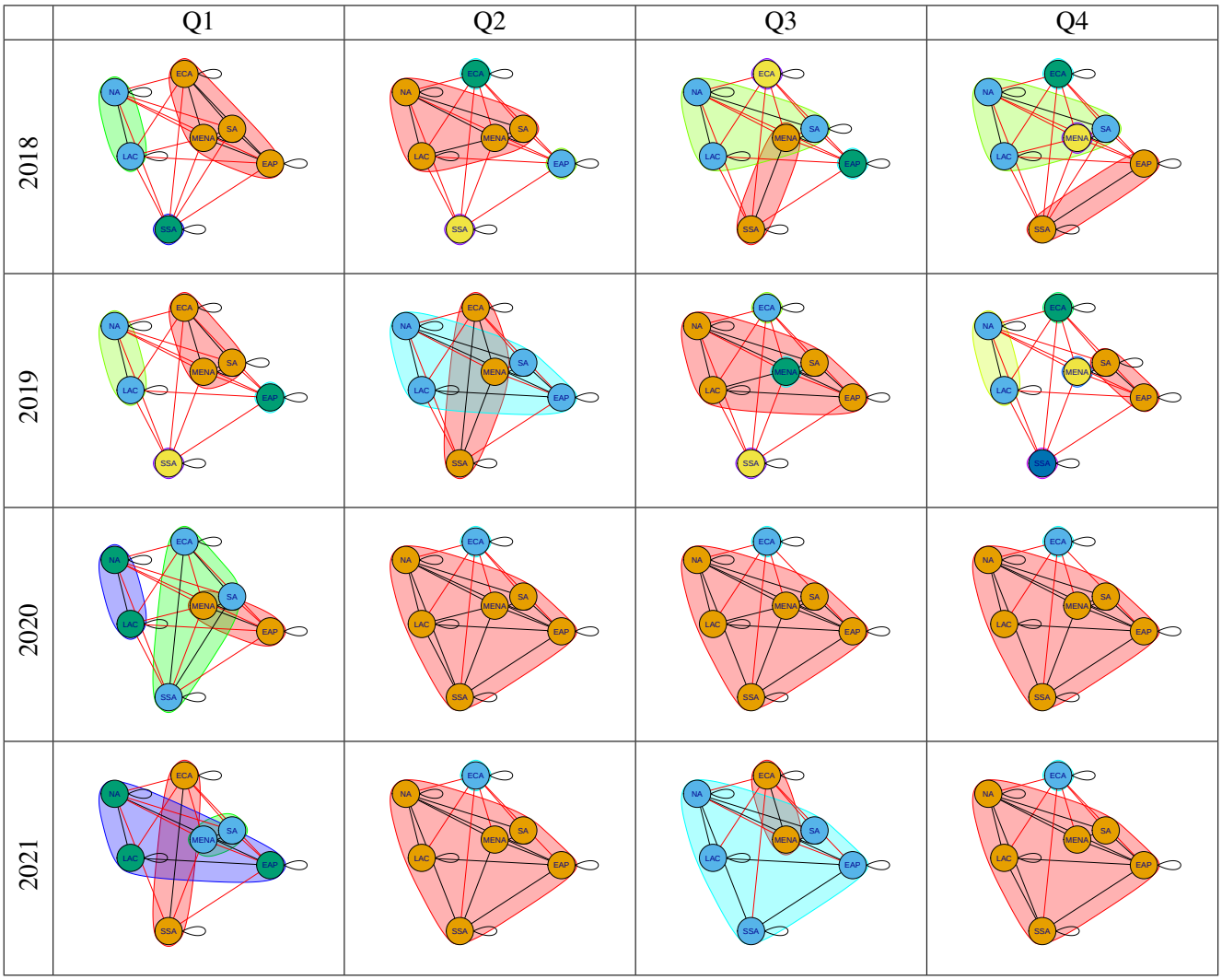


Fig. C.20 Communities in the world regions networks based on the number of catches recorded. The countries within a community are indicated by vertices of the same colour. Colours are allocated to communities by the community detection function. Edges are black when linking vertices in the same community and red when linking vertices in different communities. EAP - East Asia & Pacific, ECA - Europe & Central Asia, LAC - Latin America & Caribbean, MENA - Middle East & North Africa, NA - North America, SA - South Asia, SSA - Sub-Saharan Africa

C.7.5 Edge density

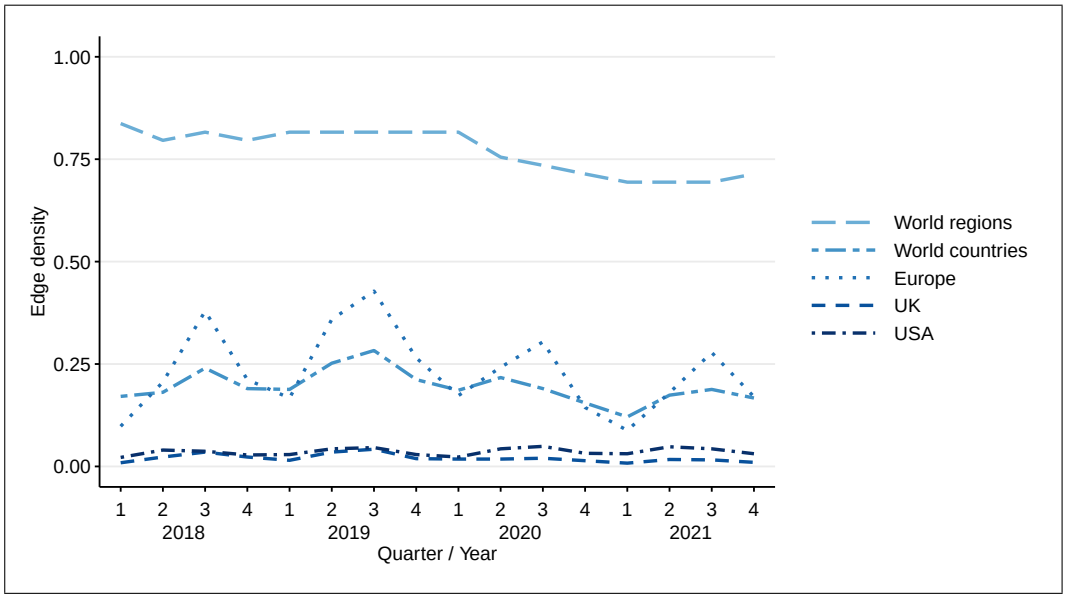


Fig. C.21 Edge density (max vertices) for the networks based on the number of catches.