



# Kent Academic Repository

**Darby, Keith V. (1986) *A statistical examination of British bird observatories data*. Doctor of Philosophy (PhD) thesis, University of Kent.**

## Downloaded from

<https://kar.kent.ac.uk/94295/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.22024/UniKent/01.02.94295>

## This document version

UNSPECIFIED

## DOI for this version

## Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

## Additional information

This thesis has been digitised by EThOS, the British Library digitisation service, for purposes of preservation and dissemination. It was uploaded to KAR on 25 April 2022 in order to hold its content and record within University of Kent systems. It is available Open Access using a Creative Commons Attribution, Non-commercial, No Derivatives (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) licence so that the thesis and its author, can benefit from opportunities for increased readership and citation. This was done in line with University of Kent policies (<https://www.kent.ac.uk/is/strategy/docs/Kent%20Open%20Access%20policy.pdf>). If you ...

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

### Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

A STATISTICAL EXAMINATION OF  
BRITISH BIRD OBSERVATORIES DATA

BY

KEITH V. DARBY

A thesis submitted for the degree of  
Doctor of Philosophy in Statistics at  
the University of Kent at Canterbury.

September 1986

## ABSTRACT

British bird observatories data are examined from a statistical viewpoint. Consideration is given to the forming of indices of migration volume which can be used to monitor fluctuations in breeding population levels.

The collection of daily counts of grounded migrants is discussed and simple graphical techniques are used to explore migratory behaviour of some common nocturnal migrants. The timing of migration is found to be related to population size as measured by the Common Bird Census.

Statistical models are fitted to sequences of daily counts of grounded migrants but migration indices developed from these models are shown to offer no significant improvements over those based on the timing of migration.

Linear discriminant analysis is shown to be a useful method of linking weather data with observatories data, and statistical models are developed for daily counts which utilise weather information.

### ACKNOWLEDGEMENTS

This research has been supported by studentships from the Science and Engineering Research Council and the University of Kent.

I would like to thank Dr. B.J.T. Morgan and Dr. P.M. North for their supervision of this work and for their comments on earlier drafts.

I would also like to thank the British Trust for Ornithology for their advice and in particular, Mr. Bob Spencer and Dr. Raymond O'Connor for access to their archives of Observatories data.

Much of the knowledge of the working of observatories was gained by partaking in the data collection. The hospitality of the wardens, in particular Nick Riddiford (formerly of Dungeness), Mick Rogers (Portland), Dick Lambert (Gibraltar Point) and Barry Spence (Spurn) was appreciated.

## CONTENTS

Page No.

### CHAPTER 1 - INTRODUCTION

1.1	Main objectives	1
1.2	Previous studies of observatories data	2
1.3	The scope of the thesis	3
1.4	Collaboration with ornithologists	6
1.5	Plan of the thesis	7

### CHAPTER 2 - BACKGROUND TO OBSERVATORIES DATA

2.1	The observatories	9
2.2	The types of data collected	12
2.3	Linking migration data with breeding bird censuses : The Common Bird Census	14
2.4	Canadian migration indices	25
2.5	A summary of other attempts at population monitoring using migration data	38
2.6	Radar studies of migration and weather	41
2.7	Other studies of British observatories data	43

### CHAPTER 3 - DESCRIPTION OF OBSERVATORIES DATA AND ASSOCIATED WEATHER DATA

3.1	Introduction	46
3.2	Daily counts of grounded migrants	47
3.2.1	General overview	47
3.2.2	The observatory day	49
3.2.3	Examples of daily counts	50
3.2.4	Timing of migration seasons	63
3.2.5	Fitting curves to sequences of autumn counts	87
3.2.6	Weekend bias	100
3.2.7	Large falls of counts	105
3.2.8	Missing data	110
3.3	Weather data from Dungeness and Portland Bill	112
3.3.1	Introduction	112
3.3.2	The data available	115
3.3.3	A principal component study of weather data	119
3.3.4	A time series analysis of weather data	135
3.3.5	Principal component study of detrended weather data	150

CHAPTER 4 - RELATING MIGRATION DATA TO WEATHER  
(I) REGRESSION

4.1	Introduction	154
4.2	Preparation of the data	156
4.2.1	Introduction	156
4.2.2	The bird data	157
4.2.3	Removing trend from daily counts	158
4.2.4	The weather data	161
4.3	The regression procedure	162
4.3.1	Methodology	162
4.3.2	Computer package used	163
4.4	Results for British observatories data	164
4.4.1	Regression on all weather variables	164
4.5	A review of Hussell's method	171
4.5.1	Description of the method	171
4.5.2	Migration indices	175
4.5.3	Results using British data	176
4.6	Conclusions	179

CHAPTER 5 - RELATING MIGRATION DATA TO WEATHER  
(II) DISCRIMINANT ANALYSIS

5.1	Introduction	181
5.2	Subdividing migration counts into groups	183
5.3	Preliminary tests	184
5.3.1	Multivariate normality	184
5.3.2	Equal covariance matrices	188
5.4	Results	190
5.5	Migration indices	200
5.6	Conclusions and discussion	201

CHAPTER 6 - MODELLING SEQUENCES OF DAILY COUNTS

6.1	Introduction	204
6.2	Immigration - death models	204
6.2.1	Introduction	204
6.2.2	A Poisson - binomial model	206
6.2.3	Results and seasonal comparisons	209
6.2.4	Goodness of fit	216
6.2.5	Alternative models	231
6.2.6	Migration indices from model (6.1)	235

	<u>Page No.</u>
6.3	Parslow's transformation 249
6.3.1	Introduction 249
6.3.2	Serial correlation under model (6.1) 250
6.3.3	The effect of Parslow's transformation 251
6.4	Conclusions and discussion 252
Appendices 6.1, 6.2 255	
<u>CHAPTER 7 - WEATHER-DRIVEN MODELS FOR DAILY COUNTS</u>	
7.1	Introduction 258
7.2	Defining the migration season 258
7.3	'Bird days' and 'Non-Bird' days 261
7.3.1	Discriminant functions 261
7.3.2	A model for daily counts (model (7.1)) 266
7.3.3	Goodness of fit 272
7.3.4	Precision of estimates 277
7.3.5	Model (7.1) with serial dependence : Model (7.2) 279
7.3.6	A generalisation of Model (7.2) : Model (7.3) 289
7.4	Migration indices 298
7.4.1	An index derived from Model (7.1) 298
7.4.2	An index derived from Model (7.2) 302
7.5	Concluding remarks 304
<u>CHAPTER 8 - RECOMMENDATIONS FOR FURTHER RESEARCH AND CONCLUSIONS</u>	
8.1	Introduction 305
8.2	The collection of data at observatories 306
8.3	The storage of data from observatories 307
8.4	Recommendations for further data analysis 308
8.5	Conclusions 312

REFERENCES.

## CHAPTER 1

### INTRODUCTION

#### 1.1 Main objectives

This thesis examines data from British Bird Observatories using statistical methods. The work continues and deepens the link between ornithology and statistics in two ways. Firstly, there is the application of statistical methodology - for example, model building and hypothesis testing - to the field of migration research, and secondly, the opening up of a complex biological area to applied statistics.

The benefits for both disciplines include the verification of common-knowledge features of ornithology by objective methods, the extension of existing knowledge of migration for the British Isles, and the demonstration of how biological mechanisms can be translated into statistical models, thereby enabling certain hypotheses to be tested.

A more specific objective is to assess the possibility of constructing an annual index of migration for species of birds that are common summer visitors to Britain, and of then using the index for population monitoring. If such an index can be constructed then it may be used to complement existing indices of population levels and also to provide a means of monitoring population levels of species not at present covered by a regular census.

One method of validating migration indices is by using results from other population censuses such as the Common Bird Census (CBC), which is run by the British Trust for Ornithology (BTO). The manner in which this census is conducted is also examined from a statistical viewpoint in this thesis.

In order to provide a full understanding of observatories data, some time is spent on examining how bird observatories are run from day to day, and the usefulness of present recording methods is discussed.

#### 1.2 Previous studies of observatories data

In the past, small subsets of observatories data have been studied, and the details are reviewed more thoroughly in chapter 3. It is sufficient at this stage to note that the points of interest have usually been the timing and length of migration seasons and the migratory routes used. This thesis provides the first analysis of observatories data on a much larger scale, due in part to the cheap and efficient computing facilities now available.

Much has been written in the past on the influence of weather on bird migration, and chapters 4 and 5 follow in this vein. The use of multiple regression in this field is not new, but most of the previous regression studies were carried out on data other than British data. Furthermore, little attention has been paid in the past to the validity of using regression analysis in this context, and this point is investigated in chapter 4.

Chapter 5 provides the first extensive study of the use of discriminant analysis in this field, and enables a direct comparison between regression and discriminant analysis to be made in the context of migration research.

This thesis represents the first major attempt at monitoring population levels by using British migration data, although the idea is not new - two small-scale studies of individual species were made where a crude index was based on annual totals of daily counts of migrants at particular observatories. A certain amount of success was achieved in these studies, but the main inspiration behind the present thesis came from overseas work, in particular from Canada and Sweden, which is discussed later.

The work of this thesis differs from previous analyses of observatories data in that attention is focused on daily counts of grounded migrants rather than daily totals of birds trapped and ringed. Section 2.2 describes how both data sets have disadvantages and advantages as regards statistical analysis.

### 1.3 The scope of the thesis

Throughout this thesis, emphasis is placed on the daily counts of grounded migrants at coastal observatories. The extensive data on the ringing and trapping of birds has been the main source of knowledge of migration for many decades, and most ornithological journals contain papers using these data.

One reason for concentrating on daily counts of birds observed rather than caught is the size of the samples involved. As will

be seen, even the daily counts data set imposes severe limitations for thorough analysis for many species, simply because of the small numbers involved. Ringing totals and daily counts are not independent data sets however, since the latter are based to a certain extent on the former, as is shown in chapters 2 and 3.

Over the period of this research a considerable amount of time was spent in extracting data from the BTO archives at their headquarters in Tring, Hertfordshire. The data, which have accumulated over several decades, are stored on record sheets (three per month) and therefore much work was needed to obtain data for a number of years from several observatories, and for several species. Although data from several observatories were obtained, the work of this thesis concentrates on data from two observatories (Dungeness and Portland). These observatories are known by ornithologists to record birds visiting Britain and not birds of passage heading for or coming from places such as Scandinavia. The presence of continental birds in the data would clearly affect conclusions concerning population levels of British breeding birds.

Data from other observatories are, however, studied in early sections of the thesis but one should remember that the counts of migrants may be inflated at certain and unknown times due to the presence of non-British-breeding birds.

Dungeness and Portland observatories are also ideal for the purpose of this thesis since weather data are available from adjacent meteorological stations. It was possible to obtain

from the Meteorological Office Archives data for the period 1968-70 (Portland; spring and autumn) and spring weather data for the period 1961-72 (Dungeness). Autumn weather data for Dungeness were obtained for 1967-70. Analyses which feature the use of weather data are therefore restricted to these time periods.

Daily counts of grounded migrants are available for the entire period 1961-72, and therefore analyses which do not use weather data have been carried out on the full data set. Because of the need for comparisons with CBC data, counts prior to 1961 were not considered due to the absence of CBC data before 1962 (counts for 1961 are studied since the autumn counts may be related to the CBC index of 1962).

At the request of the BTO, the species of birds studied in this thesis are common summer visitors. For most of these species, CBC data are available for comparison. For species without CBC data, analysis is also restricted to post-1961 due to difficulties in obtaining early records. Only species that migrate during the night are studied since diurnal migrants present more complex recording problems - see later. Similarly, only passerines i.e. small land birds, are studied since the recording of seabirds, for example, is heavily dependent on visibility and manpower - see chapter 2.

Thus we are concerned with species such as Warblers, Chats and Flycatchers rather than Swallows, Martins and Wagtails for example. Particular species such as the Redstart, Sedge Warbler, Whitethroat

and Willow Warbler are studied in detail throughout the thesis since they are common migrants at a number of observatories during both spring and autumn migration seasons. Some consideration is given to other less common species, for example, Blackcap and Flycatchers (Pied and Spotted). Latin names appear at the end of the thesis.

In the future, more extensive studies may be easier to carry out, as the BTO is currently expanding its computer storage of data. It is hoped that this thesis will help to suggest which data sets are more urgently needed on computer filestore.

The data (daily counts of grounded migrants and weather data from Dungeness and Portland) that are used in this thesis are held at the BTO headquarters at Tring, Hertfordshire. Individuals who are interested in analysing these data should contact the Ringing and Migration Section, Beech Grove, Tring.

#### 1.4 Collaboration with ornithologists

Much of the background knowledge required for analysing observatories data was obtained from discussions with staff of the BTO and wardens of observatories. Many visits to observatories were made, from which a clear understanding of the mechanics of migration was obtained. This is a good example of how statisticians can benefit by participating in the collection of the data, and discussing problems with those who work in the field from day to day.

## 1.5 Plan of the thesis

In chapter 2 the work of bird observatories is outlined, with particular aspects of the data collection methodology being examined in detail in chapter 3. Chapter 2 also puts the work of this thesis into context by reviewing other studies of observatories data, including studies made at overseas observatories. The Common Bird Census (CBC) is considered, since throughout the thesis comparisons are made between annual migration indices and CBC indices, for appropriate species.

Chapter 3 is divided into two parts, the first being a detailed study of how observatories data are collected and the implications that these methods have on subsequent analyses. Many diagrams of samples of daily counts are presented so that the reader may appreciate the variety of interesting features that these data exhibit.

The second part of chapter 3 concerns weather data which originate from lighthouses close to Dungeness and Portland observatories. In later chapters the weather data are linked with migration data, and here we simply explore the structure of the weather observations. The work of this section enables us to choose subsets of weather variables from the total number of variables available, while also being aware of properties such as correlation between variables and auto-correlation of observations taken over time.

In chapter 4 we ask the question: can we explain the numbers of birds recorded each day in terms of population level and local weather conditions? If we can isolate these two factors then we

may be able to construct a migration index that measures fluctuations in population level. Much of this chapter points to the fact that regression is not very useful in this context, but some time is spent in examining the work of Hussell (1981), who has attempted a regression approach to the formulation of indices, with reference to Canadian east coast migration.

Linear discriminant analysis is presented as an alternative to multiple regression analysis in chapter 5. Discriminant analysis is seen to be a useful tool for the study of relationships between migration counts and local weather conditions since it offers straightforward interpretations of results, and it is based on assumptions that appear to be upheld by the data. The way in which the results of chapter 5 may be used in constructing migration indices which are adjusted for weather is outlined.

Chapters 6 and 7 consider the modelling of sequences of daily counts, the latter chapter involving weather data and results from chapter 5. Simple models are proposed and the difficult problem of testing their goodness of fit is considered.

Chapter 8 sets out some recommendations arising from the thesis for ornithologists and statisticians who may wish to continue research in this area. The recommendations cover the collection and storage of data and the statistical analysis of the data.

## CHAPTER 2

### BACKGROUND TO OBSERVATORIES DATA

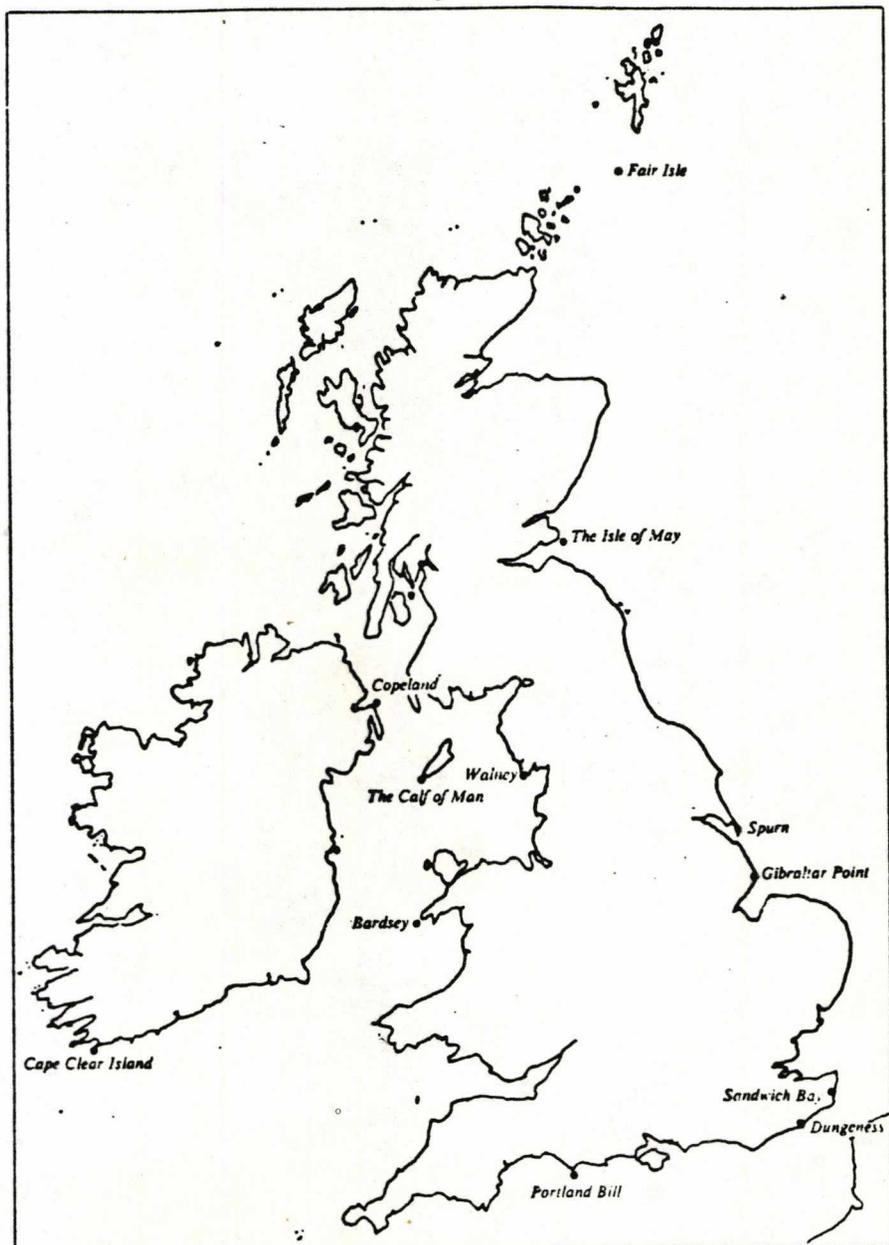
#### 2.1 The observatories

Fig. 1 shows the observatories which were operating in Britain in 1985. The map shows a widespread distribution. They are all situated on coastal headlands or on islands near to the coast. Headlands tend to 'funnel' migrating birds in the autumn as they head south, i.e. a flight direction nearest to the desired direction is taken when a sea-crossing is the only alternative. In spring, headlands form the first sight of land for birds approaching Britain, and again produce concentrations of resting migrants. Observatories are mostly situated in remote areas where human disturbance is minimised. Islands offer the same advantages for both ornithologists and birds, but with even less human disturbance.

The aims of the observatories are primarily the study of migration and the numbers of birds involved. The data that accrue at each site provide an opportunity for studying many aspects of migration, including the influence of weather, the occurrence of rarities, migratory routes involved and the biology of migration, i.e. weights of birds and the behaviour of birds that are about to complete or commence a long journey.

Each observatory has an area of land around, or adjacent to the main buildings which is termed the 'recording area'. Dungeness, for example, has a large shingle area with scrub patches and Portland has arable land with hedgerows and scrub patches.

Fig. 1 Accredited Bird Observatories in Britain  
and Ireland, 1985



( from Durman, 1976 )

Other examples are Sandwich Bay with golf courses, sand dunes, arable land and hedgerows, Gibraltar Point with scrub and wetland areas, while Fair Isle has little scrub, being a barren island.

The observatory buildings include accommodation for the warden and visitors, and the ringing laboratories. Most observatories have vantage points overlooking the sea from which sea-bird movements can be studied using fixed-point telescopes.

All observatories have at least one regular ringing site. These are usually adjacent to scrub patches or within scrub areas where paths are cut in order to set up trapping equipment. At least one Heligoland trap is usually found. These are large funnels of netting into which birds are driven by ringers, until the birds reach a collecting box at the neck of the funnel.

The recording area must obviously be kept the same size from year to year in order for censuses to be of use. Continuity of habitat is equally important. Dungeness suffered a major alteration with the building of two nuclear power stations nearby. Building commenced in 1960 and is still in progress today. Ornithologists are confident that the time period of interest in this thesis falls within a sustained period of building, and therefore any influence on observatory records will apply equally in each year studied.

The day to day running of observatories is carried out by an appointed warden together with visitors. For visitors, who are

almost entirely amateur bird enthusiasts, observatories provide an inexpensive holiday where ornithological knowledge and expertise can be improved. In particular, observatories enable bird ringers to gain experience with species not found in their own locality.

The British observatories are financed by donations from the public and local beneficiaries. These funds pay the warden and provide equipment such as rings and nets. Annual reports are also published using this money.

Each observatory is financed independantly but their activities are guided by the BTO ringing and migration section. Each year the Bird Observatories Council meets in order to discuss the organisation of observatories. The Council consists of wardens, benefactors and research ornithologists.

The Isle of May observatory was founded in 1933, but most were formed in the 1950's. In most years the observatories are manned throughout the spring and autumn migration seasons, if not throughout the year. Details of individual observatories can be found in Durman (1976).

## 2.2 The types of data collected

The daily census data available from observatories consist of counts of grounded migrants (mostly nocturnal migrants counted during the day after flight), visible migration counts (mostly diurnal migrants) and sea-bird counts. All three censuses are carried out by the warden and visitors. They are carried out

each day the observatory is manned, but the sea-bird census depends on there being interested visitors present who are able to identify birds at long range. All three censuses aim to include only migrant birds, and the experience and discretion of the warden helps to eliminate counts of local resident birds from counts of migrants.

Ringling data form the other major source of records from observatories. Not surprisingly, the number of birds caught each day depends to some extent on the manpower available. Visitors need not be qualified ringers. If many ringers are available extra ringing sites may be set up on the recording area thus increasing the capture rate. If the capture rate becomes too high for the available manpower then steps may be taken to catch fewer birds and thus reduce the time that each bird spends in captivity.

Most captures are made by mist-nets which are walls of very fine netting with horizontal pockets into which birds fall and become entangled. They are usually set in scrub patches along deliberately cut paths. Wet weather and wind result in the nets becoming visible to birds, and also cause harm to trapped birds. Mist-netting is therefore only possible in certain weather conditions.

Heligoland traps were described in the introduction, and are less dependent on manpower. Two people could cope adequately with driving birds through most traps.

Daily ringing totals clearly represent a much smaller sample of birds than counts of observed birds, but more information is available on ringed birds. If birds are trapped more than once in a given day, they will be identifiable by their ring number and therefore will only feature once in the daily total. The number of retraps gives an idea as to whether birds are staying in the area or quickly moving on. The weight of a bird may indicate whether or not a long journey has recently been undertaken, and the wing-length may give an idea as to the geographical origin of the bird.

Perhaps the most rewarding aspect of trapping birds is the retrieval of rings placed on the bird elsewhere. In some cases it is possible to learn about the migration of species by the recovery of rings, e.g. the route taken, the timing of the journey and the speed of flight. It is worth noting that identification of trapped birds is likely to be very precise. All ringers are either highly trained or are supervised by people who are.

### 2.3 Linking migration data with breeding bird censuses : the CBC

We now move temporarily away from migration data and review a census of birds breeding rather than migrating. The CBC takes place on a number of sample plots throughout the U.K. This is of interest in the context of comparing annual variation in breeding numbers with migration counts. If the latter data show similar trends to the former, then populations of birds which are not covered by the CBC may be monitored.

The CBC was started in 1962 by the Populations Section of the BTO, on behalf of the Nature Conservancy Council (NCC). The NCC was particularly interested in population changes on farmland, but in 1964 the BTO started a similar census of woodlands. The aim of both censuses is to monitor population levels of common breeding birds, by estimating the number of territory holding males at a number of farmland and woodland sites. The censuses are, therefore, restricted to species that are territorial in behaviour, and that are conspicuous enough by sight or sound to be identified and counted reliably.

The censuses are carried out by volunteers who regularly visit particular sites throughout the breeding season. Maps are constructed of sightings of birds or instances of singing birds which are not seen. The maps are examined by BTO staff who arrive at an estimate of the number of territories present. An objective method of making this estimate using cluster analysis was proposed by North (1977, 1979).

For the census to be of any use, a number of assumptions must be made. One must assume that the effectiveness of observers is constant from visit to visit, and from year to year. Mapping techniques must be consistent and if sites are dropped from the census, replacement sites should be of similar size and habitat. One must also assume that for a given species, the sites used are a representative sample of sites over the whole of Britain. The possibility of regional differences obscuring the national picture is investigated below.

In order to assess year to year changes, only sites censused in consecutive years are used. Annual indices are calculated using percentage changes from year to year. Indices are scaled so that the 1966 index is equal to 100.

The BTO calculate confidence intervals for year to year percentage changes using a method described by Bailey (1965). Upton and Lampitt (1981) point out that this method is incorrect and offer an alternative. The modelling of CBC data has also featured in Upton and Lampitt (1981), Mountford (1982) and North (1983).

The CBC indices for woodlands and farmlands are published annually in Bird Study or BTO News, together with percentage changes from year to year and confidence intervals. The number of sites used for each index is also given.

In order to use the CBC index for a given species as a national index, the sites used for the CBC must be a fair representation of all the sites used by that species throughout Britain. If this is not the case, then only regional indices will be of interest. In the past, only two authors have considered regional CBC indices. Bailey (1967) and Taylor (1965) both used contingency tables to examine differences among four geographical groups of sites over a number of years. Their results for various species suggested consistency among the regional groups, but the use of the chi-squared statistic is not appropriate here since the year categories are not independent. The serial correlation inherent in the data will tend to deflate the chi-squared statistic and perhaps fail to suggest inconsistency where

it exists.

In a recent paper which argues in favour of using regional CBC indices, Fuller, Marchant and Morgan (1985) cite examples of species whose populations are known to have declined on a local basis and for whom regional indices would be more appropriate than a national index. Although Fuller et al (1985) do not go on to present regional indices, they present a detailed study of CBC sites using a land classification technique together with agricultural statistics. They conclude that the nature of CBC sites has changed very little over the years, but that the national indices are dominated by what happens on lowland farmland habitats in the south and south-east of England.

The problem of how to detect regional differences for a particular species is now examined, using farmland data for the Sedge Warbler. This species is an example of a summer migrant that occurs in large numbers at several observatories each year, allowing comparisons to be made between migration and breeding censuses. In order to examine the CBC index for geographical consistency, we divide the census plots into three arbitrary and broad regions and calculate population indices for each region. Fig. 2.3.1 shows an example of such a breakdown where the regions roughly represent southern Britain (67 sites), Wales and Northern England (17 sites) and Scotland (12 sites). Most sites are in region 1. This is due to two reasons. Firstly, CBC plots are usually near to centres of high human population, and secondly, the Sedge Warbler is a lowland dwelling bird. Table 2.3.1 shows the population indices for each region, with the number of

Fig. 2.3.1 Regional Breakdown of Sedge Warbler CBC plots.

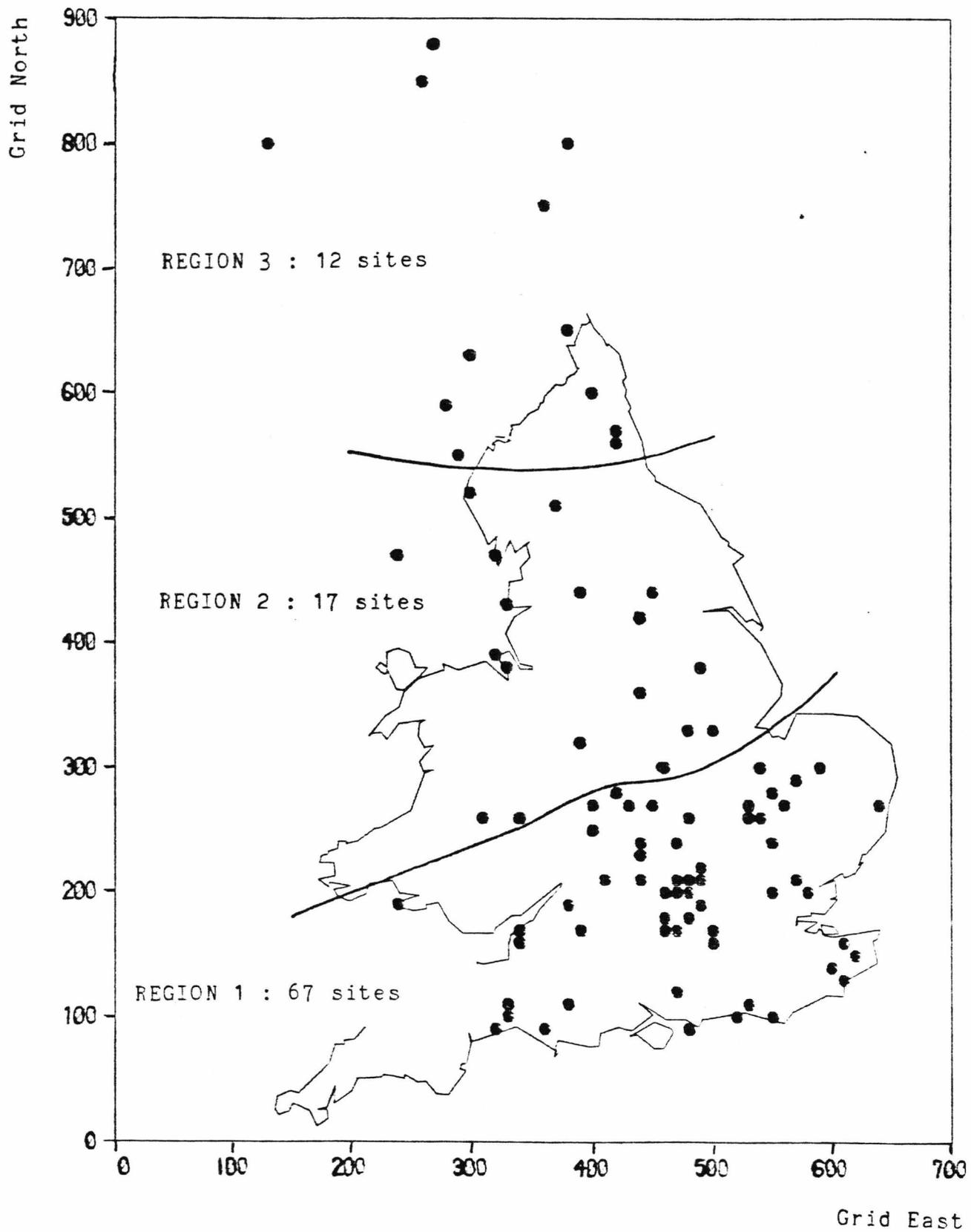


Table 2.3.1 Regional and National CBC indices for the Sedge Warbler. ( The number of territories used for calculating the index is given in brackets )

<u>Year</u>	<u>National Index</u>	<u>Region 1</u>	<u>Region 2</u>	<u>Region 3</u>
1962	60.4 ( 42)	60.5 ( 39)	120.0 ( 2)	32.0 ( 1)
1963	59.0 ( 69)	57.4 ( 59)	120.0 ( 7)	64.0 ( 3)
1964	80.3 (111)	78.8 ( 95)	120.0 ( 8)	106.7 ( 8)
1965	78.3 (130)	80.3 (108)	75.0 ( 75)	66.7 ( 10)
1966	100.0 (170)	100.0 (100)	100.0 ( 10)	100.0 ( 15)
1967	68.8 (116)	68.6 ( 85)	100.0 ( 21)	66.7 ( 10)
1968	118.0 (269)	110.9 (191)	181.0 ( 46)	93.3 ( 32)
1969	66.9 (105)	63.9 ( 71)	86.5 ( 23)	60.2 ( 11)
1970	84.3 (150)	74.3 ( 48)	150.3 ( 36)	65.7 ( 12)
1971	68.0 (108)	61.5 ( 77)	112.7 ( 26)	82.1 ( 5)
1972	55.4 ( 79)	58.2 ( 61)	56.4 ( 12)	32.8 ( 6)
1973	39.3 ( 36)	39.8 ( 24)	56.4 ( 12)	13.1 ( 0)
1974	42.6 ( 39)	41.5 ( 25)	65.8 ( 14)	13.1 ( 0)

territories involved, for the first thirteen years of the census.

The indices shown in Table 2.3.1 are calculated in the standard way, relating the percentage increase or decrease in the number of territories in consecutive years to a base value of 100 in 1966. The indices are plotted in Fig. 2.3.2. One can see that the three arbitrary regions show the same general pattern of variation as the national index.

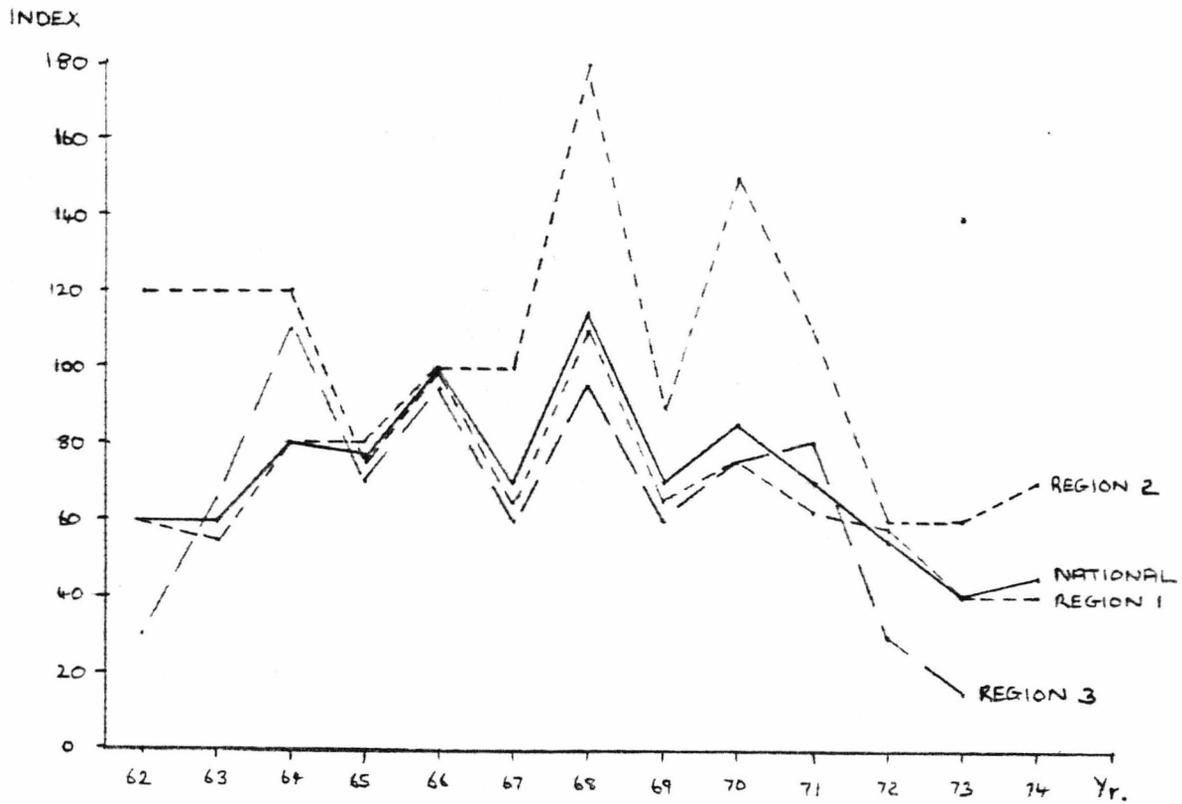
The degree of consistency between the regions may be roughly assessed by the correlation matrix given below:

	Region 1	Region 2	Region 3
Region 2	0.65		
Region 3	0.82	0.62	
National Index	0.99	0.72	0.54

All the above correlation coefficients (based on 13 observations) are significant at the 5% level. The high number of sites found in region 1 is reflected in the correlation coefficient of 0.99 with the national index. This supports the view of Fuller et al (1985) that the national index is dominated by lowland Britain.

One problem with calculating regional indices is that of the small number of sites available in areas other than lowland Britain. For example, even the national index for the Sedge Warbler over the years 1962-74 was calculated on no more than 40 sites. In some years the number of sites censused was as low as 11, limiting the number of regional breakdowns available.

Fig. 2.3.2 3-Region breakdown for Sedge Warbler CBC index.



Since the total numbers of territories each year are used to calculate the indices, the size of individual sites is ignored. Just as regions with a large number of sites dominate the final index, so too do sites that hold large numbers of territories. The BTO try to encourage observers to census sites of similar size each year, if not the same sites, but a glance at the full data set for the Sedge Warbler shows some sites containing only one territory for a number of years, while another may average approximately twenty territories. The variation on this latter site will influence the final index more than the consistency of the former site. One could, of course, overcome this problem by using an estimate of density, rather than an estimate of the total number of territories at each site.

A simple means of examining data for each inter-year comparison is to produce a scatter plot, as shown in Fig. 2.3.3, where the pair of years 1971 and 1972 are considered. There appears to be a cluster of six sites which contain approximately three times as many territories, on average, than the remaining 28 sites.

Fig. 2.3.3 Scatter plot of territory counts for 34 sites censused in both 1971 and 1972. (Sedge Warbler)

(numbers on plot indicate multiple points)

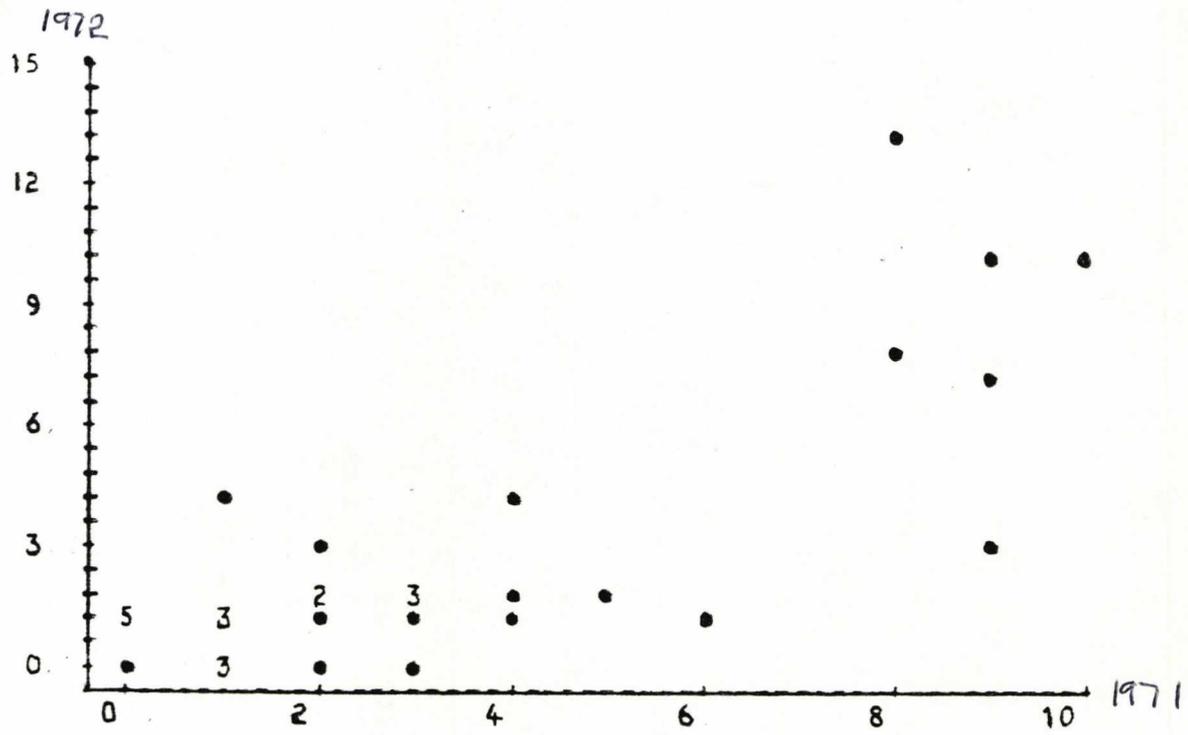
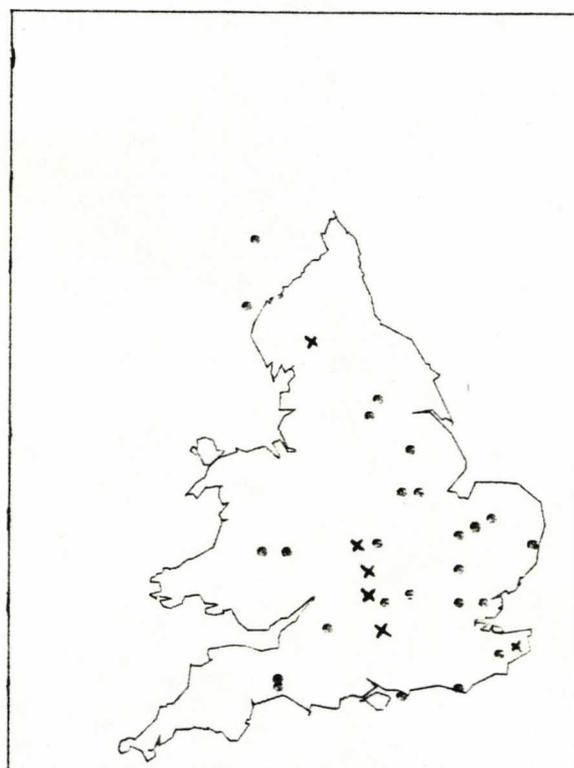


Fig. 2.3.4 Sedge Warbler sites censused in both 1971 and 1972.



The six 'large' sites are shown on the map of Fig. 2.3.4. No obvious geographical region is represented by these sites, and the Populations Section of the BTO are not aware of any likely factor linking the six sites. All six sites are near to, or include water margins but the type of farmland associated with the sites varies. Clearly, many more investigations of this kind could be carried out.

Computerisation of all the CBC data for all species would enable a critical examination of the current method of producing national indices based on territory counts to be made quite easily. Possible starting points for such an investigation are found in this section, i.e. the influence of subsamples of sites on the national index, and the influence of individual sites with large numbers of territories.

Regional indices would offer further possibilities of linking migration data to CBC data - e.g. migrants to particular areas of Britain may pass through particular observatories.

For the purpose of this thesis however, we have to use the available national indices while bearing in mind that these will be dominated by population dynamics of the southern part of Britain.

#### 2.4 Canadian Migration Indices

The main objective of this thesis is to examine population monitoring using migration data. This problem was tackled by the Canadian Wildlife Service (Hussell and Risley, 1978) using data from Long Point Bird Observatory, Ontario. Daily counts of grounded migrants were used which were taken on a more formal basis than those taken at British Observatories, i.e. regular censuses were made each day in the same areas of the observatory recording area.

The indices that were calculated were compared with results from Breeding Bird Surveys (BBS) of North America, and also Christmas Bird Counts - see Hussell and Risley (1978).

Five indices were considered, for each species studied:

(i) MIS

MIS is the total number of birds observed in a particular year divided by the mean number of birds observed each year over a number of years. The denominator is calculated after smoothing the data within each year by taking a 7-day moving average. Allowances are made for missing values by weighting and iteration methods (Hussell and Risley, 1978).

(ii) MIL

MIL is calculated as MIS, but the transformation  $\log(y_i + 1)$  is used where  $y_i$  is the count for day  $i$ . This is used as an attempt to make the distribution of observed

counts symmetrical.

(iii) MIL2

MIL2 is an index based on data from 2 sites at Long Point, calculated as MIL.

(iv) MI-SW and (v) MI-LW

These are weather-corrected indices and are discussed in chapter 4 where a detailed study of weather and observatories data is presented.

The species considered by Hussell and Risley are not found in Europe and with such a marked geographical difference between British observatories and Long Point the relevance of the Canadian results is limited. However, a brief report of the Canadian results is presented below.

For most species, the spring MIL and MI-LW indices showed high correlation with BBS indices. Autumn indices of any type showed low correlation with Christmas Bird Count indices but the latter showed high correlations with the previous spring migration indices.

Examples of the indices MIS and MIL are presented below using British observatories data. Table 2.4.1 summarises the correlations between MIS and MIL indices for a sample of British observatories data. CBC farmland indices are used - woodland indices are unavailable for some species and in some years.

Table 2.4.1 Correlation coefficients of Canadian migration indices vs. CBC indices.

n = number of years.

Observatory and Species	Spring				Autumn			
	MIS (n)		MIL (n)		MIS (n)		MIL (n)	
<u>Dungeness</u>								
Blackcap	-.33	12	.01	12	.69*	12	.73*	12
Sedge Warbler	-.23	12	-.22	12	.35	12	.37	12
Spotted Flycatcher	.06	12	.20	12	-.16	12	-.24	12
Whitethroat	.81*	12	.78**	12	.97**	12	-.97**	12
Willow Warbler	.04	12	-.04	12	-.13	12	-.05	12
<u>Portland</u>								
Sedge Warbler	.19	12	.04	12	.02	12	.29	12
Spotted Flycatcher	-.05	7	.05	7	.44	7	-.01	7
<u>Spurn</u>								
Sedge Warbler	.73	7	.68	7	.70	7	.69	7
<u>Bardsey</u>								
Sedge Warbler	-.22	7	-.38	7	-.08	7	.30	7

\* Significant at the 5% level, \*\* Significant at the 1% level

Table 2.4.1 clearly shows that for only two species (Blackcap and Whitethroat) do the Canadian indices reflect the population trend as recorded by the CBC. The Whitethroat correlation is not surprising since the population suffered a severe decline after 1968 (Winstanley, Spencer and Williamson, 1974) and the correlation coefficient is dominated by this feature. We would expect such large fluctuations to be picked up even by crude indices of the type used here.

Table 2.4.1 also shows a correlation between autumn Blackcap migration at Dungeness and the CBC index. Blackcap indices showed a steady increase over the period 1961-72. Ornithologists are not able to offer an explanation as to why spring data fail to show similarly high correlations.

Looking at sample correlation coefficients between sequences of indices gives an idea as to how long-term trends in the indices correspond. The overall picture from the above work is that, with two exceptions, migration indices are unrelated to CBC indices. However, the sequences are short - a maximum length of 11 years - and when more data are available it may be possible to detect relationships between migration and CBC indices.

The Canadian workers appeared to have based the success or failure of their migration indices solely on correlation coefficients. Alternative techniques of comparing indices are investigated below.

In order to examine year-to-year changes in both migration and CBC indices, one can convert each sequence into a sequence of

binary observations where each observation represents either an increase or a decrease from the previous year. The two binary sequences can be summarised in a 2 x 2 table:

	CBC decrease	CBC increase	total
migration index decrease	a	b	a + b
migration index increase	c	d	c + d
total	a + c	b + d	N

A test statistic which measures the association between the two indices is calculated by the standard formula

$$X^2 = \frac{(ad - bc)^2 N}{(a+b)(c+d)(a+c)(b+d)}$$

If we assume that each binary sequence is an independent trials process, then under the null hypothesis of no association  $X^2$  has asymptotically a chi-square distribution on one degree of freedom. The assumption that successive observations within each sequence are independent is not valid in this context since the second of any pair of years is used again as the first year of the next pair. In this situation  $X^2$  will be inflated and the chi-square approximation will lead to the null hypothesis being rejected more often than expected. For data of the type considered here, Tavaré and Altham (1983) give a deflation factor for  $X^2$  which leads to a statistic with an asymptotic chi-square distribution with one degree of freedom. The assumptions which one needs to make concerning the nature of the dependence between successive

observations and the calculation of the deflation factor are given in Tavaré and Altham's paper. For the data of this section, there is no need to deflate  $X^2$  since all comparisons of indices gave a very low  $X^2$ , suggesting no association between indices.

The above finding should not be viewed negatively however since the sequences of binary observations are a maximum of 10 observations in length. If one sequence has approximately equal numbers of both outcomes of the binary process then, to show a 'significant' association the other sequence needs to show identical outcomes for all but one case. This, of course, is before the application of Altham's deflation factor, which would reduce the possibility of observing association further. While this technique of comparing indices is not very useful with the short sequences considered here, it does offer ornithologists a new approach to relating one index with another, in addition to the traditional correlation coefficient.

Table 2.4.2 presents the  $X^2$  statistics (no deflation factor applied) for the data used in this section.

With short sequences of observations that are considered in this section, distributional assumptions on which the significance of the product-moment correlation coefficient are based are difficult to check. One way round this problem is to use a non-parametric test such as Spearman's rank correlation coefficient. For each index the observations are ranked and any two indices

may be compared by the rank correlation coefficient for which exact P-values are available. For the data of this section the incidence of significant correlation coefficients is exactly as in Table 2.4.1 apart from the fact that no coefficients reached significance at the 1% level.

Table 2.4.2  $\chi^2$  Statistics for measuring association between  
CBC indices and migration indices.

Observatory and Species	Spring		Autumn	
	MIS	MIL	MIS	MIL
<u>Dungeness</u>				
Blackcap	.02	.42	.42	.02
Sedge Warbler	.01	.18	1.41	3.27
Spotted Flycatcher	.18	.02	.18	.00
Whitethroat	.00	.18	.01	.02
Willow Warbler	.97	.00	.97	.97
<u>Portland</u>				
Sedge Warbler	.42	.00	.01	.18
Spotted Flycatcher	.70	1.70	.05	.05
<u>Spurn</u>				
Sedge Warbler	.31	.31	.31	.31
<u>Bardsey</u>				
Sedge Warbler	1.70	.06	.31	.05

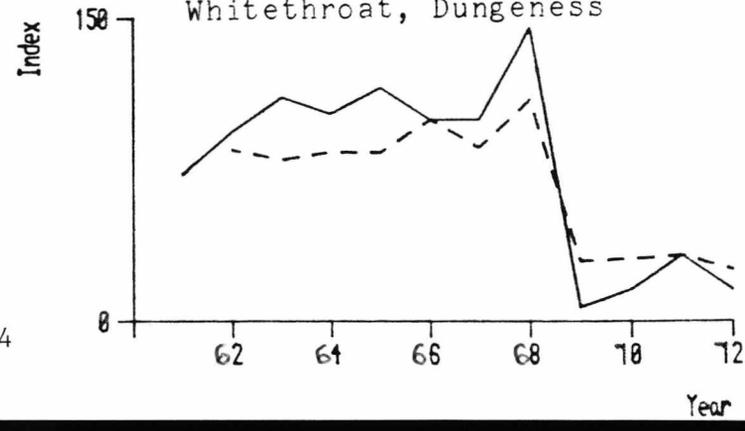
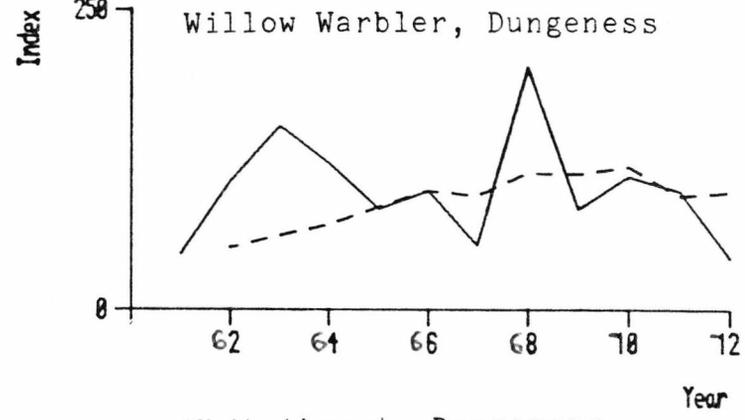
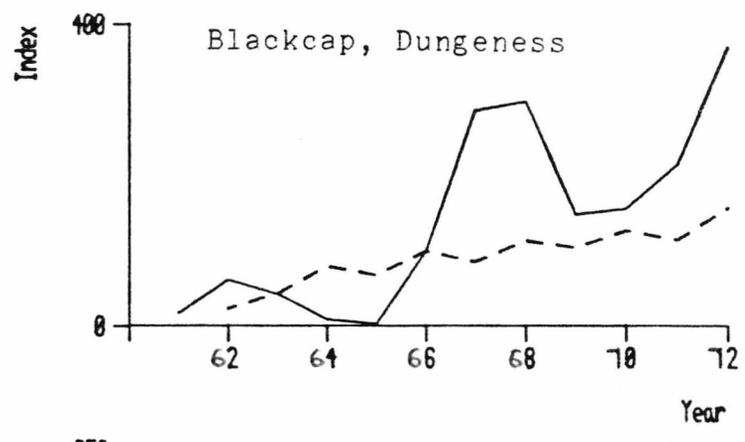
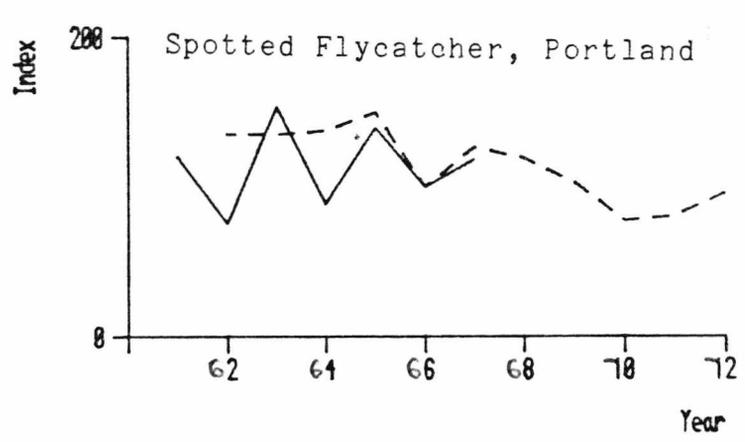
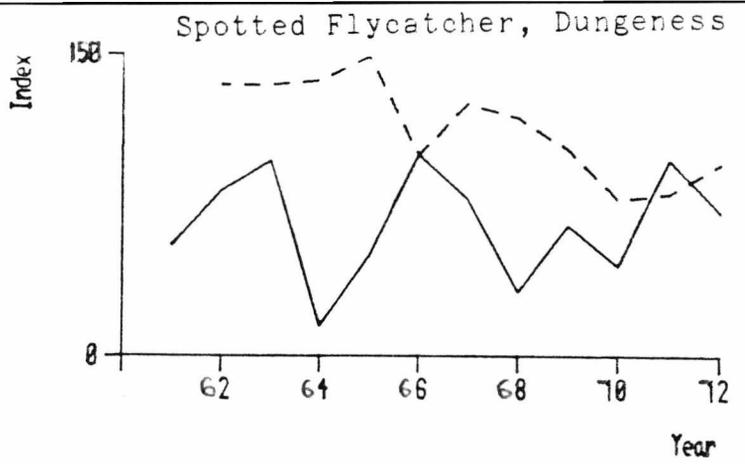
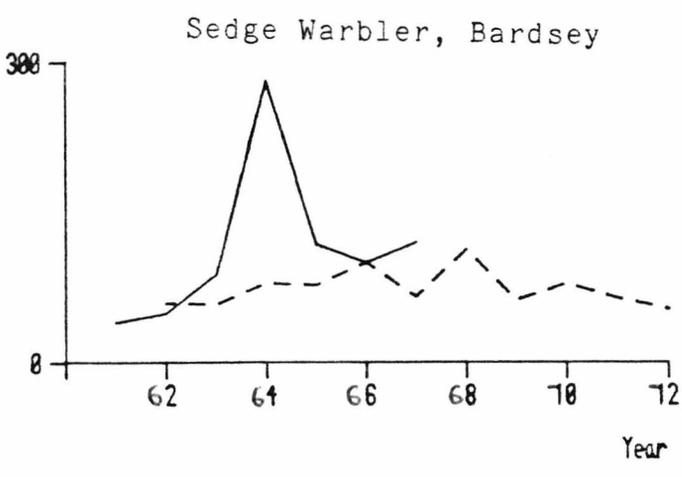
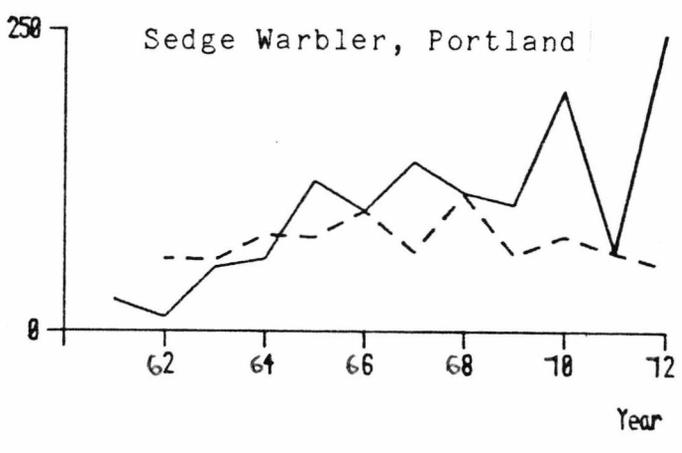
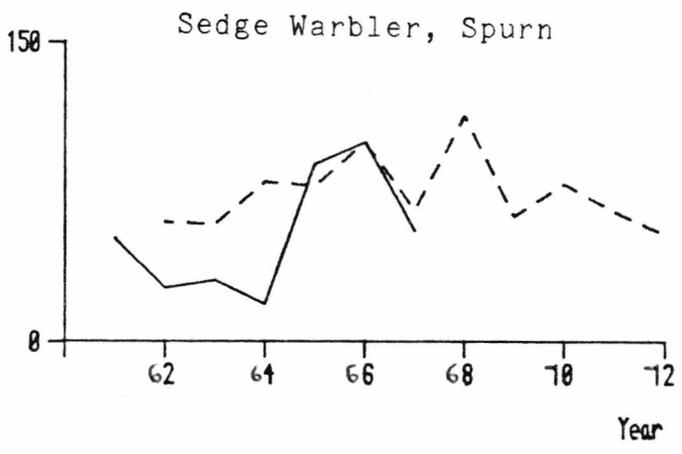
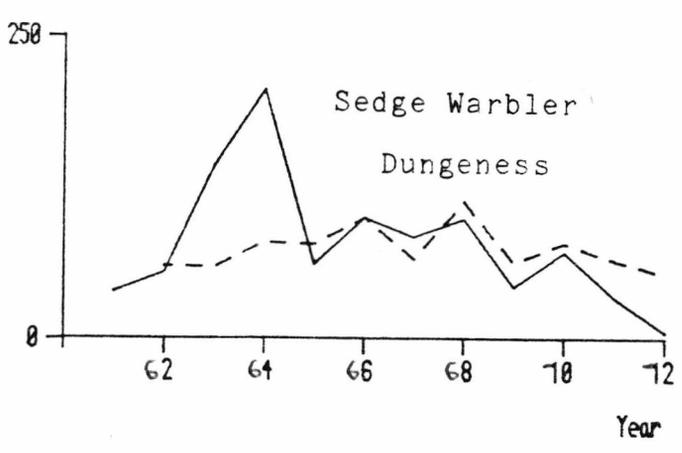
We have seen that Canadian-type migration indices calculated from individual observatories are, in most cases, uncorrelated with CBC indices. Sedge Warbler counts are available from four observatories in all, and an investigation was carried out in order to see whether or not more success could be achieved by combining indices from Dungeness, Portland, Bardsey and Spurn. The CBC index for the Sedge Warbler was regressed on the MIS and MIL indices (spring and autumn) for the four observatories, but no further success was achieved. It was not possible to find any linear relationship between migration indices and CBC indices.

It was thought possible that the CBC index might be related to the migration index of the preceding autumn as well as the preceding spring, thereby introducing a measure of winter mortality into the index, but again no relationship was found.

The indices which are studied in this section are displayed in Figs. 2.4.1 - 2.4.4. In each plot, the dashed line represents the CBC farmland index and the solid line represents the migration index.

Multiple regression analysis was used as a means of combining different indices for other species in order to see whether a relationship with the CBC index could be found. Spring and autumn indices were included in the regressions and, for the Spotted Flycatcher, indices from Dungeness and Portland. Where sufficient degrees of freedom were available, indices of the previous autumn were also included. For each species, the combined migration index was calculated by including

Fig. 2.4.1 Canadian migration indices (MIS, spring) and CBC indices.



(MIL, autumn) and CBC indices.

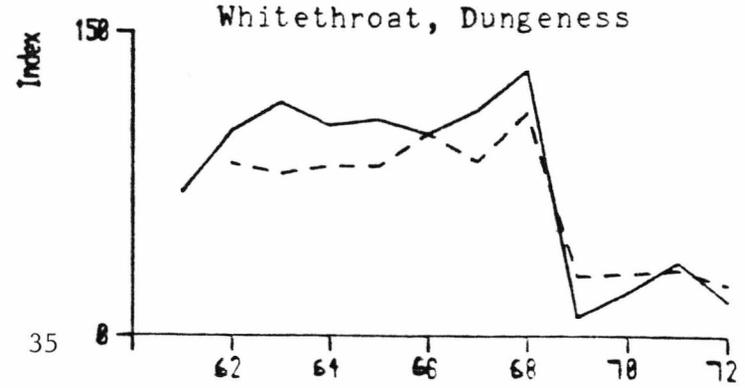
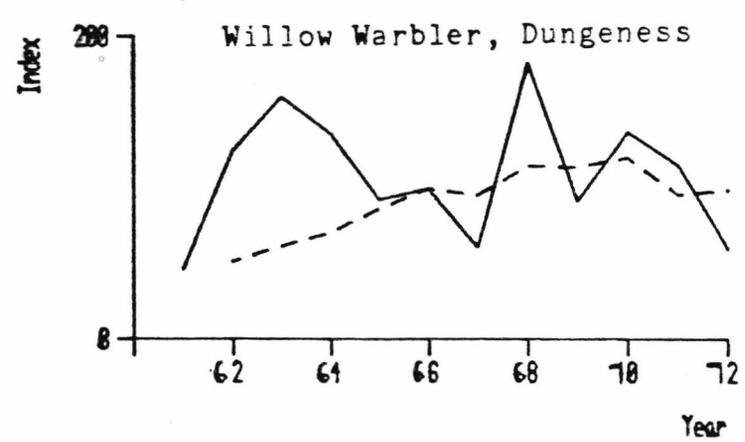
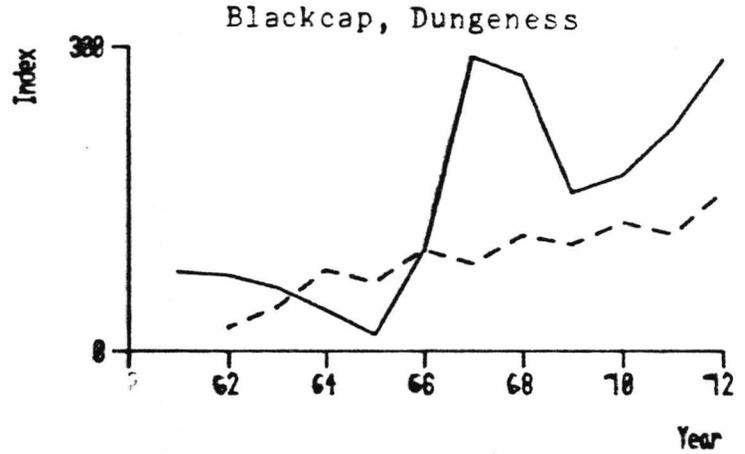
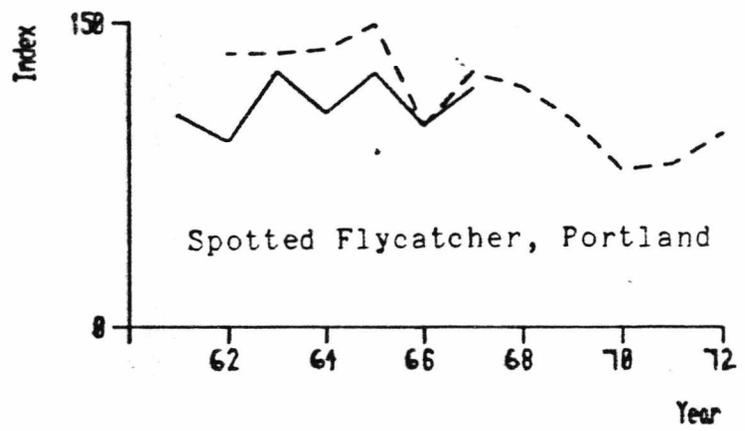
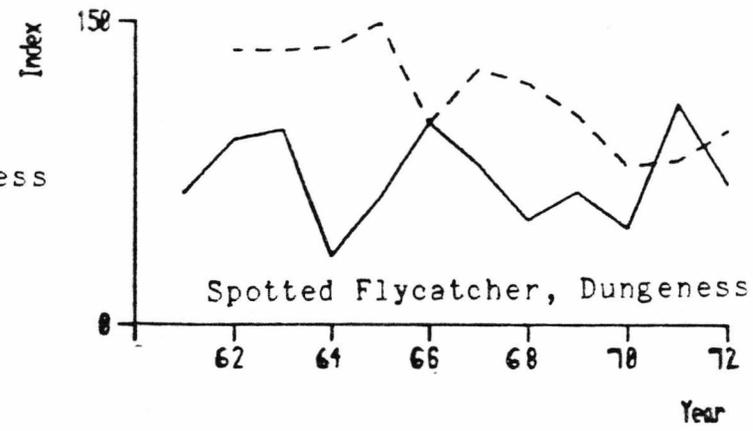
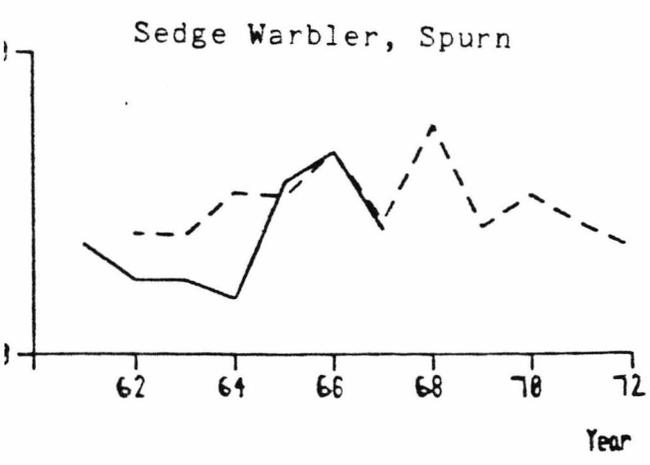
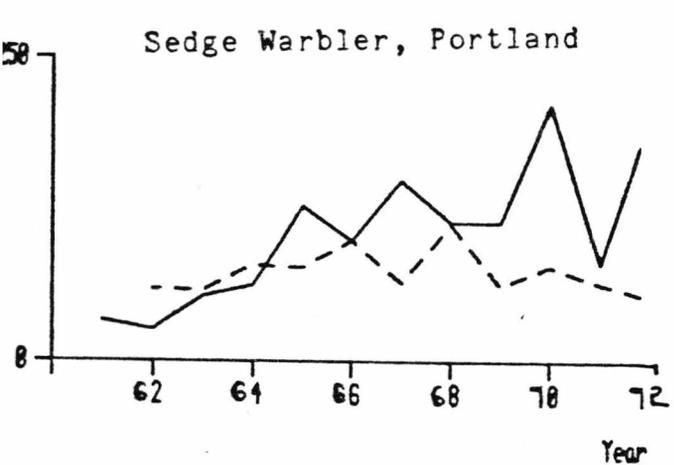
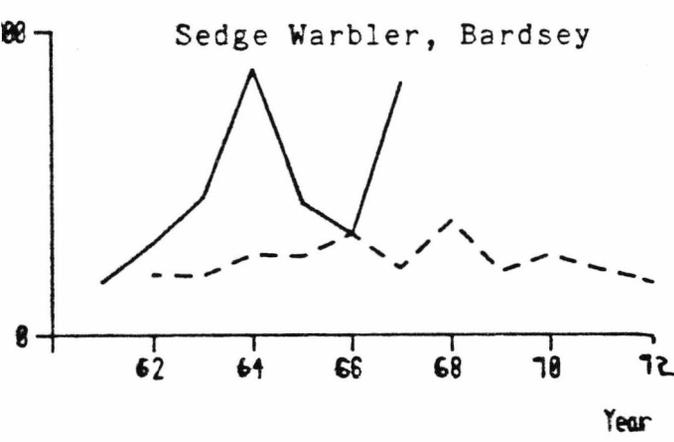
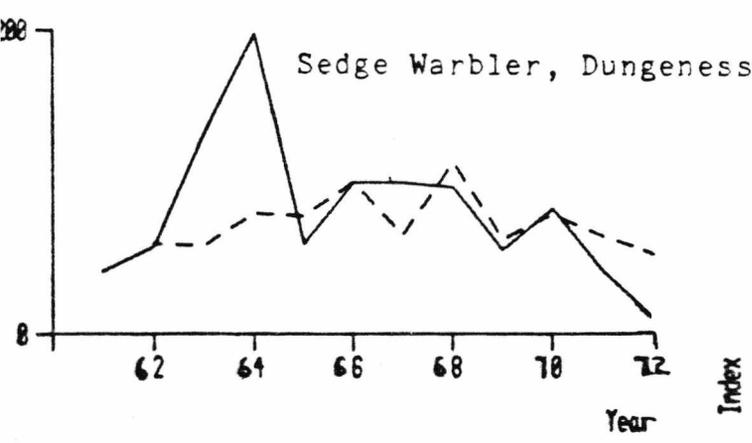


Fig. 2.4.3 Canadian Migration Indices (MIS, spring) and CBC indices.

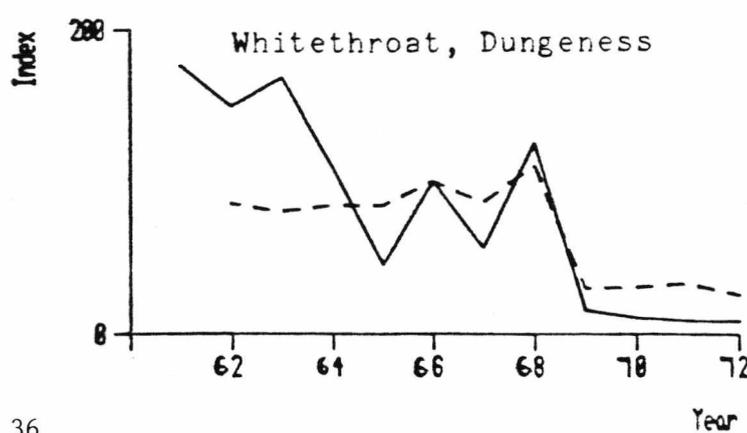
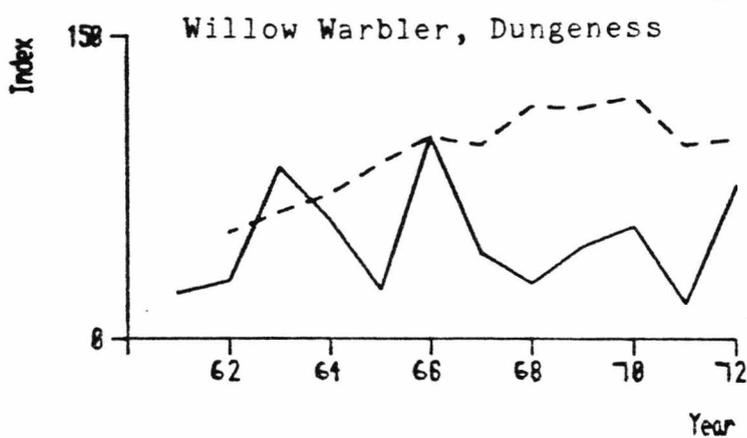
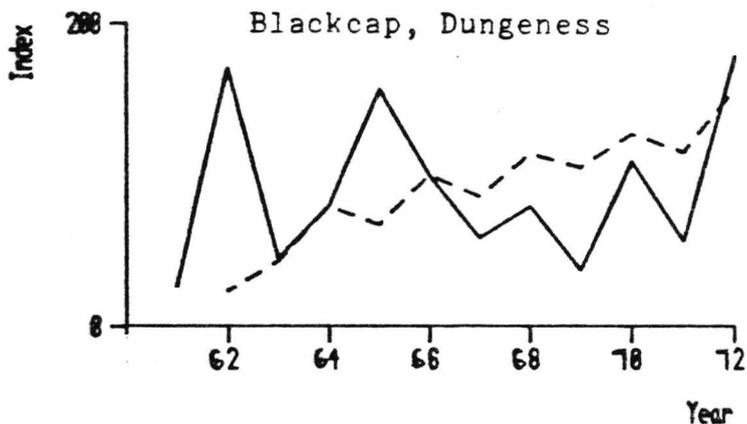
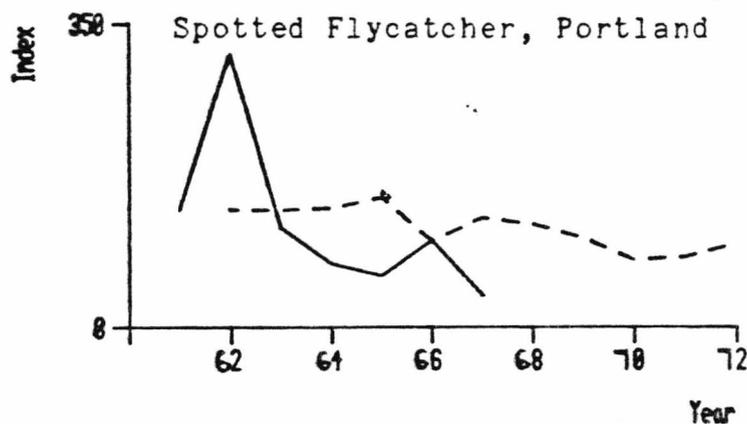
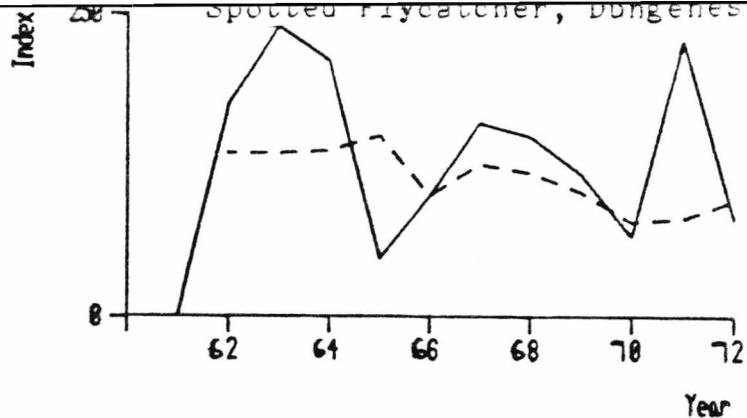
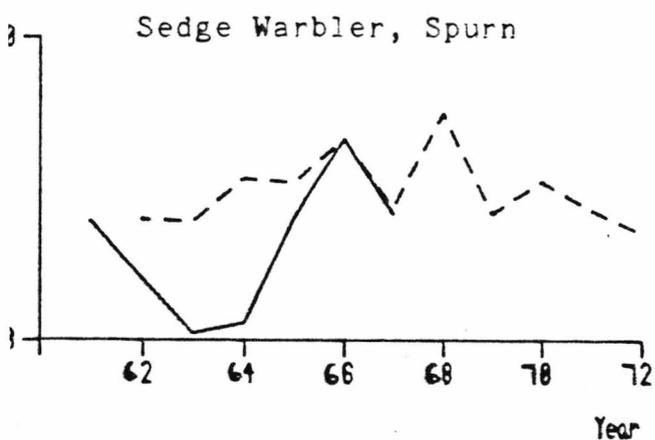
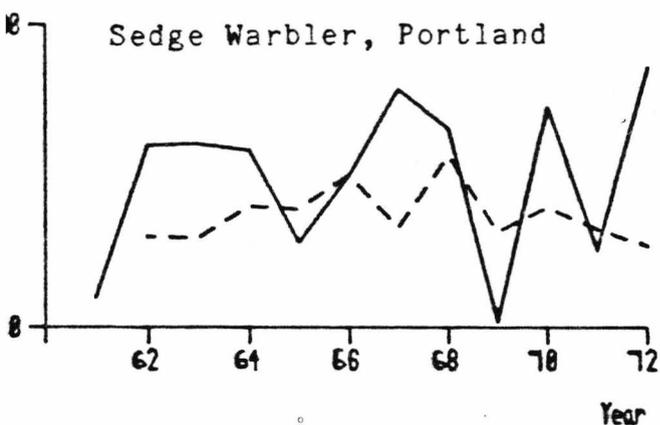
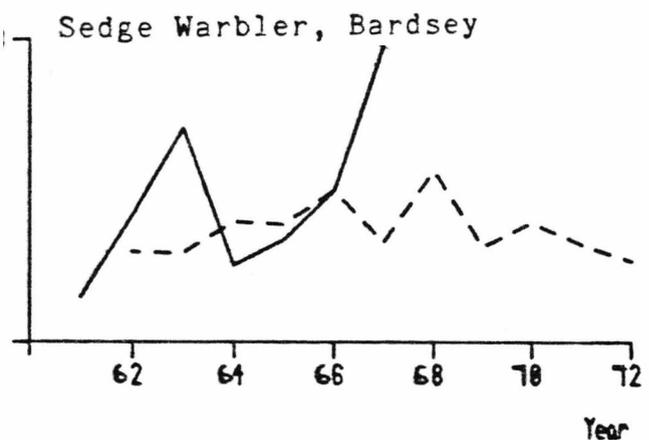
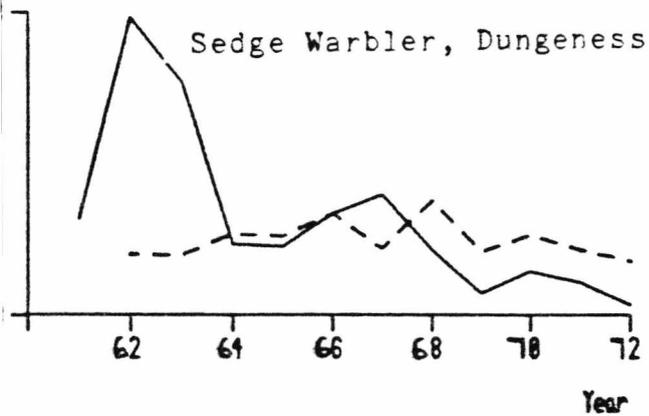
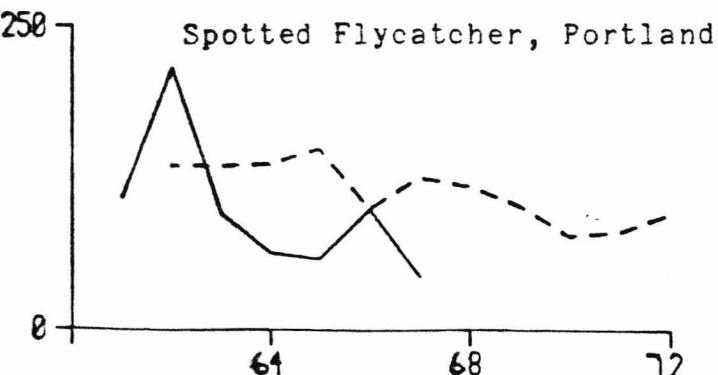
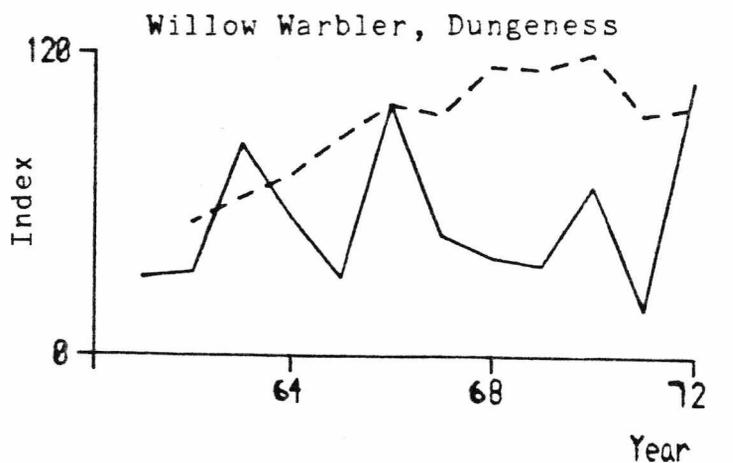
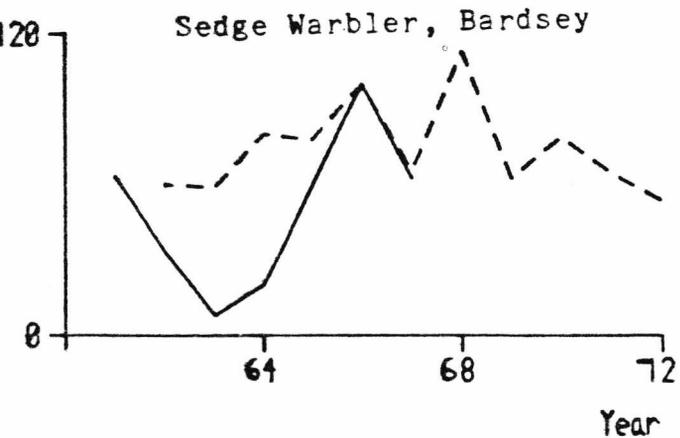
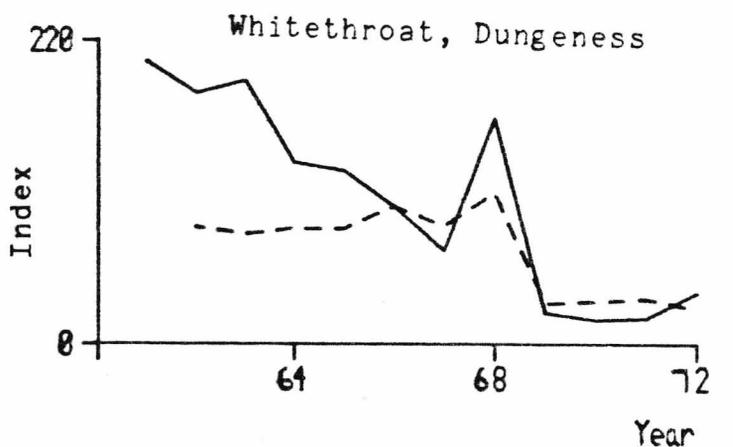
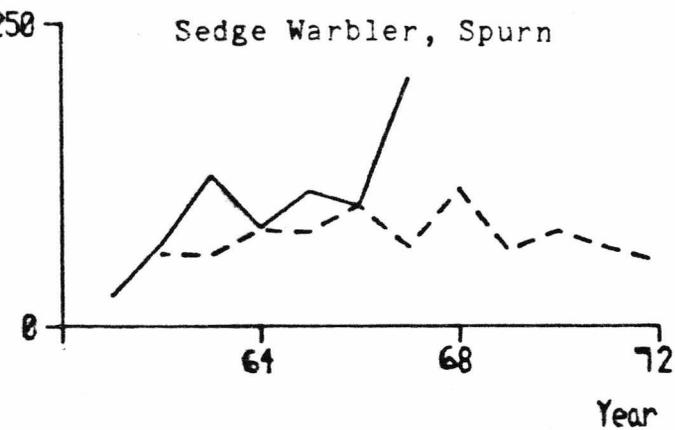
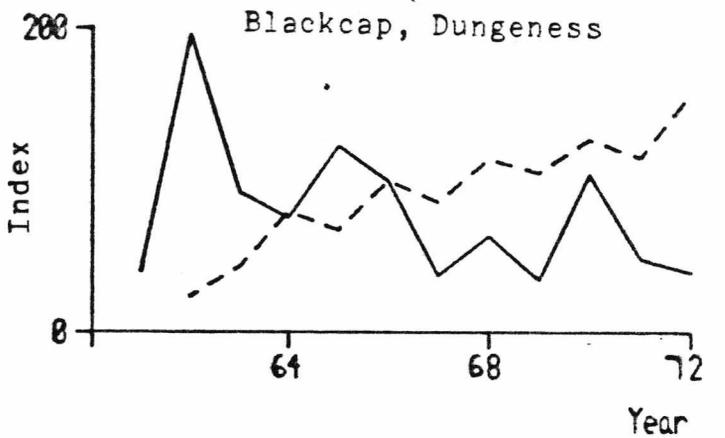
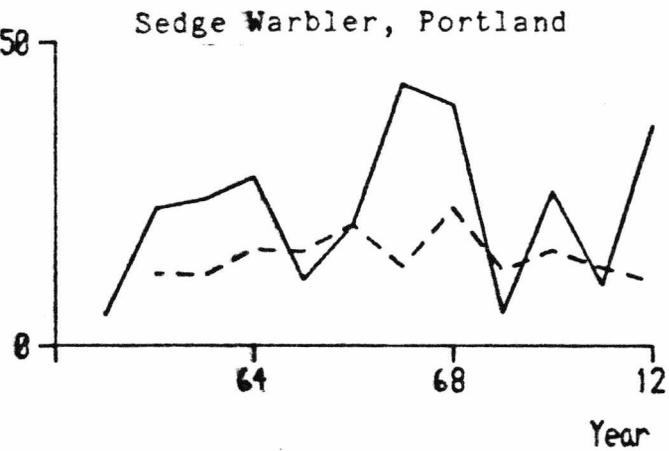
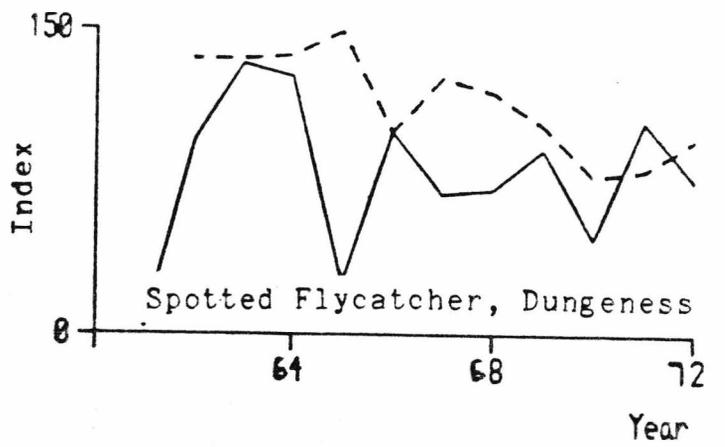
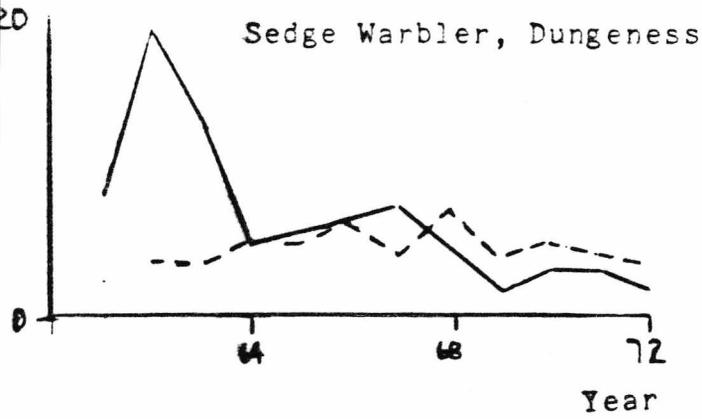


Fig. 2.4.4 Canadian migration indices (MIL, autumn) and CBC indices.



all possible indices in order to achieve the optimum correlation with the CBC index. The combined indices for the Spotted Flycatcher and Willow Warbler showed no significant improvement in correlation with CBC indices over the individual indices. A similar picture emerged from the regressions using Whitethroat and Blackcap data. Significant correlations were found, but no significant improvement was achieved over individual indices.

One interesting point to emerge from the regression is that there appears to be no advantage in comparing CBC indices with migration indices of the previous autumn. In the following chapters of the thesis however, this comparison will again be made in case this result is specific to the indices studied in this section.

In this section strong evidence has been found to suggest that Canadian-type migration indices are not suitable for use with British data for population monitoring. Further chapters of this thesis examine alternative ways in which population monitoring may be achieved by using observatories data.

#### 2.5 A summary of other attempts at population monitoring using migration data

The first mention of this topic in British literature was by Williamson (1963). He classified different types of migratory movement and suggested that one such movement - 'irruption movement' is caused by a population size reaching a critical level. No investigation of this theory using data appears to

have been made however.

Sharrock (1969) studied annual totals of Grey Wagtails from several observatories and proposed a crude migration index based on weighted totals of daily counts. The weighting gave observatories which record large numbers of Grey Wagtails more influence than others. Sharrock found that the index reflected population fluctuations detected by the Nest Record Scheme (NRS) of the BTO. For the work of this thesis the yardstick by which migration indices are assessed is the CBC index, but results from the NRS may provide an alternative way forward for future research into this topic. Sharrock's results are certainly encouraging since Grey Wagtail counts will be more influenced by variations in manning levels at observatories than counts of nocturnal migrants - Grey Wagtail counts are made as the birds fly over the observatory giving loud and distinctive calls.

Langslow (1978) examined graphically the numbers of Blackcaps migrating through ten British observatories over the period 1970-76. He found that Blackcaps were becoming more common at observatories, particularly during autumn. In section 2.4 of this thesis it was noticed that both the crude Canadian-type index and the CBC index showed a similar trend for the period 1961-72. Apart from the Whitethroat, this was the only species to show a correlation between migration data and CBC indices but other species did not show any well-defined population trends such as the Blackcaps, which seem to have continued beyond the period of study of section 2.4.

Cowley (1979) describes how annual totals of visible migration counts of Sand Martins show similar fluctuations from year to year as estimates of colony sizes at several British colonies. This is another encouraging example of migration data showing similar features to data from other sources, particularly as the Sand Martin is a diurnal migrant. The point made above that more success is expected with nocturnal migrants is also made by Cowley.

Hussell and Risley (1978) give a number of references of work on this topic from various parts of the world. They found that most studies used seasonal totals or mean daily counts as migration indices, or mean daily ringing totals.

Mueller and Berger (1967) and Mueller, Berger and Allez (1977) looked at movements of Goshawks in relation to their population size in the US, and other hawks were studied by Hackman and Henny (1971) in Maryland.

Busse (1973) studied several species that were ringed on the Polish Baltic coast and compared the numbers with estimates of breeding numbers.

Very little attention seems to have been paid to statistical aspects of migration data and associated weather conditions. The indices mentioned in the above references can therefore only hope to show long-term changes in population size.

Swedish ornithologists have recently begun work in this area, see e.g. Hjort and Lindholm (1978). Svensson (1978) concludes that monitoring population levels by migration sampling offers no advantages over breeding bird surveys of birds that can be censused by other means.

From the above summary it is clear that the population monitoring area is wide open. We now examine briefly other uses of observatories data in order to establish what other information has been gained from these data in the past.

#### 2.6 Radar studies of migration and weather

Many studies have been made of particular species and the weather conditions associated with their occurrence at observatories and other sites in Britain. This section summarises these studies. A detailed description of how radar data are used in studying migration is given in Eastwood (1967).

Lack (1960a) reviewed over a hundred papers on this subject, paying particular attention to the types of weather found to be influencing different species movements.

In the late 1950's, methods of studying migration and weather changed dramatically with the availability of radar equipment. This new technology enabled flocks of birds to be traced over considerable distances and in most types of weather. Although individual species cannot be identified by radar, birds previously invisible to observatories could now be studied. It is sometimes

possible to infer from ground records which species are being recorded by radar. Only rough guesses can be made however.

An important feature of radar studies is that one is able to assess the extent to which observatories data reflect the volume of migration as a whole. Riddiford (1985) looked at Willow Warbler records at Dungeness in relation to radar records (Parslow, 1962, 1969) and found that large numbers of grounded birds occurred when overhead migration was at its greatest. By studying only large counts of grounded birds, Riddiford throws no light on the problem of large numbers of migrants passing over the top of observatories when weather conditions permit. This problem is returned to later.

In Britain, two major studies of migration have been made using both radar data and observatories data. Parslow (1962, 1969) studied passerine movements over the English Channel and Lack (1959, 1960b, 1963a,b,c) made an extensive study of many types of migration across the North Sea. Other studies using radar data include those of Evans (1966b) and Wilcock (1965) using data from NE England and eastern Scotland respectively.

Radar has also been used abroad, e.g. North America (Nisbet and Drury, 1967) and Sweden (Alerstam, 1973).

The only statistical analyses that have been performed in the study of radar data and weather are the multiple regression analyses of Lack (1960b) and Nisbet and Drury (1968). In both of these analyses a measure of migration, taken from radar data,

forms the response variable and weather variables form the explanatory variables. In calculating regression equations, little attention is paid to the implications of using a set of explanatory variables which are likely to be correlated with one another, as well as successive observatories being auto-correlated. Such situations can lead to multicollinearity, a problem that is considered in more detail in chapter 4.

Since the publication of Lack's review in 1960, many smaller studies have been undertaken. This section has only reported the major studies involving observatories data.

## 2.7 Other studies of British observatories data

Many subsets of observatories data have been analysed by ornithologists. Most of these studies have been of individual species at particular observatories and published in observatory annual reports. Those appearing in ornithological journals are summarised below.

The forerunner of a series of papers on the movements of birds through observatories was that by Davis (1967) who used daily counts of grounded migrants from seven observatories in order to assess the length and pattern of migration seasons for five Warblers of the genus *Sylvia*. One of the observatories studied was Skokholm which was the subject of a similar study by Lack and Lack (1966) who used daily ringing totals. *Sylvia* Warblers were also studied using data from Cape Clear (Sharrock, 1968).

Hope-Jones (1975) examined Redstart data from eight observatories. By incorporating ring recovery details into his analysis he was able to postulate the migratory routes of both British Birds and continental passage birds. Durman (1976) took a similar approach in a study of Ring Ouzel migration. Durman also studied Ring Ouzel data from Fair Isle observatory and was able to identify different passage times for male and female birds.

Hope-Jones et al (1977) studied Pied Flycatcher migration using data from seven observatories.

Langslow (1979) analysed ringing data on Blackcaps and found evidence that birds passing through east coast observatories in autumn are of continental origin, suggesting that south coast observatories are likely to offer more reliable information on British birds. In the opinion of ornithologists this feature of east coast observatories generalises to most migrant species.

The species mentioned above are common migrants at most observatories. Scarcer birds have also been studied, e.g. Black Redstart (Langslow, 1977) and Red-Breasted Flycatcher (Radford, 1968), the latter being purely a passage bird.

Insley and Boswell (1978) combined daily ringing totals from Sandwich Bay and Dungeness observatories with similar data from eight other south coast ringing sites. They attempted a cross-correlation analysis in order to compare arrival times of Reed and Sedge Warblers among the ten sites. Missing data, however, prevented their drawing of any conclusions from significance tests.

One has to be careful when performing cross-correlation analyses on data of this type since long-term trend in the series of observations may cause the correlation coefficients to be inflated.

Very few studies have been made of birds seen migrating over observatory recording areas. Most observatories make some effort to record such birds though and species which receive most attention include hirundines and wagtails. This is probably due to the ease of identifying these species by both sight and call.

The ability to identify diurnal migrants depends on the height of flight. Radar studies have revealed that with following winds, birds tend to fly above the visible range, thus not being recorded.

Evans (1966a) proposed a model which enabled estimates to be made of the numbers of birds 'coasting' during the day, along the NE coast of England. No indication of how successful his complex deterministic model was given, however. A brief summary of other studies of visible migration data is given by Snow (1953). Alerstam (1978) investigates similar data from Sweden.

In Britain, Spurn Head observatory is the best suited to the study of diurnal migration since it is situated at the southern tip of a narrow peninsula which concentrates birds moving south. Even at Spurn however, the collection of such data requires a constant watch to be kept. At other observatories a watch needs to be kept over a broad area for the data to be of any use and clearly the level of manpower is a critical factor.

## CHAPTER 3

### DESCRIPTION OF OBSERVATORIES DATA AND ASSOCIATED WEATHER DATA

#### 3.1 Introduction

This chapter represents the first detailed account of daily counts of grounded migrants. An examination of how counts are made, from a statistical viewpoint, is presented and inherent defects and biases found in the data are set out. These features of the data are dealt with in later chapters where statistical models for daily counts are proposed.

The first half of the chapter considers features of the migration data that may influence the development of indices which monitor fluctuations in population level. Such features include the daily census methodology, the timing of migration, waves of migration, the influence of manning levels at observatories and the occurrence of occasional very large counts. It will be seen that individual features lead to interesting discoveries from the data. The presentation of the use of simple statistical methods in this chapter helps ornithologists to see how they can best extract information from observatories data.

Weather data from Dungeness and Portland observatories are also described, the aim being to assess which data may usefully be analysed together with migration counts in later chapters.

### 3.2 Daily counts of grounded migrants

#### 3.2.1 General overview

We concentrate on common nocturnal passerine migrants which are insectivorous birds wintering in or to the south of the Mediterranean Basin. In spring the birds move north to where daylight hours are longer, enabling extended feeding time, which is needed as adults rear young birds. As autumn approaches in northern Europe the days shorten and temperatures fall, causing birds to return south. Biological aspects of migration can be found in many books, e.g. Griffin (1974). Extensive research has been performed on navigation - see, for example, Matthews (1968).

Since observatories are situated on the coast, the recorded migrants are nearly all on passage in both spring and autumn. Although it is generally impossible to witness the actual arrival of birds at observatories without radar equipment, it is occasionally possible to observe dawn arrivals off the sea in spring. Most arrivals take place overnight, as shown by Parslow's (1969) radar studies on the south coast. In spring, one can track a broad front of migrants on a radar screen as they cross the Channel, often reproducing the outline of the French coast for two to three hours. This movement is obviously influenced by weather conditions over the Channel, and these are discussed in detail in Parslow's paper. Generally, if weather conditions are unfavourable when they approach the coast of Britain, i.e. wet, windy and cloudy, then birds are likely to land at the first opportunity at sites which

inevitably include observatories. If calm, clear and dry conditions prevail, then the birds will usually continue into Britain, flying over observatories. Here we see a fundamental problem in the interpretation of observatories data - if one records few or no migrants at observatories, then this is not always indicative of the volume of migration taking place. Also, when large numbers (in terms of observatory counts) are recorded, one does not know the extent to which the counts reflect the volume of migration taking place. For example, a large count of Willow Warblers at Dungeness in spring may be of the order of 1000, but one does not know whether this is a very small proportion of the total number of Willow Warblers passing overhead, or whether nearly all Willow Warblers in the vicinity of Dungeness have been forced to land. Another problem is that in spring, birds may stop at observatories for only a short time before continuing their journey into inland Britain, and may therefore leave without being recorded. There is a strong urgency for birds to arrive at breeding grounds as soon as possible in order to establish territories and find mates. In later chapters statistical methods are used to examine whether or not this 'urgency' is stronger in spring than in autumn.

In autumn the picture is again complicated since birds accumulate throughout the day on the coast, and leave at dusk if weather conditions permit. Birds from inland Britain may start their journey at dusk, and if possible, fly over observatories towards France and beyond. A

further complication is that birds requiring large food supplies on arrival at the coast may travel along the coast in search of suitable feeding grounds. This behaviour is generally called 'coasting'.

For birds that are present at dawn at observatories, the day begins with frantic feeding. This activity enables the birds to be observed more easily than later in the morning, when activity declines as birds retire to undergrowth in order to preen.

### 3.2.2 The observatory day

Visitors to observatories and observatory wardens usually make the first exploration of the recording area at dawn when peak counts are generally made. Bad light can make identification difficult, but the warden and other experts present are able confidently to identify birds by the smallest of calls or briefest of glimpses. The observers do not always stay together, and thus some birds may be recorded more than once. The observers make detailed lists of birds seen on all trips around the recording area, but no regular pattern of censusing is adhered to.

Clearly, the data from observers constitute a complex and irregular mass of species abundance estimates and a 'data-smoothing' conference is held each evening when the warden studies the records and finally enters a figure for each species in the migration log. The daily ringing totals are also taken into consideration.

The day-to-day continuity of the census depends heavily on the warden. His knowledge of the area, of birds, and of migration is intended to overcome the haphazard conduct of the census and variability among the observers who range from beginners to highly skilled ornithologists and ringers.

One hopes that wardens remain at observatories for as long as possible since new wardens will take time to become familiar with the mechanics of migration at different localities. In practice one finds that wardens move around from observatory to observatory every few years, but on the other hand each observatory generally has a group of regular, local ornithologists who would assist in smoothing the transition between wardens.

After the migration log has been completed each evening, written descriptions of the day's events, including weather, are made. Summary sheets are sent monthly to the BTO headquarters.

The daily procedure described above takes place throughout the migration seasons and, in some cases, throughout the entire year.

### 3.2.3 Examples of daily counts

Much can be learned about daily counts of grounded migrants by examining plots of the data. A useful type of plot is a bar-chart where time features on the horizontal axis and

the vertical axis represents the size of the count. The time axis is discrete since daily counts are considered. By placing several such diagrams in close proximity one can also examine variations in pattern over a number of years. In order to present all diagrams for a particular species on the same scale, the occasional very high counts need to be truncated. These are indicated on the diagrams.

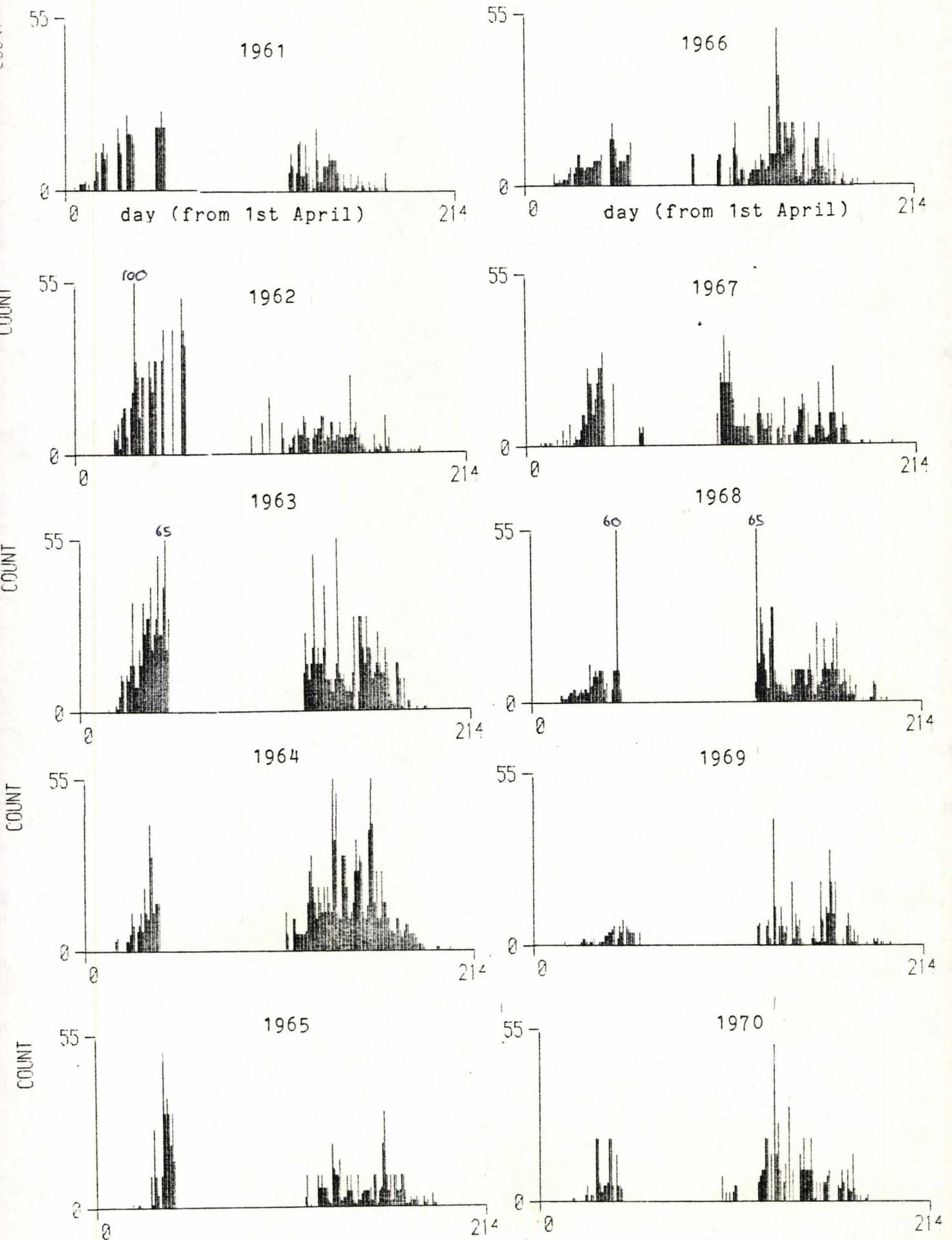
The following examples are a sample of the available data, selected on the basis of certain interesting features which are described below.

(i) Sedge Warbler - Dungeness, 1961-70 (Fig. 3.2.3.1)

The spring data generally show a build up to a peak, followed by an abrupt end to the passage. This pattern is seen in varying lengths of the season. 1961 shows a long season (approximately 7 weeks) while 1965 shows a short season (approximately 3 weeks). One can see conspicuous gaps in the spring data of 1961, 1966 and 1968 - probably due to weather conditions enabling migrants to proceed without landing at observatories. These gaps could, of course, be due to the absence of birds migrating and without further information from, e.g. radar studies, one cannot be more specific. In a later chapter we pick up this point and attempt to relate counts of grounded migrants with local weather conditions.

There is some evidence to suggest that the number of Sedge Warblers migrating through Dungeness each year is falling,

Fig. 3.2.3.1 Daily counts of Sedge Warblers, Dungeness



particularly as far as the spring data are concerned. Since this feature is not so pronounced in the autumn counts, it may be that the migratory route is changing, rather than the population level. An examination of data from other observatories has suggested that this is the case.

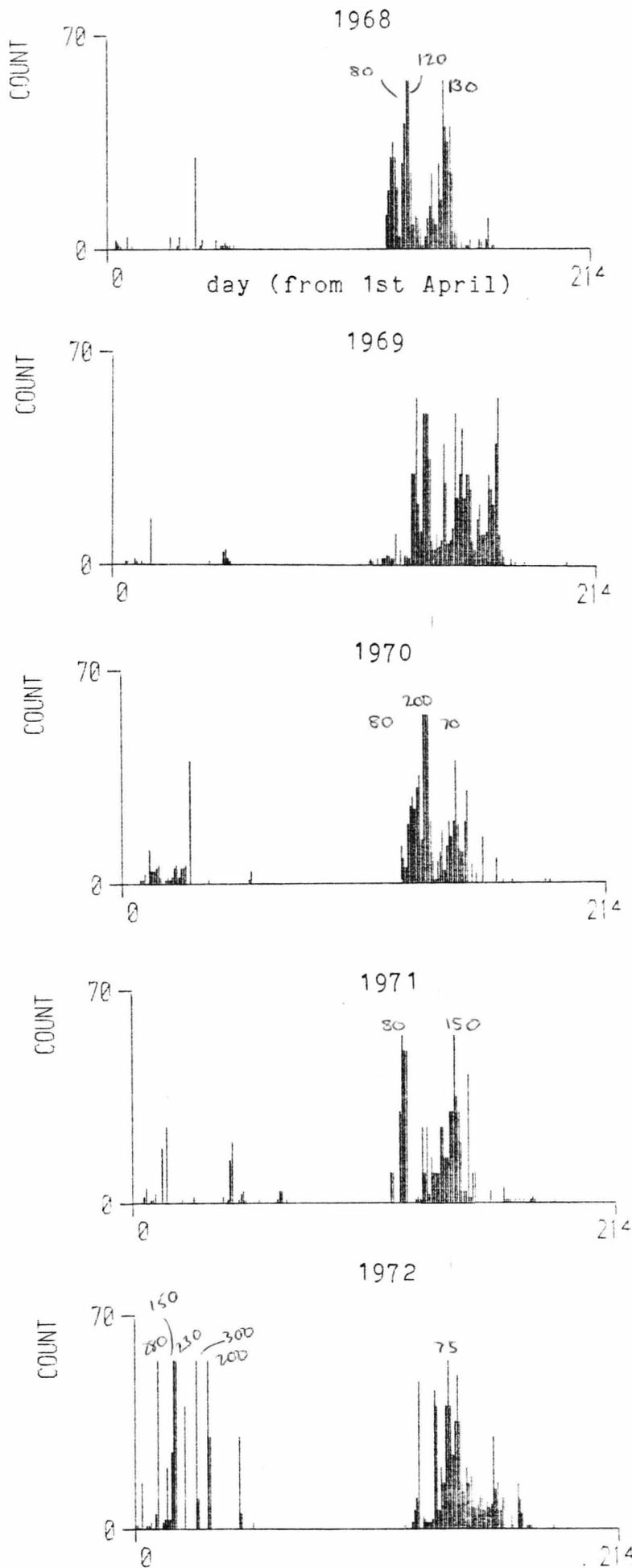
There is the occasional extremely high count in both spring and autumn series, for example, see 1968. This is a feature seen on many other plots and is discussed further in later chapters.

(ii) Willow Warbler - Portland, 1968-72 (Fig. 3.2.3.2)

The spring Willow Warbler counts from Portland appear to be increasing in size over this five year period. A noticeable feature is that the spring counts are only a small proportion of the annual total. This may be explained by the inclusion of young birds in the autumn counts, and also by the effects of winter mortality. Alternatively, this feature may be due to Willow Warblers using a different migratory route in the autumn. The autumn counts suggest the existence of waves of migration - see the 1969 diagram for a clear example. This feature is seen in data for other species and is studied in more detail in section 3.2.5.

A noticeable feature of the 1972 spring counts is the apparent truncation of counts at a value of 50. This is due to rounding on the part of observers. As one would expect with data of this type, the larger the observation, the more coarsely it is likely to be rounded. This feature is seen on other diagrams - see below.

Fig. 3.2.3.2 Daily counts of Willow Warblers, Portland



(iii) Sedge Warbler - Portland, 1965-72 (Fig. 3.2.3.3)

Fewer Sedge Warblers are recorded at Portland than at Dungeness but from the counts alone one cannot say whether this is due to fewer birds migrating through Portland or whether it is due to fewer birds being forced to land.

In some years, the bulk of the spring records are from a single day. This means that the majority of the migration seasons consist of very low counts - mostly fewer than ten birds per day.

(iv) Sedge Warbler - Spurn Point, 1961-67 (Fig. 3.2.3.4)

The pattern of migration for the Sedge Warbler at Spurn is quite different from other observatories in that the two seasons are barely discernible. Ornithologists believe that this is due to local breeding birds in the area being recorded as well as migrant birds. Also, ornithologists believe that continental migrants may be included in the counts as they travel to and from Scandinavia. This has the effect of lengthening the migration seasons.

(v) Whitethroat - Dungeness, 1961-70 (Fig. 3.2.3.5)

These diagrams show that Whitethroats occur in large numbers in both seasons at Dungeness, compared with the Sedge Warbler, for example. The most striking feature however is the sudden drop in numbers after 1968. This population decrease is well known among ornithologists and is discussed in Winstanley, Spencer and Williamson (1974). They found that the cause of the decline appeared to be connected with

Fig. 3.2.3.3 Daily counts of Sedge Warblers, Portland

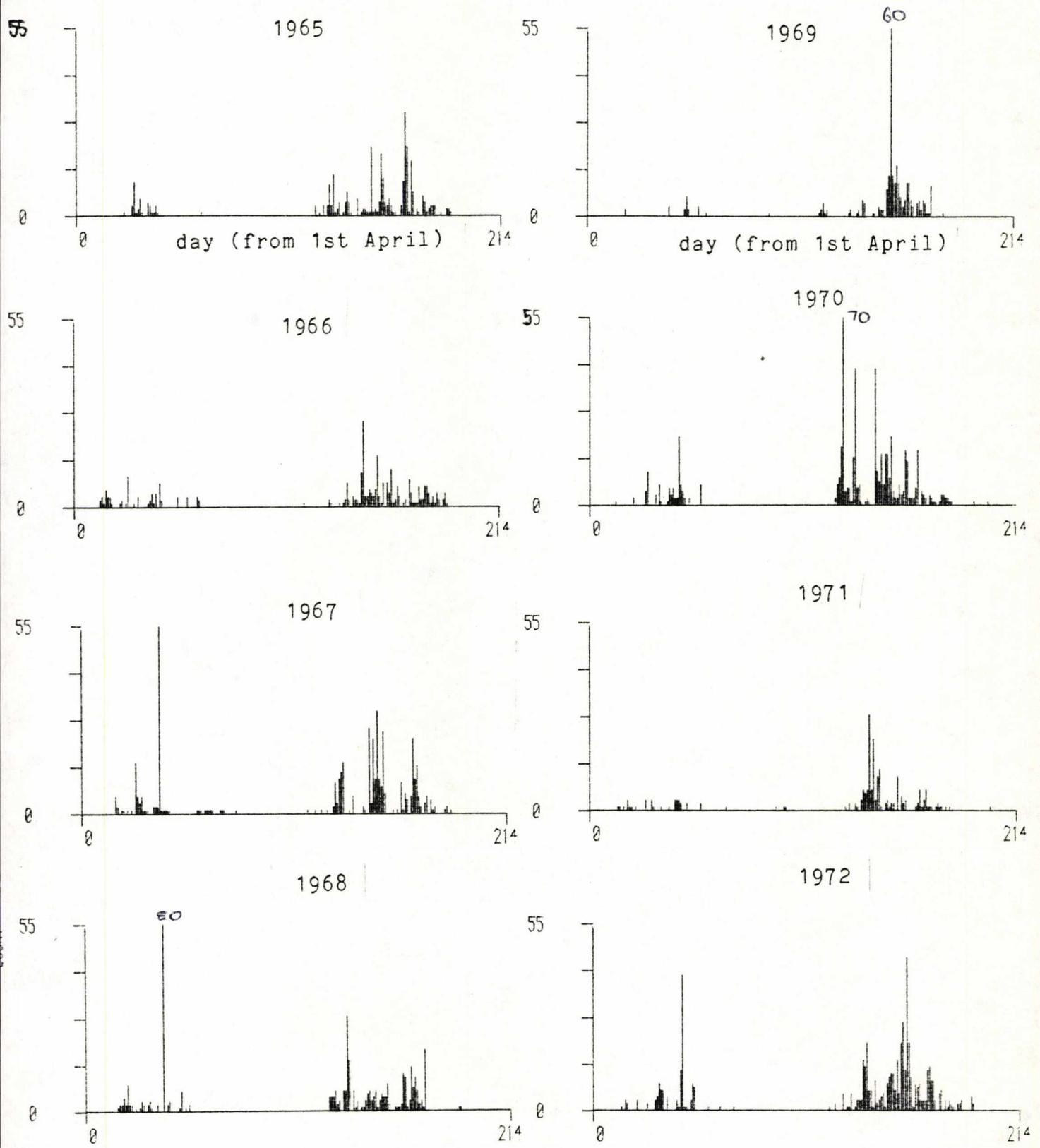


Fig. 3.2.3.4 Daily counts of Sedge Warblers, Spurn

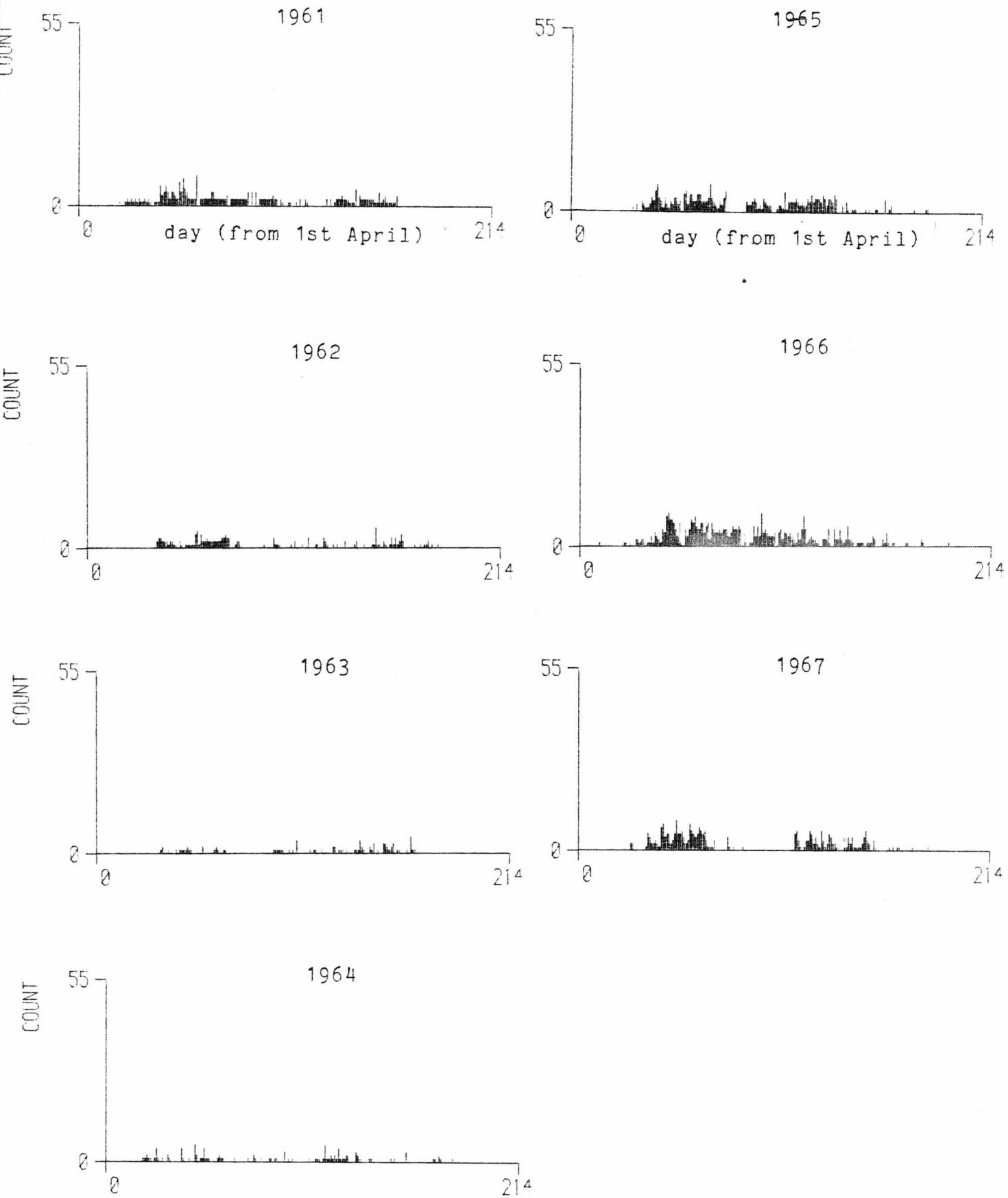
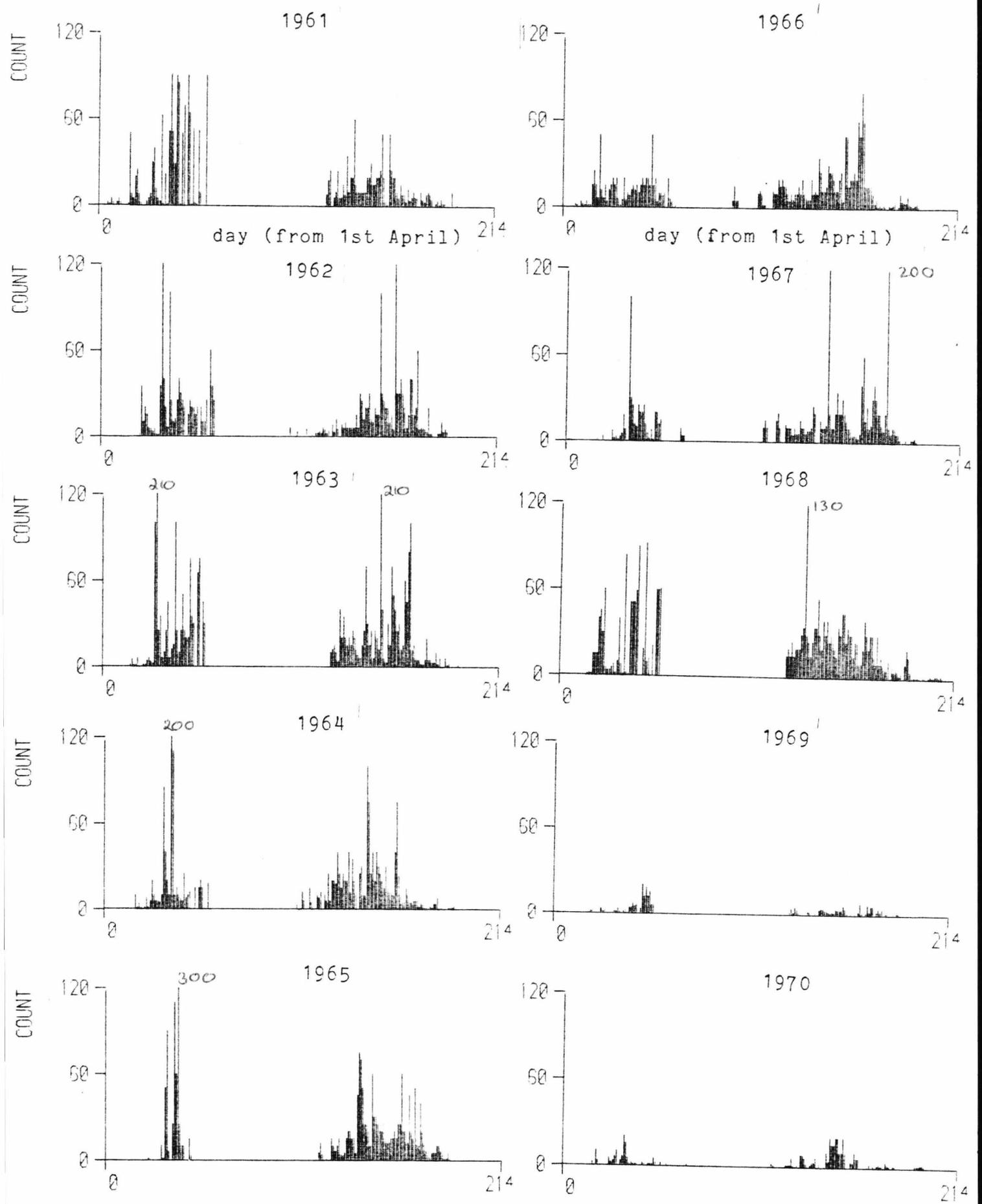


Fig 3.2.3.5 Daily counts of Whitethroats, Dungeness



climatic conditions in the birds' wintering areas. This introduces another fundamental problem of the interpretation of observatories data - the numbers of migrants that are recorded at observatories may be dependent, to an unknown extent, on circumstances elsewhere, for example, the wintering grounds. However, this does suggest that, if only in the crudest fashion, observatories data may be able to be used to monitor population levels.

(vi) Sedge Warbler - Sandwich Bay, 1961-65 (Fig. 3.2.3.6)

A variety of patterns is seen in these five diagrams which range from a small, steady trickle of birds in 1962, to much larger numbers, rather like Dungeness counts, in 1963.

The variation in counts from year to year is thought to be due to continental migrants being drifted over the south-east corner of Kent from their more usual route over Scandinavia and the low countries. This drift is caused by strong prevailing easterlies. It is interesting to note that ornithologists regard Sandwich Bay observatory as an 'east coast' observatory rather than a 'south coast' observatory, although it is only approximately 30 miles east of Dungeness.

(vii) Willow Warbler - Sandwich Bay, 1961-65 (Fig. 3.2.3.7)

The two striking features of these diagrams are that spring migration is negligible compared with autumn migration, and that the numbers of Willow Warblers recorded each year is gradually increasing. It is likely that these are

Fig. 3.2.3.6 Daily counts of Sedge Warblers, Sandwich Bay

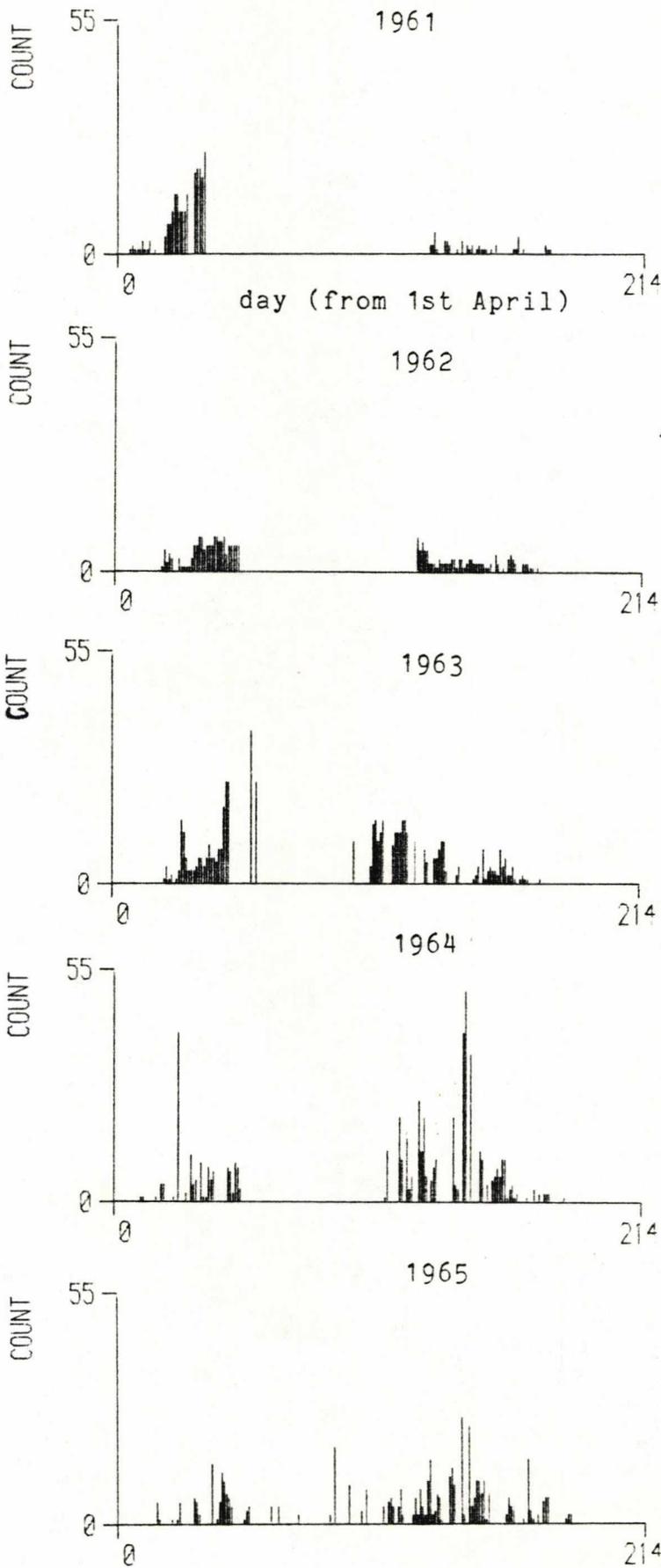


Fig. 3.2.3.7 Daily counts of Willow Warblers, Sandwich Bay

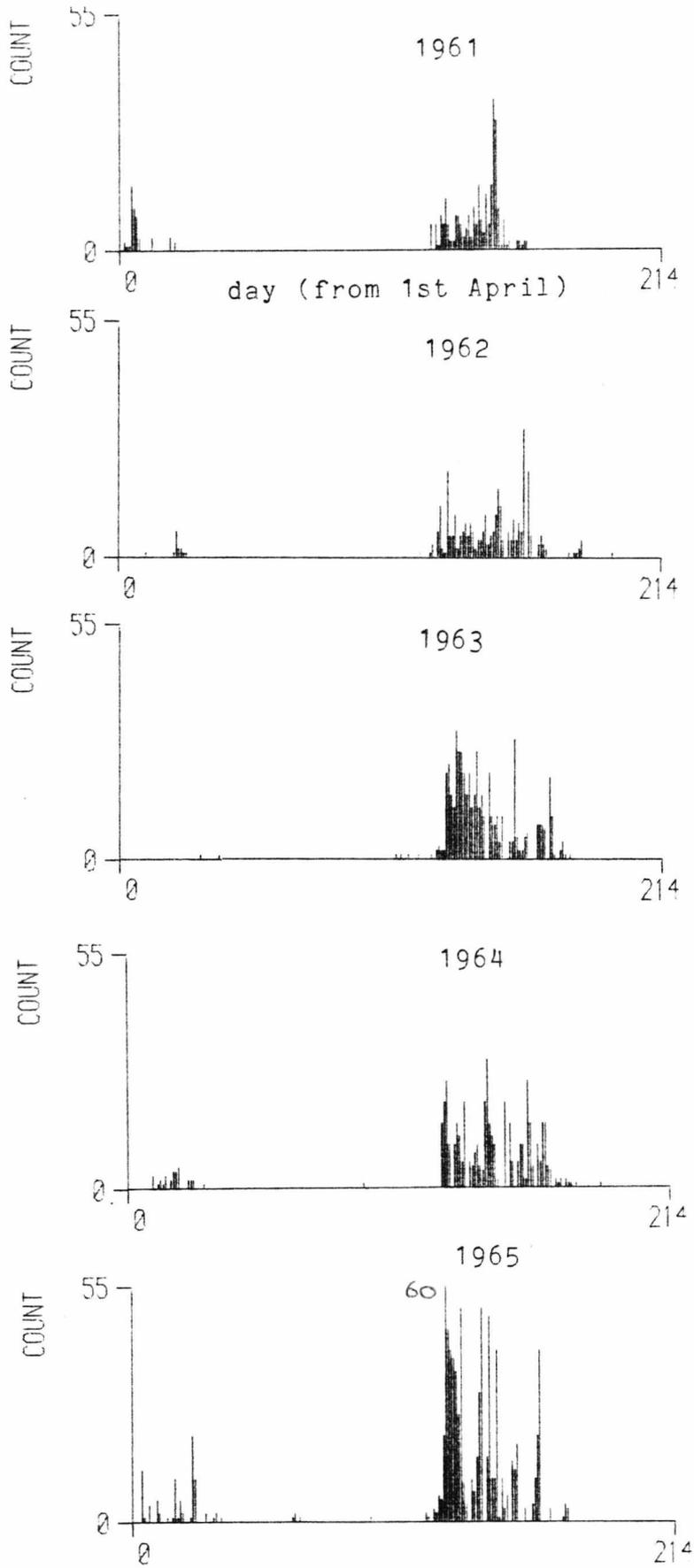
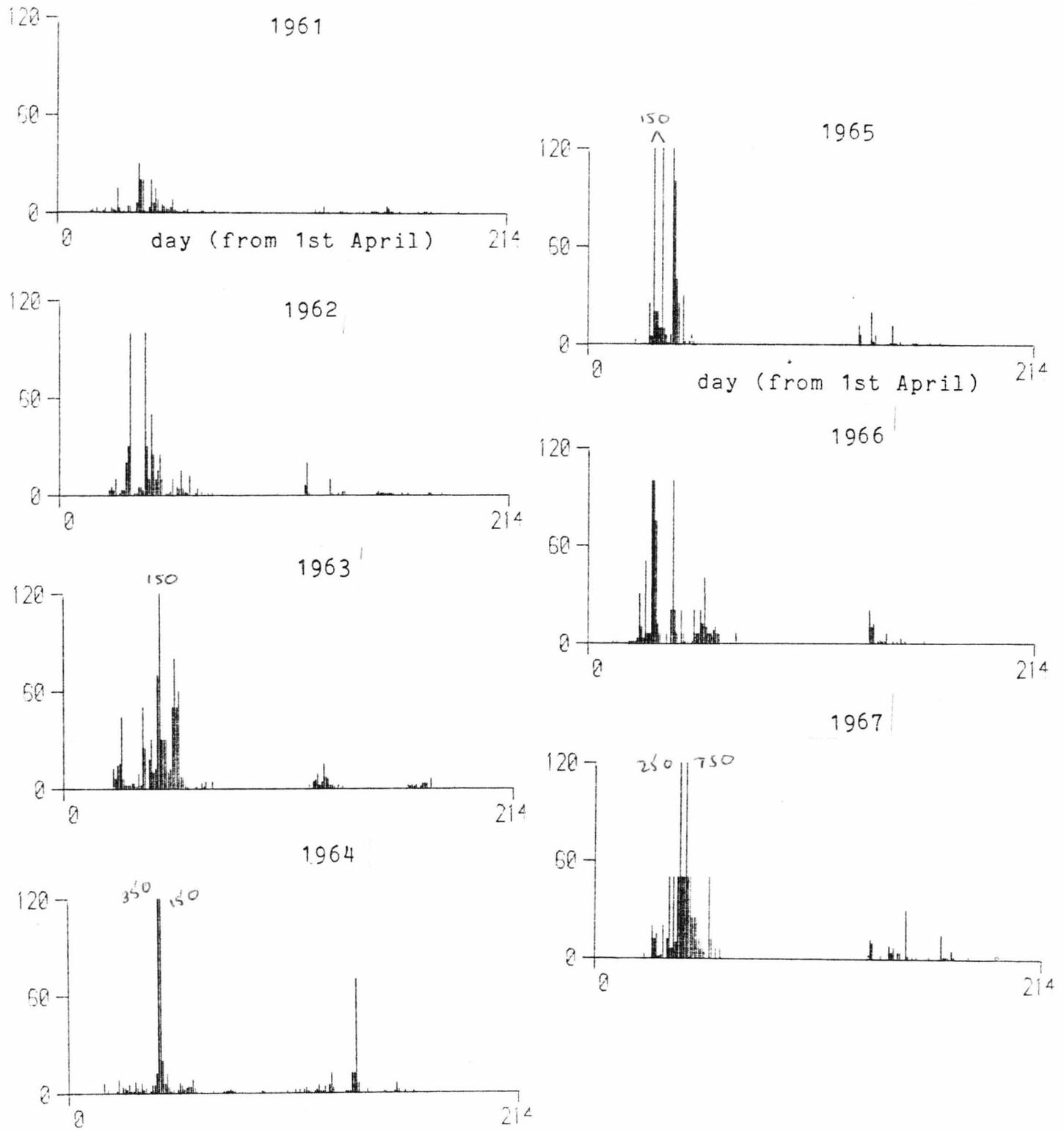


Fig 3.2.3.8 Daily counts of Sedge Warblers, Bardsey



continental birds returning south and being drifted over to Sandwich Bay (see (vi) above).

(viii) Sedge Warbler - Bardsey Is. 1961-67 (Fig. 3.2.3.8)

For this data set, autumn migration is negligible compared with spring migration. The spring migration season is a short one with most birds being recorded on two or three days.

3.2.4 Timing of migration seasons

In this section the timing of migration is examined in order to see whether it is related to population levels of breeding birds.

One method of examining the timing of migration seasons is to plot the migration data for each year on separate diagrams and superimpose on each plot a curve which represents the 'average' season. A suitable curve to use is that used by Hussell and Risley (1978) who examined data on Canadian migrants. The curve is constructed by calculating the mean count for each day using data from a number of years. The means are then smoothed by calculating a moving average, using a suitable number of days. For the examples of this section the number of years used is the maximum possible given the available data. The number of days to use for calculating the smoothed curve depends on the degree of smoothing required. Examples of using 3, 7 and 11 days are presented.

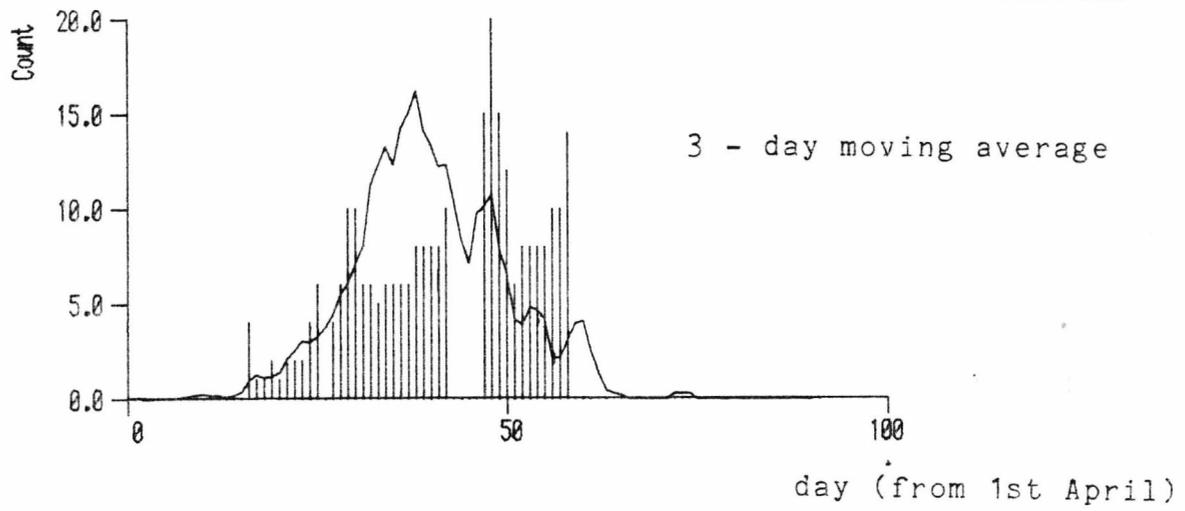
The smoothed series of daily means, hereafter referred to as a long-term moving average, forms a standard against which each individual year can be compared. It is possible to compare each year with the standard as regards both timing and volume.

The examples presented in this section are counts of Sedge Warblers at Dungeness and Portland. This species was chosen since in most years, large numbers of Sedge Warblers are recorded. The first example shows the Dungeness counts for the spring of 1966 (Fig. 3.2.4.1). The long-term moving average is based on 12 years. The three diagrams show progressively smoother curves using 3, 7 and 11 days for the moving average.

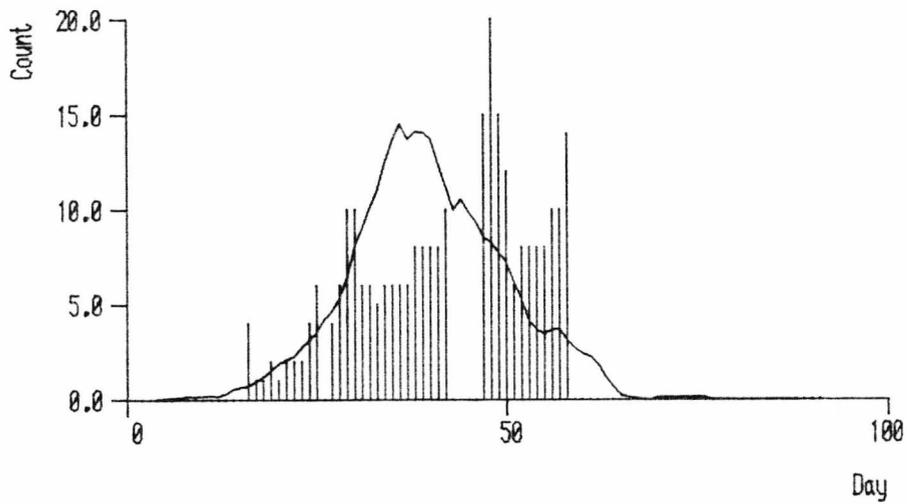
The curve using 3 days reflects features specific to 1966, i.e. the absence of birds between days 43 and 46, and the high counts of the following 4 days. The second and third diagrams show that such features are smoothed out by using more days in the moving average. The last curve represents the average pattern of migration over this 12 year period, which is a single smooth peak.

Statements that may be made concerning the migration of the spring of 1966 are (i) the central part of the season consisted of fewer birds than usual, (ii) an unusual absence of birds occurred during the last week of May, (iii) the migration finished more abruptly, and earlier than usual, and (iv) two periods of 3 or 4 days each,

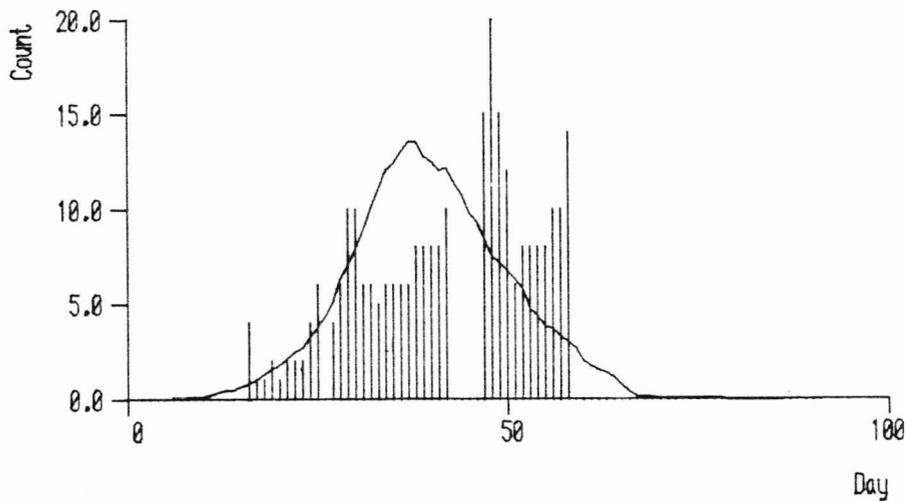
Fig. 3.2.4.1 Dungeness Sedge Warbler counts and long-term (12 year) moving average curve (1966).



7 - day moving average



11 - day moving average



towards the end of the season, saw higher than average numbers of birds recorded.

The diagram for the spring Sedge Warbler data at Dungeness, 1963 (Fig. 3.2.4.2), shows higher than average counts for nearly all of the season, but the migration appears to cease rather abruptly on the 18th of May.

The diagram showing the 1965 spring counts (Fig. 3.2.4.3) reveals that the migration for this particular year was much more condensed than usual. High numbers of birds were recorded during a two-week period which coincides with the usual peak period. The migration appeared to start later and finish earlier than usual.

We have considered a very small subset of the available diagrams, but even so, many interesting features have come to light. After using this simple technique, ornithologists would then be able to investigate the reasons behind each departure from the long-term moving average, and possibly interpret the data in terms of weather and population dynamics, for example, or at least they could postulate plausible explanations for what is observed.

In order to investigate whether or not population levels are related to the timing of migration seasons, we need a summary statistic that quantifies 'timing', i.e. a single date. The simplest statistics that we may use include the mid-season day (half-way between the first and last count),

Fig. 3.2.4.2 Dungeness Sedge Warbler counts and long-term (12 year) moving average curve (1963).

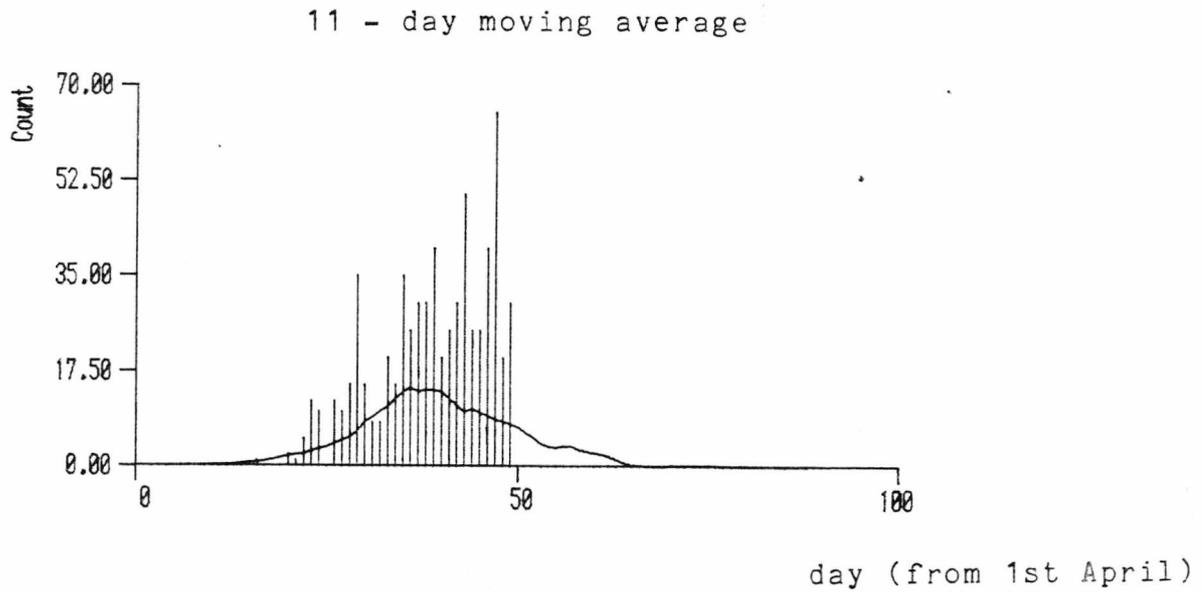
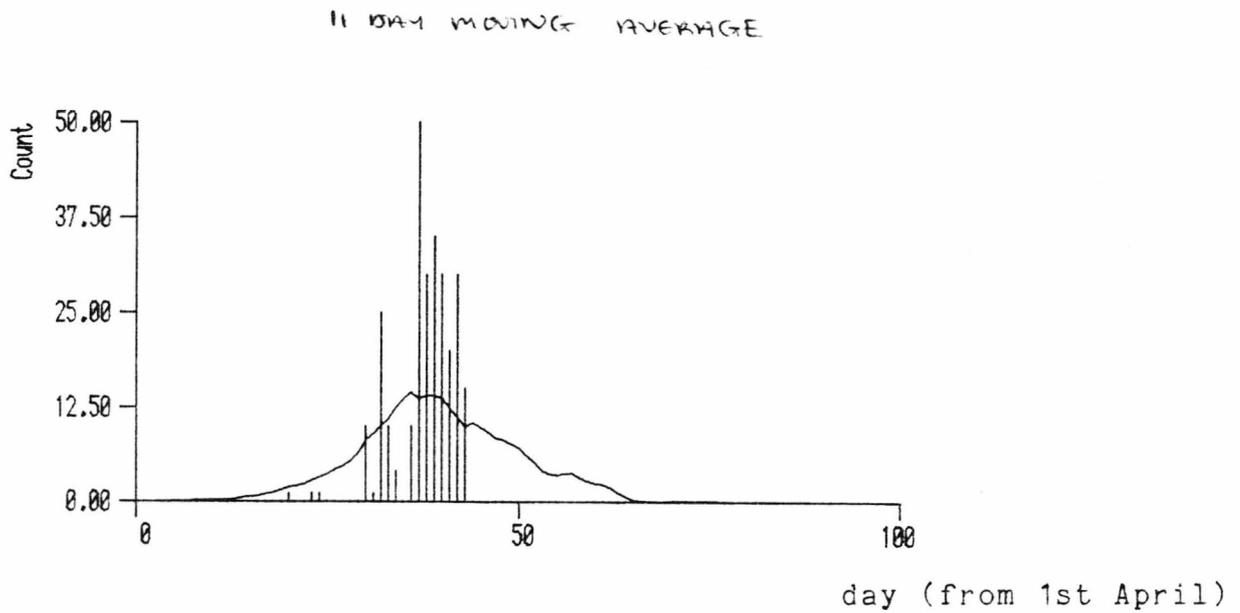


Fig. 3.2.4.3 Dungeness Sedge Warbler counts and long-term (12 year) moving average curve (1965).



the 'mean' day (calculated by treating each observation as a frequency pertaining to the day in question, numbered from some origin), and the median day (the day before which and after which half the birds are recorded).

The mean day has been used by other authors in the past, although only for species comparisons over a number of years, i.e. a long term mean day, see Lack and Lack (1966), Davis (1967) and Sharrock (1968). The Lack's paper studied migrants passing through Skokholm observatory (no longer operating), using daily ringing totals, and the papers by Davis and Sharrock studied daily counts of Sylvia Warblers at several British observatories. No investigation of trends was undertaken however.

Tables 3.2.4.1, 2 and 3 show the dates of the first and last counts of four common migrants at Dungeness, and the Sedge Warbler at Portland. (Sedge Warbler data only are shown for Portland owing to the limited number of years of data available). Also shown are the median dates and mid-season dates. The latter are midway between the first and last dates. The mean dates are very close to the median dates and the latter are used hereafter since they are easier to calculate.

The first and last record dates vary from year to year more than the median or mid-season dates, which is to be expected since they will be more sensitive to random variation which is absorbed, to a certain extent, in the

Table 3.2.4.1 Spring migration dates : 1961-72 : Dungeness

Dates are numbered from 1st April.

Year	Willow Warbler				Sedge Warbler			
	first record	mid-season date	last record	median date	first record	mid-season date	last record	median date
1961	31	45	59	35	5	30	55	36
62	8	33	59	33	22	42	61	42
63	10	42	75	29	16	32	49	40
64	5	35	64	34	17	29	41	36
65	4	32	61	37	20	31	43	38
66	1	31	61	23	16	37	58	42
67	2	30	58	34	8	36	64	38
68	2	29	56	32	16	32	49	39
69	10	35	60	35	17	37	58	47
70	5	31	57	31	18	32	45	37
71	7	40	74	38	22	40	58	43
72	2	37	73	18	15	44	73	39
mean date	7	35	63	32	16	35	55	40
s.d.	8	5	7	6	5	5	9	3
Year	Whitethroat				Redstart			
	first record	mid-season date	last record	median date	first record	mid-season date	last record	median date
1961	5	32	58	42	5	27	50	29
62	21	41	61	37	12	28	43	26
63	15	35	55	39	10	37	65	29
64	17	37	56	36	9	36	64	31
65	23	35	46	39	4	22	40	32
66	7	35	62	38	5	34	62	21
67	17	40	64	37	4	28	52	35
68	15	35	54	40	12	30	47	23
69	20	37	54	49	25	40	54	43
70	16	42	67	33	16	43	71	18
71	15	45	75	49	13	34	55	25
72	30	47	64	59	14	40	66	37
mean date	17	38	60	42	11	33	56	29
s.d.	7	5	8	7	6	6	10	7

Table 3.2.4.2 Autumn migration dates : 1961-72 : Dungeness

Dates are numbered from 1st July.

Willow Warbler					Sedge Warbler			
Year	first record	mid-season date	last record	median date	first record	mid-season date	last record	median date
1961	32	66	100	51	32	59	86	49
62	27	64	101	57	5	52	98	48
63	33	63	93	51	32	66	99	57
64	20	56	92	43	20	65	110	53
65	27	61	94	48	24	60	96	59
66	23	58	93	53	15	58	101	50
67	16	60	104	48	14	62	110	41
68	17	69	121	48	32	68	104	55
69	32	73	114	53	32	69	105	65
70	16	60	103	30	9	49	89	46
71	7	40	74	38	22	40	58	43
72	24	59	94	47	21	50	79	53
mean date	24	63	102	48	22	60	98	53
s.d.	6	5	10	7	10	7	10	6

Whitethroat				Redstart				
Year	first record	mid-season date	last record	median date	first record	mid-season date	last record	median date
1961	32	68	103	58	33	66	99	69
62	11	54	96	68	46	77	107	86
63	32	67	101	63	29	64	98	75
64	13	61	108	52	29	67	104	68
65	24	66	108	56	14	75	108	71
66	15	59	102	60	48	81	114	80
67	16	58	99	63	34	70	105	75
68	32	75	117	57	23	72	47	77
69	37	74	111	66	32	73	113	75
70	22	64	105	57	37	70	102	78
71	17	65	112	55	39	67	95	67
72	10	53	96	56	45	81	116	80
mean date	22	64	105	59	36	72	107	75
s.d.	10	7	7	5	8	6	8	6

Table 3.2.4.3 Spring and Autumn migration dates : 1961-72

Sedge Warbler, Portland .

Spring dates are numbered from 1st April

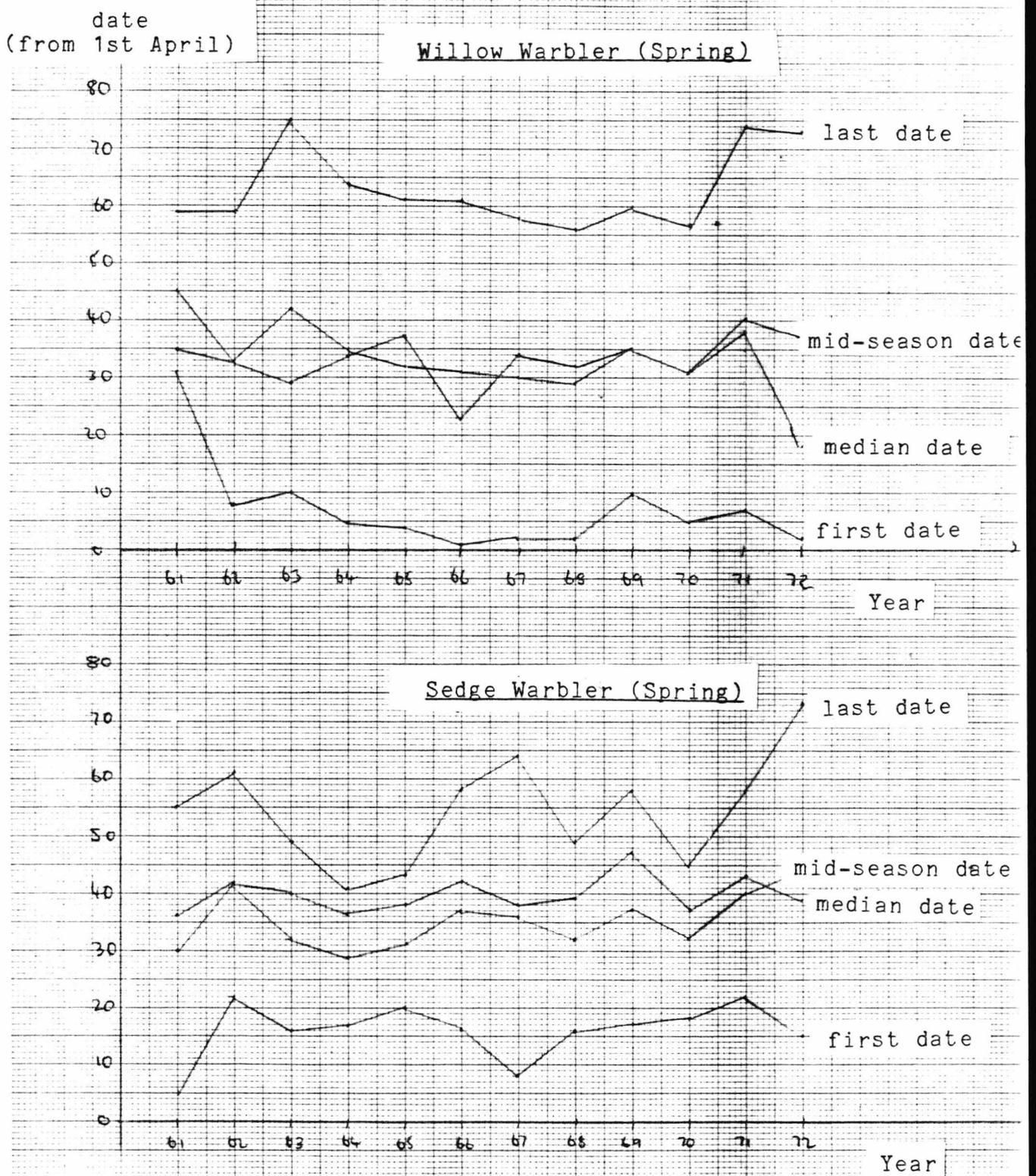
Autumn dates are numbered from 1st July.

Year	Spring				Autumn			
	first record	mid- season date	last record	median date	first record	mid- season date	last record	median date
1961	33	47	61	46	38	72	105	65
62	17	41	64	45	39	62	84	61
63	19	36	53	29	36	63	90	74
64	17	36	54	32	33	62	90	60
65	24	43	62	32	29	64	98	65
66	13	38	63	32	24	61	97	64
67	17	48	78	39	22	59	95	59
68	16	34	52	39	32	65	98	55
69	19	40	60	50	25	57	88	62
70	13	35	56	29	33	71	109	53
71	14	41	68	36	38	74	110	51
72	14	42	70	45	27	64	100	65
mean date	18	40	62	38	31	65	97	61
s.d.	6	5	8	7	6	5	8	6

median and mid-season dates. Also, the extreme dates will be more sensitive to weather conditions.

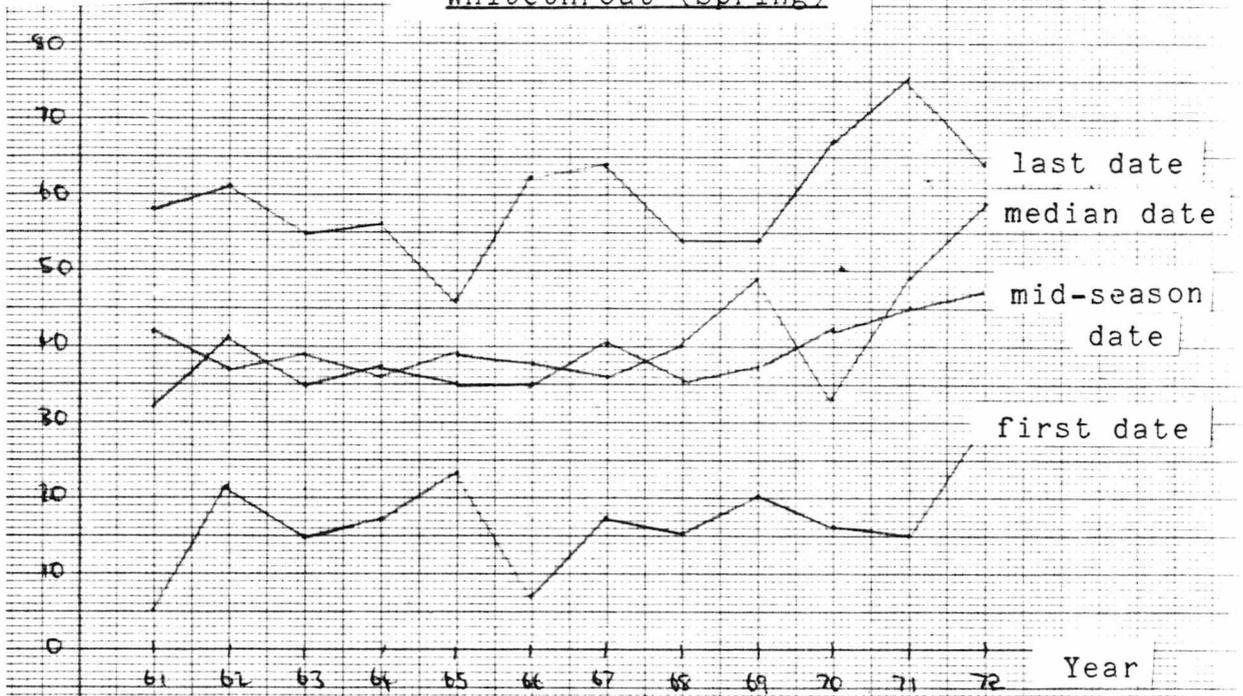
The dates are plotted in Fig. 3.2.4.3 (Dungeness) and Fig. 3.2.4.4 (Portland). In all diagrams, one can clearly see that the first and last record dates vary more than the median or mid-season dates. There appear to be no noticeable long term changes in timing for these species, apart from the Whitethroat which shows a suggestion of spring migration becoming later towards the end of the 12 year period. The variation in timing from year to year also appears to increase after 1968, and this is undoubtedly due to the sudden drop in population level experienced by this species over the 1968-69 winter.

Fig. 3.2.4.3 Migration dates : Dungeness



date  
(from 1st April)

Whitethroat (Spring)



date  
(from 1st April)

Redstart (Spring)

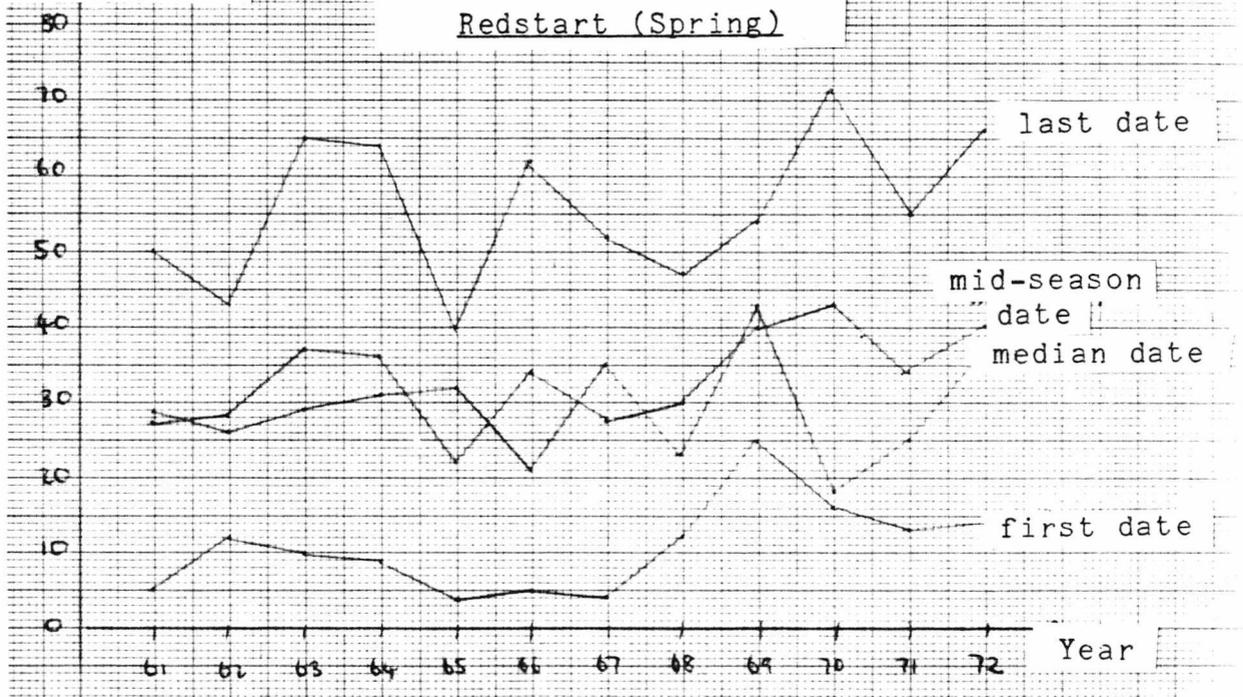


Fig. 3.2.4.3 ( continued )

date  
(from 1st July)

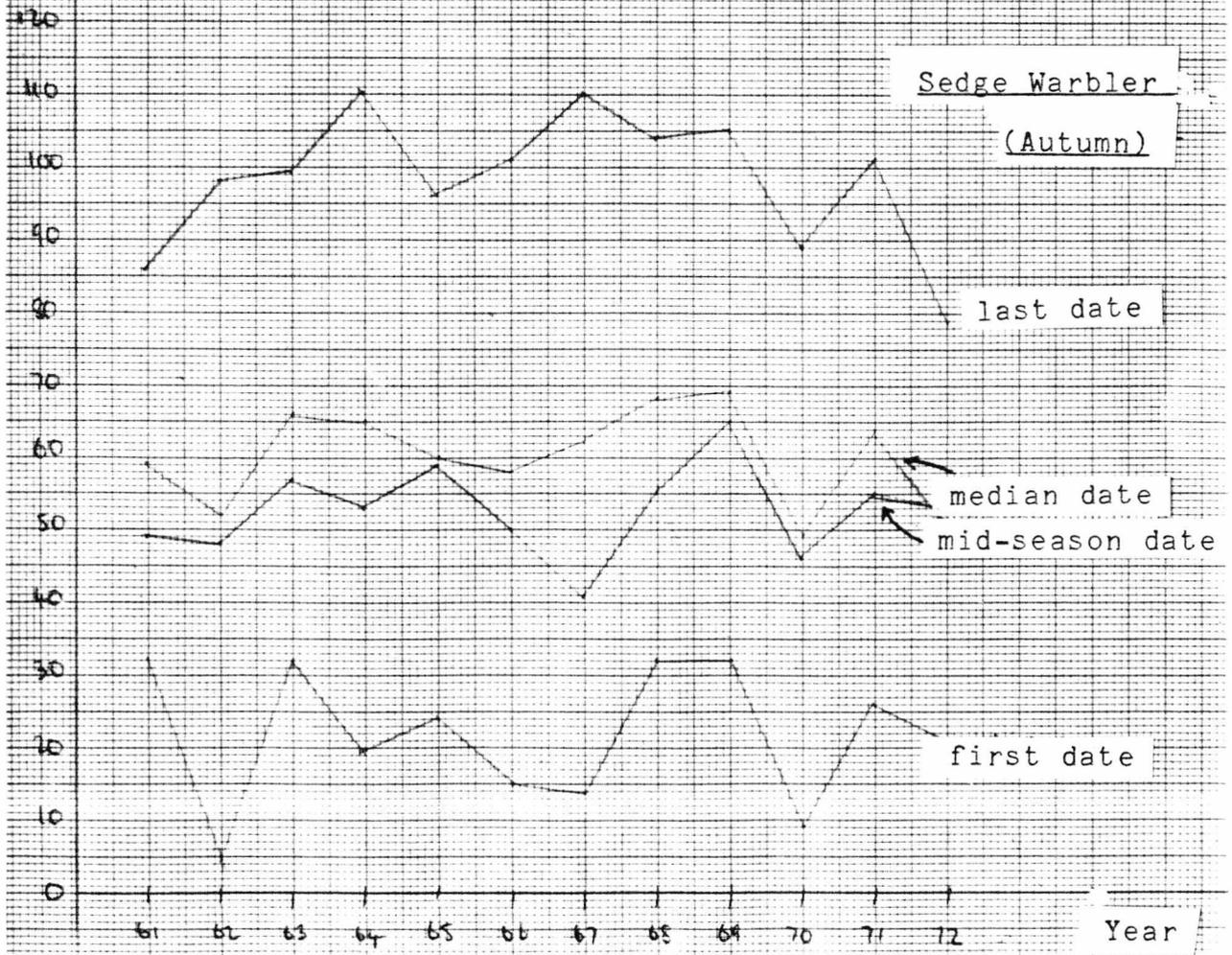
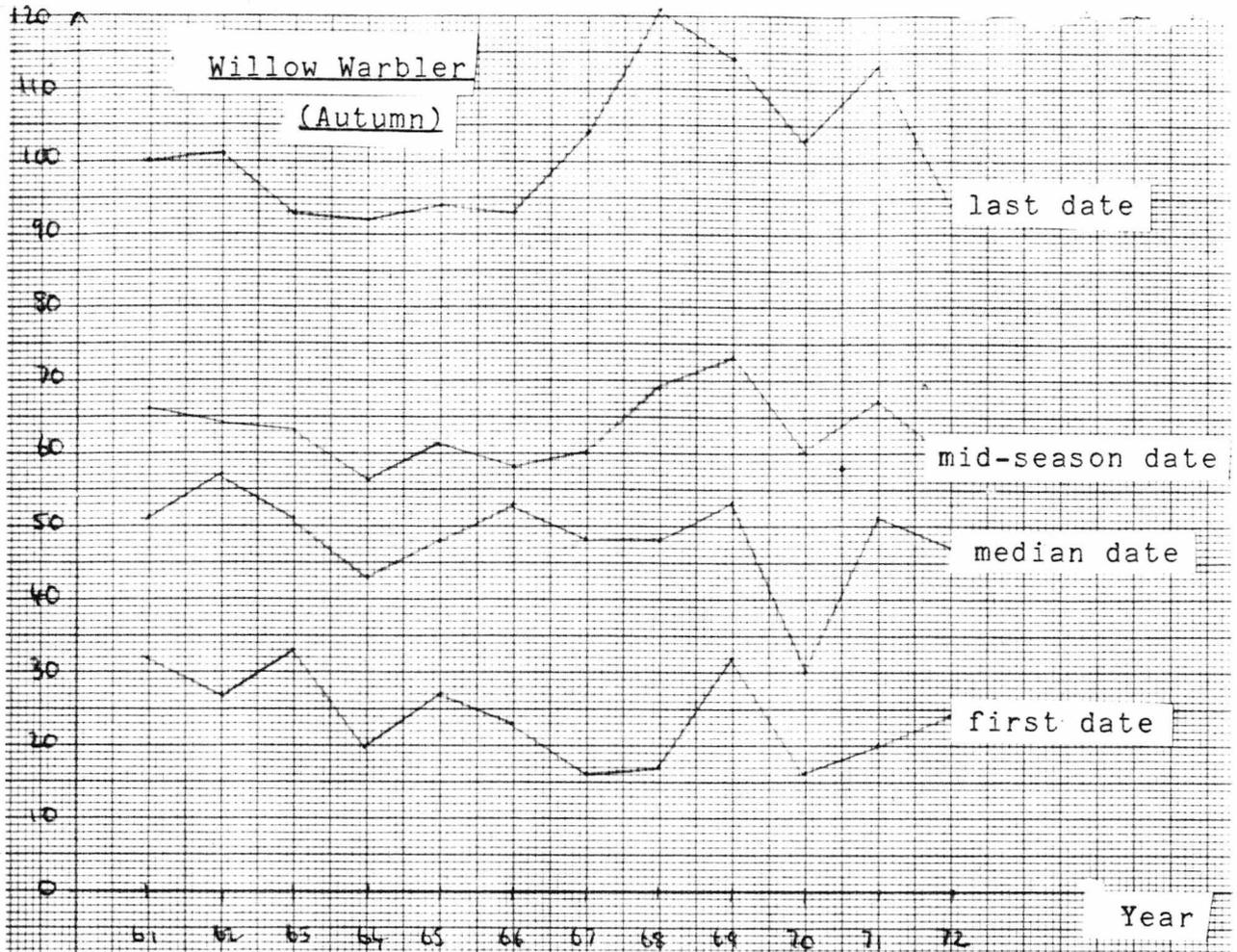


Fig. 3.2.4.3 ( continued )

date  
(from 1st July)

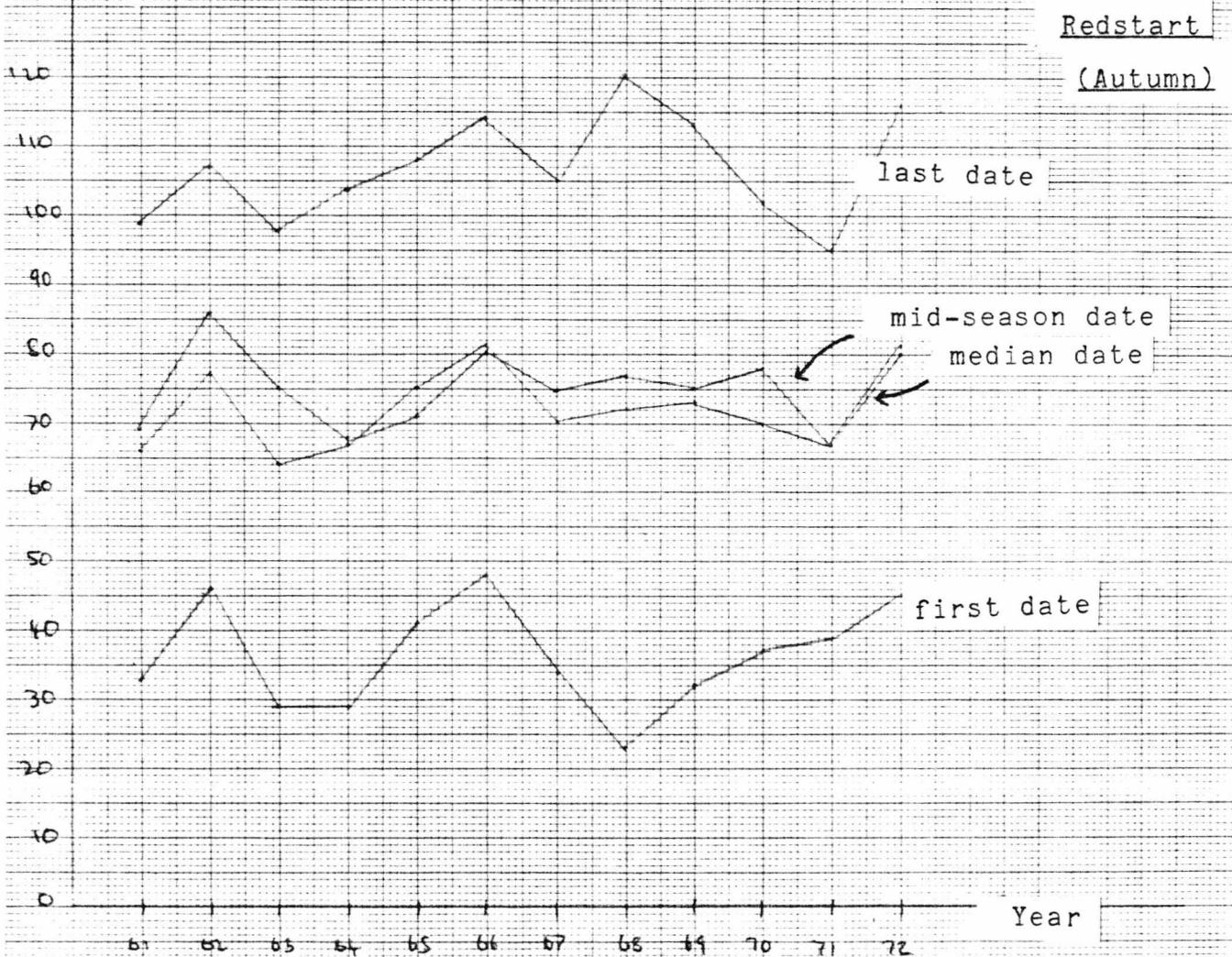
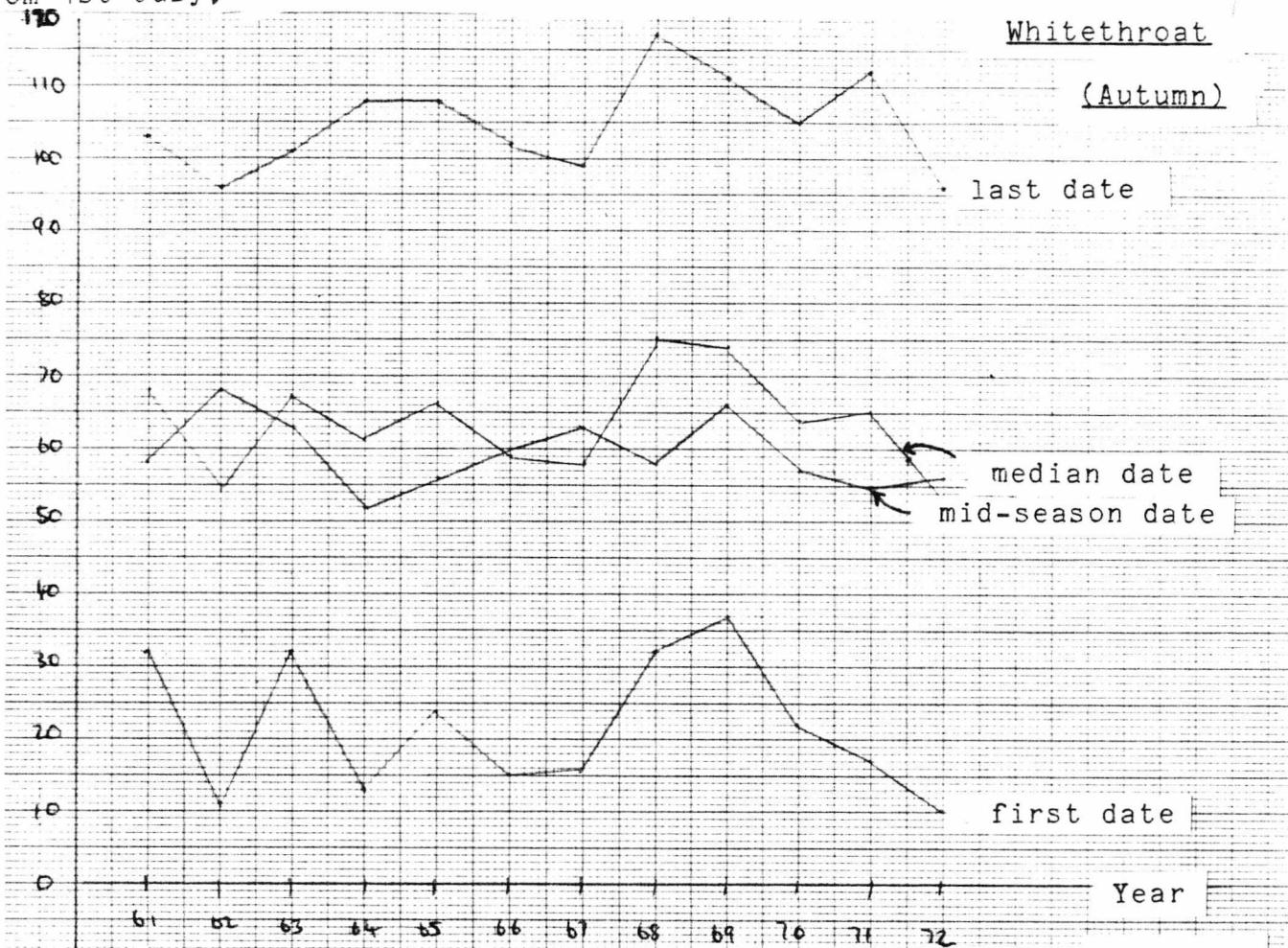
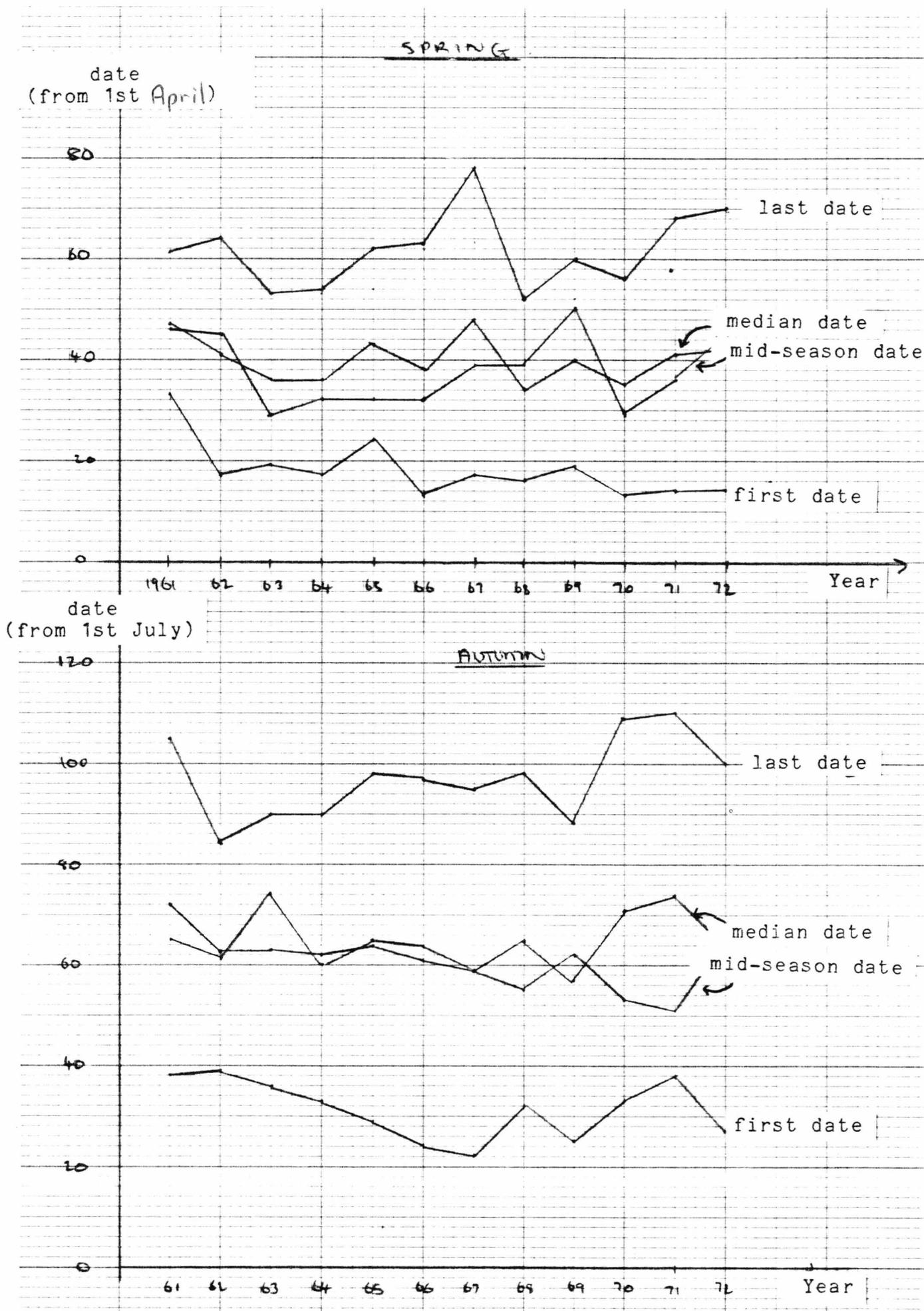


Fig. 3.2.4.4 Migration dates : Portland



A number of questions concerning the timing of migration of the four species considered can be answered by examining correlation coefficients between different dates within species, between certain dates across species, between spring dates and autumn dates and, for the Sedge Warbler, between Dungeness dates and Portland dates. A summary of the results is set out below.

- 1) The Dungeness data show no similarities between species of the timing of the beginning or ending of spring seasons. In other words, if one species is observed particularly early or late then it does not follow that other species will show similar behaviour.
- 2) The spring mid-season dates for Sedge Warblers, Whitethroats and Redstarts are positively correlated with one another (significant at 5% level). The mid-season dates for Willow Warblers are not correlated with other species. The median dates show a similar pattern of correlation, but no significant levels are reached.
- 3) The only significant correlations between species for autumn dates are between Willow Warblers and Whitethroats. For these species, mid-season dates and last dates are positively correlated.
- 4) For the four species considered, there are no significant correlations between spring mid-season dates and autumn dates. Also, there are no significant correlations between the last spring date and first autumn date.

- 5) For the Sedge Warbler, there is no significant correlation between Dungeness and Portland observatories for any of the four dates considered.

Two interesting points arise from the above work. There appears to be no connection between the spring migratory behaviour of different species but in the autumn there is some evidence to suggest that three out of the four species considered migrate at around the same time. This is what we might have expected since in spring, birds will have travelled a much greater distance over a wider area before arriving at observatories but in autumn all the species considered will have travelled from within the comparatively small area of the British Isles, and will all be subject to the weather within this area.

The second point is that any effects of particularly early or *late* *spring* migrations do not influence the timing of autumn migration. Later in this section we shall see how linear functions of spring and autumn dates are related to population levels of these four common migrants.

This work illustrates another aspect of observatories data which ornithologists could explore using data over many more species and longer time periods. Indeed, one may find more interesting correlations if more years are used for the species considered in this section. Alternative dates may also be worth considering. For example, one may use outer-quartiles or outer-deciles of the data which may be less sensitive to weather conditions than extreme dates.

We now consider how we might use summary statistics on the timing of migration for population monitoring. In order to see which dates, if any, are related to the CBC index, we can regress the CBC index on a number of dates for both spring and autumn migration seasons.

This analysis has been carried out for three species (Sedge W., and Willow W. and Whitethroat) at Dungeness, and for the Sedge Warbler at Portland. No CBC indices exist for the Redstart. In each case, the median days were found to be of no use. It was possible, however, to fit regression models using first and last dates of spring and autumn seasons.

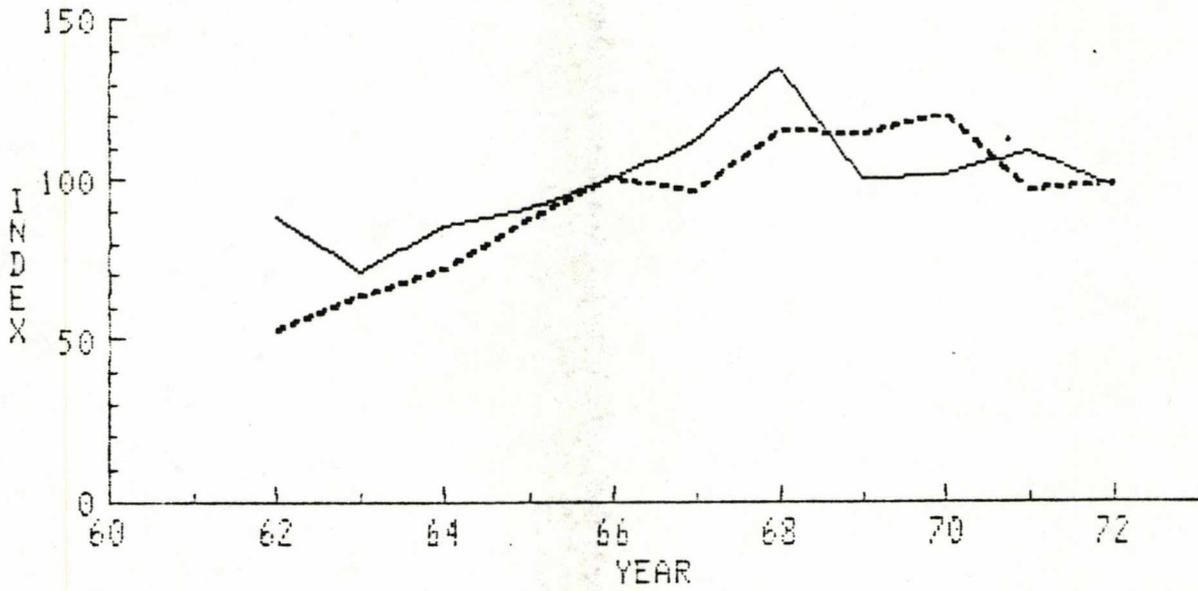
The Willow Warbler data (Dungeness) gave a model of the form:

$$\begin{aligned} \text{CBC} &= 23.20 - 3.02 (\text{first spring date}) \\ &+ 1.28 (\text{last autumn date}) \end{aligned}$$

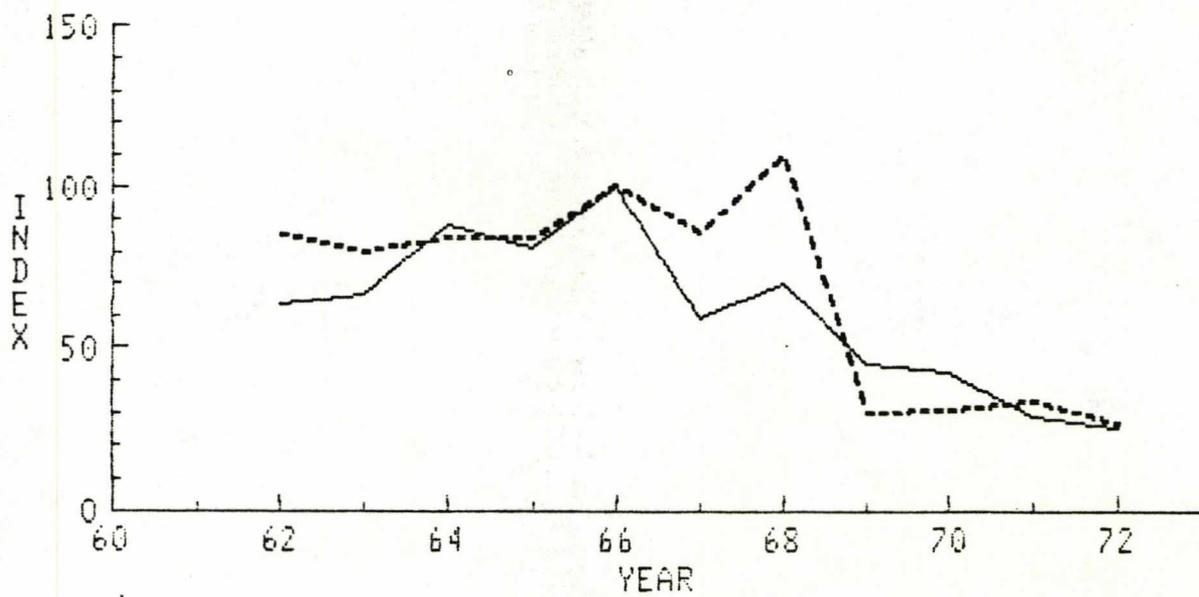
The  $R^2$  value for this regression is 49.8%, giving a correlation coefficient between CBC index and predicted value (migration index) of 0.71, which is significant at the 5% level. Last spring dates or early autumn dates were not found to provide a significant improvement in the regression model (their inclusion gave an  $R^2$  of 51%). This model suggests that for this species, a high population level is indicated by a long time span between entry to and departure from Britain. The migration index is plotted with the CBC index in Fig. 3.2.4.5, and the indices are given in Table 3.2.4.4. Note that the migration indices have been scaled so that the 1966 index is 100.

Fig. 3.2.4.5 Migration indices and CBC indices (dashed line)

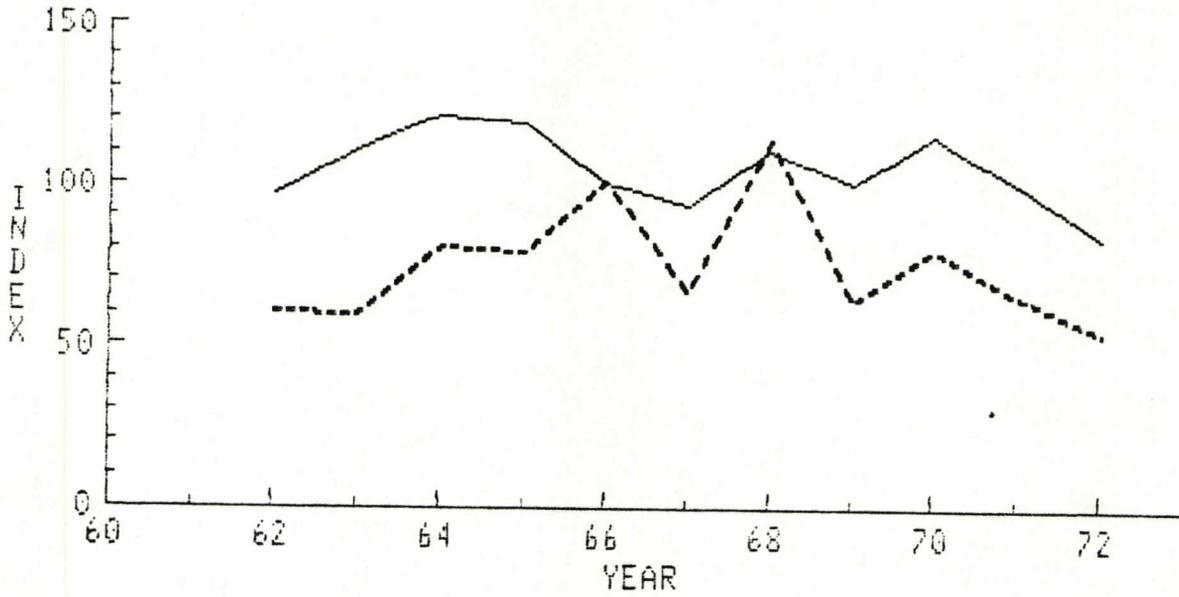
Willow Warbler



Whitethroat



Sedge Warbler ( Dungeness )



Sedge Warbler ( Portland )

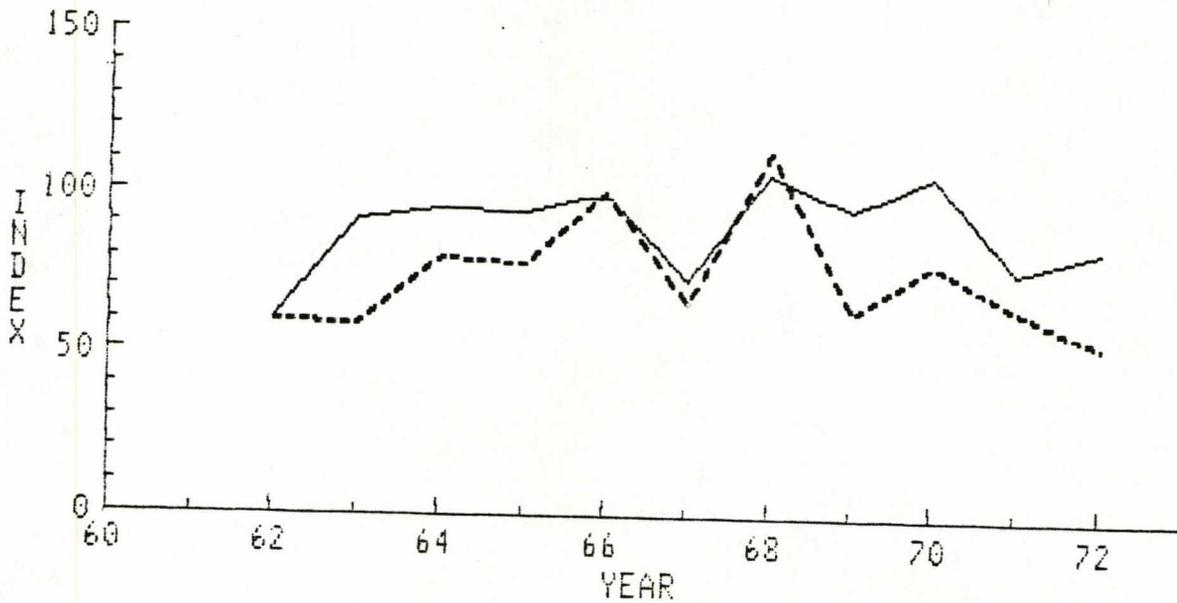


Table 3.2.4.4 Migration indices based on dates

DUNGENESS

Year	Willow Warbler		Sedge Warbler		Whitethroat	
	Migration Index	CBC Index	Migration Index	CBC Index	Migration Index	CBC Index
1962	88	53	97	60	63	85
63	71	63	111	59	67	80
64	86	72	121	80	88	84
65	91	87	118	78	81	84
66	100	100	100	100	100	100
67	112	96	93	66	59	86
68	135	115	111	113	70	110
69	100	114	100	63	45	30
70	101	120	115	79	43	31
71	109	96	100	65	29	33
72	98	99	83	54	25	26

PORTLAND

Year	Sedge Warbler	
	Migration Index	CBC Index
1962	61	60
63	92	59
64	95	80
65	94	78
66	100	100
67	74	66
68	107	113
69	96	63
70	107	79
71	77	65
72	84	54

It is not possible to find a regression model that fits the data as well for the Sedge Warbler as for the Willow Warbler. For the Sedge Warbler, the optimum model is one which contains just one explanatory variable, the last record date for spring:

$$\text{CBC} = 120 - .84 (\text{last spring date})$$

This gives an  $R^2$  of 20.6%, giving a correlation coefficient between CBC and migration indices of 0.45 (not significant at the 5% level). Although not significant, this correlation coefficient does suggest that low population levels are associated with late spring migration. The indices are plotted in Fig. 3.2.4.5 and tabulated in Table 3.2.4.4.

The regression model for the Whitethroat contains three explanatory variables:

$$\begin{aligned} \text{CBC} = & 380 - 3.76 (\text{1st spring date}) - 3.45 (\text{last spring date}) \\ & - 1.83 (\text{1st autumn date}) \end{aligned}$$

This model gives an  $R^2$  of 71.8%, giving a correlation of 0.85 between the CBC and migration index. The model suggests that for this species, low population levels are associated with late spring seasons and a late start to the autumn season. The indices are tabulated in Table 3.2.4.4 and plotted in Fig. 3.2.4.5.

One also needs three explanatory variables for predicting the Sedge Warbler CBC index from Portland data. The best model is:

$$\begin{aligned} \text{CBC} = & 148 - 1.56 (\text{last spring date}) - 1.44 (\text{1st autumn date}) \\ & + 0.70 (\text{last autumn date}) \end{aligned}$$

This model suggests that low Sedge Warbler population levels are associated with late spring migration and a late, short autumn migration season. The model gives an  $R^2$  value of 43.2%, giving

a correlation of the migration index with the CBC index of 0.66 (significant at 2½% level).

In order to see whether a better model for Sedge Warbler CBC indices can be formed by combining Dungeness and Portland data, a further regression analysis was carried out using 1st and last record dates from both sites. An  $R^2$  of 55% is obtained by including all variables, but no suitable model was found by performing a stepwise analysis, other than that found for Portland data alone. Even by including all variables in the regression, the correlation coefficient between migration indices and CBC was not significantly higher than that for Portland data alone.

Although relationships between CBC indices and indices based on dates have been found, the precision of the migration indices is poor. For example, the standard error of a Sedge Warbler (Dungeness) migration index of 75 (before dividing by the 1966 index) is approximately 40. This represents an average index - the standard error will be lower for lower indices and higher for higher indices. Confidence intervals will obviously be very wide.

This unfortunate feature of these migration indices is most likely due to the small sample on which the regression is based. For 3 of the 4 species studied there is strong evidence to suggest a linear relationship between dates and the CBC index, and when more data are available one would be able to calculate migration indices with increased precision.

In the above work, attempts were made to find models which lead to significant correlations between CBC indices and migration indices. Where such a model was found, attempts to simplify the model were made. In terms of automating this procedure for regular population monitoring by ornithologists, it would not be a problem to use the full model in order to achieve as great an accuracy level as possible.

It is very encouraging to find that such a simple approach leads to being able to predict CBC indices. As will be seen later in this thesis, more complex approaches are less successful.

### 3.2.5 Fitting curves to sequences of autumn counts

Many sequences of autumn daily counts show what might be termed waves of migration. In other words, some migration seasons consist of a number of short bursts of activity rather than one continuous flow of movement. This feature is not found with all species, nor at all observatories.

In those cases where waves appear to occur, it is worth thinking how we might investigate them more closely. In this section we suggest simple methods of describing a migration season that appears to consist of a sequence of waves. Ideally one would like to investigate the ornithological background to waves, but this would inevitably entail a detailed study of ringing data, and this is beyond the scope of this thesis.

We shall take a deterministic approach and simply fit a series of curves to each sequence of daily counts. An examination of a large number of daily counts sequences suggests that waves of migrants only occur in the autumn, and therefore we restrict our attention to this season.

Although this feature of autumn counts is of interest in its own right, we shall aim our analysis at the estimation of the numbers of birds involved in the waves in order to investigate whether or not a migration index may be constructed from autumn counts.

Fig. 3.2.5.1 shows examples of waves of Whitethroats passing through Portland in the autumns of (a) 1970 and (b) 1971. Not all diagrams of daily counts show such a clear picture as these, however.

There are several possible explanations for waves of migrants. They may simply be due to weather conditions which release large numbers of birds from some area, which then take a few days to pass through an observatory. Ornithologists have suggested that this is probably the case for east coast waves - see Fig. 3.2.5.1(c) for an example.

Other possible explanations are different geographical populations of birds migrating at different times (this may also be connected with weather) or, different age and sex categories of birds - it is a general opinion of ornithologists that adults leave Britain in the autumn before juvenile birds. There seems to be no strong evidence available to support or contradict any of these explanations, but a detailed study of ringing data from observatories may throw some light onto this subject.

The function  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$  provides a uni-modal curve (normal curve) where  $f(x)$  tends to zero as  $x$  tends to  $\pm \infty$ . The curve is characterised by two parameters :  $\mu$  (a locational parameter) and  $\sigma$  (a parameter which governs the shape of the peak). Although the use of this function imposes the restriction of

Fig. 3.2.5.1(a) Autumn Whitethroat counts - Portland 1970

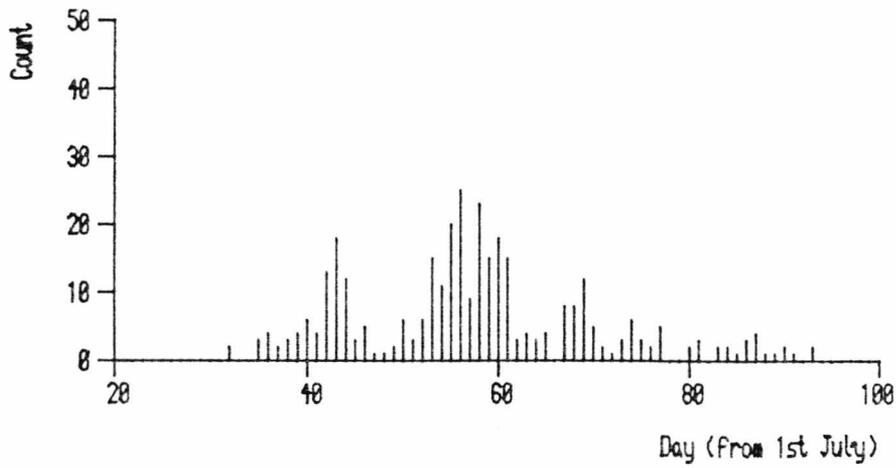


Fig. 3.2.5.1(b) Autumn Whitethroat counts - Portland 1971

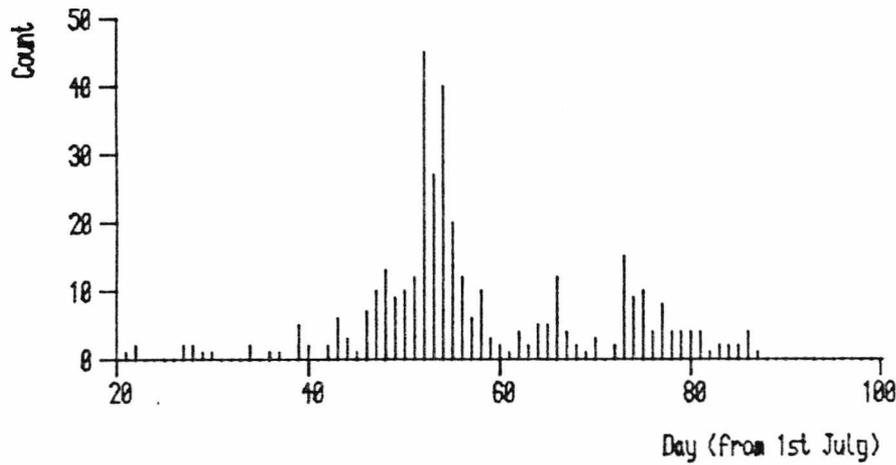
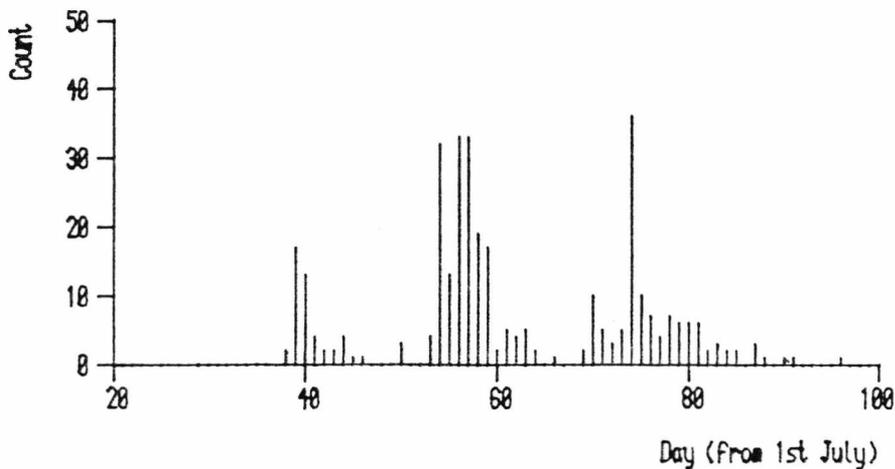


Fig. 3.2.5.1(c) Autumn Pied Flycatcher counts - Spurn 1967



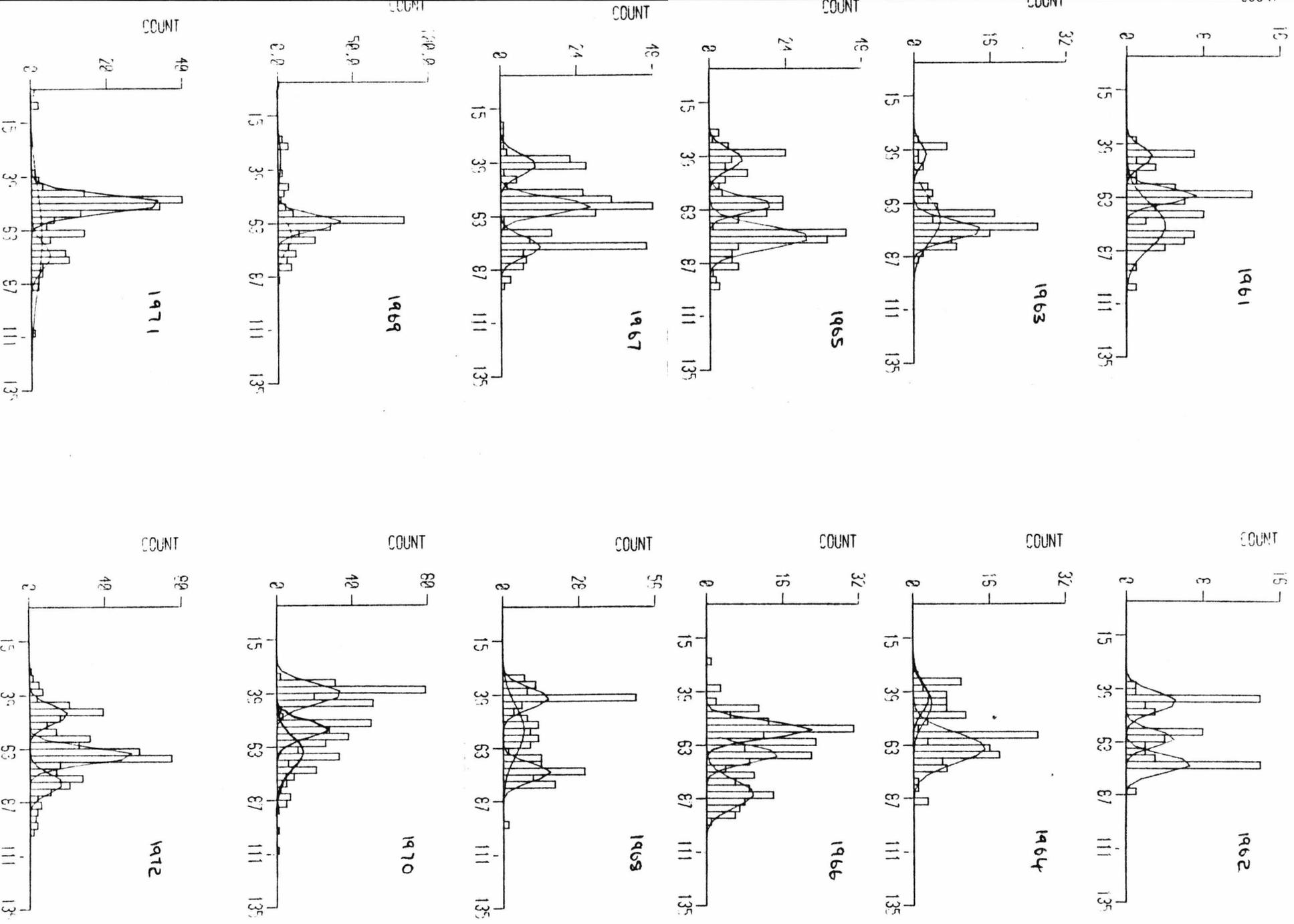
symmetry, it would seem to be a useful function to use in this situation.

It is straightforward to fit the curve, or a series of curves, to a sequence of daily counts. One can simply choose the values of  $\mu$  and  $\sigma$ , for each curve, that minimise the sum of the squares of the deviations. Further parameters need to be introduced however, in order to scale each curve to suit the patterns of counts. The function  $f(x)$  covers an area of unity, so the function  $af(x)$ , say, will cover an area of  $a$  units.

The two examples that we consider are both for the Sedge Warbler. This species appears to occur in waves more often than others, and also occurs in substantial numbers in most years. At Portland, there seems to be evidence from looking at diagrams of daily counts that three waves occur each autumn. At Dungeness, it appears that only two waves occur. If we attempt to fit more curves than there are waves, then this is no problem since redundant curves will simply serve the purpose of adjusting the shape of other curves.

Fig. 3.2.5.2 shows sets of 3 curves fitted to Portland Sedge Warbler counts. Before fitting the curves, the counts were grouped using 3-day groups. This was done in order to smooth the data to a certain extent, i.e. reducing the influence of one or two extremely large counts. In some years, one can clearly see that only two curves are

Fig. 3.2.5.2 3 curves fitted to Sedge Warbler counts from Portland. ( counts are in 3-day groups )



are necessary, e.g. 1963 and 1964. In 1969 only one curve appears to be needed. Fig. 3.2.5.3 shows the combined curves, and in most cases the fit to the data appears to be good.

Fig. 3.2.5.4 shows pairs of curves fitted to Dungeness Sedge Warbler counts, and Fig. 3.2.5.5 shows the corresponding combined curves. In some cases the fit of the curves appears rather poor, and the diagram for 1967 suggests that there is some difficulty in obtaining the optimum values of the parameters since the second curve is consistently lower than the actual counts (one would expect the curve to be lower in some places and higher in others). The reason for this may be due to the iterative minimisation process finding a local minimum, and not the global minimum. The lesson to be learnt here is that one must examine each sequence of counts carefully rather than apply an automatic procedure which deals with many sequences at once. Where possible one should try using different starting values to avoid finding local minima but sometimes, as in the case of the 1967 data, convergence on a better fitting set of curves cannot be achieved.

Tables 3.2.5.1 and 3.2.5.2 present parameter estimates of the waves of migration. Each column of the tables is divided into two or three sections, where each section refers to one wave of migrants. The estimated numbers of birds in each wave are simply the areas under each curve. The estimated dates are given as the number of

Fig. 3.2.5.3 Combined 3-wave curves fitted to Sedge Warbler counts from Portland.

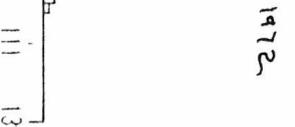
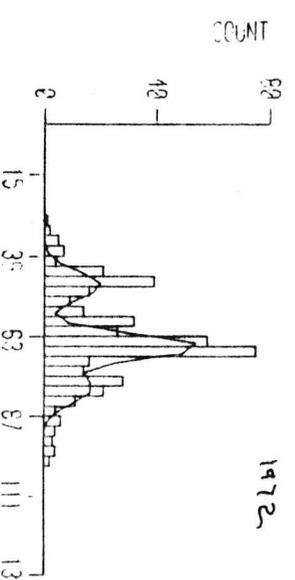
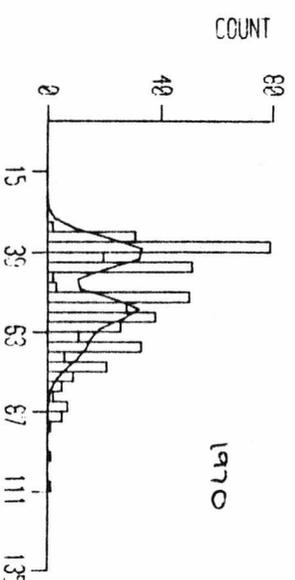
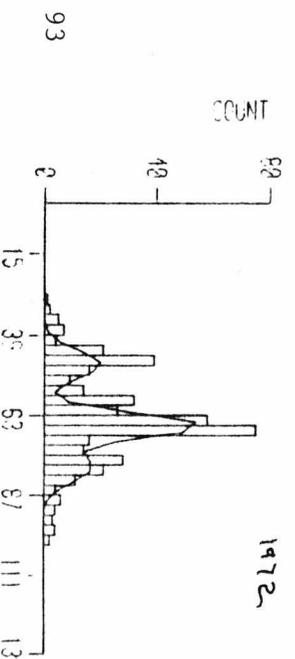
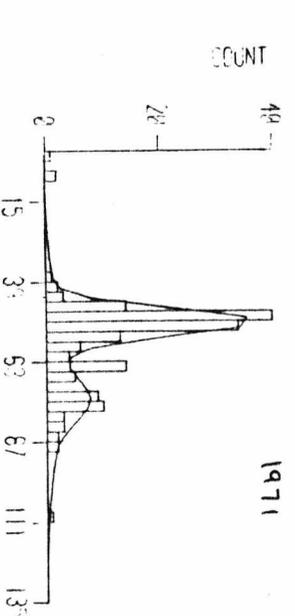
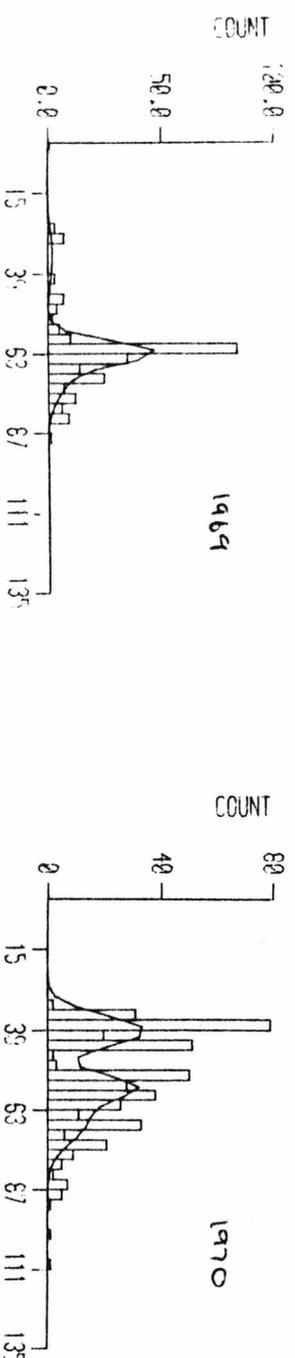
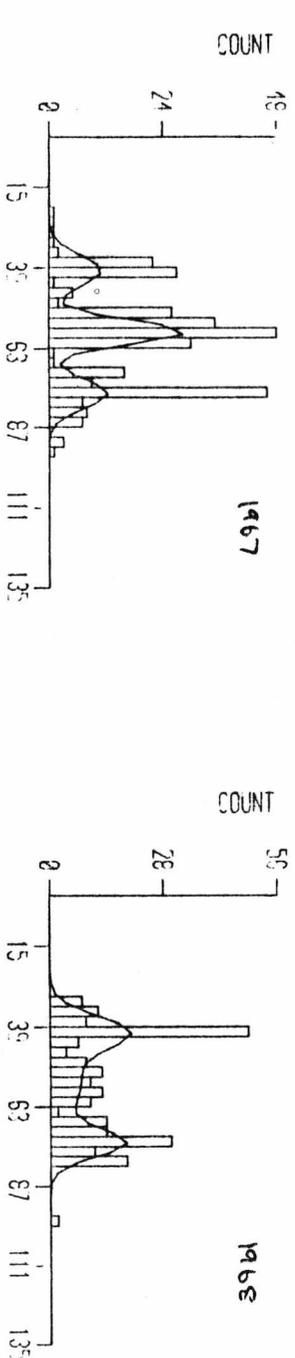
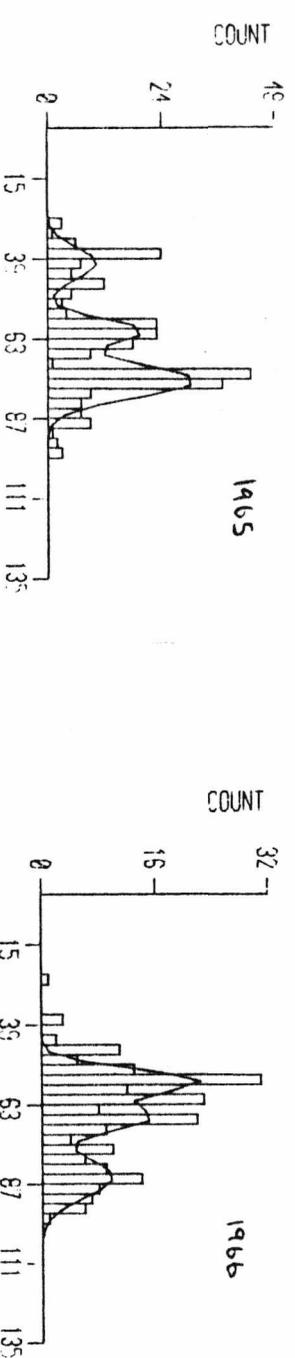
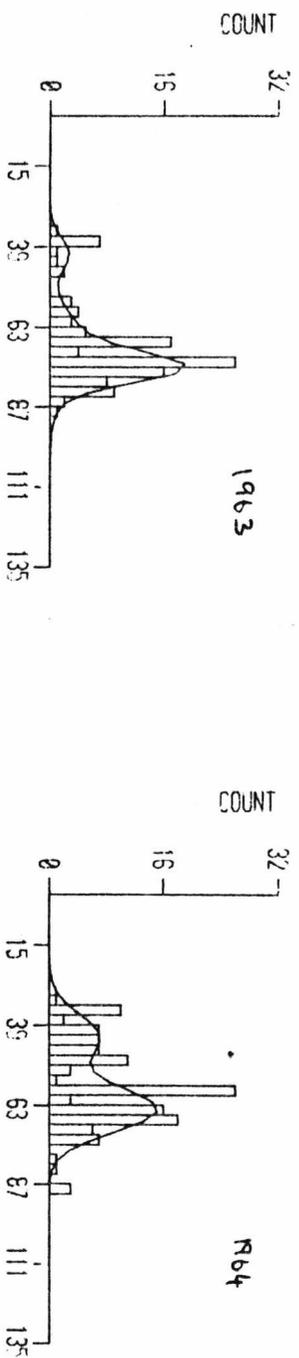
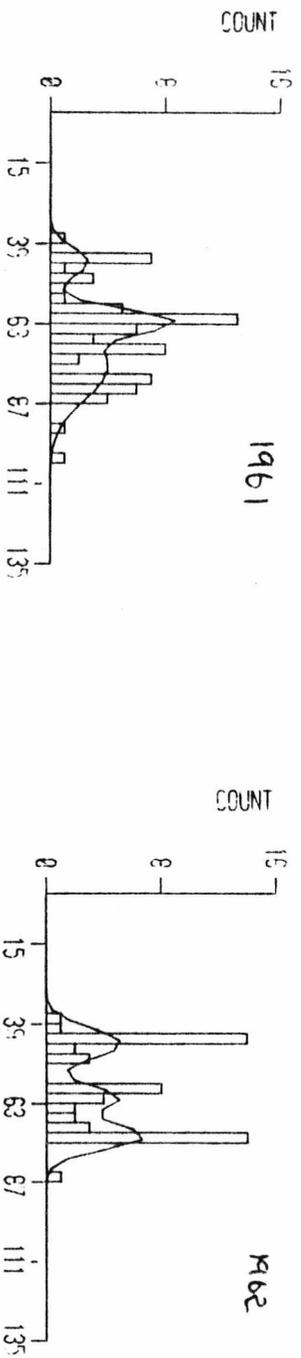




Fig. 3.2.5.5 Combined 2-wave curves fitted to Sedge Warbler counts from Dungeness.

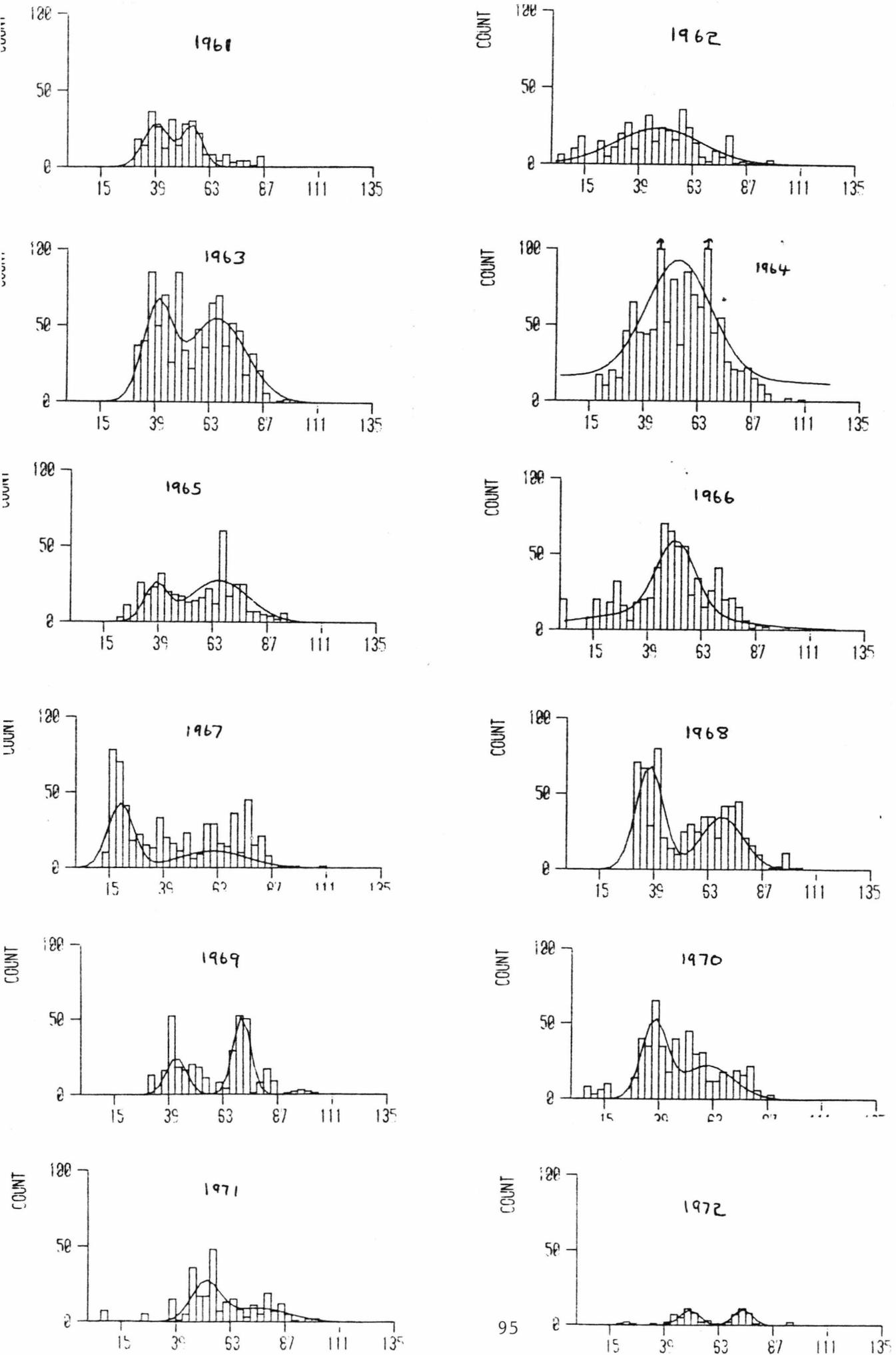


Table 3.2.5.1      3 curves fitted to Sedge Warbler counts from  
Portland. The dates are measured in days from  
1st July, and the standard deviations are  
measured in days

Year	No. of birds			Estimated date			Standard deviation		
	1	2	3	1	2	3	1	2	3
1961	1	61	22	45	63	71	4	4	7
62	23	22	23	45	61	73	4	3	4
63	1	35	30	36	73	76	5	6	4
64	46	32	54	43	57	65	9	10	5
65	31	58	99	41	62	76	5	4	4
66	120	44	39	57	73	88	8	9	5
67	25	128	75	40	58	76	5	4	4
68	102	99	7	41	73	80	5	6	5
69	110	50	93	62	65	70	3	4	34
70	151	102	46	39	57	74	5	4	10
71	103	29	20	51	65	76	4	4	4
72	39	196	48	46	66	79	11	4	4

Table 3.2.5.2      2 curves fitted to Sedge Warbler counts from  
Dungeness

Year	No. of birds		Estimated date		Standard deviation	
	1	2	1	2	1	2
1961	135	100	40	55	6	4
62	1	43	0	48	722	19
63	321	591	40	66	6	13
64	1936	936	3	55	142	14
65	97	299	38	66	5	13
66	302	355	41	52	33	9
67	202	137	20	61	6	15
68	348	272	37	69	6	9
69	92	167	43	72	4	4
70	239	311	37	61	6	11
71	137	100	52	76	6	13
72	36	35	50	74	4	4

days after the first of July. The standard deviations measure the amount of spread about the mean date, and are given in days. Most of the standard deviations are of the order of 4 or 5 days, but occasionally an extremely large value is found, suggesting that one of the expected waves was missing.

We have seen (Chapter 2) that the daily mean number of migrants recorded in each of a series of years reflect only the most noticeable fluctuations in population size. We are now in a position to decompose these counts into 'waves', and examine whether or not particular waves or combinations of waves are more useful as indices.

The sums of the areas under the two or three fitted curves are close to the raw totals of counts - any discrepancy is simply due to the lack of perfect fit of the curves.

The sample correlation coefficients between the areas of the 3 waves and the relevant CBC indices are, for Portland; .50, -.09 and .26, while for Dungeness they are .24 and .27. Combinations (pairs) of the waves were examined for the Portland data, but no significant correlation coefficients were found.

Adding together the numbers of birds estimated to be contained within each wave, in each year, offers no improvement of correlation with CBC indices over the raw

Table 3.2.5.3      Autumn totals of Sedge Warblers at Dungeness and Portland

Portland

Year	Raw Total	Sum of 3 wave areas
1961	70	84
62	55	68
63	110	66
64	130	132
65	262	188
66	205	203
67	301	228
68	236	208
69	226	253
70	412	299
71	169	152
72	422	283

Dungeness

Year	Raw Total	Sum of 2 wave areas
1961	278	235
62	376	44
63	874	912
64	1158	2872
65	403	396
66	673	657
67	612	339
68	640	620
69	351	259
70	490	550
71	261	237
72	75	71

totals of daily counts. In fact, for Dungeness data, the correlation coefficient was found to be .26, compared with .36 for the raw totals. For Portland, the respective coefficients are .16 and .02, neither of which are close to being significant.

It would appear that this rather rough and ready method of estimating the sizes of waves of migrants is not, at least for the Sedge Warbler, able to lead to migration indices that are correlated with CBC indices. This method of studying waves of migration may, however, enable ornithologists to study migration patterns in more detail than has hitherto been possible. It allows ornithologists to be objective about the timing, duration and frequency of waves which, together with a study of ringing data from observatories, would provide another interesting area of research. Ringing data may show certain waves to consist of adult or juvenile birds, male or female birds, or, if a long time period were studied, information on geographical origins of birds may be gained.

### 3.2.6 Weekend bias

Weekend bias is the term given by ornithologists to the inflation of recording effort at weekends. Whether or not weekend bias is actually apparent in observatories data is not generally known, but one thing that is known for certain is that more people visit observatories on Saturdays and Sundays than on weekdays. These extra visitors are most likely only to be day visitors though, and are unlikely to contribute to the evening meetings at which the day's records are examined. Wardens of observatories are generally of the opinion that records of commoner migrants are unaffected by these extra visitors, although if a rare bird is known to be present at an observatory then the increased disturbance caused by people looking for it may deflate the daily counts of common migrants.

Although we are concentrating our attention on common migrants it is worth noting that records of rare birds are likely to be influenced by weekend bias. The increased coverage of the recording area may yield sightings of rare birds that would otherwise have been missed. Sharrock (1966) has demonstrated weekend bias of rare bird data by counting the numbers of records made on each day of the week for the period 1958-65. He obtained his data from 'Reports of rare birds' published in British Birds. He found that 45% of records were on weekends which clearly suggests that weekend bias exists for these birds. He also demonstrated that, for island observatories, records

of rare birds are notably fewer on days on which visitors arrive or leave by boat - Sharrock (1970).

Daily ringing totals of both common and rare birds will inevitably show effects of weekend bias since ringing is heavily dependent on the effort involved.

In this section daily counts of common migrants are examined in order to see whether the widely held opinions of observatory workers are supported by the data. The simplest approach is to select a number of years to study, and examine the average numbers of birds recorded on each of the seven days of the week during migration. Obviously a Sunday in the height of the migration season will, on average, show a higher count than a Sunday at the beginning or end of a season and so in order to reduce the effects of within season time dependency we divide each season into arbitrary sections based on calendar months. A spring season thus comprises April, May and June, and an autumn season comprises July, August, September and October.

Table 3.2.6.1 shows the mean counts of four species recorded at Dungeness for the period 1961-72, for weekends and weekdays. One should, of course, include public holidays as weekend days, but in order to simplify the calculations, this point has been overlooked. Even without performing significance tests on the calculated means, it is clear that these data show no evidence of



more birds or fewer birds being recorded at weekends. In order to carry out t-tests between means the degrees of freedom need to be estimated since the variances are unequal. Under the null hypothesis of no weekend-bias, the statistic  $t = (\bar{x}_1 - \bar{x}_2)/s$  is approximately distributed with a t-distribution on  $v$  degrees of freedom, where

$$v = \frac{[(S_1/n_1)^2 + (S_2/n_2)^2]^2}{\left[ \frac{(S_1/n_1)^4}{n_1-1} + \frac{(S_2/n_2)^4}{n_2-1} \right]}$$

Table 3.2.6.1      Weekend and weekday mean counts of grounded migrants  
: 1961-72, Dungeness

SPRING

	SEDGE WARBLER			WILLOW WARBLER			WHITETHROAT			REDSTART		
	Apr	May	Jun	Apr	May	Jun	Apr	May	Jun	Apr	May	Jun
no. of weekend days	104	106	102	104	106	102	104	106	102	104	106	102
mean count	2.02	7.50	.06	8.82	11.92	.04	4.17	16.87	.09	2.36	1.70	.03
s.d. of counts	4.00	9.88	.59	22.6	22.1	.20	12.3	34.9	.57	5.93	5.09	.22
no. of weekdays	256	266	258	256	266	258	256	266	258	256	266	258
mean count	1.32	7.91	.04	9.04	9.28	.03	3.68	16.56	.12	1.49	1.18	.03
s.d. of count	3.55	13.10	.45	36.1	19.9	.26	15.1	31.4	.85	4.04	2.97	.26
t-statistic for testing weekend bias	1.55	-.33	.31	-.07	.83	.39	1.53	.08	-.39	1.37	.99	0
d.f. (approx)	131	167	120	194	133	157	157	133	175	119	117	147

\* significant at 5% level, \*\* at 1% level

AUTUMN

	Jul	Aug	Sept	Oct	Jul	Aug	Sept	Oct	Jul	Aug	Sept	Oct	Jul	Aug	Sept	Oct
no. of weekend days	107	106	103	107	107	106	103	107	107	106	103	107	107	106	103	107
mean count	1.9	11.4	7.4	.3	1.5	37.2	9.3	.1	1.9	17.1	11.8	.8	.02	1.41	6.7	1.1
s.d. of counts	4.6	16.2	11.8	.93	5.8	38.6	12.2	.4	4.2	20.1	13.2	2.7	.14	3.56	7.3	2.3
no. of weekdays	265	266	257	265	265	266	257	265	265	266	257	265	265	266	257	265
mean count	1.5	9.1	5.1	.1	1.4	28.7	8.7	.1	1.2	15.3	12.7	.5	.03	1.02	4.7	.67
s.d. of counts	4.3	10.2	6.7	.4	4.8	33.7	23.4	.3	3.7	16.3	23.1	1.4	.17	2.36	6.7	1.7
t-statistic for testing weekend bias	.77	1.36	1.86	2.15*	.16	1.99	.32	0	1.5	.82	-.46	1.09	-.58	1.04	2.4*	2.16*
d.f. (approx)	137	118	113	112	130	131	227	126	134	128	207	115	159	120	131	125

\*significant at 5% level, \*\* at 1% level

The observed t-statistics are given in Table 3.2.6.1. Only three out of 28 data sets exhibit weekend bias, and therefore no conclusive evidence is available that, in general, daily counts are affected by weekend bias. Before leaving this topic, we shall briefly examine data from other observatories.

Table 3.2.6.2 refers to the Sedge Warbler autumn records over a seven year period: 1961-67. Again we see no evidence of weekend bias. We have only examined one species at four observatories, but there is no reason to suspect that the results would be different for other species, and over different time periods.

We therefore accept our initial suspicion that daily counts of grounded common migrants are not affected by the extra visitor numbers at weekends.

Table 3.2.6.2      Weekend and weekday mean counts of Sedge Warblers :  
1961-67 Autumn only

	BARDSEY				PORTLAND				SPURN			
	Jul	Aug	Sept	Oct	Jul	Aug	Sept	Oct	Jul	Oct	Sep	Oct
no. of weekend days	64	60	61	63	64	60	61	63	64	60	61	63
mean count	.13	3.67	.33	.02	.03	2.68	2.54	.06	1.66	1.47	.59	.05
s.d. of counts	.58	10.4	.87	.13	.18	4.88	3.22	.35	1.70	1.46	.94	.22
no. of weekdays	153	157	149	154	153	157	149	154	153	157	149	154
mean count	.47	1.22	.50	.01	.03	2.49	2.68	.10	1.37	1.47	.49	0
s.d. of counts	1.97	2.89	1.51	.08	.27	4.76	4.58	.46	1.76	1.52	.93	0
t-statistic for testing weekend bias	-1.94	1.8	-1.02	.56	0	.26	-.25	-.69	1.13	0	.70	1.8
d.f. (approx)	213	60	123	70	115	76	103	100	88	78	80	62

\* significant at 5% level, \*\* at 1% level

### 3.2.7 Large falls of birds

Many visitors are attracted to observatories by the possibility of seeing a large 'fall' of migrants, i.e. a large number of grounded birds which often includes rarities.

Ornithologists believe that the ideal conditions for a fall are when weather has prevented migration for several days leading to a build up of birds at some origin. When weather conditions allow birds to proceed and if local weather at observatories changes for the worse then large numbers of migrants may be forced to land.

Another type of fall is that of an 'overshoot'. This feature is particular to south coast observatories and occurs when weather conditions are ideal for migration for several days, encouraging spring migrants to fly further north than is usual. Species which are rare to Britain are often recorded, e.g. Hoopoe and Bee-eater.

It is difficult to define objectively a large fall of migrants. In certain migration seasons one may observe one large fall among a season of low counts in which case it is clear which observation is a fall. In other seasons the picture is less clear. In order to examine the occurrence of falls at Dungeness and Portland observatories, we define falls as in Table 3.2.7.1. These arbitrary definitions were arrived at by visually examining plots of daily counts and choosing a threshold value for each species/observatory combination which

separates the main flow of migrants from the small number of particularly large counts.

Table 3.2.7.2 shows when the falls occurred during the period 1961-72. Dungeness clearly shows more falls than Portland, and both observatories show more falls in autumn than in spring.

A close examination of the dates on which falls occur suggests that it is not always the case that species occur together. Of the 69 falls at Dungeness, 44 were 'single species' falls. Although data are only presented here for 4 species, further species were examined - Spotted and Pied Flycatcher and Blackcap. No further coincidences of falls were found. At Portland, 25 of the 27 falls were single species falls.

Although observatory daily census records suggest that different species migrate together, it is clear that this is not a general rule. This feature of different behaviour between species is found in other contexts throughout this thesis, for example the influence of weather on observatories data.

<u>Dungeness</u>	<u>No. of birds</u>	<u>No. of falls 1961-72</u>
Redstart	≥ 50	2
Sedge Warbler	≥ 50	14
Whitethroat	≥ 100	18
Willow Warbler	≥ 100	35
<u>Portland</u>		
Redstart	≥ 50	1
Sedge Warbler	≥ 50	4
Whitethroat	≥ 50	2
Willow Warbler	≥ 50	20

Table 3.2.7.2 Falls of migrants at Dungeness and Portland

DUNGENESS

Year	SPRING				AUTUMN				Total
	RS	SW	WT	WW	RS	SW	WT	WW	
1961									0
62		2	2				2	3	9
63	1	2	3	2		2	2	4	16
64			2	2		3	1	3	11
65		1	2					1	4
66				2		1		1	4
67			1	1			2		4
68		1		1		1	1	4	8
69								1	1
70	1			2		1		4	8
71								1	1
72				2				1	3
Total	2	6	10	12	0	8	8	23	69

PORTLAND

1961									0
62									0
63									0
64					1				1
65									0
66									0
67		1							1
68		1	1				1	3	6
69						1		6	7
70						1		3	4
71								2	2
72				5				1	6
Total		2	2	5	1	2		15	27

RS : Redstart  
 SW : Sedge Warbler  
 WT : Whitethroat  
 WW : Willow Warbler

Table 3.2.7.2 allows one to compare Dungeness with Portland observatory as regards falls. There is clearly little similarity between the two observatories over the four species considered over the years 1961-72. A study of the dates of falls reveals only one coincidence: in the autumn of 1968 both observatories experienced a fall of Willow Warblers on the same day. As well as same-day coincidences, near-coincidences were examined. One Sedge Warbler fall at Portland was followed by a fall at Dungeness the next day and two Willow Warbler falls at Portland were followed by falls at Dungeness, also the next day. All three coincidences or near-coincidences were of autumn falls, suggesting a build up of birds along the south coast at around the same time. However, with coincidences representing such a small proportion of the total number of falls it is difficult to draw firm conclusions regarding occurrence of falls at Portland and Dungeness. This lack of similarity between data from Dungeness and Portland highlights the different geographical positions of the two sites, and is found in many of the analyses of later chapters of this thesis.

Data from other observatories were examined for coincidences of falls within species. The only suggestion of a link between two observatories was that on six days over the twelve year period large falls (2 of Whitethroats and 4 of Sedge Warblers) at Dungeness were followed by large falls at Bardsey (of the same species) either one or two days later. When one considers that six observatories, four species and twelve years of data were examined, it is clear that the timing and species involved in large falls differ between observatories. This has important

implications for the study of large falls. It would appear that one must examine individual observatories, rather than grouping data from a number of sites, and it also suggests that species must be examined individually.

If one attempts to fit a statistical model to daily counts data, large falls will inevitably show up as 'outliers', i.e. observations with residuals which are large compared with the main body of residuals. This is a common problem when modelling observed data and the action that one takes depends on the nature of the analysis. For example, one might be interested in how closely the model fits the main body of the data and not worry too much about isolated departures from the model. Another example is where one may be interested in observations which are influential in the fitting of the model, i.e. how dependent are the parameter estimates on certain data? This point is returned to in chapters 6 and 7.

### 3.2.8 Missing data

On some days it may be impossible for a daily census to be carried out at an observatory. This may be due to extreme weather conditions, or, more likely, because nobody is able to devote the necessary time. This is most unlikely during the height of the migration seasons since many visitors will be present to assist the warden.

This thesis concentrates on data from Dungeness and Portland which include very few missing observations, and where one occurs, a zero count has been inserted. This was done since the missing observations occurred at the extreme ends of the seasons, when the counts were very low. Since the counts are the result of complicated interactions between weather conditions and biological behaviour it is difficult to impute an estimate of the count. The small number of missing counts (approximately 0.01%) found in the data from these two observatories should mean that any statistical analyses performed on the data are not influenced to a great extent by these missing data.

One other source of missing observations is the tendency for some wardens to insert 'ticks' in the migration log instead of numerical estimates. These are usually intended to signify that a particular species is present but not counted. There is little that one can do here except treat the observation as missing. It hardly seems appropriate to insert a zero count for these observations, but one would expect that if the species had occurred in large

numbers, then an estimate would have been made. Again, very few of the data sets from Dungeness or Portland include ticks.

### 3.3 Weather data from Dungeness and Portland Bill

#### 3.3.1 Introduction

The aim of this section is to describe and explore weather data from meteorological stations adjacent to the observatories of Portland Bill and Dungeness, in order to prepare the way for an analysis of how weather affects the occurrence of migrants at observatories. Simple statistical procedures are used, including principal components analysis and the fitting of univariate time series models to individual weather variables. The former technique is used as a means of reducing the dimensionality of the available data by considering relationships among the variables, and the latter technique throws light on the time dependent structure of the data. This is the first instance of such techniques being applied to a study of this nature and hopefully this work will demonstrate to ornithologists that techniques other than multiple regression are applicable in this area.

An alternative approach to adopt here is to consider the set of weather variables (each measured at equally spaced points in time) as a multiple time series, and attempt to fit multivariate moving average and/or autoregressive models. Although these data do, in fact, represent a multiple time series, it is difficult to see that this approach would offer anything over and above a combined principal component and univariate time series analysis. As will be seen from the principal component analysis,

there is no evidence of changes over time of relationships between weather variables, and so there appears to be no objection to looking at the variables one-by-one. If, on the other hand, our task was one of predicting future weather conditions from an historical set of observations, then we might well improve the accuracy of our forecasts by modelling the series jointly. Also, principal component analysis often enables sensible interpretations of groups of variables, and fitting univariate time series models is, with the current availability of software, more straightforward than fitting multivariate models.

Since data from Portland and Dungeness are studied, the migration of interest is cross-channel migration. The only major study of cross-channel passerine migration has been that of Parslow (1962, 1969), who used radar to measure migration densities, and compared them with weather information extracted from Daily Weather Reports. Parslow's first paper concerns arrivals of night migrants at Portland and Hengistbury Head (Hampshire). Among his findings concerning the influence of weather on migration is the suggestion that wind direction in the upper air determines the tracks of the migrants. He gives an example where a wind from the NE deflected spring migrants onto a westerly heading. This means that under such conditions, British immigrants are unlikely to be recorded at the Kent observatories (Dungeness and Sandwich Bay) since they will have been deflected to the west.

Parslow estimated the geostrophic wind directions by studying isobars on weather maps, and he makes the point that local wind directions are likely to be quite different. Owing to the large time period considered here, we work only with local wind directions which are readily available and may thus expect different results from Parslow.

As regards falls of migrants at Portland and Hengistbury Head, Parslow found that large falls were not necessarily indicative of a large migration across the channel. On overcast nights with light winds and rain, the radar picture showed a relatively small migration across the channel, but on the following mornings large numbers of grounded migrants were reported. Strong, opposed winds gave little migration and no falls, suggesting that birds were prevented from leaving France. Calm anticyclonic conditions gave rise to dense migration, but produced only small falls, suggesting that migrants flew over the observatories into Britain.

Differences between the timing of falls at the two sites were frequently noted, but then the situation of the two observatories is different. As one might expect, such differences occur between Dungeness and Portland. Dungeness is situated at the tip of a low lying peninsula, in a relatively narrow stretch of the English Channel. Portland Bill is effectively an island off the Dorset coast at a wide section of the channel.

Parslow (1969) extended his study using data from Dungeness and Hengistbury Head, and confirmed his original findings. He also noted that the temperature appeared to be correlated with heavy immigration in spring, but this may just be due to the onset of the migration season coinciding with the general increase in temperature as spring gives way to summer.

Parslow's studies are further discussed by Riddiford and Augur (1980) and Riddiford (1985).

### 3.3.2 The data available

Three hourly observations on a number of weather variables are available from Dungeness from 1957 to 1972 inclusive, and from Portland from 1968 to 1972 inclusive. The data were obtained from the Meteorological Office Archives, on computer-readable magnetic tapes. For both sites, the observations were made at lighthouses situated on the observatory recording areas.

Eleven variables are available at each observation time:

wind direction (degrees from north)

wind speed (knots)

total and low cloud amounts (oktas)

visibility (distances to known landmarks in metres)

air and wet bulb temperatures ( $^{\circ}\text{C} \times 10^{-1}$ )

dew point ( $^{\circ}\text{C} \times 10^{-1}$ )

vapour pressure (millibars  $\times 10^{-1}$ )

humidity (measured as a percentage)

atmospheric pressure (millibars)

The wind direction variable needs careful consideration, since it is recorded as an angle in degrees measured from north. The problem is simply that, for example,  $350^{\circ}$  and  $10^{\circ}$  both represent the same degree of deviation from north, while having very different numerical values. It seems reasonable at this stage to introduce knowledge of the direction of flight over the channel of birds occurring at south coast observatories. Parslow (1969) found, by using radar, that the direction of immigration into Britain in spring was, under calm conditions, NW or NNW. A transformation of the form  $\sin(\sigma - 45^{\circ})$  where  $\sigma$  is the recorded angle gives a new variable that has a maximum value (+1) when the wind is SE, and a minimum (-1) when the wind is NW. Crosswinds give a value of zero. The effect of this transformation is to give a quantitative variable measuring the approximate favourability of wind direction to spring migration, and unfavourability for autumn migration.

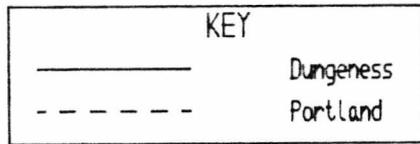
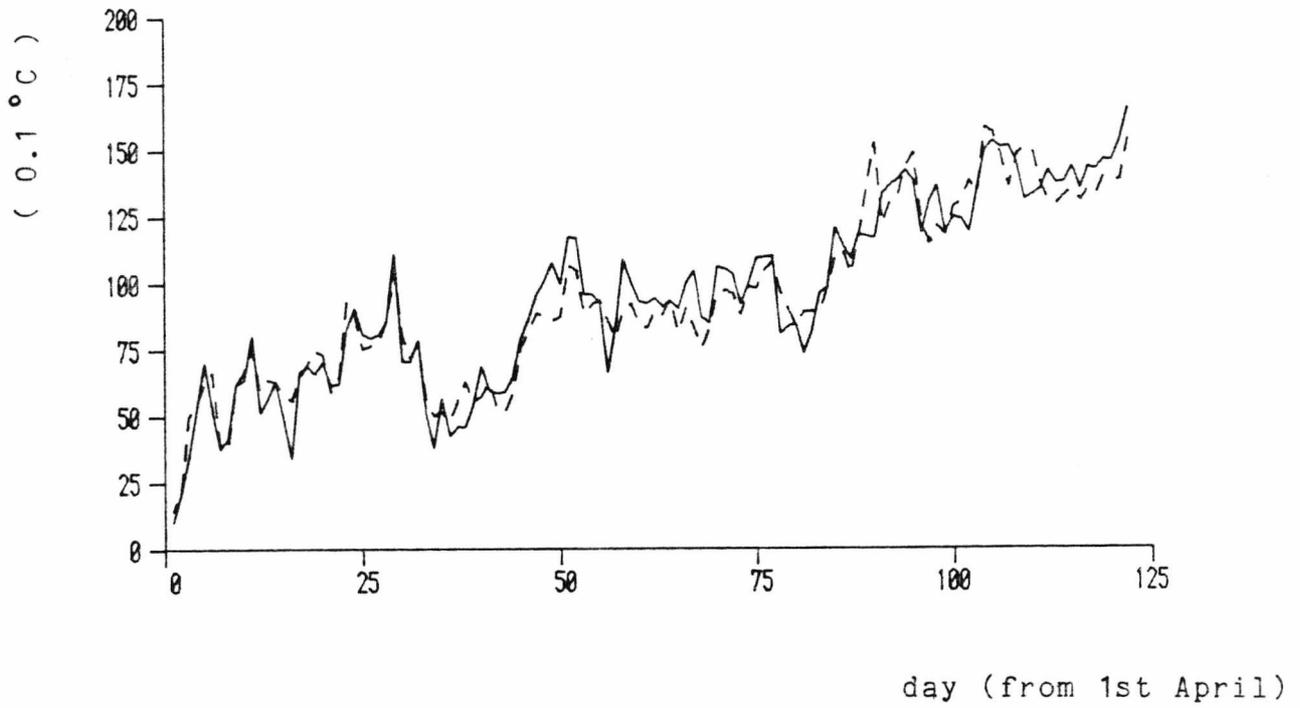
One notable absentee from the list of weather variables is that of rainfall. It is not possible to calculate the amount of rainfall in such short intervals as three hours, but since we have variables such as cloud cover, humidity and vapour pressure, we can assume that we have information on rainfall, however indirect this information is.

An important point to note here is that birds recorded at observatories were, prior to their arrival, influenced by weather elsewhere. However, the weather at observatories does, we expect, influence these arrivals and in any case these local weather records are all we have to work with at present. Future research in this area would no doubt benefit from the collection of weather data from, for example, France.

Before commencing statistical analyses of the data, we illustrate two of the variables by time series plots of daily means. Fig. 3.3.1 shows the air temperature over the months of March to June, 1968, at both Dungeness and Portland. The similarity of the two sites as regards temperature is clearly seen. The curves for autumn (July - October) show two interesting features. There appears to be a slight increase in temperature over the first 50 or 60 days, followed by a continual decline as winter approaches. This pattern has important consequences when attempting to detrend the series, since the trend is clearly non-linear. A second feature of these autumn data is that the temperature experienced at Portland, often occurs at Dungeness a couple of days later, suggesting a westward moving weather pattern in autumn. This feature is interesting in the context of daily counts of migrants, since it may provide an explanation of the differences between these sites as regards migration data.

Fig. 3.3.1 Air temperature data

Spring 1968 - Air Temperature



Autumn 1968 - Air Temperature

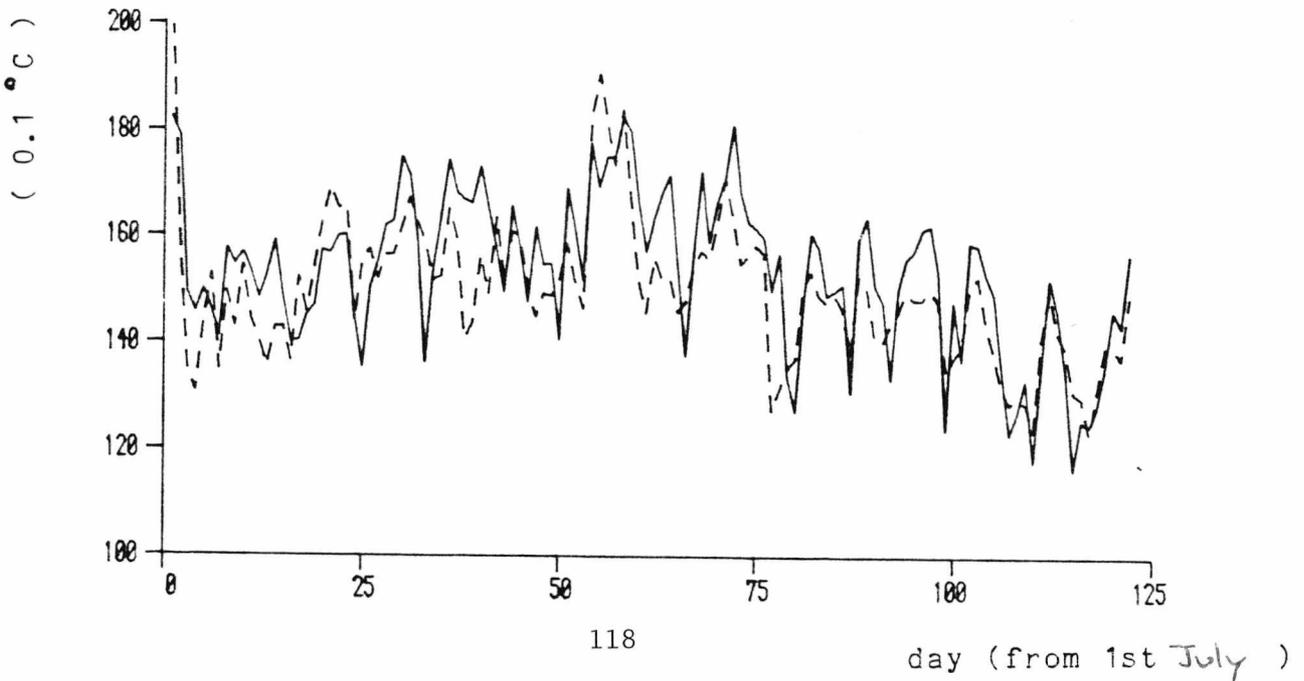


Fig. 3.3.2 shows the total cloud amount for both observatories over the spring of 1968. Large day to day variation is seen and differences between the observatories are noticeable, particularly in early July, and one can again see that a time lag of a few days occurs in the data.

A comparison of daily counts of grounded migrants from Dungeness and Portland showed no evidence of a 'lag' of the form seen in the weather data.

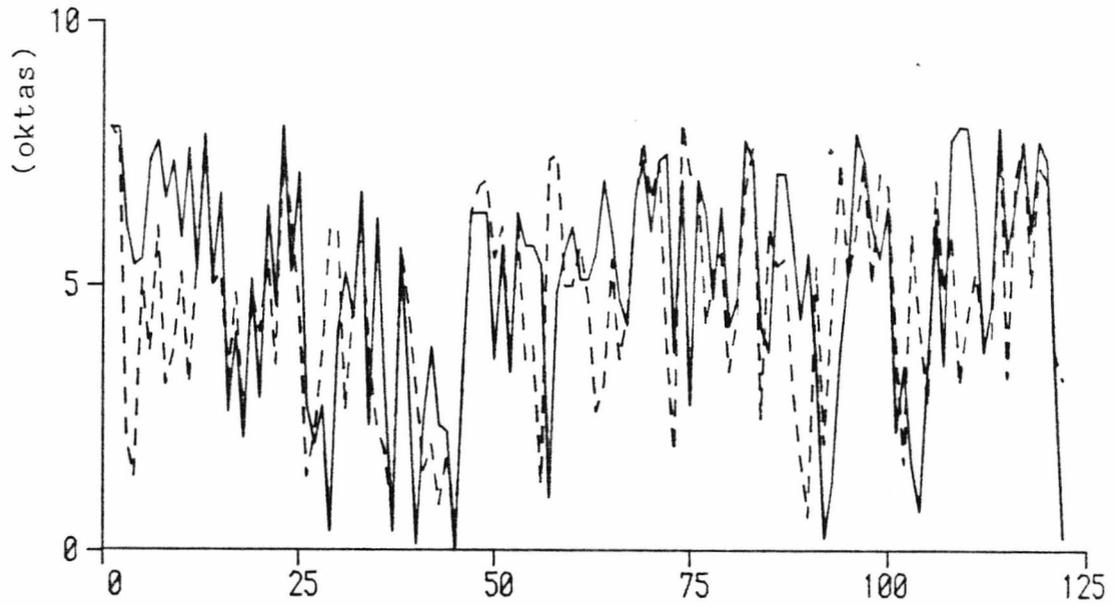
A more detailed study of the structure of the data as a whole is now considered.

### 3.3.3 A principal component study of weather data

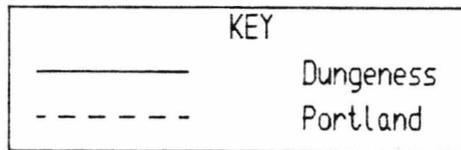
In this section we carry out a principal component analysis of a set of weather variables. We are interested here in relationships between variables. In the following section we examine the autocorrelation structure of the data. It should be remembered that we have, in effect, a multivariate time series. Usually, one performs a principal component analysis on a set of independent observations on some multivariate random variable. The successive observations on weather variables are not independent as they are measured sequentially at, for example, daily or three-hourly intervals. It is not clear from the literature whether principal components analysis is appropriate for data of this type, but as this section demonstrates, one obtains quite sensible results.

Fig. 3.3.2 Total cloud cover data

Spring 1968 - Total Cloud



day (from 1st April)



The motivation for the analyses reported in this section is largely one of exploration of a data set which will be used in further chapters. We would do well to study which weather variables or groups of variables fluctuate most from day to day, as opposed to variables which are nearly constant over the time period involved. By carrying out the analysis of weather data from Portland and Dungeness separately, we may gain insight into the types of weather which occur at each site, and the nature of any differences that may be found.

For the first analyses reported below, we examine four different months (March to June inclusive) separately. This is in order to check for consistency over a migration season. One might find that if variations in the structure of the weather data occur over time, then the influence of weather on migration data may also vary over time. The analyses of this section will not, of course, fully answer this question, but they should provide some insight into the problem. Spring data only are studied at first, but analyses of autumn data gave very similar results, some of which are reported later in this section.

Principal components analyses were carried out on data from Portland and Dungeness separately, in order to check for site differences. No differences were found, however.

Further motivation for principal components analysis is given by the fact that this technique often enables data

reduction, i.e. unimportant variables may be discovered and subsequently discarded from future analyses.

As a preliminary analysis, three-hourly observations on all eleven weather variables are studied, for both sites.

The eleven variables are:

- wind direction (transformed from an angle, as described in section 3.3.2)
- wind speed (knots)
- total cloud amount (oktas)
- total amount of low cloud (oktas)
- visibility (metres)
- air temperature ( $0.1^{\circ}\text{C}$ )
- wet-bulb temperature ( $0.1^{\circ}\text{C}$ )
- dew-point ( $0.1^{\circ}\text{C}$ )
- vapour pressure (10mb)
- humidity (%)
- pressure (mb)

The cloud variables and visibility are measured on a discrete scale.

The analyses were performed using GENSTAT, and any observation containing a missing value is automatically deleted. Over the 976 observations for the spring of 1968 - March to June, (the year chosen for this preliminary analysis), dew point was not recorded on 37 occasions at Portland, and 57 at Dungeness. The temperature was not recorded on one occasion at Portland, and 17 occasions at Dungeness.

The first step in a principal component analysis is usually the examination of the correlation matrix. With eleven variables one is looking at 55 coefficients and this can prove a difficult exercise, particularly as one expects approximately 2 or 3 coefficients to appear significant at the 5% level, even when no correlation exists.

8 correlation matrices were examined (2 observatories x 4 months : March, April, May and June), and the 28 pairs of matrices were tested for differences by the likelihood ratio test. The test statistic is given by

$$x^2 = n \log_e |\underline{s}| - n_1 \log_e |\underline{s}_1| - n_2 \log_e |\underline{s}_2|$$

where  $\underline{s}_1$  and  $\underline{s}_2$  are the correlation matrices being tested,  $n_1$  and  $n_2$  are the sizes of the samples from which  $\underline{s}_1$  and  $\underline{s}_2$  were calculated, and

$$\underline{s} = (n_1 \underline{s}_1 + n_2 \underline{s}_2) / (n_1 + n_2)$$

Under the null hypothesis of no difference,  $x^2$  follows a chi-square distribution with  $\frac{1}{2}p(p+1)$  degrees of freedom. None of the 28 test statistics was found to be significant at the 5% level. This consistency over time is a feature which is seen again when we consider the principal components.

Significant correlations (5% level) were found for the following pairs of variables (all positive correlations):

- total and low cloud amount
- wet bulb and air temperature
- dew point and air temperature

vapour pressure and air temperature  
dew point and wet bulb temperature  
vapour pressure and wet bulb temperature  
vapour pressure and dew point .

The meteorological explanations behind these correlations will not be entered into here, but some of the pairs of variables shown above, would be expected to show high correlations, e.g. wet bulb and air temperature.

This examination shows that little information would be lost if, say 2 or 3 variables out of air temperature, wet bulb temperature, vapour pressure and dew point were discarded from the data set.

Several methods exist for discarding variables using principal components knowledge, and descriptions can be found in Jolliffe (1972a,b). The most convenient methods are those where the number of variables to be retained is decided beforehand. One simply examines the last principal component, and discards the variable with the largest (in magnitude) coefficient. The next to last component is then examined, and this procedure continues until the required number of variables have been discarded. Although we have no fixed number of variables to discard in these analyses, this method does suggest that four variables can be discarded without too much loss of information, since in all 8 analyses the first four variables that would be discarded are very nearly the same four for each data set.

Table 3.3.1 illustrates this feature, where the order of deletion reads from left to right. This table also shows that there appears to be little variation in the structure of the weather data from month to month.

Table 3.3.1      Variables discarded (see below for key)

Portland	:	March	WB	VP	DP	LC	WS	V
"	:	April	WB	DP	VP	TC	V	WD
"	:	May	WB	VP	DP	LC	P	WS
"	:	June	WB	VP	DP	LC	V	P
Dungeness	:	March	WB	VP	DP	LC	WS	WD
"	:	April	WB	VP	DP	TC	P	WD
"	:	May	WB	VP	DP	TC	V	P
"	:	June	WB	VP	DP	LC	V	P

Key: WB = wet bulb temperature  
 VP = vapour pressure  
 DP = dew point  
 LC = low cloud  
 TC = total cloud  
 WS = wind speed  
 P = pressure  
 V = visibility  
 WD = wind direction

Although the fifth and sixth variables show no pattern of consistency over the eight analyses, the first four are very nearly the same.

One of the eight correlation matrices is given in Fig. 3.3.3. We see that wet bulb temperature is highly correlated with air temperature, dew point and vapour pressure, and the principal components analysis appears to be telling us that just one of these four variables is sufficient for practical purposes, this being air-temperature. Further, the fourth column of Table 3.3.1 clearly shows that only one of the two cloud variables needs to be retained. Note that the

correlation between these variables is very high: 0.76  
for the given example.

Fig. 3.3.3      An example correlation matrix for weather variables  
(Portland, March 1968)

wind dir.	1.00											
wind speed	.02	1.00										
total cloud	.17	.15	1.00									
low cloud	.01	.25	.76	1.00								
visibility	-.27	.12	-.32	-.23	1.00							
air temp.	.25	.18	.13	-.13	.01	1.00						
wet bulb temp	.24	.20	.17	-.09	-.06	.96	1.00					
dewpoint	.24	.20	.21	-.02	-.14	.81	.94	1.00				
vapour press.	.24	.20	.21	-.02	-.16	.81	.94	.99	1.00			
humidity	.01	.07	.14	.14	-.24	-.07	.21	.51	.52	1.00		
pressure	-.15	-.60	-.24	-.31	.02	-.08	-.08	-.07	-.07	-.02	1.00	

The number of variables that can be discarded without appreciable loss of information can also be determined by examining the eigen values of the correlation matrix. When expressed as a percentage of the trace of the matrix, each eigen value gives the amount of total variation accounted for by its corresponding eigen vector, or principal component. If we find that the last eigen value is small in comparison with the penultimate one, then this suggests that the data could be reduced by one more dimension.

One aspect of principal component analysis is that of examining the data by means of plotting principal component scores for various pairs of components. These two-dimensional plots can often reveal such features as multi-modal distributions, i.e. when clusters of points are found, or outliers, i.e., when points lie far away from the main

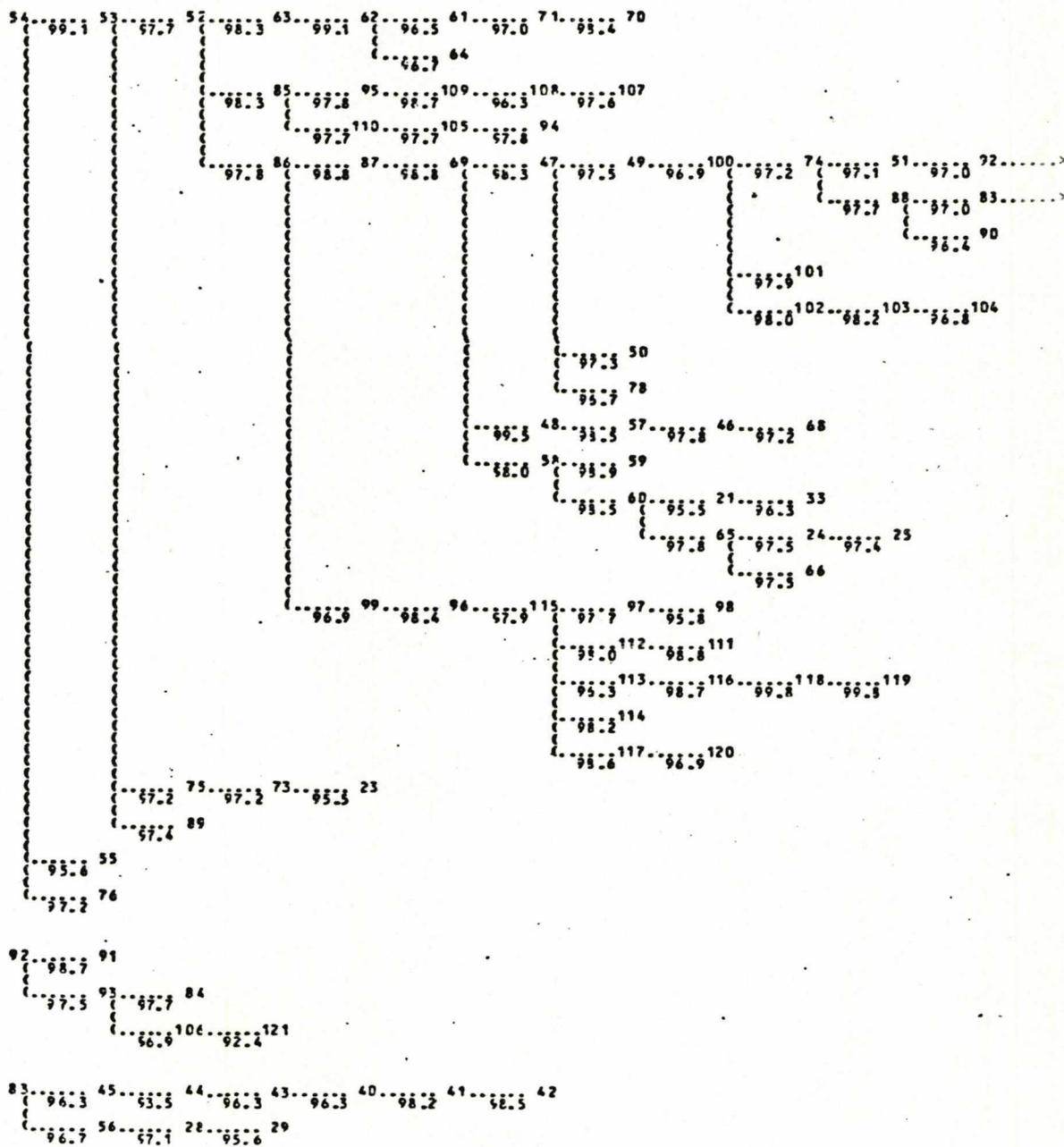
group. In all 8 analyses performed, plots of the first two components were examined, but no such features were found. Since we are looking at the data in only two dimensions, there may well be distortion introduced which can make observations appear close together, but which are in effect far apart. The construction of a minimum spanning tree can lead to a closer examination of how closely related observations are, but for such large numbers of observations they can be difficult to interpret. A part of the tree for the Portland data of March 1968 is shown in Fig. 3.3.4. Each integer represents one observation (1 - 248) and the real numbers separating each observation give the distances (Mahalanobis distance) in eleven dimensional space between neighbouring observations. (Note that the Mahalanobis distance is just one of a number of possible distance measures that may be used here.) From this example, one can see that observations that are close together in time, are often close together in space, a feature that we return to later in this chapter.

A second analysis is now described, where the weather data are transformed to daily means from 3 hourly observations. Since we have daily observations of birds, it seems sensible to investigate the weather data on a daily basis, and in doing so we can combine data for, for example, the four months of spring and end up with 122 observations.

Fig. 3.3.4 A section of the Minimum Spanning tree,

Portland, March (1968).

( see text for details )



The correlation matrices for Dungeness and Portland were found to be very similar, and the likelihood ratio test gave no evidence of a difference between the two sites. There is again a suggestion that at least four variables could be discarded without appreciable loss of information.

Table 3.3.2 shows the eigen vectors for the Dungeness data, together with the % variance accounted for by each one, to one decimal place.

The eigen vectors for the Portland data were found to be very similar to those for Dungeness.

Frequently one finds that the first few principal components can be interpreted practically, see for example, Jolicoeur and Mosimann (1960). We now make a brief attempt to interpret the components found in the above example.

The first component (see Table 3.3.2) accounts for 39.4% of total variation. Whether or not this proportion represents a significant effect or whether such a figure could have arisen purely by chance is difficult to assess. Krzanowski (1979) gives percentage points for the ratio of the largest eigen-value to the trace of the correlation matrix, but only for the cases of three or four variables. From his tables there is evidence to suggest that for 11 variables, 39.4% would represent a significant effect when a large number of observations is used (here, we have 122). The

Table 3.3.2 Principal components for Dungeness, 1968 data

P.C.	Weather variable number ( see key below )										
	1	2	3	4	5	6	7	8	9	10	11
1	-.16	.00	-.10	-.03	.16	-.44	-.46	-.47	-.47	-.27	.06
2	-.17	.24	.54	.56	-.16	-.19	-.12	-.04	-.04	.37	-.32
3	-.15	-.63	.12	.09	-.59	-.06	-.05	-.04	-.03	.09	.44
4	.74	-.15	-.13	-.17	-.30	-.18	-.14	-.09	-.12	.24	-.40
5	.61	.11	.29	.35	.23	.02	-.00	-.05	-.03	-.17	.57
6	-.03	.70	-.29	-.11	-.50	-.08	-.05	-.02	.00	.17	.36
7	-.05	-.12	-.11	-.16	.45	-.20	-.08	.04	.02	.78	.29
8	-.00	.12	.70	-.70	-.04	-.01	-.02	-.02	-.05	-.03	.06
9	-.01	.02	-.01	.03	-.01	.47	.31	-.05	-.80	.19	.00
10	-.00	.00	-.00	.01	.00	-.31	-.17	.86	-.33	-.14	.02
11	-.00	-.00	.00	-.00	-.00	-.61	.78	-.10	-.06	-.07	-.00

P.C.	% variance accounted for	( key for above table ) Weather variable
1	39.4	1. Wind direction
2	23.8	2. Wind speed
3	12.5	3. Total cloud amount
4	9.2	4. Low cloud amount
5	5.4	5. Visibility
6	5.2	6. Air temperature
7	3.4	7. Wet-bulb temp.
8	1.0	8. Dew-point
9	0.1	9. Vapour pressure
10	0.0	10. Humidity
11	0.0	11. Pressure

eigen vector for the first principal component (see Table 3.3.2) is essentially an average of air and wet bulb temperatures, dew point and vapour pressure. In meteorological terms, this component could be a measure of atmospheric stability.

The second component is an average of the two cloud measures, suggesting again that just one variable could be retained. Both humidity and pressure feature in this component, which are relevant to the formation of cloud.

The third component is an average of wind speed and visibility contrasted with pressure. The fourth component is a measure of wind direction against pressure. Already we see that interpretation is difficult.

The first four components comprise almost 85% of the total variation in the data.

The last few components often reveal those features of the data which are approximately constant over the whole data set, i.e., in this case, over four months of spring. In our example, the last component represents the difference between the two temperature measures, while the next to last component highlights the consistency of dew point over time.

We have seen from more than one viewpoint that at least four variables may be deleted from the data set, and this

we now do. We then make a further attempt at interpreting the resultant components.

Fig. 3.3.5 shows the correlation matrix for the seven remaining variables for the spring data at Dungeness. Again, the correlation matrices for the autumn data, as well as the data from Portland, were very similar, and no significant differences were found among the 4 matrices.

Fig. 3.3.5      Correlation matrix for Dungeness, 1968 Spring

Wind direction	1.00						
Wind speed	.05	1.00					
Low cloud	.09	.27	1.00				
Visibility	.15	.23	-.22	1.00			
Air temp.	-.07	-.05	-.18	-.15	1.00		
Humidity	.00	.16	.44	-.46	.23	1.00	
Pressure	.01	-.35	-.28	-.01	.04	-.38	1.00

The eigen vectors of the correlation matrices are shown in Table 3.3.3. As for the dimensionality of the reduced data sets, we see that the first four variables that would be discarded by using the procedure outlined earlier are as follows:

```
Dungeness (spring):humidity , cloud, pressure , wind dir
" (autumn): " , " , " , " "
Portland (spring): " , " , wind speed, temperature
" (autumn):visibility, " , pressure , "
```

The consistency is by no means as striking as that seen in Table 3.3.1, and furthermore, if we were to reduce the dimensionality by discarding 3 or 4 more variables, then we would lose approximately 30% of the variation in the data.

Table 3.3.3 Eigen-vectors for reduced data sets, 1968.

Dungeness, Spring

Variable	Principal component						
	1	2	3	4	5	6	7
Wind dir.	.01	.29	.38	.84	.16	-.17	.09
Wind speed	.31	.49	-.37	-.01	-.51	-.50	.06
Low cloud	.51	.13	.40	-.07	-.38	.58	.28
Visibility	-.30	.59	-.29	.08	.01	.55	-.41
Air temp.	.52	-.42	-.62	.49	-.19	.28	.27
Humidity	.58	-.25	-.02	.15	.06	-.02	-.76
Pressure	-.45	-.26	.30	.12	-.73	-.05	-.30

Dungeness, Autumn

Variable	Principal component						
	1	2	3	4	5	6	7
Wind dir.	-.12	-.35	-.26	.87	-.20	.07	-.03
Wind speed	-.26	.59	.06	.12	-.56	-.49	-.08
Low cloud	.41	.30	-.46	-.11	-.42	.57	-.17
Visibility	-.61	.08	-.18	-.13	-.09	.36	.66
Air temp.	.03	.19	.80	.26	-.10	.49	-.00
Humidity	.61	.14	.03	.19	.02	-.23	.72
Pressure	.09	-.61	.22	-.31	-.67	-.08	.11

Portland, Spring

Variable	Principal component						
	1	2	3	4	5	6	7
Wind dir.	-.03	.00	-.89	-.38	.02	.23	.07
Wind speed	-.34	-.52	-.07	-.09	-.51	-.51	.28
Low cloud	-.55	-.06	.25	-.20	-.34	.65	-.23
Visibility	.27	-.51	-.08	.53	-.04	.50	.37
Air temp.	-.11	.50	-.29	.61	-.51	-.07	-.14
Humidity	-.54	.32	.04	.11	.35	.03	.69
Pressure	.45	.34	.21	-.37	-.49	.13	.48

Portland, Autumn

Variable	Principal component						
	1	2	3	4	5	6	7
Wind dir.	.09	.14	-.82	.51	.10	-.02	.14
Wind speed	-.24	-.52	.09	.47	-.54	-.33	-.18
Low cloud	-.56	.16	-.17	-.05	-.29	.67	-.32
Visibility	.16	-.61	-.31	-.29	.35	.07	-.55
Air temp.	.41	.21	.37	.57	.16	.28	-.47
Humidity	-.49	.38	.01	.02	.35	-.55	-.43
Pressure	.43	.35	-.22	-.32	-.59	-.23	-.37

Deciding how many variables to retain for future use is clearly very subjective. We have reached a stage, however, where no obvious decisions can be made, and since it is preferable to include too much information than too little in future analyses, no further steps to discard data are taken.

Interpreting the principal components is again difficult. From Table 3.3.3 we see that the first component is essentially a component featuring cloud cover and humidity. This may be indicative of rainfall, or at least whether the weather is overcast or clear.

The second component involves wind speed, and in three cases out of four, visibility. This component appears to be measuring weather on a synoptic scale, rather than just local conditions, since wind speed and visibility are important variables for identifying atmospheric conditions such as anticyclones and low pressure systems.

The third component shows, for the first time, a difference between the two sites. For Dungeness, the component involves temperature and cloud cover, while for Portland the component is dominated in both cases by wind direction.

Clearly there is much scope for further study of weather data in this manner, and no doubt meteorologists would be able to interpret the components in more detail than has been attempted here.

We have so far been concerned with analysing relationships between the variables, but we now turn to the analysis of the autocorrelation structure of the data. We have, in effect, a multivariate time-series. One usually, however, performs principal components analysis on a set of independent observations on a multivariate random variable. Our successive observations are obviously not independent as they are measured in a fixed order at daily intervals. It is not clear from the literature whether principal components analysis is appropriate for this type of data, and this topic is considered further in section 3.3.5.

Firstly though, we examine each variable in turn for its autocorrelation properties.

#### 3.3.4 Time series models and weather variables

By performing a time-series analysis of the weather variables, we hope to gain insight into the time dependency of variables which may later be linked with series of bird counts. Another objective of this study is to explore the effect of autocorrelation of individual variables on the principal components analysis described in the previous section.

The data used in this section are again the daily means of seven variables. Since we have so far seen little evidence of differences in weather data between Dungeness and Portland, and between spring and autumn, we restrict this analysis to one data set, that of the spring data for Dungeness.

We use only the daily means of the weather variables for two reasons. Firstly, diurnal variation within each variable will increase the complexity of the models required, and secondly we eventually want to compare results with those obtained from modelling sequences of bird counts which are also recorded on a daily basis.

There is a vast amount of literature available on the subject of fitting ARIMA type models to observed data. Box and Jenkins (1968, 1970) give an iterative procedure, but this depends upon a correct interpretation of sample autocorrelation and partial-autocorrelation functions, and involves a certain amount of subjectivity. Chatfield and Prothero (1973a) discuss the shortcomings of the Box-Jenkins procedure, and their paper together with subsequent replies by Box and Jenkins (1973) and Chatfield and Prothero (1973b) provide an extensive account of the problem of finding the optimum model for a given data set.

Ozaki (1977) gives a clear account of a relatively new approach to deciding on an optimal model, i.e. the MAICE procedure. This procedure was developed by Akaike (1972, 1973), and as Akaike shows, gives similar results to the Box-Jenkins procedure. The advantage of this method is that it is objective, the criterion for choosing a particular model being based on Akaike's Information Criterion (AIC). Several models are fitted to the data, and the model giving the minimum value of AIC is chosen to be the optimum. The AIC takes into account the

likelihood as well as the number of parameters in the model. One generally needs to estimate parameters more frequently than with Box-Jenkin's procedure, but with improved computing facilities, this is becoming less of a problem.

Very often one finds that models involving only one or two parameters are adequate to explain the data. For each variable studied here, we try five models for each level of differencing of the data. The models are denoted "(p,d,q)", where p represents the order of the autoregressive part of the model, d represents the level of differencing and q the order of the moving average part of the model. These models are described in most text books on time series analysis, see for example, Box and Jenkins (1970). In most practical situations one generally finds that d rarely needs to be greater than 1, and so for each variable we shall only look at 10 2 parameter models:

$$(1,0,0), (0,0,1), (1,0,1), (2,0,0), (0,0,2)$$
$$(1,1,0), (0,1,1), (1,1,1), (2,1,0), (0,1,2)$$

These models are usually represented algebraically by the equation:  $\phi_p(B)\nabla X_t = \Phi_q(B)a_t$ , where

B = backward difference operator

$\nabla$  = forward difference operator

$X_t$  = observation

$a_t$  = random component (independently distributed Normal r.v.'s)

and  $\phi_p$  and  $\Phi_q$  are polynomials of order p and q respectively.

Six of the seven variables were able to be explained by a model of this type, but atmospheric pressure requires the more complex Box-Jenkins seasonal model.

Even when using the MAICE procedure, it is wise to inspect the results for each model and if, say, two models give close values of the AIC, then one can opt for the model with fewer parameters in the interests of simplicity. Furthermore, for each model, we perform the 'portmanteau' test of the residual autocorrelation function in order to assess whether or not the 'best' model does give a good fit to the data. The statistic calculated here is  $n \sum r_i^2$  where  $r_i$  is the  $i^{\text{th}}$  autocorrelation coefficient of the residuals,  $k$  is the number of lags used (usually about 20) and  $n$  is the number of observations. This statistic follows a chi-squared distribution with  $k-p-q$  degrees of freedom under the null hypothesis of random residuals.

We now consider each of the several variables in turn.

#### 1. Wind direction

We begin each analysis by inspecting the sample autocorrelation coefficients, in the form of a correlogram. Fig. 3.3.6 shows the correlogram for the wind direction data, and an unusual pattern is seen where the a.c.f. attains a low value (only just significant at 5% level) at lag 2, but then increases again at lags 3 and 4 before eventually dying out.

Fig. 3.3.6 Correlogram for wind direction data

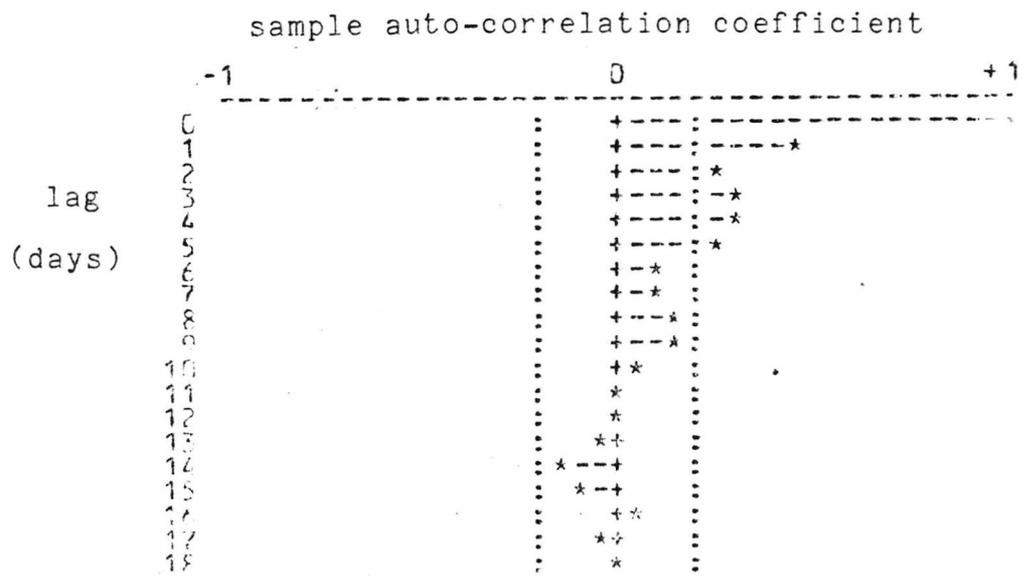
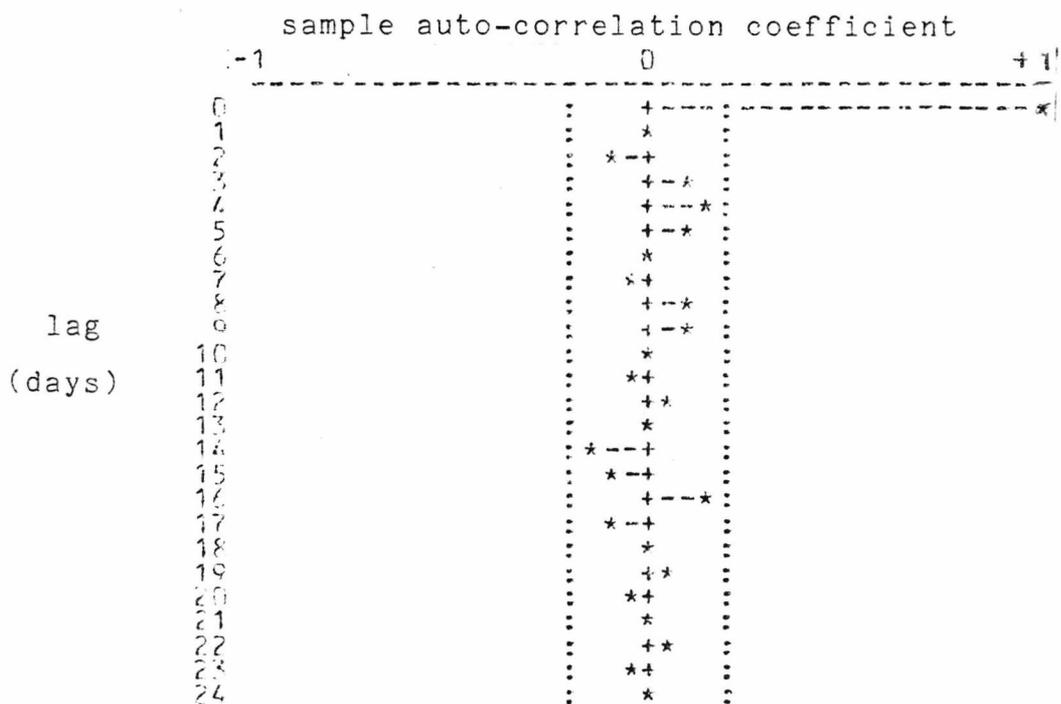


Fig. 3.3.7 Correlogram for residuals - wind direction data



It is not clear from Fig. 3.3.6 whether trend or cyclic variation is present or not, and so we investigate all 10 models. Table 3.3.4 lists the AIC values with the portmanteau test statistics.

Table 3.3.4

<u>Model</u>	<u>AIC</u>	<u><math>n \sum r^2</math></u>	<u>dF</u>
1,1,0	-202.7	43.9	23
2,1,0	-218.5	25.0	22
0,1,1	-223.4	29.8	23
0,1,2	-229.4	22.6	22
1,1,1	-227.6	24.8	22
1,0,0	-232.5	23.9	24
2,0,0	-230.5	24.0	23
0,0,1	-229.3	30.5	24
0,0,2	-229.0	26.5	23
1,0,1	-230.7	27.5	23

The model giving the lowest AIC value is (1,0,0) and the residual a.c.f. test is non-significant at the 5% level. We conclude that a first order autoregressive model is suitable for these data, where the residuals are given by  $Z_t = X_t - \cdot 12 X_{t-1} - \cdot 12$ . The correlogram for the residuals is shown in Fig. 3.3.7, and the portmanteau test suggests, no evidence is seen to suggest that the residuals are other than white noise.

## 2. Wind speed

The correlogram for wind speed shows high coefficients at lags 1 and 2, but at higher lags all coefficients are very close to zero. Again, all 10 models were tried, and again the lowest AIC value was achieved by model (1,0,0). The test statistic for the residual

a.c.f. was 9.59 on 23 degrees of freedom, suggesting a good fit. The residuals are given by

$$Z_t = X_t - 0.61 X_{t-1} - 4.77.$$

### 3. Low cloud

Fig. 3.3.8 shows the correlogram for the observations, and one can see a suggestion of cyclic variation, with a period of about 10 days. One must be careful in interpreting such diagrams though, since even by pure chance, one would expect approximately one in twenty coefficients to appear significant at the 5% level even when no such pattern existed.

Table 3.3.5 shows the AIC values, and the portmanteau test statistics for the 10 models. Model (0,0,2) gives both the minimum AIC and the minimum residual test statistic, although the statistic is just significant at the 5% level (not at the 2½% level).

If we were to fit a more complicated model, i.e. one with seasonal components, a better fit may be obtained, although more parameters would be involved. The residuals from this second order moving average model are given by  $Z_t = X_t - 0.31 Z_{t-1} - 0.3 Z_{t-2} - 3.75$ .

### 4. Visibility

The correlogram of the data looks much like that for wind speed, i.e. dying out very quickly. However the optimum model appears to be a first order mixed model,

Fig. 3.3.8 Correlogram for low cloud data

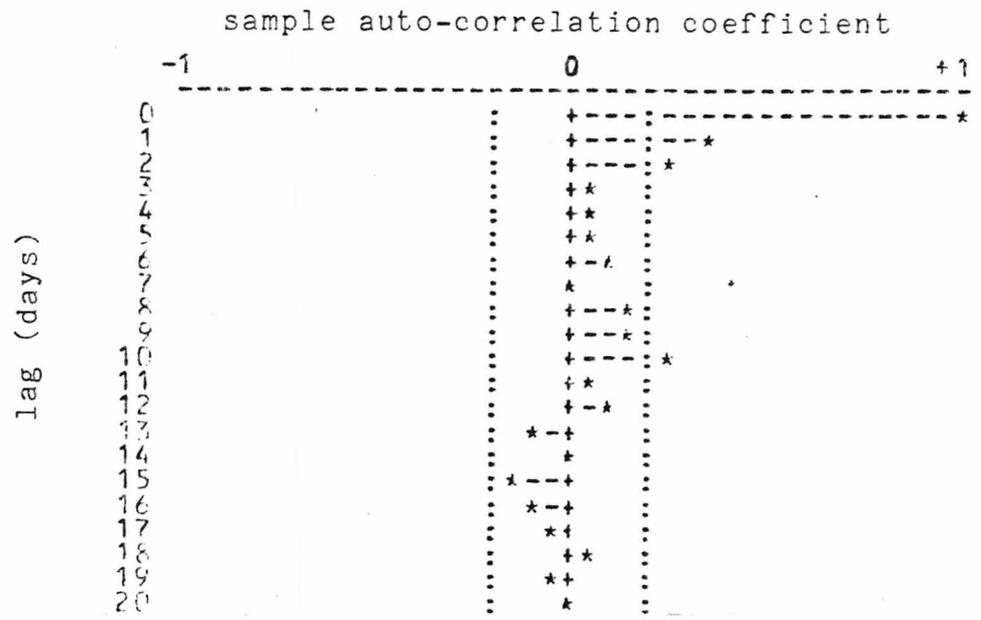
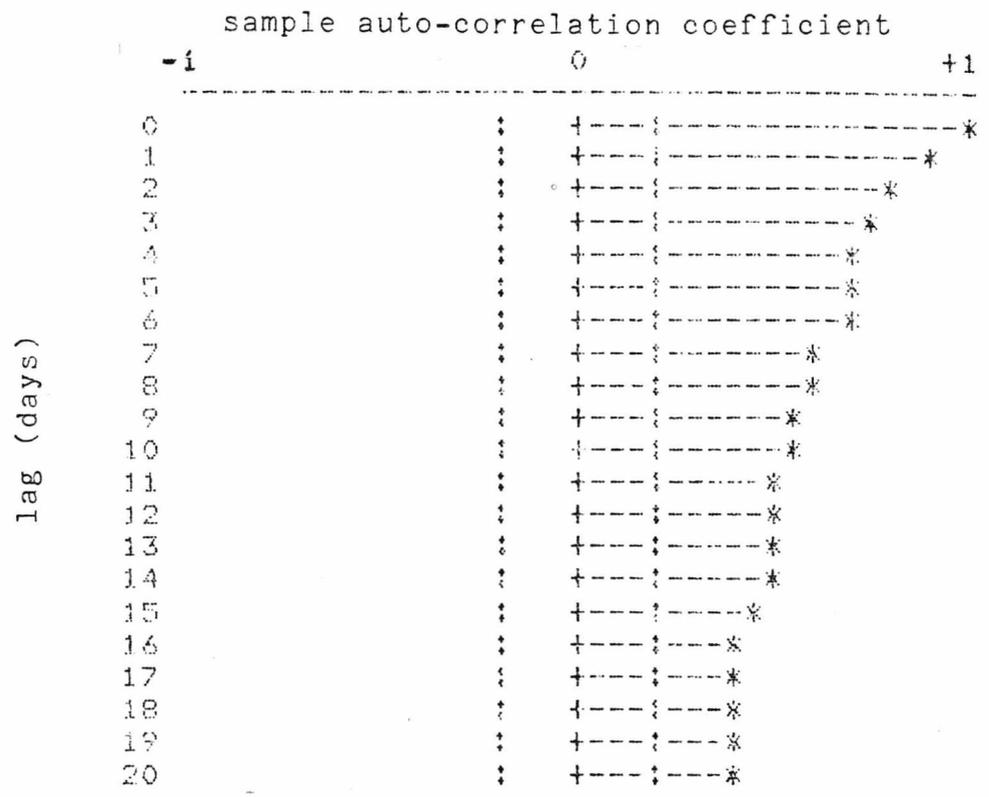


Fig. 3.3.9 Correlogram for air temperature data



i.e. (1,1,1). The residual test statistic is non-significant, and the residuals are given by

$$Z_t = X_t - 1.56 X_{t-1} + .56 X_{t-2} + .99 Z_{t-1}.$$

Table 3.3.5      Low cloud data

<u>Model</u>	<u>AIC</u>	<u><math>n\sum r^2</math></u>	<u>dF</u>
(1,1,0)	214.4	39.0	23
(2,1,0)	215.6	42.1	22
(0,1,1)	201.3	43.3	23
(0,1,2)	199.7	34.6	22
(1,1,1)	198.5	38.4	22
(1,0,0)	198.1	47.8	23
(2,0,0)	197.4	37.3	22
(0,0,1)	203.4	57.1	23
(0,0,2)	196.6	34.7	22
(1,0,1)	198.2	41.4	22

5. Air temperature

The correlogram of the data (Fig. 3.3.9) shows a typical data set that is dominated by trend, just as we might expect for such a variable. Even at lag 20 the autocorrelation coefficient is highly significant. Fig. 3.3.10 shows the correlogram for the differenced data set, and clearly the trend has been removed. Of the five models fitted to the data, the second order autoregressive model gave the lowest AIC value and also the lowest residual test statistic. The residuals are given by  $Z_t = X_t - .87 X_{t-1} + .14 X_{t-2} - .27 X_{t-3}$ .

6. Humidity

Fig. 3.3.11(a) and (b) show the correlograms for the raw data and differenced data. The cyclic variation appears to be removed by differencing, and of the five

Fig. 3.3.10 Correlogram for air temperature - differenced data

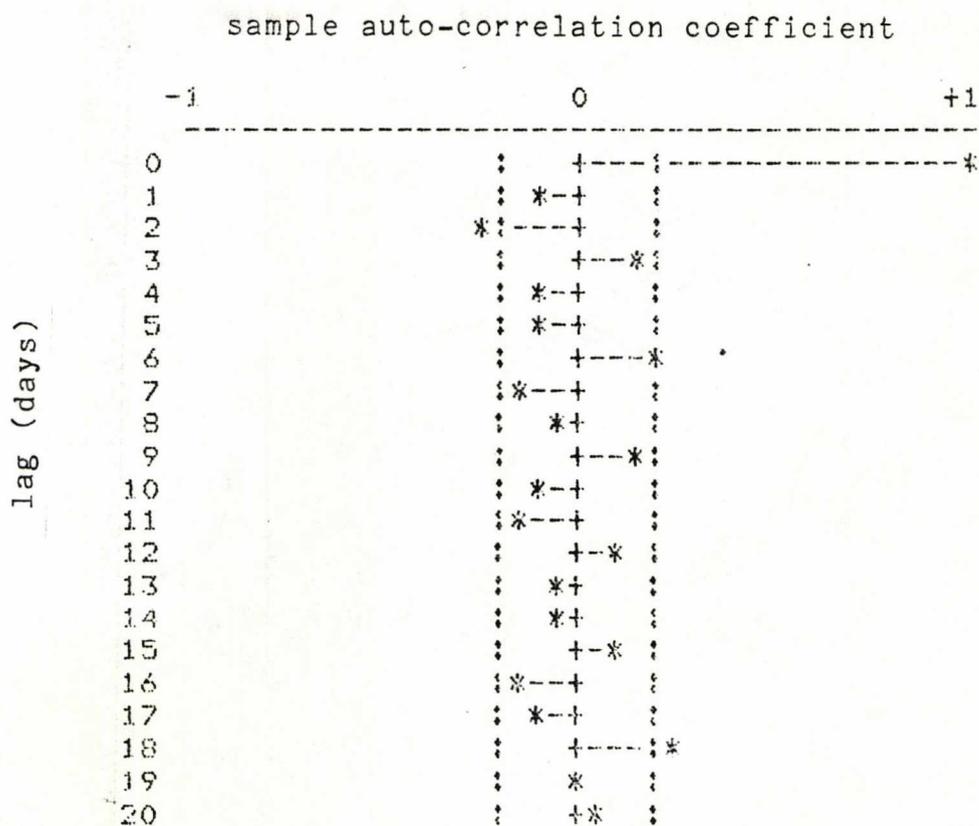
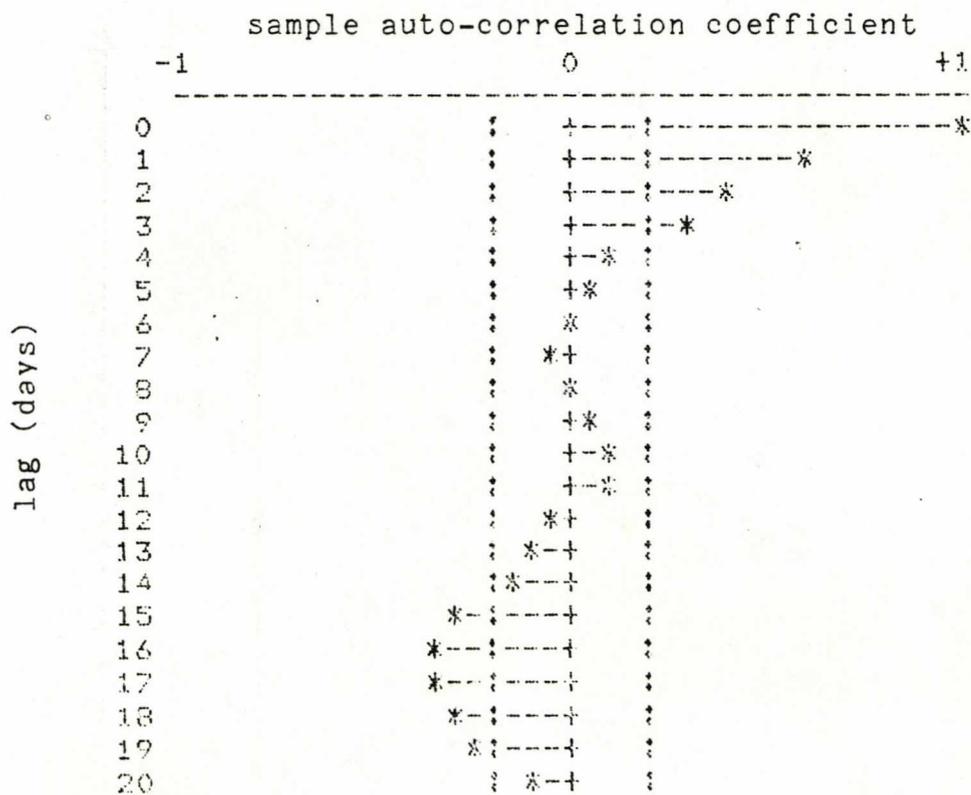
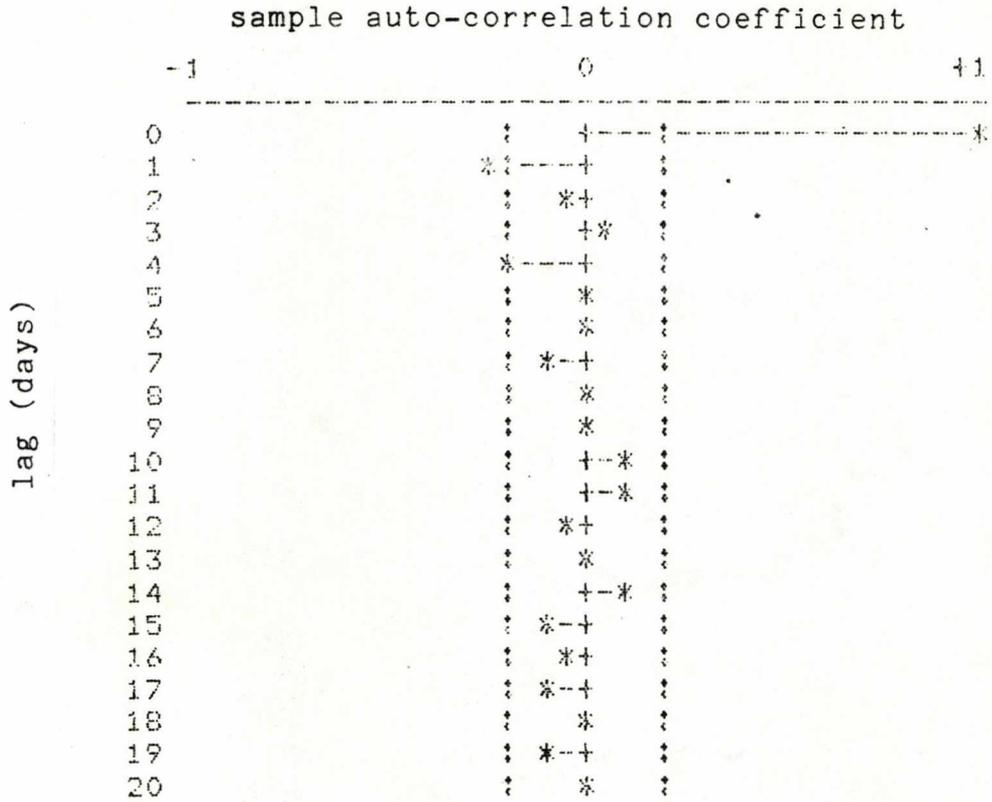


Fig. 3.3.11 Humidity correlograms

(a) Raw data



(b) Differenced data



models fitted to the data, the mixed model (1,1,1) was found to be the best in terms of fit. The residuals are given by  $Z_t = 1.61 X_{t-1} + .61 X_{t-2} + .98 Z_{t-1}$ .

7. Atmospheric pressure

Fig. 3.3.12 shows the correlogram for the pressure data, and clearly one can see that cyclic variation is present. Unlike the case of humidity, differencing the data doesn't remove this variation as Fig. 3.3.13 shows. Denoting  $X_t - X_{t-1}$  by  $\nabla X_t$ , we next examine the effect of taking the transformation  $\nabla\nabla_{14} X_t$ , since the period of the cycle appears to be 14 days. As Fig. 3.3.14 shows, the series  $W_t = \nabla\nabla_{14} X_t$  still shows evidence of a cycle of period 14 days. In light of this we shall fit a seasonal model to the data. The model will be of the form:

$$\phi_p(B) \Phi_p(B_{14}) W_t = \theta_q(B) \Theta_Q(B_{14}) a_t$$

Where the  $a_t$  are the residuals, and the polynomials are the standard form for Box-Jenkins models. It is straightforward to fit models of this type using the Minitab statistical package (Ryan, et al 1981), and 16 models were examined ( $p, q = 0$  or  $1, P, Q = 0$  or  $1$ ). The model corresponding to the minimum AIC value was found to be a seasonal ARIMA (1,1) model, with  $p = q = 0, P = Q = 1$ . The parameters were estimated to be  $\phi = - .99$  and  $\Theta = - .81$ . This model can thus be written:

$$(1 - \Phi B_{14}) W_t = (1 - \Theta B_{14}) a_t$$

or  $1.99W_t - .99W_{t-14} = 1.81a_t - .81a_{t-14}$

Fig. 3.3.12 Correlogram for Pressure data (raw data)

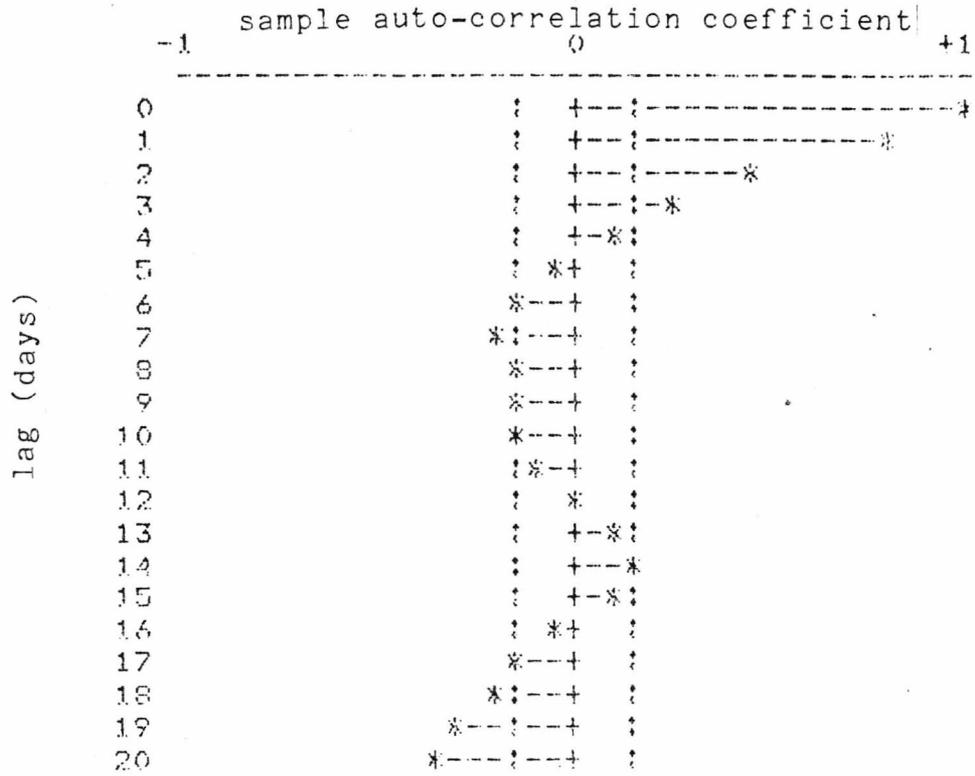
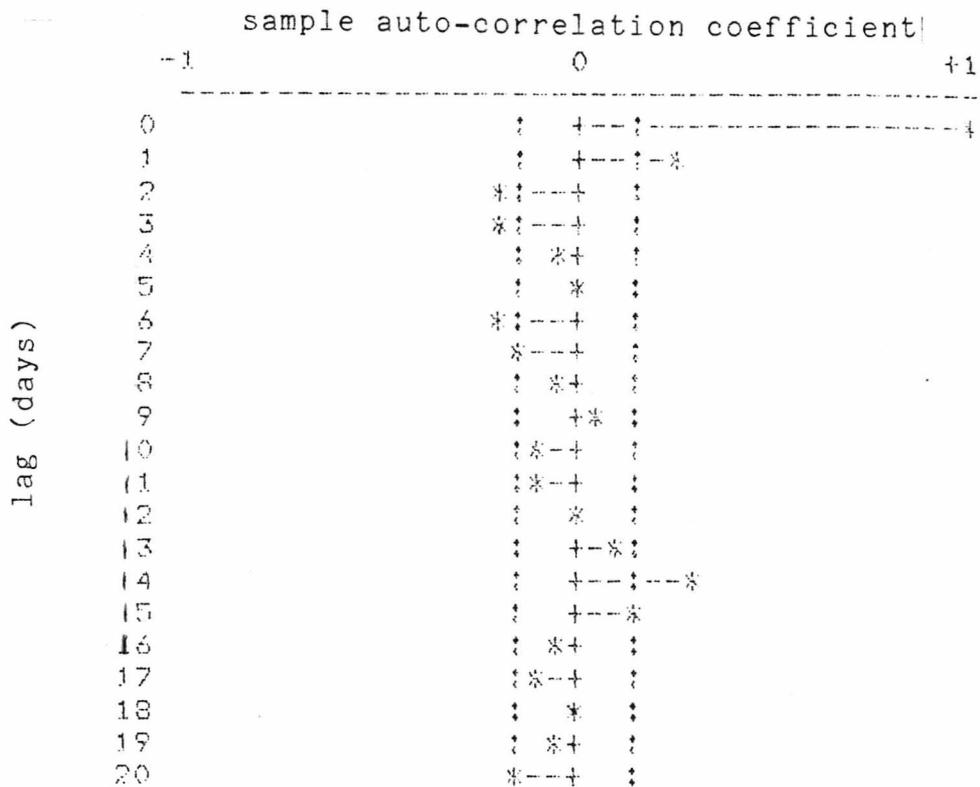


Fig. 3.3.13 Correlogram for Pressure data (differenced data)



The above model appears to give a good fit to the data since the chi-square statistic for the residual autocorrelation coefficients is not significant at the 5% level (the statistic is 24.88, on 18 degrees of freedom). Fig. 3.3.15 shows how the expected values of the time series correspond with the observations. There is clearly a close correspondence between the two series, save for a small lag between the series, which is to be expected due to the autoregressive component of the model.

We have seen how simple time series models of the Box-Jenkins type can be fitted to series of daily means of various weather variables. For all but one variable, the next observation can be predicted by reference to

Fig. 3.3.14 Correlogram for Pressure (  $\nabla \nabla_{44} X_t$  )

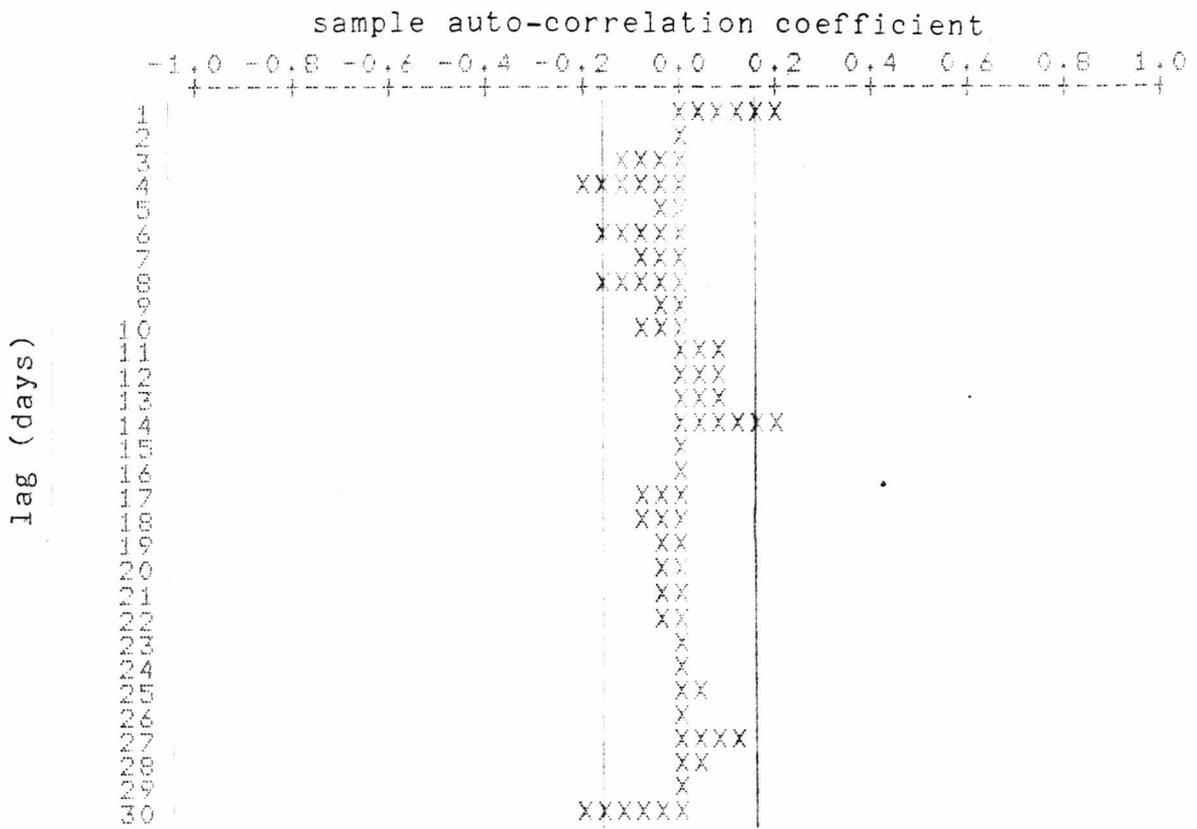
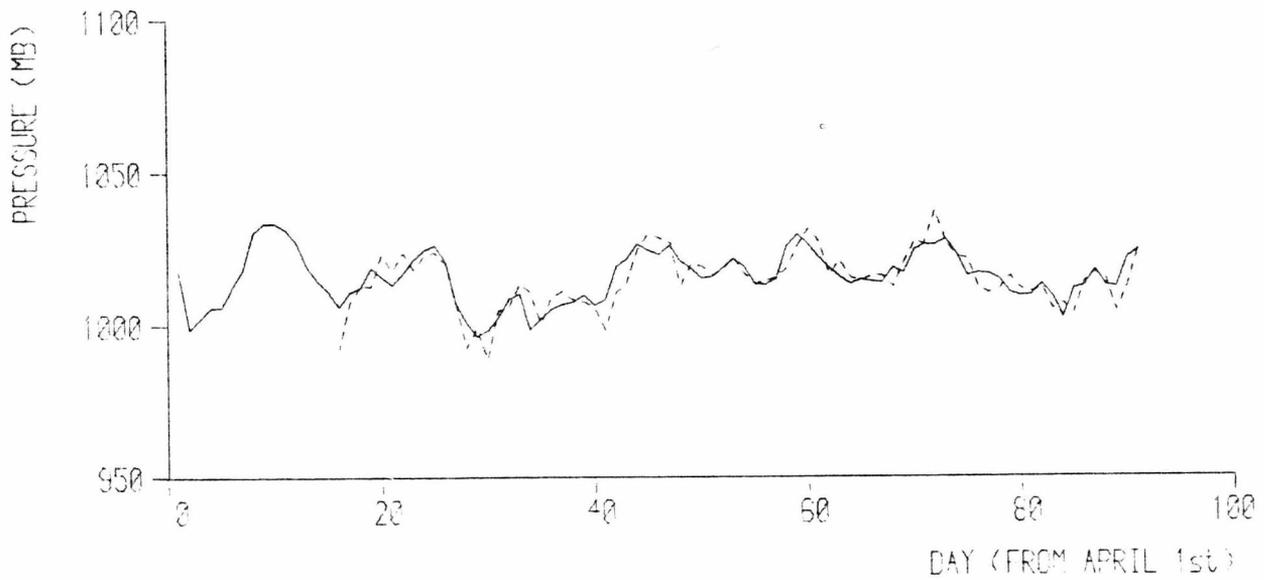


Fig. 3.3.15 Atmospheric pressure data



— observed values

- - - expected values

the observation or residual (or both) of up to three days before. For pressure, one needs to refer further into the past, in order to account for the seasonal variation.

Since it is so straightforward to calculate the residuals, we can readily investigate the effect on the principal components analysis of section 3.3.3, of time dependency. By definition, the residuals calculated in this section represent random fluctuations of the weather variables about some mean value. It is likely that such fluctuations represent the conditions that influence birds and their occurrence at observatories, and thus a further examination of the data in this manner may prove useful.

#### 3.3.5 Principal component analysis of detrended weather data

When calculating a sample correlation coefficient between two series of observations, serial correlation within the series can have the effect of inflating the correlation coefficient. Long term trend in the series, such as that observed in the air temperature data, itself causes serial correlation, and may thus distort the correlation matrix of the seven weather variables studied in previous sections.

In order to examine the effect of autocorrelation on the correlation structure of a set of weather variables, we can fit time series models and examine the residuals. In this section we perform a principal components analysis on

the residuals obtained from fitting the models described in the previous section. Clearly, physical interpretation of these series of observations is difficult, although broadly speaking, each residual is simply the observed value with some adjustment for recent observations.

Fig. 3.3.16 gives the correlation matrix for the seven sets of residuals, and by testing this against the correlation matrix of the data (Fig. 3.3.5), one finds that the two matrices are not significantly different.

<u>Fig. 3.3.16</u>	<u>Correlation matrix of residuals</u>						
Wind direction	1.00						
Wind speed	.07	1.00					
Low cloud	.14	.37	1.00				
Visibility	.15	.03	-.22	1.00			
Air temp.	-.07	-.08	-.20	-.07	1.00		
Humidity	-.01	.15	.39	-.45	.28	1.00	
Pressure	.07	-.20	-.20	.19	-.06	-.14	1.00

This suggests that the observed correlations between the variables can be treated as though they are free of effects of autocorrelation. We would expect that the principal components evaluated from the residuals would be similar to those from the data, since the components are simply the eigen-values of the correlation matrices. The 'residual' components are shown in Table 3.3.5, together with the percentage variation accounted for by each component.

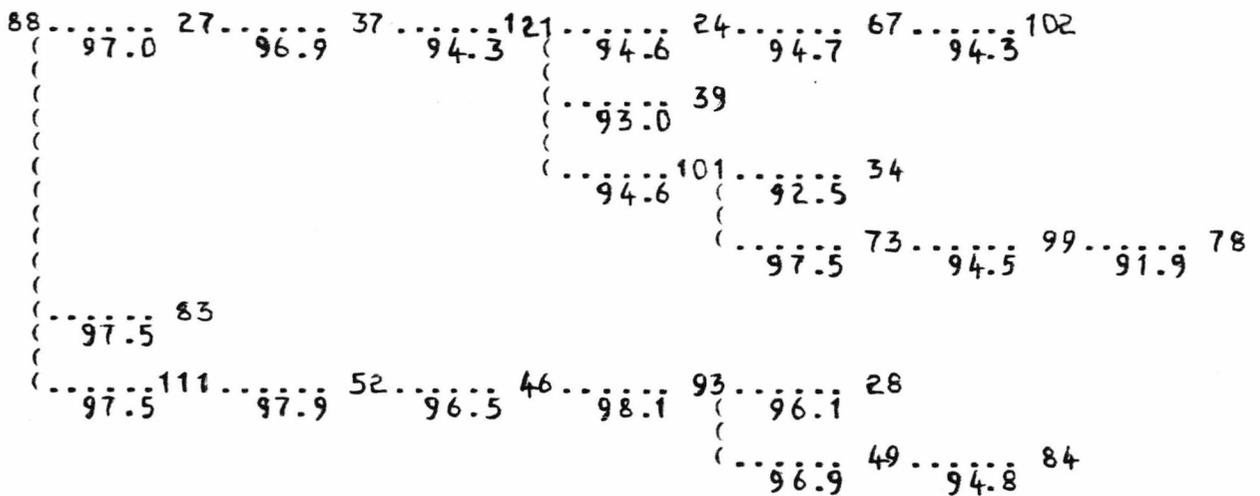
By comparing these components with those of Table 3.3.3, we can see that the eigen vectors are mostly unchanged.

Before leaving this topic, we examine the minimum spanning tree for the set of residuals. Fig. 3.3.17 shows a section

Table 3.3.5 Principal components of residuals

Weather variable	Principal Component						
	1	2	3	4	5	6	7
Wind direction	-.03	-.62	.08	.68	.34	.14	.07
Wind speed	-.33	-.24	.58	-.21	-.43	.38	.35
Low Cloud	-.54	-.26	.09	-.02	-.14	-.76	-.20
Visibility	.40	-.51	.19	-.45	.04	.07	-.58
Air temp.	.19	.40	.59	.48	-.21	-.07	-.42
Humidity	-.55	.06	-.33	.09	-.11	.50	-.56
Pressure	.32	-.24	-.40	.25	-.78	-.06	.04
% variance accounted for	28.3	18.3	17.3	12.1	11.0	7.3	5.7

Fig. 3.3.17 A section of the Minimum Spanning tree for residuals.



of the tree, and unlike the tree shown in Fig. 3.3.4, we see that the observations that are adjacent to one another are not necessarily close together in time. This feature emphasises the random nature of the residuals, and suggests once more that the simple time series models described in the previous section are suitable for representing weather variables. Furthermore, it seems that, from the principal components themselves, there is little difference in character between the original data components and those from the residuals, suggesting that correlation structures are similar. This is important when we come to consider multiple regression models and linear discriminant analyses that are reported in later chapters. It appears that we may go ahead and perform such analyses using the weather data in their original form, and not worry too much about autocorrelation.

## CHAPTER 4

### RELATING MIGRATION DATA TO WEATHER (I) REGRESSION

#### 4.1 Introduction

In this chapter the use of multiple regression as a means of including weather information in the construction of migration indices is considered. Hussell (1981) adopted this approach where the set of explanatory variables included a number of weather variables and a set of dummy variables which related to year effects. His study is reviewed in section 4.5.

Regression studies of this sort have been undertaken in the past, but all of these have used radar data rather than counts of grounded migrants, and most of these studies were undertaken outside Great Britain.

The work of this chapter represents the first attempt to examine the connection between weather and British observatories data in an objective manner.

There is little doubt among ornithologists that the occurrence of migrants at observatories is related to weather conditions.

Ornithologists are familiar with the basic rules, for example, wet and cloudy nights generally produce large numbers of grounded migrants on the following day, whereas clear calm nights enable the migrants to pass overhead.

The weather data used in this chapter consist of daily means of seven variables - see section 3.3. We also consider the same variables,

lagged by one day, giving a total of fourteen variables. The use of lagged variables is sensible for two reasons. First, the weather that determines whether or not a bird will land at an observatory may be the weather experienced prior to arrival. This point is illustrated by the following examples of situations which may occur:

- (i) In spring, bad weather over France may prevent birds from setting off, thus leading to an absence of birds at British south coast observatories irrespective of the weather conditions when the birds would have arrived.
- (ii) If birds set off from France in spring, and fly sufficiently high, they may fly over any bad weather when they arrive at the south coast of Britain.
- (iii) In autumn, bad weather over Britain may prevent birds from leaving their breeding grounds until conditions improve. The occurrence of birds at observatories will therefore be related to weather at the time of departure as well as the time of arrival at the south coast.

In the situations described above, we would ideally like weather data from sites other than observatories as well as at observatories themselves. To meet this requirement we make the assumption that, to a greater or lesser extent, the weather data obtained from observatories are indicative of the general synoptic situation. For variables such as temperature and pressure, this will be the case. For variables such as wind direction and cloud cover, local variations are bound to occur, particularly as both Dungeness and Portland observatories are situated on peninsulas which are distinct from the surrounding coastline.

The second justification for using lagged variables is that we allow for changes in recorded values, i.e. a value for pressure may not be important but a drop in pressure from one day to the next may be.

A closely related and widely published area of research is that of radar studies and these have been briefly described in section 2.6. Radar studies involve very different problems from those that face observatory workers. At observatories, birds are recorded on the ground whereas radar records birds in flight. The influential weather variables, if any, will clearly be different. Further, it is difficult to distinguish between species by using radar echoes which means that inferences can only be made on groups of species.

Although there is little scope for comparing radar results with the results presented in this chapter, it is worth noting that the most popular statistical technique has been multiple regression analysis. Examples include Lack (1960, 1963a, b), Nisbet and Drury (1968), Able (1973), Alerstam et al (1973) and Richardson (1974).

The main aims of this chapter are to determine which weather variables, if any, are related to counts of grounded migrants and to assess the validity of regression models in this context.

## 4.2 Preparation of the data

### 4.2.1 Introduction

The data used for the regression analyses of this chapter

consist of bird counts, used as the response variable, and weather data, used as explanatory variables. The weather variables which are used were selected on the basis of information found in chapter 3.

The time period covered by this study and the observatories considered were determined by the availability of weather data: at Dungeness from 1961 to 1972 and 1967 to 1970 for spring and autumn respectively, and from 1968 to 1970 at Portland for both spring and autumn.

#### 4.2.2 The bird data

The response variable for the regression analyses is the number of grounded migrants of a particular species seen on the recording area in one day.

We restrict our attention to four species: Redstart, Sedge Warbler, Whitethroat and Willow Warbler. These were selected since they are among the commonest of migrants occurring at both Dungeness and Portland.

In most of the regression studies that have previously appeared in the literature, the response variable has been transformed, usually by a logarithmic transformation. This is done as an attempt to make the distribution of counts less skewed.

Authors took account of seasonal variation in counts by including 'date' variables in the explanatory variable set

which enables polynomials to be fitted to the counts. In this chapter we take a different, more straightforward approach which involves removing the trend from the counts over each season and then using detrended counts as the response variable. Although not entirely objective, this does overcome the problem of deciding what degree of polynomial to use - we are still left with the problem of how to detrend the counts. One advantage of this approach is that we are more likely to obtain a set of responses that are normally distributed since a high proportion of a typical season's counts are very low, due mainly to the two 'ends' of the season. Further, the distribution of raw counts is truncated at the lower end of the scale.

By detrending the counts we hope to meet the assumptions required for a multiple regression analysis. It is doubtful whether such assumptions are met in the published examples of work in this area, a point also made by Richardson (1974) and often one finds little or no evidence of the assumptions having been checked.

#### 4.2.3 Removing trend from daily counts

As many of the examples of chapter 3 have shown, the number of birds recorded each day varies according to the time of season. Since we are looking for the influence of weather on a day-to-day basis, we must first remove the underlying trend from the data. There are two approaches that we could take here:

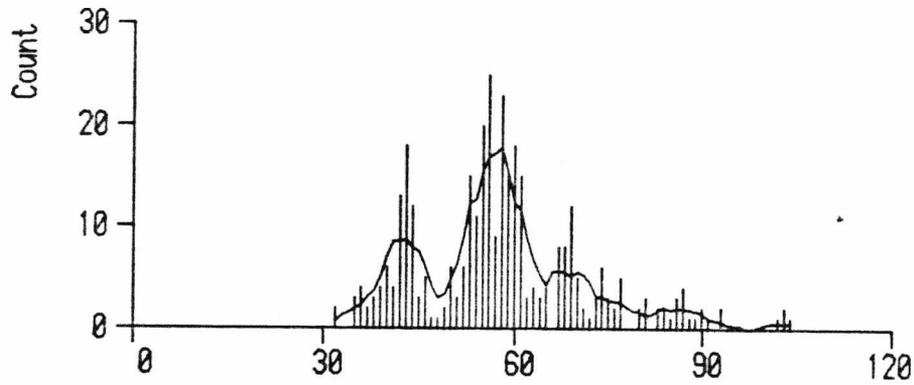
- (i) We could fit a time-series model, for example a Box-Jenkins model, to each series of counts and base the regression analysis on the residuals. A drawback to this method is that we often have rather short series, sometimes only fifteen days.
- (ii) A simple approach related to (i) is to use the deviations of the observed counts from a moving average. This approach is easily applied and with a suitably chosen moving average, the pattern of each individual season should be removed. Although an element of subjectivity is introduced here, in section 3.2.4 we found that a seven-day moving average produced a curve that broadly summarises the pattern of migration and we shall use such a moving average here.

A useful feature of (ii) is that all series of counts can be treated in the same manner. In other words, a series that initially shows no trend may be 'detrended' without distorting this fact - we will still pick up day-to-day variations in migration.

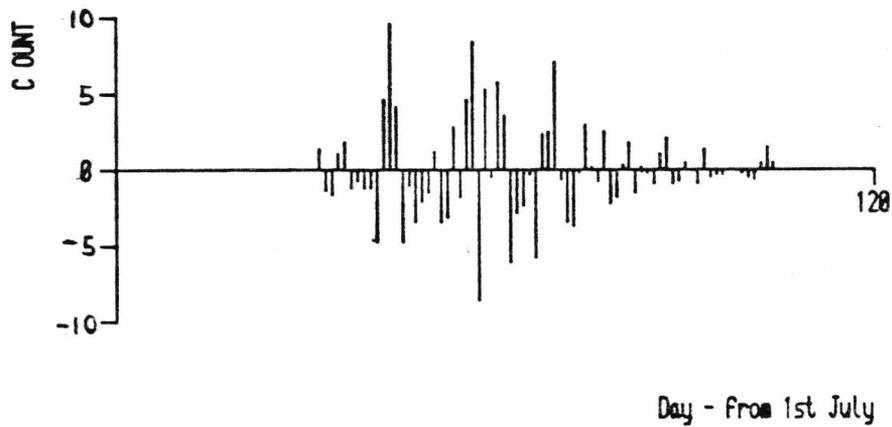
This method is also independent of the timing of the migration season as a whole, which may vary from year to year - according to climatic factors in the birds' wintering areas for example. Fig. 4.2.1 shows an example of the detrending process. The original counts and the moving average curve are shown in Fig. 4.2.1(a) and the deviations from this curve are shown in Fig. 4.2.1(b) - note the difference in

Fig. 4.2.1 Example of detrending a series of daily counts  
Portland ( 1970 ) Whitethroat counts

(a) Daily counts and Moving Average



(b) Detrended data



vertical scale. It is deviations such as these which we use in the regression analyses. Even with the detrending method described above, some element of seasonality is still present in the deviations, however we have been able to highlight day-to-day changes from the seasonal norm of counts in both positive and negative directions.

#### 4.2.4 The weather data

We use seven weather variables as the basic set of explanatory variables. These are described in detail in section 3.3, but are summarised again below.

1. Wind direction.
2. Wind speed
3. Low cloud
4. Visibility
5. Air temperature
6. Humidity
7. Atmospheric pressure

Our observations on the above variables are daily means, which are derived from 3-hourly records.

These variables were selected from a larger set on the basis of their correlation structure (see section 3.3). The seven variables tend to exhibit low correlations between one another which is a desirable property when performing multiple regression. If two or more variables are highly correlated, numerical problems may arise when

the covariance matrix is inverted.

Also, if we use a set of independent variables, we minimise the amount of work involved, i.e. we do not use two variables that are highly correlated where one will do satisfactorily.

### 4.3 The regression procedure

#### 4.3.1 Methodology

In the following sections of this chapter we report the results of various attempts at fitting linear regression models to the data. We briefly describe here the approach taken which has led to the later results.

The preparation of the data for analysis has been described in section 4.2. Initially we consider the model containing the set of fourteen variables. If a suitable fit is obtained then variable selection will be of interest.

If the full model shows a poor fit then the results are likely to yield information as to why this is so. For example, an examination of residuals may point to a transformation being necessary. One may also wish to try including interactions of the explanatory variables or power transformations, for example; an explanatory variable may need to be raised to some power before contributing significantly to the regression.

If one arrives at a model which appears to fit the data, one must check the assumptions on which the model is based. In the case of linear regression we need to check for normality of the residuals and responses, stability of the variance and also for the presence of outliers.

The above points are returned to in the following sections where the results are presented. First, however, we briefly outline the computer package used.

#### 4.3.2 The computer package used

The BMDP library of programs (Dixon, 1981) was used to obtain the results of this chapter and also the next chapter where discriminant analysis is studied. Several factors led to this choice, not least the local advantages : one can very easily run BMDP programs on line at the University of Kent and where such runs take a large amount of central-processor time, the work can be done in the 'batch' system. The package is not interactive, but the EMAS operating system at Kent enables alterations to BMDP commands files to be made with the minimum of work. Further, unlike the package MINITAB, large data sets can be analysed by BMDP.

As regards the facilities within the regression programs, BMDP produces a comprehensive array of statistics, tests and plots from a very limited and simple set of commands. Some of these statistics are referred to in the following sections.

The full regression models were fitted using the program P9R, which also considers, if required, all possible subsets of the explanatory variables. For variable selection, the program P1R is available with a number of possible stepwise strategies.

#### 4.4 Results for British observatories' data

##### 4.4.1 Regression on all weather variables

Throughout this section we consider sixteen data sets - four species recorded at two observatories and over two seasons (spring and autumn).

Goodness of fit is assessed by using the quantity  $R^2$  which can be viewed either as the proportion of the total variation accounted for by the model fitted, or the square of multiple correlation coefficient of the explanatory variables with the response variable.

The ratio of the regression mean square to the error mean square is also useful since a significant value indicates that the data suggest a linear relationship. However, in our situation, where we have a large number of observations for each analysis, we may well find a significant ratio associated with a low value of  $R^2$ . This is not a contradiction, but merely suggests that some other, more complicated, model is necessary to explain the data.

Table 4.4.1.1 shows the values of  $R^2$  and F (ratio of regression to error mean square) for the sixteen data sets,

using the full set of 14 weather variables. Recall that the response variable is the deviation of daily count from a moving average.

Table 4.4.1.1      Regression on 14 variables

	SPRING				AUTUMN			
	n	R <sup>2</sup>	F	P <sub>F</sub>	n	R <sup>2</sup>	F	P <sub>F</sub>
Sedge W (D)	462	.05	1.74	.05	325	.05	1.26	.24
Willow W (D)	613	.05	2.41	.00	345	.06	1.39	.15
Whitethroat (D)	526	.05	1.87	.03	329	.06	1.55	.09
Redstart (D)	498	.07	2.59	.00	343	.05	1.23	.25
Sedge W (P)	123	.13	1.12	.35	171	.10	1.24	.25
Willow W (P)	131	.07	.63	.83	169	.12	1.50	.12
Whitethroat (P)	132	.11	1.08	.38	197	.10	1.45	.13
Redstart (P)	126	.10	.89	.58	222	.05	.82	.65

n = no. of observations

D = Dungeness

P<sub>F</sub> = significance of F

P = Portland

We see from Table 4.4.1.1 that, apart from the Dungeness spring data, there is no evidence that a linear function of weather variables is able to explain the response variable. The Dungeness spring data are based on very high numbers of observations and so one might expect a significant F-ratio. In all 16 cases, the value of R<sup>2</sup> is extremely low which suggests that very little of the total variation is being accounted for by the linear model.

Clearly there is little point in looking for subsets of variables with practical interpretations at this stage. Instead, we consider alternative models.

An examination of the coefficients of skewness and kurtosis

of the 14 explanatory variables and the response variable gives strong evidence to suggest that the response variable requires transformation, while the other variables appear to be distributed normally. This is the case in all 16 analyses, and the noticeable feature is that the response variable has a highly skew distribution (to the right), and also tails which are much longer than the normal distribution. This is despite our attempt to overcome non-normality of the counts by detrending them.

A brief examination of the residuals also supports the case for a transformation. The frequency distribution of the residuals exhibits long tails in every case, and normal probability plots of the residuals suggest that apart from the extreme values, the residuals are more or less normally distributed.

Unfortunately it is doubtful that transforming the response variable alone will have such a marked effect on the regression that  $R^2$  rises to an acceptable level. We might also try, at this stage, interactions of weather variables and weather variables raised to the powers of 2 and 3 and so on. This is rather difficult with 14 explanatory variables since there are 91 two-variable interactions alone. We shall thus consider only interactions of two variables with zero lag, i.e. 21 pairs. Further, we shall only consider the weather variables raised to the powers of 2 and 3, and again with lag zero. Thus we now have a set of 49 explanatory variables.

The transformation is chosen by the Box-Cox method (Box and Cox, 1964). This approach involves fitting the regression model using a range of transformations of the form:

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1) / \lambda & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases}$$

and choosing  $\lambda$  such that the likelihood of the data is maximised. For a given  $\lambda$ , this maximum likelihood is given by:

$$L_{\max}(\lambda) = -\frac{n}{2} \log S_{\min}^{(\lambda)} + (\lambda-1) \sum_{i=1}^n \log y_i ,$$

where  $S_{\min}^{(\lambda)}$  is the minimised sum of squares. One can plot values of  $L_{\max}(\lambda)$  against  $\lambda$  and choose  $\lambda$  such that  $L_{\max}$  is a maximum. Fig. 4.4.1.1 shows an example of such a plot, and the appropriate value of  $\lambda$  to use is approximately 0.5. This value may not be the optimum value, but there is clearly little change in  $L_{\max}$  over a range of values of  $\lambda$  between 0 and 1.

As a guide to how useful the Box-Cox transformation is in this context, Table 4.4.1.2 gives the coefficients of skewness and kurtosis for the response variable in all sixteen cases, both before and after transformation.

The estimated standard errors of these sample coefficients are given by  $(6/n)^{\frac{1}{2}}$  and  $(24/n)^{\frac{1}{2}}$  for skewness and kurtosis respectively. All the coefficients listed in Table 4.4.1.2 are greater than three standard deviations from the mean (zero), and thus this variable appears to be highly skew

Fig. 4.4.1.1 Choosing a transformation : an example using Sedge Warbler data from Dungeness ( autumn counts ).

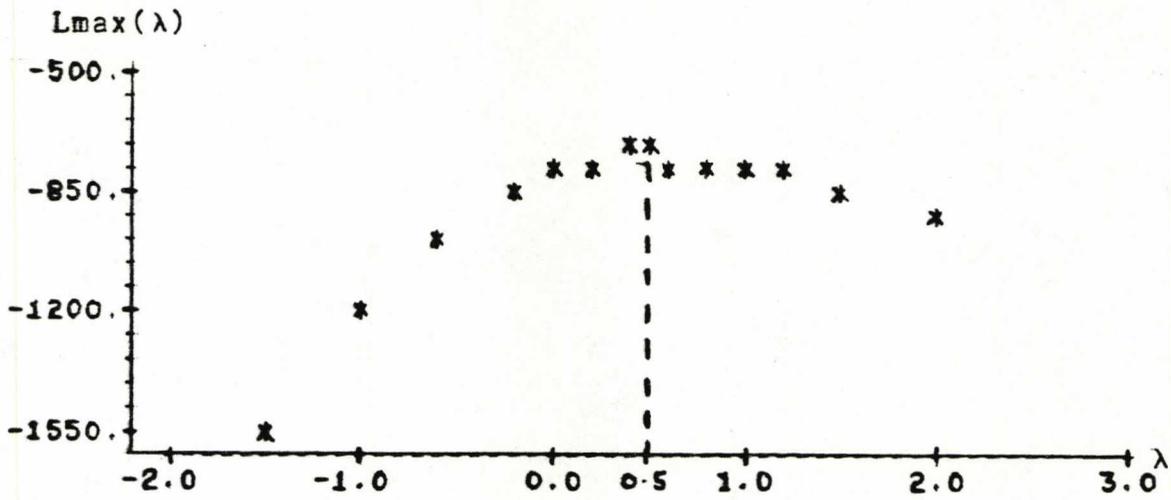


Table 4.4.1.2      Coefficients of skewness ( $k_1$ ) and kurtosis ( $k_2$ ) for the response variable. Coefficients in parentheses are for the untransformed data

Data set		SPRING			AUTUMN		
		n	$k_1$	$k_2$	n	$k_1$	$k_2$
Sedge W.	(D)	462	1.2 (2.7)	7.2 (19.5)	325	1.1 (2.4)	5.5 (13.4)
Willow W.	(D)	613	3.2 (7.2)	22.7 (92.9)	345	2.0 (3.1)	8.9 (18.3)
Whitethroat	(D)	526	2.9 (4.2)	14.4 (35.1)	329	3.7 (6.8)	30.8 (72.3)
Redstart	(D)	498	2.1 (4.2)	18.9 (35.4)	343	1.1 (3.0)	15.4 (20.5)
Sedge W.	(P)	123	3.0 (7.5)	30.0 (71.5)	171	1.4 (3.9)	9.4 (21.3)
Willow W.	(P)	131	3.5 (5.0)	21.1 (33.0)	169	2.1 (3.4)	9.3 (21.7)
Whitethroat	(P)	132	9.4 (8.6)	96.2 (90.5)	197	4.2 (6.6)	33.7 (69.0)
Redstart	(P)	126	2.2 (3.4)	14.0 (21.7)	722	0.5 (1.4)	6.5 (8.5)

and kurtose, even after transformation.

It seems that transforming the response variable is not sufficient to produce a set of normally distributed responses, and so one of the assumptions of the analysis is not met. There is little that one can do in this event apart from proceeding with a cautionary analysis or choosing some alternative method of relating migration counts with weather data. We shall take the former course of action first of all, and study a different approach to the problem in the following chapter. The next stage in the regression analysis is to consider the model which includes interactions of weather variables and weather variables raised to the powers 2 and 3. If it is not possible to improve upon the results shown in Table 4.4.1.1, then we must accept that linear regression analysis is not a suitable tool for dealing with the problem at hand.

Table 4.4.2.1      Regression on polynomials of weather variables and two-variable interactions. Figures in parenthesis are for regression on basic set of 14 variables

		SPRING				AUTUMN			
		R <sup>2</sup>	F	P <sub>F</sub>	R <sup>2</sup>	F	P <sub>F</sub>		
Sedge W.	(D)	.06 (.05)	1.22	.23	.07 (.05)	1.31	.18		
Willow W.	(D)	.06 (.05)	2.82	.00	.08 (.06)	1.50	.08		
Whitethroat	(D)	.05 (.05)	1.27	.20	.08 (.06)	1.59	.07		
Redstart	(D)	.08 (.07)	1.79	.01	.07 (.05)	1.32	.17		
Sedge W.	(P)	.13 (.13)	0.68	.83	.14 (.10)	1.36	.16		
Willow W.	(P)	.10 (.07)	0.57	.93	.12 (.12)	1.16	.30		
Whitethroat	(P)	.11 (.11)	0.54	.95	.14 (.10)	1.55	.08		
Redstart	(P)	.17 (.10)	1.01	.46	.10 (.05)	1.46	.12		

P<sub>F</sub> = significance level of F<sub>REG</sub>.

The results for the regression analyses are shown in Table 4.4.2.1 and apart from a very small increase in R<sup>2</sup> for most analyses, the picture remains one of total inadequacy of the regression model in this context.

Although other authors have, in the past, presented results from successful regression analyses of bird data and weather data, it is clear that it is not possible to do the same using British observatories' daily counts. This does not necessarily imply that weather has no effect on daily counts at observatories or even that the two types of data cannot be related by some simple model. In the next chapter we see how linear discriminant analysis can usefully be employed in this context.

#### 4.5 A review of Hussell's method

##### 4.5.1 Description of the method

Hussell (1981) describes an attempt at formulating migration indices from counts of migrants at Long Point Bird Observatory, Ontario. He used data from a period of 18 years from which he constructed annual indices and attempted to validate them by comparisons with indices from Breeding Bird Surveys. An earlier report by Hussell and Risley (1976), described briefly in section 2.4, suggested that the counts of migrants alone were not sufficient to reproduce population trends as reflected by breeding surveys. They found that, by incorporating weather information into the analysis, more encouraging results were obtainable. Hussell (1981) describes in detail how this was done, and an outline of the method is given here. We are primarily concerned with the methodology rather than the ornithological details such as which species are studied for example, but these are to be found in Hussell's paper.

Hussell's method involves regressing the bird counts onto a number of explanatory variables which include dummy variables for year effects. It is the coefficients of these dummy variables that are used to form annual migration indices.

The response variable is based on the daily count of grounded migrants for a given species. Each count is increased by 1, and then transformed by a logarithmic transformation. Hussell chose this transformation for the

following reasons:

- (i) The distributions of counts from Long Point are skewed to the right and that the variances increase with the means. This is typical of data of this type, and is clearly seen in the British data.

It is doubtful, however, that a simple logarithmic transformation is able to produce a set of normally distributed counts - certainly, for the British data at least, it was not able to do so. Unfortunately, no mention of the usefulness of this transformation, in this particular situation, is reported by Hussell.

- (ii) Hussell assumes that the year effects are multiplicative in nature, as regards the migration counts. In other words, he is assuming that if the population doubles from one year to the next, then each daily count in the second year will, on average, be twice that in the first year. An implicit assumption that is made here is that the overall pattern of migration remains the same from year to year. We have already seen that this is not the case for British data.

If, however, year effects were in fact multiplicative, then a logarithmic transformation would convert these effects to additive ones.

Without re-analysing the Canadian data it is difficult to assess the usefulness of this transformation, but in the

next section we shall see how it applies to the British data.

The year effects are included in the model by means of dummy variables which take the values 1 or 0. For a given observation or a given row of the design matrix, the value of the dummy variable corresponding to the year in which the observation was made is unity, and all other dummy variables have the value zero. We see below how these variables are used to construct migration indices.

The within-season variation of migration is built into the model by means of a set of variables based on the date. The days of each season are numbered sequentially to form the basic date variable,  $V_1$  say. Also included are variables such as  $V_1^2$ ,  $V_1^3$ ,.....etc. Variables up to and including the tenth power of  $V_1$  were used by Hussell, so that an extremely complex polynomial is being fitted to each season's counts.

Hussell used five weather variables : temperature, cloud cover, visibility, wind direction and wind speed. The size of the problem with which Hussell dealt was clearly very large. Since he used bird counts from two areas of Long Point, and therefore considered site-weather interactions and site-date interactions, his set of explanatory variables numbered 61. The sample sizes for the six species that he considered ranged from 828 to 971.

Hussell estimated the parameters of the model by using the BMDP2R program which performs a stepwise regression procedure. The year variables were forced into the regression throughout.

Broadly speaking, Hussell found that the model gave a suitable fit to the data for the six species that he considered. Values of  $R^2$  ranged from 0.27 to 0.63, but with so many variables it is difficult to interpret the final regression equations.

Examination of residual plots showed that the distribution of residuals became increasingly distorted at predicted values of less than 1.5. The observed, triangular pattern of residuals when plotted against predicted values suggests that the assumption of constant variance over the residuals is not met. Often one finds that a further transformation is able to stabilize this variance. Residual plots of this type often show diagonal bands of points (see Hussell, 1981, Fig. 1), which are attributable to the discrete nature of the bird counts.

A problem with linear regression in this type of situation is that sometimes the predicted values are negative. This is obviously meaningless in terms of the nature of the data, i.e. non-negative counts, but often happens with birds that occur only in small numbers. In order to overcome this problem, Hussell repeated the regression analyses after removing observations with negative predicted values.

This, as Hussell admits, does not completely overcome the problem as the revised regression equations may still give rise to negative predicted values, although the situation is somewhat improved.

So far in this chapter, we have seen that, for British observatories' data at least, linear regression analysis is not suitable for relating weather and migration data. From the evidence presented in Hussell (1981), it is not convincing that regression performs any better on the Long Point data, particularly as regards the validity of the assumptions made in performing regression. It is difficult to say more about Hussell's work here, since only a limited amount of information is provided in his paper.

Hussell's method of constructing migration indices is examined in the next section.

#### 4.5.2 Migration Indices

Hussell (1981) used as his migration indices, the mean numbers of birds seen per day in each year. In terms of the regression model that he fitted, the mean count in year  $j$ ,  $\hat{Y}_j$  say, is given by,

$$\hat{Y}_j = A_j + \sum_m C_m \bar{d}^m + \sum_w b_w \bar{W}_w$$

where  $A_j$  is the coefficient for the dummy variable relating to year  $j$ , the second term of the equation is the sum of the means of all date variables multiplied by their respective regression coefficients, and the final term is the sum of weather variable means multiplied by their regression

coefficients. The means for the last two terms of the equation are taken over all days in all years.

In order to present the indices in terms of bird numbers, the inverse of the transformation of the response variable is used, producing the index  $M_j$  for year  $j$

$$M_j = e^{\hat{Y}_j} - 1$$

In order to avoid overparameterisation of the model, one year needs to be designated the reference year, and its dummy variable omitted. If all dummy variables were included, then the least squares estimates would be unobtainable since a singular matrix would need to be inverted. For testing the difference between any year's index and the index of the reference year, the F-to-remove value of the dummy variable for the year in question is compared with  $F_{1,n-v-1}$  where  $n$  is the number of observations and  $v$  the number of explanatory variables. For other comparisons of indices, different reference years need to be used.

#### 4.5.3 Results using British data

The models considered in this section are extensions of those considered in section 4.4.1. Thus our basic set of 14 weather variables (7 at lags 0 and 1) are included, but now we introduce date variables and year variables, as described in section 4.5.1. Also, the response variable consists of log-transformed counts, rather than deviations from a moving average. The effect of including year variables, and of accounting for seasonal variation in

migration by means of polynomials can be measured by comparing the observed values of  $R^2$  with those of Table 4.4.1.1, which, for convenience, are shown in Table 4.5.3.1 in parentheses.

Table 4.5.3.1      Regression on 14 weather variables with date and year variables included.  
The response variable is log (observed count + 1)

		SPRING		AUTUMN	
		$R^2$		$R^2$	
Sedge W.	(D)	.34	(.05)	.38	(.05)
Willow W.	(D)	.28	(.05)	.56	(.06)
Whitethroat	(D)	.36	(.05)	.67	(.06)
Redstart	(D)	.23	(.07)	.41	(.05)
Sedge W.	(P)	.28	(.13)	.46	(.10)
Willow W.	(P)	.30	(.07)	.59	(.12)
Whitethroat	(P)	.52	(.11)	.63	(.10)
Redstart	(P)	.27	(.10)	.50	(.05)

Table 4.5.3.1 shows, in all cases, that  $R^2$  is substantially increased by including date and year variables, but, if we examine which variables contribute significantly to the regressions, then we again find that there is little or no sign of a connection between the response variable and weather variables. The high values of  $R^2$  are largely due to the date and year variables, and so fitting models of this type is again unable to establish a linear relationship between migration counts and weather variables.

Table 4.5.3.2 lists the variables that gave significant t-statistics (5% level).

Table 4.5.3.2 Significant variables for extended regression models

		SPRING	AUTUMN
Sedge W.	(D)	VEL,Y4,Y8,Y10,Y11	VEL,Y2,Y3
Willow W.	(D)	VEL,CL,TEM,D1-D3,Y1-Y11	DIR,VEL,HUM,D1-D3,Y1,Y3
Whitethroat	(D)	CL,TEM,PR,D1,D2,Y5-Y11	DIR,VEL,D1,D2,Y1-Y3
Redstart	(D)	CL,HUM,PR,D1-D3,Y5	VEL,TEM,D1-D3,Y1
Sedge W.	(P)	TEM,Y1	VEL,VIS',TEM',D3,Y1,Y2
Willow W.	(P)	Y2	VEL,VEL',VIS',D1-D3,
Whitethroat	(P)	VEL,TEM,TEM',D1-D3,Y1,Y2	VEL,VIS,D1-D3,Y1,Y2
Redstart	(P)	D1-D3,Y1	8 weather var's,D1-D3,Y1

key: DIR : wind direction                      VEL : wind velocity  
 CL : cloud cover                                VIS : visibility  
 TEM : temperature                              HUM : humidity  
 PR : pressure                                    D1 : date  
 D2 : date (squared)                            D3 : date (cubed)  
 Y<sub>i</sub> : ith year variable

A lagged variable is indicated by a single prime.

Table 4.5.3.2 may only be used as a rough guide as to which variables contribute significantly to the regression models since, if one deleted those variables that are not significant, then the situation as regards the other variables may change. It seems to be fairly clear, however, that the apparent 'good fit' of the models is obtained by including the date variables and, to a certain extent, the year variables. If one constructed migration indices from these results, then they would merely reflect the degree to which each season of counts can be represented by a cubic polynomial of the date, rather than migration volume. It would appear that, although Hussell (1981) reported a certain amount of success using these models, we need some alternative method of constructing migration indices in order to monitor population levels.

#### 4.6 Conclusions

This chapter has presented strong evidence to suggest that multiple linear regression analysis has little to offer in the study of the effects of weather on bird migration, as recorded at British observatories. One might attempt to fit more complex models using many more variables but it has already been shown that the use of polynomials of weather variables and a range of transformations failed to suggest suitable models. Also, the practical interpretation of more complex models would be very difficult.

The results of this chapter do not necessarily open to question the results of other author's work in this area since, in most cases, the analyses were based on radar data or data obtained from very different geographical situations. However, one might question the extent to which the assumptions of the analyses are met. One would assume that the absence of a discussion on the statistical assumptions and their validity in most of the published studies was due to either the unfamiliarity of authors with multiple regression analysis or possibly the lack of space in the particular journals concerned.

The results of this chapter can, however, be viewed positively. As a consequence of the work reported here, ornithologists may think again before following the example of past authors and using multiple regression analysis in a situation that demands alternative statistical methods.

This chapter has also illustrated a simple method of allowing for time dependency within migration seasons, which may be of use when examining daily counts by methods other than regression.

We remain with the subject of weather and migration for the next chapter where the usefulness of linear discriminant analysis is investigated.

## CHAPTER 5

### RELATING MIGRATION DATA TO WEATHER (II) DISCRIMINANT ANALYSIS

#### 5.1 Introduction

The previous chapter explored the possibility of relating migration counts to weather data by using multiple regression analysis. Strong evidence was found to suggest that daily counts from British observatories are not able to be explained by a linear combination of weather variables.

In this chapter we use linear discriminant analysis in an attempt to find linear functions of weather variables that best discriminate between high and low counts of migrants. If it is possible to discriminate successfully between categories of migration counts in this way then we have a means of weighting annual totals of migrants according to weather conditions. This can be explained in the following manner: if the raw total of migrants for a given season is used as an index then the index is, to a certain extent, dependent on weather conditions as well as the number of birds migrating. For example, if the season in question experienced an unusually large number of nights of fine weather then the index will be unduly low whereas a season with many cloudy and rainy nights may inflate the index. An adjustment to the raw total may be made by comparing each season with an 'average' season, in terms of numbers of days on which migrants were or were not expected to be recorded. This idea is illustrated at the end of this chapter by adjusting migration indices of Hussell and Risley which were calculated in chapter 2.

The basic problem of discriminant analysis is to assign an observation, in this case a multivariate weather observation for a particular day, of unknown origin, to one of a number of distinct groups on the basis of the value of the observation. A simple measure of the success of the analysis is provided by the proportion of observations that are assigned to the correct group - we assume that the data consist of correctly assigned observations which can be re-assigned using results from the analysis. Assignments are made using discriminant functions, the derivation of which can be found in Morrison (1976) and Lachenbruch (1975).

Note that discriminant analysis has been used successfully with weather data in the past, though in a different context - see Glahn (1965).

The data that we use in this chapter are the same as those used for the regression analyses of chapter 4. Thus we consider all combinations of four species (Sedge Warbler, Willow Warbler, Whitethroat and Redstart), two observatories (Dungeness and Portland) and two seasons (spring and autumn).

The results of this chapter were obtained from the statistical package BMDP, using program P7M. This program allows one to perform discriminant analyses using all variables, or subsets of variables selected by one of a number of criteria. Recent reviews of variable selection methods in linear discriminant analysis are found in McKay and Campbell (1982) and Kempson (1983), but there appears to be no recommended 'best' procedure. Where stepwise analysis was performed in this chapter, the BMDP routine based on

F-tests was employed, Dixon (1981).

## 5.2 Subdividing migration counts into groups

The object of the analysis is to divide the counts of migrants into two groups representing high and low counts and then to attempt to discriminate between the groups using the weather data. Only two groups are considered in the interests of simplicity. Recall that each count has been corrected for seasonal effects.

A ~~classification~~ function is derived for each group which enables us to calculate, for each day's observation, the probability of that observation belonging to each of the two groups. Each observation may then be allocated to the group corresponding to the highest probability. If the analysis has been successful, then we will find that low counts are allocated to the group representing low counts, and high counts are allocated to the group representing high counts. If the analysis has not been successful, then many counts will have been wrongly classified.

We would like to divide the counts into groups of approximately the same size since the allocation of observations is carried out partly on the basis of prior probabilities which, in the absence of a suitable alternative are usually taken to be the relative group sizes. If we had one group much larger than the other, this group may receive an unduly large number of allocations.

Many criteria may be used for defining the groups, but one method which helps ensure that the two groups contain only extreme deviations is to define extreme as meaning greater than one

standard deviation from the mean. This is purely an ad hoc method, but it does produce 'low' and 'high' groups of approximately equal size. From now on, groups 'low' and 'high' are defined in this way.

Fig. 5.2 shows two examples of histograms of the two extreme groups, together with the group sizes. These data are the amalgamation of three years' counts in each case. In all future analyses, the counts are pooled over a number of years, the precise time periods having been listed in section 3.2.

### 5.3 Preliminary tests

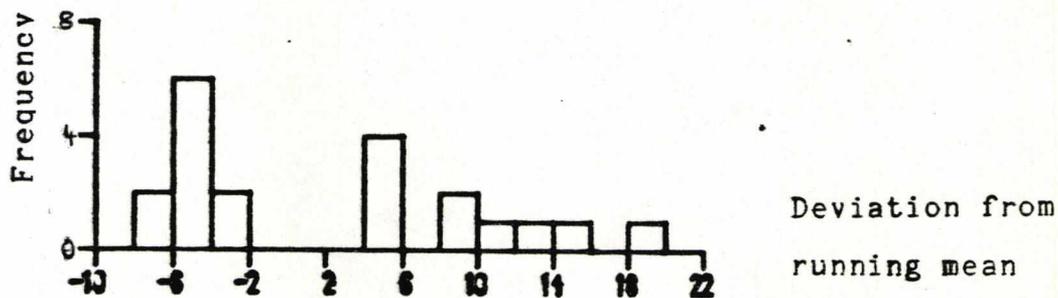
We first consider the underlying assumptions of linear discriminant analysis, namely multivariate normality of the discriminating variables and equal covariance matrices of the two groups of observations. The former assumption is needed for the significance tests of section 5.3.2 and both are needed for the stepwise procedures of the following section.

#### 5.3.1 Multivariate normality

In order to test the equality of two covariance matrices we assume multivariate normality among the discriminating variables. Mardia (1974) reports that testing covariance matrices is seriously affected by multivariate kurtosis, but not multivariate skewness. Also, Mardia (1974) states that tests of sample mean vectors such as Hotelling's  $T^2$  are sensitive to skewness rather than kurtosis. In the light of these remarks, the form of multivariate normality testing adopted here is based on the coefficients of skewness and

Fig. 5.2 Histograms of detrended daily counts

(a) Redstart, Portland, Spring

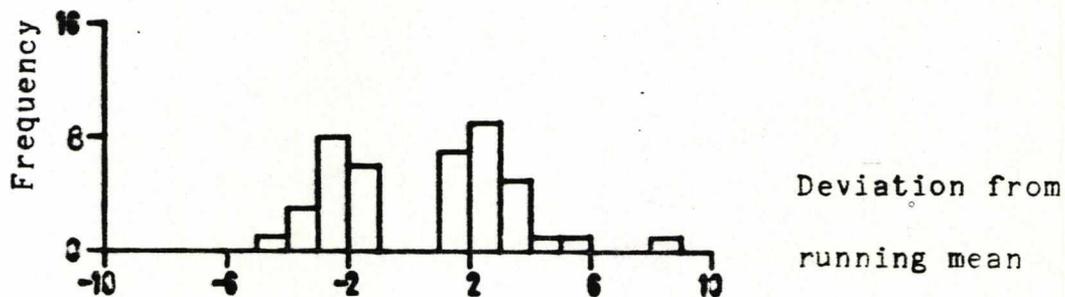


Group sizes:

10

10

(b) Redstart, Portland, Autumn



Group sizes:

18

24

kurtosis. There are other forms of testing such as those based on the univariate Shapiro-Wilk test (see Royston, 1983), but in the absence of a suitable algorithm, this method involves a formidable amount of computing. A recent review of other forms of testing is given in Mardia (1981). See also Gnanadesikan (1977) and Cox and Small (1978) for reviews.

The coefficient of skewness for a set of data consisting of  $n$  observations on  $p$  variables is estimated by

$$b_{1,p} = \frac{1}{n^2} \sum_{i,j=1}^n r_{ij}^3$$

where  $r_{ij} = (\underline{x}_i - \bar{\underline{x}})' \underline{S}^{-1} (\underline{x}_j - \bar{\underline{x}})$  and where  $\underline{x}_i$  is the vector of data for the  $i^{\text{th}}$  observation, and  $\bar{\underline{x}}$  is the vector of means of the discriminating variables.  $r_{ij}$ , therefore, is the Mahalanobis angle between the vectors  $(\underline{x}_i - \bar{\underline{x}})$  and  $(\underline{x}_j - \bar{\underline{x}})$  where  $\underline{S}$  is the sample covariance matrix. Mardia (1974) shows that, asymptotically,  $\frac{nb_{1,p}}{6}$  has a chi-square distribution with  $p(p+1)(p+2)/6$  degrees of freedom if the data come from a multivariate normal population.

Table 5.3.1.1 shows the test statistics for the 16 species/observatory/season combinations.

Table 5.3.1.1

Test statistics for assessing multivariate normality of the discriminating variables.

In each case,  $p = 14$ .  $K$  represents the standardised statistic for  $b_{2,p}$ .

i.e.  $K \sim N(0,1)$  under the hypothesis of normality

Species		SPRING		AUTUMN	
		$nb_{1,p}/6$	$K$	$nb_{1,p}/6$	$K$
Sedge W.	(D)	742.7 **	0.50	632.4 *	-1.55
Willow W.	(D)	571.1	-2.55 **	587.9	-2.74 **
Whitethroat	(D)	556.4	-2.43 **	762.7 **	0.50
Redstart	(D)	623.4 *	-0.61	550.7	-2.79 **
Sedge W.	(P)	369.9	-4.88 **	432.4	-4.08 **
Willow W.	(P)	393.1	-4.61 **	422.8	-4.27 **
Whitethroat	(P)	403.5	-4.47 **	454.2	-3.75 **
Redstart	(P)	389.9	-4.67 **	596.2	-1.84 **

\* significant at 5% level, \*\* significant at 1% level

In the majority of cases, the coefficient of skewness is non-significant, suggesting that  $T^2$  tests on the mean vectors associated with these data are feasible - these tests form the basis of two-group discriminant analysis. On the other hand in all but 5 cases, the coefficient of kurtosis is highly significant. This means that tests of equality of covariance matrices will be approximate tests, and little importance should be attached to the significance levels obtained. Unlike the univariate case, it is difficult to overcome the non-normality by choosing a suitable transformation of the data since there are numerous combinations of variables that may need to be transformed. There is little that one can do other than proceed with caution under the assumption of normality.

### 5.3.2 Equality of covariance matrices

The test statistic for comparing two covariance matrices is given by the formula

$$X^2 = n \log_e |\underline{S}| - n_1 \log |\underline{S}_1| - n_2 \log |\underline{S}_2|$$

where  $\underline{S}_1$ ,  $\underline{S}_2$  and  $\underline{S}$  are the sample covariance matrices for the two groups and the pooled estimate of the covariance matrix respectively. ( $\underline{S}_1$  refers to the low count group and  $\underline{S}_2$  to the high count group). This test is a likelihood ratio test and under the null hypothesis  $X^2$  is asymptotically distributed as a chi-square variate with  $\frac{1}{2}p(p+1)$  degrees of freedom, where  $p$  is the number of variables. We do not use lagged variables here since in several cases there are insufficient data to estimate both  $\underline{S}_1$  and  $\underline{S}_2$ . Even with seven variables, in two cases one group was too small to

allow estimation of the covariance matrix.

Table 5.3.2.1 shows the observed test statistics for each case. The critical value of the chi-square distribution at the 1% level of significance with  $\frac{1}{2}(p+1) = 28$  degrees of freedom is 48.3. Three of the fourteen observed statistics exceed this value, while none of the other statistics is significant at the 5% level.

Table 5.3.2.1      Testing the equality of within group covariance matrices ( $\chi^2$ )

		<u>SPRING</u>	<u>AUTUMN</u>
Sedge W.	(D)	25.8	35.6
Willow W.	(D)	36.5	28.6
Whitethroat	(D)	28.8	54.6 **
Redstart	(D)	27.9	31.8
Sedge W.	(P)	-	69.1 **
Willow W.	(P)	36.5	34.4
Whitethroat	(P)	-	51.2 **
Redstart	(P)	40.5	31.5

\*\* significant at 1% level (5% level critical value = 41.3)

There is little that one can do about the three significant cases since the implication of the results is that one should use quadratic discriminant analysis which is known to be unsuitable on small data sets (Lachenbruch and Goldstein, 1979). We should, however, treat further results pertaining to these data sets with a certain amount of caution.

#### 5.4 Results

There is little point in trying to discriminate between two groups of observations if the group means are approximately equal. Before examining results from discriminant analyses it is as well to check the separation of group means. The appropriate measure to use is the Mahalanobis squared distance between the two groups of observations which is given by

$$D^2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$$

The significance of  $D^2$  can be tested by calculating the F-ratio

$$F = \frac{n_1 n_2 (n_1 + n_2 - p - 1)}{(n_1 + n_2) (n_1 + n_2 - 2) p} D^2$$

where  $n_1$  and  $n_2$  are the group sizes, and  $p$  the number of discriminating variables. Under the null hypothesis ( $D^2 = 0$ ),  $F$  has an F-distribution with  $p$  and  $n_1 + n_2 - p - 1$  degrees of freedom.

We first examine the separation of group means using the full set of seven variables, and we introduce lagged variables later on.

Table 5.4.1 shows the values of  $D^2$  and the associated P-values for the set of 16 data sets, i.e. all combinations of species, observatory and season.

Table 5.4.1 Mahalanobis distances between group means  
using 7 weather variables

		<u>SPRING</u>		<u>AUTUMN</u>	
		D <sup>2</sup>	P	D <sup>2</sup>	P
Sedge W.	(D) <sup>+</sup>	0.35	0.223	1.30	0.002 **
Willow W.	(D)	2.09	0.000 **	1.63	0.001 **
Whitethroat	(D)	0.70	0.036 *	1.63	0.000 **
Redstart	(D)	0.56	0.131	0.51	0.389
Sedge W.	(P) <sup>+</sup>	3.11	0.494	4.40	0.077
Willow W.	(P)	5.89	0.138	1.93	0.294
Whitethroat	(P)	2.23	0.571	3.89	0.004 **
Redstart	(P)	3.28	0.237	1.34	0.149

+ (D) = Dungeness, (P) = Portland

\* significant at 5% level, \*\* 1% level

It appears that discriminant analysis is likely to be more successful for autumn data than spring, particularly on Dungeness data.

Table 5.4.2 shows the classification success rates evaluated using the jackknife method. This method entails classifying each observation using discriminant functions derived from the data set with that particular observation omitted. This estimate of the success rate is known to be less biased than certain other estimates; see Lachenbruch and Mickey (1968).

Table 5.4.2 Success rates for discriminant analyses  
using seven variables - all variables included

		<u>SPRING</u>	<u>AUTUMN</u>
Sedge W.	(D)	56.0%	67.9%
Willow W.	(D)	73.6%	66.7%
Whitethroat	(D)	59.8%	72.7%
Redstart	(D)	62.2%	52.3%
Sedge W.	(P)	50.0%	59.1%
Willow W.	(P)	61.1%	59.3%
Whitethroat	(P)	57.9%	74.3%
Redstart	(P)	55.0%	57.1%

The success rates shown in Table 5.4.2 show that only in three cases is the success rate greater than 70%. These cases correspond to data sets with a significant  $D^2$  - see Table 5.4.1. Success rates of the order of 75% or above are generally considered to be high, but in one case, the Sedge Warbler data from Portland (spring) a success rate of 50% was obtained. This means that the allocation of observations would be just as successful if we used a random procedure.

We now move to a stepwise analysis where the aim is to include only the important discriminatory variables. We also introduce lagged variables (each of the variables is repeated, with a lag of one day) in order to try to achieve more successful discrimination for all the data sets. In the interests of simplicity we use the BMDP stepwise procedure where variables are entered or deleted from the analysis on the basis of F-tests. One can set the critical values for the tests and thus control the number of variables that are permitted in the analysis: see Dixon (1981, page 553). One can set the critical F-value so that a conventional significance level is used, e.g. 5%, but this entails using a different value for each data set since the sample sizes vary from data set to data set. In the light of this an ad hoc approach was adopted viz. the choice of an F-value that allows a 'small' number of variables in the analysis, while also considering the results with other F-values to see whether a marked increase in the success rate was obtainable or not. 'Small' in this context usually meant 4 or 5 variables out of 14, although the minimum and maximum numbers used turned out to be 1 and 7 respectively. Table 5.4.3 shows the success rates for the

stepwise analyses starting from 14 variables. The figures in parentheses are the success rates obtained in the initial analyses.

Table 5.4.3      Success rates for stepwise analyses

		<u>SPRING</u>		<u>AUTUMN</u>	
Sedge W.	(D)	62.1%	(56.0)	69.0%	(67.9)
Willow W.	(D)	75.8%	(73.6)	76.9%	(66.7)
Whitethroat	(D)	68.0%	(59.8)	73.7%	(72.7)
Redstart	(D)	65.6%	(62.2)	61.5%	(52.3)
Sedge W.	(P)	81.2%	(50.0)	90.9%	(59.1)
Willow W.	(P)	88.9%	(61.1)	70.4%	(59.3)
Whitethroat	(P)	68.4%	(57.9)	71.4%	(74.3)
Redstart	(P)	90.0%	(55.0)	76.2%	(57.1)

In 15 of the 16 cases either the reduction in the number of variables or the inclusion of lagged variables has led to an improved success rate, and in some cases a considerable improvement: see, for example, the Redstart at Portland in both spring and autumn. Table 5.4.4 shows the corresponding values of  $D^2$ , together with an indication of the significance of each value.

Table 5.4.4      Mahalanobis distances between group means :  
stepwise analyses with lagged variables

		<u>SPRING</u>	<u>AUTUMN</u>
Sedge W.	(D)	0.29 *	1.18 **
Willow W.	(D)	1.99 **	3.04 **
Whitethroat	(D)	1.03 **	1.47 **
Redstart	(D)	0.27 *	0.34 *
Sedge W.	(P)	10.25 *	11.27 **
Willow W.	(P)	5.30 **	1.75 *
Whitethroat	(P)	6.38 *	2.81 **
Redstart	(P)	11.25 *	1.60 *

\* significant at 5% level, \*\* at 1% level

Table 5.4.4 shows that we are able to form linear functions of the variables that discriminate between groups. The separation of the group means is significant at least at the 5% level in every case.

One method of investigating which variables are important in the discrimination is to form, for each data set, a single linear combination that best discriminates between groups. This linear combination is referred to as a canonical variate, and since we are discriminating between only 2 groups, only one such canonical variate exists in each case. The variate is such that the between group variability in canonical scores is maximised, while the variability of within group scores is minimised. In effect, we reduce the dimensionality of the data to one.

Table 5.4.5 shows the coefficients of the weather variables that form the canonical variates. We can see from this table which variables are important discriminators and, of those that are important, we can see how each variable is weighted. For example, the coefficients for the variate for the Sedge Warbler (spring) data from Portland include the following components:

.92 (wind direction, no lag) + .42 (wind direction, lag 1) and .85 (wind speed, no lag) - .76 (wind speed, lag 1). The wind direction component appears to be an average taken over two days with extra weighting on the first day, while the wind speed component appears to represent a sudden increase in mean wind speed over the two-day period. This feature is repeated in several cases, as Table 5.4.5 shows, and emphasises the usefulness of including lagged variables. One could, of course, include

variables with a lag of, say, two days, enabling local maxima and minima to be detected - for example a rise in pressure after several days of decreasing pressure may indicate the passing of a low pressure system which may coincide with a wave of migrants. This, unfortunately, would raise the number of variables under consideration to an unmanageable level and has not been attempted here.

Table 5.4.5 Canonical variate coefficients. For each case, left and right hand columns represent variables lagged by 0 and 1 day respectively

SPRING

Variable	Sedge W. (D)		Willow W. (D)		Whitethroat (D)		Redstart (D)	
Wind direction						.42		
Wind speed	.14		.64			.63		
Low cloud	-.70		-1.07				-1.00	
Visibility								
Temperature			-.83					
Humidity						-.85		
Pressure						-.50		

Variable	Sedge W. (P)		Willow W. (P)		Whitethroat (P)		Redstart (P)	
Wind direction	.92	.42						
Wind speed	.85	-.76	1.56		1.34	-.83	1.33	
Low cloud			-1.07		-1.66			
Visibility					-1.95	1.34		-.68
Temperature	-.84	1.29	-.68					
Humidity								1.17
Pressure							1.42	

AUTUMN

Variable	Sedge W. (D)		Willow W. (D)		Whitethroat (D)		Redstart (D)	
Wind direction			-.56	.44	-.42			
Wind speed	.95				.76		1.00	
Low cloud								
Visibility			-.68	.56	-.42			
Temperature	.41		.48					
Humidity				-.32				
Pressure			-.85		-.54			

Variable	Sedge W. (P)		Willow W. (P)		Whitethroat (P)		Redstart (P)	
Wind direction	-1.38		-.64		-.65			
Wind speed	1.58	.68	.87		.75			.62
Low cloud			.65				.54	
Visibility							.93	
Temperature		1.54			-.55		1.05	
Humidity								
Pressure								

It is difficult to pick out any interesting patterns from Table 5.4.5, save for the fact that there is such a wide variety of weather variables included. There appears to be great variability between species, and there seems to be no noticeable similarities or differences between observatories or seasons. Also, several cases involve no lagged variables at all, while one case, the Whitethroat (spring) at Dungeness consists entirely of lagged variables. The suggestion that lagged variables may be more important for some species than others suggests that we may be observing different types of migration. For example, some species may migrate with a series of short journeys, while others may, when possible, migrate with a single long journey. This is merely conjecture, but an examination of ring recoveries may be a useful follow up here since one may gain some insight into each species' migratory behaviour.

The variable that occurs in most cases in Table 5.4.5 is wind speed (14 out of 16 cases), whether lagged by one day or not. This is followed by temperature (8 cases), wind direction (7 cases) and low cloud cover (7 cases). The fact that these variables appear in approximately half of the cases might suggest that this subset of the seven variables may be able to discriminate between groups satisfactorily for all cases, although not being the optimal subset in each case. If this was the case, then this would greatly simplify any applications of this work, such as the construction of migration indices using totals of migration counts weighted according to the number of 'bird days' in each season.

Table 5.4.6 shows the results of attempting to discriminate between groups using the same set of four variables (and associated lagged variables) in every case. In all but one case the jackknife success rate has fallen with the use of common subset of variables and indeed, the Mahalanobis squared distance between groups is not significant in over half of the cases. There may possibly exist a subset of the 14 variables that is suitable for discrimination in each case, although sub-optimal, but the question is how to find it?

Table 5.4.6      Mahalanobis distances and jackknife error rates for analyses using a common subset of 4 variables

		<u>SPRING</u>		<u>AUTUMN</u>	
		D <sup>2</sup>	Success rate	D <sup>2</sup>	Success rate
Sedge W.	(D)	0.56	57.8%	1.38 **	60.7%
Willow W.	(D)	2.17 **	73.6%	1.35 **	66.7%
Whitethroat	(D)	0.78 *	57.7%	1.59 **	70.7%
Redstart	(D)	0.57	57.8%	0.57	52.3%
Sedge W.	(P)	10.99	62.5%	13.5 **	90.9%
Willow W.	(P)	7.15	66.7%	3.03	55.6%
Whitethroat	(P)	2.10	47.4%	3.28 *	68.6%
Redstart	(P)	5.09	65.0%	0.80	57.1%

\* significant at 5% level, \*\* 1% level

In the above example we have taken an educated guess and found that the subset is not suitable for general use. Ideally one would examine every subset but, even for seven variables, there are 127 possible subsets to examine over 16 sets of data. It would appear that in order to achieve good discrimination in each case, the data samples need to be treated individually. However, we now briefly return to Table 5.4.6 and attempt to give simple interpretations to some of the variables that occur in the canonical variables.

The most frequently occurring variable is wind speed (variable 2). From the data it is clear that high wind speeds are associated with low counts of migrants. This is contrary to what we may have expected since calm weather is usually regarded by ornithologists as being associated with low counts. For those cases where wind direction is found to be an important discriminator, we see that the high winds associated with low counts are in fact favourable winds, i.e. blowing in the direction of the birds' flight. It is not surprising that such conditions encourage birds to fly straight over the coastal observatories. An important variable in spring, and to a lesser extent in autumn, is low cloud cover. In all 5 cases featuring this variable in spring, low amounts of low cloud cover are associated with low counts of migrants. This again confirms the opinions of observatory workers and visitors which were hitherto based largely on practical experience as opposed to scientific investigation.

Taking the data sets as a whole, it is difficult to determine any other consistent features as regards weather variables.

There are no obvious differences between the results for spring and autumn apart from a suggestion that lagged variables are slightly more important in spring than autumn. The success rates for individual species are much the same for both seasons, although the power of discrimination differs between Portland and Dungeness observatories for the Sedge Warbler and Redstart. In both cases the success rates are considerably higher for the Portland data. The average success rate is also higher for Portland: 80% compared with 69% for Dungeness. This may be due to the geographical

differences between the two sites. The distance travelled over sea by those spring arrivals passing through Dorset is considerably larger than the distance from France to Kent. If bad weather is met by birds heading for Dorset then by the time they arrive at the coast one would expect them to be more exhausted and disorientated than those birds taking a shorter sea crossing. Similarly, birds that are about to depart in the autumn may be more reluctant to do so from Dorset in poor weather since the sea crossing is that much longer. Thus, at Portland, one may find that records of birds are more sensitive to particular weather conditions than at Dungeness.

#### 5.5 Migration indices

The discriminant functions of the previous section can be used to classify each migration season as either 'one where we would expect to record more birds than usual', or 'one where we would expect to record fewer birds than usual'. This is achieved by designating each day either a bird-day or a non-bird day.

As an example, over the years 1961-72, the average proportion of bird days at Dungeness each spring was 43%. In 1961, the proportion was 46% which suggests that one would have expected to have observed more birds than usual due to weather conditions.

In order to adjust migration indices for weather, one simply multiplies the calculated index by 43% divided by the proportion of bird days in each individual year. Table 5.5.1 presents the correlation coefficients of the indices of Hussell and Risley - see section 2.4 - before and after adjusting for weather.

Table 5.5.1      Dungeness spring counts and CBC farmland indices - correlation coefficients

Migration Index	Sedge W.	Willow W.	Whitethroat
MIS - unadjusted	-.23	.04	.81 **
MIS - weather adjusted	-.10	-.22	.70 **
MIL - unadjusted	-.22	-.04	.78 **
MIL - weather adjusted	-.09	-.30	.84 **

\*\* significant at 1% level

Table 5.5.1 shows that there is no improvement of correlation with CBC indices after the migration indices have been adjusted for weather.

#### 5.6 Conclusions and discussion

The work of this chapter has shown that linear discriminant analysis enables one to study relationships between weather and migration data, if only at a very simple level. The conclusions that one draws from this work serve to reinforce some widely held opinions of ornithologists.

Evidence was found to suggest that discriminant analyses are more successful when using autumn data rather than spring data. This supports the view of ornithologists that birds are more determined to complete their migration in the spring than in autumn, and are

less deterred by adverse weather conditions.

Discrimination was, in general, improved by using lagged weather variables. This is what we may expect since it is not only weather at observatories that influences observatories data. By using lagged variables we hope to introduce information on weather which birds experience prior to arrival.

The success of the discriminant analyses varies from species to species and between the two observatories. For example, a high success rate was obtained for the Portland Redstart data (autumn), but not for Dungeness data. Ornithologists are not able to give explanations for these inconsistencies, but it seems likely that the geographical difference between the two sites is responsible.

Different weather variables were found to be important for different data sets, although wind (direction and/or velocity) appears to be an important feature for most data sets. High winds in the direction of the birds' flight appear to give rise to low counts at observatories, suggesting that birds are able to continue their flight.

Low amounts of cloud cover also appear to coincide with low counts which reinforces another widely held opinion of ornithologists that clear skies rarely lead to large counts of grounded migrants.

Discriminant analysis has been used in this context by Able (1973) and Richardson (1974), but both authors used radar data. Able (1973) divided his observations into eight groups according to the

direction of flight of the migrant passerines, and obtained a significant canonical variate for discriminating between groups. He used seven weather variables selected on the basis of univariate F-tests, but employed no further variable selection techniques so it is not clear whether this subset is optimal in any sense or not. Although Able's work is not comparable with the results of this chapter - he used autumn radar counts of passerine migrants - it is encouraging to see that wind direction was found to be an influential variable.

Richardson (1974) also used autumn radar counts, and reported overall success rates of less than 60%. He used three groups for his analysis, and so one would expect lower success rates than those reported in this chapter. Note that Richardson does not make clear whether or not his success rates are apparent or jackknife estimates, and so the true success rates may well be lower.

In chapter 7 we explore ways in which the results of this chapter may be put to use in fitting models to sequences of daily counts with the aim of constructing migration indices.

## CHAPTER 6

### MODELLING SEQUENCES OF DAILY COUNTS

#### 6.1 Introduction

In this chapter the serial dependence of daily counts is examined. With an understanding of the serial dependence of counts of migrants it is hoped that models can be formulated which lead to migration indices for monitoring population levels. The models considered in this chapter are simple Markov-type models and in the following chapter these models are further examined with the inclusion of information on weather conditions.

Not all of the material that is presented in this chapter is followed up in chapter 7, but as will be seen, a number of interesting problems and features come to light by attempting to fit simple models to sequences of daily counts.

#### 6.2 Immigration - death models

##### 6.2.1 Introduction

The serial correlation of a sequence of observations is a measure of the extent to which each observation is dependent on the previous observation. One could, of course, consider alternative auto-correlation structures but these will clearly involve more complex models, for example, those of the Box-Jenkins type - see section 3.3.4.

The motivation for exploring serial correlation within sequences of daily counts comes from the possibility that some birds may not depart from observatories during the

day or night following their arrival.

In autumn the weather conditions may prevent the birds from setting out on the channel crossing, or perhaps the birds have been unable to build up sufficient fat reserves for the migration and therefore need to spend more time foraging. In spring, the delay to departure may again be due to weather conditions, or possibly the physical exhaustion of the birds as they arrive. It is generally thought among ornithologists that if birds remain at observatories for any length of time, then they are less likely to do so in the spring, since it is vital for the birds to reach their breeding grounds as soon as possible in order to establish territories and find mates. Riddiford and Augur (1980) briefly discuss the behaviour of Willow Warblers at Dungeness during spring migration, and it appears from their analyses (based on ringing data) that nearly all Willow Warblers disperse during the day after their arrival. One objective of our study is to see whether a statistical analysis of daily counts (rather than ringing data) suggests similar behaviour.

For the purposes of this chapter, we shall treat the time scale as discrete, i.e. a series of days. Although in ornithological terms this may not be entirely appropriate. We shall assume that in general, immigrants arrive during the early hours of each day, remain until at least two or three hours after dawn when they are recorded, and then depart (if appropriate) during the remainder of the day.

The models that we consider in this section may be described as 'immigration - death' models, where death, in this context, refers to an exodus of birds from the observatory.

Our models reflect our interest in the assessment of migration volume, i.e. we seek to separate newly arrived birds from those that arrived on a previous day, for each daily count.

#### 6.2.2 A Poisson-binomial model

The first model that we consider is of the form

$$X_n = Y_n + I_n \quad (6.1)$$

where  $X_n$  is the count on day  $n$ ,  $Y_n$  is the number of birds that have stayed from day  $n-1$  and  $I_n$  represents the influx of birds immediately prior to the start of day  $n$ .  $n$  runs over the length of the migration season which starts on the first day that a bird is observed, and finishes on the last day on which a bird is observed. Even without any distributional assumptions, we need to make assumptions concerning  $Y_n$  and  $I_n$ . First, we assume that all the birds that are included in the  $Y_n$  were recorded during the day  $n-1$ , and so we have  $Y_n \leq X_{n-1}$ . Any birds that were present and unobserved on day  $n-1$ , but observed on day  $n$  will be classed as new arrivals and be included in  $I_n$ . Further, we make no specific attempt at this stage to build into the model the possibility of some birds remaining at the observatory for two days, three days, and so on, since this is built into the model automatically. A second assumption

that we make is that  $Y_n$  and  $I_n$  are independent, i.e. the rate of immigration is independent of the 'death' rate. This seems a reasonable assumption to make, since only when birds arrive at observatories will they be in a position to assess whether or not they should stay or move on as soon as possible.

Since  $Y_n$  is some proportion of  $X_{n-1}$ , we start by assuming that  $Y_n$  is a binomial random variable with index  $X_{n-1}$  (the observed count on day  $n-1$ ) and parameter  $p$ . The parameter  $p$  measures the probability of each bird remaining at an observatory from one day to the next, and  $1-p$  is the probability that each bird leaves the observatory soon after it is recorded. Notice that we are also assuming that (i) the probability  $p$  is the same for all birds, and (ii) the probability  $p$  is independent of time. In practice we might expect  $p$  to vary over a migration season. For example, towards the end of a spring migration season,  $p$  might be expected to decrease as birds become more urgent in their desire to reach the breeding grounds.

The distribution of  $I_n$  will be taken as a poisson distribution (an obvious choice as the data are counts), with parameter  $\lambda$ . This simplistic suggestion is based on the assumption that immigration is constant over time. We examine this assumption in greater depth later.

Under the two distributional assumptions, the model (6.1) is related to the bivariate poisson distribution. If we

consider the convolution of two poisson random variables, then the distribution of one variable conditional on the other is given by

$$\begin{aligned}
 P_r(X_n = x_n/x_{n-1}) &= \sum_{r=0}^{\min(x_n, x_{n-1})} P_r(r \text{ 'old' birds recorded}) \times \\
 &\quad P_r(x_{n-r} \text{ 'new' birds recorded}) \\
 &= \sum_{r=0}^{\min(x_n, x_{n-1})} P_r(Y_n = r/x_{n-1}) \cdot \Pr(I_n = x_n - r) \\
 &= \sum_{r=0}^{\min(x_n, x_{n-1})} \binom{x_{n-1}}{r} p^r (1-p)^{x_{n-1}-r} \frac{e^{-\lambda} \lambda^{x_n-r}}{(x_n-r)!}
 \end{aligned}$$

The likelihood function, given a series of n counts, is of the form

$$\begin{aligned}
 P_r(x_2, x_3, \dots, x_n/x_{n-1}) &= P_r(x_2/x_1) P_r(x_3/x_2) \dots P_r(x_n/x_{n-1}) \\
 &= \prod_{i=2}^n P_r(x_i/x_{i-1})
 \end{aligned}$$

The likelihood function involves two unknown parameters,  $\lambda$  and  $p$ , which can readily be found by an iterative method (explicit formulae for  $\lambda$  and  $p$  are not obtainable).

In the following section, we present some estimates of  $\lambda$  and  $p$  using samples of observatories data.

Note that this model has been used before in the study of ornithological data, although in another context - see Upton and Lampitt (1981).

### 6.2.3 Results and seasonal comparisons

The maximum likelihood parameter estimates presented in this section were obtained by using the iterative Nelder-Mead simplex method (Nelder and Mead, 1965). The estimates of the asymptotic variance-covariance matrix of the parameters were obtained using a FORTRAN subroutine written by S. Watts, formerly of the University of Kent.

We shall examine data for a number of species and from various observatories. The combinations of species, observatories and seasons that are considered in this section were selected since they provide suitable numbers of birds over several weeks each season. In some cases, species are recorded in only one of the two migration seasons and so therefore no spring-autumn comparisons are possible for these species. There follow below 10 tables which show, for each species - observatory - season combination, the estimates of  $p$  and  $\lambda$ , the estimated asymptotic standard errors and correlation coefficient, and the length of each sequence of counts.

In some tables, entries are missing. This is due to numerical problems encountered when attempting to estimate standard errors of the parameters. The estimation involves the inversion of the information matrix, which can sometimes be close to singular. In one case (Redstart, Spurn, *spring* of 1961) no birds were recorded at all.

Tables 6.2.1-6.2.10 Parameter estimates for Model(6.1)

Table 6.2.1 Sedge Warbler, Dungeness

Year	SPRING						AUTUMN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	.50	3.17	.04	.30	-.40	53	.52	1.90	.05	.22	-.36	57
62	.53	4.14	.03	.37	-.35	42	.26	2.90	.04	.23	-.55	96
63	.59	3.10	.04	.34	-.30	36	.48	5.15	.03	.34	-.46	70
64	.66	.69	.10	.17	-.18	27	.56	5.54	.02	.35	-.57	93
65	.27	.09	.30	.06	-.16	26	.39	2.89	.04	.25	-.49	75
66	.69	.98	.06	.20	-.47	45	.42	4.32	.03	.29	-.52	89
67	.56	2.18	.04	.23	-.31	59	.44	3.52	.03	.25	-.51	99
68	.73	.84	.07	.19	-.38	36	.48	3.87	.03	.29	-.47	75
69	.81	.28	.08	.09	-.17	44	.26	3.03	.04	.23	-.35	76
70	.25	.34	.16	.12	-.27	30	.13	5.12	.03	.30	-.48	83
71	.17	1.46	.07	.20	-.23	42	.22	2.09	.04	.17	-.41	97
72	.16	1.04	.06	.14	-.28	61	.22	.73	.08	.11	-.34	76

Table 6.2.2 Willow Warbler, Dungeness

Year	SPRING						AUTUMN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	.38	1.07	-	-	-	31	.44	4.30	.03	.30	-.42	71
62	.16	5.40	.03	.36	-.34	54	.56	6.00	.02	.36	-.45	77
63	.46	5.89	.03	.39	-.43	58	.68	6.07	.02	.38	-.40	64
64	.43	5.69	.04	.45	-.43	44	.52	7.94	.02	.42	-.48	75
65	.50	3.58	.04	.36	-.31	41	.57	4.90	.03	.33	-.42	70
66	.21	11.60	.03	.57	-.52	58	.45	6.99	.02	.37	-.47	86
67	.27	4.76	.03	.33	-.35	57	.42	4.27	.03	.26	-.38	91
68	.30	4.06	.04	.34	-.30	44	.66	5.54	.02	.29	-.45	107
69	.11	6.40	.04	.44	-.54	53	.42	6.61	.02	.34	-.42	85
70	.27	7.46	.03	.43	-.37	55	.34	8.87	.02	.43	-.45	73
71	.00	2.87	.02	.14	-.99	70	.45	8.43	.02	.41	-.41	73
72	.28	5.61	.02	.28	-.32	89	.45	4.37	.03	.29	-.37	73

$l$  : estimate of parameter  $\lambda$   
 $s(\hat{p})$  : standard error of parameter estimate  $\hat{p}$   
 $s(l)$  : standard error of parameter estimate  $l$   
 $c(\hat{p},l)$  : estimated correlation between  $p$  and  $l$   
 $n$  : length of migration season in days

Table 6.2.3 Whitethroat, Dungeness

Year	SPRING						AUTUMN					
	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$
1961	.12	14.50	.02	.62	-.49	56	.44	5.62	.03	.35	-.46	74
62	.50	5.69	.03	.43	-.36	43	.44	6.23	.03	.34	-.49	88
63	.46	6.58	.03	.46	-.35	43	.53	6.54	.03	.38	-.44	72
64	.46	4.78	.04	.40	-.37	42	.46	6.34	.02	.33	-.48	98
65	.48	4.17	.37	.04	-.06	26	.52	6.61	.02	.37	-.52	87
66	.23	8.54	.04	.55	-.63	57	.55	5.52	.02	.35	-.56	90
67	.49	3.78	.04	.33	-.37	50	.40	7.20	.03	.40	-.57	86
68	.46	8.60	.03	.56	-.42	42	.67	5.33	.02	.33	-.50	88
69	.33	.27	.16	.09	-.23	37	.48	.55	.07	.10	-.31	77
70	.35	1.37	.06	.18	-.31	54	.52	1.41	.04	.15	-.30	86
71	.05	1.69	.05	.18	-.40	63	.41	2.48	.03	.18	-.36	108
72	.37	.74	.12	.15	-.12	37	.46	1.30	.05	.15	-.41	89

Table 6.2.4 Redstart, Dungeness

Year	SPRING						AUTUMN					
	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$
1961	.06	1.86	.06	.22	-.40	48	.51	.58	.07	.10	-.18	69
62	.23	1.60	.08	.25	-.32	34	.15	.23	.13	.07	-.31	64
63	.20	2.27	.04	.22	-.26	58	.46	1.72	.04	.16	-.26	89
64	.03	1.36	.07	.18	-.53	58	.25	1.06	.05	.12	-.20	78
65	.20	4.32	.04	.38	-.35	39	.56	.67	.06	.11	-.16	70
66	.22	1.03	.07	.15	-.38	66	.33	1.35	.07	.15	-.25	69
67	.21	1.25	.07	.20	-.27	38	.63	.97	.03	.11	-.18	99
68	.55	.72	.09	.17	-.41	38	.46	1.82	.04	.16	-.41	100
69	.48	.03	.37	.03	-.05	32	.41	.70	.06	.10	-.15	84
70	.00	3.93	.00	.35	-.99	33	.22	.49	.09	.09	-.26	68
71	.15	1.25	.12	.25	-.46	29	.25	.27	.11	.07	-.18	59
72	.32	.76	.08	.13	-.30	55	.57	.59	.05	.08	-.21	108

Table 6.2.5 Redstart, Portland

Year	SPRING						AUTUMN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	.31	2.18	.05	.22	-.27	55	.22	.17	.16	.06	-.27	61
62	.45	1.21	.09	.23	-.46	37	.23	.21	.19	.06	-.19	72
63	.29	1.41	.06	.18	-.39	38	.49	.03	.36	.02	-.02	62
64	.00	1.79	.04	.19	-.33	57	.47	.22	.11	.06	-.19	63
65	.09	3.98	.05	.37	-.50	42	.33	.83	.07	.12	-.26	69
66	.40	4.57	.04	.39	-.39	43	.36	.78	.06	.10	-.38	101
67	.15	2.60	.08	.34	-.56	37	.64	.52	.05	.09	-.20	85
68	.43	1.11	.07	.20	-.27	37	.51	.76	.05	.10	-.22	92
69	.10	.43	.12	.10	-.33	50	.69	.26	.09	.06	-.12	68
70	.06	2.01	.05	.23	-.34	45	.53	.57	.06	.09	-.16	73
71	.29	1.33	.07	.17	-.46	70	.72	.23	.09	.07	-.05	54
72	.21	2.57	.06	.29	-.38	42	.70	.26	.09	.06	-.12	68

Table 6.2.6 Sedge Warbler, Portland

Year	SPRING						AUTUMN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	.50	.00	-	-	-	31	.10	.65	.08	.10	-.27	70
62	.23	.89	.08	.15	-.27	51	.15	.34	.11	.09	-.20	49
63	.04	1.69	.06	.23	-.34	37	.58	.13	.15	.05	-.09	57
64	.02	2.06	.04	.25	-.33	40	.37	.89	.08	.14	-.29	60
65	.29	.62	.10	.14	-.27	42	.23	1.58	.05	.17	-.36	71
66	.09	1.03	.08	.16	-.46	53	.29	1.38	.05	.15	-.31	75
67	.04	1.66	.04	.17	-.25	64	.29	2.23	.04	.19	-.26	75
68	.00	2.31	.04	.25	-.99	39	.39	1.49	.06	.17	-.36	69
69	.00	.12	.00	.05	-.99	44	.42	1.37	.05	.16	-.20	65
70	.22	1.26	.07	.18	-.26	46	.26	3.72	.03	.25	-.38	79
71	.16	.49	.10	.10	-.36	58	.29	1.10	.05	.11	-.26	110
72	.22	1.56	.05	.18	-.26	60	.46	2.50	.04	.21	-.35	75

Table 6.2.7 Redstart, Spurn

Year	SPRING						AUTUMN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	-	-	-	-	-	-	.41	1.18	.06	.14	-.39	83
62	.29	.39	.12	.11	-.25	41	.23	.11	.19	.04	-.14	61
63	.24	.24	.14	.08	-.21	48	.83	1.06	.02	.13	-.17	69
64	.07	.69	.09	.12	-.47	67	.17	.07	.19	.03	-.08	61
65	.21	.79	.09	.16	-.30	41	.73	1.08	.03	.13	-.17	72
66	.39	.37	.09	.08	-.24	72	.77	1.34	.02	.14	-.21	81
67	.37	.23	.15	.09	-.11	34	.70	.18	.09	.06	-.09	62

Table 6.2.8 Sedge Warbler, SPRING

Year	BARDSEY						SPURN					
	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n	$\hat{p}$	l	s( $\hat{p}$ )	s(l)	c( $\hat{p},l$ )	n
1961	.32	2.31	.04	.22	-.32	62	.54	.86	.07	.15	-.50	72
62	.38	4.49	.03	.34	-.31	52	.84	.09	.09	.05	-.07	46
63	.43	5.48	.03	.39	-.35	50	.36	.09	.28	.05	-.06	37
64	.44	2.28	.04	.21	-.29	65	.23	.47	.20	.09	-.38	73
65	.13	1.00	.07	.19	-.10	31	.61	.43	.08	.09	-.19	52
66	.30	6.52	.03	.38	-.37	62	.67	1.04	.05	.15	-.46	80
67	.18	4.38	.04	.37	-.26	39	.70	.66	.05	.12	-.31	62

Table 6.2.9 Pied Flycatcher, Autumn

Year	DUNGENESS						SPURN					
	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$
1961	.00	.06	.36	.04	-.41	53	.66	1.51	.04	.16	-.10	65
62	.39	.96	.12	.13	-.10	63	.28	1.54	.04	.16	-.15	70
63	.04	.40	.06	.09	-.16	52	.63	2.23	.02	.18	-.26	92
64	.35	1.61	.05	.17	-.22	69	.57	.55	.07	.11	-.19	56
65	.00	.02	.98	.02	-.28	52	.64	.67	.05	.12	-.11	56
66	.40	1.77	.05	.19	-.17	57	.73	2.28	.02	.19	-.27	87
67	.50	.00	-	-	-	44	.55	1.58	.04	.18	-.20	61
68	.36	1.90	.04	.18	-.27	73						
69	.12	.19	.16	.06	-.41	66						
70	.67	1.04	.04	.13	-.12	67						
71	.65	.00	-	-	-	43						
72	.33	.77	.07	.12	-.25	62						

Table 6.2.10 Blackcap, Dungeness

Year	SPRING					
	$\hat{p}$	$l$	$s(\hat{p})$	$s(l)$	$c(\hat{p}, l)$	$n$
1961	.19	.57	.09	.12	-.16	45
62	.31	1.97	.05	.22	-.24	48
63	.36	1.20	.06	.19	-.13	38
64	.28	1.18	.08	.20	-.32	37
65	.22	.44	.16	.13	-.15	30
66	.28	1.03	.07	.15	-.29	57
67	.29	.48	.11	.12	-.33	47
68	.15	.86	.11	.16	-.49	49
69	.20	.40	.21	.12	-.23	32
70	.03	1.56	.05	.19	-.37	50
71	.17	.46	.07	.09	-.22	72
72	.12	1.84	.06	.21	-.38	54

A number of general points arising from Tables 6.2.1 - 6.2.10 are listed below.

1. There is no convincing evidence that  $p$  is higher for autumn than for spring throughout the data sets considered. Indeed, the Dungeness *Sedge* Warbler data suggest the opposite (Table 6.2.2). There is some evidence to suggest that  $p$  is higher in autumn for the Redstart (Dungeness, Portland and Spurn), but even in these cases a wide range of values of  $p$  is found.
2. There is a wide range of values for  $\lambda$  in most tables. Note that the Dungeness Whitethroat data show an encouraging feature: the population crash of the winter of 1968 is clearly reflected in the estimates of  $\lambda$ , but it would be worrying if such a feature was not detected.
3. The estimated asymptotic standard errors of the parameters are fairly consistent both within and across the ten tables. The standard errors for  $p$  appear to be of the order of 10% of the estimate, and those for  $\lambda$  are slightly higher: approximately between 10% and 20% of  $\lambda$ .
4. In every case the estimated correlation coefficient between  $p$  and  $\lambda$  is negative, and in many cases greater

than 0.5 in magnitude. We might expect this since the more birds that we put into the category of 'new arrivals', then the fewer that are put into the category of birds that have stayed at the observatory from previous days.

The main interest in modelling daily counts in this manner is the possibility of constructing migration indices using information obtained from the models. Therefore in section 6.2.6 we make further use of the results presented in this section.

#### 6.2.4 Goodness of fit

In this section we show that model (6.1) provides an adequate description of the data, by using monte-carlo simulation. This method of assessing the goodness of fit of the model was chosen since other methods are difficult to apply in this situation. For example, the Pearson goodness-of-fit test requires that observations are serially independent. This is clearly not the case if model (6.1) is a suitable model, since it has first-order Markov dependence. In some situations, however, it is possible to use Pearson's test with serially correlated data. Tavaré and Altham (1983) show that if one can estimate the matrix of transition probabilities then one is able to adjust the usual Pearson-statistic, and assess its significance. The problem with our data is that each individual sequence of counts is very short (often less than 50 observations), and therefore one is unable to estimate transition probabilities unless

counts are grouped into, say, two or three groups, thus losing much of the day-to-day variability in the data.

Another problem with the Pearson goodness of fit statistic in this context is that the test statistic may be dominated by one or two observations and therefore give a significant result while the model gives a good fit for the rest of the data. This problem stems from the fact that if an observation is particularly larger than its predecessor, then its expected value is likely to underestimate the observed value. In other words, we cannot expect the model to be able to predict sudden large fluctuations in bird counts.

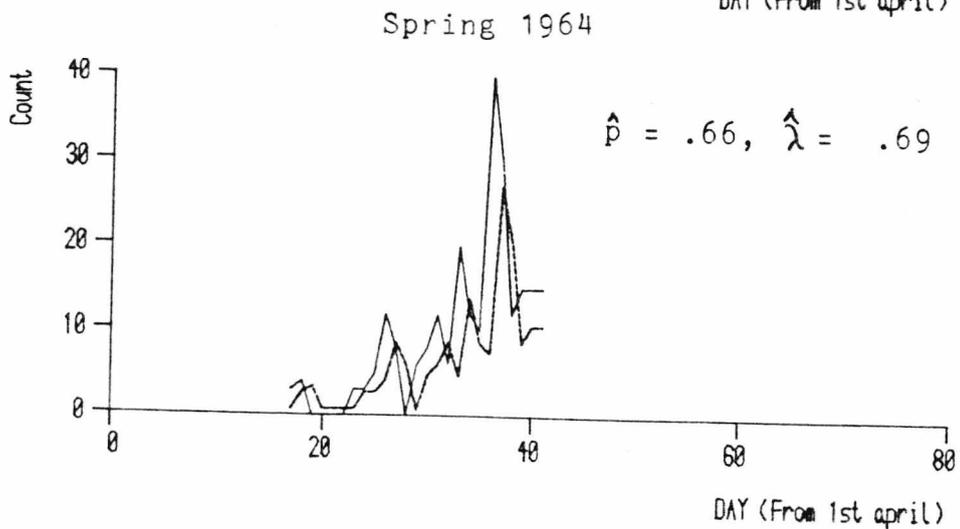
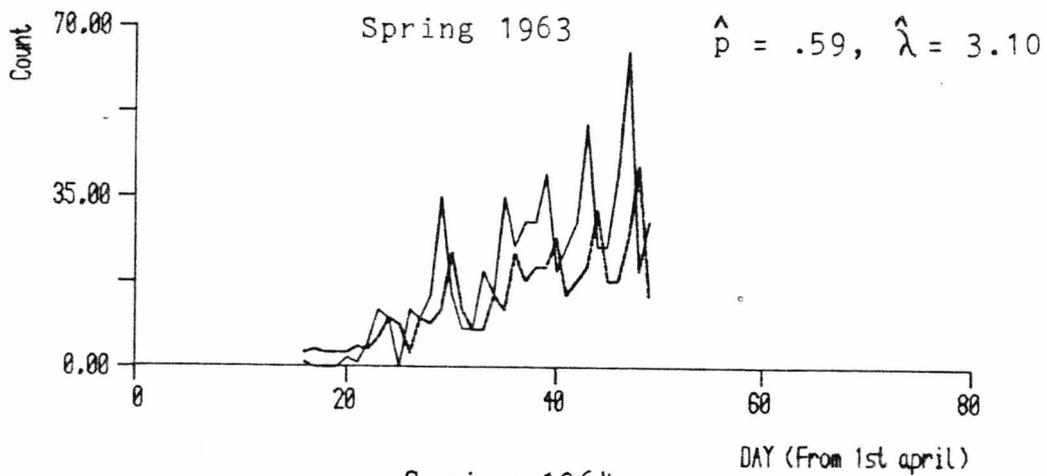
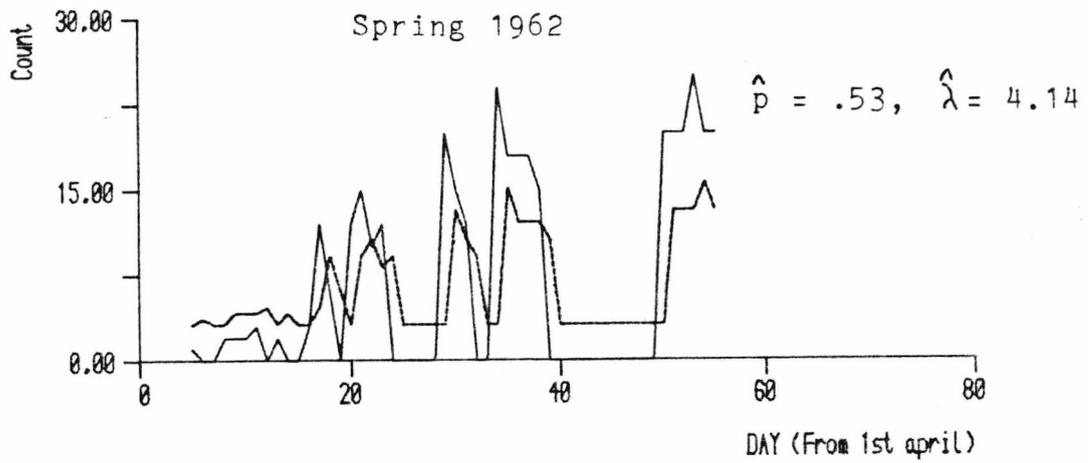
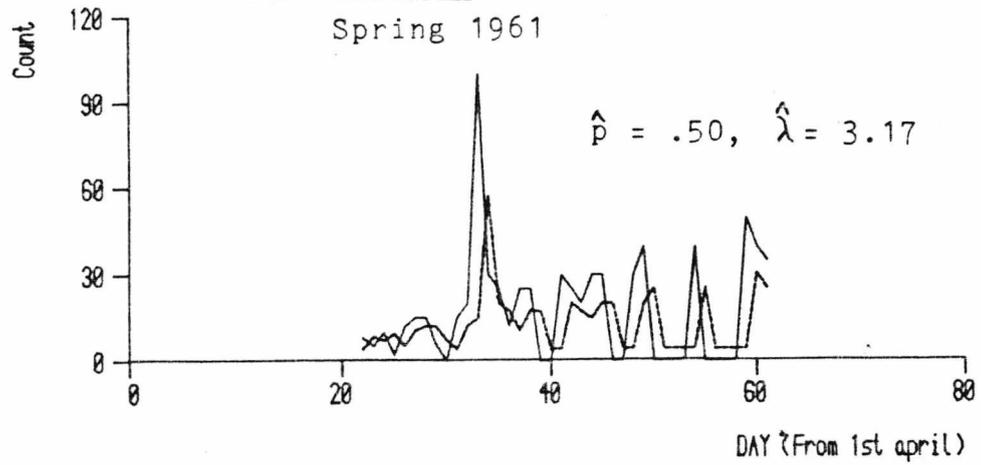
A subjective but helpful method of checking the goodness of fit is to plot the observed and expected counts of migrants against time, and compare the two curves. Some examples are shown in Figs. 6.2.4.1 and 6.2.4.2. As one would expect, when  $p$  is low, the curve of expected counts is almost constant over time, while when  $p$  is high, greater variability is seen. In the latter case, one can clearly see a lag of one day between the two curves.

The residuals (Dungeness data) are plotted against time in Figs. 6.2.4.3. In some diagrams, for example 1962, 1963 and 1967, runs of negative residuals are seen. These coincide with periods when few or no birds were observed, although the model is still giving a constant daily input. Large falls of birds are seen (1961 and 1965) as large

positive residuals, and in most diagrams the variability of the residuals tends to increase towards the end of the season. This last point suggests that this very simple model is not sufficient to explain the counts over a whole migration season, and that one must either consider a more complex model or perhaps introduce information on local weather conditions - see chapter 7.

The residuals for the Portland data are shown in Fig. 6.2.4.4. Although there is no evidence of non-stationarity in the residuals, one again sees long runs of negative residuals where no or very few birds are seen. One again sees large positive residuals which coincide with falls of migrants. The diagram for 1967 is an extreme example of these last two features.

Fig. 6.2.4.1 Model(6.1) expected values (dashed curve) and observed values (solid curve). Dungeness Sedge Warbler data.



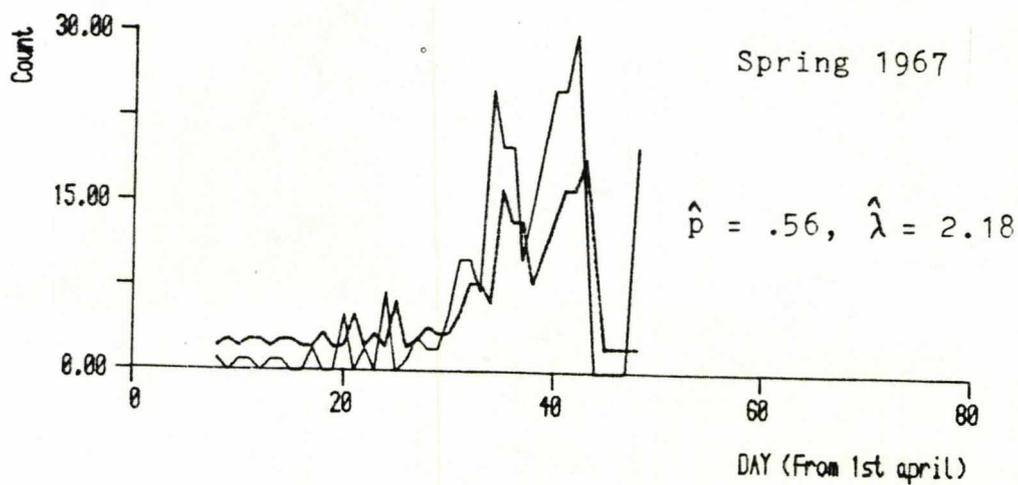
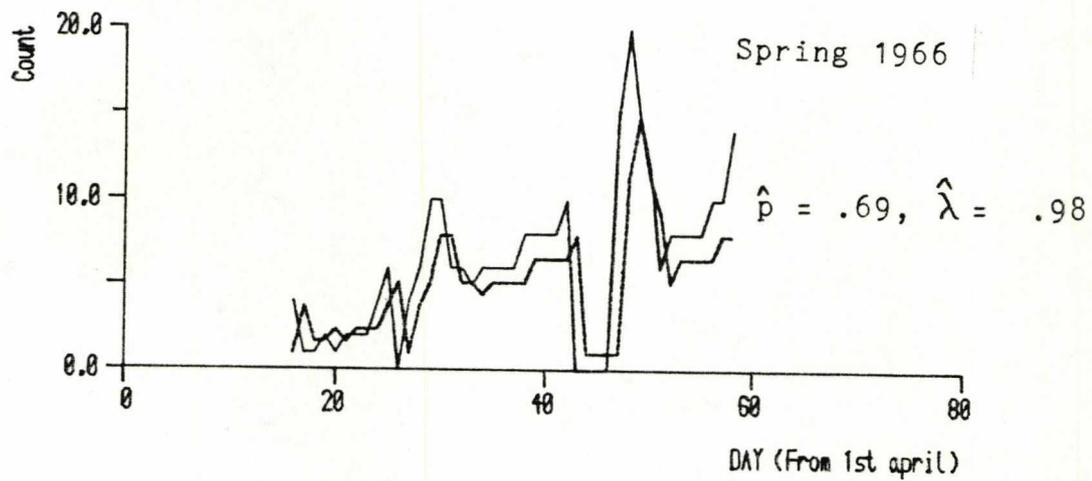
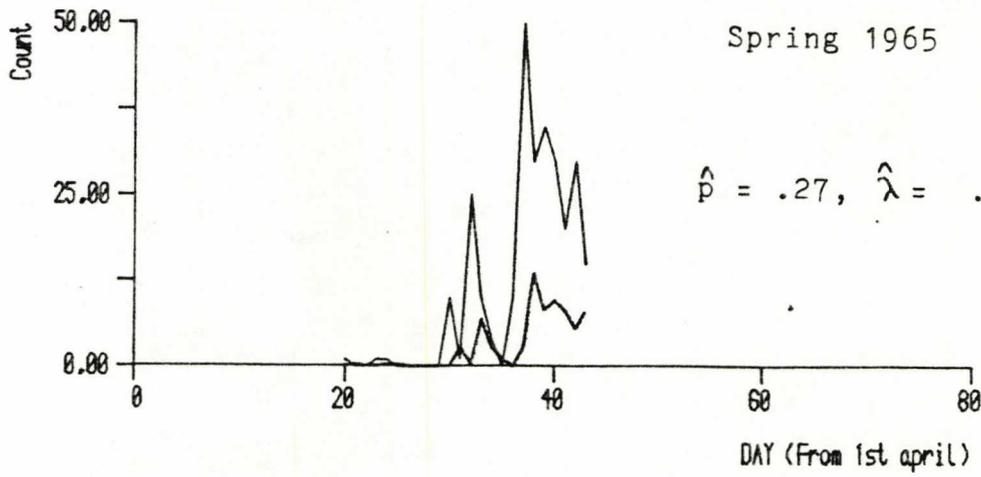
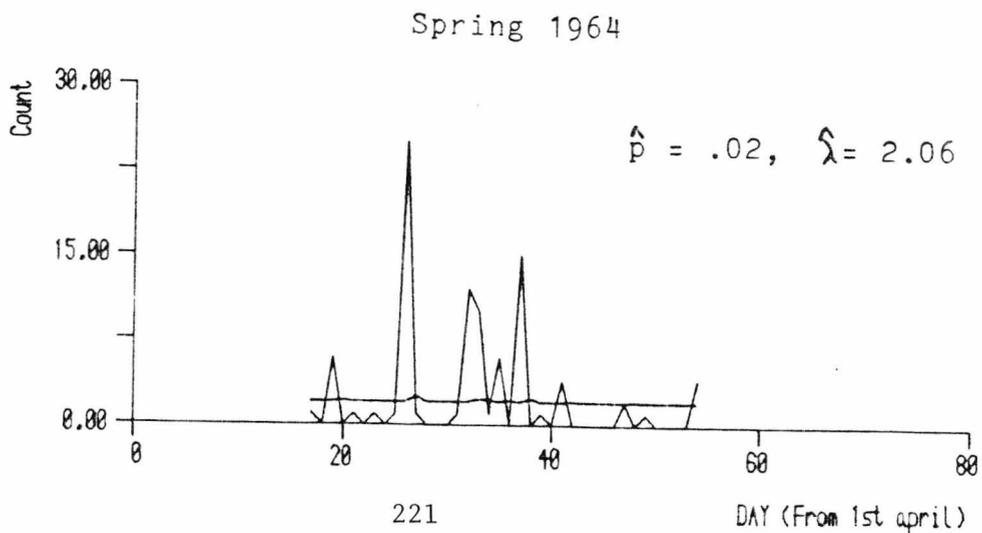
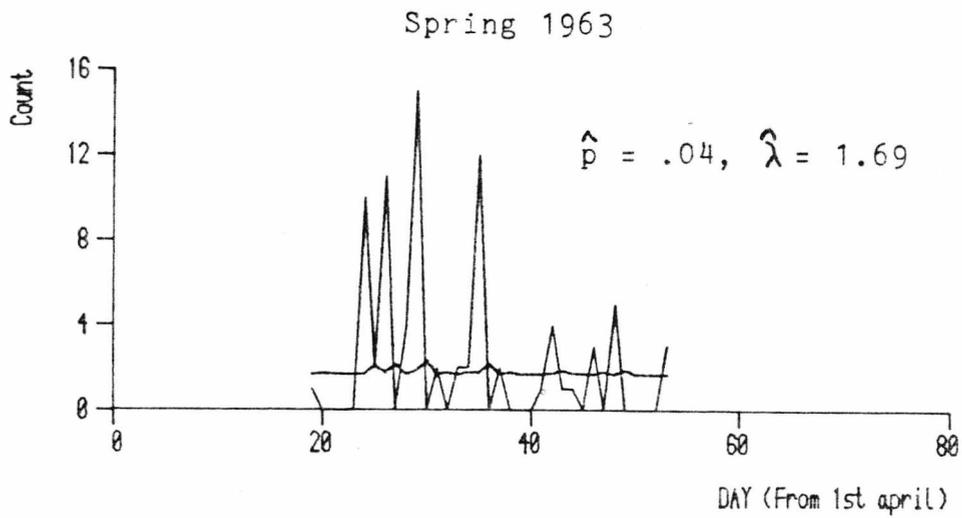
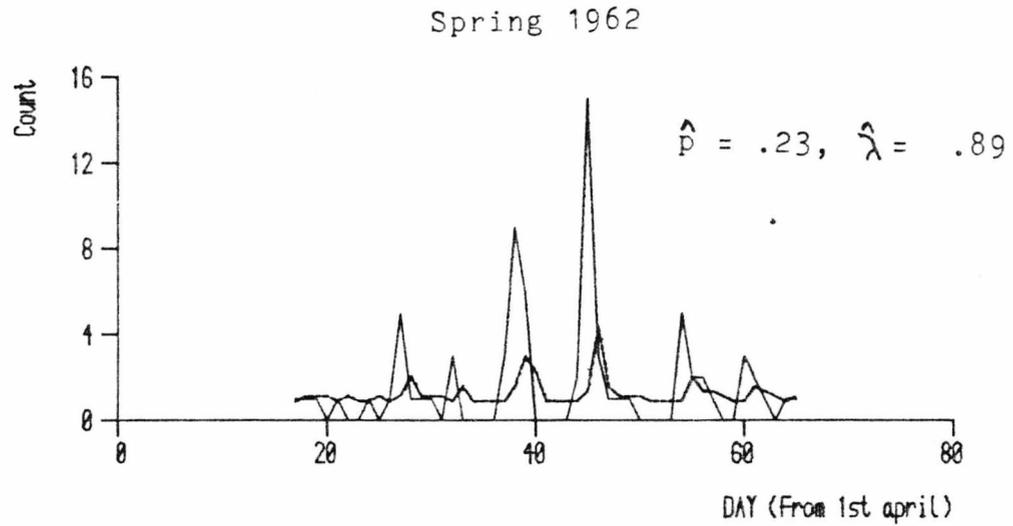


Fig. 6.2.4.2 Model(6.1) expected values (dashed curve) and observed values (solid curve). Portland Sedge Warbler data.



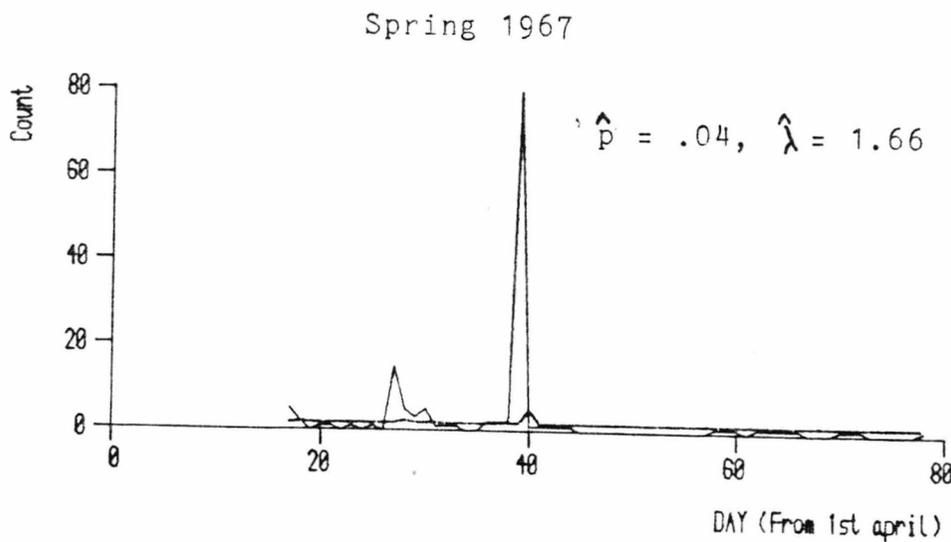
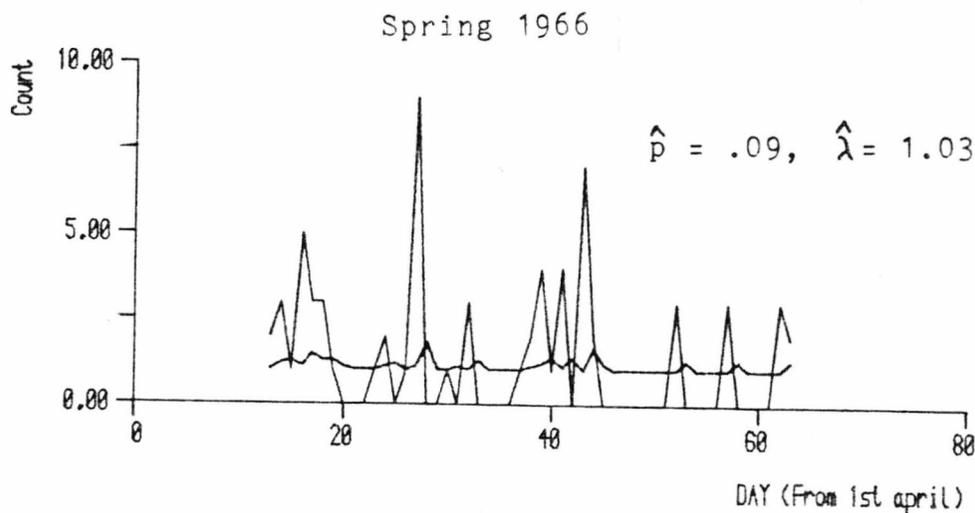
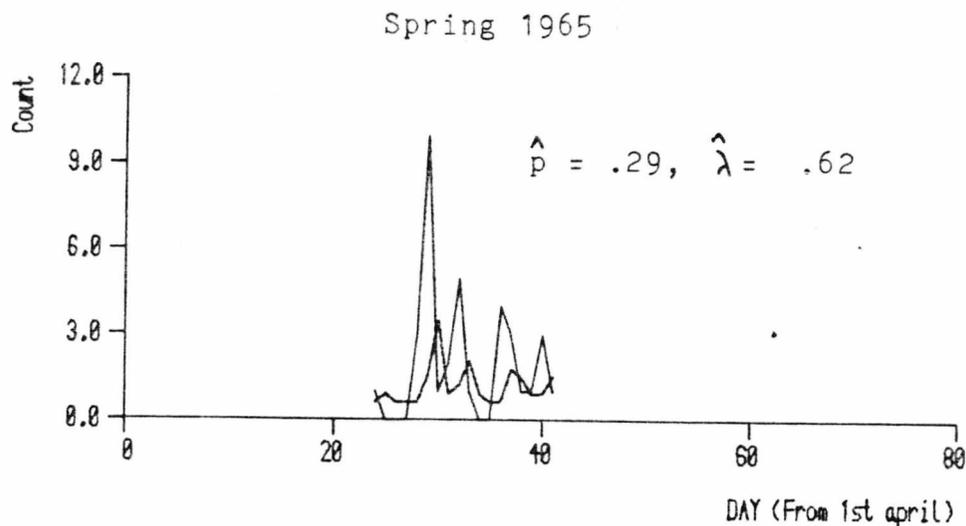
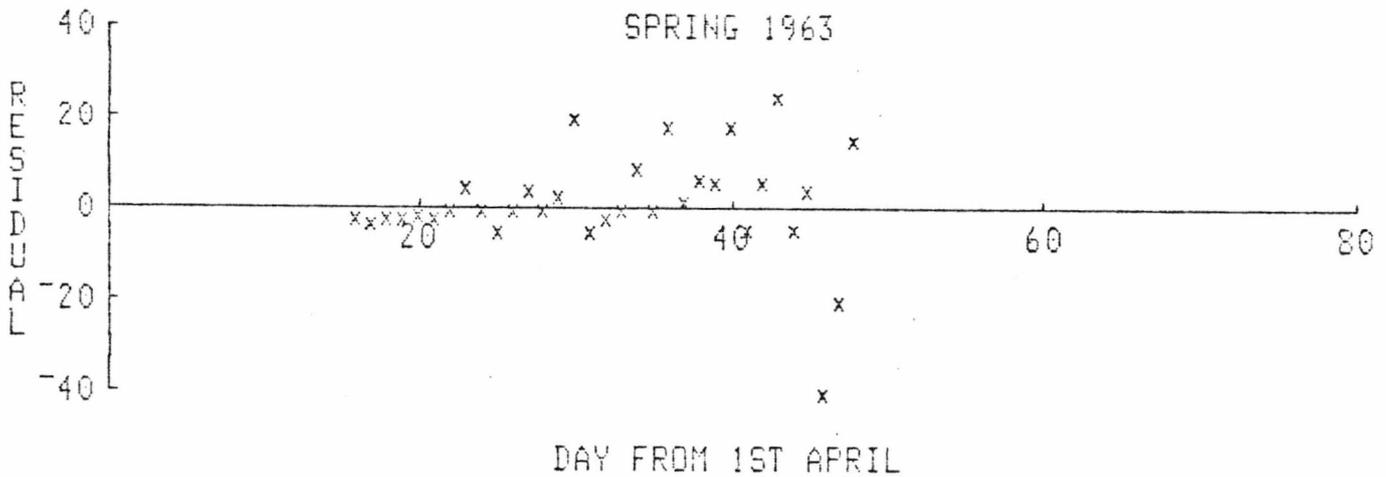
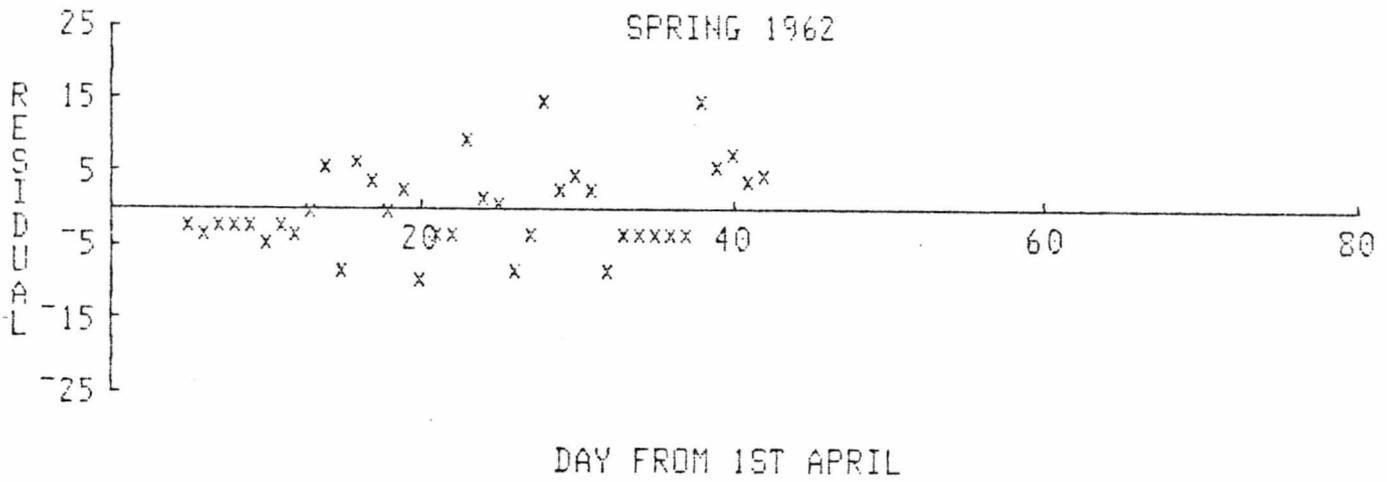
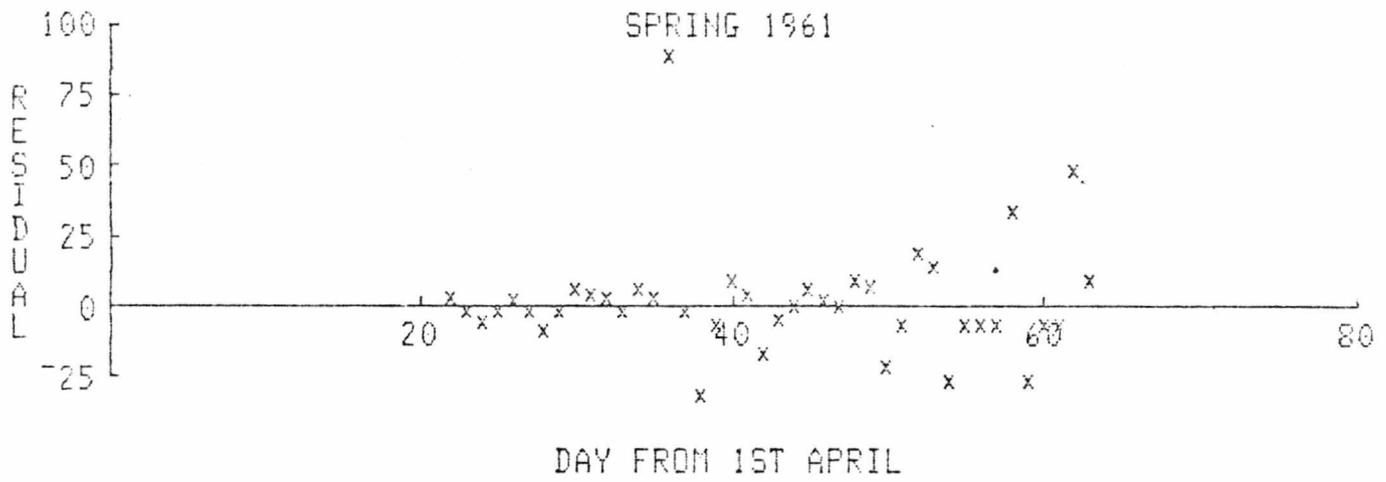


Fig. 6.2.4.3 Model(6.1) Residuals - Dungeness Sedge Warbler data.



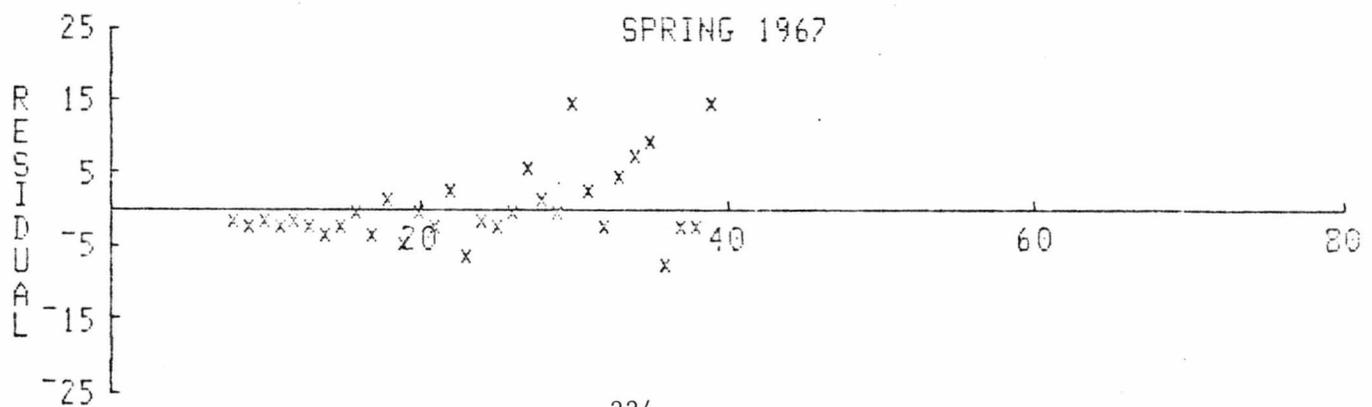
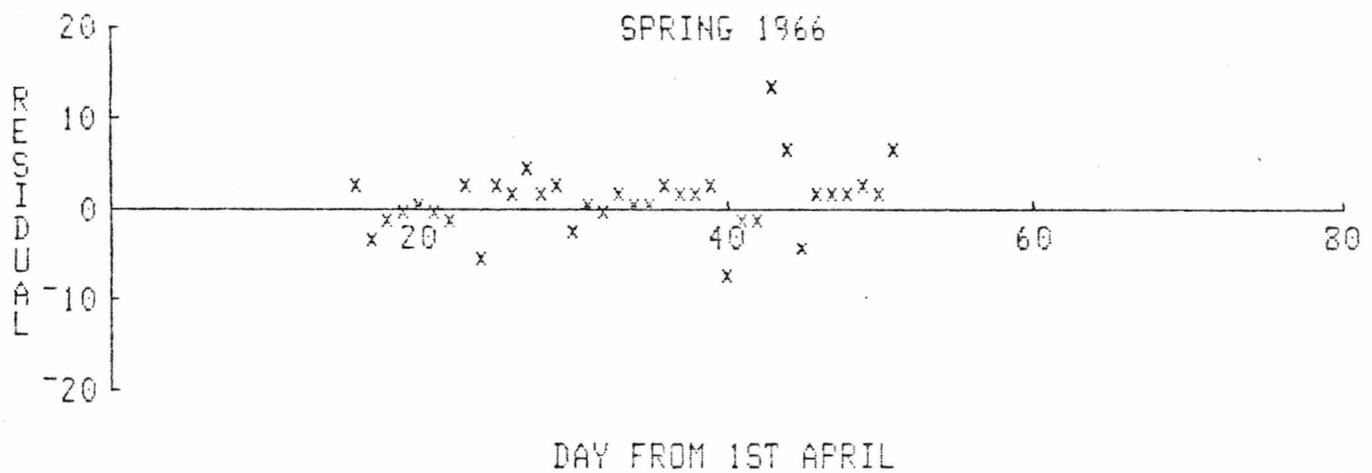
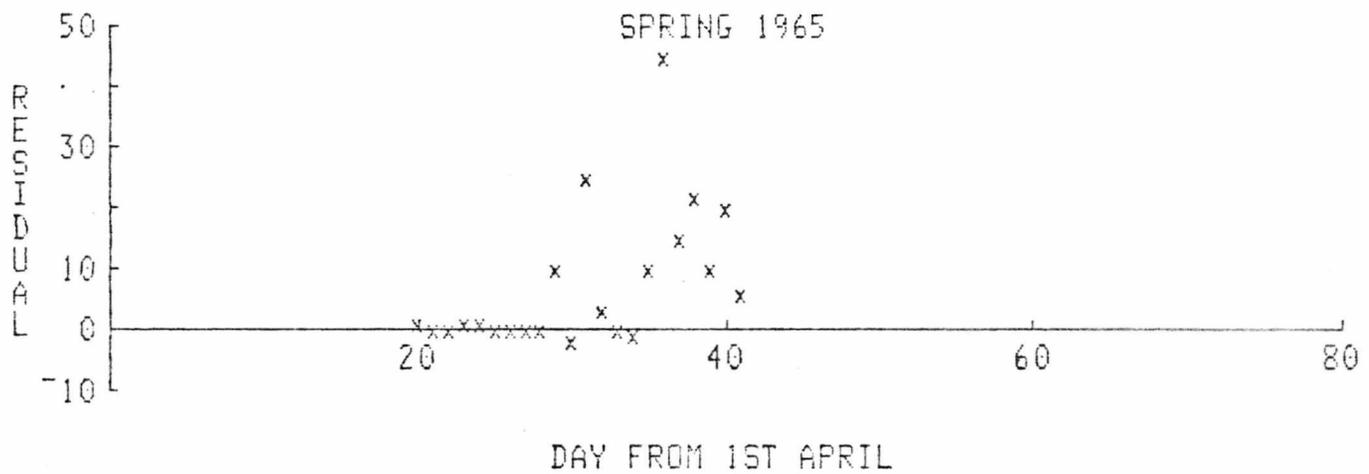
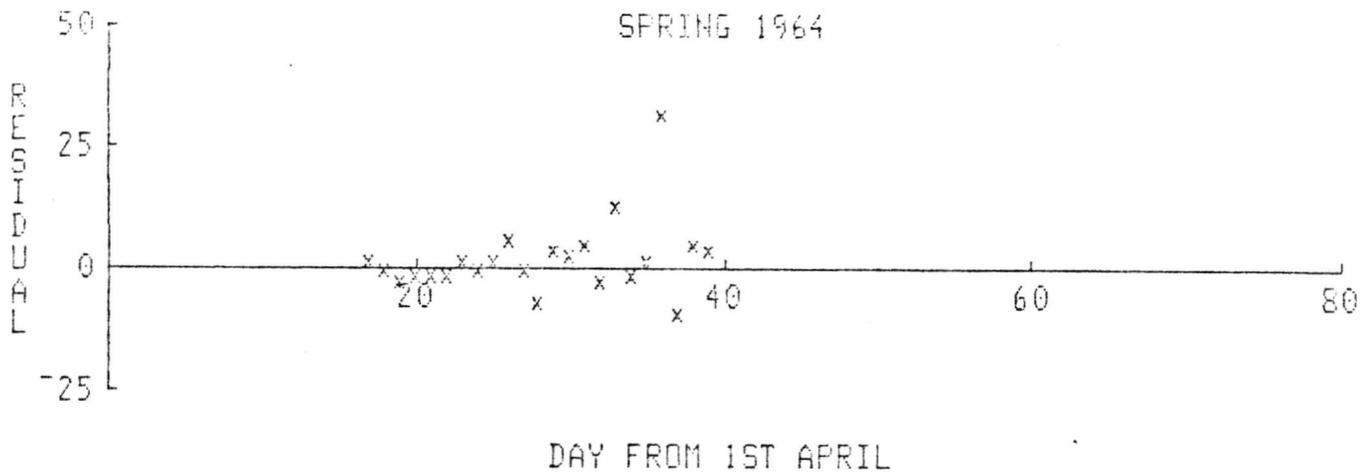
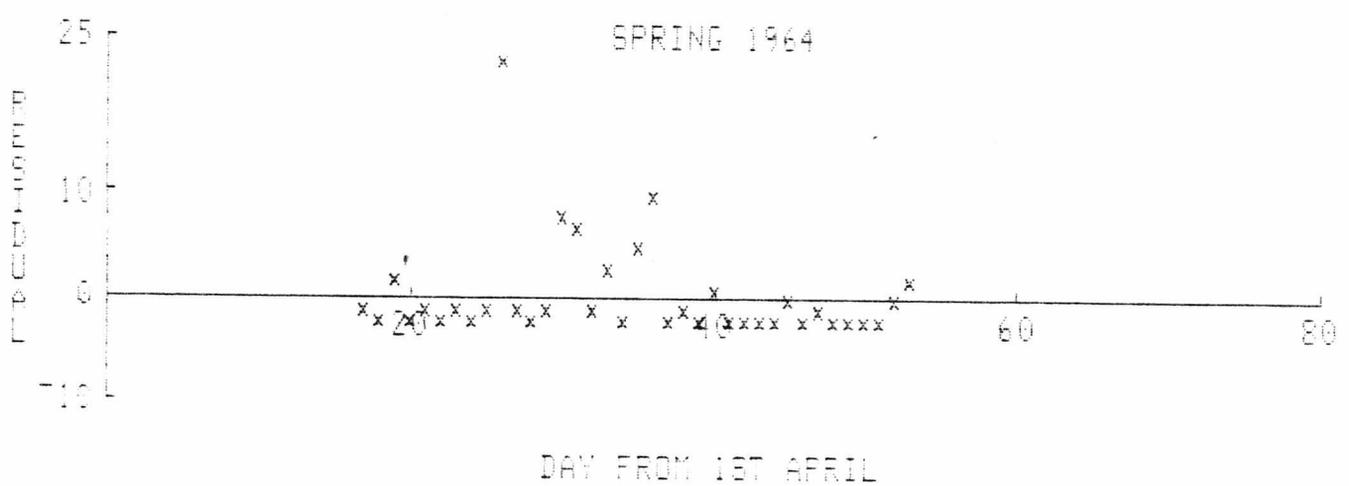
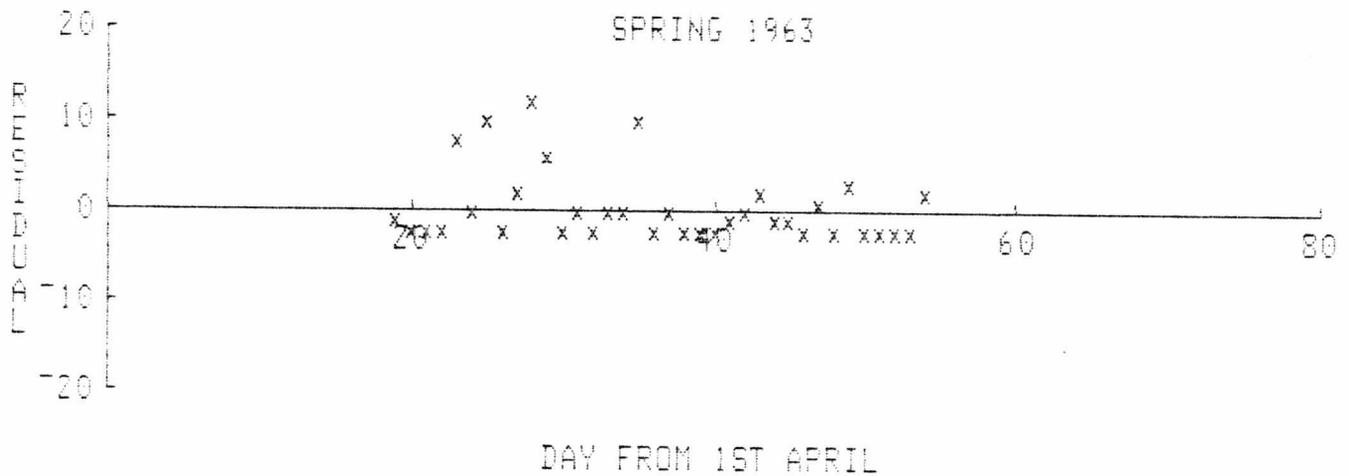
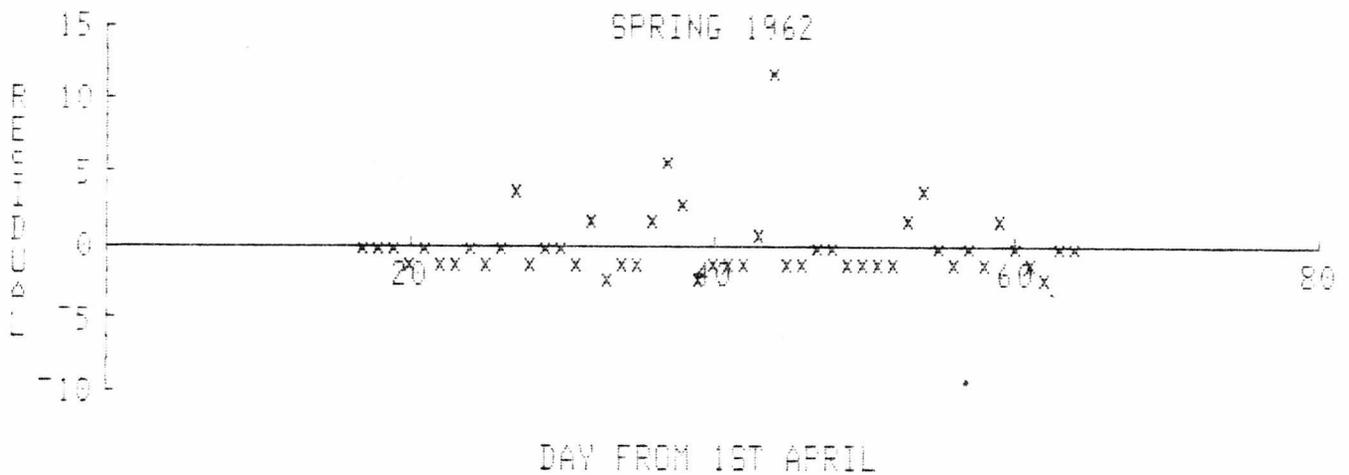
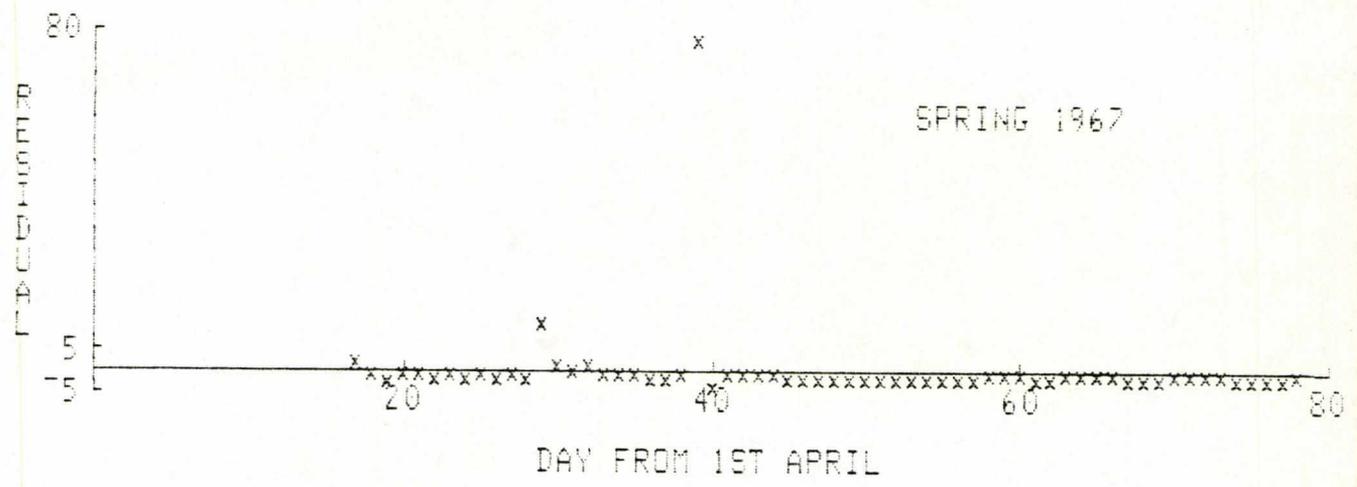
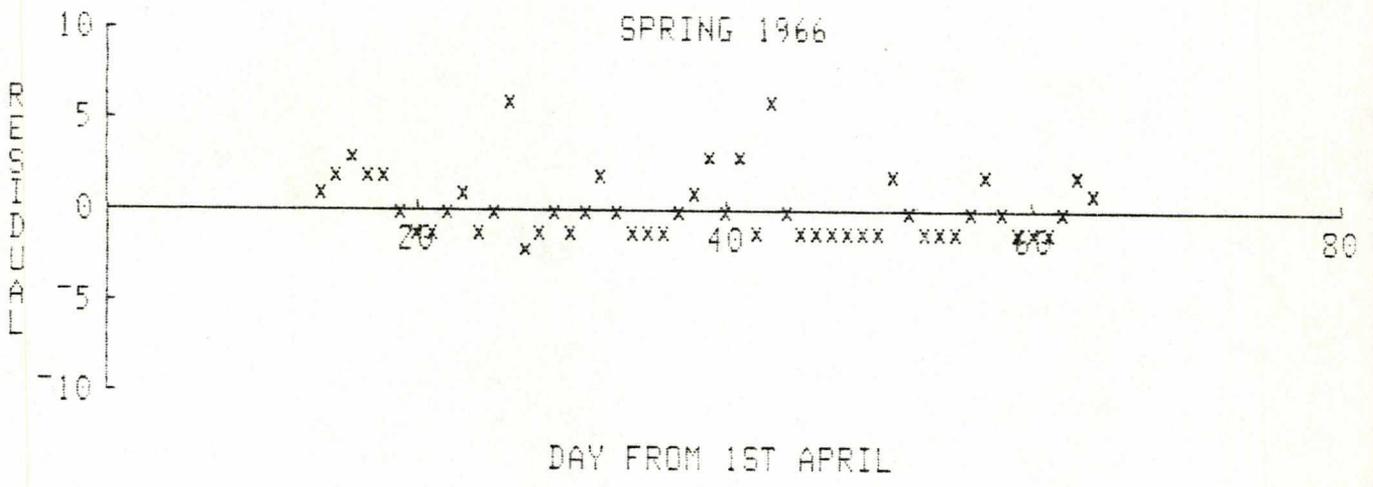
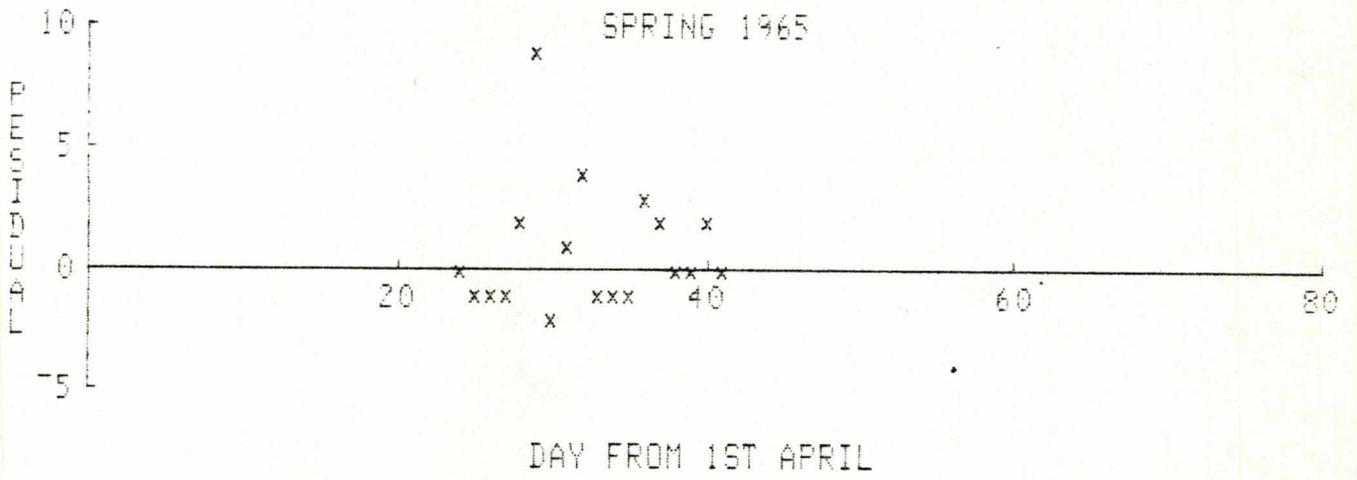


Fig. 6.2.4.4 Model(5.1) Residuals - Portland Sedge Warbler data.





We now examine the goodness of fit of model (6.1) by using the Pearson test, but, as we are unsure what the true distribution of our test statistic is under the null hypothesis, we use monte-carlo simulation to test its significance. The statistic that we test is  $X^2 = \sum (O-E)^2/E$  where  $O$  and  $E$  are observed and expected frequencies of transition between specified states. For model (6.1) the 'states' are daily counts of the order of 0-5, 5-10 and  $> 10$  birds. These groups were selected since if one uses more than 3 groups then matrices of transition probabilities soon become sparse and transition frequencies are very low. If one uses only 2 groups, then one is dispensing with much of the information in the data. The group sizes listed above gave, for most sequences of counts, transition frequencies of at least 3. For some sequences, however, the shortage of birds over the entire season led to transition frequencies of zero or unity in all but the category of  $\{0-5\} \rightarrow \{0-5\}$ .

The test procedure is carried out by calculating transition probabilities using the ML estimates of  $p$  and  $\lambda$ , and using these probabilities to calculate expected transition frequencies. The expected frequencies are compared with the observed frequencies, and the statistic  $X^2$  computed. Simulation is used to generate a large no. of sequences of counts of the same length as the observed sequence, and using  $p$  and  $\lambda$ . For each generated sequence,  $X^2$  is computed, and by establishing the position of the observed  $X^2$  in the distribution of simulated values of  $X^2$ , one can estimate

the probability of observing a value at least as high as  $\chi^2$  under the model. If this probability is less than 0.05, say, then there is evidence that the model is not appropriate for that particular sequence of counts. For this study, 1000 simulations of each sequence of counts were used.

The results of the goodness of fit study are presented in Table 6.2.11.

Table 6.2.11

Table entries are P-values for  $\chi^2$ .

Year	SWD(S)	SWD(A)	SWP(S)	SWP(A)	SWB(S)	SWS(S)
1961	0.49	0.63	<0.01	0.10	0.61	0.41
62	<.01	0.40	0.31	0.24	<.01	<.01
63	0.44	0.41	0.98	0.48	<.01	0.04
64	0.30	0.66	0.40	0.71	0.33	0.71
65	0.70	0.71	0.49	0.60	<.01	0.70
66	0.06	0.82	0.10	0.01	<.01	0.60
67	0.95	0.09	0.11	0.51	0.26	0.52
68	0.18	0.71	0.01	0.09	-	-
69	0.02	0.90	0.06	0.02	-	-
70	0.19	0.87	0.06	0.19	-	-
71	0.30	0.05	0.21	0.71	-	-
72	0.43	0.14	0.39	0.45	-	-

Year	WWD(S)	WWD(A)	WTD(S)	WTD(A)	PFD(A)	PFS(A)
1961	0.80	0.71	0.41	0.84	<0.01	0.01
62	0.74	0.49	0.29	0.79	0.31	0.30
63	0.99	0.50	0.31	0.50	<0.01	0.41
64	0.71	0.79	0.44	0.39	0.42	0.51
65	0.60	0.40	0.31	0.01	0.56	0.09
66	0.99	0.33	0.70	<.01	0.81	0.17
67	0.94	0.91	0.49	0.56	0.03	0.20
68	0.14	0.84	0.61	0.16	0.50	-
69	0.98	0.31	0.68	0.01	0.10	-
70	0.71	0.39	0.11	0.09	0.61	-
71	0.90	0.51	0.09	0.20	<0.01	-
72	0.82	0.18	0.33	0.71	0.17	-

Year	RSD(S)	RSD(A)	RSP(S)	RSP(A)	RSS(S)	RSS(A)
1961	0.12	0.48	0.51	0.05	0.10	0.21
62	0.40	0.31	0.61	0.14	0.01	0.01
63	0.31	0.40	0.08	0.02	0.04	0.19
64	0.32	0.21	<0.01	0.38	<0.01	0.61
65	0.75	0.30	0.07	0.14	0.09	0.01
66	0.25	0.70	0.28	0.25	0.01	0.08
67	0.30	0.04	0.37	0.39	0.28	0.12
68	0.13	0.14	0.77	0.29	-	-
69	0.01	0.81	0.03	0.31	-	-
70	0.04	0.03	0.04	0.01	-	-
71	0.28	0.41	0.90	0.09	-	-
72	0.23	0.32	0.85	0.14	-	-

Year	BCD(S)
1961	0.31
62	0.42
63	0.38
64	0.41
65	0.51
66	0.49
67	0.71
68	0.61
69	0.04
70	<0.01
71	0.51
72	0.09

We find that for the majority of sequences of counts, the model appears to fit the data. Approximately 18% of the calculated P-values are less than 0.05.

Table 6.2.11 shows that the model fits more sequences of counts of common migrants (e.g. Willow Warbler, Whitethroat and Sedge Warbler) than less common migrants such as Pied Flycatcher and Redstart. An examination of the parameter estimates  $p$  and  $\lambda$  in conjunction with Table 6.2.11 suggests that when either or both of  $p$  and  $\lambda$  are close to zero, then the fit is poor.

Overall, there is no strong evidence to reject model (6.1), although one would have hoped for more consistency over a range of different sequences of counts. If one is seeking a universal method of obtaining migration indices from sequences of daily counts, then one might be reluctant to use a method based on model (6.1) since there will be years in which the model does not fit the data, therefore leading to an incorrect migration index.

In section 6.2.6 we see that migration indices based on model (6.1) show little or no correlation with CBC indices.

#### 6.2.5 Alternative models

We have seen that model (6.1) is a suitable model for most sequences of daily counts that are considered in this thesis. Other models were also examined, and in this section we summarise the findings.

Model (6.1) assumes a constant parameter  $p$  over time. A model was considered which allowed  $p$  to switch between two values from day-to-day, the values being estimated from the data. This model (model (6.2)) is such that the binomial parameter  $p$  follows a two-state Markov chain with a transition matrix which is also estimated from the data. Parameter estimates were obtained by the Nelder-Mead simplex method. In chapter 5 we found that it was possible to divide bird counts into two groups on the basis of their weather observations. With this model we divide days into two groups on the basis of the serial correlation structure of the bird counts. In the next chapter we examine this connection in more detail. Model (6.2) was fitted to the same sequences of counts as model (6.1), and the parameters were estimated with standard errors of the same order as for model (6.1). In most cases one of the binomial probabilities was estimated to be close to zero and the other one close to unity. This is what we may have expected since ornithologists are of the opinion that, depending on the weather, either all the birds will leave the observatory

as soon as possible, or they will all stay. It is thought unlikely that some birds will choose to stay while others leave.

One interesting feature of the examination of model (6.2) parameter estimates is that the values of  $\lambda$  for model (6.2) are, in most cases, higher than for model (6.1). This suggests that model (6.2) is attributing more birds to the category of 'movers' and fewer birds to the category of 'stayers'. This is a consequence of allowing more freedom in the choice of value for  $p$ , i.e. a consequence of including extra parameters in the model.

There seems to be little more information to be gained from fitting model (6.2) rather than model (6.1), but it is encouraging that the behaviour of birds at observatories suggested by ornithologists is supported by statistical modelling. An extensive study of goodness of fit of model (6.2) was not carried out, but some examples of plots of observed and expected counts were compared with those for model (6.1), and there seems to be little evidence to suggest that either model gives a significantly better fit than the other.

We now consider how we might relax the assumption of model (6.1) that the parameter  $\lambda$  is independent of time. This seems to be a sensible course of action since we would not expect the same number of new arrivals on each night of the migration season. The diagrams of daily counts presented

in earlier chapters show a variety of 'patterns' of migration, but ideally we would like a method that could deal with the general case rather than specific examples. One could, for example, fit a polynomial to a series of  $\lambda$ 's, where

$$\lambda_n = a + \beta_1 n + \beta_2 n^2 + \beta_3 n^3 + \dots,$$

(where  $n$  refers to the day of the season). The parameters  $a$  and  $\beta_1, \beta_2, \dots$  could be estimated from the data using an iterative maximum likelihood procedure, but an obvious problem here is how high a degree of polynomial should one fit? Also, fitting a large number of parameters introduces more problems such as low precision in the estimators (a consequence of trying to extract too much information from too few data) and the need for suitable starting values for the iterative estimation procedure (Nelder-Mead simplex method).

A possible way round these problems is to take the following approach. We consider the Poisson parameter  $\lambda$  to be a linearly increasing function of time at the start of the migration season, until some time  $t_1$ , when  $\lambda$  is constant for the central part of the season. After some time  $t_2$ ,  $\lambda$  is considered to be a linearly decreasing function of time until the end of the season, i.e. the day after the last bird was recorded. The values of  $t_1$  and  $t_2$  can be estimated by trying various combinations and choosing that combination giving the maximum likelihood. Even by estimating  $t_1$  and  $t_2$  in this way, the amount of computer time required to fit the model is very large, and so only

a few results are available. Table 6.2.12 shows some results using the Dungeness Sedge Warbler data (spring). The last two columns of the table give the total estimated number of new arrivals seen in each season for the basic model (model 6.1) and the revised model (model 6.3). Clearly there are substantial differences. Another striking feature about these results is that the values of  $p$  are much lower under the revised model than under model (6.1) - see Table 6.2.1. This suggests that serial correlation in the data is not as high as suggested by the original model. This is a common feature of time series data where serial correlation is found in data sets with trend, but the corresponding detrended series exhibit different patterns of autocorrelation.

Table 6.2.12      Parameter estimates for model with varying  
 $\lambda$ . Dungeness Sedge Warbler data (spring)  
 $t_1$  and  $t_2$  are numbers of days from 1st April

Year	n	p	$t_1$	$t_2$	$n\lambda$	$n\lambda$ (model 6.1)
1963	36	.00	31	32	643	112
64	27	.08	21	24	227	19
65	26	.24	21	23	195	2
66	45	.12	32	45	310	44
67	59	.52	33	34	205	129
68	36	.08	30	32	187	30
69	44	.07	38	43	95	12

For the Sedge Warbler data at Dungeness, the two cut off points  $t_1$  and  $t_2$  appear to be close together, suggesting a triangular shape to the spring migration season - see Fig. 6.2.5.4. One would expect a different shape to the

autumn migration season, which is usually much longer than the spring season. This feature is clearly seen in an example using Whitethroat Dungeness data - see Fig. 6.2.5.1.

Figs. 6.2.5.2 to 6.2.5.5 show the 1963 (spring) Dungeness Sedge Warbler together with fitted values from models (6.1) and (6.3). Also shown are the residuals. Fig. 6.2.5.2 shows fitted values for model (6.1), and the one-day lag between fitted and actual values is clearly seen.

Fig. 6.2.5.4 shows the actual counts and the curve representing the expected new arrivals. The residuals for model (6.3) show much the same pattern as those for model (6.1), but are, in general, smaller in magnitude. There is some evidence therefore, but no convincing evidence that model (6.3) gives a better fit than model (6.1).

Unfortunately it is not practicable to carry out a more extensive analysis of model (6.3), and this, in itself, suggests that it is not worth pursuing this model here.

#### 6.2.6 Migration indices from model (6.1)

The object of fitting model (6.1) to sequences of daily counts is to obtain indices of migration volume by eliminating counts of birds observed more than once. The number of newly arrived birds each day is estimated by  $\lambda$  which is assumed constant throughout a migration season. A simple migration index is given by  $n\lambda$  where  $n$  is the number of days comprising that particular season.

Fig. 6.2.5.1 Example of fitting Model(6.3)

1965 WHITETHROAT (DUNGENESS)

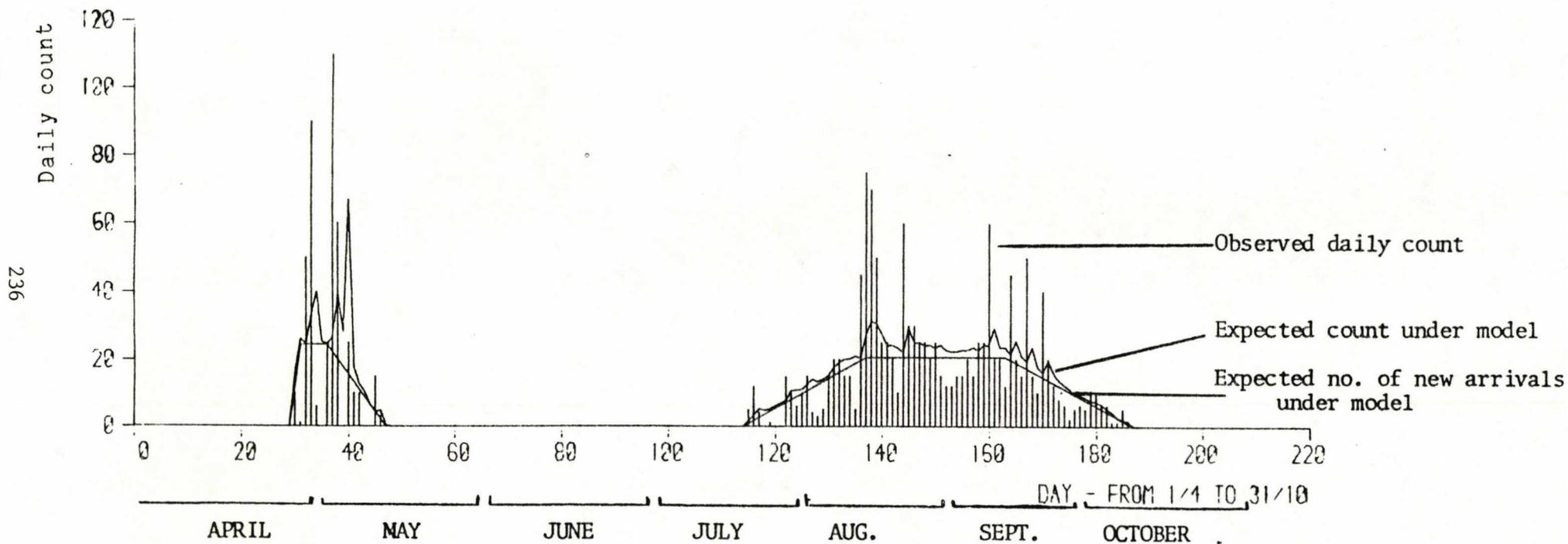


Fig. 6.2.5.2 Model(6.1) expected values and observed values

Dungeness Sedge Warbler data, Spring 1963.

$p = .59, \quad = 3.90$

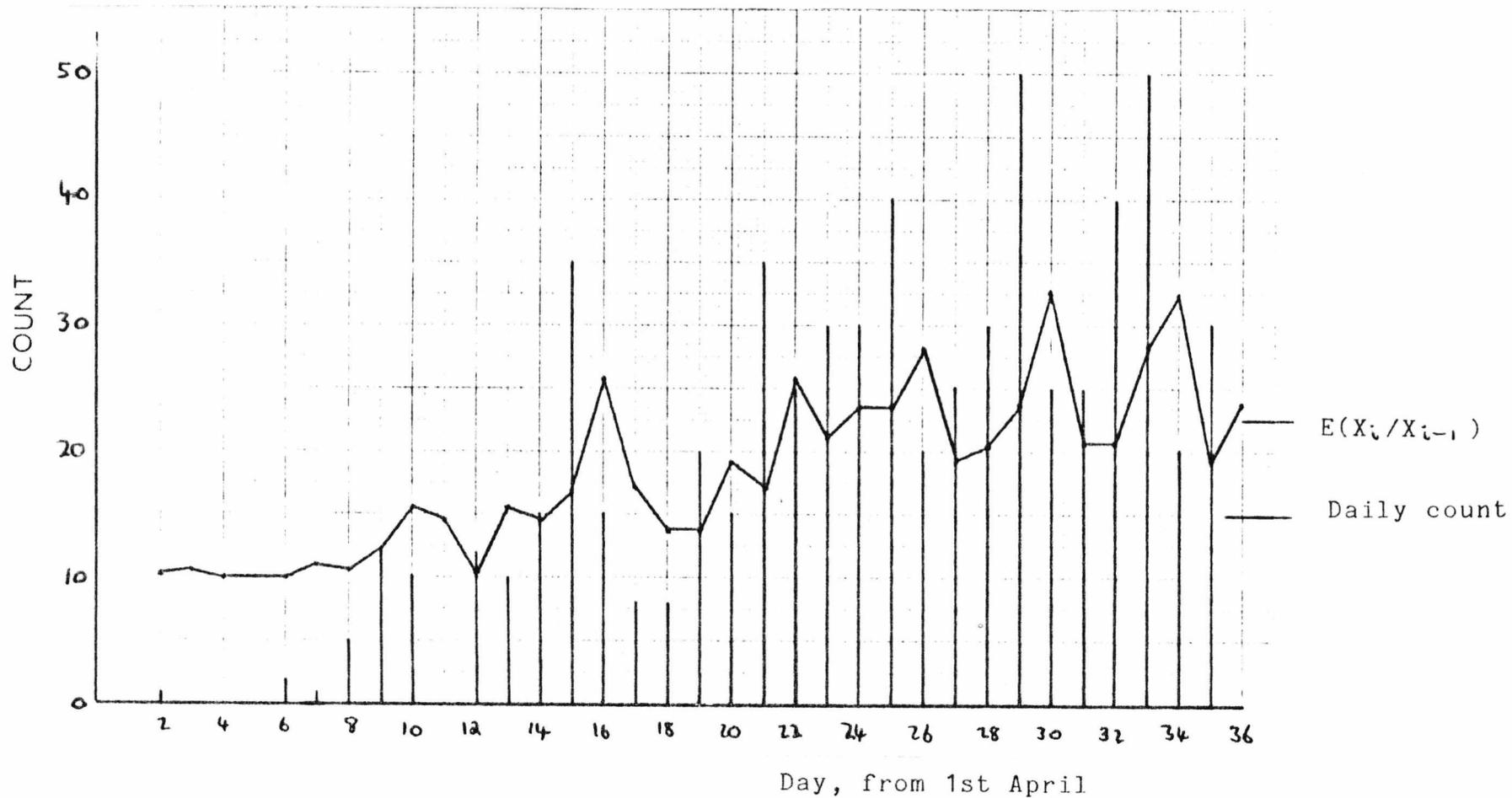


Fig. 6.2.5.3 Model(6.3) Residuals, Dungeness Sedge Warbler  
data, Spring 1963.

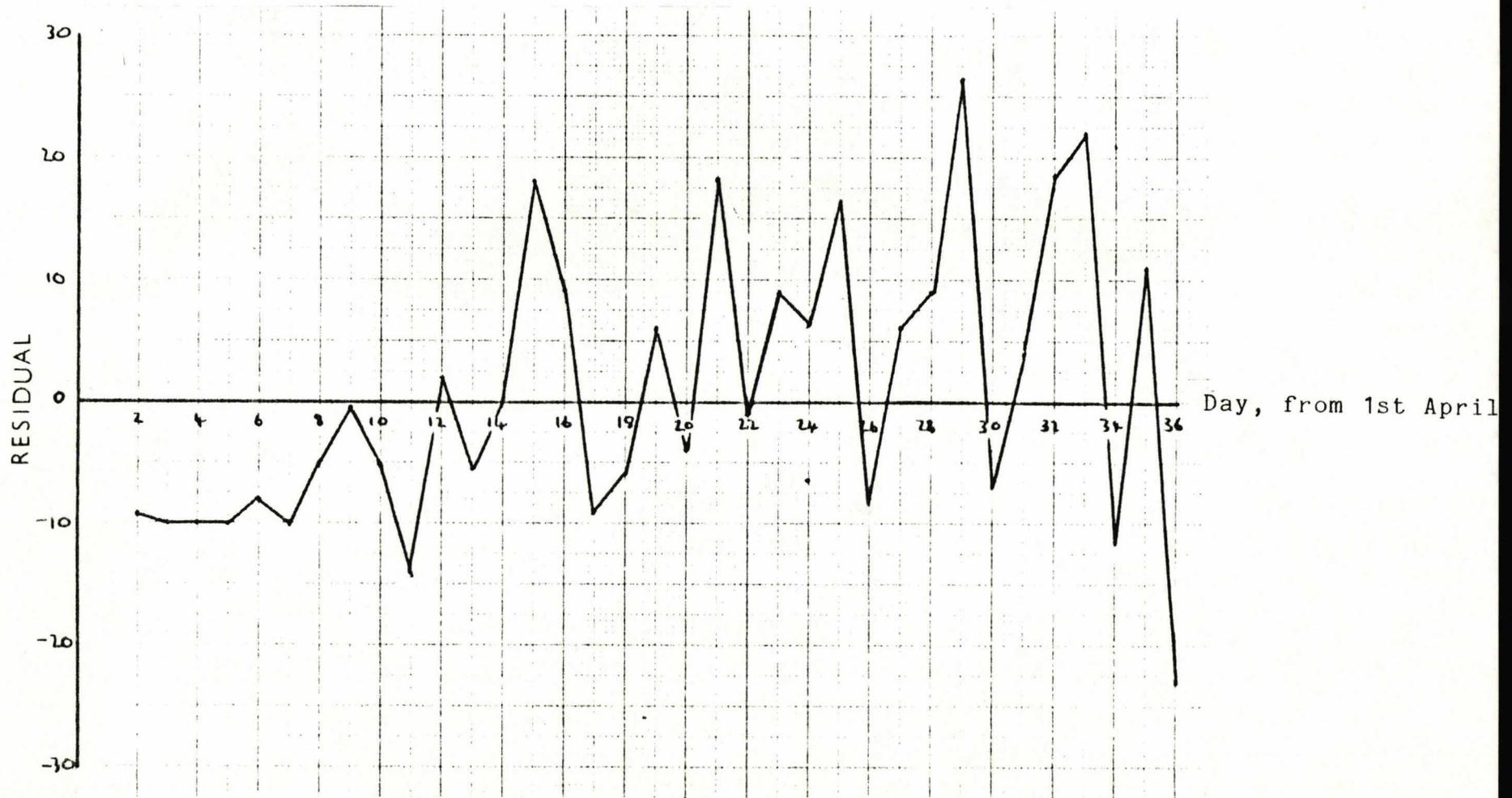


Fig. 6.2.5.4 Model(6.3) New Arrivals, Dungeness Sedge Warbler  
data, Spring 1963.

$p = 0.0$      $t = 31, t = 32$

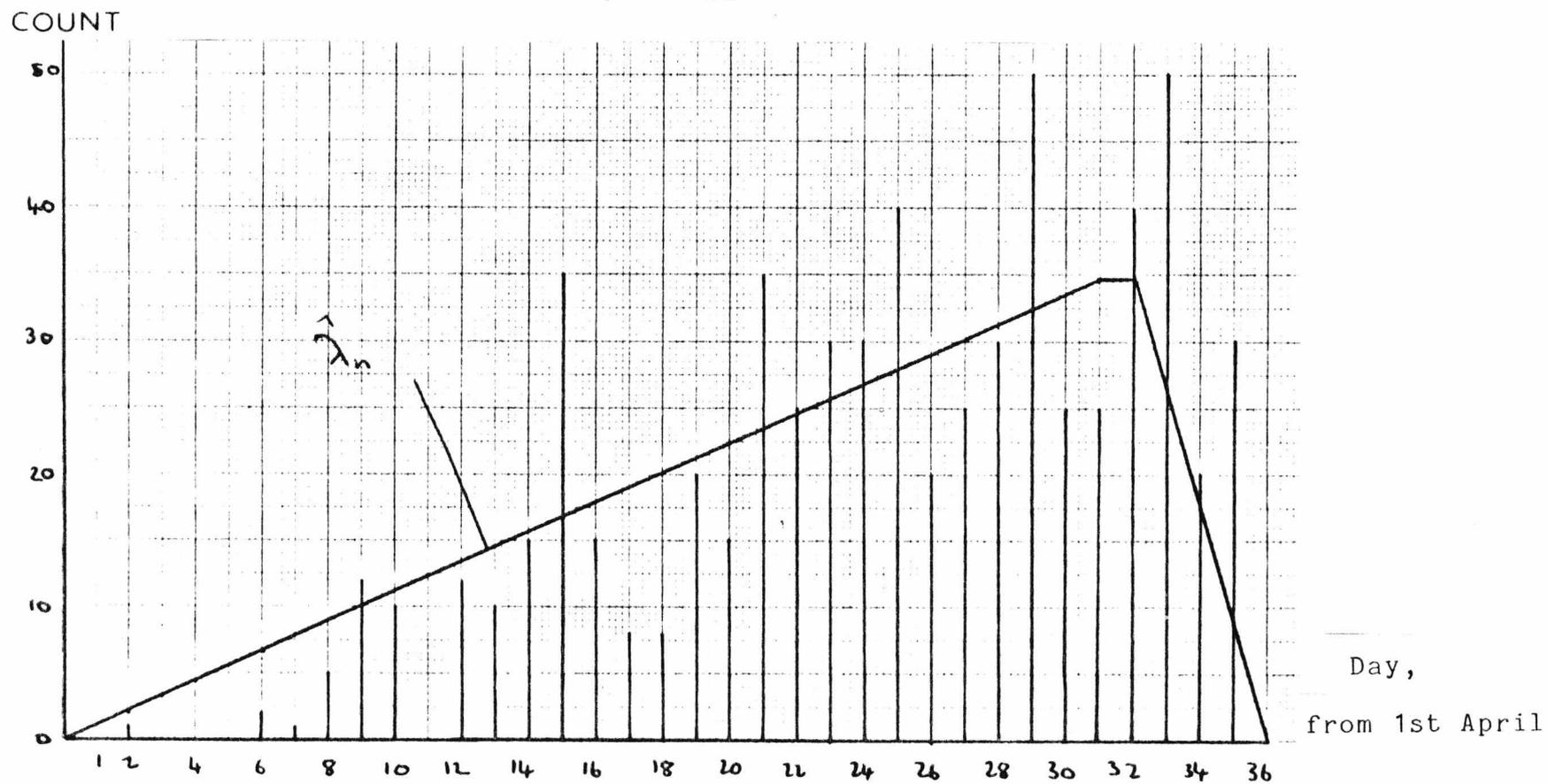
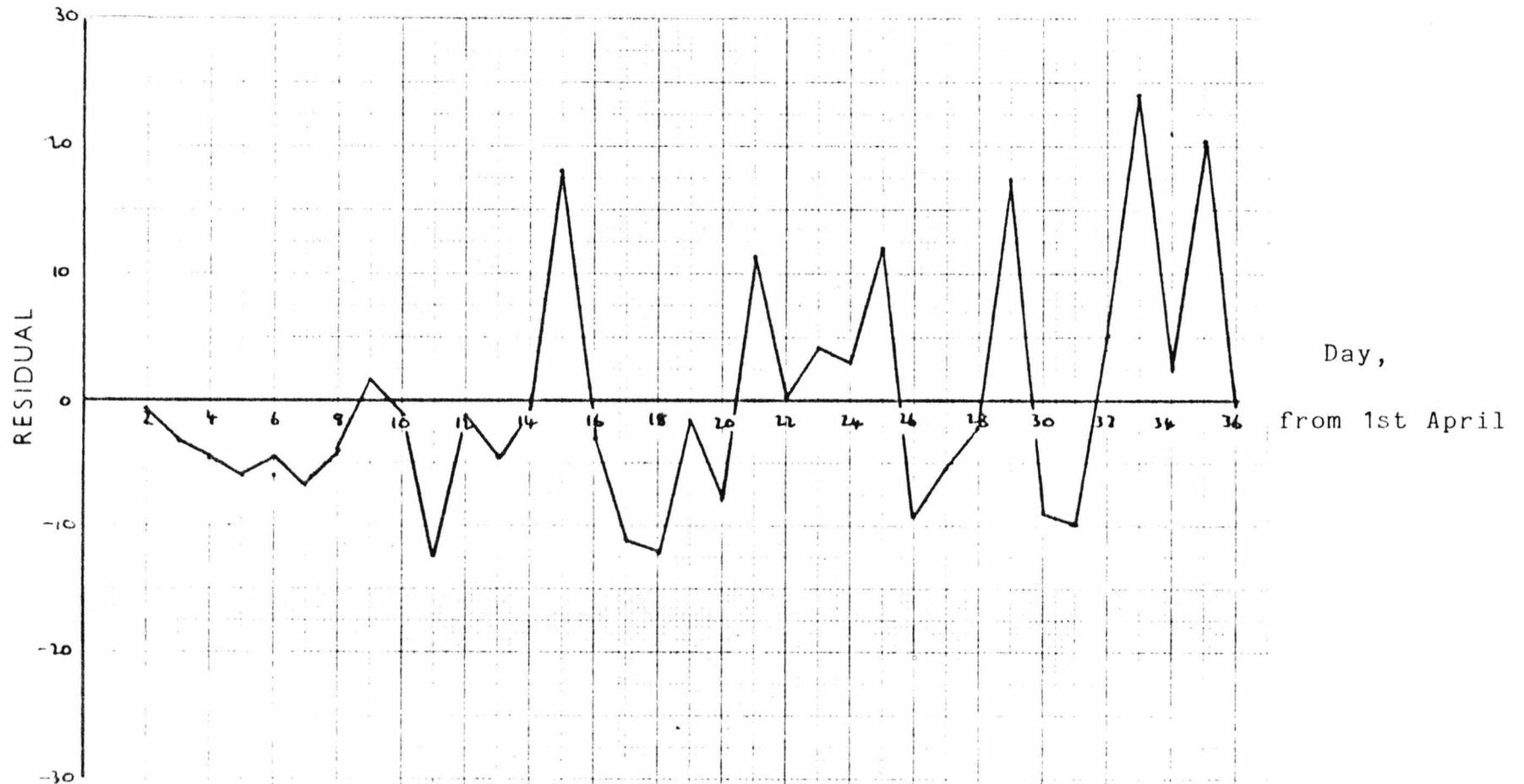


Fig. 6.2.5.5 Model(6.3) Residuals, Dungeness Sedge Warbler  
data, Spring 1963.



Tables 6.2.13 - 6.2.16 present the model (6.1) indices with the relevant CBC indices. All indices are adjusted to have a value of 100 in 1966. Comparisons with CBC indices for the Pied Flycatcher and Redstart are not possible due to the absence of CBC data for these species. Estimates of standard errors of the migration indices are also given in the tables. These estimates were arrived at by assuming that the true value of the index for 1966 is the estimate of that index. The standard errors are therefore only approximate. Although the standard errors decrease in size with decreasing size of index, in years when few birds were recorded, the coefficient of variation is over 50%. However, 70% of the indices given in the tables have a coefficient of variation of less than 15%.

Table 6.2.17 shows that for most species/observatory combinations considered, there is no evidence that model (6.1) migration indices reflect year-to-year changes as measured by the CBC. The migration indices do, however, show the major population crash of the Whitethroat (1968-69) and also the Blackcap indices appear to reflect the trend of the woodland CBC index.

In order to see whether or not the migration indices are more useful than raw totals of birds observed the relevant correlation coefficients are presented in Table 6.2.17. There is no consistent pattern of improvement over raw totals for the Sedge Warbler, but for the Blackcap there is a small improvement. There is a suggestion that for the

Table 6.2.13

Sedge Warbler CBC and model (6.1) indices

Year	CBC(F)	SPRING				AUTUMN	
		Dungeness	Portland	Bardsey	Spurn	Dungeness	Portland
1961	-	381 (36)	0 -	35 (3)	74 (13)	28 (3)	44 (7)
62	60	394 (35)	83 (13)	58 (4)	5 (3)	72 (6)	16 (4)
63	59	253 (28)	115 (16)	58 (5)	4 (2)	94 (6)	7 (3)
64	80	42 (10)	151 (18)	37 (3)	41 (8)	134 (8)	52 (8)
65	78	5 (4)	48 (11)	8 (1)	27 (6)	56 (5)	108 (12)
66	100	100 -	100 -	100 -	100 -	100 -	100 -
67	66	292 (31)	195 (20)	42 (4)	49 (9)	91 (6)	162 (14)
68	113	69 (16)	165 (18)	-	-	75 (6)	99 (11)
69	63	28 (9)	10 (4)	-	-	60 (5)	86 (10)
70	79	23 (8)	106 (15)	-	-	111 (6)	284 (19)
71	65	139 (19)	52 (11)	-	-	53 (4)	117 (12)
72	54	144 (19)	171 (20)	-	-	14 (2)	181 (15)

(Estimated standard errors in parentheses)

Table 6.2.14

Whitethroat CBC and model (6.1) indices

Year	CBC(F)	CBC(W)	SPRING	AUTUMN
			Dungeness	Dungeness
1961	-	-	167 (7)	84 (5)
62	85	-	50 (4)	110 (6)
63	80	-	58 (4)	95 (6)
64	84	92	41 (3)	125 (7)
65	84	97	22 (1)	116 (6)
66	100	100	100 -	100 -
67	86	91	39 (3)	125 (7)
68	110	128	74 (5)	94 (6)
69	30	45	2 (1)	9 (2)
70	31	49	15 (2)	24 (3)
71	33	42	22 (2)	54 (4)
72	26	35	6 (1)	23 (3)

Table 6.2.15

Willow Warbler CBC and model (6.1) indices

Year	SPRING		AUTUMN	
	CBC(F)	CBC(W)	Dungeness	Dungeness
1961	-	-	8 -	51 (4)
62	53	-	43 (3)	77 (5)
63	63	-	51 (3)	65 (4)
64	72	81	37 (3)	99 (5)
65	87	84	24 (2)	57 (4)
66	100	100	100 -	100 -
67	96	92	40 (3)	65 (4)
68	115	103	27 (2)	99 (5)
69	114	108	50 (3)	93 (5)
70	120	100	61 (4)	108 (5)
71	96	88	30 (1)	102 (5)
72	99	97	74 (4)	53 (4)

Table 6.2.16

Blackcap CBC indices and model (6.1) indices

Year	SPRING		
	CBC(F)	CBC(W)	Dungeness
1961	-	-	44 (9)
62	24	-	161 (18)
63	44	-	77 (12)
64	80	91	74 (13)
65	68	84	22 (7)
66	100	100	100 -
67	86	92	38 (10)
68	114	99	72 (13)
69	105	104	22 (7)
70	127	101	133 (16)
71	115	95	56 (10)
72	157	115	169 (19)

Table 6.2.17 CBC vs. model (6.1) indices

a) Sedge Warbler

	SPRING				AUTUMN	
	Dungeness	Portland	Bardsey	Spurn	Dungeness	Portland
CBC(F)	-.47 (-.20)	.10 (.19)	.30 (-.23)	-89* (.74)	.39 (.38)	.07 (.02)

(Correlations between raw totals and CBC in parenthesis)

\* significant at 5% level

b) Whitethroat

	Dungeness (SPRING)	Dungeness (AUTUMN)
CBC(F)	.83** (.82)	.88** (.97)
CBC(W)	.78** (.95)	.82** (.97)

\*\* significant at 1% level

c) Willow Warbler

	Dungeness (SPRING)	Dungeness (AUTUMN)
CBC(F)	.18 (.03)	.56 (-.18)
CBC(W)	.44 (.22)	.25 (.06)

d) Blackcap

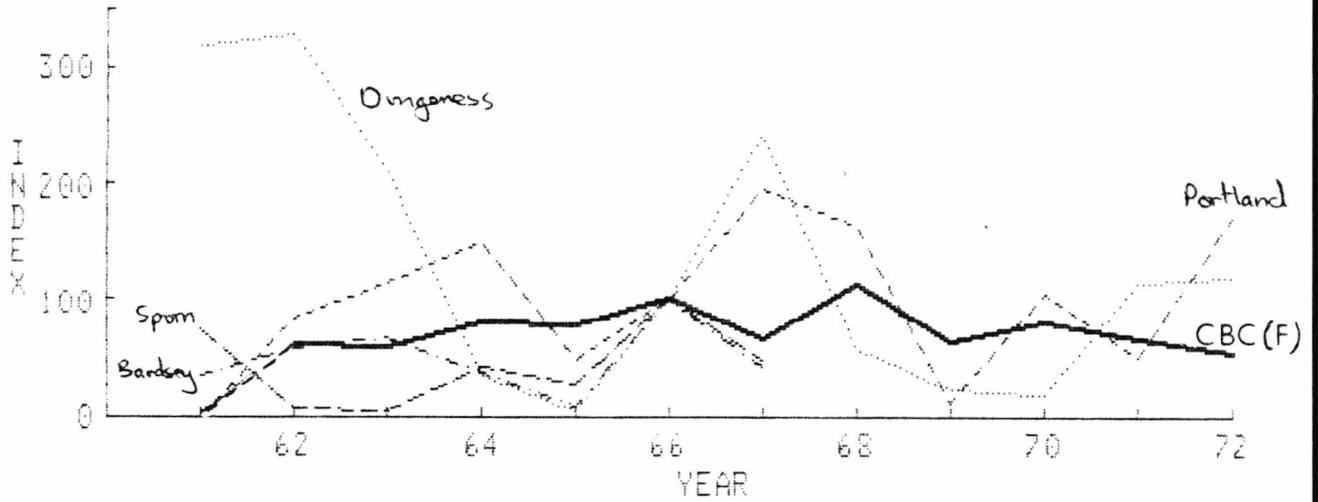
	Dungeness (SPRING)
CBC(F)	.14 (-.32)
CBC(W)	.71* (.71)

Whitethroat, the raw totals perform better than the model-based indices but both give significant correlations with CBC indices.

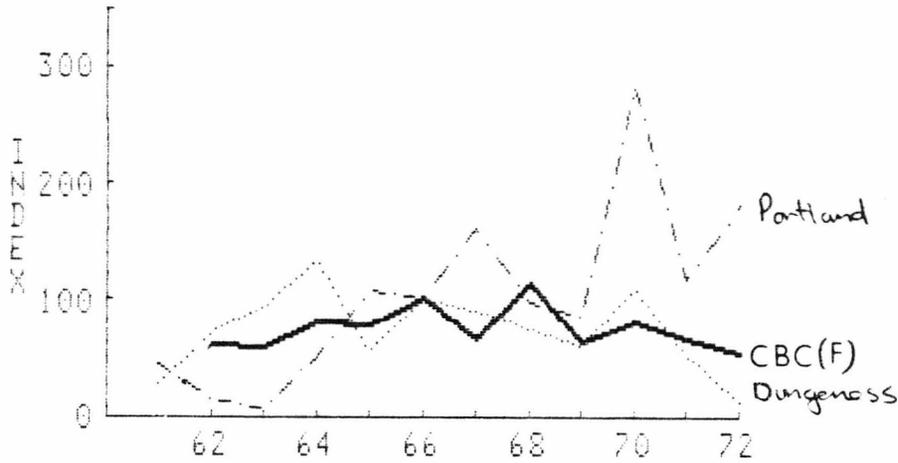
The migration indices from model (6.1) are plotted in Fig. 6.2.6.1 with the relevant CBC indices. An encouraging feature is that although the correlation coefficients for the Willow Warbler (spring and autumn) are not significant, the long-term trend of the CBC index is clearly reflected in the migration index.

Although individual seasonal migration indices from Dungeness and Portland show no correlation with the Sedge Warbler CBC index, an investigation into the possibility of forming a combined index was carried out. In order to have sufficient data to perform a multiple regression analysis of CBC index on migration indices, only indices from Dungeness and Portland are used - other observatories provide too few years with which to work. Spring and autumn indices were included in the regression as well as indices from the previous autumn. Using all variables the regression is not significant although the correlation coefficient between CBC and combined migration index is significant at the 5% level. This suggests that, given more data, a significant fit may be obtained, giving migration indices with a lower level of precision. In the above regression, six explanatory variables are used with only eleven observations and so it is not surprising that a good fit is found.

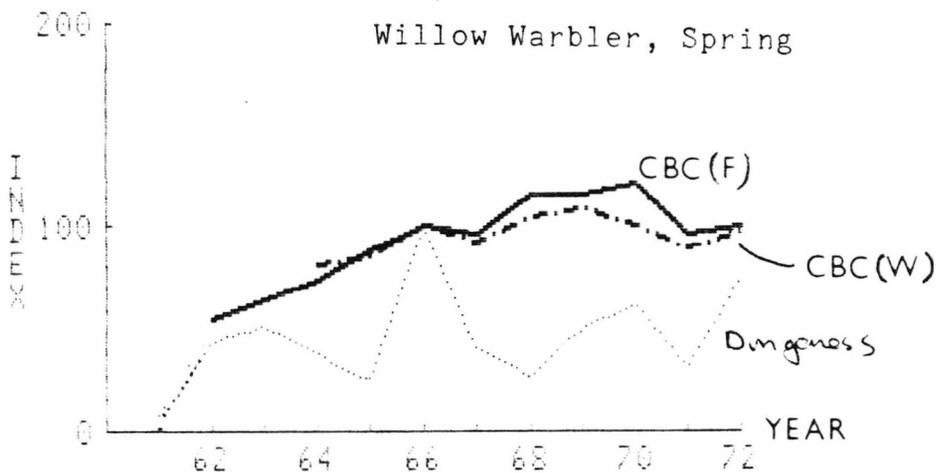
Sedge Warbler, Spring

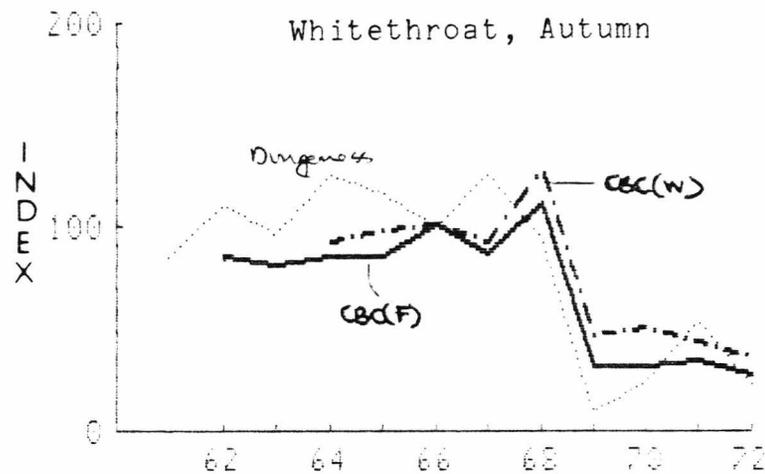
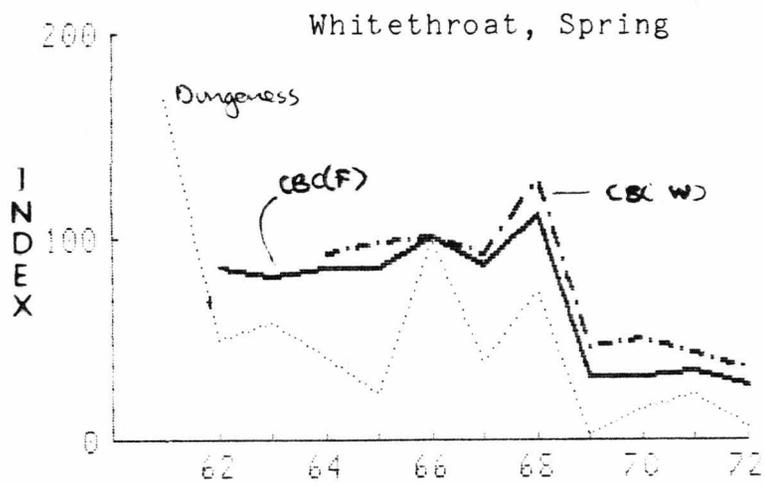
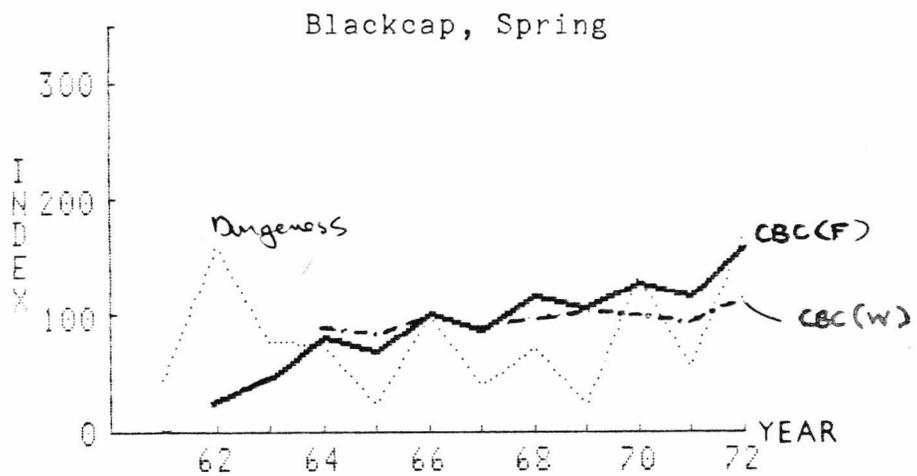
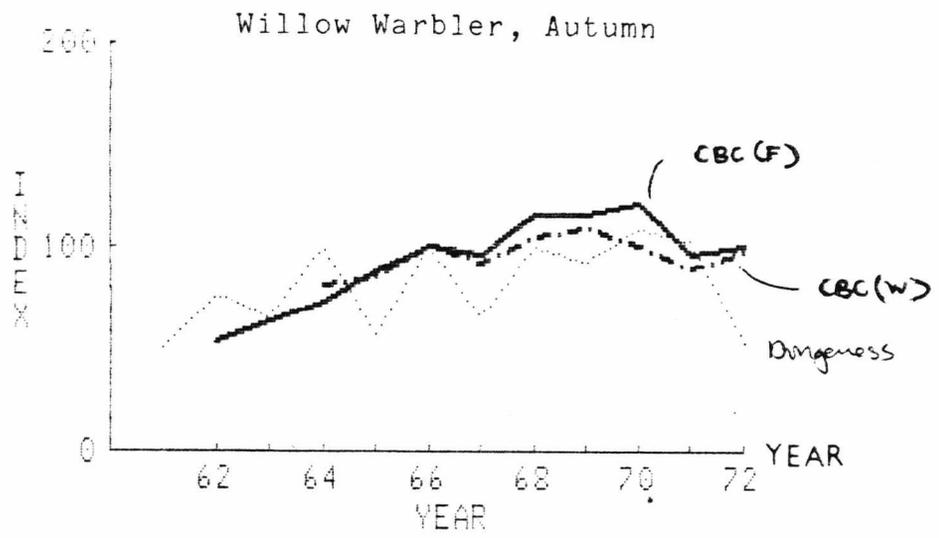


Sedge Warbler, Autumn



Willow Warbler, Spring





For the Dungeness Willow Warbler data, it was found that by combining spring, autumn and previous-autumn indices, using multiple regression, a combined index shows positive correlation with the CBC index. A significant fit is obtained and the correlation coefficient between CBC and combined migration index is 0.75 (significant at 5% level). All three indices are needed to achieve this correlation.

For the Dungeness Whitethroat data the combination of spring, autumn and previous autumn indices leads to a combined migration index which has a correlation coefficient with the CBC index of 0.98. The very good fit of this linear model suggests that by combining migration indices one can very accurately reflect population fluctuations as measured by the CBC.

Although the proposed migration indices of this chapter have so far been tested on a limited amount of data, the results are very encouraging. It appears that once model (6.1) is fitted to sequences of counts, each species requires individual attention in order to combine indices from different observatories and seasons. In other words there is no general rule for combining indices in order to achieve a migration index which follows the behaviour of the relevant CBC index.

### 6.3 Parslow's transformation

#### 6.3.1 Introduction

In a study of migration on the east coast of Great Britain, Lack (1960) used as his estimates of the numbers of newly arrived birds at various sites, the excess (if any) of each day's count over that of the previous day. Lack's attempt to eliminate 'stayers' from counts obscures arrivals of similar magnitude on successive days.

Parslow (1962) suggested using the average of the minimum possible and maximum possible number of arrivals on each day. The transformation is illustrated below, where  $W_n$  is the estimated number of new arrivals included in the count for the day  $n$  ( $X_n$ ):

$$\text{If } X_{n-1} \geq X_n \text{ then } W_n = (0 + X_n)/2 = X_n/2$$

$$\text{If } X_{n-1} < X_n \text{ then } W_n = (X_n - X_{n-1} + X_n)/2 = \frac{X_n - X_{n-1}}{2}$$

In this section we examine the extent to which this transformation accounts for birds staying from one day to the next at an observatory. We show that, if model (6.1) is appropriate for a sequence of counts, then any serial correlation in the data is caused by 'stayers'. Therefore, to check whether or not Parslow's transformation eliminates 'stayers' from the counts, we can examine the change in the sample serial correlation coefficient before and after transformation.

### 6.3.2 Serial correlation under model (6.1)

Recall that model (6.1) is given by  $X_n = Y_n + I_n$ , where  $Y_n \sim \text{BIN}(x_{n-1}, p)$  and  $I_n \sim \text{POI}(\lambda)$ , where  $Y_n$  and  $I_n$  are independent, for all  $n = 1, \dots, t$ . The serial correlation coefficient  $\rho_1$  is given by

$$\rho_1 = \frac{\text{Cov}(X_{n+1}, X_n)}{\sqrt{V(X_{n+1})V(X_n)}}$$

In order to obtain expressions for variances and covariances we let  $n$  tend to infinity, i.e. we assume that the series has settled down to an equilibrium state where variances and covariances are independent of time.

We first find  $V(X_n)$  which, since  $Y_n$  and  $I_n$  are independent, is given by  $V(X_n) = V(Y_n) + V(I_n)$ . The second term is straightforward:  $V(I_n) = \lambda$ . To find the first term, we consider the formula

$$V(Y_n) = E(Y_n^2) - \{E(Y_n)\}^2$$

and derive expressions for  $E(Y_n^2)$  and  $E(Y_n)$  using conditional expectations as follows:

$$E(Y_n) = E \left[ E(Y_n/X_{n-1}) \right]$$

$$\text{where } E(Y_n/X_{n-1}) = \sum_{Y_n} Y_n \text{Pr}(Y_n = x_n/X_{n-1} = x_{n-1})$$

$$\begin{aligned} \therefore E(Y_n) &= \sum_{x_{n-1}} \left\{ \sum_{Y_n} y_n \text{Pr}(y_n/x_{n-1}) \right\} \text{Pr}(X_{n-1} = x_{n-1}) \\ &= \sum_{x_{n-1}} \left\{ p x_{n-1} \right\} \text{Pr}(X_{n-1} = x_{n-1}) \\ &= pE(X_{n-1}) \end{aligned}$$

The derivation of  $E(Y_n^2)$  is more difficult, and the algebra is presented in Appendix 6.1. We find that  $E(Y_n^2)$  is given by

$$E(Y_n^2) = p^2 E(X_{n-1}(X_{n-1} - 1)) + p E(X_{n-1})$$

and that the variance of  $Y_n$  is given by

$$V(Y_n) = p^2 V(X_{n-1}) + p(1-p)E(X_{n-1}),$$

$$\therefore V(X_n) = p^2 V(X_{n-1}) + p(1-p)E(X_{n-1}) + \lambda$$

This is a recurrence relation which yields (see Appendix 6.1), as  $n \rightarrow \infty$ ,

$$V(X_n) = E(X_n) = \lambda/(1-p)$$

We now require the covariance between successive observations, and the algebra is set out in Appendix 6.2. We obtain the following result:

$$\text{Cov}(X_{n+1}, X_n) = \frac{p\lambda}{1-p}, \text{ which gives}$$

$$p_1 = p, \text{ as } n \rightarrow \infty.$$

An obvious estimate of  $p_1$  is thus given by  $\hat{p}$ , the maximum likelihood estimator of  $p$ .

### 6.3.3 The effect of Parslow's transformation

For all the daily counts sequences to which model (6.1) was fitted, the serial correlation coefficient was calculated before and after using Parslow's transformation.

As one would expect from the work of section 6.3.2, the sample serial correlation coefficients were close to the estimates of the binomial parameter  $p$  of model (6.1). Any discrepancy will be due to the influence of the second parameter,  $\lambda$ , of model (6.1), which is not independent of

p - see correlations of Tables 6.2.1 - 6.2.10.

For all sequences of counts, the serial correlation coefficient was reduced after using Parslow's transformation, but the reduction varied from less than 10% to over 90%. This suggests that in some instances, the transformation effectively removes 'stayers' from the counts, but sometimes it does not. An examination of the results suggested no reason for this inconsistency - there was no evidence of a link with the goodness of fit, the value of p or any ornithological features such as seasonal differences. It appears that if one is seeking an automatic method of eliminating 'stayers' from sequences of counts, then the fitting of model (6.1) provides a more reliable method since the model has been shown to fit most sequences of daily counts.

#### 6.4 Conclusions and discussion

In this chapter simple models for daily counts are proposed, with a view to constructing an index of migration volume.

The first model, model (6.1), was found to be straightforward to fit to data and an extensive simulation study showed that for most sequences of counts, the fit of the model was good. For some sequences of counts however, the model did not fit. These sequences were from seasons where few birds were recorded. The implication of this is that although migration indices may be available for some years, in years when the population falls to a low level no indices are available. However, finding that the

model fails to fit a set of data for a particular year is useful in itself since the attention of ornithologists would be drawn and investigations of the reasons behind the lack of fit could be made.

A very encouraging feature of model (6.1) is that it is possible to combine parameter estimates from different seasons or different observatories and form a migration index which is correlated with the CBC index. A limited number of examples is presented in this chapter but the success of the methodology should encourage further research using more species and observatories over longer time periods. If sufficient data from several observatories could be obtained, then the construction of a 'national' migration index may obviate the need for regional CBC indices which were discussed in chapter 2.

One surprising feature of the parameter estimates of model (6.1) is that no consistent seasonal differences are found. The serial correlation of the counts is of a similar magnitude in spring and autumn. One would have expected less serial correlation in spring since birds are thought to be very keen to reach their breeding grounds but this hypothesis is not supported by the analysis of this chapter.

Alternative models to model (6.1) have been studied. One model (model (6.2) ) was suggested where the assumption of constant  $p$  over the migration season was relaxed. Another model (model (6.3) ) was studied where the parameter  $\lambda$  was allowed to vary over the season. Neither of these two models suggested significant

improvements in goodness of fit or the correlation of migration indices with CBC indices. Further, models (6.2) and (6.3) are difficult to fit to data in terms of computing complexities and time, and would not offer a simple and automatic procedure which ornithologists could easily use.

Appendix 6.1

E(X<sub>n</sub>) and V(X<sub>n</sub>)

(see section 6.3.2)

$$E(Y_n^2) = E \left\{ E(Y_n^2/X_{n-1}) \right\}$$

where

$$\begin{aligned} E(Y_n^2/X_{n-1}) &= \sum_{Y_n} y_n^2 P_r(Y_n = y_n/X_{n-1} = x_{n-1}) \\ &= \sum_{X_{n-1}} \left\{ \sum_{Y_n} y_n P_r(y_n/x_{n-1}) \right\} P_r(X_{n-1} = x_{n-1}) \\ &= \sum_{X_{n-1}} \left\{ x_{n-1}(x_{n-1}-1)p^2 + x_{n-1}p \right\} P_r(X_{n-1} = x_{n-1}) \\ &= p^2 E \left\{ X_n(X_{n-1}) \right\} + pE(X_{n-1}) \end{aligned}$$

$$\begin{aligned} \therefore V(Y_n) &= p^2 E(X_{n-1}^2) - p^2 E(X_{n-1})^2 + pE(X_{n-1}) - p^2 E(X_{n-1})^2 \\ &= p^2 V(X_{n-1}) + p(1-p)E(X_{n-1}) \end{aligned}$$

and so  $V(X_n) = p^2 V(X_{n-1}) + p(1-p)E(X_{n-1}) + \lambda$

We form from the above recurrence relation,

$$\begin{aligned} V(X_n) &= p^{2(n-1)} V(X_1) + (1-p) \sum_{i=1}^{n-1} p^{2n-1-2i} E(X_i) \\ &\quad + \lambda \sum_{i=1}^{n-1} p^{2(i-1)} \end{aligned} \tag{A6.1}$$

Considering each term of (A6.1) individually,

as  $n \rightarrow \infty$ ,  $p^{2(n-1)} V(X_1) \rightarrow 0$ .

The second term of (A6.1) can be rewritten as

$$\begin{aligned} \sum_{i=1}^{n-1} p^{2n-1-2i} E(X_i) &= p^{2n-1} \sum_{i=1}^{n-1} p^{-2i} E(X_i) \\ &= p^{2n-1} E(X_1) \sum_{i=1}^{n-1} p^{-(i+1)} \\ &\quad + p^{2n-1} \lambda \sum_{i=1}^{n-1} \left\{ p^{-2i} \sum_{k=0}^{i-2} p^k \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{p^{2n-1} E(X_1) (1-p^{n-1})}{p^n (1-p)} + \frac{p^{2n-1} \lambda}{(1-p)} \cdot \sum_{i=1}^{n-1} p^{-2i} (1-p^{i-1}) \\
&= \frac{p^{2n-1} E(X_1) (1-p^{n-1})}{p^n (1-p)} + \\
&\quad \frac{\lambda}{(1-p)} \left\{ \frac{p p^{2(n-1)} - p}{(p^2-1)} + \frac{p^{n-1} (p^{n-1}-1)}{(1-p)} \right\}
\end{aligned}$$

and as  $n \rightarrow \infty$

$$\begin{aligned}
&= 0 + \frac{\lambda}{(1-p)} \left\{ \frac{p}{1-p^2} + 0 \right\} \\
&= \frac{\lambda p}{(1-p) (1-p^2)}
\end{aligned}$$

Thus, as  $n \rightarrow \infty$ , the second term of (A6.1)

is  $\frac{\lambda p}{(1-p^2)}$ .

Finally, the third term of (A6.1) =  $\frac{\lambda}{1-p^2}$  as

$n \rightarrow \infty$ , which gives

$$V(X_n) = \lambda / (1-p) .$$

It is straightforward to obtain an expression for  $E(X_n)$  from the recurrence relation  $E(X_n) = pE(X_{n-1}) + \lambda$ , which leads to  $E(X_n) = \lambda / (1-p)$ .

Appendix 6.2

Cov( $X_{n+1}$ ,  $X_n$ ) and  $\rho_1$

(see section 6.3.2)

$$C_{ov}(X_{n+1}, X_n) = E(X_{n+1}X_n) - E(X_{n+1}) E(X_n)$$

where

$$\begin{aligned} E(X_{n+1}X_n) &= \sum_{x_n=0}^{\infty} x_n E(X_{n+1}/X_n = x_n) P_r(X_n = x_n) \\ &= \sum_{x_n=0}^{\infty} x_n \{x_n p + \lambda\} P_r(X_n = x_n) \\ &= \sum_{x_n=0}^{\infty} \{p x_n^2 P_r(X_n = x_n) + \lambda x_n P_r(X_n = x_n)\} \\ &= p E(X_n^2) + \lambda E(X_n) \\ &= p \{V(X_n) + E(X_n)^2\} + \lambda E(X_n) \end{aligned}$$

As  $n \rightarrow \infty$ ,  $E(X_{n+1}X_n) \rightarrow p \left\{ \frac{\lambda}{1-p} + \frac{\lambda^2}{(1-p)^2} \right\} + \frac{\lambda^2}{1-p}$

and so  $C_{ov}(X_{n+1}, X_n) = \frac{p\lambda(1-p) + p\lambda^2 - \lambda^2(1-p) - \lambda^2}{(1-p)^2}$

$$= p\lambda/(1-p)$$

Hence,  $\lim_{n \rightarrow \infty} \rho_1 = \frac{p\lambda}{(1-p)} \cdot \frac{(1-p)}{\lambda} = p$ .

Note: it is straightforward to extend these results;

for example,  $\rho_2 = \text{corr.}(X_n, X_{n-2}) = p^2$ , etc.

## CHAPTER 7

### WEATHER DRIVEN MODELS FOR DAILY COUNTS

#### 7.1 Introduction

In this chapter we examine a new method of including information on local weather conditions into models for daily counts of grounded migrants.

We start by considering a simple model where weather information is condensed into a binary form, i.e. we consider two types of weather condition. This idea follows from the work of chapter 5 where two-group discriminant analysis was studied. We later re-examine the Poisson-binomial models of chapter 6 with the introduction of weather information.

The models of this chapter are constructed with a view to the formulation of migration indices for population monitoring. The weather data are used as a means of distinguishing between high and low counts while taking into account that on certain days, no birds would be expected and on other days, conditions are ideal for large numbers of migrants to be grounded. In order to proceed in this way, we need to be precise about the definition of a migration season, i.e. when does it start and finish? This problem is considered in section 7.2.

#### 7.2 Defining the migration season

In the previous chapter, models were fitted to sequences of daily counts which spanned the entire observed migration season, i.e. from the first count to the last count. Initially, no account

was taken of the fact that one often observes a slow start to the season as well as a gradual decline towards the end of the season. Later, more complex models were examined where the 'shape' of the migration season was considered.

In this chapter the aim is to calculate expected numbers of birds on certain categories of days, on the basis of weather and migration data. It is important, as will become clear in later sections, that these calculations are based on periods of time during which the migration is at a reasonably steady rate. One therefore needs to choose carefully the period of time for analysis.

In order to fix the time period for the analysis of each season's counts, we could follow one of two simple approaches. We could use a fixed number of weeks or months, starting from the same date each year, where the length of the period is chosen on the basis of, for example, the all-time earliest and latest records for a given species. However, if for some reason the supply of birds is drastically reduced in a particular year, then we may end up with a period for analysis that includes mostly zero counts. This method is also susceptible to changes in the timing of migration in each year, although in section 3.2.4 we found no evidence from our brief investigation that this would, in fact, be a problem. An alternative and more objective method is to choose the period for analysis on the basis of the observed counts. This means that the period used for analysis will vary from year to year, and so may accommodate changes in climatic conditions for which we have no data. Within each chosen period we can then examine the numbers of birds recorded in relation to

local weather conditions for which we have data available. A simple method of choosing the period for analysis is to count up the total number of birds seen over each season, and then study the central, say, 4/5<sup>ths</sup> of the season. Thus our chosen period starts on the day on which the  $n/10^{\text{th}}$  bird is recorded ( $n$  is the total number) and ends on the day on which the  $9n/10^{\text{th}}$  bird is recorded. An advantage of this method is that extremely early and late counts, which generally consist of only one or two birds are excluded. Thus we reduce the possibility of our analyses being influenced by extreme counts, which may be of birds behaving in a manner which is different from the large majority of birds.

We have suggested above that the outer deciles of the data (relative to time) could be discarded, but this is only one such possibility. If we discarded, for example, the outer quartiles, then experience has shown that the remaining period is rather short - perhaps of the order of two weeks or less. This is a consequence of extremely large counts which 'draw' in the outer percentiles. A survey of some data sets has suggested that by using the central 80% of counts in each season, then one is mainly dealing with sequences approximately three or four weeks long in the spring, and four or five weeks long in the autumn. Unless stated otherwise, the sequences of counts that are examined in this chapter have been selected in the manner described above.

### 7.3 Bird-days and Non-bird days

#### 7.3.1 Discriminant functions

In chapter 5 we saw that it was possible, in most cases, to distinguish between high and low counts on the basis of a series of weather observations. The success of the two-group discriminant analyses suggests that one could regard each daily count as originating from one of two sampling distributions where the appropriate distribution for each count is determined by local weather conditions. This idea for a simple model is developed further in the next section. In this section we investigate how the observed counts behave on days that are designated as 'bird-days' or 'non-bird' days on the basis of the discriminant functions. The discriminant functions that we use here are those that led to the most successful discriminant analyses of chapter 5. The coefficients are tabled in table 5.4.5. The discriminant scores are straightforward to obtain by multiplying the coefficients by the relevant weather observations, for every day under consideration. The success rates for the discriminant analyses may not be reflected in the diagrams presented in this section, since we now consider only one year at a time, rather than the combined data for a number of years.

Each diagram that is presented in this section displays the observed daily counts together with an indication as to whether each day was designated as a bird day or a non-bird day. (An asterisk on the time-axis indicates a bird-day, and corresponds to a discriminant score of less than

zero). Of the 88 available plots, we only present a small sample here, chosen in order to exhibit interesting features.

Fig. 7.3.1.1 shows two examples where asterisks appear to coincide with days on which birds were actually observed, and sequences of non-bird days coincide with days on which relatively few birds were observed. These examples are from Dungeness (Sedge Warbler, autumn 1967 and Willow Warbler, spring 1966).

On the whole, most of the diagrams were of this form, and it is difficult to find examples where bird-days coincide with zero or low counts of birds throughout most of the migration season. Fig. 7.3.1.2 shows three examples where few birds are observed over a period of bird days, but even on these diagrams the largest counts do coincide with bird days, (Dungeness; Willow Warbler, spring 1970, Sedge Warbler, autumn 1969 and Redstart, autumn 1968).

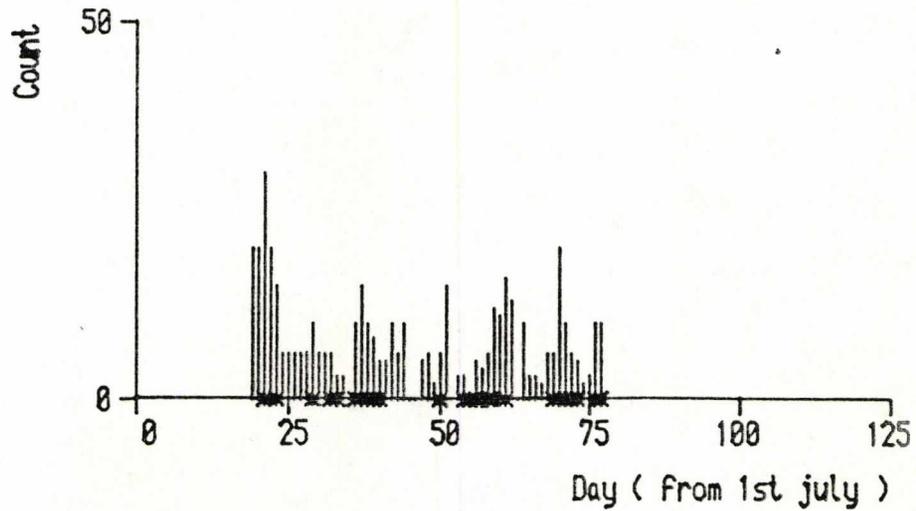
In some cases, for example Whitethroat data, very few birds were recorded in some years, and so even on bird days, few birds were seen - see Fig. 7.3.1.3 (Portland, Whitethroat, spring 1968).

The overall picture that emerges from a graphical study of the discriminant analyses is that, by and large, one is able to predict periods of high counts and low counts by considering a small set of weather variables, and therefore

Fig. 7.3.1.1 Two examples where predicted bird-days coincide with high counts of grounded migrants.

( asterisks on the horizontal axis indicate bird-days )

Sedge Warbler, Dungeness 1967



Willow Warbler, Dungeness 1967

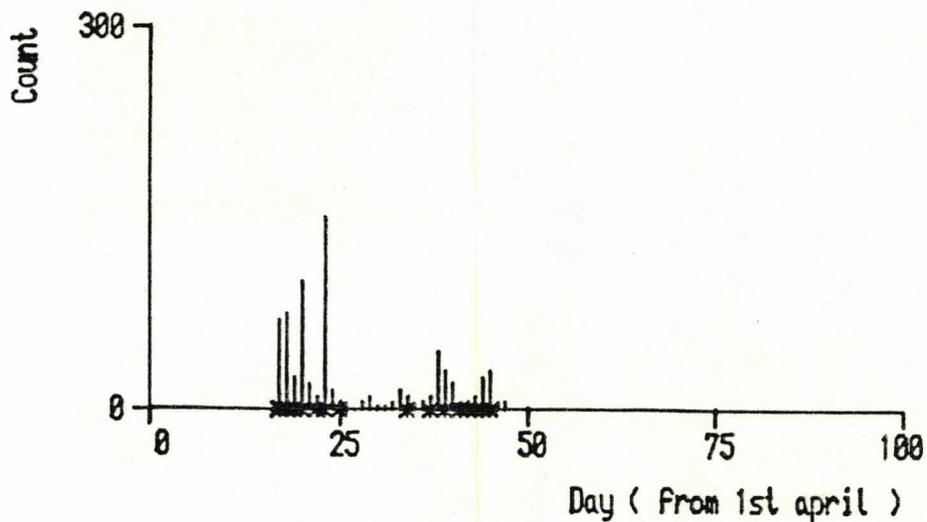


Fig. 7.3.1.2 Three examples where predicted bird-days coincide with low counts of grounded migrants.

( asterisks on the horizontal axis indicate bird-days )

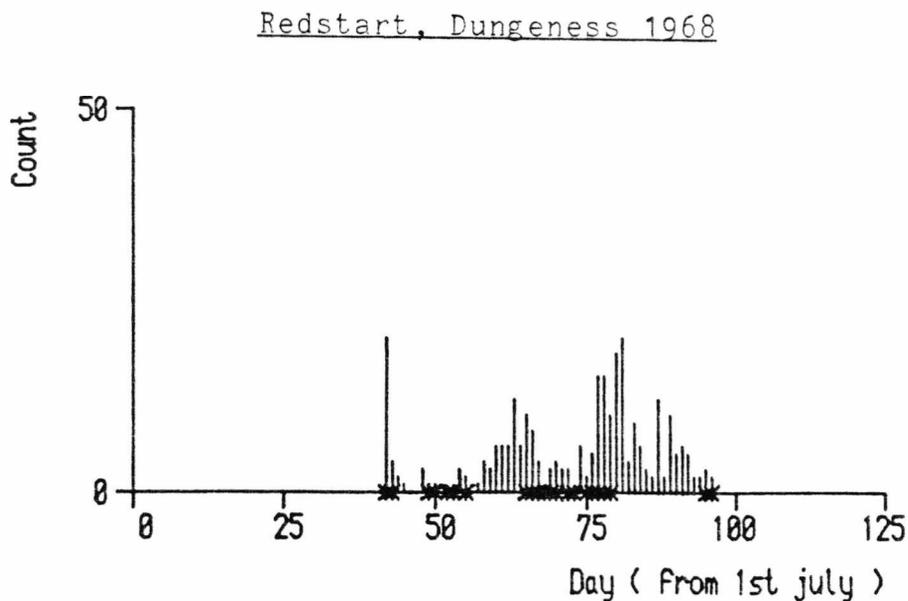
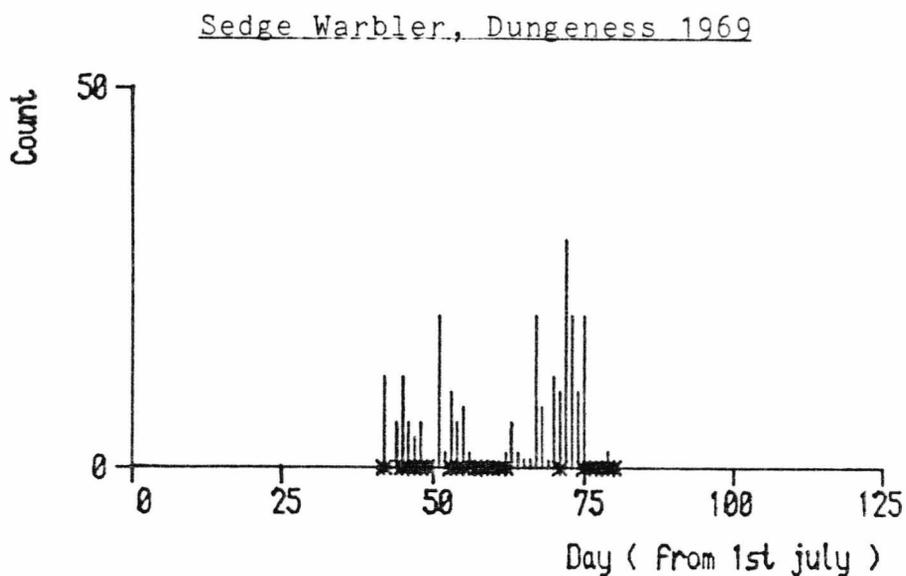
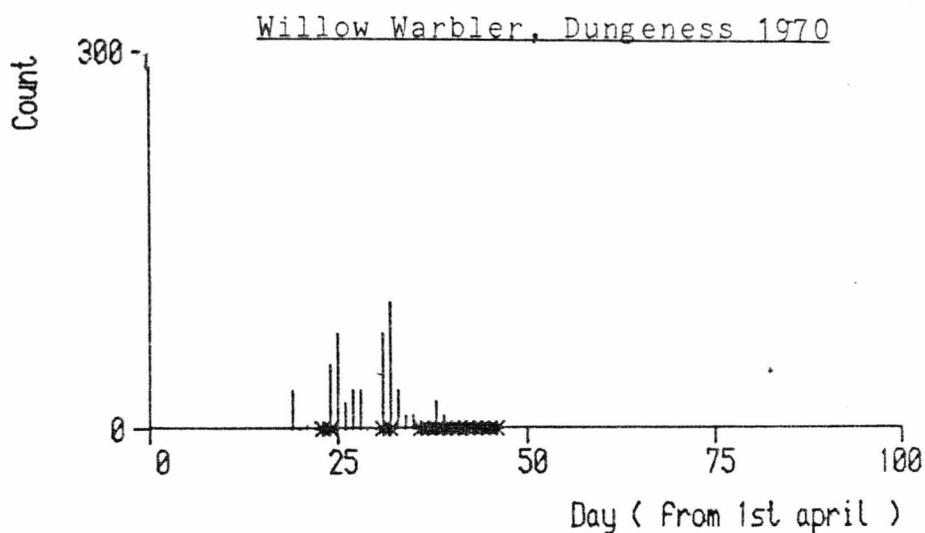
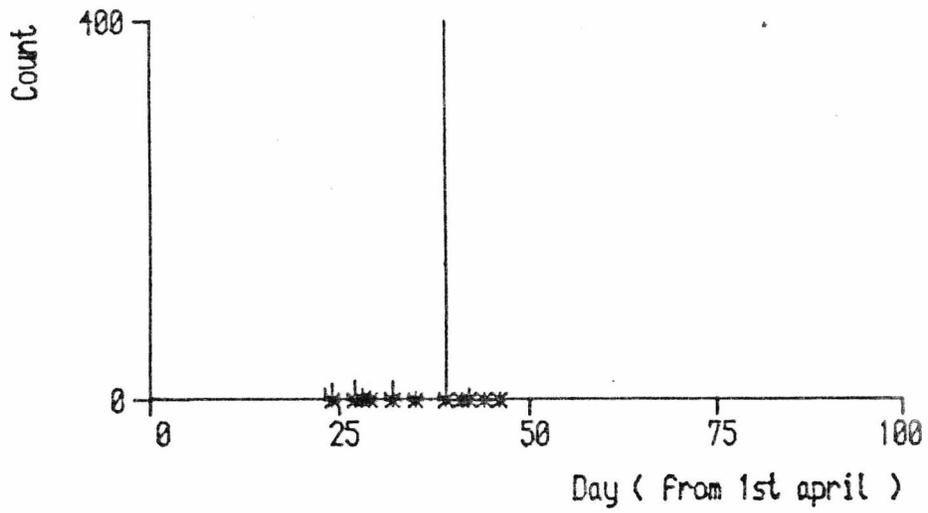


Fig. 7.3.1.3 An example where few birds are recorded on both predicted bird-days and non-bird-days.

Whitethroat, Portland 1968



the next step is to attempt to develop this idea in terms of statistical models.

### 7.3.2 A model for daily counts (model 7.1)

In this section we propose and examine a simple model for daily counts. Each count is considered to be an observation from one of two possible distributions of counts. We shall examine the case where both distributions are Poisson, this being the simplest choice when data are counts. The distribution pertaining to each count shall be governed partly by the data, and partly by weather variables. By using the appropriate discriminant function, we can calculate, for each count, a discriminant score. Depending on whether or not the score for a particular day is greater than or less than some threshold value (a parameter of the model), then the day is designated a bird day or a non-bird day.

The model may be written as follows:

$$\Pr (X_i = x) = \begin{cases} e^{-\lambda_B} \lambda_B^x / x! & W_i < W_0 \\ e^{-\lambda_N} \lambda_N^x / x! & W_i > W_0 \end{cases}$$

where  $X_i$  is the count on day  $i$ ,  $\lambda_N$  is the mean count on non-bird days,  $\lambda_B$  is the mean count for bird-days and  $W_0$  is the 'weather threshold'.  $W_i$  is the observed discriminant score on day  $i$ . In fitting this model, we are assuming that the counts are independent of time. We later consider how this assumption might be relaxed. Note that in modelling the data in this way we are assuming that on each

night, there is in fact a supply of birds migrating that are able to land at observatories if the local weather conditions are such that the day is designated a bird day. This will clearly not always be the case, but the discriminant analyses have shown sufficiently encouraging results to merit an attempt at considering weather-driven models.

The parameter for the weather threshold,  $W_0$ , is included in the model in order, as much as possible, to allow for differences in behaviour from year to year (the discriminant scores were evaluated using data amalgamated over a period of years). The estimated value of  $W_0$  is, however, expected to be close to zero.

Given a sequence of daily counts and the associated discriminant scores, it is straightforward to estimate the parameters  $\lambda_B$ ,  $\lambda_N$  and  $W_0$  by maximum likelihood. Rather than using the Nelder-Mead simplex method, a global search in three dimensions was carried out in order to find the maximum likelihood estimates. This method, while relatively inefficient compared to the simplex method, enables contour plots to be produced with little difficulty, using NAG graphical facilities.

Table 7.3.2.1 and 2 show the parameter estimates, rounded to the nearest integer, for the combinations of species, observatory, season and year that were studied in previous chapters. Because of the way in which the estimates were

Table 7.3.2.1 Parameter estimates (maximum likelihood)  
for model (7.1)

$\hat{\lambda}_N$  = mean no. of birds on non-bird days  
 $\hat{\lambda}_B$  = mean no. of birds on bird days  
 $\hat{W}_0$  = weather threshold value

SPRING

a) Sedge Warbler

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$
1961	6	11	0			
62	11	27	-1			
63	1	26	2			
64	11	21	0			
65	1	21	2			
66	7	9	0			
67	12	18	0			
68	3	7	1	1	6	-2
69	3	3	-4	1	1	0
70	5	9	1	2	11	-2
71	1	5	1			
72	1	3	1			

mean: 5.2 13.3 .3 1.3 18.0 -1.3

$\hat{\lambda}_B/\hat{\lambda}_N$ : 2.6 13.8

b) Willow Warbler

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$
1961	9	48	-2			
62	2	11	1			
63	17	200	-1			
64	9	44	1			
65	7	25	-1			
66	11	35	0			
67	5	75	0			
68	6	13	0	1	3	-2
69	9	35	-2	1	1	0
70	1	19	2	2	5	0
71	3	7	-1			
72	11	40	0			

mean: 7.5 46.0 -.3 1.3 3.0 -.7

$\hat{\lambda}_B/\hat{\lambda}_N$ : 6.1 2.3

c) Whitethroat

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$
1961	1	29	3			
62	23	84	-1			
63	14	50	0			
64	13	35	1			
65	13	59	1			
66	12	15	1			
67	16	21	1			
68	1	27	3	7	87	-1
69	4	7	0	1	3	-1
70	2	5	-1	3	13	-2
71	2	3	1			
72	3	5	2			

mean: 8.7 28.3 .9 3.7 34.3 -1.3

$\hat{\lambda}_B/\hat{\lambda}_N$ : 3.3 9.3

d) Redstart

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$
1961	1	5	0			
62	4	4	0			
63	2	23	-1			
64	1	5	0			
65	3	5	0			
66	1	5	-1			
67	1	5	1			
68	3	3	0	3	13	-2
69	1	1	0	1	1	0
70	1	7	1	1	7	-1
71	1	2	1			
72	1	3	-1			

mean: 1.7 5.7 0.0 1.7 7.0 -1.0

$\hat{\lambda}_B/\hat{\lambda}_N$ : 3.4 4.1

Table 7.3.2.2 Parameter estimates (maximum likelihood)  
for model (7.1)

$\hat{\lambda}_N$  = mean no. of birds on non-bird days  
 $\hat{\lambda}_B$  = mean no. of birds on bird days  
 $\hat{W}_0$  = weather threshold value

AUTUMN

a) Sedge Warbler

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$
1961						
62						
63						
64						
65						
66						
67	4	9	1			
68	10	13	0	2	7	2
69	6	9	-1	3	13	1
70	1	10	2	5	15	1
71						
72						

mean: 5.3 41.0 .5 3.3 35.0 1.3

$\hat{\lambda}_B/\hat{\lambda}_N$ : 7.7 10.6

b) Willow Warbler

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$
1961						
62						
63						
64						
65						
66						
67	17	33	-1			
68	1	49	3	14	39	0
69	17	26	1	17	34	-1
70	1	35	2	20	79	-1
71						
72						

mean: 9.0 35.8 1.3 17.0 50.7 -.7

$\hat{\lambda}_B/\hat{\lambda}_N$ : 4.0 3.0

c) Whitethroat

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$
1961						
62						
63						
64						
65						
66						
67	8	21	1			
68	1	31	3	20	29	0
69	1	3	1	6	11	1
70	1	7	2	5	13	0
71						
72						

mean: 2.8 15.5 1.8 10.3 17.7 0.3  
 $\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 5.5 1.7

d) Redstart

Year	Dungeness			Portland		
	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_O$
1961						
62						
63						
64						
65						
66						
67	11	22	0			
68	4	5	1	1	3	2
69	2	3	0	1	9	1
70	2	3	0	3	3	0
71						
72						

mean: 4.8 8.3 .3 1.7 5.0 1.0  
 $\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 1.7 2.9

arrived at, no estimates of standard errors or correlations are readily available. We shall gain some insight into the precision of the estimators however, when we examine contour plots of the likelihood surface with  $W_0$  fixed - see section 7.3.4.

For every data set studied, we find that  $\hat{\lambda}_B \geq \hat{\lambda}_N$ , as we would expect. Data sets that give  $\hat{\lambda}_B = \hat{\lambda}_N$  are those where discrimination (see chapter 5) was poor, or that no bird-days or no non-bird-days occurred.

One can examine the tables of estimates and compare spring and autumn, Dungeness and Portland, different species etc., but one finds no noticeable differences in the ratio of  $\hat{\lambda}_B$  to  $\hat{\lambda}_N$ , nor in their absolute values.

Having obtained a set of parameter estimates, which appear to be of the correct order of magnitude and relative magnitudes, we need to examine how well the model is fitting the data, and this is considered in the next section.

### 7.3.3 Goodness of fit

In order to see how well the model is fitting the data, we examine the magnitude of the residual variation, and also compare observed values with expected values.

As an indication of how one might proceed and investigate goodness of fit, we shall present two examples. These examples were chosen because plots of expected and observed

counts appeared to suggest that the fit was good. One of the examples includes a large 'fall' of migrants.

Fig. 7.3.3.1(a) shows a bar chart of the counts for the Whitethroat data from Portland (autumn 1970). Each count is labelled as a bird-day or a non-bird day ( $\times$  and  $\square$  respectively). Also shown are the expected counts on each day, under model (7.1), ( $*$  = bird day and  $0$  = non-bird day).

In this example one can see that bird days tend to coincide with days on which large numbers of birds are observed. The estimates of  $\lambda_N$  and  $\lambda_B$  were, for this data set, 5 and 13 respectively. There is some evidence of clustering among the residuals (Fig. 7.3.3.1(b))- i.e. the small residuals tend to be grouped together as do the larger residuals, but this is due to the serial correlation of the daily weather observations, and their combined effects in the discriminant functions. There is no evidence of particularly large residuals (ignoring sign), although this particular set of counts do not exhibit any large 'falls' of migrants.

A rough guide as to how well the model fits the data is given by the proportion of the variation within the data that is explained by the model. For the above example, 74% of the total variation is due to the model (7.1). It is not possible to carry out an F-test of the ratio of the mean squares for the model and for the error term since neither the residuals nor the data are normally distributed.

Fig. 7.3.3.1(a) Observed and expected counts under Model(7.1).

Whitethroat data, Portland, Autumn 1970.

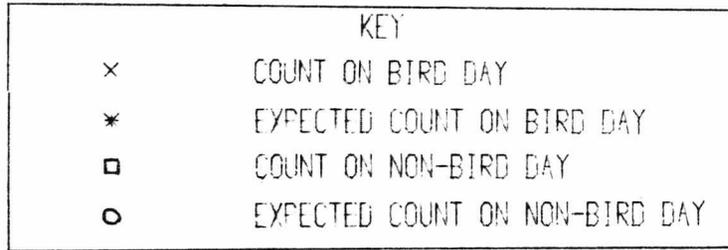
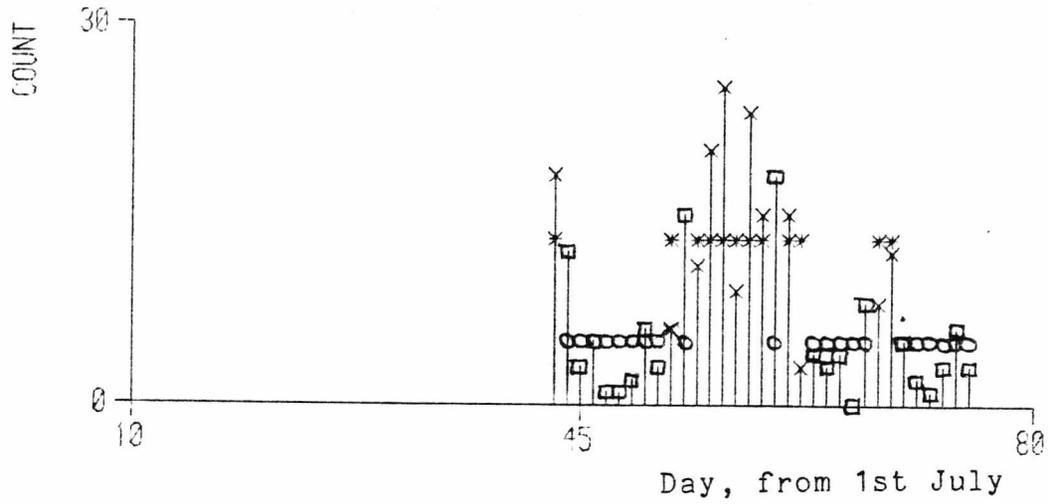
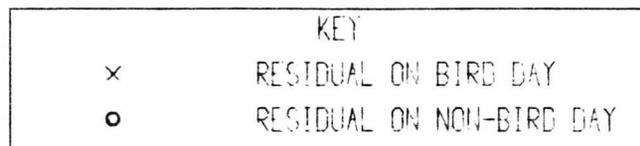
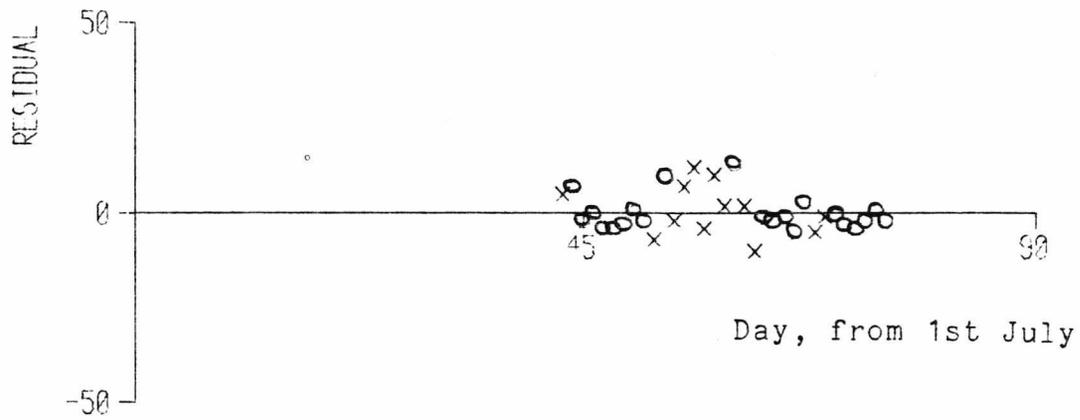


Fig. 7.3.3.1(b) Residuals



We can, however, obtain a rough guide to the significance of the residual variation by performing a chi-square goodness of fit test. (Note that the % of variation accounted for is a useful means of comparing alternative models - see section 7.3.5).

The  $\chi^2$  - statistic for the above example is 133.1 (28 degrees of freedom) which is highly significant, suggesting that the model is not fitting the data.

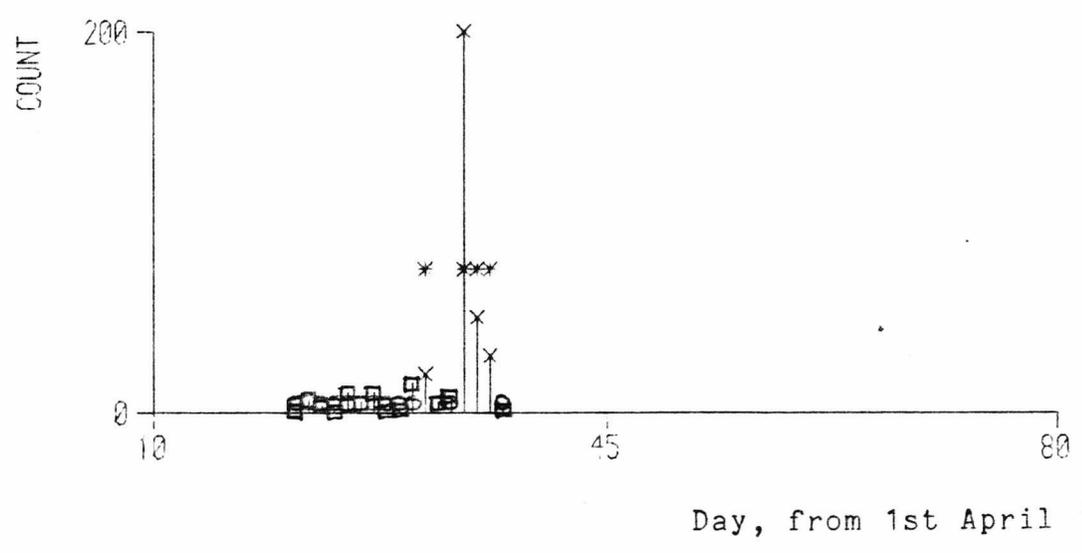
Fig. 7.3.3.2 shows a second example of observed and expected counts (Willow Warbler, Dungeness, spring 1967). These counts include one large 'fall' of birds, and consequently one residual is noticeably large.

The proportion of total variation accounted for by the model is, for this example, 51%. One would expect that this low figure is due mostly to the one particularly high count. This count corresponds to a residual of 125, and if the sum of squares of the data is reduced by  $125^2$ , then the proportion of variation accounted for becomes 84%. This is only a rough guide, since the residual would not be exactly zero, even for a more suitable model.

The  $\chi^2$  - goodness of fit statistic for this example is 352.2 (12 d.f.), or 149.9 without the large 'fall'. Either way, the model appears to be unsuitable for these data.

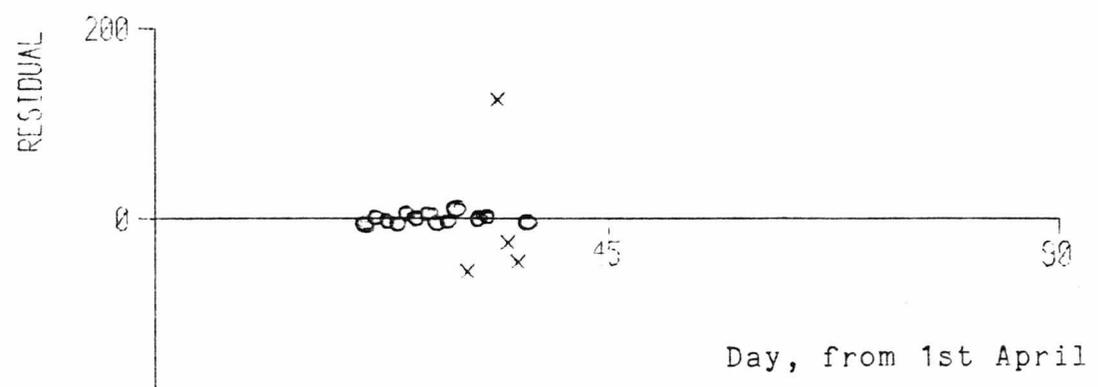
We now briefly examine the precision of the parameter

Willow Warbler data, Dungeness, Spring 1967.



KEY	
x	COUNT ON BIRD DAY
*	EXPECTED COUNT ON BIRD DAY
□	COUNT ON NON-BIRD DAY
o	EXPECTED COUNT ON NON-BIRD DAY

Fig. 7.3.3.2(b) Residuals



KEY	
x	RESIDUAL ON BIRD DAY
o	RESIDUAL ON NON-BIRD DAY

estimates and then move on to consider alternative models.

#### 7.3.4 Precision of parameter estimates

The parameter estimation procedure was carried out as follows. A series of values of  $W_0$  was examined and for each  $W_0$  the estimates of the parameters  $\lambda_N$  and  $\lambda_B$  were obtained on the basis of maximum likelihood. The optimum set of estimates  $\hat{\lambda}_N$ ,  $\hat{\lambda}_B$  and  $\hat{W}_0$  thus correspond to the overall maximum of the likelihood (or log likelihood).

In most cases  $\hat{W}_0$  turned out to be close to zero, although the maximum value of the likelihood did not alter substantially for different values of  $W_0$ . We shall examine the effect on the likelihood value when the parameters  $\lambda_N$  and  $\lambda_B$  are varied. We can do this by plotting contours of the log-likelihood surface (log-likelihood is used in order to avoid using extremely small numbers).

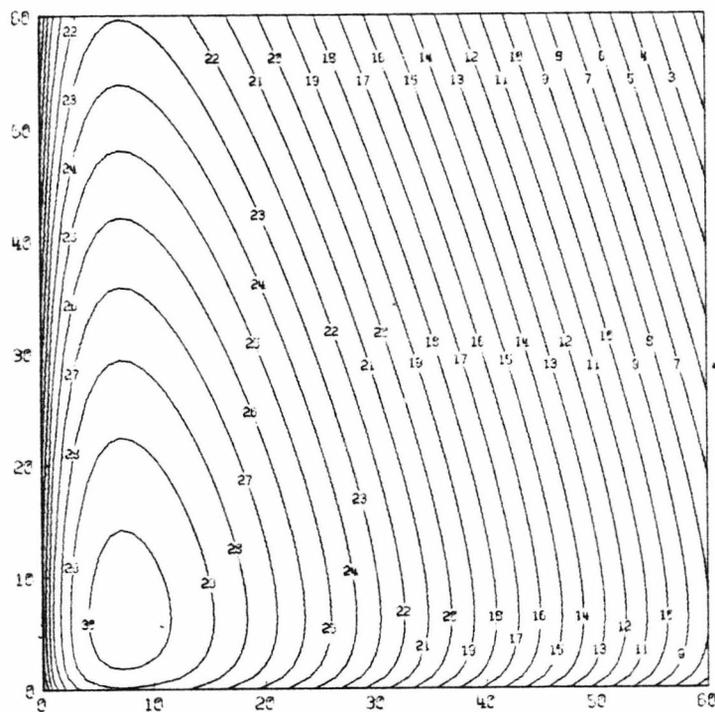
Fig. 7.3.4.1 shows two examples of such contour plots. The Sedge Warbler example suggests that  $\lambda_B$  is estimated more precisely than  $\lambda_N$  since each contour covers a larger range of values for  $\lambda_N$  than for  $\lambda_B$ . The same feature is seen with the Whitethroat data, although to a lesser extent. Indeed, this feature was found on most contour plots for this model.

In practical terms this means that on bird days (days with, for example, bad weather), we are likely to get a more accurate picture of migration volume than on non-bird days

Fig. 7.3.4.1 Contour plots of Log-likelihood surface for Model(7.1)

( Contours are numbered 1-30 in increasing order of height )

$\lambda_N$

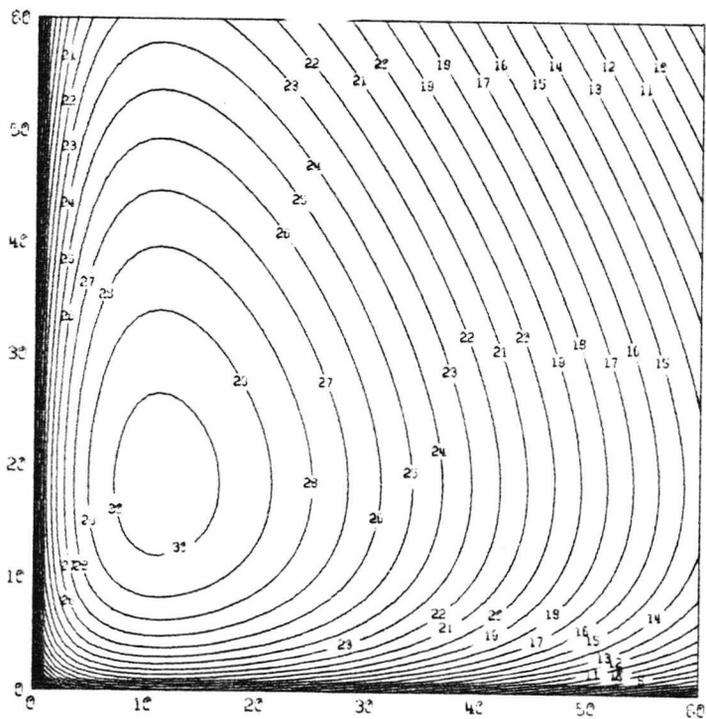


(a) Sedge Warbler  
Dungeness  
Spring, 1961

$$\hat{W}_0 = 0$$

$\lambda_B$

$\lambda_N$



(b) Whitethroat  
Dungeness  
Spring, 1966

$$\hat{W}_0 = 1$$

$\lambda_B$

when we have little idea as to whether birds are migrating overhead or not.

Fig. 7.3.4.2(a) shows an example where the value of  $\lambda_N$  has no effect on the likelihood of the data at all, this being an indication that in this particular year there were no non-bird days according to the local weather conditions.

Fig. 7.3.4.2(b) is another example of where the precision of  $\hat{\lambda}_B$  is much greater than for  $\hat{\lambda}_N$ . The highest contour on this map spans 45 units for  $\lambda_N$ , and 10 units for  $\lambda_B$ .

The information gained from looking at the contour plots is useful for when we come to form migration indices. For migration seasons which consist of more bird days than non-bird days, the migration index will be less precisely estimated than an index calculated from a migration season which consists of more non-bird days than bird days.

#### 7.3.5 Model (7.1) with serial dependency : model (7.2)

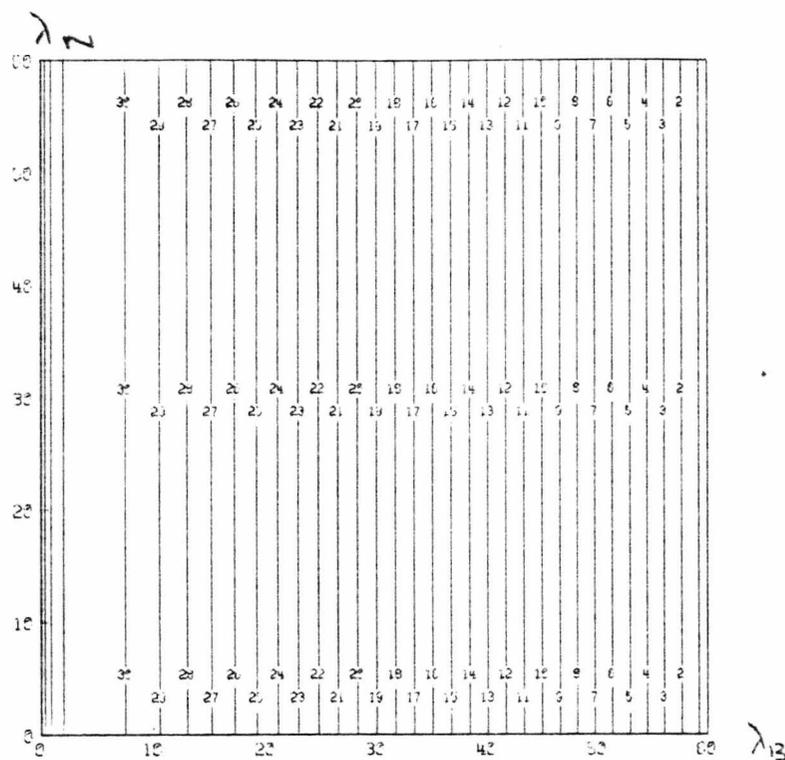
The work of chapter 6 (fitting Poisson-binomial models to sequences of daily counts) suggested that each daily count of grounded migrants is, in general, partially dependent on the count of the previous day.

We now incorporate this feature into model (7.1), and compare the fit of the new model - model (7.2) - with that of model (7.1).

Denoting the count on day  $n$  by  $X_n$ , model (7.2) can be written

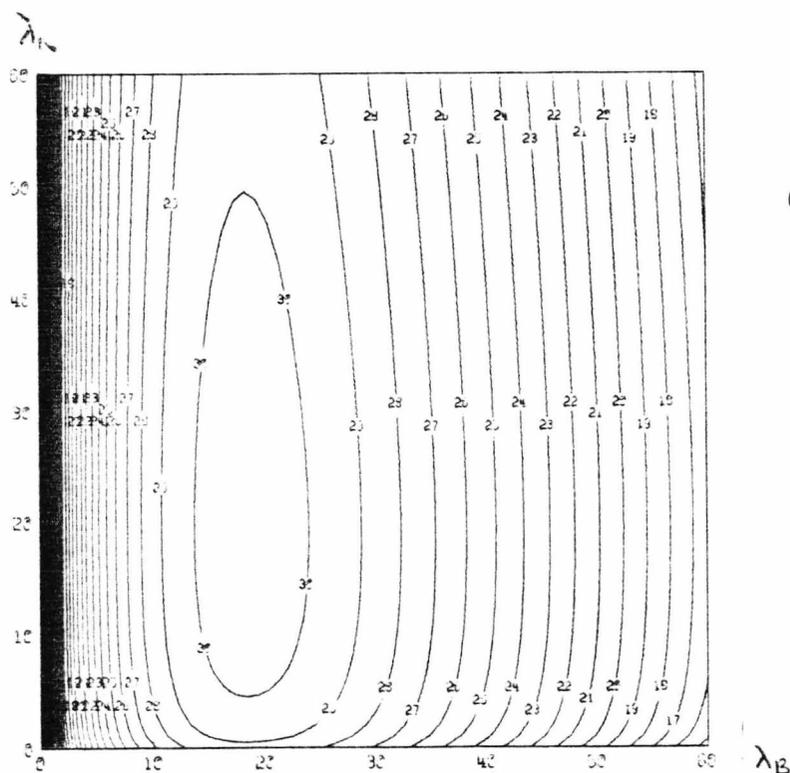
Fig. 7.3.4.2 Contour plots of Log-likelihood surface for Model(7.1)

( Contours are numbered 1-30 in increasing order of height )



(a) Sedge Warbler  
Portland  
Spring, 1968

$$\hat{W}_0 = 0$$



(b) Whitethroat  
Dungeness  
Spring, 1967

$$\hat{W}_0 = 0$$

$$X_n = Y_n + \begin{cases} I_n & \text{if } W_n \geq W_0 \\ I'_n & \text{if } W_n < W_0 \end{cases}$$

where  $Y_n \sim \text{BIN}(x_{n-1}, p)$ ,  $I_n \sim \text{POI}(\lambda_N)$ ,  $I'_n \sim \text{POI}(\lambda_B)$ ,  $W_n$  is the discriminant score for day  $n$  and  $W_0$  is the threshold value, again a parameter of the model. We now have four parameters to estimate ( $\lambda_B$ ,  $\lambda_N$ ,  $p$  and  $W_0$ ), and so we use the Nelder-Mead simplex method. The parameter estimates are given in Tables 7.3.5.1 and 7.3.5.2. The tables are set out in the same style as those tables containing the estimates for model (7.1) in order that comparisons may be made with ease.

One feature of the tables is that, in most cases, the estimates are quite different between the two models. These differences are, of course, attributable to the inclusion of the extra parameter  $p$ . Note that in some cases however, including average estimates (over all years considered), the results are much the same.

There is no consistent difference between the two models as regards the ratio of average estimates of  $\lambda_B$  and  $\lambda_N$ . In some cases model (7.1) gives a higher ratio, and in other cases model (7.2) gives a higher ratio.

Only three out of 88 estimates of  $p$ , the binomial parameter of model (7.2), are less than 0.5. This is very much in keeping with the results of chapter 6, where we fitted the simple Poisson-binomial model to the data.

Table 7.3.5.1 Parameter estimates for model (7.2)

$\hat{p}$  = binomial parameter,  $\hat{\lambda}_N, \hat{\lambda}_B$ : mean counts for non-bird days and bird days respectively,  $\hat{w}_0$  = weather threshold value.

SPRING

a) Sedge Warbler

Year	DUNGENESS				PORTLAND			
	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_0$	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_0$
1961	.59	1	6	0				
62	.11	5	17	0				
63	.24	10	20	2				
64	.14	8	20	0				
65	.37	0	16	1				
66	.49	4	5	-1				
67	.47	6	11	0				
68	.35	4	8	-3	.12	1	1	-2
69	.14	3	3	-4	.27	0	1	-4
70	.16	3	8	0	.12	2	9	-2
71	.00	4	5	-3				
72	.03	2	6	-3				

mean: .26 4.2 10.4 -1 .17 1.0 3.7 -3  
 $\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$  2.5 3.7

b) Willow Warbler

1961	.06	3	21	0				
62	.06	7	7	-3				
63	.34	1	12	2				
64	.11	6	15	1				
65	.30	0	9	1				
66	.21	7	19	0				
67	.00	5	37	0				
68	.11	5	6	-3	.19	0	2	-2
69	.00	7	11	0	.01	0	1	-3
70	.22	0	8	2	.12	2	4	-3
71	.01	2	4	0				
72	.03	2	6	-3				

mean: .12 3.8 12.9 0 .11 .7 2.3 -3  
 $\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$  3.4 3.3

SPRING

c) Whitethroat

Year	DUNGENESS				PORTLAND			
	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_0$	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_0$
1961	.15	7	17	0				
62	.08	14	18	1				
63	.11	19	20	-3				
64	.09	10	11	0				
65	.00	12	25	1				
66	.00	12	16	1				
67	.22	11	16	-2				
68	.18	6	8	-3	.19	2	9	0
69	.39	4	8	-2	.25	1	16	-3
70	.38	1	5	-1	.07	2	35	-2
71	.00	4	8	-3				
72	.28	3	5	0				
mean:	.16	8.6	13.1	-1	.17	1.7	20	-2
$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ :		1.5				11.8		

d) Redstart

1961	.04	2	10	-2				
62	.16	3	10	-3				
63	.13	1	19	0				
64	.00	2	6	-4				
65	.25	3	6	-4				
66	.15	1	6	-1				
67	.10	1	6	0				
68	.42	2	14	-1	.30	1	8	-1
69	.00	1	4	-3	.16	0	0	-4
70	.00	5	6	2	.00	1	7	-1
71	.07	1	2	0				
72	.27	1	9	-4				
mean:	.71	1.9	8.2	-2	.15	.7	5.0	-2
$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ :		4.3				7.1		

Table 7.3.5.2 Parameter estimates for model (7.2)

$\hat{p}$  = binomial parameter,  $\hat{\lambda}_N, \hat{\lambda}_B$  = mean counts for non-bird days and bird days respectively.

$\hat{W}_0$  = weather threshold value.

AUTUMN

a) Sedge Warbler

Year	DUNGENESS				PORTLAND			
	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{W}_0$
1967	.31	5	8	-3				
68	.28	7	11	0	.17	2	7	0
69	.17	5	16	-3	.52	1	4	1
70	.01	6	13	0	.09	4	15	1

mean: .19 5.8 12.0 .26 2.0 8.7 .7

$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 2.1 4.3

b) Willow Warbler

1967	.11	14	17	2				
68	.39	13	24	2	.39	9	14	0
69	.27	6	12	3	.22	12	26	0
70	.06	0	23	2	.37	5	14	2

mean: .21 5.8 19.0 2.3 .33 8.7 18.0 1

$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 3.3 2.1

c) Whitethroat

Year	DUNGENESS				PORTLAND			
	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_O$	$\hat{p}$	$\hat{\lambda}_N$	$\hat{\lambda}_B$	$\hat{w}_O$
1967	.26	7	14	1				
68	.19	19	23	3	.28	6	17	2
69	.14	8	24	2	.19	8	11	-2
70	.43	0	5	1	.23	4	13	0

mean: .26 8.5 11.3 2 .23 6.0 13.7 0

$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 1.3 2.3

d) Redstart

1967	.00	10	21	0				
68	.32	3	9	-3	.52	1	16	-2
69	.31	2	5	-2	.24	6	7	-4
70	.32	2	9	-4	.45	1	9	-2

mean: .24 4.3 11.0 -2 .40 2.7 10.7 -3

$\frac{\hat{\lambda}_B}{\hat{\lambda}_N}$ : 2.6 3.9

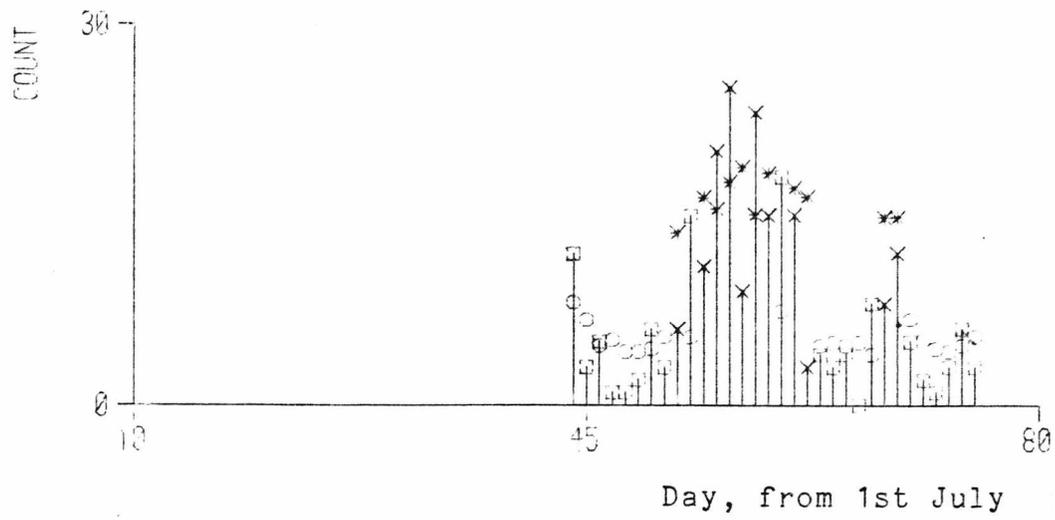
The standard errors of the estimates were estimated to be of the order of 0.05 for  $\hat{p}$ , and 0-1 for  $\hat{\lambda}_N$  and  $\hat{\lambda}_B$ . The estimator for  $W_0$  was estimated in most cases, to have an extremely large standard error, of the order of 1000. This unusual feature suggests that the likelihood is maximised for a large range of values of  $W_0$ , and is due to some seasons consisting entirely of either bird-days or non bird-days.

The examples that were used to illustrate the goodness of fit of model (7.1) are now used for model (7.2).

Fig. 7.3.5.1 (c.f. Fig. 7.3.3.1) shows the observed and expected counts of Whitethroats (Portland, autumn, 1970). The expected counts are no longer constant within bird days or non-bird days, and the fit appears to be better than for model (7.1). The residuals, however, appear to be of a similar magnitude, and the proportion of total variation accounted for by the model is also almost identical: 73% compared with 74% for model (7.1). The goodness of fit statistic is 102.1 (133.1 for model (7.1) ), which is again highly significant.

The second example (Willow Warbler data, Dungeness, spring, 1967) is shown in Fig. 7.3.5.2 (c.f. Fig. 7.3.3.2). The plot of the residuals shows that the fit is generally better than for model (7.1), but the residual corresponding to the extremely high count is larger, leading to a proportion of variation accounted for of 39%, compared with 51% for model (7.1). Note that by adjusting for the large

Fig. 7.3.5.1(a) Model(7.2) expected and observed counts ,  
Whitethroat data, Portland, Autumn 1970.



KEY	
x	COUNT ON BIRD DAY
✕	EXPECTED COUNT ON BIRD DAY
□	COUNT ON NON-BIRD DAY
○	EXPECTED COUNT ON NON-BIRD DAY

Fig. 7.3.5.1(b) Residuals

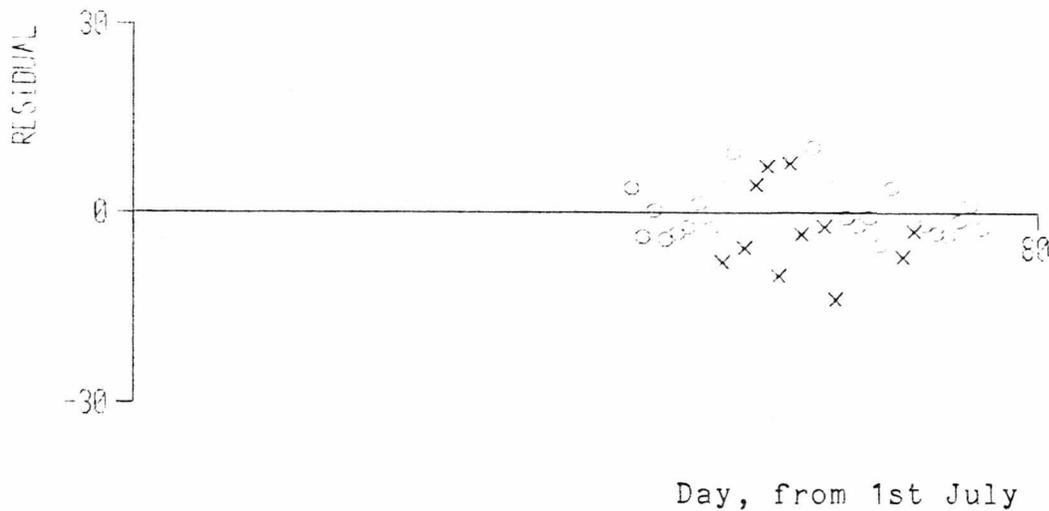
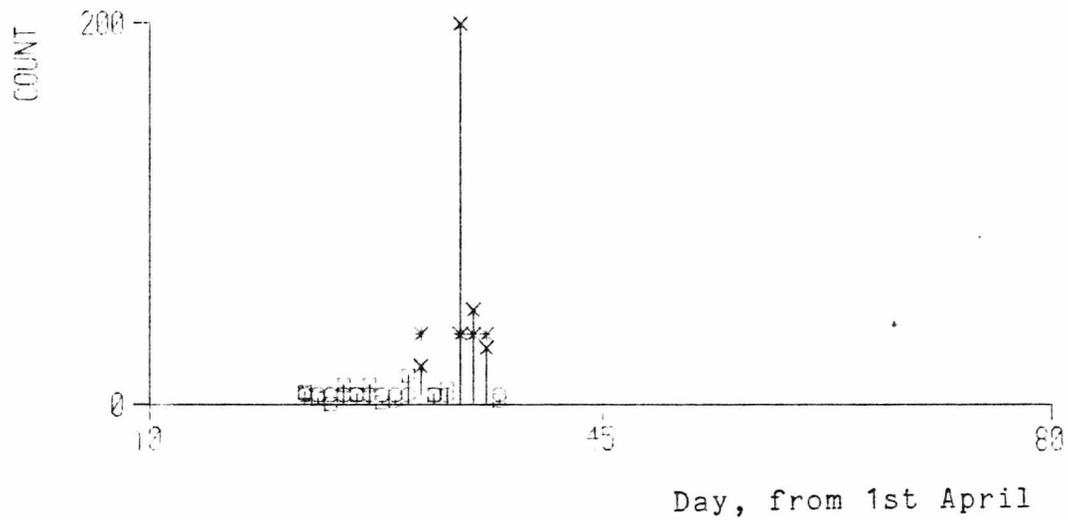


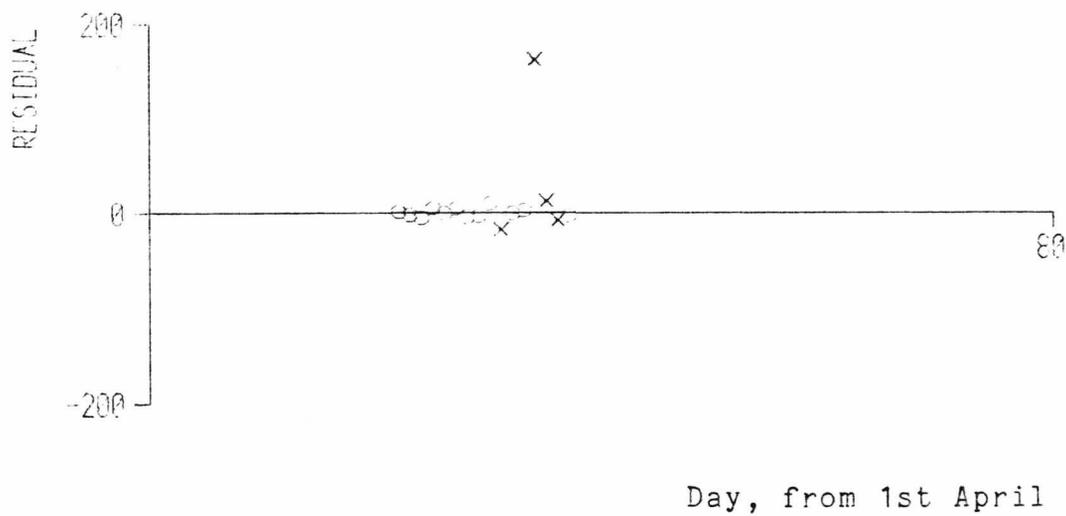
Fig. 7.3.5.2(a) Model(7.2) expected and observed counts .

Willow Warbler data, Dungeness, Spring 1967.



KEY	
x	COUNT ON BIRD DAY
*	EXPECTED COUNT ON BIRD DAY
□	COUNT ON NON-BIRD DAY
o	EXPECTED COUNT ON NON-BIRD DAY

Fig. 7.3.5.2(b) Residuals



residual, the proportion of variation accounted for becomes 98%, compared with 84% for model (7.1).

The goodness of fit statistic for this example is 124.8, or 12.7 without the large 'fall'. It appears that for most of the data set, this model fits the data well ( $\chi^2 (.05) = 19.7$  ).

Although the study of goodness of fit of models (7.1) and (7.2) is based on only two examples, it should be remembered that these examples were selected since a visual examination suggested that for these examples the fit was better than for other data sets. There is no conclusive evidence so far, that models of the type suggested in this chapter are suitable for daily counts, and we now examine one further model.

#### 7.3.6 A generalisation of model (7.2) : model (7.3)

The model that is proposed in this section is a generalisation of model (7.2) in that the poisson input component of the model is a continuous function of the discriminant score, rather than a switching process between two distinct poisson variates. Thus we no longer compress information contained in the sequence  $\{W_n\}$  into a binary process. An immediately appealing feature of this model (model (7.3) ), is that the number of parameters is reduced by one.

In order to keep the model as simple as possible, we consider linear functions of the  $\{W_n\}$ . The model is written:

$$X_n = Y_n + A_n ,$$

where  $Y_n \sim \text{BIN}(x_{n-1}, p)$  and  $A_n \sim \text{POI}(a + \beta W_n)$ . The sequence  $\{W_n\}$  is, of course, a realisation of some stochastic process determined by the discriminant analyses. However, we shall treat  $\{W_n\}$  as a fixed sequence of observations measured without error - rather like an explanatory variable in regression analysis. The three parameters that we wish to estimate are  $p$ ,  $a$  and  $\beta$ , and estimates are readily obtained by using the Nelder-Mead simplex method, as are estimates of asymptotic standard errors and correlations.

Tables 7.3.6.1 and 7.3.6.2 show the estimates of the parameters. The asymptotic standard errors of these estimators were estimated to be of the order of less than 0.1 for  $\hat{p}$ , and approximately equal to 1.0 for  $\hat{a}$  and  $\hat{\beta}$ . In many cases  $\hat{\beta}$  turns out to be zero, or close to zero, suggesting that the Poisson input to each days count is independent of weather (as measured by  $\{W_n\}$ ), and that a model of the form of model (6.1) is appropriate, i.e.

$a = \lambda$ , using model (6.1) notation. This is not to say, of course, that weather is not relevant when modelling a sequence of daily counts, but it is perhaps an indication that model (7.3) is not a suitable form of model in this case. The proportions of total variation accounted for by model (7.3) for the two example data sets studied in previous sections are 55% and 16% for the Whitethroat and Willow Warbler examples respectively. If one ignores the single large residual for the Willow Warbler data, the

Table 7.3.6.1 Parameter estimates for model (7.3)

$\hat{p}$  = binomial parameter,  
 $\left. \begin{matrix} \hat{a} \\ \hat{\beta} \end{matrix} \right\}$  parameters for linear function of  
weather threshold value.

SPRING

a) Sedge Warbler

Year	DUNGENESS			PORTLAND		
	$\hat{p}$	$\hat{a}$	$\hat{\beta}$	$\hat{p}$	$\hat{a}$	$\hat{\beta}$
1961	.57	2.9	1.1			
62	.21	12.2	0.0			
63	.43	12.9	5.6			
64	.43	7.7	0.1			
65	.58	8.2	1.8			
66	.56	3.2	0.0			
67	.47	7.7	0.0			
68	.34	3.9	0.0	.00	3.5	1.0
69	.06	2.9	0.0	.24	0.3	0.0
70	.21	7.9	0.0	.25	2.4	0.0
71	.14	3.4	0.6			
72	.14	1.7	0.0			

mean: .35 6.2 0.8 .16 2.1 0.3

b) Willow Warbler

1961	.33	7.6	0.0			
62	.11	7.7	0.4			
63	.44	11.3	0.5			
64	.44	10.4	0.0			
65	.54	5.1	0.6			
66	.37	10.9	0.6			
67	.37	8.5	0.0			
68	.08	7.3	0.0	.00	1.6	.28
69	.21	7.1	0.2	.00	0.6	.11
70	.32	11.5	4.4	.13	3.0	.00
71	.00	4.1	0.4			
72	.25	13.5	0.0			

mean: 0.29 8.8 0.6 .04 1.7 .32

SPRING

c) Whitethroat

	DUNGENESS			PORTLAND		
Year	$\hat{p}$	$\hat{a}$	$\hat{\beta}$	$\hat{p}$	$\hat{a}$	$\hat{\beta}$
1961	.11	20.7	0.0			
62	.28	15.3	0.0			
63	.23	19.8	0.0			
64	.33	13.2	0.0			
65	.78	4.5	0.0			
66	.21	11.9	0.0			
67	.48	7.5	1.2			
68	.16	16.9	1.3	.00	8.9	2.0
69	.51	2.9	0.0	.00	1.3	0.0
70	.37	2.2	0.5	.10	3.8	0.6
71	.00	4.0	1.6			
72	.30	3.0	0.0			
mean:	.31	10.1	0.4	0.03	4.7	0.9

d) Redstart

1961	.08	2.2	0.0			
62	.22	3.2	1.3			
63	.15	4.0	0.1			
64	.07	2.1	0.0			
65	.35	3.0	0.0			
66	.19	1.9	0.3			
67	.24	2.0	0.3			
68	.46	1.7	0.0	.28	3.7	0.3
69	.00	0.8	0.0	.22	0.4	0.0
70	.09	6.2	3.4	.12	2.5	0.0
71	.07	1.5	0.0			
72	.24	1.0	0.3			
mean:	.18	2.5	0.7	.21	2.2	0.1

Table 7.3.6.2 Parameter estimates for model (7.3)

AUTUMN

a) Sedge Warbler

Year	DUNGENESS			PORTLAND		
	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
1961						
62						
63						
64						
65						
66						
67	.38	4.7	0.0			
68	.41	6.7	0.0	.27	3.5	0.0
69	.27	4.6	0.9	.32	4.6	0.0
70	.15	7.6	0.0	.19	6.4	0.0
71						
72						
mean:	.30	5.9	0.2	0.26	4.8	0.0

b) Willow Warbler

1961						
62						
63						
64						
65						
66						
67	.25	14.7	0.0			
68	.54	17.1	0.5	.48	11.1	0.0
69	.30	15.3	1.0	.42	12.9	0.0
70	.19	17.0	0.0	.41	13.1	0.6
71						
72						
mean:	.32	21.3	.4	.43	12.4	0.2

AUTUMN

c) Whitethroat

	DUNGENESS			PORTLAND		
Year	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{p}$	$\hat{\alpha}$	$\hat{\beta}$
1961						
62						
63						
64						
65						
66						
67	.39	9.8	0.0			
68	.52	12.6	0.0	.39	14.3	0.0
69	.27	1.5	0.2	.37	6.7	0.0
70	.46	3.7	0.0	.39	5.1	0.0
71						
72						
mean:	0.41	7.2	0.1	0.4	8.7	0.0

d) Redstart

1961						
62						
63						
64						
65						
66						
67	.43	9.8	0.0			
68	.34	3.1	0.5	.47	1.5	0.0
69	.29	2.2	1.1	.41	4.9	0.0
70	.30	1.7	0.0	.43	1.3	0.4
71						
72						
mean:	.34	4.2	0.4	.43	2.6	0.1

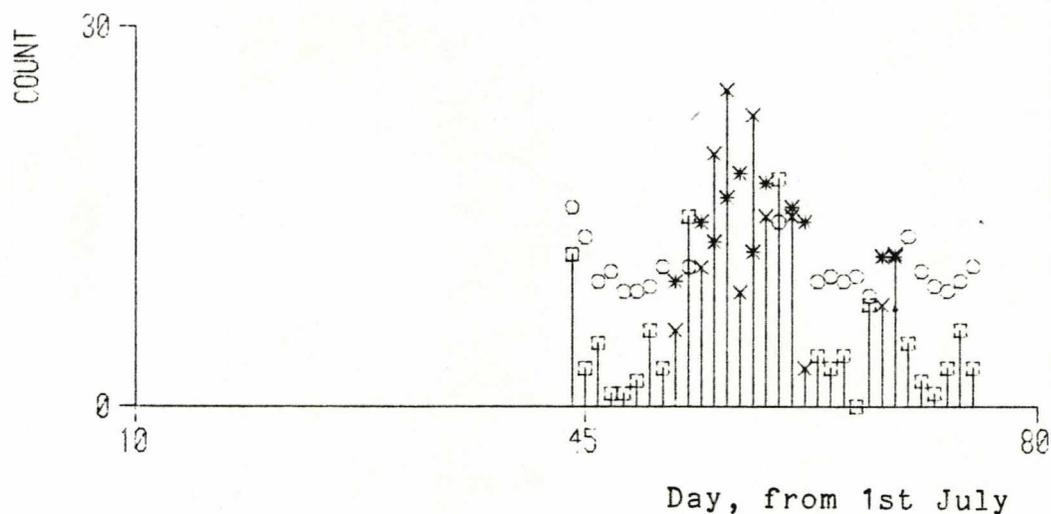
figure rises to 96%.

If one compares these proportions with those obtained using models (7.1) and (7.2), then it appears that model (7.3) gives a poorer fit, and that one can do better with a model with a 'switching' process for the Poisson component. In practical terms, this result is one which we would have expected. On a given day, each bird has to arrive at a decision - whether to land at an observatory or not, or whether to migrate from some other locality or not. There are only two options open to each bird. Further, there is no reason to suppose that, say, half the migrating birds choose to land, and the other half choose not to land. It is encouraging to see that the data back up this commonsense reasoning.

A further reason why we would have expected this model to fit less well than the other models is that it contains one less parameter.

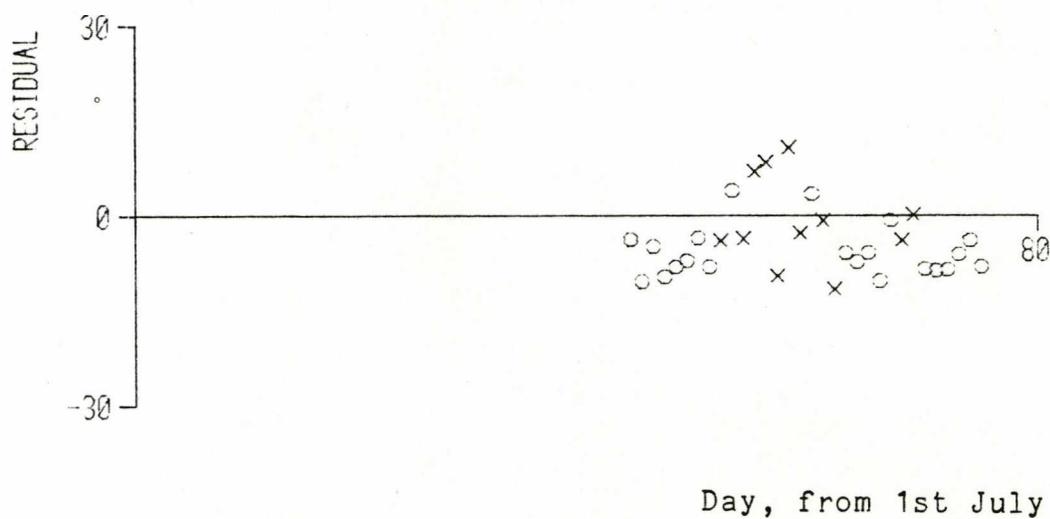
Figs. 7.3.6.1 and 2 show the observed and expected counts for the two examples, together with residuals. If one compares these with the diagrams for model (7.2), then we see evidence that model (7.3) offers a poorer fit, particularly for non-bird days where most counts are over-estimated. The goodness of fit statistics for the two examples are 117.9 and 456.2 respectively (the latter statistic becomes 44.5 after adjusting for the large 'fall').

Fig. 7.3.6.1(a) Model(7.3) expected and observed counts ,  
Whitethroat data, Portland, Autumn 1970.



KEY	
x	COUNT ON BIRD DAY
*	EXPECTED COUNT ON BIRD DAY
□	COUNT ON NON-BIRD DAY
○	EXPECTED COUNT ON NON-BIRD DAY

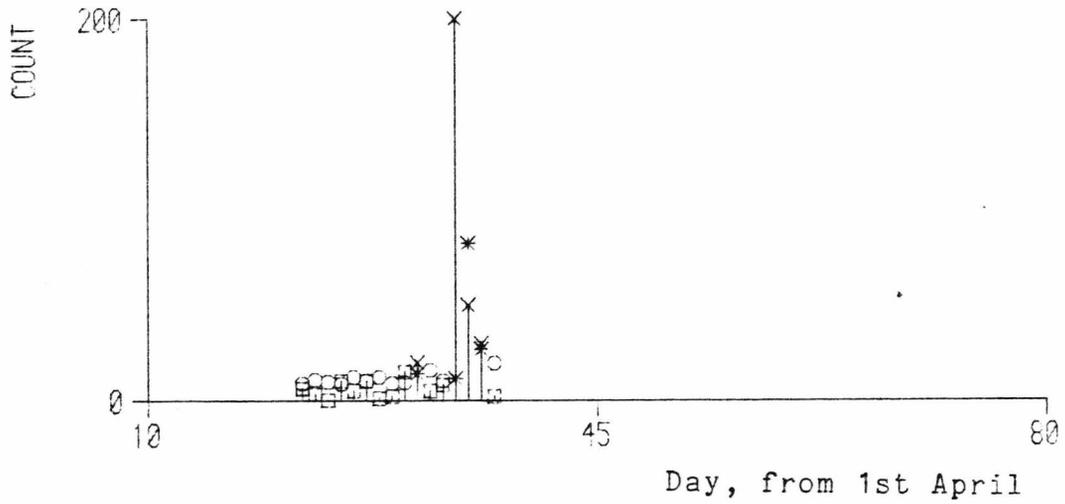
Fig. 7.3.6.1(b) Residuals



KEY	
x	RESIDUAL ON BIRD DAY
○	RESIDUAL ON NON-BIRD DAY

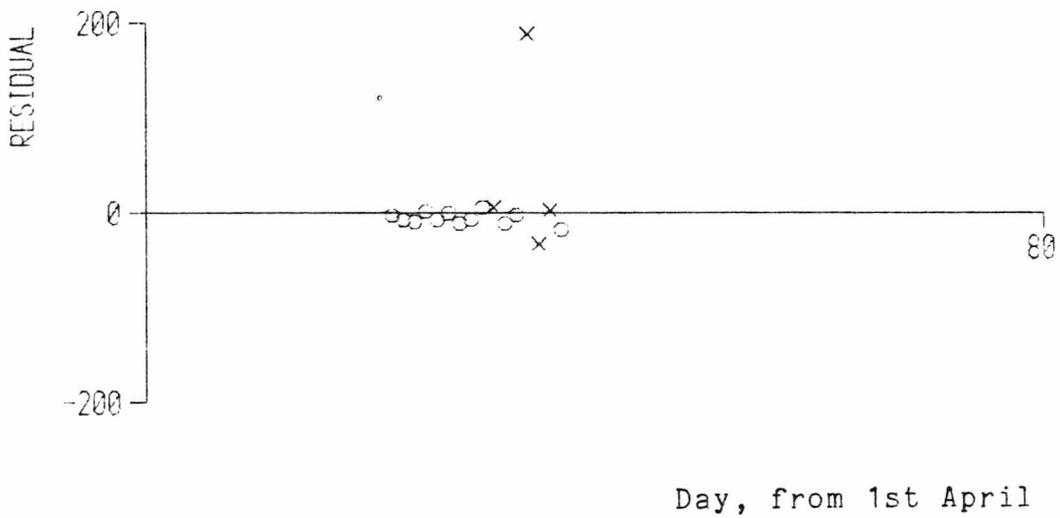
Fig. 7.3.6.2(a) Model(7.3) expected and observed counts .

Willow Warbler data, Dungeness, Spring 1967.



KEY	
x	COUNT ON BIRD DAY
*	EXPECTED COUNT ON BIRD DAY
□	COUNT ON NON-BIRD DAY
o	EXPECTED COUNT ON NON-BIRD DAY

Fig. 7.3.6.1(b) Residuals



KEY	
x	RESIDUAL ON BIRD DAY
o	RESIDUAL ON NON-BIRD DAY

In summary, there is evidence to suggest that, out of the three models considered, model (7.2) gives the best fit to the data, but even this model (switching process with serial correlation component) fails to fit the data well.

In spite of the above conclusion, we shall briefly examine migration indices obtained from the models of this chapter, in the next section. We only consider the switching-process models, since model (7.3) showed no indication of fitting the data.

#### 7.4 Migration indices

It is straightforward to obtain indices that estimate annual migration volume from the parameter estimates for models (7.1) and (7.2). In view of the way in which the models have been constructed, these indices may be thought of as 'weather corrected' indices. This section examines the correlation between migration indices and CBC indices.

##### 7.4.1 An index derived from model (7.1)

In fitting this model we assumed that, on average,  $\hat{\lambda}_B$  birds are recorded on each day that is designated a bird-day, and  $\hat{\lambda}_N$  birds are recorded on each day that is designated a non bird-day. An obvious choice of index is, for each season, to multiply  $\hat{\lambda}_B$  by the number of bird days,  $\hat{\lambda}_N$  by the number of non bird-days, add the two figures together and rescale the indices according to the B.T.O's convention (index for 1966 is equal to 100).

Table 7.4.1 shows the indices for (a) Sedge Warbler, (b) Willow Warbler and (c) Whitethroat. The data are from Dungeness, and are spring counts. These examples are those that allow comparisons with CBC indices over a suitable number of years. The corresponding CBC (farmland) indices are also shown in the tables, together with the relevant sample correlation coefficients. The story is much the same as in previous chapters: the large fluctuations in the population size of the Whitethroat is reflected in the indices, but otherwise there is little evidence of any correlation between the migration indices and the CBC indices. The indices are plotted in Fig. 7.4.1, which illustrate the lack of correlation for the Sedge Warbler and Willow Warbler data, and the positive correlation for the Whitethroat data - with the exception of 1962-63, where the migration indices were particularly high. This last feature has been found with other migration indices - see section 2.4.

An examination was also made of year-to-year changes in the indices, and no evidence was found to suggest a relationship between the CBC index and migration index for all three species. The Spearman rank correlation coefficient also suggests that for the Sedge Warbler and Willow Warbler data, no relationship between CBC and migration indices exists. The rank correlation between the two indices for the Whitethroat was, however, significant at the 5% level.

The lack of weather data from more observatories and for

Table 7.4.1 Migration indices from model (7.1)  
(Dungeness spring data)

a) Sedge Warbler

Year	Index	CBC
1961	129	-
62	314	60
63	91	59
64	86	80
65	33	78
66	100	100
67	125	66
68	36	113
69	20	63
70	35	79
71	35	65
72	18	54

Correlation -.20

b) Willow Warbler

Index	CBC
73	-
31	53
179	63
37	72
42	87
100	100
54	96
35	115
100	114
43	120
23	96
76	99

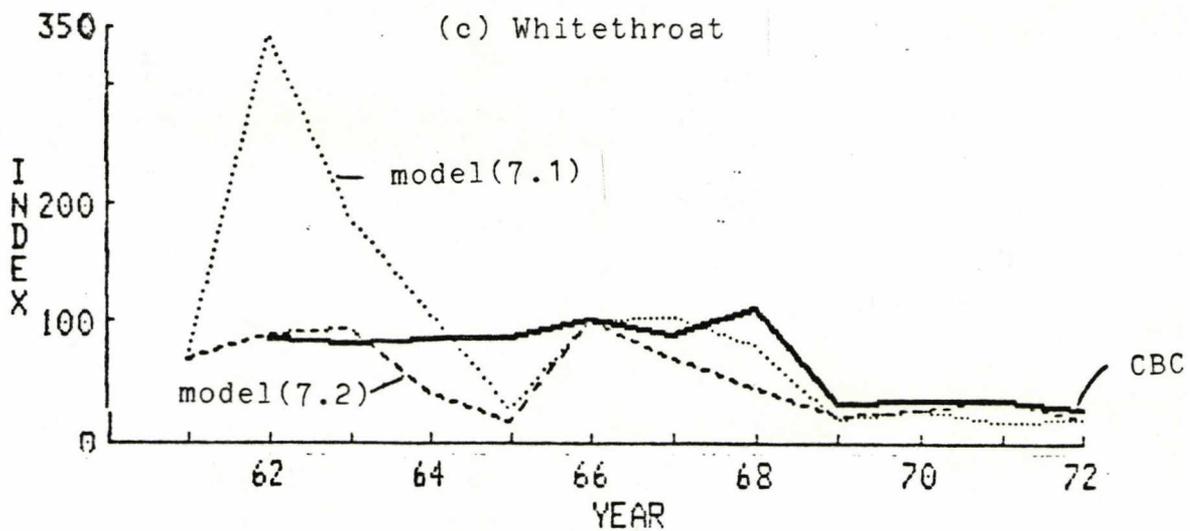
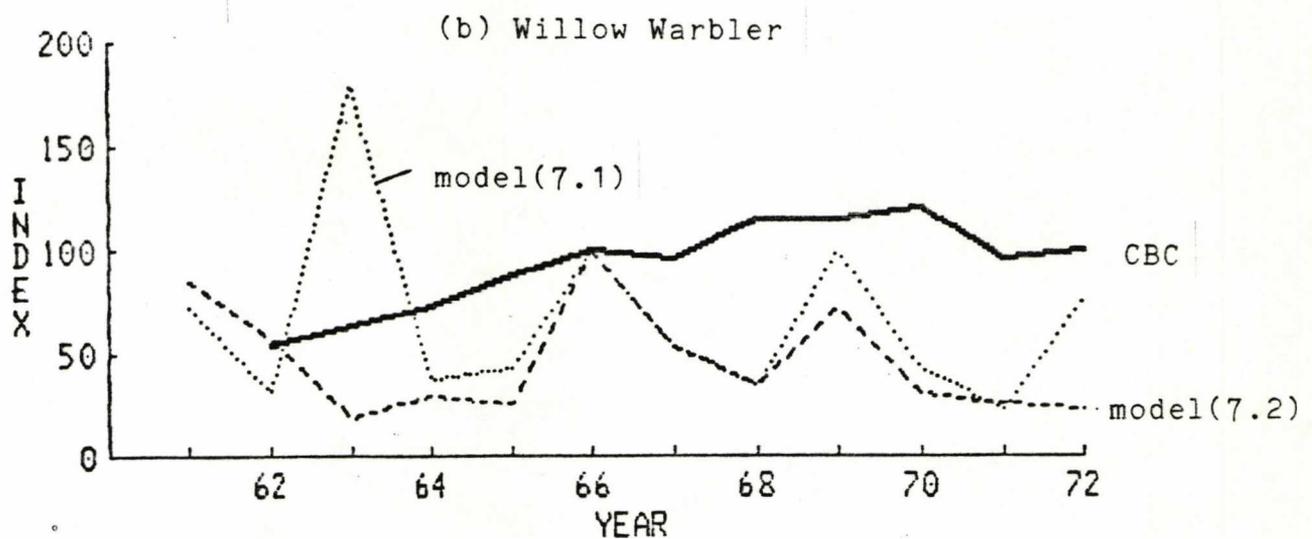
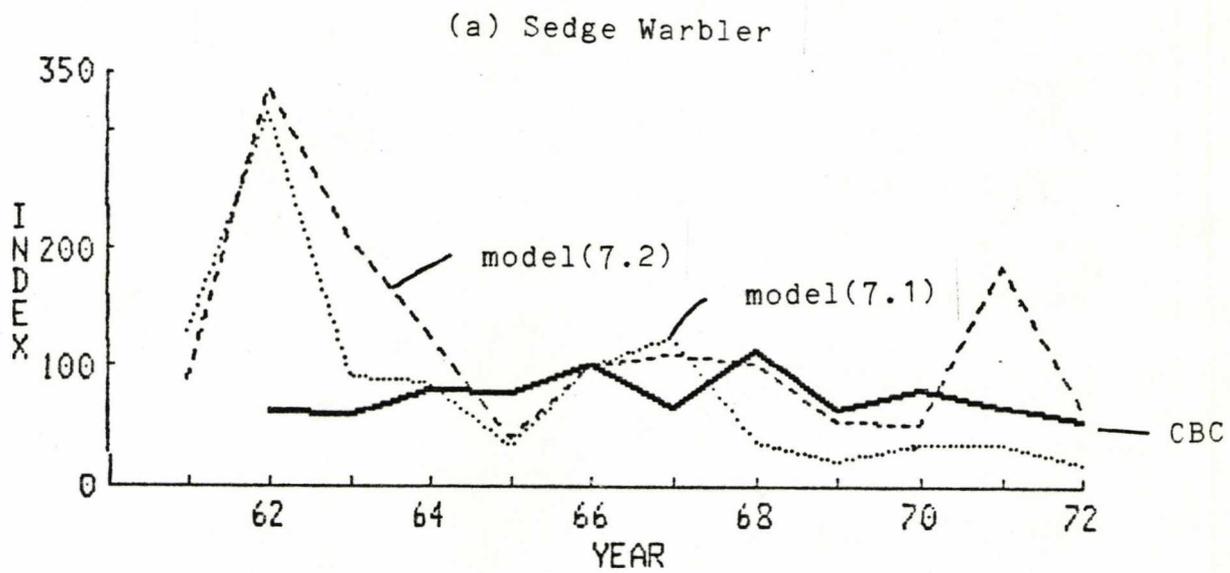
-.17

c) Whitethroat

Year	Index	CBC
1961	68	
62	343	85
63	188	80
64	108	84
65	24	84
66	100	100
67	102	86
68	80	110
69	19	30
70	26	31
71	15	33
72	19	26

Correlation .50 (not significant at 5% level)

Fig. 7.4.1 Migration indices from Models(7.1) and (7.2)



autumn migration seasons prevents the formation of a combined migration index for each species which proved successful in chapter 6. It is therefore not possible to conclude that weather data are or are not needed for the formation of migration indices. If one considers individual indices rather than combined indices, there seems to be no evidence in favour of using weather data, but it must be remembered that only spring data have been studied in detail, and only data from one observatory.

#### 7.4.2 An index of migration derived from model (7.2)

An index based on parameter estimates of model (7.2) can be derived in a similar manner to the index derived from model (7.1). In this case,  $\hat{\lambda}_B$  and  $\hat{\lambda}_N$  are calculated while correcting for serial correlation, or in other words, the possibility of the same birds being recorded on successive days. Table 7.4.2 presents the indices together with CBC indices. Again, only the index for the Whitethroat appears to be significantly correlated with the CBC index.

The indices are plotted in Fig. 7.4.1 with the model (7.1) indices and CBC indices. The two migration indices are highly correlated (significant at 1% level), and the only major difference between them appears to be that the model (7.2) index is less variable during the early part of the time period studied for the Willow Warbler and Whitethroat data.

An examination of the year-to-year changes of the indices

Table 7.4.2      Migration indices from model (7.2)  
(Dungeness spring data)

a) Sedge Warbler

Year	Index	CBC
1961	90	-
62	335	60
63	208	59
64	127	80
65	40	78
66	100	100
67	110	66
68	103	113
69	53	63
70	50	79
71	187	65
72	65	54

Correlation    -.30

b) Willow Warbler

Index	CBC
54	-
58	53
19	63
29	72
25	87
100	100
54	96
34	115
73	114
30	120
25	96
22	99

.15

c) Whitethroat

Year	Index	CBC
1961	68	-
62	89	85
63	94	80
64	40	84
65	15	84
66	100	100
67	68	86
68	43	110
69	20	30
70	24	31
71	34	33
72	18	26

Correlation    .60    (not significant at 5.5% level)

and rank correlation between indices gave the same results as in section 7.4.1.

## 7.5 Concluding remarks

In this chapter we have considered models that are driven by a concomitant sequence of discriminant scores. Each score is a function of a number of weather variables. Although we have assumed that this sequence of scores is fixed, in reality it is a realization of a stochastic process and therefore subject to error. The implication of this is that we are unfortunately compounding error in the construction of models based on the sequence of scores  $\{W_n\}$ . It is therefore not surprising that the models fail to show a good fit to the data.

There is no doubt that there is more that could be done here. For example, one may consider model (7.3) with a non-linear function of the discriminant scores, or one may investigate alternative distributions to the Poisson distribution. It is beyond the scope of this thesis to exhaust the possibilities for research of this kind, but here is an avenue that future research workers may like to follow.

The potential for obtaining useful migration indices from modelling work of this nature is not good. In order to justify going to the lengths of carrying out discriminant analyses and complex model fitting work, ornithologists would need to be sure that the resulting indices were reliable. This, as far as the work of this chapter has shown, is not the case.

## CHAPTER 8

### RECOMMENDATIONS FOR FURTHER RESEARCH

#### 8.1 Introduction

British bird observatories data have been used in this thesis for studying many aspects of migration and population dynamics. The diversity of topics covered is one reason why many questions which have arisen remain unanswered. Further progress in this field requires, in the case of forming migration indices for example, data of the type used in this thesis but over a longer time period. Other areas, for example the study of waves of migrants, require data of a different type for further research, i.e. ringing data. Ringing data are becoming more readily available as the BTO increase the computer storage of such data and therefore it should be easier in the future to use ringing data in conjunction with other types of data.

This chapter sets out recommendations for further research in the many areas of migration research that are discussed in preceding chapters. The recommendations cover the collection of data at observatories, the storage of data by the BTO, and the statistical analysis of the data.

This chapter and the thesis concludes with a summary of the main findings from the preceding chapters.

## 8.2 The collection of data at observatories

The collection of daily counts of grounded nocturnal migrants at British observatories is at present carried out in a haphazard fashion. The effect of this is that coverage of the observatory recording area may differ from day to day but, on the other hand, one usually finds that at least some observers are on the recording area throughout most of the morning.

An improvement that could be made in the collection of daily counts is to organise the available manpower each day and to conduct a census which is designed to provide optimum coverage over as much of each morning as is possible. The extent of the census will still vary from day to day but it should be possible to quantify this variability and to use this information in analyses of the daily counts. An organised daily census is carried at each day at Long Point Observatory, Canada, but no account of the level of manpower appears to have been made in analyses of the data (Hussell and Risley, 1975).

An organised daily census will not overcome the problem of birds being recorded by more than one observer and the usefulness of the census will still rely to some extent on the experience of the full-time warden who examines records each day. It is very important, therefore, that continuity of wardening from year to year should be sought as much as possible. When a new warden takes over at an observatory it would be advantageous if a transition period occurred where both out-going and in-coming wardens were working together. This would enable local knowledge of wardens to be passed on to successors.

An activity at observatories which could be improved upon is the collection of weather data in such a form that the data could be easily passed on to research workers. Most observatories include weather information in their daily logs, but much of this data is in the form of subjective description. Computer code forms would be of more use to statisticians who wish to work in this area, and the completion of such forms each day, along with forms for bird counts, would not greatly add to the wardens workload. In Chapter 3 a number of basic weather variables are discussed and the number of weather observations that would be required each day would be as few as seven. Coded weather summaries could then be forwarded to the BTO along with the counts of grounded migrants. One difficulty in examining the link between weather and observatories data in this thesis was the limited availability of coded weather data from different sites over a long period of time. The Meteorological Office provided the weather data used in this thesis, but future researchers would benefit from easily available data from observatories, or the BTO.

### 8.3 The storage of data from observatories

This thesis represents the first wide-ranging examination of British observatories data and therefore much time was spent obtaining data and entering data into a computer. If the suggestions of the previous section are followed up, then much of the preparatory work could be considerably reduced for future individuals who wish to work with bird and weather data from observatories. Ideally, the BTO would be able to assign staff to the job of entering monthly data sheets from observatories

into a computer so that future workers have immediate access to any subsets of the data. Upon request, subsets of data could be sent to individuals either as hardcopy or perhaps floppy diskettes or cartridge tapes. It seems likely that the easier it is to obtain observatories data, the more interest will be shown by both ornithological researchers and statisticians.

The above recommendation would involve the BTO in further pressures on their staff workloads, particularly if historical data were also to be entered on a computer, but this thesis has shown that many interesting features can be studied by using observatories data. In the light of this thesis it is likely that interest in this area of ornithological research will develop further, but only if data are readily available.

Since the commencement of this thesis, the BTO have entered onto their computer the data from the CBC. Future researchers will therefore be able to investigate the relevance of comparing data from observatories with CBC results with greater ease than has hitherto been possible. With the computerisation of other BTO data, for example, the Nest Record Scheme, it is logical that observatories data be treated in a similar manner.

#### 8.4 Recommendations for further data analysis

In chapter 3 it was shown that linear relationships between CBC indices and variables representing the timing of migrants passing through observatories could be formed which give good descriptions of the data. Data on four different species were examined but no common relationship could be established.

The data, however, spanned only a period of eleven years and if further data for these, and other species could be obtained over a longer time period, then it is possible that a common, underlying relationship will be found. If this is the case, then it would be possible to use this relationship with migration data for species which are not covered by the CBC, thereby providing a means of monitoring their population fluctuations.

In view of problems with using data from east coast observatories, i.e. the presence of non-British breeding birds in the data, it is recommended that attention be continued to be focused on Dungeness and Portland. Where possible, data from west coast observatories should also be obtained and used if the species under study are regular migrants at these sites. One is restricted, however, as to how many observatories can be included in an analysis of this type since the number of variables that one wishes to include in the multiple regression analysis of CBC index on migration season dates increases by four for each observatory included. CBC indices are only available from 1962, giving twenty-four years of data to the present date. One needs to balance the advantage of having migration data from a wide area with the disadvantage of having few data from which to derive a migration index. This is an important point for future researchers to investigate when more data are available.

In chapter 6 evidence was found to suggest that migration indices derived from spring and autumn seasons from Dungeness and Portland could be combined to form a migration index which monitors population levels. Using these migration indices which involve

spring, autumn and previous autumn statistics, each observatory contributes three variables to the regression of the CBC index. With the availability of data for the full twenty-four year period of the CBC, it would be possible to assess which and how many observatories need to be included in the formation of a combined migration index and to compare the reliability of the index with that based on the timing of migration. If only two or three observatories were found to be important for a particular species, then it may be possible to improve upon each individual approach by combining the index based on timing with that based on volume of migration.

The discussion of migration indices for population monitoring in this thesis has been concerned with the populations of birds of all ages. Observatories data may be useful for the study of populations of immature birds by using information on the waves of autumn migrants which some species show. In chapter 3 some methods for studying waves were given, but one would need many years of ringing data from observatories to be able to establish whether a particular wave consists of newly fledged birds or not, each year. Similarly, many years of ringing data would need to be examined in order to establish the extent to which observatories on the east coast of Britain record non-British birds which may be giving rise to one or more waves of migrants.

The statistical methods used in this thesis are those that offer, or attempt to offer, simple ornithological explanations to the data. There is clearly scope for further ideas to be tested here, for example, the models of chapters 6 and 7 are based on

simple distributional assumptions which provide a suitable starting point for the research. One could try alternative distributional assumptions in model (6.1), for example instead of a Poisson input to the model a distribution such as the negative-binomial distribution may be able to accommodate the occasional large falls of counts which occur.

In this thesis large counts have mostly been overlooked because their occurrence is rare, but when more data are available it is recommended that attention be paid to these data in order to establish whether or not allowance needs to be made in the construction of migration indices.

The study of relationships between weather data and observatories data is another area which deserves further attention when more data become available. The work of chapter 5, where discriminant analysis was used, failed to find a common relationship among four species but this situation may change if one is able to use data over a longer time period. Ideally one would like weather data from other localities as well as observatories. For example, it would be worth attempting to obtain information on weather from France in the spring and from interior Britain in the autumn. The precise nature and origins of such data would be difficult to determine but a study of synoptic charts may lead to a summary variable which broadly measures the favourability of weather for migration which could then be included in, for example, discriminant analyses. One objective of such research would be to improve upon the discrimination of chapter 5, and to avoid having to discard much of the data in the initial analyses.

If improvements of this sort were achieved, then a review of the modelling of chapter 7 may lead to more success with population monitoring.

## 8.5 Conclusions

Few attempts have been made in the past to analyse observatories data. Those that have been made have mostly been concerned with the timing of migration, length of season and migratory routes. These analyses have used ringing data rather than daily counts of grounded migrants - both data sets have their advantages and disadvantages as regards statistical analysis, and these are discussed in chapters 2 and 3.

A useful and informative way of studying daily counts of grounded migrants is to form bar-charts. These show that patterns of migration vary from year to year, as do the lengths of migration seasons. Some seasons begin and end abruptly while others build up and decline gradually. Other features revealed by bar-charts are occasional large falls of migrants and autumn waves of migration. The Sedge Warbler is a good example of a species which migrates in waves in most autumns. Patterns of migration are also seen to vary from observatory to observatory, even for the same species.

It is possible to identify specific deviations in migration from the norm by superimposing on bar-charts a curve which represents an average migration pattern. This simple graphical technique enables one to compare individual migrations both as regards volume and timing.

The calculation of summary statistics on the timing of migration enables one to examine objectively any long-term trend in the migration of particular species. The Blackcap is an example where there is some evidence to suggest that autumn migration is becoming later each year.

There is evidence to suggest that it is possible to monitor changes in population level of migrant species by forming indices based on timing statistics from observatories. As mentioned in section 8.4, more work needs to be done in this area before firm conclusions can be drawn.

It has long been thought that weather data need to be included in any study of observatories daily counts. No evidence to confirm this has been found in this thesis, but on the other hand some simple statistical procedures have been shown to provide links between the two data sets and the results of migration indices formed from such links have proved sufficiently encouraging to motivate further investigation on this subject. Linear discriminant analysis appears to be a useful technique to use in the study of weather and observatories counts. For British data at least, multiple regression analysis is not appropriate for linking weather and migration since the assumptions that one needs to make are not met by the data. It is not possible, however, to assess the reliability of multiple regression analysis in other author's work due to the lack of information given.

One feature arising from the discriminant analyses is that no common subset of weather variables could be found to give good

discrimination for all species considered. Some widely held opinions are supported by the discriminant analysis results however. For example low counts are associated with high, following winds and low amounts of cloud cover.

The modelling of sequences of daily counts with Poisson-binomial models suggests that the data do not show evidence of more hurried migration in spring than in autumn. It appears to be possible to fit simple statistical models to the data, although in years where few birds are recorded, the models do not fit satisfactorily.

No significant improvements in the fit of models or the success of migration indices derived from the models was found when weather information was introduced. Due to the limited time period used for this modelling work, firm conclusions cannot be drawn, and further work with more data is needed.

At present there appears to be no general rule for forming migration indices for different species, but as more data become available this situation may change. It is this possibility that should motivate further research on this subject since without a general rule for all species it is not possible to monitor population levels of species not covered by the CBC. If this could be achieved then this would be a major step forward for ornithologists and those concerned with protecting habitats for British birds.

The construction of migration indices is not seen to be significantly improved by the inclusion of weather data, although this approach should not be dismissed as only a preliminary examination has been possible so far due to the lack of readily available data. If a more extensive study along the lines of chapters 5 and 7 can be carried out then the relationship between weather and observatories may become clearer and the migration indices of this thesis may possibly be improved upon.

## REFERENCES

- ABLE, K.P. (1973)  
The role of weather variables and flight direction in determining the magnitude of nocturnal bird migration.  
*Ecology*, 54, p1031-1041.
- AKAIKE, H. (1972)  
Use of an information theoretic quantity for statistical model identification.  
Proc. 5th Hawaii Int. Conf. on System Sciences, p249-250.  
North Hollywood, California, Western Periodicals.
- AKAIKE, H. (1973)  
Information theory and an extension of the maximum likelihood principle.  
In 2nd Int. Symposium on Information Theory, p267-281.  
Budapest, Akademiai Kiado.
- ALERSTAM, T. (1978)  
Analysis and theory of visible bird migration.  
*OIKOS*, 30, p273-349.
- ALERSTAM, T., LINDGREN, A., NILSSON, S.G. and ULFSTRAND, S. (1973)  
Nocturnal passerine migration and cold front passages in autumn - a combined radar and field study.  
*Ornis Scandinavia*, 4, p103-111.
- BAILEY, R.S. (1967)  
An index of bird population changes on Farmland.  
*Bird Study*, 14, p195-209.
- BOX, G.E.P. and COX, D.R. (1964)  
An analysis of transformations (with discussion).  
*JRSS(B)*, 26, p211-52.
- BOX, G.E.P. and JENKINS, G.M. (1968)  
Some recent advances in forecasting and control.  
*Applied Statistics*, 17, p91-109.
- BOX, G.E.P. and JENKINS, G.M. (1970)  
Time series and analysis : forecasting and control (revised ed).
- BOX, G.E.P. and JENKINS, G.M. (1973)  
Some comments on a paper by Chatfield and Prothero and on a review by Kendall.  
*JRSS(A)*, 135, p337-352.
- BUSSE, P. (1973)  
Dynamics of numbers of some migrants caught at Polish Baltic Coast 1961-70.  
*Notatki Ornithologiczne*, 14, p1-59.
- CHATFIELD, C. and PROTHERO, D.L. (1973a)  
A reply to some comments by Box and Jenkins,  
*JRSS(A)*, 136, p345-352.
- CHATFIELD, C. and PROTHERO, D.L. (1973b)  
Box-Jenkins seasonal forecasting : problems in a case-study.  
*JRSS(A)*, 136, p295-536, (+ discussion).
- COX, D.R. and SMALL, N.J.H. (1978)  
Testing multivariate normality.  
*Biometrika*, 65, p263-272.

- COWLEY, E. (1979)  
Sand Martin population trends in Britain : 1965-78.  
Bird Study, 26, p113-117.
- DAVIS, P. (1967)  
Migration seasons of the Sylvia Warblers at British Bird Observatories.  
Bird Study, 14, p65-95.
- DIXON, W.J. (1981)  
BMDP statistical software.  
University of California Press.
- DURMAN, R.F. (1976)  
Bird Observatories in Britain and Ireland.  
edited by Durman, R.  
T. and A.D. Poyser Ltd., Hertfordshire, U.K.
- DURMAN, R.F. (1976)  
Ring Ousel migration.  
Bird Study, 23, p197-205.
- EASTWOOD, E. (1967)  
Radar Ornithology.  
Pub. Methuen.
- EVANS, P.R. (1966a)  
An approach to the analysis of visible bird migration and a comparison  
with radar observations.  
Ardea, 54, p14-54.
- EVANS, P.R. (1966b)  
Migration and orientation of passerine night migrants in NE England.  
J. Zool, London, 150, p314-369.
- FULLER, R.J., MARCHANT, J.H. and MORGAN, R.A. (1985)  
How representative of agricultural practice in Britain are CBC  
farmland plots?  
Bird Study, 32, p56-70.
- GLAHN, H.R. (1965)  
Objective methods of weather forecasting by statistical methods.  
The Statistician, 15, p111-142.
- GNANADESIKAN, R. (1977)  
Methods for statistical data analysis of multivariate observations.  
Wiley, New York.
- GRIFFIN, D.R. (1974)  
Bird migration.  
Dover publications, New York.
- HACKMAN, C.D. and HENNY, C.J. (1971)  
Hawk migration over White Marsh, Maryland.  
Chesapeake Science, 12, p137-141.
- HJORT, C. and LINDHOLM, C. (1978)  
Annual bird-ringing totals and population fluctuations.  
OIKOS, 30, p387-392.
- HOPE-JONES, P. (1975)  
The migration of Redstarts through and from Britain.  
Ringing and Migration, 1, p12-17.

- HOPE-JONES, P., MEAD, C.J. and DURMAN, R.F. (1977)  
The migration of the Pied Flycatcher from and through Britain.  
*Bird Study*, 24, p2-14.
- HUSSELL, D.J.T. (1981)  
The use of migration counts for monitoring bird population levels.  
*Studies in avian biology*, 6, p92-102.  
Cooper Ornithological Society.
- HUSSELL D.J.T. and RISLEY, C.J. (1978)  
Monitoring bird populations by migration sampling.  
Report to Canadian Wildlife Service.
- INSLEY, H. and BOSWELL, R.C. (1978)  
The timing of arrivals of Reed/Sedge Warblers at the South Coast.  
*Ringing and Migration*, 2, p1-9.
- JOLICOEUR, P. and MOSIMANN, J.E. (1960)  
Size and shape variation in the painted turtle. A p.c. analysis.  
*Growth*, 24, p339-354.
- JOLLIFFE, I.T. (1972a)  
Discarding variables in a principal component analysis.  
I. Artificial data.  
*Applied Statistics*, 21, p160-173.
- JOLLIFFE, I.T. (1972b)  
II. Real data.  
*Applied Statistics*, 22, p21-31.
- KEMPSON, R.E. (1983)  
Variable selection in linear discriminant analysis.  
PhD thesis, University of Kent.
- KRZANOWSKI, W.J. (1979)  
Some exact percentage points of a statistic useful in ANOVA and  
principal component analysis.  
*Technometrics*, 21, p261-4.
- LACK, D. (1959)  
Migration across the North Sea studied by radar.  
Part 1. Survey through the year.  
*Ibis*, 101, p209-235.
- LACK, D. (1960a)  
The influence of weather on passerine migration. A review.  
*Auk*, 77, p171-209.
- LACK, D. (1960b)  
Migration across the North Sea studied by radar.  
Part 2. Spring departure.  
*Ibis*, 102, p26-57.
- LACK, D. (1960c)  
Autumn drift migration on the English East coast.  
*British Birds*, 53, p325-353.

- LACK, D. (1963a)  
Migration across the North sea studied by radar.  
Part 3. Movements in June and July.  
Ibis, 105, p74-85
- LACK, D. (1963b)  
Migration across the North sea studied by radar.  
Part 4. Autumn.  
Ibis, 105, p1-54.
- LACK, D. (1963c)  
Migration across the North sea studied by radar.  
Part 5. Movements in August, winter and spring, and conclusion.  
Ibis, 105, p461-492.
- LACK, P. and LACK, D. (1966)  
Passerine night migrants on Skokholm.  
British Birds, 4, p129-141.
- LACHENBRUCH, P. (1975)  
Discriminant analysis.  
Hafner Press, New York.
- LACHENBRUCH, P. and GOLDSTEIN, M. (1979)  
Discriminant analysis.  
Biometrics, 35, p69-85.
- LACHENBRUCH, P. and MICKEY, M.R. (1968)  
Estimation of error rates in discriminant analysis.  
Technometrics, 10, p1-11.
- LANGSLOW, D.R. (1977)  
Movements of Black Redstarts between Britain and Europe as related  
to occurrences at observatories.  
Bird Study, 24, p169-178.
- LANGSLOW, D.R. (1978)  
Recent increases of Blackcaps at bird observatories.  
British Birds, 71, p345-354.
- LANGSLOW, D.R. (1979)  
Movements of Blackcaps ringed in Britain and Ireland.  
Bird Study, 26, p239-252.
- McKAY, R.J. and CAMPBELL, N.A. (1982)  
Variable selection techniques in discriminant analysis, I and II.  
British Journal, Math. Stat. Psych. 35, p1-41.
- MARDIA, K.V. (1974)  
Applications of some measures of multivariate skewness and kurtosis  
in testing normality and robustness studies.  
Sankhya, 36, p115-128.

- MATTHEWS, G.V.T. (1968)  
Bird Navigation (2nd edition).  
Cambridge University Press.
- MORRISON, D.F. (1976)  
Multivariate statistical methods.  
McGraw-Hill Inc.
- MOUNTFORD, M.D. (1982)  
Estimation of population fluctuations with application to the Common  
Bird Census.  
Applied Statistics, 31, p135-143.
- MUELLER, H.C. and BERGER, D.D. (1967)  
Some observations and comments on the periodic invasions of Goshawks.  
Auk. 84, p183-191.
- MUELLER, H.C. and ALLEZ, G. (1977)  
The periodic invasions of Goshawks.  
Auk. 94, p652-663.
- NELDER, J. and MEAD, R. (1965)  
A simplex method for function minimisation.  
Computer J. 7, p308-313.
- NISBET, I.C.T. and DRURY, W.H. (1968)  
Short term effects of weather on bird migration: a field study  
using multivariate statistics.  
Anim. Behaviour, 16, p496-530.
- NORTH, P.M. (1977)  
A novel clustering method for estimating numbers of bird territories.  
Applied Statistics, 26, p149-155.
- NORTH, P.M. (1979)  
A novel clustering method for estimating numbers of bird territories,  
an Addendum.  
Applied Statistics, 28, p300-301.
- NORTH, P.M. (1983)  
Modelling annual numbers of bird territories on farmland plots.  
Biometrical Journal, 25, p431-441.
- OZAKI, T. (1977)  
On the order determination of ARIMA models.  
Applied Statistics, 26, p290-301.
- PARSLOW, J.L.F. (1962)  
Immigration of night-migrants into Southern England in spring 1962.  
Bird Migration, 2, p160-175.

- PARSLOW, J.L.F. (1969)  
The migration of passerine night migrants across the English Channel studied by radar.  
*Ibis*, 111, p48-79.
- RADFORD, M.C. (1968)  
Autumn migration of the Red-Breasted Flycatcher.  
*Bird Study*, 15, p154-160.
- RICHARDSON, W.J. (1974)  
Multivariate approaches to forecasting day-to-day variations in the amount of bird migration.  
Biological aspects of the bird/aircraft collision problem.  
Clemson University, South Carolina.
- RIDDIFORD, N. (1985)  
Grounded migrants versus radar: a case-study.  
*Bird Study*, 32, p116-121.
- RIDDIFORD, N. and AUGUR, R. (1980)  
Willow Warbler migration at Dungeness.  
Unpublished paper.
- ROYSTEN, J.P. (1983)  
Some techniques for assessing multivariate normality based on the Shapiro-Wilk W.  
*Applied Statistics*, 32, p121-133.
- RYAN, T.A., JOINER, B.L. and RYAN, B.F. (1981)  
Minitab Reference Manual.
- SHARROCK, J.T.R. (1966)  
Weekend-bias, length of stay of rarities and proportion of rarities undetected.  
*British Birds*, 59, p556-558.
- SHARROCK, J.T.R. (1968)  
Migration seasons of the Sylvia Warblers at Cape Clear Bird Observatory.  
*Bird Study*, 15, p55-103.
- SHARROCK, J.T. (1969)  
Grey Wagtail passage and population fluctuations in 1956-67.  
*Bird Study*, 16, p17-34.
- SHARROCK, J.T.R. (1970)  
Weekend bias at observatories (letter).  
*Bird Study*, 17, p57-58.
- SNOW (1953)  
Visible migration in the British Isles : A review.  
*Ibis*, 95, p242-270.
- SVENSSON, S.E. (1978)  
Efficiency of two methods for monitoring bird population levels :  
Breeding bird censuses contra counts of migrating birds.  
*OIKOS*, 30, p373-386.

TAVARÉ, S. and ALTHAM, P.M. (1983)  
Serial dependence of observations leading to contingency tables and  
corrections to chi-squared statistics.  
Biometrika, 70, p139-44.

TAYLOR, S.M. (1965)  
The CBC - some statistical aspects.  
Bird Study, 12, p268-286.

UPTON, G. and LAMPITT, G.A. (1981)  
A model for interyear change in the size of bird populations.  
Biometrics, 37, p113-128.

WILCOCK, J. (1965)  
Detection by radar of autumn migration in Eastern Scotland.  
Ibis, 107, p316-325.

WILLIAMSON, K. (1963)  
Movements as an indicator of population changes.  
Bird Migration, 2, p207-223.

WINSTANLEY, P., SPENCER, R. and WILLIAMSON, K. (1974)  
Where have all the Whitethroats gone?  
Bird Study, 21, p1-15.



Scientific names of birds mentioned in the text.  
( Species are arranged in Wetmore order. )

Goshawk	Accipiter gentilis
Bee-eater	Merops apiaster
Hoopoe	Upupa epops
Sand Martin	Riparia riparia
Swallow	Hirundo rustica
Grey Wagtail	Motacilla cinerea
Sedge Warbler	Acrocephalus schoenobaenus
Reed Warbler	Acrocephalus scirpaceus
Blackcap	Sylvia atricapilla
Whitethroat	Sylvia communis
Willow Warbler	Phylloscopus trochilis
Pied Flycatcher	Ficedula hypoleuca
Red-breasted Flycatcher	Ficedula parva
Spotted Flycatcher	Muscipapa striata
Black Redstart	Phoenicurus ochruros
Redstart	Phoenicurus phoenicurus
Ring Ouzel	Turdus torquatus