



Kent Academic Repository

Cole, Diana J. (2019) *Parameter Redundancy and Identifiability in Hidden Markov Models*. Metron, 77 (2). pp. 105-118. ISSN 0026-1424.

Downloaded from

<https://kar.kent.ac.uk/75204/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1007/s40300-019-00156-3>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

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. 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>).



Parameter redundancy and identifiability in hidden Markov models

Diana J. Cole¹ 

Received: 30 November 2018 / Accepted: 8 July 2019 / Published online: 18 July 2019
© The Author(s) 2019

Abstract

Hidden Markov models are a flexible class of models that can be used to describe time series data which depends on an unobservable Markov process. As with any complex model, it is not always obvious whether all the parameters are identifiable, or if the model is parameter redundant; that is, the model can be reparameterised in terms of a smaller number of parameters. This paper considers different methods for detecting parameter redundancy and identifiability in hidden Markov models. We examine both numerical methods and methods that involve symbolic algebra. These symbolic methods require a unique representation of a model, known as an exhaustive summary. We provide an exhaustive summary for hidden Markov models and show how it can be used to investigate identifiability.

Keywords Hidden Markov models · Identifiability · Parameter redundancy

1 Introduction

Hidden Markov models (HMMs) are a flexible class of models that are used when there are unobservable states. An HMM is a type of dependent mixture model, which consists of a time series of observations, where each observation is dependent on an unobserved Markov process; see for example [40].

As with any model, it does not follow that it is possible to estimate every parameter within that model. For example, two parameters could be confounded as a product, so that there is not a unique estimate of each parameter individually; only the product has a unique estimate. This problem is known as parameter redundancy. Alternatively, we say that the parameters are non-identifiable; see for example [6,14].

In Sect. 1.1 we describe the form of HMMs. Then parameter redundancy and identifiability is discussed in Sect. 1.2. Methods for investigating parameter redundancy and identifiability

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40300-019-00156-3>) contains supplementary material, which is available to authorized users.

✉ Diana J. Cole
d.j.cole@kent.ac.uk

¹ School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7FS, UK

in HMMs are explained in Sects. 2 and 3. Examples are used to illustrate the methods; computer code for each example is available in the supplementary material.

1.1 Hidden Markov models

The observations in an HMM are represented by $\{X_t : t = 1, 2, \dots, T\}$. Each X_t is dependent on an underlying hidden Markov process, C_t , with each X_t only dependent on the current state of C_t . The hidden process C_t can be in one of m states at time t , and is only dependent on C_{t-1} . Let

$$p_i(x) = \Pr(X_t = x | C_t = i)$$

be the probability mass function or probability density function of X_t given the Markov chain is in state i at time t , and let $\mathbf{P}(x)$ be a diagonal matrix with entries $p_i(x)$. The transition between states is represented by the matrix $\mathbf{\Gamma}$, whose (ij) th element is

$$\gamma_{ij} = \Pr(C_{t+1} = j | C_t = i),$$

with $\sum_{j=1}^m \gamma_{ij} = 1$. The initial distribution of the states is $\boldsymbol{\delta} = [\delta_1, \delta_2, \dots, \delta_m]$, with $\sum_{j=1}^m \delta_j = 1$. The likelihood for a hidden Markov model is then

$$L_T = \boldsymbol{\delta} \mathbf{P}(x_1) \mathbf{\Gamma} \mathbf{P}(x_2) \mathbf{\Gamma} \mathbf{P}(x_3) \dots \mathbf{\Gamma} \mathbf{P}(x_T) \mathbf{1}$$

where $\mathbf{1}$ is a column vector of 1s. If $\boldsymbol{\delta}$ is assumed to be the stationary distribution of the Markov chain then $\boldsymbol{\delta} = \boldsymbol{\delta} \mathbf{\Gamma}$, and this can replace $\boldsymbol{\delta}$ in the above [40].

Below we give three example of HMMs that are used to illustrate different methods for investigating identifiability.

Example 1: two-state poisson HMM

Zuchini et al. [40] present a two-state Poisson HMM for earthquake data, which can be extended to three, four or more states. The data consists of the number of earthquakes magnitude 7 or greater from 1900 to 2006 with $X_1 = 13, X_2 = 14, X_3 = 8, \dots, X_{107} = 11$. The two-state transition matrix is

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & 1 - \gamma_{11} \\ \gamma_{21} & 1 - \gamma_{21} \end{bmatrix},$$

the initial state vector is

$$\boldsymbol{\delta} = [\delta_1, 1 - \delta_1]$$

and

$$p_i(x) = \frac{\exp(-\lambda_i) \lambda_i^x}{x!}.$$

The model has 4 parameters $\boldsymbol{\theta} = [\gamma_{11}, \gamma_{21}, \lambda_1, \lambda_2]$, if it is assumed that the initial distribution is stationary with $\boldsymbol{\delta} = \boldsymbol{\delta} \mathbf{\Gamma}$. Alternatively δ_1 can be treated as an additional parameter.

This simple example is a commonly used HMM, where identifiability results are well known, see for example [28]. It is included here to provide a simpler example to illustrate the different methods discussed.

Example 2: mark-recapture-recovery model

In mark-recapture-recovery experiments, animals are captured and marked with unique tags. The animal is then recaptured at later time points, or the tag is recovered from a dead animal. Data can be used to estimate the mortality in wild animal populations, see for example [32].

The data recorded consists of whether the animal is captured alive or recovered dead at different time points. For example, suppose an animal population is monitored over 5 years. One individual animal is first caught in year 2, then not recaptured in year 3, recaptured in year 4 and then recovered dead in year 5. This animal’s capture history can be represented as $h = 01012$, where each number represents one year, with 0 representing an animal is not captured, 1 representing an animal is captured alive and 2 representing an animal is recovered dead. Each capture history has an associated probability of occurring. For capture history $h = 01012$, $Pr(h) = \phi_2(1 - p_3)\phi_3p_4(1 - \phi_4)\lambda_5$, where ϕ_t is the probability an animal survives from time period t to $t + 1$, p_t is the probability an animal is captured at time t , and λ_t denotes the probability that the mark of a dead animal is recovered at time t .

Consider an individual animal that was first recaptured at time a and last recaptured alive or recovered dead at time b . Let x_k represent the capture history entry at time k . The probability for a particular capture history is then

$$Pr(h) = \begin{cases} \prod_{k=a+1}^b \phi_{k-1} \{x_k p_k + (1 - x_k)(1 - p_k)\} \chi_b & \text{if } x_b = 1 \\ \prod_{k=a+1}^{b-1} \phi_{k-1} \{x_k p_k + (1 - x_k)(1 - p_k)\} (1 - \phi_{b-1})\lambda_b & \text{if } x_b = 2, \end{cases}$$

where $\chi_k = (1 - \phi_k)(1 - \lambda_k) + \phi_k(1 - p_{k+1})\chi_{k+1}$. For N animals, the likelihood is $L = \prod_{m=1}^N Pr(h_m)$ [24]. Alternative forms for the likelihood are given in [7,9,26,31].

This model can be considered as two processes. One process represents whether the animal is alive or dead, the other process represents the observation process (the animal is recaptured or recovered) [23]. This allows the model to be represented as an HMM, see Chapter 24 of [40]. The true state of whether the animal is dead or alive is treated as a hidden state. The under-lying state process has three states: alive, recently dead and long dead. The transition matrix is

$$\Gamma_t = \begin{bmatrix} \phi_t & 1 - \phi_t & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

at time t . The observation matrix is

$$\mathbf{P}(x_t) = \begin{cases} \text{diag}(1 - p_t, 1 - \lambda_t, 1) & \text{if } x_t = 0 \\ \text{diag}(p_t, 0, 0) & \text{if } x_t = 1 \\ \text{diag}(0, \lambda_t, 0) & \text{if } x_t = 2 \end{cases}.$$

In this case the transition matrix has been extended to be time dependent, so the likelihood for individual i extends to

$$L_i = \delta \Gamma_a \mathbf{P}(x_{a+1}) \Gamma_{a+1} \mathbf{P}(x_{a+2}) \dots \Gamma_{T-1} \mathbf{P}(x_T),$$

where the individual was initially marked at time a , has history x_t at time t . The likelihood for N individual histories is $L = \prod_{m=1}^N L_i$ [40].

This example is used to illustrate that complex models can be more simply represented as an HMM. We later investigate whether it is simpler to use the HMM form to investigate identifiability.

Example 3: longitudinal HMM

In medical research a patient’s health or response to treatment can be in one of m different states. For example in a schizophrenia study, patients were classified in one of four categories: normal, mild, moderate, and severe. The state of the patient is then recorded over several time periods. It is possible that the state can be misclassified, which can be taken into account using an HMM. See [39].

For this example, the initial state vector is

$$\boldsymbol{\delta} = [\delta_1, \delta_2, \dots, \delta_m],$$

where δ_i is the probability a patient is in state i at the start of the study, with $\delta_m = 1 - \sum_{i=1}^{m-1} \delta_i$. The transition matrix is

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1m} \\ \vdots & & & \\ \gamma_{m1} & \gamma_{m2} & \dots & \gamma_{mm} \end{bmatrix},$$

where γ_{ij} represents the probability of moving from state i to state j . Again $\sum_{j=1}^m \gamma_{ij} = 1$, so we let $\gamma_{ii} = 1 - \sum_{j=1, j \neq i}^m \gamma_{ij}$. The observation matrix is

$$\mathbf{P}(x = j) = \text{diag}(b_{1j}, b_{2j}, \dots, b_{mj}),$$

where b_{ij} represents the probability of misclassification. As $\sum_{j=1}^m b_{ij} = 1$ we let $b_{ii} = 1 - \sum_{j=1, j \neq i}^m b_{ij}$.

For K patients observed at T points in time, the likelihood is

$$\prod_{k=1}^{k=K} \boldsymbol{\delta} \mathbf{P}(x_{1k}) \boldsymbol{\Gamma} \mathbf{P}(x_{2k}) \boldsymbol{\Gamma} \dots \mathbf{P}(x_{Tk}) \boldsymbol{\Gamma} \mathbf{1},$$

where $x_{tk} = 1, 2, \dots, m$ is the k th patient’s state at time t .

Identifiability of this model follows from [1]. However, real applications have a different form for the entries of the transition matrix, see for example [2,4]. Here we are interested in how identifiability can be investigated in such cases.

1.2 Parameter redundancy and identifiability

Consider a model, $M(\boldsymbol{\theta})$, which depends on q_θ parameters, $\boldsymbol{\theta}$. The model, $M(\boldsymbol{\theta})$, is defined as parameter redundant if it can be reparameterised in terms of a smaller number of parameters, $\boldsymbol{\beta} = g(\boldsymbol{\theta})$, of length $q_\beta < q_\theta$, where g is a function, and where q_β is the number of parameters in the vector $\boldsymbol{\beta}$ [6]. A model is globally identifiable if $M(\boldsymbol{\theta}_1) = M(\boldsymbol{\theta}_2)$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. A model is locally identifiable if there is a region in the parameter space where this is true. A model is non-identifiable if $M(\boldsymbol{\theta}_1) = M(\boldsymbol{\theta}_2)$ for $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. See for example [14]. It follows that a parameter redundant model will be non-identifiable [6]. A non-identifiable model can consist of both identifiable and non-identifiable parameters; the same definition can be applied to identifiability regarding a single parameter.

If an HMM is identifiable it will often be locally identifiable rather than globally identifiable, due to label switching. That is, the state parameters and the distribution parameters can be permuted and still give the same value of the likelihood [28]. An example of label switching is given in Sect. 2.2. This is frequently overcome by using an order constraint on the parameters. For example, in Example 1 below, $\lambda_1 < \lambda_2$.

Generic rules for identifiability in HMMs have been established by Petrie [33] and Allman et al. [1]. For example, in a binary HMM with m states, generic identifiability is achieved from the marginal distribution of $2m - 1$ observations (see Theorem 6 in [1]). However, the theorem is not applicable for every HMM. In Example 1 there is only a single time series and the state space of the observed variable X_t has no upper limit. Example 2 uses a different parameterisation so the theorem is no longer valid. The theorem can be applied to Example 3, resulting in an identifiable model as long as there are at least $2m - 1$ observations. However, we are also interested in a different parameterisation, so again the theorem is no longer necessarily valid.

Rather than considering generic results for all standard HMMs, we will now investigate how identifiability can be checked for a specific example. Various methods exist for checking identifiability, which include the Hessian method, the log-likelihood profile method and a symbolic algebra method [22].

As the Hessian matrix in a non-identifiable model will be singular at the maximum likelihood estimate, the Hessian matrix can be used to check identifiability [38]. However, in HMMs the Hessian matrix, when found numerically, is unreliable for longer time series and more complex models [3]. In Sect. 2.1 we demonstrate how this problem effects using the Hessian matrix method to check identifiability.

The log-likelihood profile method involves plotting a log-likelihood profile for each parameter. If the profiles are flat for any parameter then the model is non-identifiable, see [35]. In Sect. 2.2 we demonstrate how the log-likelihood method can be used to check identifiability in an HMM.

A model can be represented by an exhaustive summary, which is a vector of parameter combinations that uniquely define the model. To investigate non-identifiability a derivative matrix is formed by differentiating each term of the vector with respect to each parameter and the rank of the derivative matrix is found. If the rank is less than the number of parameters the model is non-identifiable or parameter redundant, see [14,36]. In Sect. 3 we provide an exhaustive summary and show how this symbolic algebra method can be used with HMMs.

2 Numerical methods for investigating identifiability in HMMs

In this Section we explore two numerical methods for investigating identifiability: the Hessian method in Sect. 2.1 and the log-likelihood profile method in Sect. 2.2.

2.1 Hessian methods

If a model is non-identifiable then the Fisher Information matrix will be singular [36] and the Hessian matrix at the maximum likelihood estimate is also singular [38]. This means that standard errors do not exist for a non-identifiable model, and also gives a method to check identifiability. Typically the Hessian matrix is simpler to use than the Fisher information matrix [30], but rather than the exact Hessian matrix, a numerical approximation is used. A singular matrix will have at least one zero eigenvalue. However, the approximate

Hessian matrix may no longer be singular, so instead we consider whether at least one of the eigenvalues is close to zero [38]. The eigenvalues can be standardised by dividing by the largest eigenvalue. The difficulty with this method is choosing the threshold for when a standardised eigenvalue is considered close enough to zero. Vialefont et al. [38] suggest a threshold of $\tau = q \times 10^{-9}$, for a model with q parameters. Testing indicates this can be too strict. Simulation studies suggest a better threshold is $\tau = q \times 10^{-6}$. Other alternative criteria are discussed in Konstatinides and Yao [27] and Little et al. [29].

It is also possible that an identifiable model still causes issues with parameter estimation for specific data sets, which is called statistical estimability or sloppiness [5,12]. This can occur because the model is very similar to a model that is parameter redundant for a particular data set, which is known as near redundancy [10,14]. The Hessian matrix will also have small eigenvalues in such case. To distinguish sloppiness, Chis et al. [12] use a threshold of $\tau = 10^{-3}$.

Example 1: two-state poisson HMM

To demonstrate how the Hessian method works we consider the two-state HMM for the earthquake data set. In this example we assume that δ is stationary. We use both the full data set with 107 observations and a small data set consisting of only the first three observations, $X_1 = 13$, $X_2 = 14$, $X_3 = 8$. We note that the small data set is only used for a simple illustration; it is not sensible to fit an HMM with such a small sample size. The R code `Example1.R` in the supplementary material shows how to find the eigenvalues of the Hessian matrix. The eigenvalues are then standardised by dividing by the largest eigenvalue. For the full data set the standardised eigenvalues are 1, 0.33, 0.0040, 0.0027. For the smaller data set the standardised eigenvalues are 1, 7.70×10^{-9} , 4.60×10^{-11} , 0. The full data set is identifiable, whereas the smaller data set is non-identifiable.

There are two potential problems with this method. The first is the choice of threshold, and the second is the accuracy of the numerical Hessian approximation for HMMs (see, for example, [3]). We explore these issues considering three different simulations, labelled Sim 1, Sim 2 and Sim 3:

- Sim 1: The two-state Poisson HMM with $\lambda_1 = 10$, $\lambda_2 = 20$, $\gamma_{11} = 0.7$, $\gamma_{21} = 0.2$. This model should be identifiable for $T \geq 4$.
- Sim 2: The two-state Poisson HMM, simulated from a single state model, so that $\lambda_1 = 10$, $\lambda_2 = 10$, $\gamma_{11} = 0.7$, $\gamma_{21} = 0.3$. This model should be identifiable for $T \geq 4$, but is near redundant due to the data set used.
- Sim 3: The two-state Poisson HMM reparameterised to be obviously non-identifiable, with $\gamma_{11} = ab$. The simulation values are $\lambda_1 = 10$, $\lambda_2 = 10$, $a = 0.5$, $b = 0.9$, $\gamma_{21} = 0.2$.

Each simulation study consists of 100 simulations. Similar to `Example1.R`, the optimisation procedure used was the R function `nlm`, initialised using the true parameter values. We record how many of the simulations have their smallest standardised eigenvalue below each of three thresholds: $\tau = 10^{-3}$, $\tau = q \times 10^{-6}$ and $\tau = q \times 10^{-9}$.

The simulation study is summarised in Table 1. It is clear that the Hessian method is not always accurate at detecting identifiability. The threshold of $\tau = q \times 10^{-9}$ is too strict to pick up all cases of non-identifiability. A threshold of $\tau = q \times 10^{-6}$ is better, except for small values of T . A threshold of $\tau = 10^{-3}$ seems to pick up near-redundancy in most, but not all, cases.

Table 1 Simulation study, showing the percentage of simulations where the smallest standardised eigenvalue is below each of three thresholds: $\tau = 10^{-3}$, $\tau = q \times 10^{-6}$ and $\tau = q \times 10^{-9}$

	T	3 (%)	5 (%)	10 (%)	20 (%)	50 (%)	100 (%)	500 (%)
Sim 1	$\tau = 10^{-3}$	100	99	47	14	3	1	0
	$\tau = q \times 10^{-6}$	100	97	47	13	0	0	0
	$\tau = q \times 10^{-9}$	70	70	36	13	0	0	0
Sim 2	$\tau = 10^{-3}$	100	<i>94</i>	<i>97</i>	<i>89</i>	<i>93</i>	<i>93</i>	<i>100</i>
	$\tau = q \times 10^{-6}$	100	<i>94</i>	<i>95</i>	<i>88</i>	<i>87</i>	<i>71</i>	<i>67</i>
	$\tau = q \times 10^{-9}$	75	<i>80</i>	<i>66</i>	<i>56</i>	<i>59</i>	<i>45</i>	<i>30</i>
Sim 3	$\tau = 10^{-3}$	100	100	100	100	100	100	100
	$\tau = q \times 10^{-6}$	100	100	100	100	100	100	100
	$\tau = q \times 10^{-9}$	99	92	73	49	19	10	16

T is the length of the time series. Sim 1 is the two-state Poisson HMM, which is identifiable when $T \geq 4$. Sim 2 is the two-state Poisson HMM with near-redundant data. Sim 3 is a reparameterised two-state Poisson HMM that is non-identifiable. Non-identifiable models, indicated by bold, would ideally have 100% of simulation over the thresholds. In the simulation where the data is chosen to cause near redundancy, indicated by italic, would ideally have 100% of simulations over the threshold $\tau = 10^{-3}$. Identifiable models would ideally have 0% of simulations over the threshold

For Sim 1 we also constructed confidence intervals based on the Hessian matrix and found that only 5% of the simulations contained the true value of the parameter for $T = 500$. As found by Visser et al. [37], this is because the confidence intervals were too narrow. Even though the approximate Hessian matrix is a poor approximation for finding the standard error and confidence intervals, the Hessian method is still giving the correct result about identifiability in this case. In other cases it is not clear whether the wrong result is due to the poor Hessian approximation or the threshold.

Overall, because of the difficulty in choosing a threshold and because the numerical approximation of the Hessian matrix can be inaccurate, this method is not recommended for HMMs.

2.2 Log-likelihood profile method

As parameter redundant models can be reparameterised in terms of a smaller number of parameters, so that $\beta = g(\theta)$, it follows that there are multiple maximum likelihood estimates satisfying $\hat{\beta} = g(\hat{\theta})$ [19]. This results in a flat ridge in the log-likelihood surface [6]. Therefore a method of detecting non-identifiability is to examine log-likelihood profiles for each parameter. If the profile for a parameter is flat, the parameter is non-identifiable. If there is a single optimum, the parameter is globally identifiable and if there are two or more optima, the parameter is locally identifiable.

Example 1: two-state poisson HMM

The R code `Example1.R` also gives code to create log-likelihood profiles. These are shown for the parameter λ_1 in Fig. 1. We consider four different examples:

- (a) The standard two-state model for the earthquake data set. This example is locally identifiable.

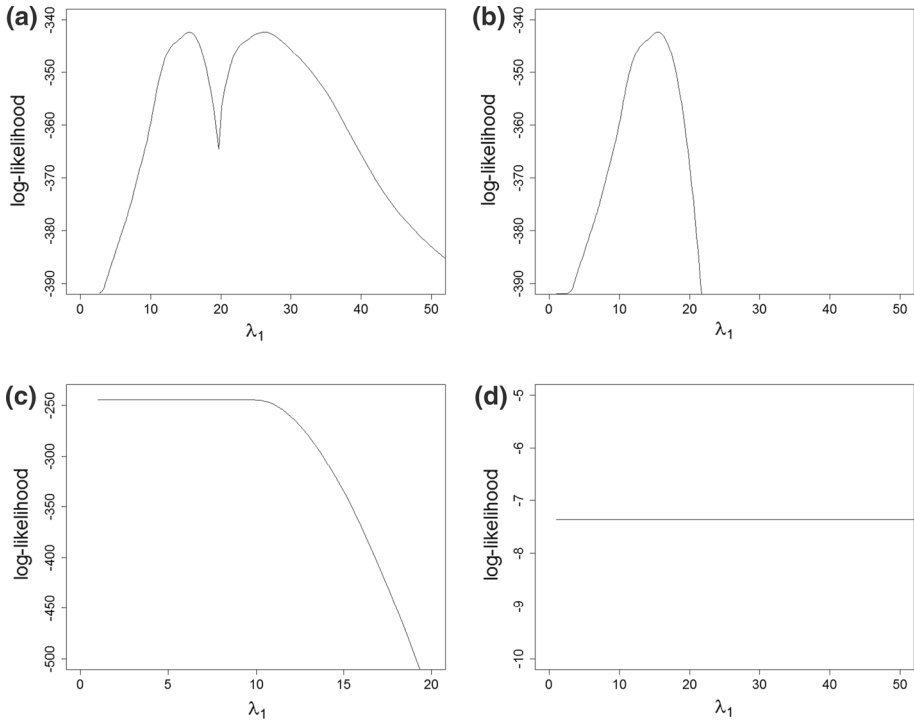


Fig. 1 Log-likelihood profiles for Example 1 for **a** two-state Poisson HMM (locally identifiable) **b** two-state Poisson HMM with constraint $\lambda_1 < \lambda_2$ (globally identifiable) **c** two-state Poisson HMM with data that causes near redundancy **d** two-state Poisson HMM with only $T = 3$ observations (non-identifiable)

- (b) The two-state model for the earthquake data set with the constraint $\lambda_1 < \lambda_2$. This example is globally identifiable.
- (c) The two-state model for a simulated data set following Sim 2, with $T = 100$. This example is near redundant.
- (d) The standard two-state model with only $T = 3$ observations. This example is non-identifiable.

When all the data are used there are two optima, therefore the parameter λ_1 is locally identifiable (Fig. 1a). The other parameters are all locally identifiable as well. In this case there are two sets of maximum likelihood estimates: $\hat{\lambda}_1 = 15.47$, $\hat{\lambda}_2 = 26.13$, $\hat{\gamma}_{11} = 0.93$, $\hat{\gamma}_{21} = 0.13$ (with $1 - \hat{\gamma}_{11} = 0.07$, $1 - \hat{\gamma}_{21} = 0.87$) and $\hat{\lambda}_1 = 26.13$, $\hat{\lambda}_2 = 15.47$, $\hat{\gamma}_{11} = 0.87$, $\hat{\gamma}_{21} = 0.07$ (with $1 - \hat{\gamma}_{11} = 0.13$, $1 - \hat{\gamma}_{21} = 0.93$). This is an example of label switching, see for example [34]. Local identifiability can be changed to global identifiability using an order constraint; here we use $\lambda_1 < \lambda_2$. This is shown in Fig. 1b, where there is now only one optimum.

For the simulated data set, the data causes near redundancy. This can be seen in Fig. 1c as the profile is flat for part of the log-likelihood. For the small data set with 3 observations the log-likelihood profile is flat, showing λ_1 is non-identifiable.

We note that it is difficult to distinguish between whether a log-likelihood profile is flat, or is near flat. It is therefore not always possible to distinguish between non-identifiability and near redundancy.

3 Symbolic algebra method for investigating identifiability in HMMs

To investigate identifiability it is often simpler to use an exhaustive summary rather than the model. An exhaustive summary, $\kappa(\theta)$, is a vector of parameter combinations that uniquely define the model. That is $\kappa(\theta_1) = \kappa(\theta_2) \Leftrightarrow M(\theta_1) = M(\theta_2)$ for all θ_1 and θ_2 [14].

The symbolic algebra method involves forming a derivative matrix by differentiating the exhaustive summary with respect to the parameters, so that

$$D = \frac{\partial \kappa}{\partial \theta}.$$

If the rank, r , of the derivative matrix is less than the number of parameters, q_θ , the model is parameter redundant and therefore non-identifiable. If the rank, r , of the derivative matrix is equal to the number of parameters then the model is at least locally identifiable, that is, the model could be either locally or globally identifiable [6,14,36]. We let $d = q_\theta - r$ represent the deficiency of the model; a parameter redundant model will have $d > 0$. If a model is parameter redundant it is possible to find whether any of the individual parameters are identifiable by solving $\alpha'D = 0$. There will be d solutions to $\alpha'D = 0$, which we label α_j , for $j = 1, \dots, d$, with individual entries α_{ij} . If $\alpha_{ij} = 0$ for all j then θ_i will be individually identifiable. By solving the system of linear first-order partial differential equations $\sum_{i=1}^p \alpha_{ij} \partial f / \partial \theta_i = 0, j = 1 \dots d$ for arbitrary differentiable function f , we can find a reparameterisation of the model which is not parameter redundant, known as estimable parameter combinations, or a locally identifiable reparameterisation [8,11,14,21].

To use the symbolic algebra method in Hidden Markov models we require a suitable exhaustive summary. In a similar way to the expansion exhaustive summary developed for state-space models in Cole and McCrea [18], we start with the likelihood for just X_1 , and then extend to the likelihood for X_1 and X_2 , etc. This gives an exhaustive summary of

$$\kappa = \begin{bmatrix} \delta P(x_1)\mathbf{1} \\ \delta P(x_1)\Gamma P(x_2)\mathbf{1} \\ \delta P(x_1)\Gamma P(x_2)\Gamma P(x_3)\mathbf{1} \\ \vdots \\ \delta P(x_1)\Gamma P(x_2)\Gamma P(x_3) \dots \Gamma P(x_T)\mathbf{1} \end{bmatrix}.$$

If δ is assumed to be stationary then the first term of the exhaustive summary can be replaced by $\delta \Gamma P(x_1)\mathbf{1}$. This will change the exhaustive summary's algebraic representation, but not the identifiability results.

In some models the exhaustive summary length will depend on the number of observations. For HMMs, the exhaustive summary length will depend on the length of the time series, T . We can use this method to obtain general results about a model for any T by using the Extension Theorem [6,14]. We start from the lowest value of T for which the general result is valid, $T = T_1$, and find the rank of the derivative matrix for the exhaustive summary with $T = T_1$. If the derivative matrix has $d = 0$, we then extend the model to $T = T_1 + 1$. The extra exhaustive summary terms are then differentiated with respect to the extra parameters to form a second derivative matrix. If that derivative matrix has $d = 0$ as well then the model will always be identifiable, by the Extension Theorem. If the model with $T = T_1$ has $d > 0$ we first reparameterise the model in terms of the estimable parameter combinations, to be able to apply the same result and show the model will always be non-identifiable.

The symbolic algebra method is executed in a symbolic algebra package, such as Maple. In more complex models the package will run out of memory trying to find the rank symbolically

[15,16,25]. Cole et al. [14] have developed an extension to the symbolic method that involves reparameterisation. The alternative is the hybrid symbolic-numerical method, which involves calculating the derivative matrix symbolically and then finding the rank at several random points in the parameter space [13]. Choquet and Cole [13] demonstrate that it is sufficient to use 5 random points.

We can distinguish between global and local identifiability by solving the equations $\kappa_i(\theta) = k_i$ for $i = 1, \dots, n$, where n is the number of terms in the exhaustive summary. If there is a single solution the model is globally identifiable. If there are a countable number of solutions, greater than 1, then the model is locally identifiable [14].

We demonstrate how this symbolic algebra method is executed in three examples below.

Example 1: two-state poisson HMM

For the two-state Poisson HMM, when δ is stationary, the exhaustive summary is

$$\kappa_{p2} = \begin{bmatrix} \delta \Gamma \mathbf{P}(x_1) \mathbf{1} \\ \delta \mathbf{P}(x_1) \Gamma \mathbf{P}(x_2) \mathbf{1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \frac{(\delta_1 \gamma_{11} + \bar{\delta}_1 \gamma_{21}) \exp(-\lambda_1) \lambda_1^{x_1}}{x_1!} + \frac{(\delta_1 \bar{\gamma}_{11} + \bar{\delta}_1 \bar{\gamma}_{21}) \exp(-\lambda_2) \lambda_2^{x_1}}{x_1!} \\ \frac{(\delta_1 \gamma_{11} + \bar{\delta}_1 \gamma_{21}) \gamma_{11} \exp(-2\lambda_1) \lambda_1^{x_1} \lambda_1^{x_2}}{x_1! x_2!} + \frac{(\delta_1 \bar{\gamma}_{11} + \bar{\delta}_1 \bar{\gamma}_{21}) \gamma_{21} \exp(-\lambda_1 - \lambda_2) \lambda_2^{x_1} \lambda_1^{x_2}}{x_1! x_2!} + \\ \frac{(\delta_1 \gamma_{11} + \bar{\delta}_1 \gamma_{21}) \bar{\gamma}_{11} \exp(-\lambda_1 - \lambda_2) \lambda_1^{x_1} \lambda_2^{x_2}}{x_1! x_2!} + \frac{(\delta_1 \bar{\gamma}_{11} + \bar{\delta}_1 \bar{\gamma}_{21}) \bar{\gamma}_{21} \exp(-2\lambda_2) \lambda_2^{x_1} \lambda_2^{x_2}}{x_1! x_2!} \\ \vdots \end{bmatrix},$$

where $\bar{\delta}_i = 1 - \delta_i$ and $\bar{\gamma}_{ij} = 1 - \gamma_{ij}$. There are 4 parameters $\theta = [\lambda_1, \lambda_2, \gamma_{11}, \gamma_{21}]$. We show in the Maple file `Example1.mw` that when there are $T = 3$ observations then the derivative matrix $\partial \kappa_{p2} / \partial \theta$ has rank 3. As there are 4 parameters the model is non-identifiable. When $T = 4$ the derivative matrix has rank 4, so the model is at least locally identifiable. Increasing T to 5 adds one extra exhaustive summary term, but no extra parameters, so we can therefore apply the Extension Theorem without the need to examine another derivative matrix, concluding that the model will have rank 4 and be at least locally identifiable for any $T \geq 4$.

It is not possible to show whether the modal is only locally identifiable symbolically, as Maple cannot solve the equations $\kappa = \mathbf{k}$.

We can extend this model by increasing the number of states. However, as we increase the number of states the exhaustive summary becomes more complex. We also need a larger number of exhaustive summary terms to achieve identifiability, and each exhaustive summary term is successively more complex. It is recommended instead to use the hybrid symbolic-numerical method. For example when $m = 3$ there are 9 parameters. When $T = 8$ the ranks of the derivative matrix evaluated at 5 random points in the parameter space were 7, 8, 8, 8, 8. The model rank is the maximum of these 5 ranks, so is 8. However, as there are 9 parameters the model is parameter redundant with deficiency 8. When $T = 9$ the model rank is 9, so the model is at least locally identifiable. The code for the hybrid symbolic-numerical method is given in the Maple code for this example.

For this example we require $T \geq m^2$ to ensure model is at least locally identifiable. As m^2 is the number of parameters, the limiting factor here is that there must be at least as many

exhaustive summary terms as parameters for identifiability, which is a requirement in any example, see for example [17].

Example 2: mark-recapture-recovery model

For the mark-recapture-recovery model the contribution to the exhaustive summary for the i th observation with capture history h_i is

$$\kappa_{h_i} = \begin{bmatrix} \delta \Gamma_1 \mathbf{P}(x_2) \mathbf{1} \\ \delta \Gamma_1 \mathbf{P}(x_2) \Gamma_2 \mathbf{P}(x_3) \mathbf{1} \\ \vdots \end{bmatrix}.$$

For example the capture history $h_i = 112$ would have exhaustive summary contribution

$$\kappa_{h_i=112} = \begin{bmatrix} \delta \Gamma_1 \mathbf{P}(x_2 = 1) \mathbf{1} \\ \delta \Gamma_1 \mathbf{P}(x_2 = 1) \Gamma_2 \mathbf{P}(x_3 = 2) \mathbf{1} \end{bmatrix} = \begin{bmatrix} \phi_1 p_2 \\ \phi_1 p_2 (1 - \phi_2) \lambda_3 \end{bmatrix}.$$

The exhaustive summary then combines the exhaustive summary contribution for each individual capture history in the study. To check the identifiability of the model we consider an exhaustive summary with every possible history combination. Ignoring repeated terms for $T = 3$ years of data the exhaustive summary is

$$\kappa_{MRR2} = \begin{bmatrix} (1 - \phi_1) \lambda_2 \\ \phi_1 p_2 \\ \phi_1 p_2 (1 - \phi_2) \lambda_3 \\ \phi_1 p_2 \phi_2 p_3 \\ \phi_1 p_2 \phi_2 (1 - p_3) + \phi_1 p_2 (1 - \phi_2) (1 - \lambda_3) \\ \phi_1 (1 - p_2) + (1 - \phi_1) (1 - \lambda_2) \\ \phi_1 (1 - p_2) (1 - \phi_2) \lambda_3 \\ \phi_1 (1 - p_2) \phi_2 p_3 \\ \phi_1 (1 - p_2) \phi_2 (1 - p_3) + \phi_1 (1 - p_2) (1 - \phi_2) (1 - \lambda_3) + (1 - \phi_1) (1 - \lambda_2) \\ (1 - \phi_2) \lambda_3 \\ \phi_2 p_3 \\ \phi_2 (1 - p_3) + (1 - \phi_2) (1 - \lambda_3) \end{bmatrix}.$$

The parameters are $\theta_{MRR} = [\phi_1, \phi_2, p_2, p_3, \lambda_2, \lambda_3]$. In the Maple code, `Example2.mw`, given in the supplementary material, we show that the derivative matrix $\mathbf{D} = \partial \kappa_{MRR2} / \partial \theta_{MRR}$ has rank 5. However, there are 6 parameters, so the model is non-identifiable.

It is also possible to use the Extension Theorem to show that this model is always non-identifiable with deficiency 1 for $T \geq 3$, as we demonstrate in the Maple code.

The Maple code also demonstrates how the probabilities of each capture history could also be used to create an exhaustive summary. Any alternative exhaustive summary will give the same results on identifiability. The advantage of HMMs exhaustive summary is that the Extension Theorem is more straightforward to apply. However, a simple exhaustive summary has been given in [24].

Example 3: longitudinal HMM

For the longitudinal HMM we start by considering the case with $m = 2$ states. Suppose there are $T = 2$ observations and every possible combination of states is observed at least once, then the exhaustive summary is

$$\kappa_L = \begin{bmatrix} \delta\mathbf{P}(1)\mathbf{1} \\ \delta\mathbf{P}(2)\mathbf{1} \\ \delta\mathbf{P}(1)\Gamma\mathbf{P}(1)\mathbf{1} \\ \delta\mathbf{P}(1)\Gamma\mathbf{P}(2)\mathbf{1} \\ \delta\mathbf{P}(2)\Gamma\mathbf{P}(1)\mathbf{1} \\ \delta\mathbf{P}(2)\Gamma\mathbf{P}(2)\mathbf{1} \end{bmatrix} = \begin{bmatrix} \delta_1\bar{b}_{21} + \bar{\delta}_1b_{21} \\ \delta_1b_{12} + \bar{\delta}_1\bar{b}_{12} \\ \delta_1\bar{b}_{21}^2\bar{\gamma}_{12} + \bar{\delta}_1b_{21}\bar{b}_{21}\gamma_{21} + \delta_1b_{21}\bar{b}_{21}\gamma_{12} + \bar{\delta}_1b_{21}^2\bar{\gamma}_{21} \\ \delta_1\bar{b}_{21}b_{12}\bar{\gamma}_{12} + \bar{\delta}_1b_{21}b_{12}\gamma_{21} + \delta_1\bar{b}_{21}\bar{b}_{12}\gamma_{12} + \bar{\delta}_1b_{21}\bar{b}_{12}\bar{\gamma}_{21} \\ \delta_1b_{12}b_{12}\bar{\gamma}_{12} + \bar{\delta}_1\bar{b}_{21}\bar{b}_{12}\gamma_{21} + \delta_1b_{21}b_{12}\gamma_{12} + \bar{\delta}_1b_{21}\bar{b}_{12}\bar{\gamma}_{21} \\ \delta_1b_{12}^2\bar{\gamma}_{12} + \bar{\delta}_1b_{12}\bar{b}_{12}\gamma_{21} + \delta_1b_{12}\bar{b}_{12}\gamma_{12} + \bar{\delta}_1\bar{b}_{12}^2\bar{\gamma}_{21} \end{bmatrix},$$

where $\bar{\gamma}_{ij} = 1 - \gamma_{ij}$ and $\bar{b}_{ij} = 1 - b_{ij}$. The parameters are $\theta_L = [\gamma_{12}, \gamma_{21}, b_{12}, b_{21}, \delta_1]$. The derivative matrix $\mathbf{D} = \partial\kappa_L/\partial\theta_L$ has rank 4. As there are 5 parameters the model is non-identifiable. For $T = 3$ the appropriate derivative matrix has rank 5, so the model is at least locally identifiable. We show in the Maple code, by solving the equations $\kappa = \mathbf{k}$, that this model is locally identifiable. Increasing T by 1 would add no extra parameters, therefore by the Extension Theorem this model is always locally identifiable for $T \geq 3$. Similarly, as demonstrated in the Maple code, when there are $m = 3$ states, there are 14 parameters and the rank of the appropriate derivative matrix is 14 when $T \geq 3$.

Bureau et al. [4] and Benoit et al. [2] have a different form for the transition rates. However, we can infer the identifiability results using the reparameterisation Theorem from [14]. This Theorem states that if the exhaustive $\kappa(\theta)$ is reparameterised in terms of $\mathbf{s} = g(\theta)$ for some function g , to give $\kappa(\mathbf{s})$, then the rank of $\mathbf{D}_s = \partial\kappa(\mathbf{s})/\partial\mathbf{s}$ is the same as the rank of $\mathbf{D} = \partial\kappa(\theta)/\partial\theta$ as long as $\text{rank}(\partial\mathbf{s}/\partial\theta) = q_s$, where q_s is the length of \mathbf{s} .

For example in the three state model, suppose the amount of time a patient stays in category i follows an exponential distribution with mean $(\sum_{j=1, i \neq j}^3 \lambda_{ij})^{-1}$ and the probability a patient changes state is $\lambda_{ij} / (\sum_{j=1, i \neq j}^3 \lambda_{ij})^{-1}$, and we have equally spaced time periods in increments of 1, then

$$\Gamma_2 = \begin{bmatrix} 1 - \exp\left(-\frac{1}{\lambda_{12} + \lambda_{13}}\right) & \frac{\lambda_{12}}{\lambda_{12} + \lambda_{13}} \exp\left(-\frac{1}{\lambda_{12} + \lambda_{13}}\right) & \frac{\lambda_{13}}{\lambda_{12} + \lambda_{13}} \exp\left(-\frac{1}{\lambda_{12} + \lambda_{13}}\right) \\ \frac{\lambda_{21}}{\lambda_{21} + \lambda_{23}} \exp\left(-\frac{1}{\lambda_{21} + \lambda_{23}}\right) & 1 - \exp\left(-\frac{1}{\lambda_{21} + \lambda_{23}}\right) & \frac{\lambda_{23}}{\lambda_{21} + \lambda_{23}} \exp\left(-\frac{1}{\lambda_{21} + \lambda_{23}}\right) \\ \frac{\lambda_{31}}{\lambda_{31} + \lambda_{32}} \exp\left(-\frac{1}{\lambda_{31} + \lambda_{32}}\right) & \frac{\lambda_{32}}{\lambda_{31} + \lambda_{32}} \exp\left(-\frac{1}{\lambda_{31} + \lambda_{32}}\right) & 1 - \exp\left(-\frac{1}{\lambda_{31} + \lambda_{32}}\right) \end{bmatrix}.$$

Γ_2 is a reparameterisation of Γ with

$$\gamma_{ij} = \frac{\lambda_{ij}}{\sum_{j=1, i \neq j}^3 \lambda_{ij}} \exp\left(-\frac{1}{\sum_{j=1, i \neq j}^3 \lambda_{ij}}\right) \quad i \neq j.$$

The 14 new parameters are

$$\theta_{L2} = [\lambda_{12}, \lambda_{13}, \lambda_{21}, \lambda_{23}, \lambda_{31}, \lambda_{32}, b_{12}, b_{13}, b_{21}, b_{23}, b_{31}, b_{32}, \delta_1, \delta_2].$$

By the reparameterisation Theorem the rank will also be 14, so again the model is identifiable.

4 Discussion

In this paper we have demonstrated some of the tools that can be used to check identifiability in HMMs.

We recommend that you do not use the Hessian method to determine identifiability in HMMs, as it can be inaccurate.

The log-likelihood profile method is a suitable numerical method for HMMs. However, the log-likelihood profile method can be difficult to distinguish non-identifiability and near redundancy, and would require a profile for every parameter. Theoretically it can also be used to find estimable parameter combinations, using a method called subset profiling; see [20].

The symbolic method for checking identifiability is always accurate, but it requires a symbolic algebra package to execute. We have provided a new exhaustive summary for HMMs, however this exhaustive summary gets consecutively more complex for larger T . The symbolic algebra package can run out of memory calculating the rank of the appropriate derivative matrix, in which case we recommend using the hybrid symbolic-numerical method. This method can be used to get results for any T using the Extension Theorem, and can also give estimable parameter combinations in non-identifiable models.

Although a symbolic method exists for distinguishing between local and global identifiability, for most HMMs the exhaustive summary will be symbolically too complex for a symbolic algebra package to solve the appropriate equations. It was only possible for the simplest case in Example 3. We therefore recommend using the log-likelihood profile method to check for local identifiability. We note that if a model is found to be globally identifiable from a profile it is only a valid result for the region of the profile examined. A second optimum could potentially exist outside of the region of the graph. Local identifiability can typically be changed to global identifiability using an order constraint.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Allman, E.S., Matias, C., Rhodes, J.A.: Identifiability of parameters in latent structure models with many observed variables. *Ann. Stat.* **37**, 3099–3132 (2009)
2. Benoit, J.S., Chan, W., Luo, S., Yeh, H.-W., Doody, R.: A hidden Markov model approach to analyze longitudinal ternary outcomes when some observed states are possibly misclassified. *Stat. Med.* **35**, 1549–1557 (2016)
3. Bulla, J., Berzel, A.: Computational issues in parameter estimation for stationary hidden Markov models. *Comput. Stat.* **23**, 1–1 (2008)
4. Bureau, A., Shiboski, S., Huges, J.P.: Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. *Stat. Med.* **22**, 441–462 (2003)
5. Campbell, D., Lele, S.: An ANOVA test for parameter estimability using data cloning with application to statistical inference for dynamic systems. *Comput. Stat. Data Anal.* **70**, 257–267 (2014)
6. Catchpole, E.A., Morgan, B.J.T.: Detecting parameter redundancy. *Biometrika* **85**, 187–196 (1997)
7. Catchpole, E.A., Freeman, S.N., Morgan, B.J.T., Harris, M.P.: Integrated recovery/recapture data analysis. *Biometrics* **54**, 33–46 (1998)
8. Catchpole, E.A., Morgan, B.J.T., Freeman, S.N.: Estimation in parameter redundant models. *Biometrika* **85**, 462–468 (1998)
9. Catchpole, E.A., Morgan, B.J.T., Coulson, T.N., Freeman, S.N., Albon, S.D.: Factors influencing soay sheep survival. *J. R. Stat. Soc. Ser. C* **49**, 453–472 (2000)
10. Catchpole, E.A., Kgos, P.M., Morgan, B.J.T.: On the near-singularity of models for animal recovery data. *Biometrics* **57**, 720–726 (2001)
11. Chappell, M.J., Gunn, R.N.: A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach. *Math. Biosci.* **148**, 21–41 (1998)
12. Chis, O.-T., Villaverde, A.F., Banga, J.R., Balsa-Canto, E.: On the relationship between sloppiness and identifiability. *Math. Biosci.* **282**, 147–161 (2016)
13. Choquet, R., Cole, D.J.: A hybrid symbolic-numerical method for determining model structure. *Math. Biosci.* **236**, 117–125 (2012)

14. Cole, D.J., Morgan, B.J.T., Titterton, D.M.: Determining the parametric structure of models. *Math. Biosci.* **228**, 16–30 (2010)
15. Cole, D.J., Morgan, B.J.T.: A note on determining parameter redundancy in age-dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *J. Agric. Biol. Environ. Stat.* **15**, 431–434 (2010)
16. Cole, D.J.: Determining parameter redundancy of multi-state mark-recapture models for sea birds. *J. Ornithol.* **152**, 305–315 (2012)
17. Cole, D.J., Morgan, B.J.T., Catchpole, E.A., Hubbard, B.A.: Parameter redundancy in mark-recovery models. *Biom. J.* **54**, 507–523 (2012)
18. Cole, D.J., McCrea, R.S.: Parameter redundancy in discrete state-space and integrated models. *Biom. J.* **58**, 1071–1090 (2016)
19. Dasgupta, A., Self, S.G., Gupta, S.D.: Non-identifiable parametric probability models and reparameterization. *J. Stat. Plan. Inference* **137**, 3380–3393 (2007)
20. Eisenberg, M.C., Hayashi, M.A.L.: Determining identifiable parameter combinations using subset profiling. *Math. Biosci.* **256**, 116–126 (2014)
21. Evans, N.D., Chappell, M.J.: Extensions to a procedure for generating locally identifiable reparameterizations of unidentifiable systems. *Math. Biosci.* **168**, 137–159 (2000)
22. Gimenez, O., Viallefont, A., Catchpole, E.A., Choquet, R., Morgan, B.J.T.: Methods for investigating parameter redundancy. *Anim. Biodivers. Conserv.* **27**, 1–12 (2004)
23. Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Dorris, B., Varella, H., Vila, J., Pradel, R.: State-space modelling of data on marked individuals. *Ecol. Model.* **206**, 431–438 (2007)
24. Hubbard, B.A., Cole, D.J., Morgan, B.J.T.: Parameter redundancy in capture-recapture-recovery models. *Stat. Methodol.* **17**, 17–29 (2014)
25. Jiang, H., Pollock, K.H., Brownie, C., Hightower, J.E., Hoenig, J.E., Hearn, W.S.: Age-dependent tag return models for estimating fishing mortality, natural mortality and selectivity. *J. Agric. Biol. Environ. Stat.* **12**, 177–194 (2007)
26. King, R., Brooks, S.P.: Closed-form likelihoods for Arnason-Schwarz models. *Biometrika* **90**, 435–444 (2003)
27. Konstantinides, K., Yao, K.: Statistical analysis of effective singular values in matrix rank determination. *IEEE Trans. Acoust. Speech Signal Process.* **36**, 757–763 (1998)
28. Leroux, B.G.: Maximum likelihood estimation for hidden Markov models. *Stoch. Process. Appl.* **40**, 127–143 (1992)
29. Little, M.P., Heidenreich, W.F., Li, G.: Parameter identifiability and redundancy: theoretical considerations. *PLoS ONE* **4**, e8915 (2009)
30. Little, M.P., Heidenreich, W.F., Li, G.: Parameter identifiability and redundancy in a general class of stochastic carcinogenesis models. *PLoS ONE* **4**, e8520 (2010)
31. McCrea, R.S.: Sufficient statistic likelihood construction for age- and time-dependent multi-state joint recapture and recovery data. *Stat. Prob. Lett.* **82**, 357–359 (2012)
32. McCrea, R., Morgan, B.J.T.: *Analysis of Capture–Recapture Data*. Chapman and Hall/CRC Press, Cambridge (2014)
33. Petrie, T.: Probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* **40**, 97–115 (1969)
34. Redner, R.A., Walker, H.F.: Mixture densities, maximum likelihood and the EM algorithm. *SIAM Rev.* **26**, 195–239 (1984)
35. Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J.: Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929 (2009)
36. Rothenberg, T.J.: Identification in parametric models. *Econometrica* **39**, 577–591 (1971)
37. Visser, I., Raijmakers, M.E.J., Molenaar, P.C.M.: Confidence intervals for hidden Markov model parameters. *Br. J. Math. Stat. Psychol.* **53**, 17–327 (2000)
38. Viallefont, A., Lebreton, J.-D., Reboulet, A.-M.: Parameter identifiability and model selection in capture-recapture models: a numerical approach. *Biom. J.* **40**, 313–325 (1998)
39. Yeh, H.-W., Chan, W., Symanski, E.: Intermittent missing observations in discrete-time hidden Markov models. *Commun. Stat. Simul. Comput.* **41**, 167–181 (2012)
40. Zucchini, W., MacDonald, I.L., Langrock, R.: *Hidden Markov Models for Time Series*. CRC Press, Cambridge (2016)