Individual based population inference using tagging data

IMM-Technical-Report-2010-11
Department for Informatics and Mathematical Modelling, Technical University of Denmark
Martin W. Pedersen, Uffe H. Thygesen, Henrik Baktoft, Henrik Madsen
This version compiled: September 28, 2010.

Abstract

A hierarchical framework for simultaneous analysis of multiple related individual datasets is presented. The approach is very similar to mixed effects modelling as known from statistical theory. The model used at the individual level is, in principle, irrelevant as long as a maximum likelihood estimate and its uncertainty (Hessian) can be computed. The individual model used in this text is a hidden Markov model. A simulation study concerning a two-dimensional biased random walk is examined to verify the consistency of the hierarchical estimation framework. In addition, a study based on acoustic telemetry data from pike illustrates how the framework can identify individuals that deviate from the remaining population.

Contents

1 Introduction

The development and availability of electronic tags have revolutionised the study of individual animal movement. Often, however, the purpose of tagging studies is to investigate movement and behaviour patterns in the population rather than at the individual. Models with random effects is the common statistical tool for population analysis of individual measurements. Unfortunately they are not straightforward to employ in the context of animal movement, since the movement of an individual is not easily parametrised such that meaningful population level patterns are captured.

Some studies have integrated state-space models (SSMs) for individual analysis into population frameworks. In drug development SSMs are used to model the dynamics of the concentration of chemical compounds in the blood. Nonlinear mixed effects models have been used to provide improved parameter estimates because variability between individuals is captured. This enables joint analysis of data from multiple and possibly unbalanced studies (Tornøe, 2005). It is therefore tempting to take a similar approach and combine individual SSMs for animal movement to infer population trends.

Using Bayesian methods, Jonsen et al. (2003) implemented a hierarchical model for combining multiple individual SSMs for simulated movement data. Their inference focused on a parameter which related movement rate to the sea surface temperature experienced by turtles. The results of the study clearly illustrated the inferential strength of sharing information between individuals to improve estimation. The same hierarchical approach was taken in Jonsen et al. (2006) to reveal diel variation in travel rates of migrating leatherback turtles. Few other studies are found in the literature that deal with the difficult task of jointly analyse movement data from multiple individuals.

Aarts et al. (2008) present the, perhaps, most extensive (non state-space) attempt to model population space use using individual tagging data. The paper examines grey seal habitat preference with a case-control model. Outliers present in the telemetry data are removed with a heuristic scheme and the remaining locations are smoothed temporally with a generalized additive model (GAM). A number of static environmental variables (sediment type, depth, distance from haul-out) are related to the number of observed locations in a region as covariates. Thus, the model can be used for predicting the spatial usage of the species as a function of the covariates. The model, as discussed by the authors, ignores that the location data used for estimation is autocorrelated.

The present text studies the use of mixed effects models to combine data from multiple electronic tags with the aim to draw conclusions about the population. The focus is not on explicitly modelling the movement of the population, but rather on parameters that are related to the population movement, e.g. movement rate. First the theory for hierarchical models based on likelihood functions from multiple individuals is reviewed. This framework is similar to the empirical Bayesian method presented by Efron (1996) or the mixed effects framework as described in Pawitan (2001). In a simulation study the hierarchical model is used to merge individually estimated SSMs for movement data with observation error. In another study accurate real acoustic telemetry data from pike were used to distinguish individuals that displayed a deviating behaviour as compared to the remaining population.

2 The hierarchical model

The population has the parameter vector $\boldsymbol{\theta}$. Then, individual $i \in \{1, \ldots, M\}$ of the population has a parameter vector given by

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{w}_i, \tag{1}$$

where

$$\boldsymbol{w}_i \sim N(\mathbf{0}, \boldsymbol{W}).$$

In mixed-effects modelling $\boldsymbol{\theta}$ are referred to as the fixed effects and $\boldsymbol{w}_i$ are the mutually independent random effects. The dataset related to individual i has N_i observations. A general model for the observed data $\mathcal{Z}_{N_i}^{(i)} = \left\{\boldsymbol{z}_1^{(i)}, \ldots, \boldsymbol{z}_{N_i}^{(i)}\right\}$ from individual i is

$$\mathcal{Z}_{N_i}^{(i)} = f(\boldsymbol{\theta}_i, \boldsymbol{\Theta}), \tag{2}$$

where Θ covers other parameters required to generate data. Here, we assume that Θ is known (i.e. it can be estimated from independent data). The form of f is arbitrary, however here only models with noise (randomness) are considered, for example f could be a stochastic SSM. In this case the parameters must be estimated using the probability density of the data conditional on the parameters is $p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}_i\right)$. For time series data the observation density is typically obtained by a filtering procedure.

When viewed as a function of $\boldsymbol{\theta}_i$ the observation density is the likelihood function for the parameters of individual i, i.e. we have

$$L(\boldsymbol{\theta}_i) = p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}_i\right), \tag{3}$$

and therefore that the maximum likelihood (ML) estimate of $\boldsymbol{\theta}_i$ is

$$\widehat{\boldsymbol{\theta}}_i = \arg\max_{\boldsymbol{\theta}_i} L(\boldsymbol{\theta}_i), \tag{4}$$

which can be determined independently of the other individuals. The uncertainty of $\widehat{\boldsymbol{\theta}}_i$ is described by covariance $\boldsymbol{\Sigma}_i$ of the parameter estimate, which is computed as the inverse Hessian evaluated at the optimum of the likelihood function.

The joint probability density of the random effects and individual observations conditional on $\boldsymbol{\theta}$ and $\boldsymbol{W}$ is

$$p\left(\boldsymbol{w}_i, \mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \boldsymbol{W}\right) = p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \boldsymbol{w}_i\right) p\left(\boldsymbol{w}_i|\boldsymbol{W}\right), \tag{5}$$

by the definition of conditional densities. In (5) the first term on the right-hand side is equal to (3) since $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{w}_i$. The joint likelihood function related to the random effects and the model parameters is therefore

$$L(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{w}_i) = p\left(\boldsymbol{w}_i, \mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \boldsymbol{W}\right).$$

Then, the ML estimate of the random effects for individual i with fixed $\boldsymbol{\theta}$ and $\boldsymbol{W}$ is

$$\widehat{\boldsymbol{w}}_i = \arg\max_{\boldsymbol{w}} L(\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{w}_i). \tag{6}$$

The population parameters are also of interest so we marginalise over the random effects and get

$$p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta},\boldsymbol{W}\right)=\int p\left(\boldsymbol{w}_i,\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta},\boldsymbol{W}\right)d\boldsymbol{w}_i. \tag{7}$$

This leads to the likelihood function for the population parameter given data from the i'th individual

$$L\left(\boldsymbol{\theta},\boldsymbol{W}|\mathcal{Z}_{N_i}^{(i)}\right)=p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta},\boldsymbol{W}\right). \tag{8}$$

Individuals are conditional independent given $\boldsymbol{\theta}$ and $\boldsymbol{W}$. Thus, the full population likelihood, i.e. the likelihood given data from all individuals, is the product of the individual likelihood contributions

$$L\left(\boldsymbol{\theta},\boldsymbol{W}|\mathcal{Z}\right)=\prod_{i=1}^{M}\int p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta},\boldsymbol{w}_i\right)p\left(\boldsymbol{w}_i|\boldsymbol{W}\right)d\boldsymbol{w}_i, \tag{9}$$

where $\mathcal{Z}=\left\{\mathcal{Z}_{N_1}^{(1)},\ldots,\mathcal{Z}_{N_M}^{(M)}\right\}$. Therefore, the ML estimate of the population parameters is

$$(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{W}})=\arg\max_{\boldsymbol{\theta},\boldsymbol{W}}\left\{L(\boldsymbol{\theta},\boldsymbol{W}|\mathcal{Z})\right\}. \tag{10}$$

2.1 Excluding deviating individuals

Say a population of M individuals has been analysed with the framework described above. Then a new dataset becomes available from a new individual, which possibly belongs to the same population. The parameter estimate and parameter covariance matrix for the new individual are $\widehat{\boldsymbol{\theta}}_a$ and $\boldsymbol{\Sigma}_a$ respectively. Two hypotheses are defined:

H_0**:** The new individual comes from the same population as the other individuals.

H_1**:** The new individual does not come from the same population as the other individuals.

Under H_0 it holds that

$$\boldsymbol{\theta}_a\sim N(\boldsymbol{\theta},\boldsymbol{W}),\quad \widehat{\boldsymbol{\theta}}_a|\boldsymbol{\theta}_a\sim N(\boldsymbol{\theta}_a,\boldsymbol{\Sigma}_a),$$

which leads to

$$\widehat{\boldsymbol{\theta}}_a\sim N(\boldsymbol{\theta},\boldsymbol{W}+\boldsymbol{\Sigma}_a),$$

using the rules for conditional mean and variance. The H_0 hypothesis can be tested with

$$S_a=(\widehat{\boldsymbol{\theta}}_a-\boldsymbol{\theta})^T(\boldsymbol{\Sigma}_a+\boldsymbol{W})^{-1}(\widehat{\boldsymbol{\theta}}_a-\boldsymbol{\theta})\sim\chi^2(n),$$

where n is the number of parameters in $\boldsymbol{\theta}$, i.e. the dimension of the parameter space. So, H_0 is rejected if

$$S_a>\chi^2(n)_{1-\alpha},$$

where the conventional level of significance, $\alpha=0.05$, is chosen.

This simple test can be used to eliminate individuals that deviate from the population by setting $a=i$ and comparing with the population comprised by all individuals except i. This procedure is carried out for all i. The individual that deviates the most (smallest p-value) is eliminated from the population.

Then for the remaining $M-1$ individuals the procedure is repeated. The scheme runs until no individuals deviate from the population. Before the population comprised by the remaining individuals is accepted forward selection of the eliminated individuals may be performed. That is, using the above test to ensure that none of all the eliminated individuals can be included in the reduced population. This might be the case as the population composition has changed since the first individual was eliminated.

3 Algorithm for estimating the hierarchical model

It is difficult to estimate the random effects and the population parameters simultaneously because their respective likelihood functions are coupled. That is, when estimating $\boldsymbol{w}_i$ values of $\boldsymbol{\theta}$ and $\boldsymbol{W}$ are required, when estimating $\boldsymbol{\theta}$ the value of $\boldsymbol{W}$ is required and finally for estimating $\boldsymbol{W}$ values for $\boldsymbol{\theta}$ and $\boldsymbol{w}_i$ for all i are required. Instead of direct numerical optimisation of all parameters, an iterative algorithm (Pawitan, 2001) is employed:

1. Set $\boldsymbol{W} = \widehat{\boldsymbol{W}}$, where $\widehat{\boldsymbol{W}}$ is a starting guess.
2. Compute the estimate $\widehat{\boldsymbol{\theta}}$ using $\widehat{\boldsymbol{W}}$.
3. Compute the estimate $\widehat{\boldsymbol{w}}_i$ for all i using $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{W}}$.
4. Update $\widehat{\boldsymbol{W}}$ using $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{w}_i$.
5. Iterate step 2 to 4 until convergence.

It is clear, however, that step 2 and 4 require that the integral (7) over the random effects be computed. This integral is the main challenge of parameter estimation in a nonlinear mixed-effects model. The optimisation routine that maximises (9) requires for each function evaluation that (8) be computed for all individuals. It not possible in general to compute the integral on closed form and therefore approximation schemes must be employed. The computing effort required to evaluate (8) with quadrature based algorithms grows rapidly in n (the number of parameters and dimension of the integral). Therefore, even for moderate values of n these methods are not suitable. Alternative approaches to solving the integral are Monte Carlo simulation, first-order conditional estimation (FOCE) and the Laplacian approximation. Here, an approach similar to the latter is employed.

The individual log-likelihood function

$$l(\boldsymbol{\theta}_i) = \log p\left(\mathcal{Z}_{N_i}^{(i)} | \boldsymbol{\theta}, \boldsymbol{w}_i\right) \tag{11}$$

is assumed to take a quadratic form, i.e. it satisfies

$$l(\boldsymbol{\theta}_i) = K_i + \log\left(|2\pi\boldsymbol{\Sigma}_i|^{1/2}\right) - \log\left(|2\pi\boldsymbol{\Sigma}_i|^{1/2}\right) - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta} - \boldsymbol{w}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta} - \boldsymbol{w}_i), \tag{12}$$

where $\boldsymbol{\Sigma}_i$ comes from the inverse Hessian of the individual likelihood function at $\widehat{\boldsymbol{\theta}}_i$, i.e. the observed Fisher information, and K_i is value of the individual log-likelihood function at its maximum. Note that the two latter terms of (12) comprise a Gaussian density in the log-domain. The quadratic form is a reasonable assumption since the likelihood function is asymptotically Gaussian around the maximum likelihood estimate (Wasserman, 2005). Assuming a quadratic form for $l(\boldsymbol{\theta}_i)$ is equivalent to developing

the second-order Taylor expansion of $l(\boldsymbol{\theta}_i)$ around its maximiser $\widehat{\boldsymbol{\theta}}_i$. As mentioned above this technique is similar to the Laplace approximation (Vonesh, 1996).

By assumption, the log-density of the random effects also has a quadratic form. Therefore, it is evident that the log of (5) is

$$\begin{aligned} l_i &= \log\left[p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta},\boldsymbol{w}_i\right)p\left(\boldsymbol{w}_i|\boldsymbol{W}\right)\right] \\ &= l(\boldsymbol{\theta}_i) + \log p\left(\boldsymbol{w}_i|\boldsymbol{W}\right) \\ &= K_i + \log\left(|2\pi\boldsymbol{\Sigma}_i|^{1/2}\right) - \log\left(|2\pi\boldsymbol{\Sigma}_i|^{1/2}\right) - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta} - \boldsymbol{w}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta} - \boldsymbol{w}_i) \\ &\quad - \log\left(|2\pi\boldsymbol{W}|^{1/2}\right) - \frac{1}{2}\boldsymbol{w}_i^T\boldsymbol{W}_i^{-1}\boldsymbol{w}_i. \end{aligned}$$

For $\boldsymbol{\theta}_i = \widehat{\boldsymbol{\theta}}_i$ and $\boldsymbol{W} = \widehat{\boldsymbol{W}}$, the estimate for the random effects is found by taking the derivative of l_i with respect to $\boldsymbol{w}_i$:

$$\frac{\partial l_i}{\partial \boldsymbol{w}_i} = -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta} - \boldsymbol{w}_i) - \frac{1}{2}\widehat{\boldsymbol{W}}_i^{-1}\boldsymbol{w}_i.$$

Equating to the zero-vector and solving for $\boldsymbol{w}_i$ gives the random-effects estimate

$$\widehat{\boldsymbol{w}}_i = (\boldsymbol{\Sigma}_i^{-1} + \widehat{\boldsymbol{W}}^{-1})^{-1}\boldsymbol{\Sigma}_i^{-1}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}). \tag{13}$$

The covariance of the random effects is therefore

$$\boldsymbol{S}_i = (\boldsymbol{\Sigma}_i^{-1} + \widehat{\boldsymbol{W}}^{-1})^{-1}.$$

Now, while dropping unimportant constant terms, the population likelihood (9) can be rewritten as

$$\begin{aligned} l\left(\boldsymbol{\theta},\boldsymbol{W}|\mathcal{Z}\right) &= \sum_{i=1}^{M}\log\left(\int\exp\left(l_i\right)d\boldsymbol{w}_i\right) \\ &= \sum_{i=1}^{M} -\frac{1}{2}\log\left(|\boldsymbol{\Sigma}_i + \boldsymbol{W}|\right) - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})^T(\boldsymbol{\Sigma}_i + \boldsymbol{W})^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta}). \end{aligned} \tag{14}$$

This log-likelihood is similar to that of a linear mixed-model with the exception that the individuals have different covariance matrices $\boldsymbol{\Sigma}_i$ whereas for the standard linear model they are normally assumed equal across individuals (Pawitan, 2001).

For known $\boldsymbol{W} = \widehat{\boldsymbol{W}}$ and $\boldsymbol{V}_i = \boldsymbol{\Sigma}_i + \widehat{\boldsymbol{W}}$, a closed-form expression for the maximum likelihood estimate of $\boldsymbol{\theta}$ is now available by

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{\theta}} l\left(\boldsymbol{\theta}, \widehat{\boldsymbol{W}} | \mathcal{Z}\right) \\
\mathbf{0} &= \sum_{i=1}^{M} -\frac{1}{2} \boldsymbol{V}_i^{-1} (\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}) \\
\widehat{\boldsymbol{\theta}} &= \left[\sum_{i=1}^{M} \boldsymbol{V}_i^{-1}\right]^{-1} \left[\sum_{i=1}^{M} \boldsymbol{V}_i^{-1} \widehat{\boldsymbol{\theta}}_i\right]. \qquad (15)
\end{aligned}$$

The Hessian of $l\left(\boldsymbol{\theta}, \widehat{\boldsymbol{W}} | \mathcal{Z}\right)$ at the optimum is

$$\boldsymbol{H}_{\widehat{\boldsymbol{\theta}}} = \sum_{i=1}^{M} \boldsymbol{V}_i^{-1},$$

so the covariance matrix of $\widehat{\boldsymbol{\theta}}$ is $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}} = \boldsymbol{H}_{\widehat{\boldsymbol{\theta}}}^{-1}$.

The estimation procedure for the variance component $\boldsymbol{W}$ is not immediately tractable via (14) owing to the $\boldsymbol{V}_i$ terms which involve a sum of two covariances. With $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, $\boldsymbol{w}_i = \widehat{\boldsymbol{w}}_i$, and using equation (17.14) in Pawitan (2001), (14) can be rewritten as

$$\begin{aligned}
l\left(\widehat{\boldsymbol{\theta}}, \boldsymbol{W} | \mathcal{Z}\right) = \sum_{i=1}^{M} &-\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{w}}_i)^T \boldsymbol{\Sigma}_i^{-1} (\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{w}}_i) \\
&- \frac{1}{2}\log|\boldsymbol{W}| - \frac{1}{2}\widehat{\boldsymbol{w}}_i^T \boldsymbol{W}^{-1} \widehat{\boldsymbol{w}}_i - \frac{1}{2}\log|\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{W}^{-1}|. \qquad (16)
\end{aligned}$$

It is not possible in general to find an expression for $\boldsymbol{W}$ from (16). Therefore $\boldsymbol{W}$ has to be estimated numerically. Alternatively (16) can be simplified by assuming that $\boldsymbol{W} = \sigma_{\boldsymbol{w}}^2 \boldsymbol{R}$, i.e. that the structure of the covariance matrix of the random effects is known. Then (as in Pawitan, 2001) define the objective function

$$\begin{aligned}
Q = \sum_{i=1}^{M} &-\frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{w}}_i)^T \boldsymbol{\Sigma}_i^{-1} (\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{w}}_i) \\
&- \frac{n}{2}\log\sigma_{\boldsymbol{w}}^2 - \frac{1}{2\sigma_{\boldsymbol{w}}^2}\widehat{\boldsymbol{w}}_i^T \boldsymbol{R}^{-1} \widehat{\boldsymbol{w}}_i - \frac{1}{2}\log|\boldsymbol{\Sigma}_i^{-1} + \sigma_{\boldsymbol{w}}^{-2}\boldsymbol{R}^{-1}|.
\end{aligned}$$

With n parameters

$$\begin{aligned}
\frac{\partial Q}{\partial \sigma_{\boldsymbol{w}}^2} = \sum_{i=1}^{M} &-\frac{n}{2\sigma_{\boldsymbol{w}}^2} + \frac{1}{2\sigma_{\boldsymbol{w}}^4}\widehat{\boldsymbol{w}}_i^T \boldsymbol{R}^{-1} \widehat{\boldsymbol{w}}_i \\
&+ \frac{1}{2\sigma_{\boldsymbol{w}}^4}\mathrm{tr}\{(\boldsymbol{\Sigma}_i^{-1} + \sigma_{\boldsymbol{w}}^{-2}\boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\}. \qquad (17)
\end{aligned}$$

By equating (17) to zero it can be shown that $\sigma_{\boldsymbol{w}}^2$ can be updated via

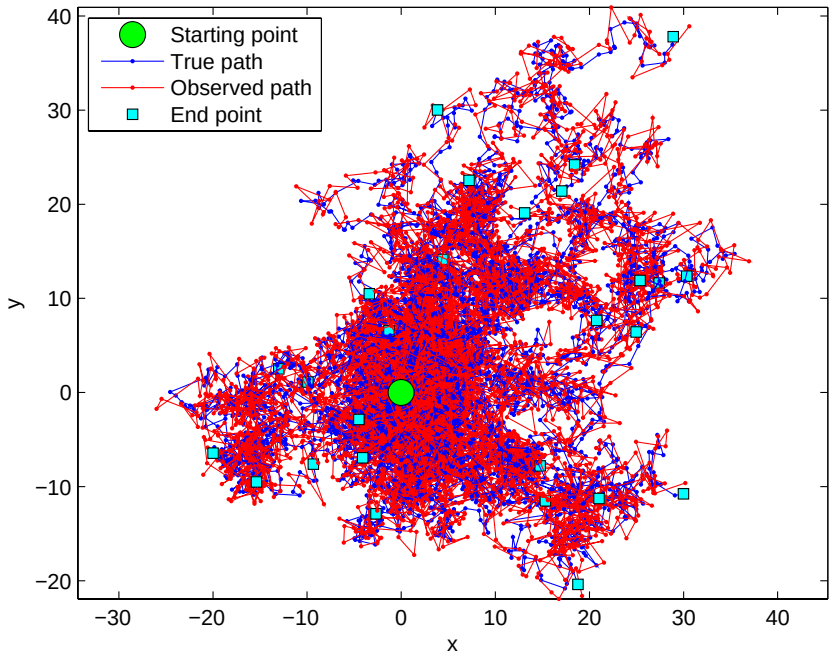

Figure 1: Simulated data from $M = 30$ individuals with a biased random walk in two dimensions.

$$\sigma_{\boldsymbol{w}}^2 = \frac{1}{Mn}\sum_{i=1}^{M} \widehat{\boldsymbol{w}}_i^T \boldsymbol{R}^{-1} \widehat{\boldsymbol{w}}_i + \mathrm{tr}\{(\boldsymbol{\Sigma}_i^{-1} + \sigma_{\boldsymbol{w}}^{-2}\boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\}. \tag{18}$$

It is not necessarily straightforward to determine the structure matrix $\boldsymbol{R}$. In the simplest case it may be set to the identity matrix ($\boldsymbol{I}$), however this may be a too rough approximation. An alternative and somewhat heuristic approach to get a more reasonable $\boldsymbol{R}$ is to do one iteration of the loop described in the beginning of this section with $\boldsymbol{R} = \boldsymbol{I}$. Then, using the estimated random effects it possible to empirically calculate $\boldsymbol{R}$, which can be used in subsequent iterations.

4 Examples

Here we use the presented methodology to analyse data from multiple individuals. First a simulation study is considered. Then real tagging data is analysed.

4.1 Simulation

In the simulation study data were generated from a two-dimensional SSM for $M = 30$ individuals (see Figure 1). The aim was to mimic an object moving in the plane. Specifically, data for the i'th individual were simulated from a biased random walk model

$$\boldsymbol{x}_{k+1}^{(i)} = \boldsymbol{x}_k^{(i)} + \boldsymbol{u}_i + \boldsymbol{\nu}_k^{(i)}, \tag{19}$$

Param.	D	u_x	u_y	$\sigma^2_{\boldsymbol{\epsilon}}$	$\sigma^2_{\boldsymbol{w}}$	$\boldsymbol{R}$	M	N_i
Value	log(10)	1	0	1	0.3^2	$\boldsymbol{I}$	30	200

Table 1: Parameter values used for generating data for the simulation study.

where $\boldsymbol{x}_k^{(i)}$ is the two-dimensional location vector at time t_k, $\boldsymbol{u}_i$ is the drift (or advection) vector and $\boldsymbol{\nu}_k^{(i)} \sim N(\boldsymbol{0}, 2D_i\boldsymbol{I}dt)$. The time-step dt is constant in time and for all individuals. The observation equation is

$$\boldsymbol{y}_k^{(i)} = \boldsymbol{x}_k^{(i)} + \boldsymbol{\epsilon}_k^{(i)}, \tag{20}$$

where $\boldsymbol{y}_k^{(i)}$ is the observed location at time t_k and $\boldsymbol{\epsilon}_k^{(i)} \sim N(\boldsymbol{0}, \sigma^2_{\boldsymbol{\epsilon}}\boldsymbol{I})$. In this example it is assumed that $\sigma^2_{\boldsymbol{\epsilon}}$ is independent of i and known. Equations (19) and (20) comprise the mapping f in (2).

The individual parameters $\boldsymbol{\theta}_i = \{D_i, \boldsymbol{u}_i\}$ are generated from the population parameters $\boldsymbol{\theta} = \{D, \boldsymbol{u}\}$ as described by (1), restated here

$$\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{w}_i,$$

with $\boldsymbol{w}_i \sim N(\boldsymbol{0}, \sigma^2_{\boldsymbol{w}}\boldsymbol{R})$ with $\boldsymbol{R} = \boldsymbol{I}$. Data were generated with the parameter values shown in Table 1.

4.1.1 Estimation scheme

The only known parameters are $\boldsymbol{R} = \boldsymbol{I}$ and the variance of the observation noise $\sigma^2_{\boldsymbol{\epsilon}}$. All other parameters are estimated. First, all individual parameters $\boldsymbol{\theta}_i$ are estimated separately and independently of each other such that $\widehat{\boldsymbol{\theta}}_i$ and $\boldsymbol{\Sigma}_i$ is computed for all i, see (4). This estimation is carried out with a hidden Markov model (HMM), which discretises the two-dimensional domain into grid cells and solves the filtering equations on this grid. For further details see Thygesen et al. (2009). Note that the simple SSM considered here could be estimated using the Kalman filter. However, the purpose of the simulation study is to show the use of mixed effect modelling together with HMMs because this framework generalises to nonlinear and non-Gaussian SSMs.

The model parameters are estimated with the recursive scheme described in Section 3. With the starting guess $\sigma^2_{\boldsymbol{w}} = 1$ the population parameters are estimated with (15). The random effects are then estimated with (13) using the previous values for $\widehat{\boldsymbol{\theta}}$ and $\sigma^2_{\boldsymbol{w}}$. The final step in the recursion is to update the value of $\sigma^2_{\boldsymbol{w}}$ with (18). This loop continues until the parameter values converge. The recursive scheme is very similar to an Expectation-Maximization algorithm, which is a derivative-free approach to ML estimation. It is guaranteed that the likelihood will increase with every iteration, however sometimes the algorithm converges slowly. Fortunately, all the estimation steps in the algorithm have closed-form solutions (subject to some assumptions). This allows the recursion to converge rapidly.

4.1.2 Estimation results

Estimation time of one individual was approximately three minutes on a standard desktop computer. Obviously, this time depends on the resolution of the discrete grid in the HMM, which in turn depends on the parameter values (or rather the path of the simulated data). The computing time spent to estimate random effects, random effects variance, and population parameters was around one second. This estimation was quick because only analytical expressions are part of the estimation procedure.

Test	D	u_x	u_y	$\sigma^2_{\boldsymbol{w}}$
1	(8.85 9.87 11.02)	(0.35 0.80 1.25)	(−0.38 0.06 0.50)	0.27^2

Table 2: Results from simulation study. Estimated population parameter values with 95% confidence bounds. Estimated of diffusivity are transformed back from log.

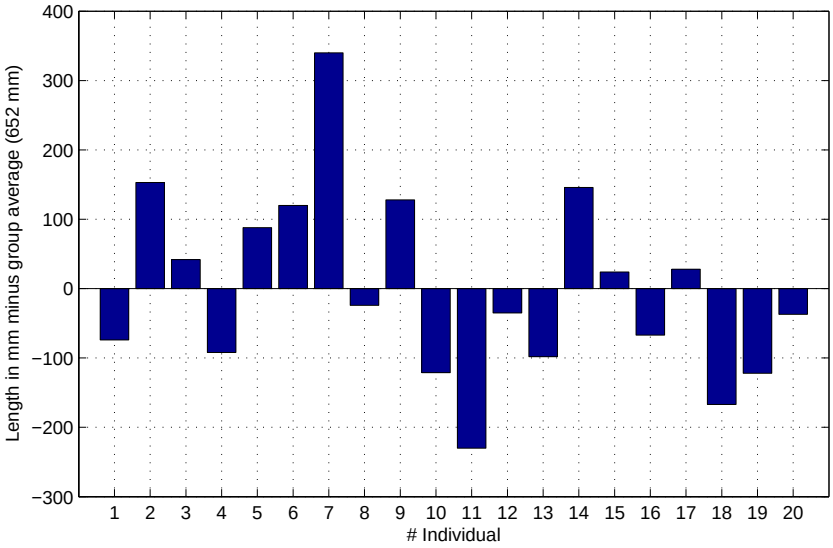

Figure 2: Length distribution of the tagged pike.

Estimation of the individuals can be parallelised to obtain further speed-up since they are conditional independent.

The estimation results for the simulation study are shown in Section A.1 and summarised in Table 2. All confidence intervals for the population parameters contained the true parameter values. The 95% confidence intervals for the individual parameters also behaved as expected (approximately 5% did not contain the true parameter values). The individual estimates of the advection parameters were relatively uncertain. The random effects therefore had a large influence on the updated estimates, i.e. the estimates of $\boldsymbol{u}_i$. That is, $\boldsymbol{u}_i$ were close to $\boldsymbol{u}$ in general. In contrast, the diffusivity estimates were only modified slightly by the random effects. Overall the estimation performance of the HMM with mixed effects was satisfactory.

4.2 Acoustic data from pike

Here we use the mixed effects framework to estimate the behaviour of $M = 20$ pike with length distribution as shown in Figure 2. Data are recorded using acoustic tags and hydrophones (listening stations) in a lake. Via triangulation, the location of the pike is measured. The location data are accurate, but prone to outliers. Therefore, data are pre-filtered with a robust SSM (using t-distributed observation noise). After filtering we assume that locations are known without error.

The aim of the study is to investigate the movement behaviour of the pike and to identify individuals that deviate from the rest of the population. Our approach is to set up a three-state HMM where each

state corresponds to either "resting", "cruising", or "aggressive". First, the location data is converted to speed data by differencing. This is only possible when the location error is small, otherwise the speed becomes uncertain. The speed data pertaining to individual i are denoted $\mathcal{Z}_{N_i}^{(i)} = \{\boldsymbol{z}_1^{(i)}, \ldots, \boldsymbol{z}_k^{(i)}, \ldots, \boldsymbol{z}_{N_i}^{(i)}\}$.

4.2.1 Estimation scheme

For each data point the likelihood of having one of the three behaviours can be computed using the following scheme:

1. Resting (no movement),
$$L_{1,k}^{(i)} = 1 - \Phi\left(\frac{z_k^{(i)} - \mu_1}{\sigma_1}\right),$$
where $\mu_1 = 0.025$ m/s and $\sigma_1 = 0.002$ m/s.

2. Cruising,
$$L_{2,k}^{(i)} = \Phi\left(\frac{z_k^{(i)} - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{z_k^{(i)} - \mu_3}{\sigma_3}\right),$$
where $\mu_2 = 0.03$ m/s, $\mu_3 = 0.25BL_i$ m/s, $\sigma_2 = 0.01$ m/s and $\sigma_3 = 0.02$ m/s. Here BL_i is the body length of individual i.

3. Aggressive,
$$L_{3,k}^{(i)} = \Phi\left(\frac{z_k^{(i)} - \mu_3}{\sigma_3}\right).$$

Here $\Phi(\cdot)$ is the cumulative density function of a standard Gaussian distributed random variable. The likelihood scheme is illustrated in Figure 3. The data likelihood (Thygesen et al., 2009; Zucchini and MacDonald, 2009) vector to be used in the HMM is then

$$\mathbf{L}_k^{(i)} = \text{diag}(L_{1,k}^{(i)}, L_{2,k}^{(i)}, L_{3,k}^{(i)}).$$

The data sampling interval was 45 seconds. However, with acoustic data many transmissions are lost so the resulting data are very unevenly sampled. It is therefore necessary to formulate the HMM in continuous time. Then the dynamics of the Markov process is described by its generator

$$\mathbf{G} = \begin{pmatrix} -\lambda_{12} - \lambda_{13} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -\lambda_{21} - \lambda_{23} & \lambda_{23} \\ \lambda_{31} & \lambda_{23} & -\lambda_{31} - \lambda_{32} \end{pmatrix},$$

where λ_{ab} is the rate of jumping from state a to state b.

We are also interested in if the fish display different behaviours at day and night so we setup an HMM for the (approximately) twelve hours of darkness and one for the twelve hours of daylight. This corresponds to considering time as a covariate with two levels (day and night). The generators pertaining to daytime and night time are $\mathbf{G}^d$ and $\mathbf{G}^n$ respectively. The probability transition matrices needed in the HMM iterations are $\boldsymbol{P}_k = \exp(\mathbf{G}\Delta_k)$, where $\Delta_k = t_{k+1} - t_k$. The parameter vectors of the model for individual i for day and night are respectively

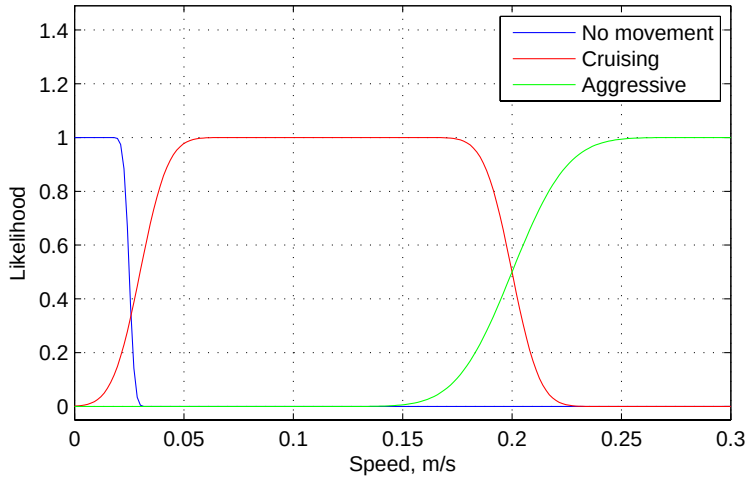

Figure 3: The likelihood of each of the three movement behaviours as a function of the observed speed.

$$\boldsymbol{\theta}_i^d = (\lambda_{12}, \lambda_{13}, \lambda_{21}, \lambda_{23}, \lambda_{31}, \lambda_{32})_i^{(d)}$$
$$\boldsymbol{\theta}_i^n = (\lambda_{12}, \lambda_{13}, \lambda_{21}, \lambda_{23}, \lambda_{31}, \lambda_{32})_i^{(n)}.$$

Thus the total parameter vector for individual i is $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_i^{(d)}, \boldsymbol{\theta}_i^{(n)})$.

The state probability distribution of the HMM at time t_k conditional on $\mathcal{Z}_k$ is $\boldsymbol{\phi}(t_k, \boldsymbol{x}_k|\mathcal{Z}_k) = \boldsymbol{\phi}_{k|k}$. This distribution is updated (omitting the i index) with

$$\boldsymbol{\phi}_{k+1|k+1} = \psi_k^{-1} \boldsymbol{\phi}_{k|k} \mathbf{P}_k \mathbf{L}_k, \tag{21}$$

where

$$\psi_k = [\boldsymbol{\phi}_{k|k} \mathbf{P}_k \mathbf{L}_k] \cdot \mathbf{1}_n,$$

where $\mathbf{1}_n$ is a column vector of ones of length n and '$\cdot$' is the dot-product. The likelihood of the HMM parameters (Zucchini and MacDonald, 2009) is then calculated using

$$p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}_i\right) = p\left(\boldsymbol{z}_1^{(i)}|\boldsymbol{\theta}_i\right) \prod_{k=2}^{N_i} \psi_k.$$

For a faster and more accurate likelihood estimation we also implement the recursion for calculating the gradient of the likelihood function (see Section A.2).

Now, the mixed effects procedure explained in Section 3 can be utilised to estimate population parameters and random effects for the transition rates. However, we are interested in the stationary distribution of the Markov chain rather than the state transition rates in the generators because these have a more intuitive interpretation (Patterson et al., 2009). Note, though, that time series are not stationary. Still, the stationary distributions can provide useful information on how the fish spent their time, but should not be used for prediction under different conditions.

The stationary distribution of a Markov chain is a function of the estimated transition rates. Specifically, the estimated stationary distribution is the vector $\widehat{\boldsymbol{\mu}}$, which fulfills

$$\widehat{\boldsymbol{\mu}}\widehat{\mathbf{G}} = \mathbf{0}.$$

Knowing the uncertainty of $\widehat{\mathbf{G}}$ (from the Hessian of the likelihood function), the uncertainty of $\widehat{\boldsymbol{\mu}}$ can be calculated with the delta method (Wasserman, 2005). Then, setting $\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\mu}}_i$ with estimated covariance matrix $\boldsymbol{\Sigma}_i$ found with the delta method, we perform mixed effects estimation on the stationary distributions for day and night with the scheme of Section 3.

For this application it is unrealistic to assume that the elements of the stationary distribution are uncorrelated. In other words $\boldsymbol{R} \neq \boldsymbol{I}$, but other than that the structure of $\boldsymbol{R}$ is unknown. Instead, the empirical estimate of the covariance matrix $\boldsymbol{W}$ is used. Specifically, the first three steps of the algorithm stated in Section 3 are performed using $\boldsymbol{I}$ as starting guess for $\boldsymbol{W}$. After step 3 the empirical estimate of $\boldsymbol{W}$ is computed from the residuals of the model. Thereafter $\boldsymbol{W}$ remains fixed to its empirical estimate. Then the fixed and random effects are estimated as before. While this scheme is somewhat ad hoc it does provide a much higher likelihood value than directly using the algorithm in Section 3.

4.2.2 Estimation results

The numbering and individual parameter estimates are shown in Section A.3. The backward-elimination procedure outlined in Section 2.1 was used for the $M = 20$ pike with the estimated parameters for day time and night time. For the day time parameters no deviating individuals were found. For the night time parameters individuals were eliminated in the following order: #7, #2, #14, #18, #11. None of these five individuals could be included in the remaining population by forward selection (see Section A.3).

It is important to note that the three largest fish and the two smallest were excluded from the group. This suggests that the size of a pike influences its behaviour, which seems plausible from a biological point of view. Further study of the excluded individuals and the remaining group is required to enable detailed biological conclusions about the pike population to be made.

5 Discussion

The modelling framework presented here is similar to the hierarchical Bayes approach presented in Jonsen et al. (2003) with (at least) two important differences: first, prior information about parameters is not required, and second, our framework allows the investigator to test if individuals deviate from the rest population using backward elimination and forward selection. A Bayesian alternative to the latter point has been investigated by Efron (1996) based on so-called parameter relevance. The technique requires a prior probability that an individual belongs to the population and then provides the posterior probability.

A limitation of our framework is that the individual log-likelihood functions must be approximatively quadratic. The degree to which this assumption holds has not been dealt with in depth here, instead the reader is referred to Vonesh (1996); Mortensen (2009). It is known, though, that the log-likelihood is asymptotically quadratic as the number of observations approach infinity, however the order of convergence is problem dependent.

Similar to previous individual based population models (Aarts et al., 2008) explanatory covariates can be incorporated into the model presented here. In the study of pike this was done simplistically by

letting parameters depend on time of day (night or day time). Naturally, functional links could also be used as in Bestley et al. (2008). The use of environmental covariates improves the model's ability to make predictions in other but similar environments. Furthermore, can inference based on covariates provide ecological insights into animal's usage of space and indicate possible behavioural responses to changes in the environmental variables.

As discussed by Aarts et al. (2008) the broader terms of the inference the higher the uncertainty of the results. Inference within the estimation dataset can be carried out with high confidence in the conclusions. Using the estimated model to predict behaviour for other populations of the same species in a similar environment seems reasonably safe also. Extrapolation, on the other hand, to different environmental properties, other species, different seasons etc. should only be carried out if this can be justified empirically. One should also be aware than even within seemingly similar environments unmodelled covariates may differ such as prey distribution, risk of predation or other influential information, which is unavailable to the modeller.

Explicit modelling of space use with data from electronic tags is difficult because data are temporally and spatially correlated. Ignoring correlation will possibly bias conclusions. On the other hand, the high temporal resolution of tagging data can, if correlation is accounted for, provide unique insights into behavioural responses of the animal. Archival tags are becoming increasingly advanced measuring not only temperature and depth, but also salinity, oxygen levels, magnetic field, and physiological variables such as visceral warming, and heart rate. These explanatory variables will become important for future studies of individual and population behaviour.

References

Aarts, G., M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos. 2008. Estimating space-use and habitat preference from wildlife telemetry data. Ecography **31**:140–160.

Bestley, S., T. Patterson, M. Hindell, and J. Gunn. 2008. Feeding ecology of wild migratory tunas revealed by archival tag records of visceral warming. Journal of Animal Ecology **77**:1223–1233.

Efron, B. 1996. Empirical Bayes Methods for Combining Likelihood. Journal of the American Statistical Association **91**:538–550.

Jonsen, I. D., R. A. Myers, and J. M. Flemming. 2003. Meta-analysis of animal movement using state-space models. Ecology **84**:3055–3065.

Jonsen, I. D., R. A. Myers, and M. C. James. 2006. Robust hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles. Journal of Animal Ecology **75**:1046–1057.

Mortensen, S. B., 2009. Markov and mixed models with applications. Ph.D. thesis, Technical University of Denmark (DTU), Kgs. Lyngby, Denmark.

Patterson, T., B. M., B. M.V., and J. Gunn. 2009. Classifying movement behaviour in relation to environmental conditions using hidden markov models. Journal of Animal Ecology **78**:1113–1123.

Pawitan, Y. 2001. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, USA.

Thygesen, U. H., M. W. Pedersen, and H. Madsen, 2009. Geolocating fish using hidden Markov models and data storage tags. Pages 277–293 *in* J. Nielsen, H. Arrizabalaga, N. Fragoso, A. Hobday, M. Lutcavage, and J. Sibert, editors. Tagging and Tracking of Marine Animals with Electronic Devices, volume 9 of *Reviews: Methods and Technologies in Fish Biology and Fisheries*. Springer.

Tornøe, C. 2005. IMM-PhD. 154. Population pharmacokinetic/pharmacodynamic modelling of the hypothalamic-pituitary-gonadal axis. IMM, Informatik og Matematisk Modellering, Danmarks Tekniske Universitet.

Vonesh, E. F. 1996. A note on the use of laplace's approximation for nonlinear mixed-effects models. Biometrika **83**:447–452.

Wasserman, L. 2005. All of statistics. Springer-Verlag.

Zucchini, W. and I. MacDonald. 2009. Hidden Markov Models for Time Series. Chapman & Hall/CRC, London.

A Appendix

A.1 Simulation results

```
sigma_b: 0.2727169 (0.300), niter: 86, likval: 72.4821775

- Pop D:  [ 8.85  9.87 11.02]
  # 1, w rand D: [10.24 12.76 15.91] (D indv: [10.20 13.08 16.77]) (true: 13.23)
  # 2, w rand D: [10.10 12.56 15.62] (D indv: [10.28 13.06 16.60]) (true: 11.37)
  # 3, w rand D: [ 7.44  9.46 12.03] (D indv: [ 7.16  9.37 12.26]) (true:  9.81)
  # 4, w rand D: [ 6.17  8.01 10.40] (D indv: [ 5.54  7.47 10.08]) (true:  7.27)
  # 5, w rand D: [ 5.38  7.02  9.17] (D indv: [ 4.62  6.28  8.55]) (true:  8.96)
  # 6, w rand D: [ 7.34  9.35 11.90] (D indv: [ 7.01  9.19 12.05]) (true:  9.65)
  # 7, w rand D: [ 8.71 10.90 13.64] (D indv: [ 8.43 10.81 13.86]) (true: 12.08)
  # 8, w rand D: [12.61 15.67 19.47] (D indv: [13.57 17.21 21.83]) (true: 15.25)
  # 9, w rand D: [10.17 12.58 15.57] (D indv: [10.43 13.16 16.60]) (true: 10.16)
  #10, w rand D: [ 6.21  7.98 10.27] (D indv: [ 5.65  7.51 10.00]) (true:  6.95)
  #11, w rand D: [ 7.22  9.13 11.53] (D indv: [ 6.92  8.98 11.66]) (true:  9.18)
  #12, w rand D: [14.40 17.76 21.90] (D indv: [15.72 19.75 24.80]) (true: 18.19)
  #13, w rand D: [ 6.33  8.12 10.42] (D indv: [ 5.80  7.69 10.19]) (true:  7.34)
  #14, w rand D: [ 6.37  8.11 10.34] (D indv: [ 5.88  7.71 10.12]) (true:  8.09)
  #15, w rand D: [ 6.62  8.55 11.04] (D indv: [ 6.12  8.20 10.98]) (true: 10.32)
  #16, w rand D: [11.67 14.35 17.65] (D indv: [12.09 15.14 18.95]) (true: 14.17)
  #17, w rand D: [ 7.47  9.67 12.52] (D indv: [ 7.16  9.62 12.93]) (true: 11.07)
  #18, w rand D: [ 8.82 11.08 13.90] (D indv: [ 8.75 11.27 14.52]) (true: 10.92)
  #19, w rand D: [ 7.77 10.02 12.93] (D indv: [ 7.52 10.05 13.43]) (true:  9.24)
  #20, w rand D: [ 7.24  9.22 11.75] (D indv: [ 6.90  9.06 11.88]) (true:  8.08)
  #21, w rand D: [ 7.02  9.04 11.64] (D indv: [ 6.61  8.81 11.73]) (true:  9.72)
  #22, w rand D: [ 6.97  8.89 11.33] (D indv: [ 6.60  8.67 11.38]) (true:  8.47)
  #23, w rand D: [ 7.89 10.01 12.70] (D indv: [ 7.80 10.19 13.31]) (true:  7.65)
  #24, w rand D: [10.61 13.05 16.06] (D indv: [10.94 13.70 17.16]) (true: 12.41)
  #25, w rand D: [ 5.72  7.58 10.05] (D indv: [ 4.92  6.86  9.55]) (true:  7.08)
  #26, w rand D: [ 7.97 10.23 13.14] (D indv: [ 7.78 10.33 13.70]) (true: 10.30)
  #27, w rand D: [ 6.69  8.49 10.78] (D indv: [ 6.25  8.16 10.66]) (true:  8.23)
  #28, w rand D: [ 6.08  7.88 10.21] (D indv: [ 5.47  7.35  9.88]) (true:  8.52)
  #29, w rand D: [ 4.24  5.67  7.58] (D indv: [ 3.17  4.48  6.34]) (true:  6.04)
  #30, w rand D: [10.23 12.51 15.30] (D indv: [10.45 12.99 16.15]) (true: 13.31)

- Pop Ux:  [ 0.35  0.80  1.25]
  # 1, w rand Ux: [ 0.37  0.89  1.42] (Ux indv: [ 0.48  3.28  6.08]) (true:  1.54)
  # 2, w rand Ux: [ 0.32  0.85  1.37] (Ux indv: [-0.86  2.24  5.34]) (true:  1.02)
  # 3, w rand Ux: [ 0.27  0.79  1.31] (Ux indv: [-1.97  0.61  3.19]) (true:  1.11)
  # 4, w rand Ux: [ 0.16  0.68  1.21] (Ux indv: [-4.16 -1.76  0.64]) (true:  0.59)
  # 5, w rand Ux: [ 0.30  0.82  1.34] (Ux indv: [-0.76  1.23  3.22]) (true:  0.97)
  # 6, w rand Ux: [ 0.24  0.75  1.27] (Ux indv: [-2.05 -0.02  2.01]) (true:  0.76)
  # 7, w rand Ux: [ 0.36  0.88  1.40] (Ux indv: [ 0.57  3.09  5.61]) (true:  0.11)
  # 8, w rand Ux: [ 0.32  0.85  1.37] (Ux indv: [-0.71  2.63  5.96]) (true:  0.93)
  # 9, w rand Ux: [ 0.25  0.76  1.28] (Ux indv: [-1.68  0.31  2.30]) (true:  1.01)
```

```
  #10, w rand Ux: [ 0.33  0.85  1.37] (Ux indv: [-0.41  1.92  4.26]) (true:  0.90)
  #11, w rand Ux: [ 0.33  0.85  1.37] (Ux indv: [-0.01  2.31  4.63]) (true:  1.39)
  #12, w rand Ux: [ 0.23  0.76  1.29] (Ux indv: [-5.02 -1.27  2.48]) (true:  1.20)
  #13, w rand Ux: [ 0.24  0.76  1.28] (Ux indv: [-1.76  0.20  2.16]) (true:  1.06)
  #14, w rand Ux: [ 0.32  0.85  1.37] (Ux indv: [-0.67  1.78  4.23]) (true:  0.83)
  #15, w rand Ux: [ 0.21  0.73  1.25] (Ux indv: [-2.68 -0.59  1.50]) (true:  1.00)
  #16, w rand Ux: [ 0.32  0.85  1.38] (Ux indv: [-0.58  2.89  6.37]) (true:  0.99)
  #17, w rand Ux: [ 0.30  0.83  1.35] (Ux indv: [-1.13  1.52  4.18]) (true:  1.36)
  #18, w rand Ux: [ 0.23  0.75  1.28] (Ux indv: [-3.75 -0.78  2.20]) (true:  0.82)
  #19, w rand Ux: [ 0.27  0.79  1.31] (Ux indv: [-1.34  0.86  3.05]) (true:  0.79)
  #20, w rand Ux: [ 0.35  0.87  1.40] (Ux indv: [ 0.11  2.71  5.31]) (true:  0.96)
  #21, w rand Ux: [ 0.29  0.81  1.33] (Ux indv: [-1.44  1.02  3.47]) (true:  0.47)
  #22, w rand Ux: [ 0.21  0.73  1.25] (Ux indv: [-3.07 -0.71  1.65]) (true:  0.28)
  #23, w rand Ux: [ 0.34  0.86  1.39] (Ux indv: [-0.38  2.46  5.30]) (true:  1.06)
  #24, w rand Ux: [ 0.21  0.73  1.26] (Ux indv: [-5.10 -1.84  1.43]) (true:  0.93)
  #25, w rand Ux: [ 0.30  0.82  1.34] (Ux indv: [-0.88  1.28  3.43]) (true:  0.88)
  #26, w rand Ux: [ 0.31  0.83  1.35] (Ux indv: [-0.79  1.70  4.19]) (true:  1.15)
  #27, w rand Ux: [ 0.28  0.80  1.32] (Ux indv: [-1.27  0.99  3.24]) (true:  0.96)
  #28, w rand Ux: [ 0.26  0.78  1.30] (Ux indv: [-1.61  0.59  2.80]) (true:  1.22)
  #29, w rand Ux: [ 0.26  0.77  1.28] (Ux indv: [-1.53  0.36  2.25]) (true:  1.02)
  #30, w rand Ux: [ 0.24  0.76  1.29] (Ux indv: [-3.80 -0.64  2.52]) (true:  1.17)

- Pop Uy:  [-0.38  0.06  0.50]
  # 1, w rand Uy: [-0.46  0.06  0.59] (Uy indv: [-3.27 -0.45  2.38]) (true:  0.15)
  # 2, w rand Uy: [-0.49  0.04  0.57] (Uy indv: [-3.75 -0.59  2.58]) (true: -0.12)
  # 3, w rand Uy: [-0.44  0.08  0.60] (Uy indv: [-1.70  0.36  2.42]) (true: -0.24)
  # 4, w rand Uy: [-0.42  0.09  0.61] (Uy indv: [-1.24  0.75  2.74]) (true:  0.08)
  # 5, w rand Uy: [-0.60 -0.08  0.43] (Uy indv: [-4.71 -2.52 -0.33]) (true: -0.11)
  # 6, w rand Uy: [-0.49  0.04  0.56] (Uy indv: [-3.30 -0.68  1.94]) (true: -0.45)
  # 7, w rand Uy: [-0.38  0.14  0.66] (Uy indv: [-0.11  2.20  4.51]) (true:  0.09)
  # 8, w rand Uy: [-0.44  0.09  0.62] (Uy indv: [-2.01  1.64  5.30]) (true:  0.18)
  # 9, w rand Uy: [-0.49  0.04  0.57] (Uy indv: [-3.81 -0.62  2.57]) (true: -0.20)
  #10, w rand Uy: [-0.41  0.10  0.62] (Uy indv: [-1.21  0.82  2.86]) (true: -0.13)
  #11, w rand Uy: [-0.37  0.15  0.67] (Uy indv: [-0.22  2.05  4.32]) (true:  0.09)
  #12, w rand Uy: [-0.39  0.14  0.67] (Uy indv: [ 0.41  4.25  8.09]) (true: -0.11)
  #13, w rand Uy: [-0.45  0.07  0.59] (Uy indv: [-2.20  0.25  2.69]) (true: -0.11)
  #14, w rand Uy: [-0.42  0.09  0.61] (Uy indv: [-1.42  0.54  2.50]) (true:  0.09)
  #15, w rand Uy: [-0.52 -0.01  0.51] (Uy indv: [-3.61 -1.41  0.78]) (true: -0.06)
  #16, w rand Uy: [-0.56 -0.04  0.49] (Uy indv: [-7.22 -3.84 -0.46]) (true: -0.45)
  #17, w rand Uy: [-0.42  0.09  0.61] (Uy indv: [-1.40  0.67  2.73]) (true: -0.03)
  #18, w rand Uy: [-0.44  0.09  0.61] (Uy indv: [-1.88  0.78  3.44]) (true:  0.05)
  #19, w rand Uy: [-0.51  0.01  0.54] (Uy indv: [-3.66 -1.03  1.60]) (true: -0.02)
  #20, w rand Uy: [-0.51  0.00  0.52] (Uy indv: [-3.10 -1.05  1.00]) (true: -0.29)
  #21, w rand Uy: [-0.48  0.04  0.56] (Uy indv: [-2.45 -0.29  1.87]) (true:  0.11)
  #22, w rand Uy: [-0.47  0.04  0.56] (Uy indv: [-2.82 -0.57  1.67]) (true: -0.34)
  #23, w rand Uy: [-0.41  0.11  0.63] (Uy indv: [-1.34  1.13  3.60]) (true:  0.52)
  #24, w rand Uy: [-0.49  0.04  0.57] (Uy indv: [-3.66 -0.53  2.59]) (true:  0.16)
  #25, w rand Uy: [-0.39  0.13  0.65] (Uy indv: [-0.88  1.30  3.48]) (true: -0.09)
```

```
#26, w rand Uy: [-0.51  0.01  0.53] (Uy indv: [-3.45 -1.09  1.27]) (true: -0.36)
#27, w rand Uy: [-0.49  0.02  0.54] (Uy indv: [-2.87 -0.61  1.64]) (true: -0.34)
#28, w rand Uy: [-0.43  0.09  0.61] (Uy indv: [-1.62  0.58  2.77]) (true:  0.24)
#29, w rand Uy: [-0.43  0.08  0.60] (Uy indv: [-1.53  0.40  2.34]) (true:  0.11)
#30, w rand Uy: [-0.52  0.01  0.54] (Uy indv: [-4.85 -1.69  1.48]) (true: -0.78)
```

A.2 Gradient of likelihood function for HMMs

The optimum ($\widehat{\boldsymbol{\theta}}$) found by a numerical optimising routine is only a value close to the true optimum $\boldsymbol{\theta}$, that is

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta} + e,$$

where e is the approximation error. The size of e depends on the termination criteria for the optimising routine. The curvature of the likelihood function around $\boldsymbol{\theta}$ is approximated by the Hessian calculated around $\widehat{\boldsymbol{\theta}}$. For some problems the approximation of the Hessian is quite sensitive to the point around which it is calculated.

For likelihood estimation it is common to optimise the likelihood function only using evaluations of the function itself. However, in some cases it is possible to calculate the gradient of the likelihood function analytically and provide this as input to the optimiser along with the function value. This will typically lead to a faster and more accurate estimation of the optimum and therefore also a more accurate Hessian estimate. Below, the recursions for calculating the likelihood value and its gradient with respect to the model parameters are derived.

The parameter vector for individual i is $\boldsymbol{\theta}_i = \{\theta_1, \ldots, \theta_{n_{par}}\}_i$. Define the short-hand notation

$$\psi_k = p\left(\boldsymbol{z}_k^{(i)} | \mathcal{Z}_{k-1}^{(i)}, \boldsymbol{\theta}_i\right),$$

for $k > 1$. The gradient of the likelihood function (3) with respect to θ_j is

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta}_i)}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j}\left[\log \psi_1 + \sum_{k=2}^{N_i} \log \psi_k\right] \\ &= \frac{1}{\psi_1}\frac{\partial \psi_1}{\partial \theta_j} + \sum_{k=2}^{N_i} \frac{1}{\psi_k}\frac{\partial \psi_k}{\partial \theta_j},\end{aligned} \tag{22}$$

where $\psi_1 = p(\boldsymbol{z}_1^{(i)}|\boldsymbol{\theta}_i)$. The way to compute $\frac{\partial \psi_k}{\partial \theta_j}$ is through a recursion similar to that for computing the likelihood value itself. For a continuous-time Markov chain the following relation holds

$$\dot{\boldsymbol{\phi}}_{k|k} = \boldsymbol{\phi}_{k|k}\mathbf{G}_k, \tag{23}$$

where $\dot{\boldsymbol{\phi}}_{k|k} = \frac{\partial \boldsymbol{\phi}_{k|k}}{\partial t}$. Taking the partial derivative of (23) with respect to θ_j gives

$$\frac{\partial \dot{\boldsymbol{\phi}}_{k|k}}{\partial \theta_j} = \frac{\partial \boldsymbol{\phi}_{k|k}}{\partial \theta_j}\mathbf{G}_k + \boldsymbol{\phi}_{k|k}\frac{\partial \mathbf{G}_k}{\partial \theta_j}.$$

Define the derivative of the state probabilities

$$\boldsymbol{s}_k = \frac{\partial \boldsymbol{\phi}_{k|k}}{\partial \theta_j}$$

and concatenate $\boldsymbol{\phi}_{k|k}$ and $\boldsymbol{s}_k$ to get

$$\boldsymbol{\pi}_k = \left(\boldsymbol{\phi}_{k|k}\,,\,\boldsymbol{s}_k\right).$$

The system of differential equations analogous to (23), but including $\boldsymbol{s}_k$ is then

$$\dot{\boldsymbol{\pi}}_k = \boldsymbol{\pi}_k \boldsymbol{\Gamma}_k,$$

where

$$\boldsymbol{\Gamma}_k = \begin{pmatrix} \mathbf{G}_k & \frac{\partial \mathbf{G}_k}{\partial \theta_j} \\ \mathbf{0} & \mathbf{G}_k \end{pmatrix}.$$

The matrix $\boldsymbol{\Gamma}_k$ is the generator for the augmented system comprising both $\boldsymbol{\phi}_{k|k}$ and $\boldsymbol{s}_k$. Then the usual relation holds

$$\boldsymbol{\Pi}_k = \exp\left(\boldsymbol{\Gamma}_k \Delta_k\right), \tag{24}$$

where $\boldsymbol{\Pi}_k$ is the transition matrix for $\boldsymbol{\pi}_k$. Thus, the time-evolution of the state probabilities ($\boldsymbol{\phi}_{k|k}$) and the state probability derivatives ($\boldsymbol{s}_k$) is described by $\boldsymbol{\Pi}_k$. This matrix is not a transition probability matrix because it can have element values below zero and larger than one.

As for the standard HMM filter (21), time and data-updates of $\boldsymbol{\pi}_k$ are performed analogously

$$\boldsymbol{\mu}_k = \boldsymbol{\pi}_k \boldsymbol{\Pi}_k \boldsymbol{\Lambda}_k, \tag{25}$$

where $\boldsymbol{\Lambda}_k$ is the concatenated data likelihood matrix, i.e.

$$\boldsymbol{\Lambda}_k = \begin{pmatrix} \mathbf{L}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_k \end{pmatrix}.$$

Note that $\boldsymbol{\mu}_k$ has not yet been normalised. The normalisation constants for $\boldsymbol{\mu}_k$ are

$$\left(\psi_k\,,\,\frac{\partial \psi_k}{\partial \boldsymbol{\theta}_j}\right) = \boldsymbol{\mu}_k \begin{pmatrix} \mathbf{1}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_n \end{pmatrix},$$

which are the ones required to calculate the sum (22). To complete the recursion the normalisation of $\boldsymbol{\mu}_k$ is given by

$$\boldsymbol{\pi}_{k+1} = \boldsymbol{\mu}_k \boldsymbol{\Psi}_k, \tag{26}$$

where

$$\boldsymbol{\Psi}_k = \begin{pmatrix} \psi_k^{-1}\mathbf{1}_n & -\frac{1}{\psi_k^2}\frac{\partial \psi_k}{\partial \theta_j}\mathbf{1}_n \\ \mathbf{0} & \psi_k^{-1}\mathbf{1}_n \end{pmatrix}.$$

The matrix $\boldsymbol{\Psi}_k$ is found using the rules for differentiation of a fraction.

The steps of the filter recursion are summarised by (25) and (26). The main concern with the recursion is (24) which can be a computationally demanding operation depending on the size of $\mathbf{\Gamma}_k$.

A similar recursive scheme can also be derived for the Hessian of the likelihood function.

A.3 Individual estimates of pike data

Below StatDay and StatNight refer to the stationary distribution for the day and night time periods respectively.

```
  # 1, Length: 578 mm, StatDay: [ 0.897  0.096  0.007], StatNight: [ 0.994  0.005  0.001]
  # 2, Length: 805 mm, StatDay: [ 0.794  0.202  0.004], StatNight: [ 0.691  0.306  0.003]
  # 3, Length: 694 mm, StatDay: [ 0.731  0.261  0.008], StatNight: [ 0.981  0.018  0.001]
  # 4, Length: 560 mm, StatDay: [ 0.702  0.269  0.029], StatNight: [ 0.948  0.047  0.006]
  # 5, Length: 740 mm, StatDay: [ 0.857  0.137  0.006], StatNight: [ 0.980  0.017  0.002]
  # 6, Length: 772 mm, StatDay: [ 0.918  0.079  0.003], StatNight: [ 0.987  0.012  0.001]
  # 7, Length: 992 mm, StatDay: [ 0.732  0.265  0.004], StatNight: [ 0.768  0.230  0.002]
  # 8, Length: 628 mm, StatDay: [ 0.837  0.158  0.005], StatNight: [ 0.990  0.009  0.001]
  # 9, Length: 780 mm, StatDay: [ 0.815  0.183  0.002], StatNight: [ 0.968  0.030  0.002]
  #10, Length: 531 mm, StatDay: [ 0.698  0.278  0.023], StatNight: [ 0.958  0.036  0.005]
  #11, Length: 422 mm, StatDay: [ 0.852  0.121  0.028], StatNight: [ 0.970  0.022  0.008]
  #12, Length: 617 mm, StatDay: [ 0.720  0.267  0.014], StatNight: [ 0.981  0.017  0.003]
  #13, Length: 554 mm, StatDay: [ 0.837  0.152  0.011], StatNight: [ 0.977  0.021  0.002]
  #14, Length: 798 mm, StatDay: [ 0.747  0.248  0.005], StatNight: [ 0.926  0.073  0.001]
  #15, Length: 676 mm, StatDay: [ 0.894  0.101  0.006], StatNight: [ 0.986  0.013  0.001]
  #16, Length: 585 mm, StatDay: [ 0.679  0.311  0.010], StatNight: [ 0.989  0.010  0.002]
  #17, Length: 680 mm, StatDay: [ 0.876  0.121  0.003], StatNight: [ 0.995  0.004  0.001]
  #18, Length: 485 mm, StatDay: [ 0.923  0.064  0.013], StatNight: [ 0.987  0.009  0.004]
  #19, Length: 530 mm, StatDay: [ 0.884  0.106  0.010], StatNight: [ 0.987  0.011  0.002]
  #20, Length: 615 mm, StatDay: [ 0.933  0.062  0.005], StatNight: [ 0.992  0.008  0.001]
```

Day time population estimates:

```
theta = [ 0.8259  0.1638  0.0078]
        [ 0.2569 -0.2675 -0.0107]
    W = [-0.2675  0.2812 -0.0194] (in logit domain)
        [-0.0107 -0.0194  0.4353]
  # 1:  [ 0.8212  0.1682  0.0078], L: 578, p-val: 0.638035
  # 2:  [ 0.8268  0.1624  0.0082], L: 805, p-val: 0.685034
  # 3:  [ 0.8312  0.1584  0.0077], L: 694, p-val: 0.725120
  # 4:  [ 0.8270  0.1632  0.0073], L: 560, p-val: 0.284323
  # 5:  [ 0.8236  0.1657  0.0079], L: 740, p-val: 0.916625
  # 6:  [ 0.8203  0.1688  0.0080], L: 772, p-val: 0.350415
  # 7:  [ 0.8290  0.1603  0.0083], L: 992, p-val: 0.347690
  # 8:  [ 0.8255  0.1638  0.0080], L: 628, p-val: 0.941316
  # 9:  [ 0.8298  0.1595  0.0085], L: 780, p-val: 0.073640
  #10:  [ 0.8329  0.1575  0.0072], L: 531, p-val: 0.088945
  #11:  [ 0.8256  0.1650  0.0073], L: 422, p-val: 0.213921
  #12:  [ 0.8318  0.1581  0.0075], L: 617, p-val: 0.506483
  #13:  [ 0.8253  0.1643  0.0076], L: 554, p-val: 0.919697
```

```
  #14:  [ 0.8303  0.1592  0.0080], L: 798, p-val: 0.653147
  #15:  [ 0.8244  0.1649  0.0079], L: 676, p-val: 0.870831
  #16:  [ 0.8338  0.1560  0.0077], L: 585, p-val: 0.375426
  #17:  [ 0.8217  0.1673  0.0081], L: 680, p-val: 0.570583
  #18:  [ 0.8182  0.1720  0.0076], L: 485, p-val: 0.165754
  #19:  [ 0.8217  0.1678  0.0077], L: 530, p-val: 0.705593
  #20:  [ 0.8155  0.1747  0.0079], L: 615, p-val: 0.152295
```

Night time population estimates (with indviduals #7, #2, #14, #18, and #11 excluded):

```
theta = [ 0.9844  0.0139  0.0016]
        [ 0.4468 -0.4565 -0.3659]
    W = [-0.4565  0.4681  0.3582] (in logit domain)
        [-0.3659  0.3582  0.4425]
  # 1:  [ 0.9834  0.0149  0.0017], L: 578, p-val: 0.486218
  # 3:  [ 0.9846  0.0137  0.0016], L: 694, p-val: 0.357374
  # 4:  [ 0.9857  0.0128  0.0014], L: 560, p-val: 0.095166
  # 5:  [ 0.9846  0.0137  0.0015], L: 740, p-val: 0.897195
  # 6:  [ 0.9841  0.0141  0.0017], L: 772, p-val: 0.569789
  # 8:  [ 0.9839  0.0144  0.0016], L: 628, p-val: 0.746050
  # 9:  [ 0.9852  0.0132  0.0015], L: 780, p-val: 0.494091
  #10:  [ 0.9855  0.0130  0.0015], L: 531, p-val: 0.137100
  #12:  [ 0.9846  0.0138  0.0015], L: 617, p-val: 0.802857
  #13:  [ 0.9848  0.0135  0.0016], L: 554, p-val: 0.844332
  #15:  [ 0.9842  0.0140  0.0017], L: 676, p-val: 0.649057
  #16:  [ 0.9840  0.0143  0.0016], L: 585, p-val: 0.781717
  #17:  [ 0.9830  0.0153  0.0017], L: 680, p-val: 0.157704
  #19:  [ 0.9841  0.0142  0.0016], L: 530, p-val: 0.899792
  #20:  [ 0.9836  0.0145  0.0017], L: 615, p-val: 0.286372

- Forward selection
  # 2:  [ 0.6908  0.3058  0.0034], L: 805, p-val: 0.000000 *
  # 7:  [ 0.7684  0.2295  0.0020], L: 992, p-val: 0.000000 *
  #11:  [ 0.9700  0.0223  0.0077], L: 422, p-val: 0.000262 *
  #14:  [ 0.9257  0.0730  0.0013], L: 798, p-val: 0.000001 *
  #18:  [ 0.9871  0.0091  0.0039], L: 485, p-val: 0.000641 *
```