Hidden Markov modelling of movement data from fish

Martin Wæver Pedersen

Kongens Lyngby 2010
IMM-PHD-2010-70

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

IMM-PHD: ISSN 0909-3192

Preface

This thesis was prepared at the Technical University of Denmark (DTU), department for Informatics and Mathematical Modelling (DTU Informatics) in partial fulfillment of the requirements for acquiring the Ph.D. degree in engineering.

The work was carried out in the time period from July 2007 to September 2010 separated by a three months leave to see if there is a world outside the four office walls. The project was a collaboration between DTU Informatics and the National Institute for Aquatic Resources (DTU Aqua). The work was primarily funded by a DTU scholarship with additional travel funds provided by Otto Mønsted's foundation, the Oticon foundation and CSIRO Marine and Atmospheric Research (CMAR).

The topic of the thesis is hidden Markov modelling of movement data from fish. The thesis is comprised of a summary report and six scientific research papers written during the course of the Ph.D. Three papers are published in international journals, two papers are in review at international journals, and one paper is published as a research report at DTU Informatics.

At this point a couple of 'thank you's is befitting: to my supervisory dream duo Henrik Madsen and Uffe H. Thygesen who almost surely are the best of their kind, to Toby A. Patterson for immaculate hosting of my visit to CMAR, to people from DTU, CMAR, and Cefas, for their efforts in collecting data and being a source of inspiration, and finally to all the nice people in Tasmania and Denmark who continuously reminded me that there indeed is a world outside the four office walls.

Kgs. Lyngby, September 2010

Martin Wæver Pedersen

Abstract

Movement data from marine animals tagged with electronic tags are becoming increasingly diverse and plentiful. This trend entails a need for statistical methods that are able to filter the observations to extract the ecologically relevant content. This dissertation focuses on the development and application of hidden Markov models (HMMs) for analysis of movement data from fish. The main contributions are represented by six scientific publications.

Estimation of animal location from uncertain and possibly indirect observations is the starting point of most movement data analyses. In this work a discrete state HMM is employed to deal with this task. Specifically, the continuous horizontal plane is discretised into grid cells, which enables a state-space model for the geographical location to be estimated on this grid.

The estimation model for location is extended with an additional state representing the behaviour of the animal. With the extended model can migratory and resident movement behaviour be related to geographical regions. For population inference multiple individual state-space analyses can be interconnected using mixed effects modelling. This framework provides parameter estimates at the population level and allows ecologists to identify individuals that deviate from the rest of the tagged population.

The thesis also deals with geolocation on state-spaces with complicated geometries. Using an unstructured discretisation and the finite element method tortuous shore line geometries are closely approximated. This furthermore enables accurate probability densities of location to be computed.

Finally, the performance of the HMM approach in analysing nonlinear state-

space models is compared with two alternatives: the AD Model Builder framework and BUGS, which relies on Markov chain Monte Carlo estimation.

Key words: Atlantic cod, behaviour switching, electronic tags, hidden Markov models, movement data, nonlinear mixed effects models, nonlinear state-space models, southern bluefin tuna, stochastic differential equations.

Resumé

Bevægelsesdata fra marine dyr mærket med elektroniske mærker bliver til stadighed mere forskelligartet og rigelig i mængde. Denne tendens medfører et behov for statistiske metoder, som er i stand til at filtrere observationerne for at udtrække det økologisk relevante indhold. Denne afhandling fokuserer på at udvikle og anvende skjulte Markov modeller (SMM) til at analysere bevægelsesdata fra fisk. De primære bidrag er repræsenteret af seks videnskabelige publikationer.

Estimation af dyrets lokation fra usikre og muligvis indirekte observationer er udgangspunktet for de fleste bevægelsesdataanalyser. I denne tese anvendes en diskret-tilstand SMM til at håndtere denne opgave. Mere specifikt inddeles det kontinuerte horisontale plan i et antal af celler, hvilket muliggør estimation af en tilstandsmodel for den geografiske lokationen på dette diskrete net.

Estimationsmodellen for lokationen udvides med en tilstand yderligere, der repræsenterer dyrets adfærd. Med den udvidede model kan migratorisk og residerende adfærd relateres til geografiske områder. Angående populationsinferens kan miksed effekt modellering anvendes til at sammenkæde et antal individuelle tilstandsanalyser. Denne tilgang giver parameterestimater på populationsniveau og tillader økologer at identificere individer, som afviger fra den resterende mærkede population.

Tesen omhandler også geolokalisering på tilstandsrum med komplicerede geometrier. Ved at anvende en ustruktureret diskretisering og finit element metoden kan snørklede kyststrækningsgeometrier approksimeres tæt. Dette muliggør ydermere beregning af nøjagtige sandsynlighedstætheder.

Slutteligt sammenlignes ydeevnen af SMM tilgangen til at analysere ikke-lineære tilstandsmodeller med to alternativer: AD Model Builder og BUGS, som bygger på Markov chain Monte Carlo estimation.

Nøgleord: Atlantisk torsk, adfærdsskift, elektroniske mærker, skjulte Markov modeller, bevægelsesdata, ikke-lineære miksede effekt-modeller, ikke-lineære tilstandsmodeller, sydlig blåfinnet tun, stokastiske differentialligninger.

List of publications

The thesis is based on the following scientific research papers.

A. **Pedersen, M.W.**, Righton, D., Thygesen, U.H., Andersen, K.H., and Madsen, H. Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching. Canadian Journal of Fisheries and Aquatic Sciences, **65**: 2367-2377 (2008).

B. Thygesen, U.H., **Pedersen, M.W.**, and Madsen, H. Geolocating Fish Using Hidden Markov Models and Data Storage Tags. *In:* Tagging and Tracking of Maring Animals with Electronic Devices, Reviews: Methods and Technologies in Fish Biology and Fisheries, **9**: 277-293 (2009). *Editors*: Nielsen, J.L., Arrizabalaga, H., Fragoso, N., Hobday, A., Lutcavage, M., and Sibert, J. Springer.

C. **Pedersen, M.W.**, Thygesen, U.H., and Madsen, H. Nonlinear tracking in a diffusion process with a Bayesian filter and the finite element method. Computational Statistics and Data Analysis, **55**: 280-290 (2011).

D. **Pedersen, M.W.**, Patterson, T.A., Thygesen, U.H., and Madsen, H. Estimating animal behaviour and residency from movement data. Oikos. Submitted (June 2010).

E. **Pedersen, M.W.**, Berg, C.W., Thygesen, U.H., and Madsen, H. Estimation methods for nonlinear state-space models in ecology. Ecological modelling. Submitted (September 2010).

F. **Pedersen, M.W.**, Thygesen, U.H., and Madsen, H. Individual based population inference using tagging data. IMM-Technical Report-2010-11. Published (September 2010).

Below is a list of other publications, which present applications of the methodology described in this thesis. These papers do not contain new methodological content and will not be addressed further.

- Thygesen, U.H., Karlsen, J., Nielsen, A., and **Pedersen, M.W.**. 2008. Indirect observation of fish movements: a general methodology applied at different scales. ICES Council Meeting (ISSN: 1015-4744), vol: P:12, pages: 1-7, International Council for the Exploration of the Sea. Presented at: ICES Annual Science Conference (ICES ASC), 2008. Halifax. Published.

- Berx, B., Neat, F., **Pedersen, M.W.**, Wright, P., and Proctor, R. 2008. Increasing confidence in tidal-based geolocation models by including temperature information: the case of cod in the northern North Sea. ICES Council Meeting (ISSN: 1015-4744), vol: P:05, pages: 1-3, International Council for Exploration of the Sea. Presented at: ICES Annual Science Conference (ICES ASC), 2008, Halifax. Published.

- Righton, D., Quayle, V., Neat, F., **Pedersen, M.W.**, Wright, P., Armstrong, M., Svedang, H., Hobson, V., and Metcalfe, J. 2008. Spatial dynamics of Atlantic cod (*Gadus morhua*) in the North Sea: results from a large-scale electronic tagging programme. ICES Council Meeting (ISSN: 1015-4744), vol: P:04, pages: 1-32, International Council for Exploration of the Sea. Presented at: ICES Annual Science Conference (ICES ASC), 2008, Halifax, Canada. Published

- **Pedersen, M.W.**, Patterson, T.A., and U.H. Thygesen. 2009. Estimating spatially and temporally dependent behaviour switching of tagged marine animals. ICES Council Meeting (ISBN-978-87-7482-079-6), vol: B:03, pages: 1-9, International Council for the Exploration of the Sea. Presented at: ICES Annual Science Conference (ICES ASC), 2009. Berlin. Published.

- Baktoft, H., Aarestrup, K., Berg, S., Boel, M., Jacobsen, L., Koed, A., **Pedersen, M.W.**, Svendsen, J.C., and Skov, C. 2010. Effects of angling and manual handling on pike behaviour investigated by high resolution positional telemetry. Proceedings of the 8th Conference on Fish Telemetry, Umeaa. In review.

Abbreviations and symbols

Abbreviations

ADMB	Automatic Differentiation Model Builder
BUGS	Bayesian inference using Gibbs sampling
DST	Data storage tag
FD	Finite difference
FEM	Finite element method
HMM	Hidden Markov model
iid.	Independent and identically distributed
KF	Kalman filter
ML	Maximum likelihood
MCMC	Markov Chain Monte Carlo
PDE	Partial differential equation
PF	Particle filter
PSAT	Pop-up satellite archival tag
RD	Residency distribution
SDE	Stochastic differential equation
SMC	Sequential Monte Carlo
SSM	State-space model
SST	Sea surface temperature

Symbols

$\cdot$	Dot-product between two vectors.
$\odot$	Element-wise product between vectors or matrices of same size.
$\oslash$	Element-wise division between vectors or matrices of same size.
Δ_k	Time interval between two observations, $\Delta_k = t_{k+1} - t_k$.
$\mathbf{D}$	Diffusivity of location.
e, $\boldsymbol{e}$	Observation noise.
ϵ, $\boldsymbol{\epsilon}$	Process noise.
Φ	Full posterior distribution for states.
$\boldsymbol{\phi}_k$	Vector with state probabilities at time t_k.
ϕ	Elements of $\boldsymbol{\phi}$.
$\mathbf{G}$	Generator matrix of a continuous-time Markov chain.
I_k	Behavioural state.
$\mathbf{L}$	Data likelihood vector.
λ_{ij}	Rate for jumping from i to j.
$N(\mu, \sigma^2)$	Gaussian (normal) distribution with mean μ and variance σ^2.
Ω	State-space.
$\mathbf{P}_k$	Probability transition matrix related to Δ_k.
$S_{\Delta t}$	Expected step length over the time step Δt.
t_k	The k'th time point in data.
$\boldsymbol{\theta}$	Parameter vector.
$\boldsymbol{u}$	Advection vector.
$\boldsymbol{w}_i$	Random effects related to individual i.
$\mathbf{W}$	Covariance of random effects.
$\boldsymbol{X}$	Vector containing longitudinal and latitudinal location coordinates.
$\boldsymbol{X}^T$	Transpose of $\boldsymbol{X}$.
$\mathcal{Z}_k$	All data taken at or before time t_k.
$\boldsymbol{z}_k$	Observed data at t_k.

Contents

Chapter 1

Introduction

1.1 Background

The large scale movement[1] of marine species is difficult to monitor *in situ*. Therefore, little is known about the biological motifs for the movement and behaviour of these animals. The location of migration corridors and residency areas in relation to habitat characteristics inform about the environmental preference of a species. Thus, access to this type of knowledge would increase the biological understanding of life in the oceans and provide valuable insights into possible climate change responses of fish and other marine animals.

The stocks of numerous commercially targeted marine species are declining. Differences in abundance and size composition of predator species indicate that a likely reason for the decline is overfishing (Ward and Myers, 2005). It is therefore important for scientists and conservation commissions to understand the movement and behaviour of these species to determine the actions required to obtain sustainable levels for industrial fishing.

The prime example is the Atlantic bluefin tuna (*Thunnus thynnus*), which is heavily targeted by commercial fisheries. In managing the fisheries of the species

[1] In the context of this work, movement refers to the animal's locomotion (in contrast with movement of body parts).

it was, until recently, believed that no mixing occurred between the East Atlantic and the West Atlantic bluefin tuna stock. Block et al. (2001), however, confirmed that individuals tagged in the west performed trans-Atlantic migrations to the Mediterranean Sea. This result prompted a change in the management of the Atlantic bluefin tuna stock and underlines the fact that knowledge about movement is key in spatial management.

Historically, assessments of fish movement and stock size have relied on data from conventional tagging studies (mark-recapture), from reportings of commercial catches, and scientific surveys. During the 1990's electronic data loggers were introduced providing a new type of data. By attaching electronic tags to free-ranging fish, detailed information about the individual's ambiance, for example sea temperature and depth, can be retrieved from the tag through recapture or satellite transmission. Electronic tagging data related to movement (in short movement data) are particularly useful in the assessment of marine animal movement (Metcalfe and Arnold, 1997).

State-space models are the most popular modelling framework for the analysis of movement data (Patterson et al., 2008*b*). The quality and type of movement data vary depending on the species of interest and the type of electronic tag used for data collection. Thus, a variety of methods have been employed to estimate state-space models for movement, for example the Kalman filter (Sibert et al., 2003), Markov chain Monte Carlo (Jonsen et al., 2005), and the particle filter (Andersen et al., 2007).

1.2 Aims of the thesis

This thesis focuses on the development and application of hidden Markov models (HMMs) for analysing nonlinear state-space models. The thesis is primarily concerned with modelling of movement data from fish in scenarios where standard approaches are insufficient. Specifically, the work encompasses:

- Using HMMs and data from electronic tags to estimate the geographical location (geolocation) of fish. Such methodology is particularly useful when individual location is observed indirectly and Gaussian error distributions are inappropriate.
- Developing an HMM which estimates movement and behaviour simultaneously using Markov switching. With such a model it is possible to relate individual behaviour to geographical regions, a feature of the results, which aids the understanding of the animal's use of space.

- The use of advanced numerical tools for approximating state-space models in continuous time and space with special attention to boundary conditions. These tools are needed for geolocating fish in waters with a complicated shoreline geometry such as the western Baltic Sea.
- Combining data from multiple individuals in an integrated framework with mixed models and HMMs. Such a framework can, for example, be useful in discerning individuals that deviate from the rest of the tagged population.
- Investigating the use of HMMs in more general nonlinear state-space models within ecology.

1.3 Outline

Chapter 2 and 3 give a thorough introduction to the data and the methods that have previously been used within the analysis of movement data. Chapters 4 and 5 describe the modelling and statistical methods used in this thesis. The chapters overview the methodology used in the papers and therefore contain some recapitulation. Chapter 6 focuses on the results obtained from applying the methods to synthetic and real data. Furthermore, it discusses the different studies in broader terms and considers the future prospects of movement data analysis from the perspective of a modeller. Finally, Chapter 7 concludes the study.

Chapter 2

Movement data

This chapter clarifies some of the terms and abbreviations that are prevalent within the analysis of movement data. The term movement data simply covers data that relate to the movement of an individual animal either by direct observations of location or by some proxy that relates to the location. For example, a dataset of daylight intensities taken at the individual can be used to roughly estimate its location via day-length and times of dawn and dusk (Musyl et al., 2001). Because data often come from tags attached to the animal, movement data are sometimes referred to as tagging data.

The most primitive type of movement data come from conventional tags. These tags are used for identification of the animal e.g. by simple numbering and have no electronics. Conventional tagging of bluefin tuna *(Thunnus thynnus)* was initiated in 1954 by Frank J. Mather III who also published some of the first statistical modelling of tagging data (Mather et al., 1974). At the time, such experiments were a major advance to the study of fish migrations. The conventional tag itself is cheap so for the decades following 1960 thousands of tags were deployed. Still, conventional tags only provide information about the release and recapture of the fish and very little if any about its behaviour. In addition, the time and location of recapture is fisheries dependent which biases the data and makes discovery of unexpected fish migration improbable.

During the 1950's the first studies of fish movements using sonic tracking tech-

nology were published (Trefethen, 1956). In sonic tracking a transmitter is mounted on the fish or implanted. This allows a research vessel to follow the fish at the surface by receiving transmissions emitted by the tag. See e.g. Yuen (1970) for an in-depth description of the sonic tracking procedure. In contrast with conventional tagging data this approach gives detailed information about the movement of an individual fish. The technique, however, is expensive owing to the required amount of equipment and personnel. Thus, it is infeasible to follow an individual for more than a couple of days.

In the 1980's and 1990's micro-processor technology was reaching a stage where the size of a chip was small enough to allow implementation on board a tag. Such a development allowed researchers to record information in the time period between tag attachment and recapture without the need for a monitoring research vessel. The vast amount of information accumulated by tags equipped with electronic data loggers was in stark contrast to the relatively sparse datasets retrieved from conventional and sonic tags. Naturally, the first electronic archival tags were large compared with those of today and their application was therefore limited to larger fish species or marine mammals. Since then, the development of electronic tags has been rapid.

Many branches of electronic tags have evolved each of which is highly specialised in collecting a certain type of data. Below, the most common types of electronic tag data are mentioned.

2.1 Archival data

Archival data are recorded by tags equipped with data storage facilities. This type of data are retrieved from so-called archival tags or, equivalently, data storage tags (DSTs). Archival tags are mostly used for species that are likely to be recaptured by commercial fisheries because tags need to be recollected physically for the full data record to be downloaded. Some archival tags (PSATS) have the added ability to transmit data via satellite (see below), however substantial battery requirements only allows a subset of the full archival record to be transmitted.

Archival tags without satellite transmission functionality are relatively moderate in size (1-2 cm) and are therefore mostly used to study smaller animals. Their size enables tags to be either mounted externally on the fish (see Figure 2.1) or surgically implanted to reduce swimming drag (see Figure 2.2).

An archival tag typically records ambient water temperature, pressure (a proxy

Figure 2.1: Atlantic cod tagged externally with an archival tag. *Photo credit: Line Reeh, DTU Aqua.*

for depth), light intensity, and sometimes salinity. Data are collected at high sample rates (minutes or even seconds if required) and potentially for many years if the animal avoids recapture. Because of the high sampling frequency, archival data are useful for studying both horizontal and vertical animal movement and behaviour. Archival data have previously been used in the study of plaice (Metcalfe and Arnold, 1997; Hunter et al., 2003), Atlantic cod (Andersen et al., 2007), Atlantic bluefin tuna (Block et al., 2001), bigeye tuna (Sibert et al., 2003), and more.

2.2 Pop-up data

Pop-up data are recorded by pop-up satellite archival tags (PSATs) that are preprogrammed to detach themselves from the individual and float to the surface to transmit data via satellite (see Figure 2.3). PSATS are larger tags (10-20 cm) because a considerable battery capacity is required for the data transmission. Thus, PSATS are primarily used in the study of large pelagics such as tuna

Figure 2.2: Internal tagging of an Atlantic cod. *Photo credit: Junita Karlsen, DTU Aqua.*

(Patterson et al., 2008*a*), marlin (Graves et al., 2002), shark (Nielsen et al., 2006), and recently even on eel (Aarestrup et al., 2009).

Movement analysis of pop-up data is normally based on observation of daylight intensities and sea surface temperature recordings. Owing to bandwidth limitations data have normally been preprocessed and summarised on board the tag. Often, the preprocessed data have a reduced sample frequency which is suited mainly for analysis of horizontal movement on a larger scale. Unfortunately, the propagation of measurement uncertainty through the proprietary software used for the data preprocessing is unknown.

Some PSATs only take measurements when the animal is close to the sea surface because observations of sea surface temperature (SST) are key in geolocation algorithms. Such data have inherent uneven sample intervals, a feature which must be accounted for in the modelling phase. Pop-up data is independent of the focus of commercial fisheries, which is a major advantage and critical property if unforeseen animal behaviour is to be revealed. In fact, data can be collected from species that are not targeted by commercial fisheries at all.

Figure 2.3: A juvenile southern bluefin tuna. Note the pop-up satellite archival tag mounted on the dorsal side of the tuna. *Photo credit: CSIRO MAR.*

2.3 Argos data

Argos data are direct[1] (but noisy) observations of location transmitted via the Argos satellite system. Argos tags are also capable of measuring water salinity, temperature, pressure, etc. For transmission to be possible the tag must be above the sea surface. Therefore, the use of Argos tags is restricted to marine animals that frequent the surface, for example seals (Jonsen et al., 2005), loggerhead turtles (Polovina et al., 2000), whale sharks (Wilson et al., 2007), and more.

As for pop-up data from PSATs, Argos data are fisheries independent. The main concern when working with Argos data is that the location uncertainty is heterogeneous in time. However, Argos data are categorised into quality classes,

[1] In fact, locations are derived by Argos processing centers from the Doppler shift of transmission signals, however it is common to regard Argos data as direct.

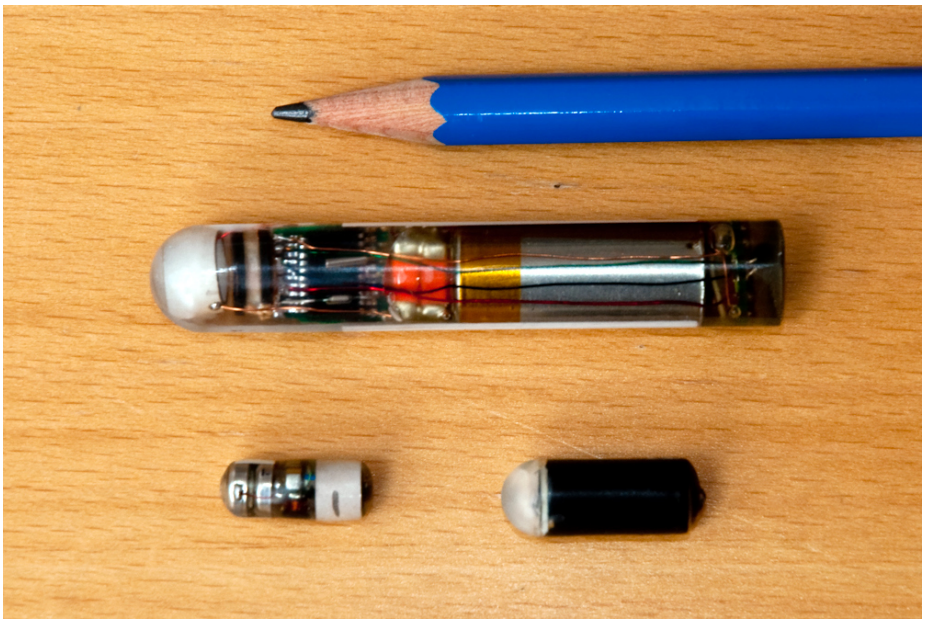


Figure 2.4: Three acoustic transmitters from Lotek. Similarly to some archival tags (see Figure 2.2) the acoustic tags are attached internally to the fish. The larger tag is, in addition to acoustic positioning, equipped with depth and temperature sensors. The smaller tags only have tracking capabilities. *Photo credit: Henrik Baktoft, DTU Aqua.*

which quantify the uncertainty of the observations. Argos tags are also prone to produce outlying locations. Therefore, the use of robust estimation methods is important when analysing these data.

2.4 Acoustic data

Acoustic data are transmitted by acoustic tags and collected by so-called listening stations or hydrophones. Acoustic tags do not require large storage or battery capacity. Thus, data can be collected from a wide range of marine animals of different sizes. Acoustic tags have some data storage capabilities and transmit depth and temperature along with a presence/absence signal when detected by a hydrophone (see Figure 2.4).

More advanced setups can provide direct observations of location by triangulating signals from multiple hydrophones. Hydrophones have a relatively limited listening range and many units are therefore required for triangulation setups. Thus, direct positioning using acoustics would typically be used to analyse smaller ecosystems such as those found in lakes or on reefs.

Acoustic positioning data can be sampled at very high frequencies (2 sec intervals) with high spatial accuracy allowing for observation of feeding or spawning behaviour. Acoustic data contain outliers because it is rarely possible to evenly cover the entire movement range of the animal with the hydrophone grid. Therefore, as for Argos data, robust estimation methods are required.

Some studies using acoustic data include the species bigeye thresher shark (Nakano et al., 2003), nurse and Caribbean reef shark (Chapman et al., 2005), and more.

Chapter 3

Methods and models for the analysis of movement data

Analysis of movement data from electronic tags is a maturing field of research. Initially, direct manual inspection of the data was possible because the amount of data was moderate. As the price for an electronic tag gradually dropped and data accumulated the need for automated analysis tools became pressing. Below, some of the methodology for movement data analysis is reviewed.

3.1 Geolocation problem

With noisy movement data the most fundamental problem is the estimation of animal location, i.e. the geolocation problem.

3.1.1 Geolocation techniques

The choice of approach to solving the geolocation problem depends on the type of movement data. For Argos and acoustic data the problem is predominantly an exercise in reducing the influence of outliers as observations are already of

location. On the other hand, for archival and pop-up data, animal location must somehow be inferred from monitored quantities that can be related to geographical regions. The choice of which environmental variables to measure depends on the characteristics of the relevant waters. Generally, however, only variables with significant spatial gradients are useful for geolocation.

In waters with a significant amphidromic system (e.g. the North Sea) the phase and amplitude of tidal fronts recorded via depth can be utilised to pinpoint the location of a tagged individual. This technique also requires that the species of interest spends enough time at the sea bed for the archival tag to record a significant part of the tidal cycle. This technique is referred to as the tidal location method (Hunter et al., 2003) and has previously been used to track plaice (Metcalfe and Arnold, 1997), thornback ray (Hunter et al., 2005), and more. If tidal variations are not present depth can still be used to roughly estimate location by comparison with bathymetry maps (Andersen et al., 2007). Depth can be supplemented with water temperature and salinity to increase the accuracy of the estimated location (Neuenfeldt et al., 2007).

Large pelagic species are typically studied using PSATs. With transmitted recordings of light intensity location is estimated from day length (latitude) and time of sunrise and sunset (longitude) (Sibert et al., 2003). Light-based geolocation, however, suffers from large uncertainties in the estimation of latitude in particular around the two equinox periods because day length is constant for all latitudes (Musyl et al., 2001). If available, recordings of sea surface temperature (SST) are used to improve the latitude estimate by comparing observed temperature with remotely sensed maps of the SST distribution (Nielsen et al., 2006).

3.1.2 Statistical implementations

While many heuristic methods have been used to solve the geolocation problem using the above techniques the focus here will be on the statistically based methods which employ state-space models (SSMs).

Today, SSMs have become the dominant statistical approach for solving geolocation problems (Patterson et al., 2008*b*). The versatility of SSMs, it turns out, is yet unrivaled by any other method when it comes to inferring location from movement data. The reason, as explained by Patterson et al. (2008*b*), is clear: the ability of SSMs to simultaneously model the stochasticity of the movement process and the observation process allows for separation of the noise at these two levels. This property is important because it mimics the actual features present in the biological system and the data acquisition process.

SSMs are suited for analysing observations of light intensity from tags attached to large pelagic species. Sibert et al. (2003) employed the Kalman filter (KF) to estimate location of bigeye tuna from archival data records of light. Using an extended KF (EKF) Nielsen et al. (2006) combined light and SST data to improve the uncertain latitude estimate. While the EKF handles mild departures from linearity through a first-order Taylor expansion, the unscented KF (UKF) is a derivative free second-order approximation (Julier et al., 2000). Lam et al. (2008) implemented the UKF to further improve light and SST-based geolocation in handling nonlinearities.

Jonsen et al. (2005) formulated an SSM using a correlated random walk (random walk on the velocity) for describing the movement of seals observed directly with Argos tags. To handle the occasional outliers in the Argos location data a t-distribution was used for modelling the observation error. The same model was applied to Argos data from leatherback turtles in Jonsen et al. (2007). The SSM was implemented in WinBUGS (Spiegelhalter et al., 1999), which is a free software for Bayesian analysis. The Bayesian modelling approach requires prior distributions to be specified for all parameters and is therefore tailored to problems where *a priori* information is available.

Argos data are collected opportunistically when the transmitter surfaces, a property which renders the time-intervals between successive observations nonuniform. Jonsen et al. (2005) linearly interpolated the Argos locations to obtain uniform time-intervals. An elegant alternative was presented in Johnson et al. (2008) where the correlated random walk model was analysed in continuous-time thus bypassing the need for interpolation. This method, however, requires Gaussian distributed errors and can therefore not exploit the t-distribution for robust handling of outliers.

The choice of geolocation method depends also on the environment of the species. For example, tuna in the Pacific Ocean rarely encounter dry land. In contrast, cod and other demersal species that live on the continental shelf are much more likely to be found in the vicinity of land masses. For these species, geolocation with KF techniques is inappropriate because a Gaussian distribution for location would often place the fish on land with non-zero probability. In a study of Atlantic cod in the Baltic Sea, Andersen et al. (2007) applied a particle filter method to estimate fish locations. Particle filters or equivalently sequential Monte Carlo methods (SMC, Cappé et al., 2007) are one of the most advanced and general implementations of SSMs. SMC is based on Monte Carlo simulations and free of distributional and linearity assumptions. This makes them appropriate for geolocation in areas with a complex shoreline geometry such as the Baltic Sea. However, despite being methodologically promising, a full SMC implementation requires much work and have therefore only gained moderate popularity as geolocation method.

The geolocation is the first step from which to proceed when analysing movement data. As will be apparent from the following sections, the geolocation serves as a building block for higher level inference regarding behaviour and population dynamics.

3.2 Movement behaviour

The term movement behaviour covers the types of behaviour that can be related to particular movement patterns, for example migratory behaviour or searching behaviour.

Animals alter their movement behaviour as a response to seasons and changes in their ambient environment, e.g. increased prey abundance or risk of predation (Bowler and Benton, 2005; Bestley et al., 2008). Movement can typically be roughly categorised as either fast and directed or slower and entangled. Ecologically, the former movement type would be interpreted as a display of migration to a more favourable habitat whereas the interpretation of the latter is less clear. Searching, foraging, and spawning behaviour would all classify as entangled movements, which are difficult to distinguish when animal location is observed on a relatively large spatial and temporal scale. Still, much information about behaviour can be inferred from the movement. It is therefore important that geolocation models incorporate animal behaviour to avoid bias in location estimates and uncertainty.

Movement data from terrestrial animals are often from accurate GPS tags. Therefore, modellers concerned with terrestrial data have somewhat bypassed the geolocation problem and focused more on the steps beyond, such as the analysis of behaviour. Morales et al. (2004) analysed GPS movement data from elk with an accuracy of 10-20 m and showed that the movement behaviour of these animals could be classified as either “encamped” or “exploratory”. Similar patterns have been observed in movement data from marine animals. A study of Argos data from seals (Jonsen et al., 2005) succeeded in classifying movement into migratory and foraging behaviour. The occasional outliers and uncertainty of Argos locations relative to GPS fixes demanded more care in the modelling phase with respect to robustness.

Recently, Patterson et al. (2009) considered movement behaviour classification of juvenile southern bluefin tuna from archival data. Data were summarised into daily position fixes by combining light and sea surface temperature information. Their study presented a comprehensive statistical machinery based on hidden Markov models for behaviour estimation using environmental data

as covariate information for the switching probabilities of the Markov process. Such an analysis enables ecologists to relate the estimated animal behaviour to characteristics of their ambiance. However, as discussed by the authors, their model did not include the location uncertainty, which for this species can be quite significant (Teo et al., 2004).

3.3 Individual based population inference

Although an electronic tag solely provides information about its host, analysis of single individual animals is rarely of great interest. In fact, the aim of most tagging studies is to achieve inference at the population level, which can provide answers to fundamental biological questions related to the species and is useful for management purposes. Thus far, however, the number of publications presenting individual based population inference related to marine animals has been limited. The main reason for this is that although much data from electronic tags have been collected, data from different studies and also data within studies are heterogeneous. Differences in data are present even for a single species, e.g. with respect to tag type, sample rate, observation uncertainty, time and location of deployment, length of time series etc. This emphasises the importance of proper experimental design prior to any tagging mission. However, even carefully planned studies cannot avoid the irregular sample rate of opportunistically transmitting tags, observations outliers, premature pop-ups etc. Instead such difficulties must be ameliorated in the modelling phase. Advanced models for handling these issues do exist and it is theoretically possible to build an integrated model which combines multiple tags. Yet, a joined model is challenging to implement and likely to be computationally cumbersome to work with.

One of the most developed models for analysing conventional tagging data from multiple individuals was presented in Sibert et al. (1999), which used the diffusion-advection-reaction equation for modelling the dispersal of skipjack tuna (*Katsuwonus pelamis*) in the southern Pacific Ocean. The data spanned the years 1977-1982 and encompassed more than 94000 releases and about 5300 recaptures. This vast amount of data from different individuals enabled the model to predict the probable spread of the population, and allowed for statistical tests of alternative hypothesis such as for the number of seasons evident in the data. Modelling of movement data from electronic tags cannot immediately follow a similar model because data collections of sufficient size are not yet available.

Using feeding data from archival tags and generalized linear mixed models, Best-

ley et al. (2008) assessed differences in behaviour among a group ($n = 19$) of juvenile southern bluefin tuna as a function of time of day and moon quarter. A similar approach could be taken to population analysis of movement data. In a study on simulated Argos turtle data, Jonsen et al. (2003) presented a hierarchical Bayesian model for joining state-space analyses of individual animal movement. Inference was on a parameter which related sea surface temperature to the movement rate of the animal. The authors showed that weak datasets with few observations could be analysed by borrowing strength from more informative datasets. They applied the hierarchical framework to real Argos data from $n = 14$ leatherback turtles to investigate diel variation in migration speed (Jonsen et al., 2006). This modelling framework incorporates many of the uncertainty aspects associated with movement data. It is implemented in the free MCMC based modelling software WinBUGS, which relies on a Bayesian model formulation.

Aarts et al. (2008) discussed the challenges related to using individual telemetry data for estimation of population space use. In addition, they presented a model which combined individual datasets to produce a map of presence. The model was fitted to telemetry (Argos location) and simulated data under a so-called case-control design. Specifically, their approach for the individual data assigns a probability of presence to each spatial unit of a gridded geographical region. Then, the individual models are joined using mixed effects modelling. The technique tackles many of the challenges of population modelling, however it does not account for temporal and spatial autocorrelation in the data. As discussed by the authors, future advancements in CPU technology should allow extending the framework to include SSMs.

Chapter 4

State-space models

Movement is a dynamical process. Consequently, movement datasets are always time series, i.e. composed of repeated observations of certain time-varying variables on the same individual. Biological processes that vary in time commonly display elements that can be explained theoretically and elements that appear random. For example, a migration path of an animal is directed but also perturbed by small scale movements that are not directed. There are many possible biological explanations for these perturbations, however explicit modelling at this level requires accurately observed data. Instead, small movements that are indistinguishable from observation noise are modelled as random.

Acquiring movement data from marine animals involves observing (with error) a dynamical system which displays a degree of randomness. State-space models (SSMs) formalise this procedure in a natural and intuitive manner by separating the biological process from the observation process.

An example of a simple, but often very useful movement model is the random walk, where the location X_k of an individual evolves according to

$$X_{k+1} = X_k + \epsilon_k, \tag{4.1}$$

where ϵ_k are independent and identically distributed (iid). The random walk is a Markov process since future states of the random process are independent of past

states conditional on the present state. A process with this characteristic is said to have the Markov property and is therefore a Markov process. Equivalently (in discrete-time), if $P(X_{k+1}|X_1, \ldots, X_k) = P(X_{k+1}|X_k)$ for all values of k then $\{X_k\}_{k\in\{1,2,\ldots\}}$ is a Markov process.

Observation errors are present when monitoring any biological process. Say a biological process (e.g. movement) X_k is monitored. What is actually observed is

$$Z_k = X_k + e_k \tag{4.2}$$

where e_k is the error inherent in the procedure of observing X_k. Often, when tracking terrestrial animals with accurate GPS telemetry, e_k will be small and sometimes negligible. Conversely, in studying marine animal movement, e_k often has significant influence on retrieved data. Failure to acknowledge this will most likely lead to false or biased conclusions about the location of the animal.

The two equations (4.1) and (4.2) constitute an example of an SSM. In general, though, SSMs can have other forms and do not necessarily include processes that are Markov or even random. This chapter outlines the class of SSMs that are relevant to the modelling of movement data from marine animals and describes the general theory behind state and parameter estimation.

4.1 General notation

The animal location at time t_k is $\boldsymbol{X}_k = (X, Y)_k^T$, where X is longitudinal coordinate, and Y is latitudinal coordinate. The state-space is denoted Ω, which in this chapter is equivalent to the model domain. Later Ω also encompasses behavioural states. The vector of observed/measured quantities at time t_k is $\boldsymbol{z}_k$ with $k \in \{1, \ldots, N\}$, i.e. observations are available at N different points in time. The time interval between two observations is $\Delta_k = t_{k+1} - t_k$.

The observation vector $\boldsymbol{z}_k$ is a generic way to refer to observations of any nature. For example, $\boldsymbol{z}_k$ can be a set of location coordinates, a measure of temperature or day light intensities etc. If data are sampled at a frequency higher than the time-stepping of the model, $\boldsymbol{z}_k$ can encompass multiple measurements that are related to time t_k. This is often the case for models using archival data. Then, say 60 measurements of depth can be gathered in a single observation, $\boldsymbol{z}_k$ related to t_k. The composition of $\boldsymbol{z}_k$ is specified separately for each model. All observations taken at or prior to t_k are jointly referred to as $\mathcal{Z}_k = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}$.

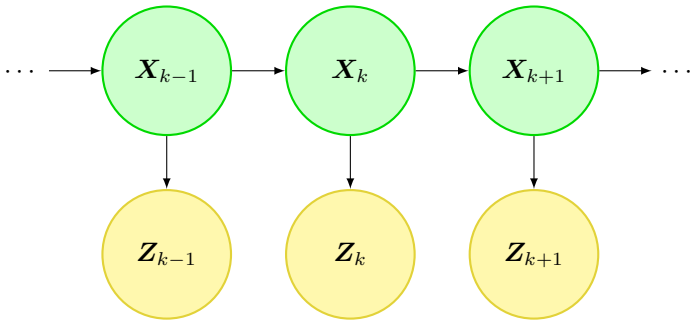


Figure 4.1: Dependence structure of a state-space model. $\boldsymbol{X}_k$ are unobservable states, $\boldsymbol{Z}_k$ are observed data.

4.2 Discrete-time dynamical model

Commonly, SSMs are specified in discrete-time in which case the general formulation of the process model (4.1) and observation model (4.2) is

$$\boldsymbol{X}_{k+1} = g(t_k, \boldsymbol{X}_k, \boldsymbol{\epsilon}_k), \tag{4.3}$$
$$\boldsymbol{Z}_k = h(t_k, \boldsymbol{X}_k, \boldsymbol{e}_k), \tag{4.4}$$

where g and h are possibly nonlinear mappings of the location $\boldsymbol{X}_k$. The noise terms $\boldsymbol{\epsilon}_k$ and $\boldsymbol{e}_k$ are mutually independent, iid., and have arbitrary but known distributions. The SSM in (4.3) and (4.4) is a way to formalise the dependence structure between the unobservable states $\boldsymbol{X}_k$ and the data $\boldsymbol{Z}_k$ (see Figure 4.1).

Often, when the objective is to estimate location, some simplifying assumptions are reasonable and necessary to implement and analyse the SSM. In this work the preferred process model is a biased random walk or, equivalently, a random walk with a drift term. Discussion of the choice of movement model is found in Chapter 6. With a biased random walk (4.3) becomes

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \boldsymbol{u}_k + \boldsymbol{\epsilon}_k, \tag{4.5}$$

where $\boldsymbol{u}_k = \boldsymbol{u}\Delta_k$ and $\boldsymbol{\epsilon}_k \sim N(\boldsymbol{0}, 2\mathbf{D}\Delta_k)$. In general, the two movement compo-

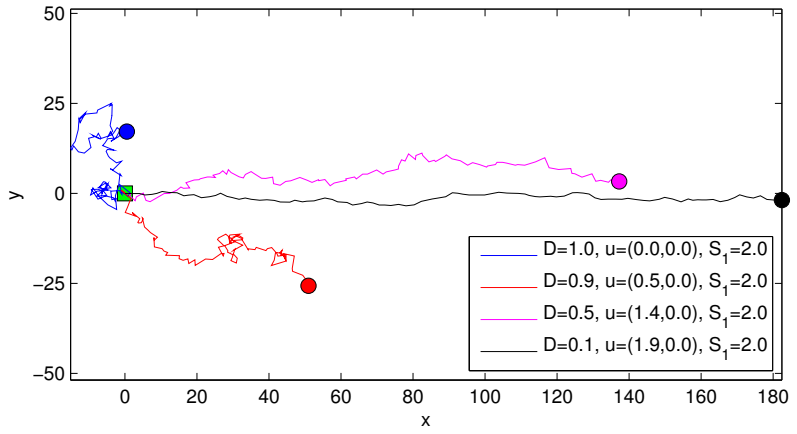


Figure 4.2: Simulated biased random walks with isotropic diffusion. All four simulated tracks have the same expected squared displacement. Clearly, as the advection becomes the dominating term the tortuosity of the path decreases.

nents take the form

$$\boldsymbol{u} = \begin{pmatrix} u_x \\ u_y \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}. \tag{4.6}$$

The deterministic part of the movement is represented by the advection or drift $\boldsymbol{u}$, whereas the random part of the movement is controlled by $\mathbf{D}$, which is the location diffusivity. It is assumed that $D_{xy} = 0$ unless stated otherwise.

Mixing deterministic and random components makes the interpretation of the movement parameters somewhat unclear. One way to summarise the movement is by the square root of the expected squared displacement $S_{\Delta t}$ over the time interval Δt: $S_{\Delta t} = \{E[(\Delta X)^2] + E[(\Delta Y)^2]\}^{1/2}$, where ΔX is the step length in the X-direction in Δt. This quantity is denoted the expected step length. Using the definition of variance

$$S_{\Delta t} = \left\{2D_{xx}\Delta t + (u_x\Delta t)^2 + 2D_{yy}\Delta t + (u_y\Delta t)^2\right\}^{1/2}. \tag{4.7}$$

On small time scales the displacement is dominated by diffusive terms whereas on larger time scales the drifts become the main contributing factors (see Figure 4.2). This formula also applies to continuous-time models.

4.3 Continuous-time dynamical model

Time intervals between movement observations are often uneven. Modelling of such data with a discrete-time model is difficult and often requires some sort of interpolation of data. More appropriate is the use of an SSM formulated in continuous-time. In this thesis the following class of continuous-time models is considered

$$d\boldsymbol{X}_t = f(t, \boldsymbol{X}_t)dt + g(t, \boldsymbol{X}_t)d\boldsymbol{B}_t, \tag{4.8}$$
$$\boldsymbol{Z}_t = h(t, \boldsymbol{X}_t, \boldsymbol{e}_t), \tag{4.9}$$

where $\boldsymbol{B}_t$ is a standard two-dimensional Brownian motion. Equation (4.8) is in general termed a stochastic differential equation. For modelling movement (4.8) is often simplified

$$d\boldsymbol{X}_t = \boldsymbol{u}dt + \boldsymbol{\sigma}d\boldsymbol{B}_t. \tag{4.10}$$

With a formulation in continuous-time the model handles time intervals between observations of arbitrary length seamlessly. In (4.10) the advection parameter $\boldsymbol{u}$ appears directly, whereas the diffusion parameters appears indirectly as $\mathbf{D} = \frac{1}{2}\boldsymbol{\sigma}\boldsymbol{\sigma}$. Then the expected step length can be calculated via (4.7).

4.4 Estimation

This section describes estimation of the probability density of the unobservable state and estimation of the model parameters.

4.4.1 State estimation

For state estimation it is assumed that all model parameters are known. Then, a complete analysis of an SSM would compute the full posterior density of the states conditional on all data

$$p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N | \mathcal{Z}_N). \tag{4.11}$$

However, the full posterior density can only be computed accurately when simplifying assumptions such as Gaussianity can be justified. Furthermore, even for moderate dimensional full posteriors it is impossible to visualise all the information it holds. Instead, state-space analyses focus on estimating time-marginal densities of the full posterior, that is

$$p(\boldsymbol{x}_k|\mathcal{Z}_N) = \int p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N|\mathcal{Z}_N) d\mathcal{X}_k, \tag{4.12}$$

where $\mathcal{X}_k = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{k-1}, \boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_N\}$. In this work, the collection of $p(\boldsymbol{x}_k|\mathcal{Z}_N)$ for all k are referred to as the posterior density and (4.11) is termed the *full* posterior density.

State estimation of the marginal densities is carried out in a filtering and a smoothing procedure. The filtering recursions for the general SSM described by (4.3) and (4.4) are

$$p(\boldsymbol{x}_{k+1}|\mathcal{Z}_k) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\mathcal{Z}_k) d\boldsymbol{x}_k, \tag{4.13}$$

$$p(\boldsymbol{x}_k|\mathcal{Z}_k) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\mathcal{Z}_{k-1})}{\int p(\boldsymbol{z}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\mathcal{Z}_{k-1}) d\boldsymbol{x}_k} \tag{4.14}$$

see e.g. Kitagawa (1987). Here, all densities are given for fixed parameters ($\boldsymbol{\theta}$) although this has been omitted from the equations for simplicity. In the context of tagging data the initial state density $p(\boldsymbol{x}_1|\boldsymbol{z}_1)$ is usually known since the release location of the tag is known. The two steps in the recursion are often termed the time-update and data-update respectively.

The smoothing step iterates backward in time and produces state densities that appear "smoother" than the filtered densities. The smoothing recursion is

$$p(\boldsymbol{x}_k|\mathcal{Z}_N) = p(\boldsymbol{x}_k|\mathcal{Z}_k) \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) \frac{p(\boldsymbol{x}_{k+1}|\mathcal{Z}_N)}{p(\boldsymbol{x}_{k+1}|\mathcal{Z}_k)} d\boldsymbol{x}_{t+1}. \tag{4.15}$$

The smoothing recursions start with $p(\boldsymbol{x}_N|\mathcal{Z}_N)$ which is the final filtered estimate and the first smoothed estimate. See e.g. Kitagawa (1987) for derivation of the smoothing step. When the smoothing step is finalised the smoothed state estimates $p(\boldsymbol{x}_k|\mathcal{Z}_N)$ are available for $k \in \{1, \ldots, N\}$. Note that the density is conditional on all available information in the data.

4.4.2 Parameter estimation

The model parameters are collectively referred to as $\boldsymbol{\theta} = \{\mathbf{D}, \boldsymbol{u}\}$. The observation noise $\boldsymbol{e}_t$ also has related parameters, however, these parameters can often be estimated from independent data and thus be omitted from $\boldsymbol{\theta}$. The likelihood function for the model parameters evaluated at $\boldsymbol{\theta}$ is given directly by the filter recursions since

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}_N) = p(\mathcal{Z}_N|\boldsymbol{\theta}) = p(\boldsymbol{z}_1|\boldsymbol{\theta}) \prod_{k=2}^{N} p(\boldsymbol{z}_k|\mathcal{Z}_{k-1}, \boldsymbol{\theta}), \tag{4.16}$$

where $p(\boldsymbol{z}_k|\mathcal{Z}_{k-1}, \boldsymbol{\theta}) = \int p(\boldsymbol{z}_k|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{x}_k|\mathcal{Z}_{k-1}, \boldsymbol{\theta}) d\boldsymbol{x}_k$ which is the denominator of (4.14). Now, the maximum likelihood (ML) estimate of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}). \tag{4.17}$$

Closed-form solutions to (4.13-4.17) are not available in general because the integrals are intractable. In the special case where $\boldsymbol{e}_k$ and $\boldsymbol{\epsilon}_k$ are Gaussian distributed and the process and observation equations are linear, the smoothed state estimates are also Gaussian and their mean and covariance have exact expressions. Likewise, an exact expression is available for $\widehat{\boldsymbol{\theta}}$. Analysis of some types of movement data lead to scenarios where Gaussian assumptions on the errors can be justified (Sibert et al., 2003), in which case the filtered and smoothed state estimates are given by the Kalman filter (KF, Harvey, 1990).

Regarding nonlinear and/or non-Gaussian SSMs where the KF is inappropriate, other approaches must be taken. Here is named a few alternatives which have been employed within movement data analysis: the extended Kalman filter (Nielsen et al., 2006), the unscented Kalman filter (Lam et al., 2008), particle filter methods (Andersen et al., 2007; Cappé et al., 2007) or Markov chain Monte Carlo methods using WinBUGS (Jonsen et al., 2005; Gilks et al., 2001). The approach taken in this work relies on hidden Markov models and is similar to that of Kitagawa (1987), however motivated differently, see Chapter 5.

4.4.3 An alternative estimation approach

The analysis of a state-space model seeks to estimate the unobservable process and the model parameters from the information in the observed data. This setup is equivalent to a mixed effects statistical model where the set of all states $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ comprises the random effects. Say the joint conditional

probability density $p(\mathcal{X}, \mathcal{Z}_N|\boldsymbol{\theta})$ can be computed. Then, the ML estimate for $\boldsymbol{\theta}$ is found by optimising

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}) = p(\mathcal{Z}_N|\boldsymbol{\theta}) = \int p(\mathcal{X}, \mathcal{Z}_N|\boldsymbol{\theta}) d\mathcal{X}. \tag{4.18}$$

This integral of dimension N is difficult to compute in general. Instead, the integrand must be approximated. To this end, the state estimates are computed

$$\widehat{\mathcal{X}}(\boldsymbol{\theta}) = \arg\max_{\mathcal{X}} \left\{ \log p(\mathcal{X}, \mathcal{Z}_N|\boldsymbol{\theta}) \right\}, \tag{4.19}$$

Now, the log of the integrand in (4.18) can be approximated by a quadratic function $q\big(\mathcal{X}, \widehat{\mathcal{X}}(\boldsymbol{\theta}), \boldsymbol{H}(\boldsymbol{\theta})\big)$ with peak in $\widehat{\mathcal{X}}(\boldsymbol{\theta})$ and curvature $\boldsymbol{H}(\boldsymbol{\theta})$, where

$$H_{ij}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(\mathcal{X}, \mathcal{Z}_N|\boldsymbol{\theta}) \Big|_{\mathcal{X}=\widehat{\mathcal{X}}}$$

are the elements of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\theta})$. Then, the integral reduces to

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}) \simeq \int \exp\left[q\big(\mathcal{X}, \widehat{\mathcal{X}}(\boldsymbol{\theta}), \boldsymbol{H}(\boldsymbol{\theta})\big)\right] d\mathcal{X} = p\big(\widehat{\mathcal{X}}, \mathcal{Z}_N|\boldsymbol{\theta}\big) \sqrt{\frac{(2\pi)^M}{\det \boldsymbol{H}(\boldsymbol{\theta})}}, \tag{4.20}$$

where M is the number of model parameters. Now, (4.20) is a known function of $\boldsymbol{\theta}$ and therefore simple to optimise. The technique of approximating $p(\mathcal{X}, \mathcal{Z}|\boldsymbol{\theta})$ by a quadratic function is called Laplace's method or Laplace's approximation (Wolfinger and Xihong, 1997).

In summary, the estimation of $\boldsymbol{\theta}$ involves iteration between an inner estimation (4.19) and an outer estimation of $\boldsymbol{\theta}$ by optimisation of (4.20). That is, for estimating $\mathcal{X}$ the parameters $\boldsymbol{\theta}$ are fixed to the previous estimate and vice versa. The iterations start with some initial guess on $\boldsymbol{\theta}$. Such an alternating estimation procedure is standard (albeit involved) when estimating nonlinear mixed models (Pawitan, 2001).

For SSMs, it is common that $N > 100$. Thus, the optimisation in (4.19) is not straightforward to compute. Using automatic differentiation, however, and the fact that the autocorrelation of $\boldsymbol{X}_k$ has a certain structure, the optimisation can be carried out surprisingly fast. The open-source AD Model Builder software provides an integrated framework which exploits these features (Skaug and Fournier, 2006).

Chapter 5

Hidden Markov models

A model where the distribution of observed quantities depends on an unobservable Markov process is a hidden Markov model (HMM, Zucchini and MacDonald, 2009). Following this definition the state-space models (SSMs) mentioned in Chapter 4 are indeed also HMMs. HMMs can operate in both continuous and discrete time and space. However, as encountered in the literature, HMMs are predominantly used to model phenomena with a finite and discrete state-space, for example in speech recognition (Rabiner, 1989). An advantage of discrete state HMMs is that exact analysis is possible. That is, closed-form expressions for state probabilities and the most probable state sequence are available.

Animal movement is a dynamical process in continuous space. However, a process in discrete space is obtained by restricting the possible locations of the animal to a finite number of cells in a grid. This procedure enables the movement process to be analysed using discrete state HMM theory. In addition, it increases modelling flexibility since aspects such as outlying observations, uneven data sampling, behaviour switching, hierarchical modelling, and avoidance of dry land, can be incorporated. As described in Chapter 3, many of these issues have been addressed by previous studies, however only by separate models. Moreover, discrete state HMMs enable illustrative visualisation of the estimation results.

In this thesis, the method of discretising a continuous state HMM is denoted the

HMM method. So, to analyse a continuous state SSM using the HMM method is to discretise the state-space of the SSM and use discrete state HMM theory for the analysis. Onward from here, the term HMM refers to an HMM with discrete and finite state-space. The remainder of this chapter details the HMM method with extensions and their use in the context of movement data.

5.1 Discrete-time

The principle of the HMM method as outlined above is now put into mathematical terms. The HMM appears when the continuous model domain (Ω) is partitioned into a finite number of grid cells. The grid cell with center in $\boldsymbol{x} = (x, y)^T$ is denoted $\Omega_{\boldsymbol{x}}$. In a uniform grid the number of grid cells in the longitudinal and latitudinal directions are n_x and n_y respectively. Therefore the total number of grid cells is $n_{xy} = n_x n_y$. The width of a grid cell is dx. HMMs with nonuniform grids are discussed in Section 5.2.1 and Chapter 6.

With the location of the animal resolved on a discrete grid movements on a scale finer than dx are no longer discernible. Still, with a sufficiently fine grid other sources of error will become the limit for the location accuracy.

The discrete analogue of the probability density $p(\boldsymbol{x}_k|\mathcal{Z}_k)$ is the row vector $\boldsymbol{\phi}(t_k, \mathcal{Z}_k)$. Each of the n_{xy} elements in $\boldsymbol{\phi}(t_k, \mathcal{Z}_k)$ corresponds to a cell in the (x, y) grid such that $\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_k)$ is the probability that the animal is at location $\boldsymbol{x}$ at time t_k conditional on $\mathcal{Z}_k$. In other words $\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_k) = P(\boldsymbol{X}_k = \boldsymbol{x}|\mathcal{Z}_k)$.

With the discrete grid the filter recursions become

$$\phi_{\boldsymbol{x}}(t_{k+1}, \mathcal{Z}_k) = \sum_{\boldsymbol{x}_k} P(\boldsymbol{X}_{k+1} = \boldsymbol{x}|\boldsymbol{X}_k = \boldsymbol{x}_k)\phi_{\boldsymbol{x}_k}(t_k, \mathcal{Z}_k), \tag{5.1}$$

$$\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_k) = \frac{P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{x})\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_{k-1})}{\sum_{\boldsymbol{x}} P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{x})\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_{k-1})} \tag{5.2}$$

analogous to (4.13) and (4.14). For short-hand reference the denominator of (5.2) is written

$$\psi_k = P(\boldsymbol{Z}_k|\mathcal{Z}_{k-1}) = \sum_{\boldsymbol{x}} P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{x})\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_{k-1}). \tag{5.3}$$

This quantity is used for parameter estimation in (4.16) and (4.17). The smoothing equation (4.15) on the discrete state-space is

$$\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_N) = \phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_k) \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{X}_{k+1} = \boldsymbol{x}_{k+1} | \boldsymbol{X}_k = \boldsymbol{x}) \frac{\phi_{\boldsymbol{x}_{k+1}}(t_{k+1}, \mathcal{Z}_N)}{\phi_{\boldsymbol{x}_{k+1}}(t_{k+1}, \mathcal{Z}_k)}. \quad (5.4)$$

The main difference between continuous and discrete state-space is that integrals have been replaced with sums.

The transition probabilities in (5.1), (5.2), and (5.4) are computed by integrating the density of the movement process over the grid cells. The density of the movement process is given by (4.5) and (4.10). Steps of equal length have equal transition probability because the movement process is homogeneous in space. With $\boldsymbol{w} = adx$ for any integer a, the probability of a transition from $\boldsymbol{x}$ to $\boldsymbol{x} + \boldsymbol{w}$ in a time period of Δ_k is then

$$P(\boldsymbol{X}_{k+1} = \boldsymbol{x} + \boldsymbol{w} | \boldsymbol{X}_k = \boldsymbol{x}) = \int_{\Omega_{\boldsymbol{w}}} N_{pdf}(\boldsymbol{s}, \boldsymbol{u}\Delta_k, 2\mathbf{D}\Delta_k) d\boldsymbol{s}, \quad (5.5)$$

where $N_{pdf}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian probability density function with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{x}$. Recall that $\Omega_{\boldsymbol{w}}$ is the grid cell with center at $\boldsymbol{w}$. In general, the integral needs to be solved numerically which can be done with a quadrature algorithm. Less accurate but faster algorithms (e.g. the trapeziodal rule) can be employed if parameters are time-varying and the transition probabilities need to be recalculated in each time step. If movement in the x and y-directions is independent, (5.5) can be computed by simply evaluating the Gaussian cumulative distribution function at the boundaries of the grid cell.

The term $P(\boldsymbol{Z}_k = \boldsymbol{z}_k | \boldsymbol{X}_k = \boldsymbol{x})$ in (5.2) is the probability of observing $\boldsymbol{z}_k$ given the location $\boldsymbol{x}$. Computing this as a likelihood (i.e. as a function of $\boldsymbol{x}$) for all states, highlights the likely locations at time t_k. The complexity of computing $P(\boldsymbol{Z}_k = \boldsymbol{z}_k | \boldsymbol{X}_k = \boldsymbol{x})$ depends on the data type and the function in (4.4) which links $\boldsymbol{x}$ and $\boldsymbol{z}_k$.

As explained by Pedersen (2007), the sums in (5.1) and (5.4) are in fact convolutions. The time-update (5.1) has the two-dimensional discrete Gaussian convolution kernel $\mathbf{K}_k$ which is a matrix with the transition probabilities (5.5) as elements (see Figure 5.1 for an example). The range for $\boldsymbol{w}$ should be chosen such that the smallest elements of $\mathbf{K}_k$ are close to zero, e.g. $< 10^{-6}$. The sum of all elements in $\mathbf{K}_k$ must, of course, be 1.

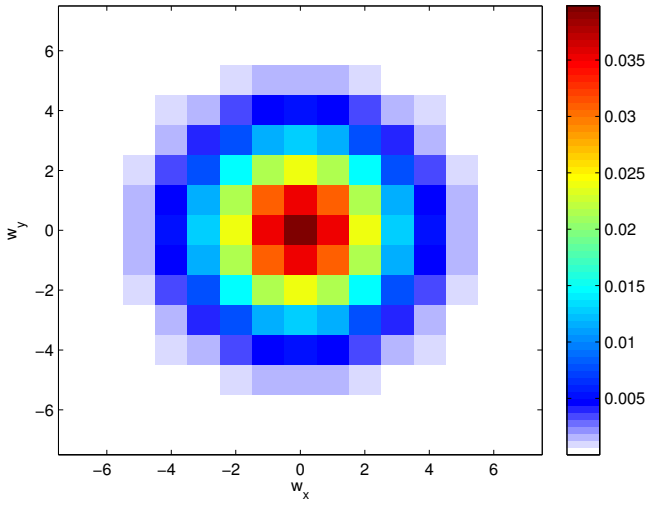


Figure 5.1: Convolution kernel with transition probabilities. Parameter values are $\Delta_k = 1$, $dx = 1$, $\boldsymbol{u} = \mathbf{0}$ and $(D_{xx}, D_{yy}) = (0.5, 0.5)^T$. In this case, if $\boldsymbol{w} = (w_x, w_y)^T = (0, 0)^T$ say, then $P(\boldsymbol{X}_{k+1} = \boldsymbol{x} + \boldsymbol{w} | \boldsymbol{X}_k = \boldsymbol{x}) = 0.0398$

Then (5.1) can be written on a compact form

$$\phi_{\boldsymbol{x}}(t_{k+1}, \mathcal{Z}_k) = \mathbf{K}_k \star \phi_{\boldsymbol{x}_k}(t_k, \mathcal{Z}_k),$$

where '$\star$' is the convolution operation. The data-update (5.2) remains the same while the smoothing step (5.4) becomes

$$\phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_N) = \phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_k) \left[\mathbf{J}_k \star \left(\frac{\phi_{\boldsymbol{x}_{k+1}}(t_{k+1}, \mathcal{Z}_N)}{\phi_{\boldsymbol{x}_{k+1}}(t_{k+1}, \mathcal{Z}_k)} \right) \right],$$

where $\mathbf{J}_k$ is constructed from (5.5), however with $\boldsymbol{u}$ replaced by $-\boldsymbol{u}$, i.e. $\mathbf{J}_k$ is the mirror image of $\mathbf{K}_k$.

In Matlab, an implementation of the filter and smoother with the convolution formulation requires very little computation time because the built-in `conv2` function makes efficient use of the fast Fourier transform. The main drawback of the convolution approach is the accuracy of the solution in the vicinity of land areas. The convolution does not distinguish between dry land and sea. Therefore, for distributions close to shore lines probability mass is likely to be convolved into states on land. This is not as disastrous as it seems, since the

data-update step (5.2) assigns zero weight to states that are invalid and thus removes the probability mass on land. Still, this has the effect that the probability distribution is "repulsed" from the shores. If data are informative, however, this is of minor concern. In the alternative case where only vague information is available the estimated probability distributions become less reliable and alternative solution approaches should be considered (see below).

5.2 Continuous-time

While the continuous-time formulation of the SSM (see Section 4.3) is more complex to analyse it is also more flexible. Furthermore, for unevenly sampled data results turn out to be more accurate as compared to a discrete-time analysis.

Given the movement process (4.8) in continuous time and space the evolution of $p(\boldsymbol{x}_t|\mathcal{Z}_k)$ for $t > t_k$ is described by the Kolmogorov forward equation (Øksendal, 2007)

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\boldsymbol{u}p - \mathbf{D}\nabla p) \tag{5.6}$$

where ∇ is the two-dimensional spatial gradient operator. The solution domain is the relevant geographical region with land masses excluded. Thus, the boundaries of the domain are the shorelines. The appropriate boundary conditions for (5.6) are Neumann boundary conditions, which reflect probability mass and therefore ensures that the solution (a probability density) integrates to unity.

On a discrete spatial grid like the one outlined in Section 5.1 a partial differential equation (PDE) is reduced to a system of ordinary differential equations (ODEs). In PDE terminology the technique of discretising all spatial dimensions is the method of lines (Schiesser, 1991). With a second-order central scheme for the spatial derivatives the rates of moving (jumping) a distance of dx to the east and west are $\frac{D_{xx}}{dx^2} \pm \frac{u_x}{2dx}$ respectively. For moving a distance of dx to the north and south the rates are $\frac{D_{yy}}{dx^2} \pm \frac{u_y}{2dx}$ respectively. It can be confirmed that for a process with these rates the expected movement in a time-interval dt is $\boldsymbol{u}dt$ with covariance $2\mathbf{D}dt$ as stated by (4.8). Since negative jump rates are not allowed (i.e. boundedness of the solution must be preserved) the resolution of a uniform grid must fulfill the requirement that

$$dx < \min\left\{\frac{2D_{xx}}{|u_x|}, \frac{2D_{yy}}{|u_y|}\right\}.$$

The jump rates are organised in the generator matrix $\mathbf{G}$ with the diagonal elements of $\mathbf{G}$ equal to the negative row sum of the corresponding row (see Figure 5.2). Now, the time evolution of the discrete state vector for $t > t_k$ is

$$\frac{d}{dt}\boldsymbol{\phi}(t, \mathcal{Z}_k) = \mathbf{G}\boldsymbol{\phi}(t, \mathcal{Z}_k). \tag{5.7}$$

This formulation is recognised from the theory of continuous-time Markov chains on a discrete state-space (Grimmett and Stirzaker, 2001). Using the generator to describe the movement process is a powerful approach because the rate of moving to states on land can be set to zero (Figure 5.2). Consequently, the boundaries (shore lines) are reflective to the animal, and non-zero probabilities are never assigned to dry land areas. In fact, states on dry land areas can be completely omitted from the state space thereby saving computing time since the size of $\mathbf{G}$ is reduced.

When data have been collected, all time-intervals Δ_k between observations are known. Then, the transition probability matrices $\mathbf{P}_k$ which describe the evolution of the state probability distribution can be calculated by solving (5.7). Theoretically, it is straightforward to compute $\mathbf{P}_k$ (Grimmett and Stirzaker, 2001) since

$$\mathbf{P}_k = \exp(\mathbf{G}\Delta_k).$$

The exponential of a matrix is defined as an infinite sum

$$\exp(\mathbf{A}h) = \sum_{i=0}^{\infty} \frac{1}{i!}(\mathbf{A}h)^i. \tag{5.8}$$

Thus, approximations are required to calculate the matrix exponential. Many numerical methods are available to this end, see Moler and Van Loan (2003). The preferred method of this thesis is to use the uniformization approximation which exploits that the generator of a continuous-time Markov chain on a large state-space is in general sparse and banded. The technique will not be detailed here, but see Grassmann (1977).

The continuous-time filter and smoothing recursions are respectively

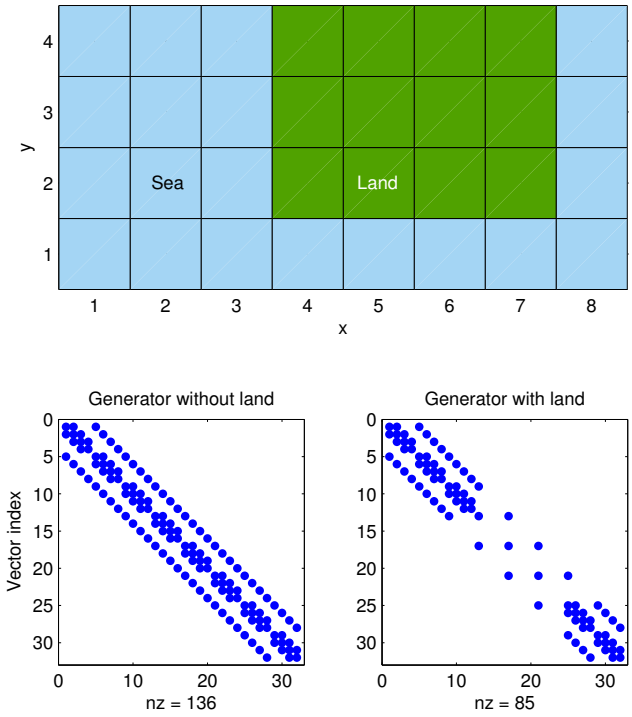


Figure 5.2: Top: An example of a discretised domain with $n_x = 8$ and $n_y = 4$. Blue indicates sea and green indicates land states. Bottom left: Dots are non-zero elements of $\mathbf{G}$. Here rates for moving to and from land states have not been set to zero. Bottom right: Dots are non-zero elements of $\mathbf{G}$. Here rates for moving to and from land states are zero. In $\mathbf{G}$ linear indexing of the states is used. For example, state $(x, y) = (1, 4)$ in the grid is number 4 in linear indexing, state (2, 1) equals linear state 5 etc.

$$\boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_k) = \boldsymbol{\phi}(t_k, \mathcal{Z}_k)\mathbf{P}_k, \tag{5.9}$$

$$\boldsymbol{\phi}(t_k, \mathcal{Z}_k) = \psi_k^{-1}\boldsymbol{\phi}(t_k, \mathcal{Z}_{k-1}) \odot \mathbf{L}(\boldsymbol{z}_k), \tag{5.10}$$

$$\boldsymbol{\phi}(t_k, \mathcal{Z}_N) = \boldsymbol{\phi}(t_k, \mathcal{Z}_k) \odot \left[\{\boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_N) \oslash \boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_k)\}\mathbf{P}_k^T\right], \tag{5.11}$$

where $\odot$ and $\oslash$ are element-wise product and division respectively. The data likelihood vector $\mathbf{L}(\boldsymbol{z}_k)$ of length n_{xy} has elements $P(\boldsymbol{Z}_k = \boldsymbol{z}_k | \boldsymbol{X}_k = \boldsymbol{x}_k)$ for all $\boldsymbol{x}_k$. The normalisation constant $\psi_k = \boldsymbol{\phi}(t_k, \mathcal{Z}_{k-1}) \cdot \mathbf{L}(\boldsymbol{z}_k)$ is used for maximum likelihood estimation of parameters as in (5.3). Here, '$\cdot$' is the dot-product.

This formulation of the continuous-time filter and smoother is compact and relatively simple to implement because all sums are replaced with vector and matrix operations. It is, however, more computationally costly than the convolution approach. In theory, $\mathbf{P}_k$ are dense, however for small time-intervals or movement rates the majority of the elements of the matrices are close to zero ($< 10^{-6}$ say). By setting these transitions to zero $\mathbf{P}_k$ become sparse. As a consequence, the computational expense of the recursions (5.9-5.11) is reduced significantly since optimised sparsity algorithms can be employed.

5.2.1 Filtering from a PDE perspective

As discussed there are several numerical approaches to solving the filtering equation. This is because the problem can be viewed as a that of solving an integral (4.13), or as that of solving a partial differential equation (PDE), i.e. (5.6). Both topics are heavily studied in the literature on numerical analysis. Approaches to solving integrals include quadrature rules, Monte Carlo simulation, functional approximations (such as Laplace's approximation), and more. Approaches to solving PDEs include finite difference methods, finite volume methods, finite element method and more.

The convolution approach described above is strongly related to the method of finite differences (FDs, see Mitchell and Griffiths, 1980). In FD methods (as for the convolution approach) it is cumbersome to implement boundary conditions. In addition, the FD method is mostly suited for problems dominated by diffusive terms. When advection is present in the problem it is important to compensate the solution for the "artificial" or numerical diffusion which is induced. Sometimes, however, numerical diffusion cannot be avoided unless the grid is altered.

An alternative to FD is the finite element (FE) method which is widely used for solving the PDEs that arise in structural mechanics (Cook et al., 2001).

The main advantage of FEM is its ability to handle boundaries with a complex geometry and to implement the conditions that apply to these boundaries in a simple manner. FEM works on an unstructured mesh, so the mesh can be refined in areas of particular importance and coarsened elsewhere. Note that refining the mesh will not (necessarily) make the estimate of the state more accurate, but it will provide a better approximation to the probability density of the location. The FE method in relation to movement data analysis is discussed further in Chapter 6.

5.3 Visualising results

When the filtering and smoothing recursions are completed in either discrete or continuous-time the estimated state probability vectors $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ can be used for answering ecologically relevant questions about the animal. To this end, the large amount of information contained in the posterior distribution can be visualised either as surface plots of the two-dimensional marginal state distributions or as sequence plots showing the animal's movement path.

5.3.1 Movement trajectories

The most common means to visualise movement is a trajectory. In the HMM framework a movement trajectory is the concatenation of all locations $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$.

5.3.1.1 Simple trajectory estimation

Two basic, but often encountered approaches to trajectory estimation are sketched here.

The mean track is calculated by connecting the means of the smoothed distributions $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$. The means in the x and y-directions are respectively

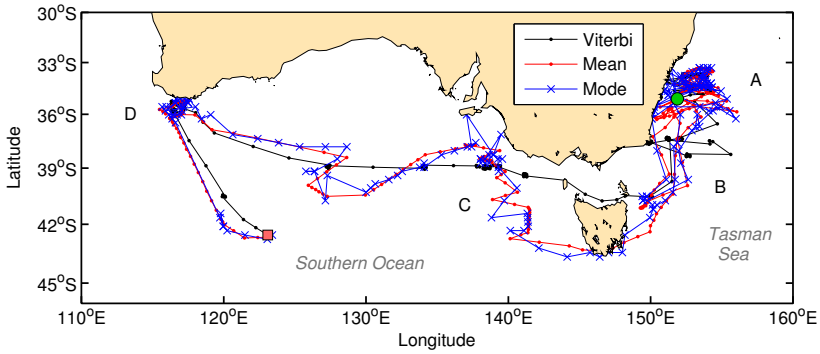


Figure 5.3: Viterbi, mean, and mode track estimated using the same dataset from a southern bluefin tuna. Green circle is release location and orange square is pop-up location of the PSAT tag. The tracks are similar close to A. At B and C the Viterbi track diverges from the two other tracks in that it moves north of Tasmania. The three tracks differ in estimating the timing of the south-ward migration from D: The mode track makes a sudden jump of approximately six degrees latitude in one time step (12 hours), the mean track averages this movement over many small steps, and the Viterbi track takes larger steps and then pauses before it moves on.

$$\bar{x}_k = \sum_x x \left[\sum_y \phi_{x,y}(t_k, \mathcal{Z}_N) \right],$$
$$\bar{y}_k = \sum_y y \left[\sum_x \phi_{x,y}(t_k, \mathcal{Z}_N) \right].$$

So the mean track $\bar{\mathcal{X}}$ is comprised of the means related to each time point, i.e. $\bar{\mathcal{X}} = \{\bar{\boldsymbol{x}}_1, \ldots, \bar{\boldsymbol{x}}_N\}$, where $\bar{\boldsymbol{x}}_k = [\bar{x}_k, \bar{y}_k]^T$ (see Figure 5.3 for an example). Using the mean track has some pitfalls which should be mentioned: The means $\bar{\boldsymbol{x}}_k$ may refer to locations with zero probability, for example between modes of the distribution or even on dry land. The mean track relies on the time-marginal distributions and thus ignores transition probabilities. So jumps with zero probability can occur in the mean track.

The mode track is identical to the mean track with the exception that the means are replaced with modes of the smoothed distributions. At time t_k the mode is

$$\tilde{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}} \boldsymbol{\phi}(t_k, \mathcal{Z}_N),$$

leading to the mode track $\tilde{\mathcal{X}} = \{\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_N\}$. In contrast to $\bar{\boldsymbol{x}}_k$, $\tilde{\boldsymbol{x}}_k$ are guaranteed to be a location with non-zero probability. However, the mode track should still be employed with caution (see Figure 5.3). In the multimodal case if $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ have alternating maxima the mode track will jump between the maximum points because transition probabilities are not accounted for. Such a scenario would result in an invalid track. Thus, it important to inspect $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ and ensure unimodality before using either the mean or the mode track.

5.3.1.2 Generating a likely movement trajectory

Given a probability distribution one can generate random outcomes of this distribution with a random number generator. Therefore, random outcomes can be generated from $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$. An outcome $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ generated from $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ for all k represents a likely movement trajectory of the animal.

The recursions for generating a random trajectory are initialised by sampling a random state $\boldsymbol{v}_N$ from $\boldsymbol{\phi}(t_N, \mathcal{Z}_N)$. Then, starting with $k+1 = N$, iterate over the following four steps:

1. Create a distribution (δ_{k+1}) with all probability mass concentrated at the location $\boldsymbol{v}_{k+1}$.
2. Compute the evolution of δ_{k+1} backward in time by

   $$\pi_k = \boldsymbol{\phi}(t_k, \mathcal{Z}_k) \odot [\{\delta_{k+1} \oslash \boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_k)\}\mathbf{P}_k^T].$$

   This equation is identical to the smoothing step (5.11) with $\boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_N)$ replaced with δ_{k+1}.
3. Sample a random state, $\boldsymbol{v}_k$, from π_k.
4. Set $k = k - 1$. If $k > 1$ go to step 1 otherwise stop.

When the recursion is finalised at $k = 1$ the random trajectory is $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$.

Sampling of random trajectories is a versatile instrument that can be used to answer questions that do not seem quantifiable at first. Questions such as: what is the probability that the animal resided in a particular area during the summer

months? or: when was the most probable time the animal initiated a migration toward a specific habitat? Generating multiple random tracks from the posterior distribution and jointly analysing these answers the questions posed. Say 800 out of 1000 random trajectories resided in an area of interest during the summer months and the remaining 200 did not. Then, the probability of the animal residing in that area during the summer months is estimated to 80% with standard deviation 0.013% (using the variance of a Bernoulli distributed random variable). For estimating the timing of migrations e.g. the departure time from a habitat, the say 1000 random tracks are analysed and the individual departure times assessed. The timing of the migration is then estimated e.g. as the mean of the individual departure times. Alternatively can the distribution of departure times be inspected to reveal possible bimodality. Random track analysis evidently posseses great potential for dealing with biologically relevant and seemingly complex scenarios.

5.3.1.3 Most probable movement trajectory

The likelihood of a trajectory is

$$L(\mathcal{X}) = P(\boldsymbol{Z}_1 = \boldsymbol{z}_1|\boldsymbol{X}_1 = \boldsymbol{x}_1) \prod_{k=2}^{N} P(\boldsymbol{X}_k = \boldsymbol{x}_k|\boldsymbol{X}_{k-1} = \boldsymbol{x}_{k-1}) P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{x}_k), \tag{5.12}$$

where $P(\boldsymbol{X}_k = \boldsymbol{x}_k|\boldsymbol{X}_{k-1} = \boldsymbol{x}_{k-1})$ can be found by look-up in the transition matrix $\mathbf{P}_{k-1}$ and $P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{x}_k)$ is the data likelihood. The most probable trajectory $\widehat{\mathcal{X}}$ is the one that maximises $L(\mathcal{X})$, i.e.

$$\widehat{\mathcal{X}} = \arg\max_{\mathcal{X}} L(\mathcal{X}).$$

The trajectory $\widehat{\mathcal{X}}$ is the most likely of all possible sequences generated by the HMM. Determining this sequence is carried out with the Viterbi algorithm (Viterbi, 2006) which relies on principles from dynamic programming. The Viterbi algorithm is outlined below:

Define the branch metric related to time t_k

$$B_k(\boldsymbol{x}, \boldsymbol{y}) = \log P(\boldsymbol{X}_k = \boldsymbol{y}|\boldsymbol{X}_{k-1} = \boldsymbol{x}) + \log P(\boldsymbol{Z}_k = \boldsymbol{z}_k|\boldsymbol{X}_k = \boldsymbol{y}).$$

The branch metric can be interpreted as the log-likelihood of the branch leading from location $\boldsymbol{x}$ to location $\boldsymbol{y}$. Using the branch metric, define the state metric related to time t_k

$$S_k(\boldsymbol{x}_k) = \max_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k} \left\{ \log P(\boldsymbol{Z}_1 = \boldsymbol{z}_1 | \boldsymbol{X}_1 = \boldsymbol{x}_1) + \sum_{l=2}^{k} B_l(\boldsymbol{x}_{l-1}, \boldsymbol{x}_l) \right\},$$

which is the log-likelihood of the most likely of all possible state sequences leading to $\boldsymbol{x}_k$ at time t_k. This optimisation is complicated to carry out directly. Fortunately, the state metric can be formulated as a recursion which exploits the Markov property of the HMM. Then

$$S_k(\boldsymbol{x}_k) = \max_{\boldsymbol{x}_{k-1}} \{ S_{k-1}(\boldsymbol{x}_{k-1}) + B_k(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k) \}. \tag{5.13}$$

This optimisation uses the state metric from the previous time step which already contains the log-likelihood of the most likely sequence at t_{k-1}. Still, computing the state metric is a demanding operation since it must be carried out for all locations in the state-space, i.e. (5.13) must be computed n_{xy} times for each time step. When the final time step is reached the state metric of the most probable trajectory is

$$S_N(\widehat{\boldsymbol{x}}_N) = \max_{\boldsymbol{x}_N} S_N(\boldsymbol{x}_N).$$

To get the actual trajectory there are two approaches: By storing the most likely sequence leading to each location throughout the recursion, along with the state metric, the most probable trajectory can simply be picked out as the one related to $S_N(\widehat{\boldsymbol{x}}_N)$ at the final time step. This approach requires that all intermediate sequences are stored in the memory. The memory requirement, however, is unlikely to become a prohibitive factor.

The alternative approach is to do an additional recursion which runs in reverse time and starts with $\widehat{\boldsymbol{x}}_N$:

$$\widehat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}_k} \{ S_k(\boldsymbol{x}_k) + B_k(\boldsymbol{x}_k, \widehat{x}_{k+1}) \}.$$

This approach requires more computation, but minimal memory. In this work, the former of the two approaches has primarily been used.

Computing the most probable trajectory is involved when compared to computing the mean or the mode track. However, the drawbacks of both of the simple

tracks are avoided because by including transition probabilities it is guaranteed that $\widehat{\mathcal{X}}$ is a valid track (see Figure 5.3).

If location uncertainty is heterogeneous in time, which is often the case for HMMs, it is difficult to visualise trajectory and location uncertainty simultaneously. Furthermore, for long data series adding uncertainty information to a trajectory plot is more likely to confuse than guide the reader since confidence bounds may overlap. Instead, by generating a number of likely trajectories, the variation around the estimated track can be visualised.

5.3.2 Distribution plots

The full posterior probability density of an SSM is the density of all locations conditional on all observations, i.e. $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N | \mathcal{Z}_N)$. The dimension of the full posterior is 2^N. For HMMs, where the state-space is discretised, $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ is a vector containing the marginal distribution of the location at time t_k. Say the full posterior distribution $\Phi = P(\boldsymbol{X}_1 = \boldsymbol{x}_1, \ldots, \boldsymbol{X}_N = \boldsymbol{x}_N | \mathcal{Z}_N)$ was calculated and stored in a vector; the length of this vector would be n_{xy}^N. Even for small problems (say $n_{xy} = 1000$, $N = 100$) it would not be possible to store Φ.

Fortunately, the stochastic process underlying the posterior distribution is Markov. Consequently, the information in Φ can be accessed through its time-marginals $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ provided by the HMM smoothing recursions. For example, random outcomes of Φ can be generated (Section 5.3.1.2) and the mode of Φ can be determined (Section 5.3.1.3).

An alternative to trajectory plots are distribution plots which highlight the spatial variability of the the posterior distribution. The simplest distribution plot is to view the time-marginals $\boldsymbol{\phi}(t_k, \mathcal{Z}_N)$ in succession for increasing k, i.e. as an animation. Alternatively a map of the space usage of the animal in the time interval $\tau = \{a, a+1, \ldots, b\} \subseteq \{1, \ldots, N\}$ can be constructed. The expected time spent at a location $\boldsymbol{x}_k$ in the time interval Δ_k is $\Delta_k \phi_{\boldsymbol{x}}(t_k, \mathcal{Z}_N)$. Therefore, the vector containing the expected time spent for all locations in τ is

$$\boldsymbol{R}_\tau = \sum_{k \in \tau} \Delta_k \boldsymbol{\phi}(t_k, \mathcal{Z}_N).$$

The residency distribution $\boldsymbol{r}_\tau$ is the normalised cumulative distribution related to $\boldsymbol{R}_\tau$. The residency distribution (RD) indicates the space usage of the animal by assigning a number between 0% and 100% to each location in the spatial grid. Then, the contour line of the RD at, say, 15% encloses the smallest region

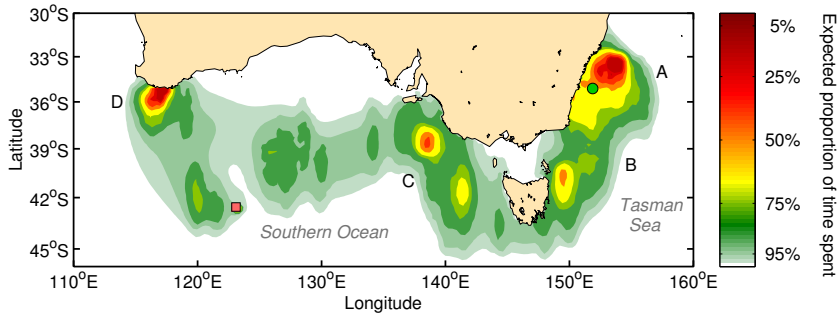


Figure 5.4: Residency distribution $\boldsymbol{r}_\tau$ for a southern bluefin tuna. Green circle is release location and orange square is pop-up location of the PSAT tag. The, say, 50% contour (orange colour) encloses the smallest region where the animal was expected to spend 50% of its time.

where the animal was expected to spend 15% of its time (see Figure 5.4). The following commands in Matlab compute $\boldsymbol{r}_\tau$ from $\boldsymbol{R}_\tau$:

```
 R_normalised     = R_tau/sum(R_tau);
[R_sort,indices]  = sort(R_normalised,'ascend');
 R_sum            = cumsum(R_sort);
 r_tau            = zeros(size(R_tau));
 r_tau(indices)   = R_sum;
```

Distribution plots are useful in showing the spatial variability of the animal's location while disregarding temporal patterns. Conversely, trajectory plots highlight temporal correlation of the movement. It is therefore important to always consult both trajectory and distribution plots when synthesising the posterior distribution.

5.4 Markov switching

Marine animals adapt their movement in response to environmental and physiological factors. To obtain reliable movement estimates it is important to implement a model which includes different types of movement behaviours such

as migratory or resident. This increases the flexibility of the model and provides a higher level of inference since focus is not only on the animal's location, but also on its behaviour.

Markov switching is a technique known from the analysis of regime shifting time series, i.e. data for which the underlying model dynamics appear to shift between n separate models. The switching dynamics is modelled by a Markov process with n states. In the HMM framework Markov switching is incorporated by simply adding a new hidden state I_k which represents the internal behavioural state of the animal at time t_k. The switching HMM has nn_{xy} states. Typically $n = 2$ with a state representing slow movements and a state representing fast movements. This has a clear biological interpretation as resident and migratory behaviour. Estimation of a model with $n > 2$ requires highly informative and frequently sampled data and is not necessarily tractable in practice.

The general movement model including the behavioural state is

$$\boldsymbol{X}_{k+1} = g(\boldsymbol{X}_k, I_k, \boldsymbol{\epsilon}_k). \tag{5.14}$$

Often, the behavioural state is "fully hidden", because it must be inferred conditional on the estimated movement (see Figure 5.5). If available, auxiliary data of the physiological state of the animal can be used to indicate possible shifts in movement behaviour. As of yet, only few tags provide this type of data. So in this work the behavioural state is always fully hidden.

The switching HMM has a set of movement parameters for each of the n behavioural states. In continuous-time the behavioural process is parametrised by the switching rates λ_{ij}, which are the probability of switching from behavioural state i to j in an infinitesimal time-step. The forward Kolmogorov equation including behavioural switching is

$$\frac{\partial p_i}{\partial t} = -\nabla \cdot (\boldsymbol{u}_i p_i - \mathbf{D}_i \nabla p_i) + \sum_j \lambda_{ji} p_j, \tag{5.15}$$

where p_i, $\mathbf{D}_i$, and $\boldsymbol{u}_i$ are the density and parameters related to behavioural state i. Analogous to the non-switching model, the generator $\mathbf{G}$ of the movement and behaviour process can be constructed from (5.15). Since $\mathbf{G}$ is a $nn_{xy} \times nn_{xy}$ matrix this approach may not be computationally feasible. Instead the behaviour and movement processes can be treated separately by their respective generator matrices $\mathbf{G}^b$ and $\mathbf{G}_i^m$ and corresponding transition matrices $\mathbf{P}_k^b$ and $\mathbf{P}_{i,k}^m$ for the time-interval Δ_k. The time-update step of the filter (5.9) is then split into a behaviour and a movement update

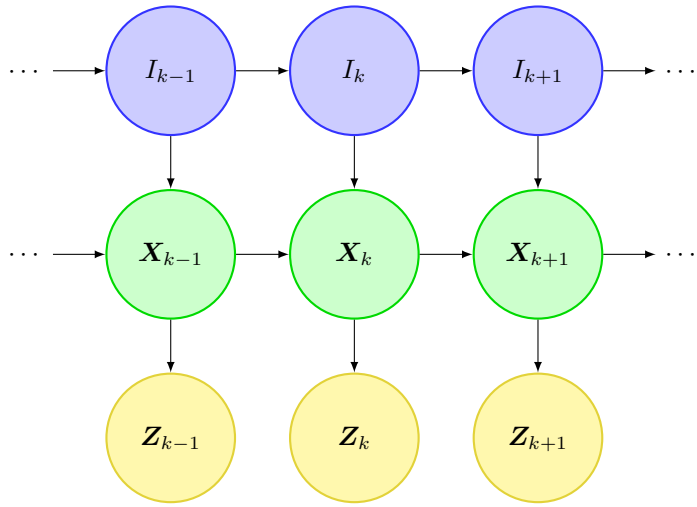


Figure 5.5: Dependence structure of a state-space model with behavioural switching. I_k are fully hidden behavioural states because they can only be estimated conditional on $\boldsymbol{X}_k$ (the unobservable location states) which are estimated from $\boldsymbol{Z}_k$ (observed data).

$$\boldsymbol{\Psi}_i(t_k, \mathcal{Z}_k) = \sum_{j=1}^{n} \boldsymbol{\phi}_j(t_k, \mathcal{Z}_k)\mathbf{P}_k^b(j,i), \qquad \text{behaviour} \tag{5.16}$$

$$\boldsymbol{\phi}_i(t_{k+1}, \mathcal{Z}_k) = \boldsymbol{\Psi}_i(t_k, \mathcal{Z}_k)\mathbf{P}_{i,k}^m, \qquad \text{movement} \tag{5.17}$$

where $\boldsymbol{\phi}_j(t_k, \mathcal{Z}_k)$ is the vector of spatial states related to behaviour j, and $\mathbf{P}_k^b(j,i)$ is probability of jumping from j to i during Δ_k. The updated state probability vectors $\boldsymbol{\phi}_i(t_{k+1}, \mathcal{Z}_k)$ related to $i \in \{1, \ldots, n\}$ are then concatenated to form $\boldsymbol{\phi}(t_{k+1}, \mathcal{Z}_k)$. The smoothing step (5.11) is changed analogously. This separation approach involves taking the matrix exponential n times of a $n_{xy} \times n_{xy}$ matrix and once of a $n \times n$ matrix. The alternative is to take the matrix exponential once of a $nn_{xy} \times nn_{xy}$ which is more computationally costly.

Including the behavioural state has many advantages with respect to the inferential potential of the model. The posterior distribution can still be visualised using trajectories and distribution plots. A state sequence now also includes the behavioural state. So the Viterbi sequence also provides the most likely behaviour switching sequence, however the fundamentals of the algorithm remain the same. The smoothed estimate of the behavioural state can be visualised as

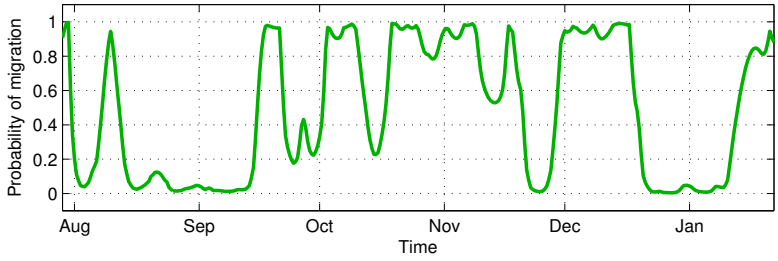


Figure 5.6: Probability of a southern bluefin tuna displaying migratory behaviour estimated using pop-up data simultaneously with estimating location (see Figure 5.3).

a simple line plot, which shows the probability of a certain behaviour through time (see Figure 5.6). Moreover, the posterior distribution can be marginalised in different ways similar to the residency distribution to reveal links between behaviour and spatial regions. This is discussed further in Chapter 6.

5.5 Hierarchical modelling

Hierarchical models are used to capture the variability between individuals of a population. In the context of tagging experiments the term population is frequently used, but its meaning is somewhat ambiguous since it can refer to both the population comprised by the tagged individuals of the specific study, or to the population of which the tagged individuals are a sub-sample. Here, the population term refers to the group of tagged individuals in the experiment.

Hierarchical modelling can be tackled from two angles: Jonsen et al. (2003) take a Bayesian approach, where the hierarchical dependencies are modelled with prior distributions. Models with mixed effects are the frequentist alternative to the hierarchical Bayesian approach and relies on random effects to describe the individual variability. The use of mixed effects models with HMMs is the approach taken here.

Say data are available from M individuals of the same population. If the population has parameter vector $\boldsymbol{\theta}$, then the parameter vector of individual i is

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

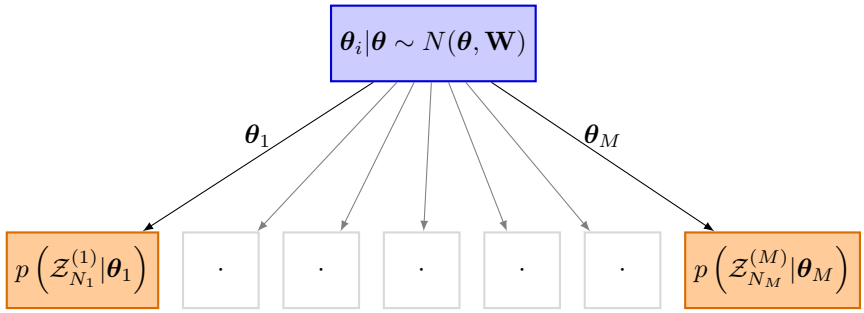


Figure 5.7: Structure of a hierarchical model. The individual parameters $\boldsymbol{\theta}_i$ are generated from the population parameters $\boldsymbol{\theta}$ and $\mathbf{W}$.

where it is assumed that $\boldsymbol{w}_i \sim N(\mathbf{0}, \mathbf{W})$. The random effects vector $(\boldsymbol{w}_i)$ represents the deviation of individual i from the population. With this setup the aim is to estimate $\boldsymbol{\theta}$, $\mathbf{W}$, and $\boldsymbol{w}_i$ for all $i \in \{1, \ldots, M\}$ with maximum likelihood.

Now, N_i refers to the number of observations from individual i and $\mathcal{Z}_{N_i}^{(i)}$ refers to all the N_i observations from i. The joint probability density of the random effects and individual observations conditional on $\boldsymbol{\theta}$ and $\mathbf{W}$ is

$$p\left(\boldsymbol{w}_i, \mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \mathbf{W}\right) = p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \boldsymbol{w}_i\right) p\left(\boldsymbol{w}_i|\mathbf{W}\right), \tag{5.18}$$

by the definition of conditional densities and where $p\left(\mathcal{Z}_{N_i}^{(i)}|\boldsymbol{\theta}, \boldsymbol{w}_i\right)$ for each individual is given by (4.16) since $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \boldsymbol{w}_i$ (see Figure 5.7). The density in (5.18) can be regarded as a likelihood function if it is viewed as a function of random effects and the model parameters, i.e.

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

It is not straightforward to maximise $L(\boldsymbol{\theta}, \mathbf{W}, \boldsymbol{w}_i)$ so an iterative approach is taken instead. With fixed $\boldsymbol{\theta}$ and $\mathbf{W}$ the ML estimate of the random effects for individual i is

$$\widehat{\boldsymbol{w}}_i = \arg\max_{\boldsymbol{w}} L(\boldsymbol{\theta}, \mathbf{W}, \boldsymbol{w}_i). \tag{5.19}$$

The contribution from individual i to the likelihood function for the population parameters is found by marginalising over the random effects

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

Since individuals are conditional independent given $\boldsymbol{\theta}$ and $\mathbf{W}$, the likelihood function conditional on data from all individuals is the product of the individual likelihood contributions

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

where $\mathcal{Z} = \left\{\mathcal{Z}_{N_1}^{(1)}, \ldots, \mathcal{Z}_{N_M}^{(M)}\right\}$. Thus, the ML estimate of $\boldsymbol{\theta}$ and $\mathbf{W}$ is

$$(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{W}}) = \arg\max_{\boldsymbol{\theta}, \mathbf{W}} \left\{L(\boldsymbol{\theta}, \mathbf{W}|\mathcal{Z})\right\}. \tag{5.20}$$

A possible algorithm for the optimisations in (5.19) and (5.20) is outlined in Pawitan (2001) and basically iterates between the two optimisation problems similarly to the procedure of an expectation-maximisation algorithm. Some modifications to this approach are required, however, because here the uncertainty of the individual parameter estimates $\widehat{\boldsymbol{\theta}}_i$ vary among individuals.

Chapter 6

Results and discussion

This chapter summarises and discusses the main results documented in Papers A-F. The aim of the thesis was to investigate the use of theory from hidden Markov models (HMMs) in analysing general nonlinear state-space models (SSMs) with emphasis on the analysis of movement data from fish. Therefore, focus in this chapter will be on both the methodological and technical contributions as well as the ecological results. Potential areas for future research are also mentioned.

6.1 Methodological contributions

This section summarises the most interesting technical solutions that have emerged during the course of this work.

6.1.1 General nonlinear state-space analysis

SSMs are used in the analysis of dynamical systems where some of the system states of interest are observed with noise via related quantities. Monitoring

dynamical systems in ecology always involves some uncertainty and these systems are therefore suited for state-space analysis. The work of this thesis was motivated by the analysis of nonlinear ecological models for movement data. A technique of discretising the continuous state-space and solving the filtering and smoothing problem on the grid was used to deal with nonlinearities. Here, this approach is termed the HMM method.

Paper E explains how the HMM method can be used to solve general nonlinear SSMs with Gaussian process and observation errors. The distributional assumption is, in fact, not a requirement for the HMM to operate. Instead, the main limitation of the HMM method is that the dimension of the state-space must be below, say, four. This is because discretisation of a high-dimensional state-space entails a large number of grid cells, which leads to prohibitively high computational demands.

The Automatic Differentiation Model Builder (ADMB, Skaug and Fournier, 2006) framework is another alternative to state-space analysis (see Paper E). The ADMB approach is efficient and is not restricted to low dimensional SSMs. ADMB relies on two techniques: 1) the integrals in the filtering and smoothing equations (4.13) and (4.15) are computed using Laplace's approximation (Wolfinger and Xihong, 1997). For this approximation to be valid the shape of the integrands must be close to quadratic in the log-domain. If the log integrands deviate from the quadratic form, ADMB can instead employ importance sampling to estimate the integrals at the expense of computing time. 2) in optimising the likelihood function automatic differentiation is utilised to significantly increase accuracy and computing speed. Note, however, that ADMB is less suited for the movement and behaviour problems of this thesis since boundaries to the state-space (shore lines) and discrete distributions (behaviour switching) cannot be incorporated.

The purpose of Paper E was to assess the estimation performance of the HMM method, ADMB, and BUGS (which is a Markov chain Monte Carlo approach). For the theta logistic population growth model (a nonlinear SSM), the timing results pointed toward ADMB as the advantageous analysis framework. The HMM method gave identical estimation results to ADMB, whereas BUGS was less accurate in its parameter estimates. Judging from the conclusion of this study, it is somewhat surprising that ADMB and HMM have been largely overlooked by published state-space analyses. The reason, as argued in Paper E, is probably that larger effort has been invested into making BUGS generally accessible via the user friendly WinBUGS (Spiegelhalter et al., 1999). In addition BUGS has fewer assumptions (e.g. state dimension, quadratic form of integrands) that limit its applicability.

Approaches similar to the HMM method have previously been considered by

other authors (Bucy and Senne, 1971; Kitagawa, 1987; Zucchini and MacDonald, 2009). However, its use in the literature on nonlinear time series analysis has thus far been outshone by sequential Monte Carlo (SMC) methods or, equivalently, particle filters (PFs). In future work it would be interesting to investigate how PFs perform in comparison with the three alternatives mentioned above. Such a study should encompass a wide range of SSMs that challenge the assumptions of the different approaches, for example the widely analysed SMC benchmark problem (Example 1 in Cappé et al., 2007), and the four dimensional bearings-only tracking problem (Gordon et al., 1993).

6.1.2 State-space analysis with the finite element method

The stochastic differential equation (4.8) is a possible model for the movement of an individual in continuous time. Since the movement, to some degree, is random its evolution must be described in probabilistic terms via for example expectations and variances. However, because the relation between movement and observed data, as defined by the observation equation (4.9), is nonlinear a moment representation is insufficient. Instead, a non-parametric version of the probability density of the location is more appropriate.

The temporal evolution of the probability density of the state is expressed by the Kolmogorov forward equation (5.6), which is a partial differential equation (PDE). As noted in Section 5.2.1, there are several methods available to solve PDEs. Paper C investigated the use of the finite element (FE) method to solve (5.6). An advantage of the FE method is that it operates on an irregular discretisation of the state-space. This feature is beneficial if it is important to accurately model boundaries to the state-space (see Figure 6.1). With the FE method it is also simple to implement any type of boundary conditions (e.g. reflecting, absorbing).

Using the FE method in a filtering and smoothing context is uncomplicated once the FE machinery is set up. All the HMM procedures remain the same with the exception that forward and backward time updates are handled by the FE solver. Setting up the FE solver, however, is a non-negligible task. The irregular mesh increases flexibility and accuracy of the solution, but can, for the same reason, be demanding to construct.

Paper C used the open-source software Triangle (Shewchuk, 1996) to triangulate the model domain. Boundary geometry (shorelines) was extracted from the mapping package in Matlab and imported into Triangle. It is advisable to start with a relatively coarse grid to compute a crude solution to the smoothing problem. Then, using Triangle the mesh can be refined in regions that require

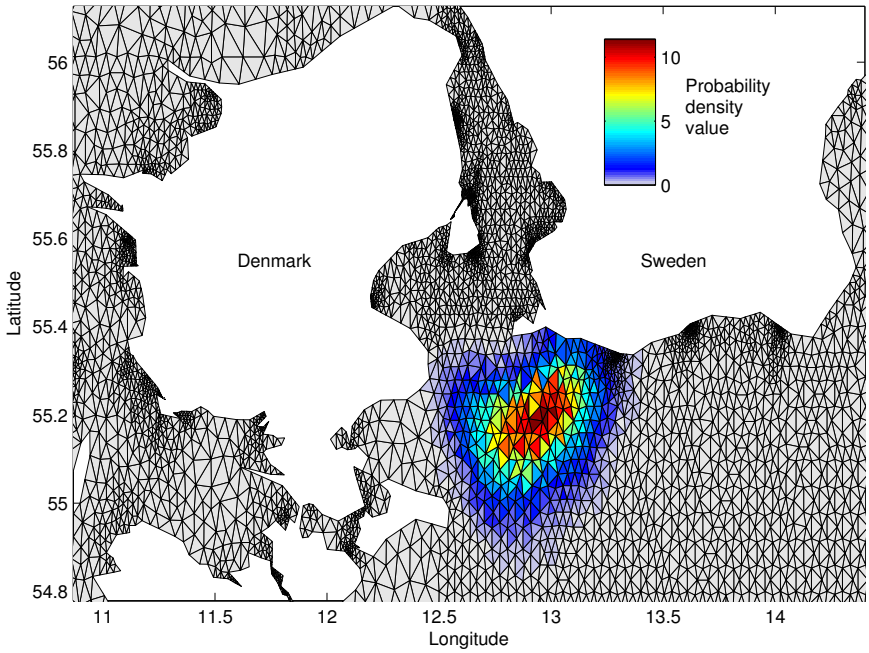


Figure 6.1: Probability density of the location of a cod in the Baltic Sea calculated with the finite element method. Note that the triangulated mesh accurately reproduces the geometry of the landmasses. Note also that the mesh has been refined in the sound between Denmark and Sweden (Øresund) because the external data fields (salinity and bathymetry) have large gradients in this region that need to be accurately resolved.

higher accuracy (see Figure 6.1).

The idea of using the FE method in filtering and smoothing of continuous-time nonlinear SSMs is novel. Paper C showed that the FE method was superior to a PF in approximating the state probability density of an SSM. The performance in estimating mean and variance for the two methods was identical.

In general, the FE computing time depends on the number of nodes in the mesh, which in turn grows exponentially in the state-space dimension for a fixed resolution. Consequently, the technique is limited to state-spaces with dimension below four. The FE approach therefore has its primary application in the analysis of one to three dimensional problems, where one wishes to compute

the full probability density.

The highly nonlinear geolocation scenario analysed in Paper C is an example of a problem which is advantageous to handle with the FE method. A somewhat different application illustrating the potential of the method is the tracking of persons inside houses (Kjærgaard et al., 2010). Here, the FE method's ability to handle boundaries is important to realistically capture the geometry of the space where movement is possible. In the case of emergency situations (e.g. conflagration, earthquake) where the interior of the building may shift, the FE method could (possibly) be used to estimate the modified interior arrangement from observed person movements inside the building. Such information would be particularly valuable in planning the line of approach for entering rescue personnel.

6.2 Ecological contributions

This section emphasises the ecological relevance of the methods developed in this work. Focus is on the geolocation problem, behaviour estimation and population level analysis.

6.2.1 The geolocation problem

Estimating the geographical location of an individual is the most fundamental problem of movement data analysis and is often termed the geolocation problem. In analysing movement data with the purpose to estimate individual location it is important to acknowledge that observations are noisy and autocorrelated. SSMs not only account for autocorrelation in data, but use it to propagate strength from informative observations to uncertain observations. In addition, SSMs separate data uncertainty into movement related uncertainty and observation related uncertainty. As discussed in Section 3.1 many different types of SSM analyses have been proposed to tackle the geolocation problem. Paper B explains how the HMM method can be used to analyse SSMs for individual animal movement.

One of the main advantages of the HMM approach is that the filtering and smoothing equations are solved non-parametrically on a spatial grid. This feature is useful when geolocating movement data with spatially or temporally heteroskedastic errors as in tidal or sea surface temperature (SST) data. With these types of data the observation equation (4.4) or (4.9) becomes nonlinear

which prohibits the use of Kalman filter (KF) approaches unless approximations are made (Nielsen et al., 2006). Often, oceanographic data such as SST or tidal variations are already provided non-parametrically on a spatial grid. It is therefore straightforward to use such data with the HMM method without the need for (further) approximations.

Paper A applied the HMM framework explained in Paper B to Atlantic cod (*Gadus morhua*) in the North Sea. The North Sea has a complex amphidromic system, i.e. there are large spatial differences in the tidal pattern, which enable accurate geolocation of bottom dwelling species. Previously, good location estimates were obtained with a heuristic approach, the tidal location method, although geolocation uncertainties were not quantifiable with this method (Metcalfe and Arnold, 1997; Hunter et al., 2004). The use of HMMs was in this respect a major advance in particular because tidal data are ambiguous, i.e. identical tidal patters can occur at multiple separate locations simultaneously (see Figure 4, Paper A). If possible the HMM method eliminates multimodality in the estimated probability distributions by supplementing information at t_k with information from before and after t_k. In this way the autocorrelation between observations becomes an advantage.

A feature of the HMM method is that random tracks can be generated from the posterior distribution. This was utilised in Paper A to estimate the probability of a cod moving between different management regions. The inferential potential of random track generation is considerable because ecological events, which from a modelling perspective seem unmanageable, can be assigned a probability of occurring. Another example in this respect is the timing of migrations from one habitat to another. This can be estimated by generating a batch of random tracks and then calculating the empirical distribution of the time the fish leaves the region of interest. Figure 6.2 shows such a departure time distribution for the southward migration from 118°E, 36°S for the southern bluefin tuna analysed in Paper D (see also Figure 6.3).

The only serious alternative to the HMM method in solving highly nonlinear geolocation problems is the particle filter (PF) method described in Royer et al. (2005); Andersen et al. (2007). The PF works in continuous space and does not require a discrete gridding of the state-space. Thus, it is relatively simple to implement in its basic form. Often, however, a complete analysis requires estimation of parameters and smoothed state probabilities. While these features are possible with the PF they require prudence to incorporate. Conversely, for the HMM method the main effort lies in the implementation of the basic components (time and data update steps) whereas incorporation of subsequent parameter estimation and smoothing is simple but important.

Paper B uses the convolution approach to propagate state probabilities in time

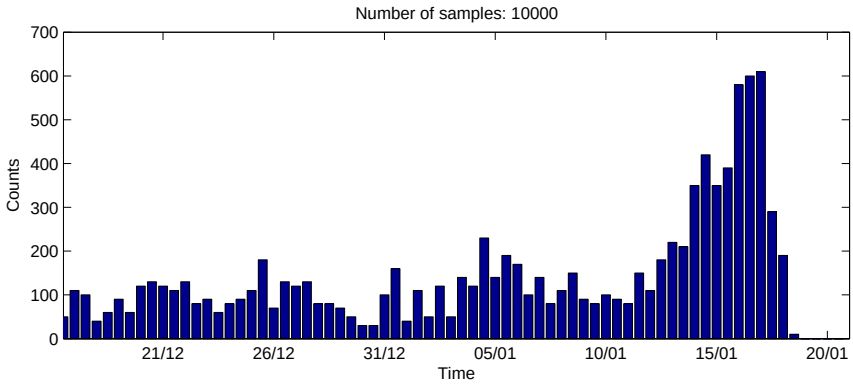


Figure 6.2: Departure time distribution of a southern bluefin tuna from a residency region at (118°E, 36°S). A total of 10000 random tracks were generated and for each track the departure time in the interval from 17 December 2003 to 22 January 2004 was noted. Clearly there is a higher probability that the fish departed the region toward the end of the time interval.

(see Section 5.1). This method is only advantageous to use if boundary conditions can be ignored because then the fast Fourier transform can provide an accurate approximate solution. Ignoring boundary conditions actually implies, in the context of the convolution operation, that absorbing boundaries are imposed at the coast lines. This leads to "leaking" of probability mass at these boundaries which is unfortunate. Reflective boundaries preserve probability mass in the domain and thus gives a more realistic solution. However, these cannot be easily implemented with the convolution approach. Having the correct boundary conditions is particularly important when data is uninformative because this increases the likelihood that boundary leaking will occur. An example of this is tracking of cod in the Baltic Sea (Paper C). Because spatial gradients in the observed data (depth and salinity) are relatively insignificant there will be a tendency to boundary leaking. Furthermore, the shoreline geometry, particularly in the Western Baltic, is tortuous with many islands and straits, which complicate matters for the convolution based HMM approach. As discussed in Section 6.1.2 the FE method is a novel technique to correctly deal with complicated boundary geometry in geolocation problems.

The distributional approach to visualising movement as presented in Paper D adds a new degree of detail to the analysis of geolocation results. The so-called residency distribution (RD, see Figure 5.4) can be calculated by summing time marginals of the posterior distribution over time. The RD summarises the overall movement of the individual and provides regional bounds corresponding

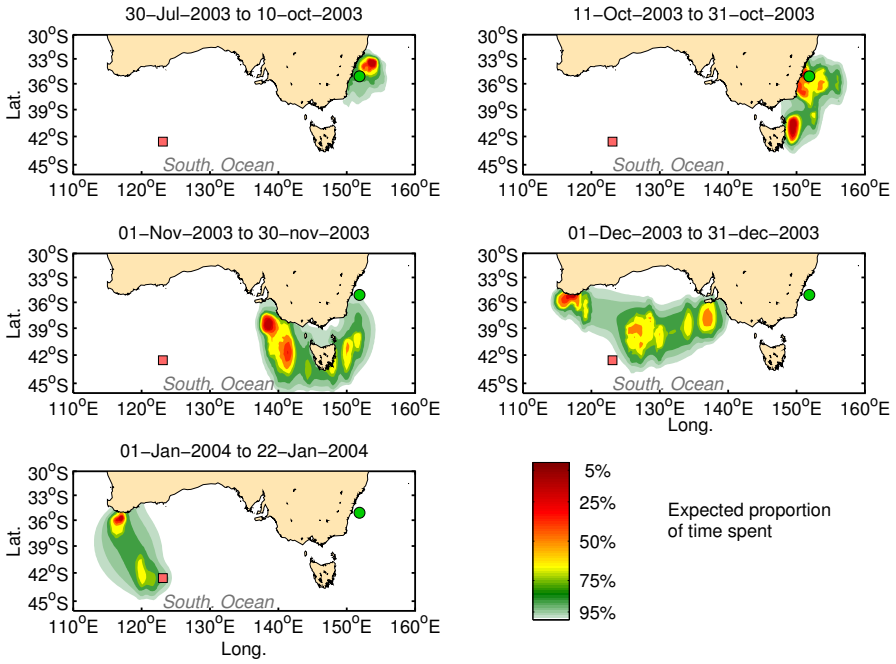


Figure 6.3: Residency distributions at different time periods for a southern bluefin tuna south of Australia. The temporally divided plots highlight the tuna's seasonal usage of space.

to the expected proportions of time spent. The RD can be separated into multiple shorter time periods to identify geographical regions that are important during different seasons (see Figure 6.3).

There is a general consensus among ecologists that the main challenges involved in the geolocation problem have been overcome by now, and that further improvements to geolocation precision are more likely to be obtained through new types of data rather than more developed methodology (Patterson et al., 2009). So, while researchers wait for acoustic, magnetic, and accelerometer data, focus in modelling has gradually shifted to the use of geolocation as a building block for behaviour and population estimation.

6.2.2 Behaviour estimation

While the location of a fish conveys ecologically useful knowledge it is also important to assess possible motifs for the animal to visit this location. Paper A categorized depth data of Atlantic cod (*Gadus morhua*) into different behaviour types prior to the geolocation analysis. This was possible because data were frequently sampled (10 minute intervals) and highly informative. For other data types (e.g. pop-up data) the situation is not so fortunate and behaviour must be estimated simultaneously with geolocation.

Methods for estimating behaviour have been developed for terrestrial animals (Morales et al., 2004) and marine mammals (Jonsen et al., 2005) that can be tracked with GPS or Argos techniques. The switching HMM method presented in Paper D is inspired by these approaches, but also works with indirect (non positional) and/or ambiguous data from archival and pop-up tags.

The simplest categorization of horizontal movement behaviour is as either resident or migratory. Commonly in the literature, behaviour estimation results are illustrated with colour coded movement trajectories. This representation does not convey the location uncertainty, which is always present in state-space modelling. Paper D conditioned the estimated RD on different behavioural states and computed realistic regional confidence bounds, which explicitly linked behaviour and location. As shown, resident behaviour often occurs in relatively confined "hot spots" whereas migratory behaviour entails higher uncertainty in location estimates and therefore wider regional bounds (see Figure 3 in Paper D).

In estimating behaviour from movement it is common in the literature to restrict estimation to only two different states even though expanding the behaviour state-space is theoretically straightforward. The migratory state has a clear biological interpretation as a transition toward a new habitat. The resident state, on the other hand, can represent several types of more specific behaviour such as localised search, spawning, resting, foraging etc. How categorization of the resident state into possible sub-states should proceed is likely to be species and data dependent. Dividing the resident behaviour into a foraging state and a resting state, would require highly detailed movement data or auxiliary covariate data (Patterson, 2009). The majority of marine animal studies lack such data, but tags with physiological sensors are becoming increasingly common.

Paper F considered a simple three state behaviour model. While the main focus of that report was on population modelling, the accurate and highly detailed (1 minute interval) acoustic telemetry data enabled pike behaviour to be categorized into "resting", "cruising" or "aggressive" (without simultaneous estimation

of movement). As in Patterson et al. (2009), the stationary distributions of the switching Markov chain was used to summarise the overall behaviour of the individuals. Resting and cruising states were much easier to detect since they are processes with longer duration and low energy cost. In contrast, aggressive movements happen in short bursts. Thus, frequent sampling is essential if such patterns are to be captured.

In the analysis of movement data the combination of SSMs and switching models is a natural way to incorporate behaviour. Since the HMM method already uses hidden discrete states in estimating location it is simple to extend it with a discrete behavioural state. It is advantageous that the switching HMM provides a probability distribution of the behavioural state instead of a strict categorization. The behaviour probabilities ensure a "soft" classification of behaviour, i.e. that the effective behaviour of the animal is a weighted average of the different behavioural states. This enables the model to capture a display of mixed behaviour.

Using covariate data (e.g. environmental, physiological, time, space) improves the ability of the framework to predict animal behaviour in different regions but under similar conditions. Estimation of increasingly complicated behaviour models, however, will entail higher computational demands and raise issues with respect to convergence and identifiability of parameters. These are important topics to address in future studies.

6.2.3 Population analysis

Ecologists are interested in jointly analysing tagging data from multiple individuals to capture population patterns. How to model a population with individual data depends on the specific purpose of the study. This thesis focused on population modelling with mixed effects models to explain variability between individual HMMs. The principle of mixed models is that when individuals from the same population are spawned they inherit the population characteristics (the fixed effect) with some individual perturbation (the random effect).

A major challenge of mixed effects modelling is how to parameterise the individual model. The simulation study in Paper F assigned random effects to the movement parameters of a biased random walk. This parameterisation was convenient in verifying the estimation properties of the mixed model HMM framework. When analysing real data, however, it can be difficult to ecologically interpret the diffusivity and advection parameters. In the three state HMM for acoustic telemetry data in Paper F the stationary state probabilities were the parameters used for population inference. These parameters can be ecologically

interpreted as indicators for the time spent in each behavioural state. With the mixed model framework it was furthermore possible to statistically identify individuals that deviated in behaviour from the rest of the tagged population. This feature is useful in the study of possibly mixed populations.

One way to link behaviour to space is via spatially dependent covariates such as spatial coordinates, SST, bathymetry, sediment type, salinity, oxygen concentration, prey availability, risk of predation etc. Which of these quantities that are relevant depends on the specific study. As shown by Patterson et al. (2009), covariate data are theoretically simple to incorporate into HMMs by allowing the transition probabilities (or transition rates in continuous time) of the Markov chain to be a function of for example SST. Including covariate information, however, leads to an increase in the number of model parameters and thus to growing computational requirements. This issue will diminish as computer technology advances, though.

An HMM supported by spatial covariates could potentially be used to identify similarities in behavioural responses to these variables among and between groups of tagged individuals. Furthermore, using covariates, population behaviour can be linked to spatial regions. All spatial covariate data are observed with error and may also be partially or entirely unobservable (for example prey field). Unfortunately, the magnitude of the observation error is rarely known except in certain cases where it can be estimated empirically (Musyl et al., 2001; Patterson et al., 2010). In addition, many of the potentially useful covariates are unevenly sampled dynamical processes with large temporal and spatial variation. For example, SST fields monitored by satellites require temporal and spatial interpolation owing to cloud cover interference. Moreover, the heterogeneity of the SST field (fronts, eddies, etc.), necessitates spatial smoothing of the satellite observations to even out small scale disturbances. Obviously, the induced error from these operations propagates through the estimation procedure and inflates the uncertainty of estimated individual movement. Unfortunately, these issues are difficult to ameliorate. Instead, their presence must be acknowledged and dealt with in the modelling phase.

Movement and behavioural processes of individual animals are also influenced by intrinsic processes (Nathan et al., 2008). The majority of the intrinsic physiological processes (at least for fishes and marine mammals) remains unobservable. However, advances in tag technology are made which, for example, enable visceral warming to be monitored and used as a covariate in behaviour modelling (Bestley et al., 2008). In this way a link between the physiological state of the animal and different environmental covariates can be mobilised and utilised for population inference.

In tagging marine animals, it is difficult, if at all possible, to take a represen-

tative sample of the population one wishes to study. Therefore, it is doubtful whether it is reasonable to make inference outside the population of sampled individuals. Still, several sub-population studies providing similar results would make extrapolation beyond the tagged population increasingly fair. Variability between individuals can be significant even at the sub-population level. It is therefore important in the design phase of the experiment to ensure that a statistically sufficient number of individuals is tagged. The problem of an insufficient sample size is crucial because if the information one seeks is not available in the data then one can never hope to extract it.

6.3 General discussion

Here, a couple of general topics relevant to the statistical analysis of movement data will be discussed.

6.3.1 Movement models

Statistical models for describing marine animal movement is a fundamental topic related to state-space modelling of movement data. In the literature the two most prevalent movement models are

- The biased random walk (BRW) (Andersen et al., 2007; Jonsen et al., 2006; Lam et al., 2008; Nielsen et al., 2006; Sibert et al., 2003). This is the movement model used in this work, see Chapter 4 and Papers A-D.
- The correlated random walk (CRW) (Johnson et al., 2008; Jonsen et al., 2005; Royer et al., 2005), which incorporates directional correlation of the movement.

The CRW has been presented as a natural descriptor of animal movement (Jonsen et al., 2005) and is sometimes parameterised by the step length and turning angle of the individual (Morales et al., 2004). An alternative parameterisation of the CRW is to extend the BRW model so the future location not only depends on the present, but also the location before that, i.e. a second-order Markov process. The step length and turning angle or alternatively the velocity are ecologically intuitive quantities. In contrast, the diffusion and advection parameters of the BRW can be more difficult to interpret, although it is possible to relate these parameters to an expected step length via (4.7).

An important difference between the two movement models is that the CRW is able to capture directional persistence. This property is advantageous in time periods where the animal displays migratory behaviour. The BRW has a similar feature in that the drift (advection) parameter will impose a trend in the movement. The difference is that the trend is constant, which makes it an inappropriate model for cyclic movements. Allowing the drift parameter to be a function of time (e.g. piecewise constant) would give more flexibility but also increase the total number of model parameters. Alternatively, by using the switching model described in Paper D temporal variation in the parameters is allowed thus increasing the realism of the model.

While the CRW is the preferred model from an ecological viewpoint it is somewhat more complicated to implement than the BRW. Using the CRW with the HMM method is possible, but requires that both location and velocity domains are discretised. Gridding four state dimensions results in a large total number of states. Even if the velocity can be coarsely discretised, memory requirements and calculation time of a CRW HMM are likely to be prohibitively high.

The simplicity of the BRW may seem an inadequate descriptor of animal movement. However, the BRW is flexible and therefore able to accommodate rapid changes in location. This makes it particularly powerful in estimating location. On the other hand, for prediction and simulation without data the BRW is too naive. Yet, by maintaining the simplicity of the fundamental building block (the movement model), estimation of complicated problems as in Paper A, C, and D is made possible.

When the information content in data increases the estimated movement is largely determined by the observations and to a smaller extent the specific model for movement. With tidal data, and to some degree pop-up data, it is unlikely that the estimated overall movement would change significantly if estimated using a CRW instead of the BRW. For the data in Paper A and Paper D though, this is a hypothetical situation since no alternative methods are able to carry out such analyses unless the problem nonlinearities are mitigated.

A third movement model also deserves attention: it has been argued that Lévy walks (random walk with Lévy distributed steps) in certain scenarios represent the optimal search strategy for animals (Viswanathan et al., 1999). The validity of this conclusion has been questioned by other authors (Benhamou, 2007; Plank and Codling, 2009) and is a topic of an ongoing debate among movement ecologists. Irrespectively, it should be noted that Lévy walks can be used in an SSM context (Sornette and Ide, 2001), however thus far no state-space analyses of marine animal movement data using Lévy walks have been published. Also, though observed movement may in some cases fit well with a Lévy model, the ecological interpretation of the estimated parameter (the Lévy exponent) is

somewhat unclear. Conversely, a switching random walk model, which is the frequently suggested alternative to the Lévy walk (Benhamou, 2007), intuitively represents two types of behaviour: extensive and intensive searching. A further advantage of switching models is that, if available, covariate data can be included to identify possible environmental or physiological triggers of behaviour shifts.

6.3.2 Prospects of tag development

Forthcoming advances in tag technology are expected to (yet again) revolutionise ecologists' understanding of the movement and behaviour of marine animals. Tags measuring magnetic field strength in three dimensions are already available and can, if gradients in the magnetic field are available, provide supplementary information to improve geolocation estimates. Along the same lines, including recordings of compass direction will relax the need to estimate turning angle when using a CRW movement model. This reduction of the state vector's dimension may significantly improve the practicability of the CRW in relation to the HMM method.

Recent developments in accelerometer technology enable electronic tags to record the acceleration of the individual and therefore also the orientation in relation to the gravity vector. This feature can aid geolocation estimates because by integrating the accelerometer signal one obtains estimates of velocity and location (Wilson et al., 2008). This, so-called dead reckoning technique, has the disadvantage that location errors cumulate over time. Therefore, correct modelling of estimation uncertainty is crucial. Accelerometer recordings could also lead to more detailed behaviour estimation since, for example for fishes, tail beat frequency could be calculated and potentially linked to different activities such as foraging or spawning.

Currently, the only way to retrieve acoustic data is via transmission to a listening buoy. Therefore, acoustic tags are primarily deployed on marine species which have a relatively limited movement range. However, recent developments have allowed tags to both transmit and receive data. These so-called business card tags (BCTs) are able to detect other acoustic tags (including BCTs) in their proximity and store the tag IDs and time of detection on board (Holland et al., 2009). Unfortunately, at the current stage of development the BCTs have considerable battery requirements. In addition the tags must be recovered physically to access the recorded data. Yet, when the technology matures these impediments are likely to be mitigated.

A BCT records encounters with other tags but is not able to spatially pinpoint

where these events occurred. Including BCT data in the geolocation method presented in this thesis would enable investigators to obtain an estimated location for the detection event. Further advances to the BCT technology would allow tags to share their information and thereby improve the probability of retrieving data from both tags. Moreover, since observations from the two tags would be correlated a joint analysis could improve individual estimates. Ultimately other tag types such as archival tags or PSATS could benefit significantly from the business card technology and greatly enhance the cost effectiveness of future tagging studies.

Studies using tags with the ability to monitor physiological variables are emerging (Bestley et al., 2008; Papastamatiou et al., 2007). Supplementing a movement and behavioural analysis with for example feeding data could potentially allow sub-categorization of the behavioural model. Then, using the visualisation techniques presented in Paper D, regional bounds for possible feeding grounds could be highlighted.

The many novel types of data also entail a growing requirement for careful experimental design of tagging studies. To facilitate the statistical analysis ground truthing and uncertainty assessment of the new data types are important to realistically evaluate the potential inference of tagging projects. Using the HMM method it is in fact relatively straightforward to assess the expected geolocation uncertainty in a region of interest prior to tagging. This can be done by discretising the region and the computing the data likelihood distribution for a hypothetical but probable observation. While this approach is simplistic it could be a first step toward rigorous methods for planning future tagging missions.

Chapter 7

Conclusions

The analysis of data from electronic tags related to the location of marine animals is a developing field of research. Numerous methods have been developed for specific species, specific types of data or specific geographical regions. Still, the field lacks a framework which integrates the facilities of the specialised methods. This thesis illustrated the versatility of hidden Markov models (HMMs) as a statistical framework for analysing movement data from electronic tags. The use of HMMs in this context also led to investigations of the relationship between stochastic differential equations (SDEs), HMMs, and partial differential equations (PDEs).

An SDE observed with noise is equivalent to an HMM in continuous-time. Thus, theory from both fields can be used in the analysis of the SDE/HMM. The Kolmogorov forward equation is a PDE which describes the evolution in time of the state probability density of an SDE. Discretising and solving this PDE with numerical methods is identical to the time-stepping procedure of a finite-state Markov chain which, in turn, is the driving mechanism of an HMM. In this work the technique of employing HMM theory to analyse a discrete state SDE observed with noise is termed the HMM method.

The HMM method was employed to analyse simulated movement data (depth recordings) to demonstrate the its ability to estimate geographical location (geolocation). Using the HMM method, a study of highly informative real data

from cod in the North Sea allowed location to be accurately estimated from tidal cycles present in the depth record retrieved from an archival tag. Furthermore, the study utilised the technique of generating random movement trajectories from the posterior distribution of the fish's location. This technique is useful for estimating the probability of events that are otherwise somewhat intangible. For example, the probability that the fish enters a specific geographical region within a certain time period.

The finite difference method is the simplest approach to solve the Kolmogorov forward equation. For geolocation problems, however, it is not always sufficiently accurate. Especially for regions close to shore lines, when analysing uninformative movement data, the probability density can be poorly approximated. Therefore, the use of the finite element (FE) method to solve the Kolmogorov equation was investigated as an alternative to finite differences. The FE method is suited for solving PDEs on irregular grids which are able to closely approximate complex boundary geometries. The approach of using the FE method to analyse SDEs is novel and is particularly suited for state-space dimensions below four.

The performance of the FE method for density estimation was compared to a sequential Monte Carlo method (particle filter) on a simple Brownian bridge problem. Archival data from a cod in the Baltic Sea contain limited movement information since gradients of depth and salinity are either small or scarce in number. Also, the shore line geometry of the western Baltic Sea is complex with numerous islands and narrow sounds. The FE method was used to analyse this difficult problem and improved the results as compared to alternative methods.

Estimation of animal movement behaviour is a complex task in general. By analysing an HMM which simultaneously estimates movement and behaviour it is possible to highlight the periods in time where the animal displays either resident or migratory behaviour. Specifically, the probability of being in either of the behavioural states can be estimated and simultaneously linked to spatial regions. This methodology was employed to classify the movement of southern bluefin tuna to reveal possible migration routes.

Most tagging studies are conducted with the intention to clarify hypotheses about certain species. Analysis of individual fish can provide this at some level. However, simultaneous analysis of data from multiple fish of the same species admits inference directly at the population level. Mixed effects modelling is an appropriate statistical method for joining individual HMM analyses. This approach was taken to gain insights into the behaviour of pike using "time of day" as covariate.

The work with HMMs inevitably branched into the study of alternative methods

for analysing nonlinear state-space models (SSMs). Arguably the most popular method within the field of ecological modelling is Markov chain Monte Carlo owing to its availability in the free software WinBUGS, which builds on the modelling language BUGS. The open-source software AD Model Builder (ADMB) is an overlooked alternative for analysing nonlinear SSMs. In a comparison of the estimation accuracy of BUGS, HMM, and ADMB it was concluded that the two latter outperformed the former. As for computational efficiency, ADMB proved to be superior.

Overall, this work illustrated the use of HMMs for advanced analysis of movement data from marine animals. In the near future increasingly informative data will be available as tag technology advances. Including physiological measures such as stomach temperature and tail beat frequency can add a new degree of detail to the estimation of behaviour. Accelerometer and compass data can improve geolocation accuracy and aid in classifying apparent resident behaviour into sub-categories such as foraging and spawning. Incorporation of such new information into existing HMM methods is uncomplicated and will hopefully provide the ecological knowledge needed to aid threatened species.

Bibliography

Aarestrup, K., F. Okland, M. Hansen, D. Righton, P. Gargan, M. Castonguay, L. Bernatchez, P. Howey, H. Sparholt, M. Pedersen, et al. 2009. Oceanic spawning migration of the European eel (Anguilla anguilla). Science **325**:1660.

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.

Andersen, K. H., A. Nielsen, U. H. Thygesen, H. H. Hinrichsen, and S. Neuenfeldt. 2007. Using the particle filter to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea, with special emphasis on determining uncertainty. Can. J. Fish. Aquat. Sci. **64**:618–627.

Benhamou, S. 2007. How many animals really do the Lévy walk? Ecology **88**:1962–1969.

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.

Block, B., H. Dewar, S. Blackwell, T. Williams, E. Prince, C. Farwell, A. Boustany, S. Teo, A. Seitz, A. Walli, et al. 2001. Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. Science **293**:1310–1314.

Bowler, D. and T. Benton. 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biological Reviews **80**:205–225.

Bucy, R. S. and K. D. Senne. 1971. Digital synthesis of non-linear filters. Automatica **7**:287–298.

Cappé, O., S. J. Godsill, and E. Moulines. 2007. An overview of existing methods and recent advances in sequential Monte Carlo. IEEE Proc. Signal Proc. **95**:899–924.

Chapman, D., E. Pikitch, E. Babcock, and M. Shivji. 2005. Marine reserve design and evaluation using automated acoustic telemetry: a case-study involving coral reef-associated sharks in the Mesoamerican Caribbean. Marine Technology Society Journal **39**:42–55.

Cook, R. D., D. S. Malkus, M. E. Plesha, and R. J. Witt. 2001. Concepts and Applications of Finite Element Analys. John Wiley & Sons Inc.

Gilks, W., S. Richardson, and D. Spiegelhalter. 2001. Markov Chain Monte Carlo in practice. CRC Press, Boca Raton, Florida.

Gordon, N., D. Salmond, and A. Smith, 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Pages 107–113 *in* IEE Proc. F. Radar Signal Process, volume 140.

Grassmann, W. 1977. Transient solutions in Markovian queueing systems. Computers and Operations Research **4**:47–53.

Graves, J., B. Luckhurst, and E. Prince. 2002. An evaluation of pop-up satellite tags for estimating postrelease survival of blue marlin (Makaira nigricans) from a recreational fishery. Fish. Bull **100**:134–142.

Grimmett, G. and D. Stirzaker. 2001. Probability and random processes. Oxford University Press, USA.

Harvey, A. 1990. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Holland, K., C. Meyer, and L. Dagorn. 2009. Inter-animal telemetry: results from first deployment of acoustic business card tags. Endangered Species Research. doi **10**.

Hunter, E., J. N. Aldridge, J. D. Metcalfe, and G. P. Arnold. 2003. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables. Mar. Biol. **142**:601–609.

Hunter, E., A. A. Buckley, C. Stewart, and J. D. Metcalfe. 2005. Migratory behaviour of the thornback ray, *Raja clavata*, in the Southern North Sea. J. Mar. Biol. Assoc. U.K. **85**:1095–1105.

Hunter, E., J. D. Metcalfe, B. H. Holford, and G. P. Arnold. 2004. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables II. Reconstruction of plaice ground tracks. Mar. Biol. **144**:787–798.

Johnson, D. S., J. M. London, M. A. Lea, and J. W. Durban. 2008. Continuous-time correlated random walk model for animal telemetry data. Ecology **89**:1208–1215.

Jonsen, I. D., J. M. Flemming, and R. A. Myers. 2005. Robust state-space modeling of animal movement data. Ecology **86**:2874–2880.

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.

Jonsen, I. D., R. A. Myers, and M. C. James. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology-Progress Series **337**:255–264.

Julier, S., J. Uhlmann, and H. Durrant-Whyte. 2000. A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Transactions on Automatic control **45**:477–482.

Kitagawa, G. 1987. Non-Gaussian state-space modeling of nonstationary time series. J. Am. Stat. Assoc. **82**:1032–1041.

Kjærgaard, M., H. Blunck, T. Godsk, T. Toftkjær, D. Christensen, and K. Grønbæk. 2010. Indoor Positioning Using GPS Revisited. Pervasive Computing pages 38–56.

Lam, C., A. Nielsen, and J. Sibert. 2008. Improving light and temperature based geolocation by unscented Kalman filtering. Fisheries Research **91**:15–25.

Mather, F., B. Rothschild, G. Paulik, and W. Lenarz. 1974. Analysis of migrations and mortality of bluefin tuna, thunnus thynnus. tagged in the northwestern atlantic ocean. Fishery Bulletin **72**:900–914.

Metcalfe, J. and G. Arnold. 1997. Tracking fish with electronic tags. Nature **387**:665–666.

Mitchell, A. and D. Griffiths. 1980. The finite difference method in partial differential equations. Wiley-Interscience.

Moler, C. and C. Van Loan. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM review **45**:3–49.

Morales, J., D. Haydon, J. Frair, K. Holsinger, and J. Fryxell. 2004. Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology **85**:2436–2445.

Musyl, M., R. Brill, D. Curran, J. Gunn, J. Hartog, R. Hill, D. Welch, J. Eveson, C. Boggs, and R. Brainard, 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. Pages 343–367 *in* J. Sibert and J. Nielsen, editors. Electronic tagging and tracking in marine fisheries. Kluwer, Dordrecht.

Nakano, H., H. Matsunaga, H. Okamoto, and M. Okazaki. 2003. Acoustic tracking of bigeye thresher shark Alopias superciliosus in the eastern Pacific Ocean. Marine Ecology Progress Series **265**:255–261.

Nathan, R., W. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. Smouse. 2008. A movement ecology paradigm for unifying organismal movement research. Proceedings of the National Academy of Sciences of the USA **105**:19052–19059.

Neuenfeldt, S., H. H. Hinrichsen, A. Nielsen, and K. H. Andersen. 2007. Reconstructing migrations of individual cod (*Gadus morhua* l.) in the Baltic Sea by using electronic data storage tags. Fish. Oceanogr. **16**:526–535.

Nielsen, A., K. A. Bigelow, M. K. Musyl, and J. Sibert. 2006. Improving light-based geolocation by including sea surface temperatures. Fish. Oceanogr. **15**:314–325.

Øksendal, B. 2007. Stochastic Differential Equations. Springer-Verlag, Berlin Heidelberg.

Papastamatiou, Y., C. Meyer, and K. Holland. 2007. A new acoustic pH transmitter for studying the feeding habits of free-ranging sharks. Aquatic Living Resources **20**:287–290.

Patterson, T., 2009. Analysis of marine animal behaviour from electronic tagging and telemetry data using state-space models. Ph.D. thesis, University of Tasmania, CSIRO Marine and Atmospheric Research.

Patterson, T., K. Evans, T. Carter, and J. Gunn. 2008*a*. Movement and behaviour of large southern bluefin tuna (Thunnus maccoyii) in the Australian region determined using pop-up satellite archival tags. Fisheries Oceanography **17**:352–367.

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.

Patterson, T., B. McConnell, M. Fedak, M. Bravington, and M. Hindell. 2010. Using gps data to evaluate the accuracy of state-space methods for correction of argos satellite telemetry error. Ecology **91**:273–285.

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008*b*. State-space models of individual animal movement. Trends in Ecology & Evolution **23**:87–94.

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

Pedersen, M. W., 2007. Hidden Markov models for geolocation of fish. Masters, Technical University of Denmark.

Plank, M. and E. Codling. 2009. Sampling rate and misidentification of Lévy and non-Lévy movement paths **90**:3546–3553.

Polovina, J., D. Kobayashi, D. Parker, M. Seki, and G. Balazs. 2000. Turtles on the edge: movement of loggerhead turtles (Caretta caretta) along oceanic fronts, spanning longline fishing grounds in the central North Pacific, 1997–1998. Fisheries Oceanography **9**:71–82.

Rabiner, L. 1989. A tutorial on hidden Markov models and selected applications inspeech recognition. Proceedings of the IEEE **77**:257–286.

Royer, F., J.-M. Fromentin, and P. Gaspar. 2005. A state-space model to derive bluefin tuna movement and habitat from archival tags. Oikos **109**:473–484.

Schiesser, W. 1991. The numerical method of lines: integration of partial differential equations. Academic Press San Diego.

Shewchuk, J. R., 1996. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. Pages 203–222 *in* M. C. Lin and D. Manocha, editors. Applied Computational Geometry: Towards Geometric Engineering, volume 1148 of *Lecture Notes in Computer Science*. Springer-Verlag. From the First ACM Workshop on Applied Computational Geometry.

Sibert, J., J. Hampton, D. Fournier, and P. Bills. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna *(Katsuwonus pelamis)*. Canadian Journal of Fisheries and Aquatic Sciences **56**:925–938.

Sibert, J. R., M. K. Musyl, and R. W. Brill. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. Fish. Oceanogr. **12**:141–151.

Skaug, H. and D. Fournier. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Computational Statistics & Data Analysis **51**:699–709.

Sornette, D. and K. Ide. 2001. The Kalman-Lévy filter. Physica D: Nonlinear Phenomena **151**:142–174.

Spiegelhalter, D., A. Thomas, and N. Best, 1999. Winbugs version 1.2 user manual.

Teo, S. L. H., A. Boustany, S. Blackwell, A. Walli, K. C. Weng, and B. A. Block. 2004. Validation of geolocation estimates based on light level and sea surface temperature from electronic tags. Marine Ecology-Progress Series **283**:81–98.

Trefethen, P. 1956. Sonic equipment for tracking individual fish. U.S. Fish. Wildlife Serv. Spec. Sci. Rep. Fish. **179**:11.

Viswanathan, G., S. Buldyrev, S. Havlin, M. Da Luz, E. Raposo, and H. Stanley. 1999. Optimizing the success of random searches. Nature **401**:911–914.

Viterbi, A. J. 2006. A personal history of the Viterbi Algorithm. IEEE Signal Processing Magazine **23**:120–142.

Ward, P. and R. Myers. 2005. Shifts in open-ocean fish communities coinciding with the commencement of commercial fishing. Ecology **86**:835–847.

Wilson, R., E. Shepard, and N. Liebsch. 2008. Prying into the intimate details of animal lives: use of a daily diary on animals. Endangered Species Research **4**:123–137.

Wilson, S. G., B. S. Stewart, J. J. Polovina, M. G. Meekan, J. D. Stevens, and B. Galuardi. 2007. Accuracy and precision of archival tag data: a multiple-tagging study conducted on a whale shark (rhincodon typus) in the indian ocean. Fisheries Oceanography **16**:547–554.

Wolfinger, R. and L. Xihong. 1997. Two Taylor-series approximation methods for nonlinear mixed models. Comput. Stat. Data An. **25**:465–490.

Yuen, H. 1970. Behavior of skipjack tuna, Katsuwonus pelamis, as determined by tracking with ultrasonic devices. J. Fish. Res. Board Can **27**:2071–9.

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

Appendix A

Geolocation of North Sea cod (Gadus morhua) using hidden Markov models and behavioural switching

---

Authors:
M.W. Pedersen, D. Righton, U.H. Thygesen, K.H. Andersen, and H. Madsen.

Published in:
Canadian Journal of Fisheries and Aquatic Sciences **65**: 2367-2377 (2008).

Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching

M.W. Pedersen, D. Righton, U.H. Thygesen, K.H. Andersen, and H. Madsen

Abstract: When geolocating fish based on archival tag data, a realistic assessment of uncertainty is essential. Here, we describe an application of a novel Fokker–Planck-based method to geolocate Atlantic cod (*Gadus morhua*) in the North Sea area. In this study, the geolocation relies mainly on matching tidal patterns in depth measurements when a fish spends a prolonged period of time at the seabed with a tidal database. Each day, the method provides a nonparametric probability distribution of the position of a tagged fish and therefore avoids enforcing a particular distribution, such as a Gaussian distribution. In addition to the tidal component of the geolocation, the model incoporates two behavioural states, either high or low activity, estimated directly from the depth data, that affect the diffusivity parameter of the model and improves the precision and realism of the geolocation significantly. The new method provides access to the probability distribution of the position of the fish that in turn provides a range of useful descriptive statistics, such as the path of the most probable movement. We compare the method with existing alternatives and discuss its potential in making population inference from archival tag data.

Résumé : Lorsqu'on fait la géolocalisation de poissons à partir de données provenant d'étiquettes à archivage, il est essentiel d'obtenir une évaluation réaliste de l'incertitude. Nous décrivons ici l'utilisation d'une méthode nouvelle basée sur l'équation de Fokker-Planck pour faire la géolocalisation des morues franches (*Gadus morhua*) dans la région de l'Atlantique Nord. Dans notre étude, la géolocalisation se base principalement sur l'appariement des patrons de marées dans les mesures de profondeur lorsqu'un poisson passe une période de temps prolongée sur le fond de la mer avec la banque de données sur les marées. Chaque jour, la méthode fournit une distribution non paramétrique de la position du poisson marqué et ainsi elle évite l'imposition d'une distribution particulière, par exemple la gaussienne. En plus de la composante tidale de la géolocalisation, le modèle incorpore deux états comportementaux, soit une activité forte et une activité faible, estimés directement à partir des données de profondeur, qui affectent le paramètre de diffusivité du modèle et améliorent significativement la précision et le réalisme de la géolocalisation. La nouvelle méthode donne accès à la distribution de probabilité de la position du poisson qui, à son tour, fournit une gamme de données statistiques descriptives utiles, telles que la piste la plus probable de déplacement. Nous comparons notre méthode avec les méthodes de rechange actuellement disponibles et discutons de son potentiel pour faire des déductions à partir de données provenant d'étiquettes à archivage.

[Traduit par la Rédaction]

Introduction

The application of advanced statistics when analysing data for tracking of marine animals has become increasingly popular during recent years. This trend is closely linked to the growing deployment of archival tags as data collectors attached to or within the tagged individual. Tags deliver highly accurate and oftentimes detailed information of the immediate environment of the host animal in the form of, e.g., depth, salinity, temperature, light, or oxygen content. These data can be used to estimate location of individuals, and so the introduction of electronic tags to the community of marine biology has spawned several geolocation studies. Heuristic methods vary in approach but typically focus on narrowing down the ensemble of possible locations by comparison of observations with outputs from environmental models (Metcalfe and Arnold 1997; Hunter et al. 2003; Neuenfeldt et al. 2007). The heuristic approaches to geolocation yield reasonable and at times accurate position estimates but do not fully exploit the autocorrelation of the observations, which, in turn, may limit the applicability of the methods when data quality is reduced.

Stochastic geolocation methods, i.e., methods assuming that the individual moves according to some stochastic process, have enabled the development of statistical tools to estimate horizontal movement of tagged fish and other animals (Nielsen 2004). The random walk process is prevalent within modelling of behavioural ecology (Okubo 1980) and has proven to be proficient in describing marine animal

Received 22 November 2007. Accepted 5 June 2008. Published on the NRC Research Press Web site at cjfas.nrc.ca on 15 October 2008.
J20280

M.W. Pedersen[1] and H. Madsen. Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
D. Righton. Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory, Lowestoft, Suffolk NR33 0HT, UK.
U.H. Thygesen and K.H. Andersen. National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark.

[1]Corresponding author (e-mail: mwp@imm.dtu.dk).

movements (Deriso et al. 1991; Sibert et al. 1999). Geolocation based on light measurements is commonly used for pelagic animals equipped with light-sensing tags (Welch and Eveson 1999) but suffers from excessive variation in latitude at periods of time close to the equinox (Musyl et al. 2001). To this end, the Kalman filter can be used to exploit the correlation of successive observations (Harvey 1989) and provide improved estimates of position. Furthermore, in assessing the uncertainty of each daily position estimate, the filter incorporates all observations to extract the maximum amount of information from the available data material (Sibert et al. 2003).

The Kalman filter relies on a Gaussian error assumption and therefore has the great advantage that it suffices to estimate the mean and variance of the position to describe the probability distribution on a given day. For geolocation of marine animals in the open ocean, the Kalman filter works well, but for fish moving close to shores, the parametric method is inadequate because it is likely to assign nonzero probability to dry land. A solution is the nonparametric particle filter method (Ristic et al. 2004), which simulates a large number of particles (fish) according to behavioural assumptions and environmental limitations, i.e., fish cannot move onto land, thereby avoiding the problem that the Kalman filter has. The framework has for geolocation purposes been applied to synthetic temperature measurements of the bluefin tuna (*Thunnus thynnus*) in the eastern Atlantic Ocean (Royer et al. 2005) and also Atlantic cod (*Gadus morhua*) in the Baltic Sea (Andersen et al. 2007). A drawback of the method is the enormous computational demands that arise owing to the number of particles that need to be simulated to obtain reliable parameter estimates (Andersen et al. 2007).

Methods of geolocation that are not based on light levels have yielded some of the longest time series of positional data to date (Hunter et al. 2005). Observations of wave patterns in depth records owing to tidal variations have proven to yield very accurate geolocations of, e.g., plaice (*Pleuronectes platessa*) (Metcalfe and Arnold 1997; Hunter et al. 2004), thornback ray (*Raja clavata*) (Hunter et al. 2005), and Atlantic cod (Righton et al. 2007; Gröger et al. 2007). These methods are based on a comparison of the observed tidal range and phase retrieved from the archival tag data with predictions from a tidal forecast model. Detection of a tidal pattern in the depth record from the tag implies an inactive fish dwelling at or very close to the seabed: the tag is thus recording the changing depth of the water column as the tide rises and falls over the fish. At other times, when fish are more active, tidal patterns are usually absent or difficult to detect with precision. Information about activity levels and changes in geographic location is fundamental to the analysis of behaviour modulation in demersal species (Righton et al. 2000) and greatly aids with objective classifications and interpretation of temporal and spatial differences in behaviour of individuals or populations (Hobson et al. 2007). It is at this point evident that, because cod and other demersal fish exhibit considerable seasonally dependent shifts in activity level (Turner et al. 2002), geolocating methods require greater sophistication to allow for time-varying changes in the behavioural state.

In the present study, we apply a direct Fokker–Planck-based methodology (FPM) using hidden Markov models to data from archival depth recorders attached to Atlantic cod in the North Sea (Pedersen 2007; Thygesen et al. 2008). The calculations were carried out on a laptop PC with the HMM geolocation toolbox (available from www.imm.dtu.dk/~mwp) for Matlab. Our aim was to obtain the most accurate recontruction of the geographic movements of cod and to describe these in terms of the estimated probability distribution of the position of the fish during its time at liberty and an estimate of the most likely route of migration. The geolocation method uses the detectable tidal patterns to partition the observed behaviour of the fish into two activity states, each with separate movement parameters that are estimated with the maximum likelihood method. Inferences on foraging–migration behaviour can then be made from the estimated parameter values, and the significance of the two-parameter model compared with the usual one-parameter model can be tested statistically in a likelihood ratio test. Using this method, we show that the uncertainty involved in fish tracking can be reduced considerably, enabling the fine-scale reconstruction of fish movements with a level of detail and information that could ultimately be used in behaviour-based models in fisheries assessment and management.

Materials and methods

In short, the geolocation technique applied here follows the principle of state–space modelling and Kalman filtering with time and data update steps but with the deviation that no assumptions are made about linearity or Gaussianity of the distribution of the states (Harvey 1989; Sibert et al. 2003; Patterson et al. 2008).

We assume that the fish performs a random walk in two-dimensional space with diffusivity D. We take the random walk to be isotropic (so that D is a scalar); we have no prior reason to believe that the fish should have a direction preference, so we elect to keep the model simple and leave the question of anisotropy to future studies. At time t, the fish has position $\boldsymbol{X}_t$ in two dimensions and we represent the estimate of this position explicitly by its probability density function $\phi(\boldsymbol{x}, t)$, a function of two-dimensional position $\boldsymbol{x} = (x_1, x_2)$ and time t. We discretize space on a quadratic grid over the North Sea; details will be given in the following section. In the filter, the time update propagates the probability density ϕ from the time of one measurement to the time of the next by solving the Fokker–Planck equation:

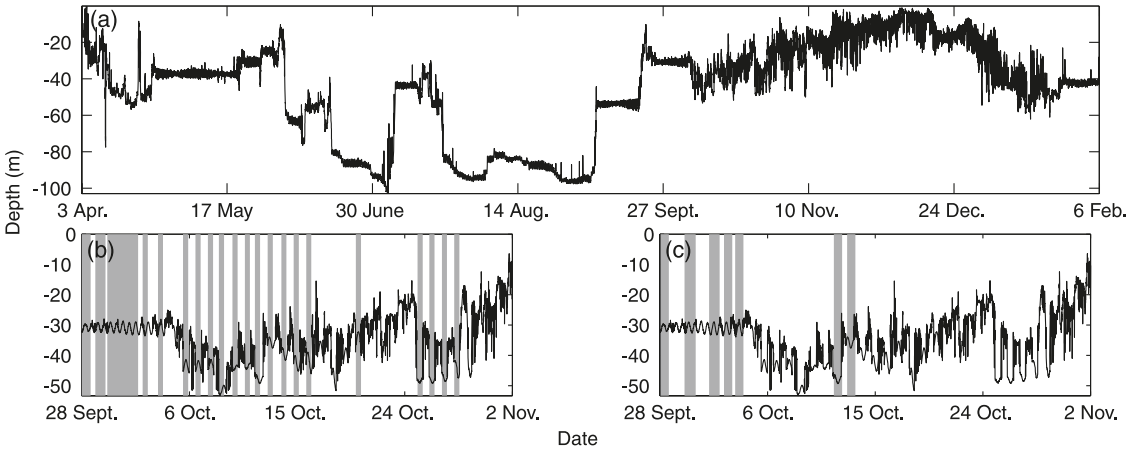
$$\frac{\partial \phi}{\partial t}(\boldsymbol{x}, t) = D\left[\frac{\partial^2 \phi}{\partial x_1^2}(\boldsymbol{x}, t) + \frac{\partial^2 \phi}{\partial x_2^2}(\boldsymbol{x}, t)\right]$$

This equation is solved numerically by finite-differences using the distribution at the previous time step as the initial condition. At time t of the next observation, some quantity $\boldsymbol{Y}$ is measured to have the value $\boldsymbol{y}$. In this paper, $\boldsymbol{Y}$ will typically be depth readings over the tidal cycle; details will be given in the following section. The filter then performs a data update using Bayes' formula to modify the probability distribution according to the information in the observation:

$$\phi(\boldsymbol{x}, t) \mapsto \frac{1}{\lambda_t}\phi(\boldsymbol{x}, t)\mathcal{L}(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X}_t = \boldsymbol{x})$$

Here, λ_t is a normalization constant, while $\mathcal{L}(\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X}_t = \boldsymbol{x})$

Fig. 1. Example of archival tag data and illustration of the tidal and behaviour classification. (*a*) Depth record from tag No. 2255. (*b*) Part of the depth record for No. 2255 classified with respect to tidal information where the shaded regions mark the detected tidal patterns. (*c*) Depth record classified with respect to activity level, where shaded area denotes low activity level, and open area denotes high activity level. Note that the fish can have a high activity level although a tidal pattern is detected, e.g., around 24 October.



is termed the data likelihood and describes the probability of the observation for each possible position. We shall elaborate on this term in the following section.

The two update steps are run recursively in a manner analogous to the Kalman filter: first forward in time and then backward to smooth the estimates (Harvey 1989). The geolocations presented here rely on observations of depth and tide (Hunter et al. 2003) complemented by the release and recapture positions. The outcome of such a geolocation is the probability distribution of the position of the fish at all time steps throughout its time at liberty. For a mathematical walk-through of the method, the reader is referred to Pedersen (2007) and Thygesen et al. (2008). Given the probability distribution, it is possible to assess simply the most probable track that the fish took during its time at liberty, i.e., the mode of the joint distribution of all positions. To this end, the Viterbi algorithm (Viterbi 2006) is applied; this recursive algorithm provides the path through the states of a hidden Markov model that has the largest overall probability given the observations.

Data

The archival tags used to collect the data material for this study were of the type DST-Centi manufactured by Star-Oddi (www.star-oddi.com) and the slightly larger LTD 1200 manufactured by LOTEK (www.lotek.com). The resolution of depth measurements from the tags is approximately 0.05 m. The tags were programmed to record temperature and pressure (converted to depth upon download) every 10 min. The temperature records were not used in the present study. Fish to be tagged were caught by hook and line and anaesthetized before tagging to minimize the traumatization of the individual. The tagging procedure is described in greater detail in Righton et al. (2006). Data were retrieved from the tags after return through the commercial or recreational fishery.

The quantity and quality of tidal patterns in the resulting cod depth records show large variation between and within individuals. For example, the data record of cod No. 2255 (Fig. 1*a*) contains periods with smooth tidal patterns, periods with noisy tidal patterns, and periods without tidal patterns, making the tag well suited for the illustrative purpose of this paper. The cod was released on 3 April 2001 at 52.44°N, 1.78°E and recaptured 87 km away on 6 February 2002 at 52.00°N, 2.85°E, yielding a total time at liberty of 311 days. To show the versatility of the method, a less optimal data set from cod No. 1186 is also geolocated, which was released on 11 March 2005 at 50.3°N, 0.5°E and recaptured 395 km away on 19 January 2005 at 53°N, 4°E, yielding a total time at liberty of 315 days.

Tidal prediction model

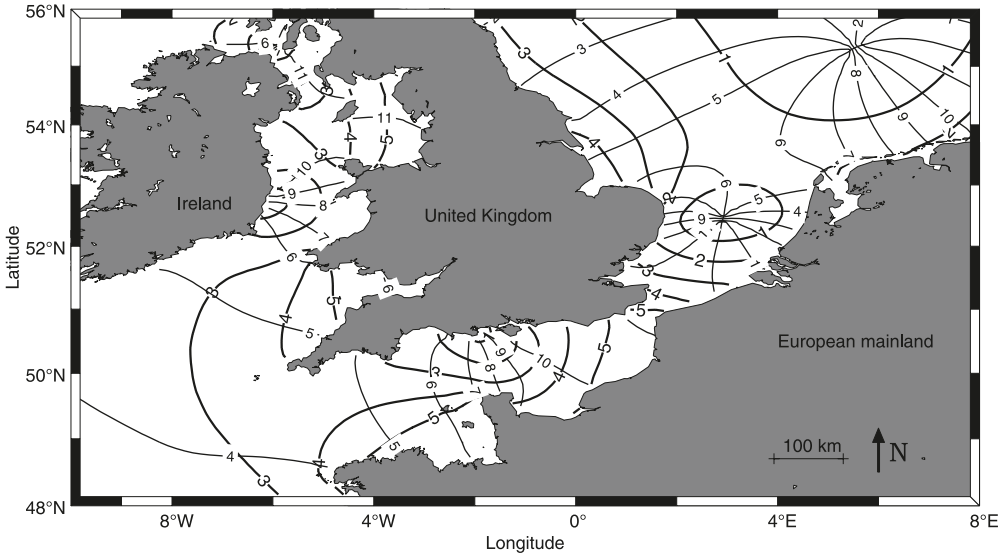
The tides observed in the North Sea are mainly due to forcing from the Atlantic Ocean through the English Channel and north of the British Isles. At a particular location and time, the tide can be predicted by numerical forecast models. Such models split the tidal variation into a number of constituents that represent the characteristic modes of the system. A superposition of all modes yields the resulting wave that approximates the one observed in practice. For a constituent k, the depth variation $z_k(t, \boldsymbol{x})$ at a fixed position, $\boldsymbol{x}$, is fully represented by the function

$$z_k(t, \boldsymbol{x}) = A_k(\boldsymbol{x}) \cos[\omega_k t - \theta_k(\boldsymbol{x}) + G_k] \quad (1)$$

where $A_k(\boldsymbol{x})$ and $\theta_k(\boldsymbol{x})$ are amplitude and phase, respectively, associated with this position $\boldsymbol{x}$, ω_k is the angular velocity, and G_k is the phase lag relative to time zero.

A forecast database from the Proudman Oceanographic Laboratory that included seven constituents (M2, S2, N2, K2, O1, K1, and M4) was used to predict tidal variations according to eq. 1. The database covers an area from 48°N to 60°N latitude and from 12°W to 8°E longitude with a reso-

Fig. 2. Amphidromic system of the North Sea here illustrated by the M2 constituent. The thick lines emanating from the amphidromic points are positions with a constant tidal phase relative to their numbering (hours). Intersecting perpendicular thin lines are positions with constant tidal range, i.e., difference between high water and low water in metres.



lution of 1/9° latitude and 1/6° longitude, approximately a 12 km × 12 km grid. The North Sea tidal system is roughly illustrated by observing its dominant constituent M2, which has a period of 12.42 h (Fig. 2).

Tidal extraction method

The observed time series was preprocessed to identify the time intervals in the depth record that contained a tidal pattern (Fig. 1). The extraction algorithm worked by sliding a 10 h window across the data and successively estimating the best least-squares fit of a sine function to the observations. The choice of window length is a trade-off: a shorter window length increases the number of successful fits but reduces the quality of each fit whereas a longer window length generates fewer fits but each with more statistical power. The window length of 10 h was chosen because it captures most of the dominant 12.4 h tidal cycle but also allows periods with tidal transport to be extracted. Of the 144 possible fits within each 24 h interval, the summary statistics root mean square error (rmse), R^2, and amplitude were extracted and the best fit (lowest rmse) was used as representative for this day. If the extracted summary statistics of the best fit fulfilled the criteria that rmse < 0.42 m, $R^2 > 0.85$, and amplitude > 0.6 m, the corresponding observed tidal pattern was stored for use in the data likelihood computation (see below). These limit values are hand-tuned parameters that were chosen so that the quality of extracted tidal signals was optimized. This tidal extraction technique bears strong resemblance to the methods applied in Hunter et al. (2003) and Gröger et al. (2007). If the criteria were not fulfilled, the maximum depth observation within the 24 h interval was stored to provide a means to confine the possible positions of the fish on the given day.

Behaviour classification

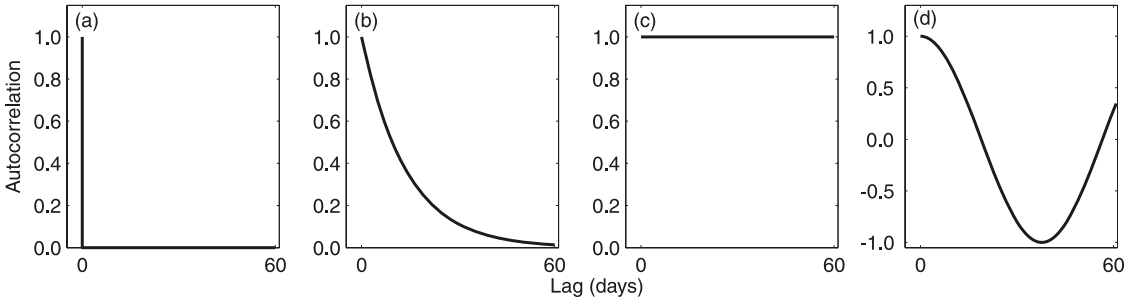
Previous studies have shown that the behaviour of cod tends to be divided into intervals of high and low activity (Righton et al. 2001). Modelling this dual-state behaviour with a single constant diffusivity would force the geolocation model to overestimate the uncertainty of the geolocation in some parts and underestimate it in other parts. In addition, if the different periods of high and low activity are not taken into account in the model, the diffusivity estimate for the entire data set will depend on the quality and type of the depth data and therefore make comparison of individuals difficult. As a partial solution to this, we extended the state–space of the system with a new state variable describing the activity level of the fish, thus making comparisons between individuals much less subjective.

The activity state is a time-dependent indicator function that, on a daily basis, is classified as either high or low. The state is, in principle, hidden (not directly observable), but to preserve the tractability of the problem, we estimated the activity state directly from the observed depth record before the actual geolocation step.

For each day in the time at liberty, the activity state of the fish is determined by testing the following hypotheses: H_0: the fish has a high level of activity (large value of diffusivity) and H_1: the fish has a low level of activity (small value of diffusivity). Only when H_0 is rejected at a sufficiently high level of significance can the small value of diffusivity be applied.

Following this thread, we construct a test to determine whether H_0 can be rejected. The test works in a way similar to the tidal extraction method by fitting a sine wave function to the observed depth in a 16 h sliding window rather than the 10 h window required for tidal data fitting. A fit con-

Fig. 3. Assumed autocorrelation contributions to the covariance structure of the observed 10 h tidal patterns: (*a*) white noise ($\mathbf{S}_E$); (*b*) autoregressive ($\mathbf{S}_\varepsilon$); (*c*) bathymetry uncertainty ($S_\eta(\boldsymbol{x})$); (*d*) tidal prediction uncertainty ($\mathbf{S}_e(\boldsymbol{x})$).



forming to the predefined limit values (see above) of rmse and R^2 rejects H_0 and implies a low level of activity in the current 24 h interval (Fig. 1). If H_0 cannot be rejected, a high level of activity is applied.

The algorithm relies on the assumption that a fish can only perform a limited migratory movement within a 24 h interval if it stays at the seabed for a continuous period of at least 16 h. Often, however, it is the case that the fish makes minor vertical excursions into the water column resulting in spikes in the tidal pattern. This makes the simple test fail, which in some cases will reject instances where a low activity would safely apply. To overcome this, the influence statistics of the 16 h fit are analysed to spot and exclude these outlying observations that deviate largely from the tidal pattern. This procedure greatly increases the detectable number of intervals with low activity within the time series and thus improves the uncertainty assessment.

Reconstructing the migration trajectory: building a data likelihood model

The data likelihood is a value computed for each position in the discrete grid describing the likelihood of the fish being in that position given the observation on the current day. The data likelihood is computed differently depending on the type of observation, i.e., it is computed either from the best extracted tidal pattern or from the maximum depth during that day if no tidal pattern was available.

Using the tidal pattern

The observed tidal pattern, denoted by the vector $\boldsymbol{Y}_j$ at day j, from a demersal fish consists of 60 depth observations (10 h fit sampled at a 10 min rate) and is assumed to follow a Gaussian distribution;

$$\boldsymbol{Y}_j \sim \mathcal{N}_{60}(\widehat{\boldsymbol{z}}_j(\boldsymbol{x}), \boldsymbol{\Sigma}(\boldsymbol{x}))$$

where $\widehat{\boldsymbol{z}}_j(\boldsymbol{x})$ is the predicted tidal pattern from the database at position $\boldsymbol{x}$ in the domain and $\boldsymbol{\Sigma}(\boldsymbol{x})$ is the covariance matrix that is a sum of four contributions:

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \mathbf{S}_E + \mathbf{S}_\varepsilon + S_\eta(\boldsymbol{x}) + \mathbf{S}_e(\boldsymbol{x})$$

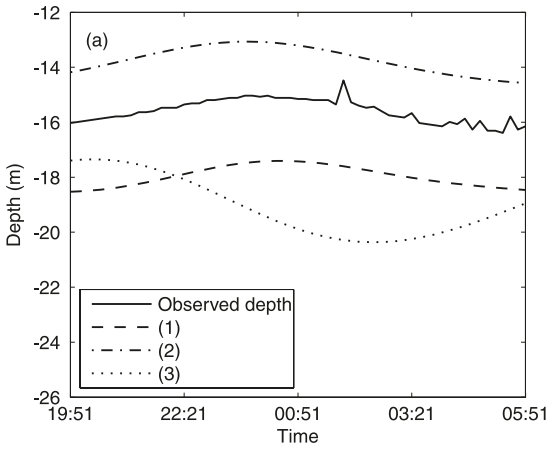
where $\mathbf{S}_E$ has a white noise structure, $\mathbf{S}_\varepsilon$ has an autoregressive structure of first order, and $S_\eta(\boldsymbol{x})$ and $\mathbf{S}_e(\boldsymbol{x})$ are the uncertainty following the discretization of the domain for the bathymetry and tide, respectively (Fig. 3). The covariance matrices are estimated from the available data material prior to maximum likelihood estimation of the movement parameters (Pedersen 2007).

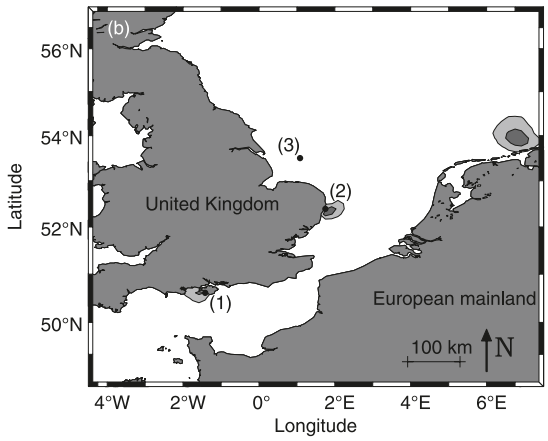
The white noise term (Fig. 3*a*) with a variance of $(0.2\ \text{m})^2$ describes the uncertainties invoked by the sensor resolution of the tag, by noise from environmental influences such as waves, and by other sources of error that are unknown and not explicitly modelled. The variance of the white noise was estimated by comparing observations from moored tags with the predicted tidal variations on the known location. In this way, the movement-related uncertainties were eliminated.

For a resident fish and a sample rate of 10 min, successive observations of depth will be correlated. Uncertainty owing to small-scale movements around rocks and holes in the seabed may therefore be modelled as an autoregressive process, $Y_i = \lambda Y_{i-1} + \varepsilon_i$ (Fig. 3*b*). As the statistical estimation of the parameter values, λ and the variance of ε_i, is not immediately feasible, we used heuristic estimates based on the assumption that the small-scale movement of the fish has decorrelated (reached an autocorrelation of <0.05) after 7 h. This results in $\lambda = 0.93$. For minor depth variations owing to small-scale movement, we conservatively set the variance of ε_i equal to $(0.4\ \text{m})^2$. More work is required to analyse the small-scale movements to confirm these assumptions, but the spatiotemporal dependence of fish movements makes this exercise complex and is best supported by more sophisticated observations, e.g., from an accelerometer; this study is beyond the scope of this paper.

A bottom-dwelling fish is likely to record a mean depth several metres off the prediction of the bathymetry when positioned in an area with a large depth gradient. This bathymetry uncertainty is estimated by comparing the depth of each grid cell with the maximal depth of its neighbouring grid cells. The estimated bathymetry variance, $S_\eta(\boldsymbol{x})$, accounts for the large-scale variation by adding a spatial-dependent but time-constant variance to the observations (Fig. 3*c*) in the range from $\sim 0\ \text{m}^2$ to $(750\ \text{m})^2$. A large value of the bathymetry variance means that the confidence of the observed depth level is reduced and hence that the geolocation at this position relies only on the tidal pattern if one is available. Analogous to the bathymetry variance, uncertainty in the tidal predictions is imposed owing to the spatial discretisation. The difference within a grid cell between tidal predictions of two distinct positions will

Fig. 4. Principle of the data likelihood computation when a tidal pattern is present. (*a*) Observed depth compared with the corresponding predictions from the tidal model at three distinct positions marked in Fig. 4*b*. (*b*) The light grey areas are the 95% confidence areas for the data likelihood. It is evident that multiple predictions fit the observed tidal signal because of the ambiguous nature of the amphidromic system. In this particular example, locations 1 and 2 fit well with the observed signal, whereas location 3 clearly does not.





show a sinusiodal waveform because of slight differences in amplitude and phase (Fig. 3*d*). This is accounted for by computing the variance of the tidal prediction, $\mathbf{S}_e(\boldsymbol{x})$, and including this in $\boldsymbol{\Sigma}(\boldsymbol{x})$. The tidal prediction uncertainty is computed by comparing neighbouring cells in a way similar to the bathymetry uncertainty. The range of $\mathbf{S}_e(\boldsymbol{x})$ is from ~ 0 m^2 to (0.97 m)^2.

The result of a data likelihood computation is an array of size equal to the discrete domain containing the likelihood of each position given the observed data (Fig. 4). Owing to the ambiguity of the amphidromic system, multiple positions typically appear equally likely although they are spatially separated. The statistical filter (Thygesen et al. 2008) and in particular the smoothing step will remove most of this multimodality by conditioning the resulting estimated probability distribution on future as well as past observations within the time at liberty.

The data likelihood at position $\boldsymbol{x}$ is written formally as

$$\mathcal{L}(\boldsymbol{Y}_j = \boldsymbol{y}_j | \boldsymbol{X} = \boldsymbol{x}) = \frac{1}{(2\pi)^{30}\sqrt{\det \boldsymbol{\Sigma}(\boldsymbol{x})}} \times \exp\left\{ -\frac{1}{2} [\boldsymbol{y}_j - \widehat{\boldsymbol{z}}_j(\boldsymbol{x})]^T \boldsymbol{\Sigma}(\boldsymbol{x})^{-1} [\boldsymbol{y}_j - \widehat{\boldsymbol{z}}_j(\boldsymbol{x})] \right\}$$

i.e., the probability density function of a 60-dimensional Gaussian distribution.

Using the maximum depth

In the absence of a tidal pattern in the observations on a given, day there remains valuable information in the depth record that can be used for geolocation. Previous studies have simply excluded positions shallower than the maximum observed depth in the tag within some uncertainty bounds (Ådlandsvik et al. 2007). Instead of a threshold, we assign in the data likelihood a value between 0 and 1 to each position dependent on its depth and bathymetry variance compared with the observed depth. This provides a more informative data likelihood than the simple indicator and exploits the important information in the bathymetry variance. This is of particular importance, as the variance of the bathymetry strongly depends on the position in that the depth of a grid cell on a slope has a high variance compared with the depth of a grid cell in a flat area.

The data likelihood computation method, inspired by the one applied in Andersen et al. (2007), assumes that the observed depth at a given position is Gaussian distributed with mean equal to the depth of the bathymetry, $z(\boldsymbol{x})$, and variance equal to the estimated bathymetry variance, $S_\eta(\boldsymbol{x})$. The likelihood of a position given a maximum observed depth (Fig. 5), $\overline{z}_j$, is found by

$$\mathcal{L}(\boldsymbol{Y}_j = \boldsymbol{y}_j | \boldsymbol{X} = \boldsymbol{x}) = \Phi\left[\frac{\overline{z}_j - z(\boldsymbol{x})}{\sqrt{S_\eta(\boldsymbol{x})}}\right] \Phi\left[\frac{-z(\boldsymbol{x})}{\sqrt{S_\eta(\boldsymbol{x})}}\right]^{-1} \quad (2)$$

where Φ is defined as the cumulated density function of a standardized Gaussian distribution with the constraint (truncation) $\overline{z}_j < 0$ and $z(\boldsymbol{x}) < 0$; hence, the normalization, at $\overline{z}_j = 0$, in eq. 2 is required because we cannot observe positive depths, and therefore, an observed depth of >0 m must always result in a data likelihood value of 1.

Supplementing with the recapture position

Typically, a recapture position is reported at the retrieval of the tag. This position may be subject to misreporting or inaccuracy. Regardless, the recapture position is of particular importance if tidal information is scarce and will effectively rule out dead ends and narrow down the estimated probability distribution towards the end of the time at liberty.

We modelled the reported recapture position as an unbiased measurement, where the error is bivariate Gaussian. We chose the variance in this distribution subjectively but conservatively. The choice of a Gaussian distribution is purely arbitrary owing to a lack of information about this

Fig. 5. Principle of the data likelihood computation when a tidal pattern is absent. (*a*) The maximum depth within the 24 h interval is used when a tidal pattern is undetectable. (*b*) Data likelihood computation for a candidate position. To determine the likelihood value, the maximum observed depth value is compared with a truncated Gaussian cumulated density function with mean and variance equal to the depth and bathymetry variance at the given position.

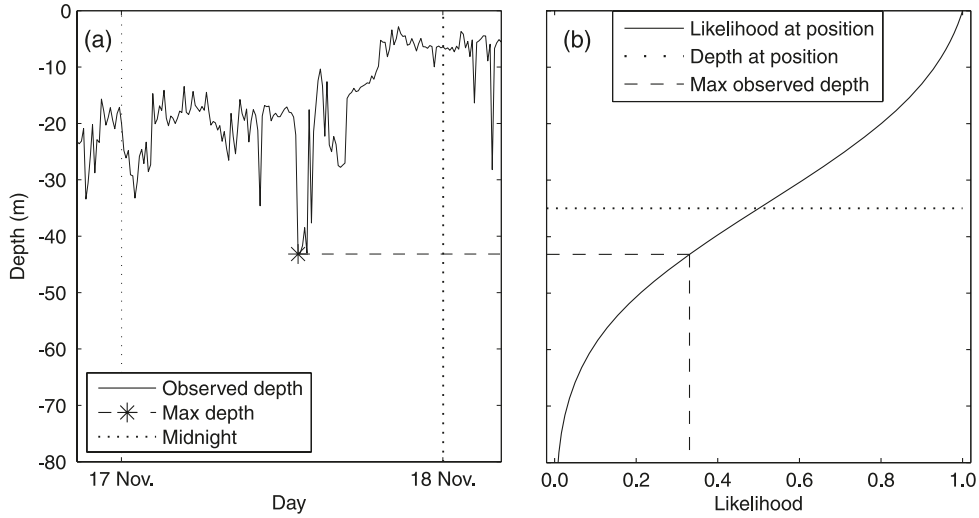
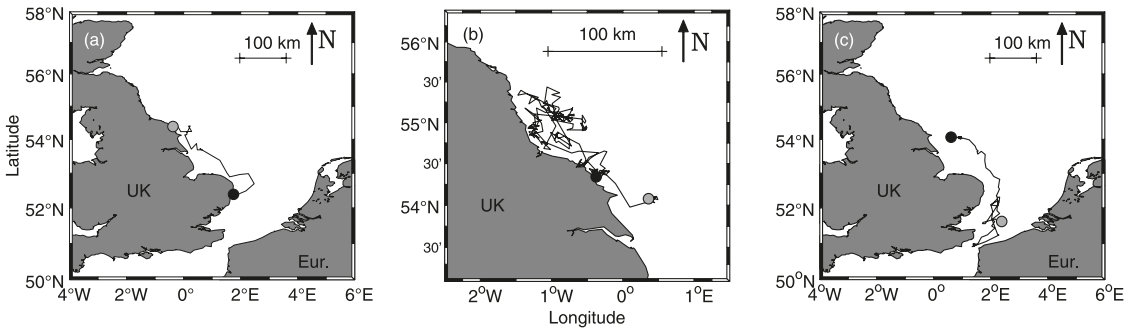


Fig. 6. Most probable track of tag No. 2255; the solid circle is the initial position of partial track and the shaded circle is the end position of the partial track. (*a*) The fish was released off Lowestoft on 3 April and migrated north to its summer residence by 30 April. (*b*) Here, it stayed until 18 November and performed only minor swimming activity during this period. (*c*) Then the fish returned to the southern North Sea and was recaptured close to the English channel on 6 February of the following year.



distribution. One might in some cases consider applying a *t* distribution with a small number of degrees of freedom to increase the probability of extreme deviations between the actual and the reported recapture.

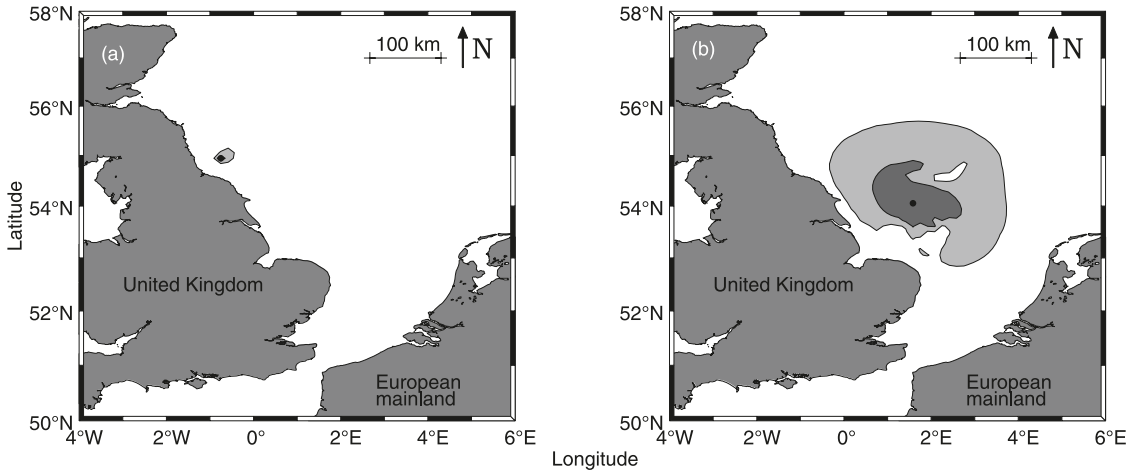
By conditioning on the position, the depth observations and the reported recapture position become independent stochastic variables. Therefore, the likelihood contributions from the two terms are simply multiplied to obtain the data likelihood function.

Results

Tag No. 2255

The cod was captured, tagged, and released close to Lowestoft, UK, and, according to the most probable track, immediately began a migration to the north, settling down after a month at approximately 54.5°N, 0.5°W (Fig. 6*a*). Here, it stayed for another month before relocating a bit farther north to an area around 55°N, 1°W, where it stayed for a prolonged period until late September at a constant depth of around 90 m (Fig. 6*b*). Then the activity level gradually increased and eventually a southwards migration brought the cod to a position at 51.75°N, 2.5°E, around 9 January and was recaptured a month later at approximately this position (Fig. 6*c*). The cod showed many long periods of inactivity, particularly during the summer when a continuous smooth tidal signal was measured spanning more than a month from late July until early September. The tidal extraction al-

Fig. 7. Probability distributions of the position of tag No. 2255 on (*a*) 23 June 2001 and (*b*) 6 December 2001 estimated in the two-mode behaviour model. Light shading is the 95% confidence region, dark shading is the 50% confidence region, and the circle is the mode of the distribution. (*a*) The distribution off the northern English coast is narrow owing to the high quality of tidal information and low activity mode of the fish in this time step. (*b*) Distribution at a time step where the fish was pelagic, i.e., no observed tidal information, thus causing the distribution to widen. The maximum recorded depth is used to exclude positions on the Dogger Bank, enforcing a hole in the distribution. This illustrates the ability of the method to estimate distributions that take on arbitrary forms.



gorithm found a total of 198 days with a tidal pattern of sufficient quality out of the 311 days that the fish was at liberty.

The maximum likelihood estimates (MLEs) of the diffusivities were 17.4 and 149 $km^2 \cdot day^{-1}$ for low and high activity, respectively, with a standard deviation of 2.69 and 28.5 $km^2 \cdot day^{-1}$, respectively. The MLEs were found using the simplex search method. The likelihood surface was smooth around the optima, which enabled us to estimate the uncertainty of the MLEs from the Fisher information. For comparison, we carried out a calculation with only one movement parameter, i.e., one behavioural state. This resulted in a MLE of the diffusivity of 57.6 $km^2 \cdot day^{-1}$ with a standard deviation of 6.6 $km^2 \cdot day^{-1}$. The statistical difference of the two parameterizations can be quantified by a likelihood ratio test. This essentially determines if a two-diffusivity model improves the likelihood of the MLE significantly compared with a one-diffusivity model, i.e., it determines whether the fish showed at least two different types of behaviour. The likelihood ratio test statistic is $Z_{LR} = 2[\ell(\widehat{D}) - \ell(\widehat{D}_0)] = 68$, where $\ell(\widehat{D}_0)$ is the log-likelihood value of the MLE in the one-diffusivity case, and $\ell(\widehat{D})$ is the log-likelihood value of the MLE in the two-diffusivity case. The test statistic Z_{LR} is χ^2 distributed with 1 degree of freedom resulting in a p value for the test of $p < 10^{-15}$, which is highly significant at all reasonable levels. This result provides evidence that No. 2255 switches its activity level in a way that is well estimated by the classification algorithm.

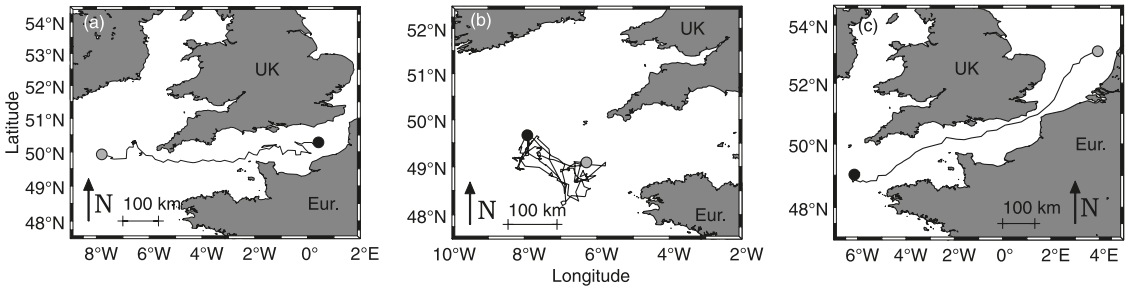
The uncertainty of the marginal distributions estimated on a daily basis depends on the diffusivity estimate and on the type and quality of data. At times, particularly in the first half of the data set when the activity level was low, the marginal distributions were very narrow (Fig. 7*a*). Often, in this period, the precision of the geolocation was limited by the discretisation (12 km × 12 km) of the domain rather than of the quality of the data. In the latter part of the data set where activity was high, the estimated marginal distributions widened owing to the lack of tidal information in the depth record (Fig. 7*b*), an unavoidable uncertainty of the position given the low quality of the information available.

Access to the probability distribution of the position allows us to sample random outcomes of this distribution, i.e., random tracks that the fish might have swum. By sampling of a batch of tracks, we can estimate the probability of the fish having visited some specific region, e.g., crossed the border of a marine protected area or the probability of the fish having picked one of many possible routes to reach a destination (Ådlandsvik et al. 2007, their figure 5*b*). Here, we sampled 1000 random paths of tag No. 2255 and found that 54 of the random paths entered ICES area VIId in the eastern English Channel towards the end of its time at liberty. Crudely, this equates to a 5.4% probability of this event, although a more robust statistical framework would be required to evaluate the significance of this result in a fisheries management context. Nonetheless, it is an indication that summary statistics of this type may be relevant to estimate if one wishes to investigate the mixing of populations and the behaviour of individuals in relation, e.g., to ICES areas.

Tag No. 1186

The estimated movement of this cod shows a similar type of periodicity as tag No. 2255 (Fig. 8). The fish was at liberty for 315 days and spent its first 3 months migrating west from its release position in the eastern English Channel (Fig. 8*a*). For 6 months, it resided at the western end of the English Channel in the Celtic Sea (Fig. 8*b*) before returning

Fig. 8. Most probable track of tag No. 1186; the solid circle is the initial position of partial track and the shaded circle is the end position of the partial track. (*a*) The fish was released in the eastern English Channel on 11 March and migrated west through the Hurd Deep area over a period of 3 months. (*b*) From 1 June to 5 December, it stayed just west of the English Channel within a relatively limited area. (*c*) Then it executed a migration east through the English Channel and into the southern North Sea in a time span of just over a month, being recaptured eventually on 19 January of the following year.



east and continuing the migration to the southern North Sea (Fig. 8*c*).

The movement away from the English Channel into the Atlantic Ocean where tidal variation is less pronounced resulted in high values of the diffusivity estimates and with large variance. In such a case of reduced data quality, it may be reasonable to use, as prior information, the diffusivity estimates from a high-quality depth record (such as No. 2255) geolocated in areas with prominent tidal variation to improve the geolocation. Without applying this correction, the apparent movement in the residing period (Fig. 8*b*) may be due to uncertainty in the tidal forecast model. To correct for this, it would be required to also consult the estimated probability distribution that expresses the geolocation uncertainty.

Again, the nature of the track with clear periods of high and low activity emphasizes the need for at least two regimes in the behaviour model. This matter is well illustrated by the return migration where a distance of 900 km is covered in 41 days as compared with the much lower activity in the middle part of the time at liberty.

Discussion

The geolocation method that we have described, termed a direct FPM, is a considerable evolution of the tidal geolocation method (TLM) described in Hunter et al. (2003). We made two fundamental advances: (*i*) successive geolocations, even those separated by many days, were linked together to create a continuous estimate of geographic location, and (*ii*) correlation of position estimates was implemented rather than treating singular positions as independent observations. This has the benefit that not only are reconstructions of migrations more precise, the reconstructions provide a genuine assessment of certainty that a sequence of independently reconstructed geographic locations can sometimes falsely convey. In addition to these advances, our method achieved greater accuracy of geolocation by taking the behaviour of individuals into account.

Accuracy of the FPM

Errors in the FPM can occur at one or more of the stages of the reconstruction process, and a considerable advantage of the method is that most of these errors can either be controlled or reported on so that an integrated assessment of the reliability of the migration reconstructions can be made. The errors that are most likely to occur are estimation of tidal parameters in the data record, error in the tidal database itself (Hunter et al. 2003), estimation of the (two different) diffusion coefficients, and error during the filtering process (Thygesen et al. 2008). For example, the mean positional error of the TLM varies between 10 and 80 km, depending on location, with the greatest errors found at locations midway between amphidromic points.

Overall, however, the errors generated by estimating tidal parameters and the uncertainty that can arise when suitable tidal data cannot be extracted are likely to be the greatest source of uncertainty in reconstructing migration pathways (Hunter et al. 2003), i.e., fish spend only brief periods of time close to the seabed and are highly mobile at other times (Hunter et al. 2006; Righton et al. 2007). In the absence of a tidal pattern, the FPM only uses the maximum depth on the given day to give some coarse geolocation. The FPM therefore addresses the problem of low information implicitly by using what knowledge that can be extracted from the depth record, and the information on prior and succeeding geolocations, to adjust the uncertainty of the geolocations. The geolocation of a given day is therefore conditioned on the information of all days and uncertainty is reduced even when the data information on the specific day is weak.

The FPM also addresses the problem that animals tend to move in irregular patterns with many small movements interspersed with the occasional large movement (Benhamou 2007; Sims et al. 2008). Our method, if necessary, derives two diffusivity values related to different activity levels of the fish to directly adjust the uncertainty of geolocation at different times. At other times, if the observations do not imply that the fish has two behaviours, the maximum likelihood estimate of the two diffusivity parameters is reduced to one diffusivity parameter. Our results show that this method is statistically robust and provides a more accurate measure of position on any given day than by using a single parameter. The model also shows that multiple activity levels are a feature of cod behaviour and migration. For data sets in which the distinction between different behaviours is less

obvious, it should still be possible to include and estimate the activity level as a hidden state within the hidden Markov model filter. However, this approach complicates matters significantly in terms of statistical implementation and computational demands and is far from a trivial task. Additionally, in theory, there is no reason why the model could not be extended to include several more levels of activity at the expense of run time, although it would not be likely to add much more information than the two-parameter model currently provides and may be very difficult to parameterize and validate. As it is, the model is already capable of providing a much more accurate and precise estimate of location and movement rate than was previously possible.

Ultimately, it is not trivial to give a single standardized measure of the accuracy of the FPM in reconstructing the migrations of cod because of the lack of data against which to validate the positional estimates. Initial studies showed that geolocation analyses of moored tags yielded consistent position estimates that were in correspondence to their true geographical position (Pedersen 2007). In this sense, the FPM produces accurate results, even though temperature data were not used to validate positional estimates. In addition, the method also produces qualitatively similar results to reconstructions made with a simulation method (Righton and Mills 2008) that uses depth and temperature data to estimate geoposition and therefore successfully captures the same overall pattern of movements of individual cod, as have been described previously (Turner et al. 2002; Righton et al. 2007). However, the relative simplicity (depth only) and transparency of the FPM gives it an advantage over the simulation method because it requires only bathymetry and a tidal database rather than a temporally and spatially resolved temperature database.

Application of the FPM

The TLM has been used to describe the migrations of plaice in the North Sea (Hunter et al. 2004), but the application of the method to other species has been limited because few other species that are large enough to be fitted with electronic tags spend sufficient time close enough to the seafloor for similar analyses to be undertaken. In addition, the TLM can be time intensive and produce multiple estimates of location that can be difficult to discriminate between. Together, these problems can hamper the reconstruction of migration pathways in cod (or any species that spends significant time away from the seabed) because suitable algorithms for processing the uncertainty have not, until now, been available (Turner et al. 2002; Hunter et al. 2003). Reconstructions of migrations of, e.g., cod to date have therefore been necessarily simplistic (Righton et al. 2007). Our reconstructions of cod migrations with the FPM showed that, even though there may be long periods of time when individuals cannot be located using the TLM, the inferential power of the FPM provides valuable daily estimates of position and the uncertainty of those estimates. The quality and frequency of the positional estimates are sufficient enough that it is easy to imagine their use within individual-based models of fish movement, therefore enabling simulations of the effect of stock movements or mixing. The nonparametric representation of the estimated probability distribution also makes the FPM a source for interesting new applications of archival tag data. The ability of the method to handle any type of archival tag data and a free choice of data likelihood computation technique can make the FPM a building block for more advanced statistical geolocation such as implementation of complex behaviour models or incorporation of robustness towards outlying position estimates, e.g., from GPS tags or from light-based tags that provide raw geolocations as output. In turn, these advances make possible new analyses of migration mechanisms and behaviours and will help to shed light on the underlying behavioural processes that govern habitat selection or foraging behaviour (Sims et al. 2006).

Models of population movement used to delineate the structure of fish stocks or changes in abundance in space and time are becoming increasingly sophisticated (Metcalfe 2006; Metcalfe et al. 2008). This has been encouraged by the requirement for "evidence-based" fisheries policies. A recurring theme of these policies, considering the difficulty of characterizing accurately the features of the marine environment, is the need for assessments of how reliable the information is and to attach an estimate of certainty to any evidence that may be used to define or support policies. At a basic level, estimating the likelihood that an individual visits a delineated area is an important first step because this has an immediate application to identifying stock identity and the risk of capture as well as to the potential utility of closed areas. A direct link to population-level models has yet to be developed for the FPM method, but a crude approach that simply averages multiple distribution estimates could be used as a first approach (Andersen et al. 2007). However, high-quality representative data sets are needed to create a statistical population model with a large number (>100) of reconstructed migrations that capture the appropriate spatial and temporal scales (Hunter et al. 2005). This applies not only at the individual level but also with respect to the experimental design of the tagging study, i.e., data spanning all seasons and possibly stratified spatially as well as with respect to age and species. One should therefore bear in mind the application of geolocation techniques when planning new studies and enhancing existing studies.

References

Andersen, K.H., Nielsen, A., Thygesen, U.H., Hinrichsen, H.H., and Neuenfeldt, S. 2007. Using the particle filter to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea, with special emphasis on determining uncertainty. Can. J. Fish. Aquat. Sci. **64**: 618–627. doi:10.1139/F07-037.

Ådlandsvik, B., Huse, G., and Michaelsen, K. 2007. Introducing a method for extracting horizontal migration patters from data storage tags. Hydrobiologia, **582**: 187–197. doi:10.1007/s10750-006-0556-7.

Benhamou, S. 2007. How many animals really do the Lévy walk? Ecology, **88**: 1962–1969. doi:10.1890/06-1769.1. PMID: 17824427.

Deriso, R.B., Punsly, R.G., and Bayliff, W.H. 1991. A Markov movement model of yellowfin tuna in the eastern Pacific Ocean and some analyses for international management. Fish. Res. **11**: 375–395. doi:10.1016/0165-7836(91)90010-D.

Gröger, J.P., Rountree, R.A., Thygesen, U.H., Jones, D., Martins, D., Xu, Q., and Rothschild, B.J. 2007. Geolocation of Atlantic cod (*Gadus morhua*) movements in the Gulf of Maine using tidal information. Fish. Oceanogr. **16**: 317–335. doi:10.1111/j.1365-2419.2007.00433.x.

Harvey, A.C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, London.

Hobson, V.J., Righton, D., Metcalfe, J.D., and Hays, G.C. 2007. Vertical movements of North Sea cod. Mar. Ecol. Prog. Ser. **347**: 101–110. doi:10.3354/meps07047.

Hunter, E., Aldridge, J.N., Metcalfe, J.D., and Arnold, G.P. 2003. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables. Mar. Biol. (Berl.), **142**: 601–609.

Hunter, E., Metcalfe, J.D., Holford, B.H., and Arnold, G.P. 2004. Geolocation of free-ranging fish on the European continental shelf as determined from environmental variables II. Reconstruction of plaice ground tracks. Mar. Biol. (Berl.), **144**: 787–798. doi:10.1007/s00227-003-1242-1.

Hunter, E., Buckley, A.A., Stewart, C., and Metcalfe, J.D. 2005. Migratory behaviour of the thornback ray, *Raja clavata*, in the southern North Sea. J. Mar. Biol. Assoc. U.K. **85**: 1095–1105. doi:10.1017/S0025315405012142.

Hunter, E., Berry, F., Buckley, A., Stewart, C., and Metcalfe, J. 2006. Seasonal migration of thornback rays and implications for closure management. J. Appl. Ecol. **43**: 710–720. doi:10.1111/j.1365-2664.2006.01194.x.

Metcalfe, J. 2006. Fish population structuring in the North Sea: understanding processes and mechanisms from studies of the movements of adults. J. Fish Biol. **69**: 48–65. doi:10.1111/j.1095-8649.2006.01275.x.

Metcalfe, J.D., and Arnold, G.P. 1997. Tracking fish with electronic tags. Nature (London), **387**: 665–666. doi:10.1038/42622.

Metcalfe, J.D., Righton, D.A., Hunter, E., and Eastwood, P. 2008. Migration and habitat choice in marine fishes. *In* Fish behaviour. *Edited by* C. Magnhagen, V.A. Braithwaite, E. Forsgren, and B.G. Kapoor. Science Publishers Inc., Enfield, N.H.

Musyl, M.K., Brill, R.W., Curran, D.S., Gunn, J.S., Hartog, J.R., Hill, R.D., Welch, D.W., Eveson, J.P., Boggs, C.H., and Brainard, R.E. 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. *In* Electronic tagging and tracking in marine fisheries reviews: methods and technologies in fish biology and fisheries. *Edited by* J. Sibert and J. Nielsen. Kluwer Academic Press, Dordrecht, the Netherlands. pp. 343–368.

Neuenfeldt, S., Hinrichsen, H.H., Nielsen, A., and Andersen, K.H. 2007. Reconstructing migrations of individual cod (*Gadus morhua* L.) in the Baltic Sea by using electronic data storage tags. Fish. Oceanogr. **16**: 526–535. doi:10.1111/j.1365-2419.2007.00458.x.

Nielsen, A. 2004. Estimating fish movement. Ph.D. thesis, Royal Veterinary and Agricultural University, Copenhagen, Denmark.

Okubo, A. 1980. Diffusion and ecological problems: mathematical models. Springer-Verlag, New York.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. 2008. State–space models of individual animal movement. Trends Ecol. Evol. **23**: 87–94. doi:10.1016/j.tree.2007.10.009. PMID:18191283.

Pedersen, M.W. 2007. Hidden Markov models for geolocation of fish. M.Sc. thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Kgs. Lyngby, Denmark. IMM Publication.

Righton, D., and Mills, C. 2008. Reconstructing the movements of free-ranging demersal fish in the North Sea: a data-matching and simulation method. Mar. Biol. (Berl.), **153**: 507–521. doi:10.1007/s00227-007-0818-6.

Righton, D., Turner, K., and Metcalfe, J.D. 2000. Behavioural switching in North Sea cod: implications for foraging strategy? ICES CM Q:09.

Righton, D., Metcalfe, J.D., and Connolly, P. 2001. Different behaviour of North and Irish Sea cod. Nature (London), **411**: 156. doi:10.1038/35075667.

Righton, D., Kjesbu, O.S., and Metcalfe, J.D. 2006. A field and experimental evaluation of the effect of data storage tags on the growth of cod. J. Fish Biol. **68**: 385–400. doi:10.1111/j.0022-1112.2006.00899.x.

Righton, D., Quayle, V., Hetherington, S., and Burt, G. 2007. Movements and distribution of cod (*Gadus morhua* L.) in the southern North Sea and English Channel: results from conventional and electronic tagging experiments. J. Mar. Biol. Assoc. U.K. **87**: 599–613. doi:10.1017/S0025315407054641.

Ristic, B., Arulampalam, S., and Gordon, N. 2004. Beyond the Kalman filter. Particle filters for tracking applications. Artech House Publishers, Boston, Mass.

Royer, F., Fromentin, J.-M., and Gaspar, P. 2005. A state–space model to derive bluefin tuna movement and habitat from archival tags. Oikos, **109**: 473–484. doi:10.1111/j.0030-1299.2005.13777.x.

Sibert, J.R., Hampton, J., Fournier, D.A., and Bills, P.J. 1999. An advection–diffusion–reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). Can. J. Fish. Aquat. Sci. **56**: 925–938. doi:10.1139/cjfas-56-6-925.

Sibert, J.R., Musyl, M.K., and Brill, R.W. 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. Fish. Oceanogr. **12**: 141–151. doi:10.1046/j.1365-2419.2003.00228.x.

Sims, D., Witt, M., Richardson, A., Southall, E., and Metcalfe, J. 2006. Encounter success of free-ranging marine predator movements across a dynamic prey landscape. Proc. R. Soc. Lond. Ser. B Biol. Sci. **273**: 1195–1201. doi:10.1098/rspb.2005.3444.

Sims, D.W., Southall, E.J., Humphries, N.E., Hays, G.C., Bradshaw, C.J.A., Pitchford, J.W., James, A., Ahmed, M.Z., Brierley, A.S., Hindell, M.A., Morritt, D., Musy, M.K., Righton, D., Shepard, E.L.C.Wearmouth, V.J., Wilson, R.P., Witt, M.J., and Metcalfe, J.D. 2008. Scaling laws of marine predator search behaviour. Nature (London), **451**: 1098–1102. doi:10.1038/nature06518. PMID:18305542.

Thygesen, U.H., Pedersen, M.W., and Madsen, H. 2008. Geolocating fish using hidden Markov models and data storage tags. *In* Tagging and tracking of marine animals with electronic devices II. Volume 8 reviews: methods and technologies in fish biology and fisheries. *Edited by* J.L. Nielsen, H. Arrizabalaga, N. Fragoso, A. Hobday, M. Lutcavage, and J. Sibert. Springer, The Netherlands. In press.

Turner, K., Righton, D., and Metcalfe, J.D. 2002. The dispersal patterns and behaviour of North Sea cod (*Gadus morhua*) studied using electronic data storage tags. Hydrobiologia, **483**: 201–208. doi:10.1023/A:1021344015515.

Viterbi, A.J. 2006. A personal history of the Viterbi algorithm. IEEE Signal Process. Mag. **23**: 120–142. doi:10.1109/MSP.2006.1657823.

Welch, D.W., and Eveson, J.P. 1999. An assessment of light-based geoposition estimates from archival tags. Can. J. Fish. Aquat. Sci. **56**: 1317–1327. doi:10.1139/cjfas-56-7-1317.

Appendix B

Geolocating fish using hidden Markov models and data storage tags

Authors:
U.H. Thygesen, M.W. Pedersen and H. Madsen.

Published in:
J.L. Nielsen et al. (eds.), *Tagging and Tracking of Marine Animals with Electronic Devices*, Reviews: Methods and Technologies in Fish Biology and Fisheries **9**: 277-293 (2009).

Geolocating Fish Using Hidden Markov Models and Data Storage Tags

Uffe Høgsbro Thygesen, Martin Wæver Pedersen and Henrik Madsen

Abstract Geolocation of fish based on data from archival tags typically requires a statistical analysis to reduce the effect of measurement errors. In this paper we present a novel technique for this analysis, one based on Hidden Markov Models (HMM's). We assume that the actual path of the fish is generated by a biased random walk. The HMM methodology produces, for each time step, the probability that the fish resides in each grid cell. Because there is no Monte Carlo step in our technique, we are able to estimate parameters within the likelihood framework. The method does not require the distribution to be Gaussian or belong to any other of the usual families of distributions and can thus address constraints from shorelines and other nonlinear effects; the method can and does produce bimodal distributions. We discuss merits and limitations of the method, and perspectives for the more general problem of inference in state-space models of animals. The technique can be applied to geolocation based on light, on tidal patterns, or measurement of other variables that vary with space. We illustrate the method through application to a simulated data set where geolocation relies on depth data exclusively.

Keywords Fish migrations · Geolocation uncertainty · Hidden Markov Model · State-space models

Introduction

A fundamental question in spatial ecology is where animals move; its answer is a prerequisite for understanding observed patterns, predicting future scenarios, or providing a knowledge base for nature management. Since movements of individual fish can rarely be observed directly on the relevant scales, we rely on indirect

U.H. Thygesen (✉)
Institute for Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark
e-mail: uht@difres.dk

J.L. Nielsen et al. (eds.), *Tagging and Tracking of Marine Animals with Electronic Devices*, Reviews: Methods and Technologies in Fish Biology and Fisheries 9,
DOI 10.1007/978-1-4020-9640-2_17,

measurements and infer positions from measurements of light, pressure, temperature, salinity, or any other environmental parameter that varies with position.

Because measurements of animal movements are indirect and often quite noisy, a geolocation algorithm is needed to infer movements from measurements. In many situations it is possible to obtain reasonable results without a formal statistical model (Metcalfe and Arnold, 1997; Shaffer et al., 2005), but even in this situation we may still prefer the statistical model because it also yields a measure of the accuracy of the geolocation.

Statistical models for geolocation are comprised of a model of movements and a model of measurements. Choosing the right model of movements is a challenge: simple models are typically unable to mimic the complex behavioral patterns of animals, while complex models typically have more parameters than can be estimated confidently. A state-space approach (Patterson et al., 2008) has the conceptual advantage that the model can be extended with additional states besides those directly linked to position, such as condition or current occupation. So far, most studies have taken the conservative approach, and used a biased random walk, e.g. (Sibert et al., 1999; Morales et al., 2004).

The statistical model of measurements depends on which parameters are measured, but specifies the likelihood of each observation for each possible position. In light-based geolocation, the measured parameter is typically the raw (unfiltered) estimate of position, which is modeled as the true position subject to a random Gaussian distributed measurement error (Sibert et al., 2003; Jonsen et al., 2005). More generally, the observation is a measurement of some parameter that varies with position, such as temperature and salinity (Andersen et al., 2007) or light levels (Nielsen and Sibert, 2007). In this case the measurement model specifies at each position a predicted measurement, and a distribution of measurement error.

When the model of movements and measurements has been formulated, the objective of the analysis is to compute the posterior distribution of the position, given the available measurements. The simplest and most efficient framework for this analysis is the linear Kalman filter (Sibert et al., 1999; see also Thygesen, 2009). The Kalman filter is restricted to the case of linear state dynamics and Gaussian errors, such as is the case when combining a biased random walk with standard light-based geolocation.

To overcome this restriction on model structure, one may use Monte Carlo methods. These techniques are, in general, flexible and straightforward, but also computer intensive. Jonsen et al. (2003) applied the Markov Chain Monte Carlo technique to geolocation. The appeal of the approach is its generality, but the resulting Markov chain may display exceedingly long mixing times. This is particularly true when the posterior probability distribution has several isolated local maxima, i.e. the animal may have followed one of several distinct routes. This means that simulation runs must be long and diagnostics must be analyzed carefully to make sure that they are long enough.

This problem is avoided with the method of particle filtering (Ristic et al., 2004; Royer et al., 2005). This is a simulation-based approach in which the distribution at each time step is represented by an ensemble of possible positions and during the

simulation, candidate positions may be removed if they do not agree well with the current observation. The method is flexible with respect to alternative movement models, but requires a large number of particles, in particular when the final position of the animal (pop-up or recapture) is known. A similar but somewhat simpler method is proposed by Ådlandsvik et al. (2007), and applied to Northeast Arctic cod.

Monte Carlo methods have an additional disadvantage when the model contains unknown parameters such as the bias (advection) and intensity (diffusion) of the random walk. In this case, we would like to do likelihood-based inference of these parameters, but Monte Carlo methods do not evaluate the likelihood function. Rather, we get a noisy measurement of the likelihood based on the stochastic simulation. In this case parameter estimation is most conveniently done in a Bayesian setting (Andersen et al., 2007).

In this paper, we discuss a different computational approach to geolocation, based on Hidden Markov Models. Our method relies on a discretisation of the state space, i.e. dividing the region of interest into small discrete cells. We model animal movement with a biased random walk and solve the corresponding advection-diffusion equations numerically on the grid. Thus, we obtain a numerical non-parametric representation of the probability distribution of the animal's position. This probability distribution illustrates the uncertainty of the estimated movement with a high degree of detail. Finally, we can draw inference about parameters in a likelihood framework, compute the most probable track of the animal, and sample random tracks that the animal may have traveled.

The technique may include information from any type of measurement. Pedersen et al. (2008) describes the details of an application to the tidal method of geolocation, but also light levels, temperature, or even recordings at listening buoys may be incorporated. In this paper, to avoid the details of the measurement process itself, we illustrate the method by applying it to a simple hypothetical example by simulating observations of pressure (depth) from a fish in the Baltic Sea.

The paper is organised as follows: In Section 'A Simulated Cod in the Baltic' we introduce the simulation model, in the context of which we shall discuss the use of Hidden Markov Models for geolocation. In Section 'The Filter Explained' we go through the steps of the filter: (1) The predictive filter, (2) the likelihood maximisation, and (3) the smoothing filter. Section 'Track Estimation' concerns estimation of tracks, i.e. the sampling of random tracks conditional on data, and the computation of the most probable track. Finally Section 'Discussion' offers a discussion.

Notation

The computational core of our method is matrix algebra. We use a matrix notation to reduce the number of subscripts and to have formulas mirror the implementation in high-level programming languages such as Matlab or R. In our notation, a function $\phi(x, y)$ of two spatial variables is discretized as a matrix Φ. See Section

'A Simulated Cod in the Baltic' for details of the discretisation. We use '$*$' to denote convolution, e.g.

$$(H * \Phi)_{ij} = \sum_{s,t} H_{st} \Phi_{i-s,j-t}.$$

We use '$\times$' to denote element-wise multiplication, $(H \times \Phi)_{ij} = H_{ij}\Phi_{ij}$, whereas '/' denotes element-wise division, $(H/\Phi)_{ij} = H_{ij}/\Phi_{ij}$. $|\Phi|$ denotes the sum of the absolute value of all elements in a matrix Φ.

A Simulated Cod in the Baltic

We consider a simulated data set, describing a demersal fish such as a cod moving in the Baltic Sea. The fish is tagged and released at a known position, after which it performs a biased random walk. We simulate the track of the fish (Fig. 1a) and construct a sequence of noisy pressure 'measurements'; since the fish is demersal this corresponds to the depth at the current position (Fig. 1b). In the end, the fish is caught at a position which is also known.

Afterwards, when we reconstruct the track, each observation thus restricts the position of the fish to the proximity of an isobath. This is a simplification of the tidal method for geolocation (Hunter et al., 2003, 2004) where pressure measurements are taken on the fish and stored in an archival tag (Righton et al., 2006).

To be more specific, we use a stochastic state-space model of the fish, where the state is the two-dimensional position $X(t)$. The displacements, or increments, $X(t+h) - X(t)$ are random vectors which are independent and each follow a Gaussian distribution with mean $u \cdot h$ and variance $2Dh$ on each coordinate. Here u is the

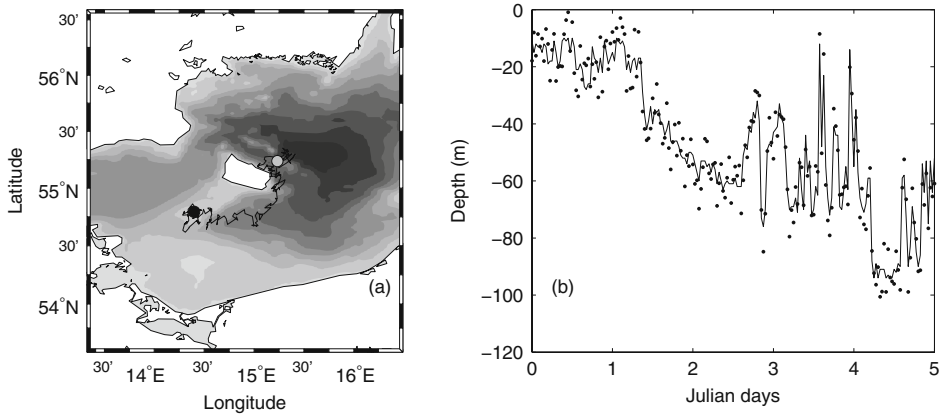


Fig. 1 The simulation model. (**a**) Bathymetry with the simulated track. *Black dot* is release position, gray dot is recapture position. The bathymetry contours are in steps of 20 m, the *darkest gray area* is −95 m. (**b**) True depth record (*solid line*) related to the simulated track and the observed depth with added noise (*dots*)

mean movement velocity vector, with units length per time, and D is the diffusivity or dispersal with units length squared per time. In general, D is a matrix or tensor, but in this paper we assume isotropic dispersal and thus D is a scalar. We use a rejection method to avoid the fish moving onto dry land.

The model can also be given in terms of the probability density $\phi(x, t)$ which determines the probability that the animal is in a given region at a given time. In absence of dry land, the time evolution of this probability distribution $\phi(x, t)$ is described by the partial differential equation

$$\dot{\phi} = -\nabla \cdot (u\phi - D\nabla\phi). \tag{1}$$

This equation is known as the Fokker-Planck equation, forward Kolmogorov equation, or the advection-diffusion equation. See Okubo (1980) for background material. Here $\dot{\phi}$ denotes time derivative while '∇' is the spatial gradient operator.

The data set contains observations of depth, $z(t_k)$, at discrete times $0 = t_1 < \ldots < t_k < \ldots < t_N = \mathrm{T}$. The measurements are simulated by adding random noise, uniformly distributed between -10 m and +10 m, to the actual depth-at-position. The main source of uncertainty is in the bathymetry database rather than in the pressure gauge or in the conversion from pressure to depth. We take the measurements to be evenly spaced in time although this is not required by our filtering method. This is important because, for example, some electronic tags are programmed to change the sample rate while deployed to optimize the use of data storage space.

The Filter Explained

In this section we turn to the question: Given a data set such as the one constructed in Section 'A simulated Cod in the Baltic', how can we estimate the actual trajectory of the fish?

To achieve this, we use a filter that follows the standard paradigm from Kalman filtering (Harvey, 1989). However, whereas traditional Kalman filtering parameterizes probability distributions in terms of mean and covariance, we discretize space and model probability distributions non-parametrically, in terms of the probability associated with each grid cell. The resulting discrete filtering problem belongs to the class of Hidden Markov Models (Cappé et al., 2005), but has a special structure because the Markov chain is a two-dimensional biased random walk.

To obtain a self-contained presentation, we describe the filtering steps explicitly. The filter involves a predictive component, running in a forward sweep over the measured time series, consisting at each time step of a time update and a data update. At the end of the forward sweep, we are able to evaluate the likelihood function of the model's unknown parameters. To perform Maximum Likelihood Estimation, we construct an outer loop around the predictive filter, maximising numerically the likelihood function. With the final estimates of the parameters, we employ the so-called smoothing filter, which performs a backward sweep over the time series. The result of this is the posterior probability distribution of the position at each time step,

based on all observations, and assuming the true parameters equals the estimated parameters. Next, we simulate random tracks from the model, conditional on the data. Finally, we compute the most probable track, which is an appropriate statistic that summarizes the data analysis.

Discretisation of the Advection-Diffusion Equation

The central element in our technique is that the equation (1) is discretised in space and time, thus obtaining a finite-state Markov chain. We discretise space using a regular quadratic grid, which means that we approximate the motion of the animal with a biased random walk on a two-dimensional lattice. At time t, the probability distribution of the position is given in terms of a matrix $\Phi(t)$ whose (i, j)-element is the probability that the individual is in cell (i, j), and which approximates the integral of the probability density function ϕ over the cell surrounding the grid point (i, j).

Since u and D are independent of position x, the probability matrix $\Phi(t)$ is governed by an evolution equation

$$\Phi(t + \Delta) = H(t, t + \Delta) * \Phi(t),$$

where '$*$' denotes convolution in 2 dimensions and $H(t, t + \Delta)$ is a matrix containing transition probabilities. This is the discrete version of the partial differential equation (1). For each grid cell, it redistributes the probability in that cell to its neighbors according to the transition probabilities in H.

We assume that the time interval is short enough that the animal can only move one cell per time step, either East, West, North or South. In this case H is a 3-by-3 matrix and it is convenient to let the row and column indeces run from -1 to $+1$. The elements of H are found from (1) by a standard finite difference discretization (Versteeg and Malalasekera, 1995)

$$H(t, t + \Delta) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{D\Delta}{h^2} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \frac{\Delta}{2h} \begin{bmatrix} 0 & u_y & 0 \\ -u_x & 0 & u_x \\ 0 & -u_y & 0 \end{bmatrix}, \tag{2}$$

where h is the spatial increment and Δ is the temporal increment. For example, the center element in the right column is $H_{0,1}\ (t,\ t + \Delta) = D\ \Delta/h^2 + u_x\ \Delta/2h$ and is the probability that the animal moves East during one time step. Since all transition probabilities must be non-negative, we require that the discretisation fulfills $|u_x|\ h/D < 2$, $|u_y|h/\mathrm{D} < 2$, and $4\Delta D/h^2 < 1$. When measurements are frequent, it is natural to take the time step equal to the distance in time between subsequent measurements, i.e. $\Delta = t_{k+1} - t_k$, but we may also use smaller time steps.

In the convolution, the boundary must be treated separately. Convolving $\Phi(t)$ with the 3-by-3 matrices $H(t,\ t + \Delta)$, we obtain a matrix with an extra row to the North and to the South, and an extra column to the East and to the West. We assume

that the domain is chosen large enough, so that we can neglect the probability that the fish passes the boundary. Thus, these extra fields are simply clipped from the matrix.

The Predictive Filter

The predictive filter is a recursion in time, running in a forward sweep. At each point of time t_k it involves two conditional probability distributions. Firstly, the distribution conditional on all observations taken at or before time t_k. We use $\Phi\ (t_k|t_k)$ to denote this distribution and note that it corresponds to the state estimate in Kalman filter terminology. Secondly, the distribution conditional on all observations taken strictly before time t_k. We use $\Phi\ (t_k|t_{k-1})$ to denote this distribution which corresponds to the state prediction in Kalman filter terminology.

The forward recursion in the predictive filter is as follows: At time t_k, we have the distribution of the estimate $\Phi\ (t_k|t_k)$ (Fig. 2a). The next observation is available at time t_{k+1}. We first perform a *time update*, which determines the next prediction (Fig. 2b)

$$\Phi(t_{k+1}|t_k) = H(t_k, t_{k+1}) * \Phi(t_k|t_k). \tag{3}$$

This corresponds to solving (1) numerically by marching time forward from t_k to t_{k+1}. Note, by comparing panels *a* and *b* in Fig. 2, that the resulting prediction $\Phi\ (t_{k+1}|t_k)$ is more diffuse than the estimate $\Phi\ (t_k|t_k)$.

We next perform a *data update*, which turns the prediction $\Phi\ (t_{k+1}|t_k)$ into an estimate $\Phi\ (t_{k+1}|t_{k+1})$ (Fig. 2d). This data update takes the observation at time t_{k+1} into account. To this end, we first construct the likelihood of the actual observation $z(t_{k+1})$, as a function of position, here termed the 'data likelihood' (Fig. 2c). With discretised space, this a two-dimensional matrix $L(t_{k+1})$ whose (i, j)-element is

$$L_{ij}(t_{k+1}) = P(Z(t_{k+1}) = z(t_{k+1})|X(t_{k+1}) = (i, j)).$$

If the observed quantity $Z(t_{k+1})$ is continuous, then this should be read as the conditional probability density function. In our case, we have measured depth with a maximal uncertainty of 10 m, and assuming that the measurement error is uniformly distributed, we obtain

$$L_{ij}(t_{k+1}) \begin{cases} 1, \text{ if depth - at - position } (i, j) \text{ differs at most 10 m from } z(t_{k+1}) \\ 0, \text{ else.} \end{cases}.$$

Note that L is a likelihood function, not a probability mass function, so needs not sum to one. Likewise, the scale of L is insignificant, so the 1 could be replaced by any other positive number. The method requires that this matrix L is available, but does not depend on how it is computed. Therefore, the method can incorporate

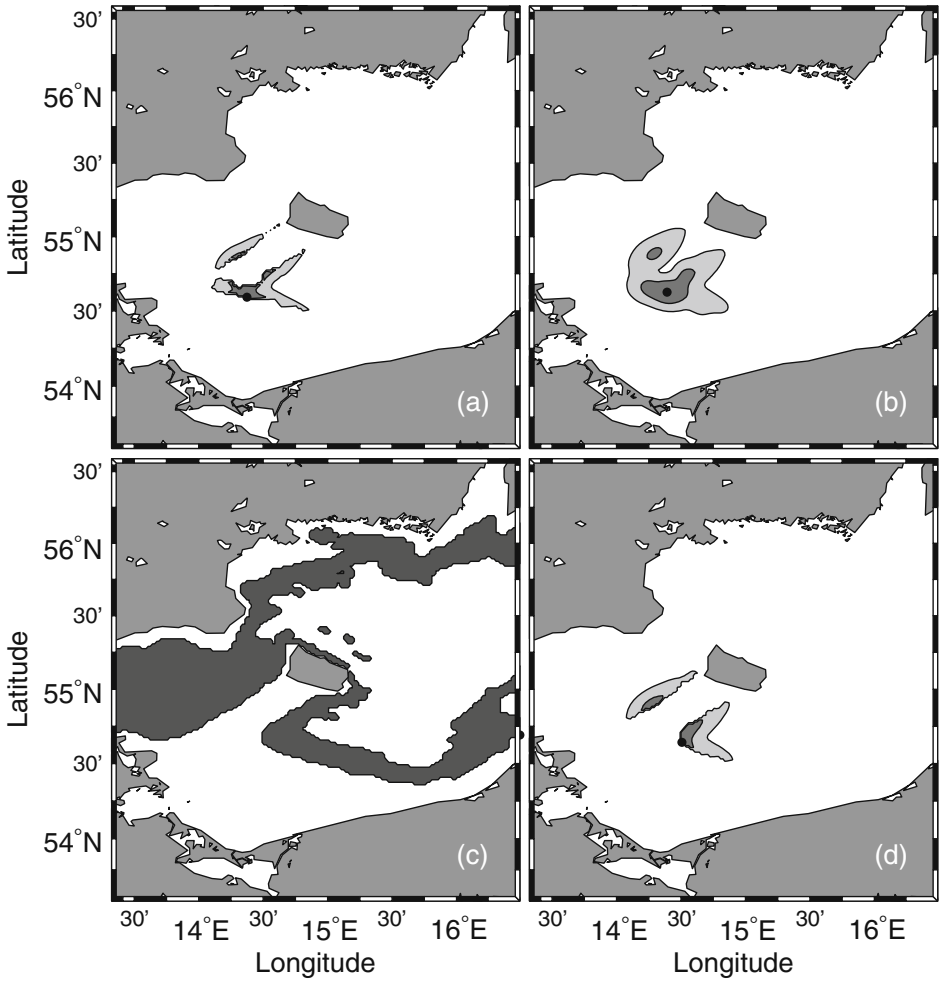


Fig. 2 The steps in the predictive filter. The tag measured at time t_{k+1} a depth of 64 meters. In (**a**), (**b**) and (**d**) dark gray is 50% confidence areas, *dark and light gray* combined is 95% confidence areas, and the *black dot* is the mode of the distribution. (**a**) Contour of $\Phi(t_k|t_k)$. (**b**) Contour of $\Phi(t_{k+1}|t_k)$, the time-update of $\Phi(t_k|t_k)$, before next observation is available. The probability field is significantly widened from (**a**) to (**b**). (**c**) The data likelihood matrix $L(t_{k+1})$. (**d**) Contour of $\Phi(t_{k+1}|t_{k+1})$, the data-update of $\Phi(t_{k+1}|t_k)$. The probability field is narrowed down due to the update of the data likelihood in (**c**)

measurements of light, tidal patterns, temperature, or any other variable that contains information about position.

Given this data likelihood, Bayes' theorem provides the conditional distribution of the state $X(t_{k+1})$, given also the new observation $Z(t_{k+1})$:

$$\Phi(t_{k+1}|t_{k+1}) = \frac{L(t_{k+1}) \times \Phi(t_{k+1}|t_k)}{\lambda(t_{k+1})}. \tag{4}$$

Here, $\lambda(t_{k+1}) = |L(t_{k+1}) \times \Phi\ (t_{k+1}|\ t_k)|$. This normalization constant is needed for the maximum likelihood estimation of the model parameters, and can be interpreted as the probability of the new observation $z(t_{k+1})$ given the observations preceding t_{k+1}.

Note that in the case of missing data, we may set all elements of L to the same constant. In this case $\Phi(t_{k+1}|t_{k+1})$ equals $\Phi(t_{k+1}|t_k)$ and the data update does nothing.

The previous describes one step in the forward iteration. We can summarize the entire predictive filter with the following:

1. Initialize $\Phi(t_1|t_1)$ with a 1 at the known position of release, and zero elsewhere.
2. Set $k = 1$.
3. Compute $\Phi(t_{k+1}|t_k)$ from $\Phi(t_k|t_k)$ and $H(t_k, t_{k+1})$ using the time update step (3).
4. Compute $\Phi(t_{k+1}|t_{k+1})$ from $\Phi(t_{k+1}|t_k)$ and $L(t_{k+1})$ using the data update step (4)
5. Set $k := k + 1$
6. If $k < N$, go to step 3. Else, stop.

Likelihood Estimation of Parameters

The underlying stochastic model typically contains unknown parameters that must be estimated. In our simulated case we use a mean swimming velocity vector u that is zero and a diffusivity D, which we shall attempt to estimate. The uncertainty parameter of ± 10 m used in the data likelihood computation is treated as known, although in real applications also such parameters must be estimated (Andersen et al., 2007).

The likelihood of a candidate parameter vector $\theta = (u_x, u_y, D)$ is

$$\mathcal{L}(\theta) = \mathbf{P}(Z_1 = z_1, \ldots, Z_N = z_N; \theta) = \prod_{k=1}^{N} \lambda(t_k),$$

where $\lambda(t_k)$ is the normalization constant obtained from a predictive filter as in Section 'The Predictive Filter' based on the parameter θ (Brockwell and Davis, 1987). This normalisation constant, in turn, is an indication of how well the observation at time t_k was predicted. Maximum likelihood estimation thus corresponds to tuning the one-step predictor.

In our simulation case, where the only unknown parameter is $\theta = D$, the maximum likelihood estimation yields an acceptable result in terms of bias and uncertainty. A more detailed study of the estimator and its statistical properties can be found in Pedersen (2007).

Smoothing of the Position

In the Kalman filtering terminology, the smoothing filter gives posterior distribution of the states $X(t_k)$ conditional on *all* data, not just previous observations. The smoothing refines the position estimate $\Phi(t_k|t_k)$ by taking all subsequent

observations into account (Fig. 3), and thus obtaining the smoothed position $\Phi(t_k|\infty)$. This is a matrix, the (i, j) element of which is the probability of being in the grid cell (i, j), conditional on all observations. $\Phi(t_k|\infty)$ is obtained as

$$\Phi(t_k|\infty) = \begin{cases} \Phi(t_k|t_k) & \text{if } k = N \\ \Phi(t_k|t_k) \times [K(t_k, t_{k+1}) * \{\Phi(t_{k+1})|\infty)/\Phi(t_{k+1}|t_k)\}] & \text{otherwise.} \end{cases} . \quad (5)$$

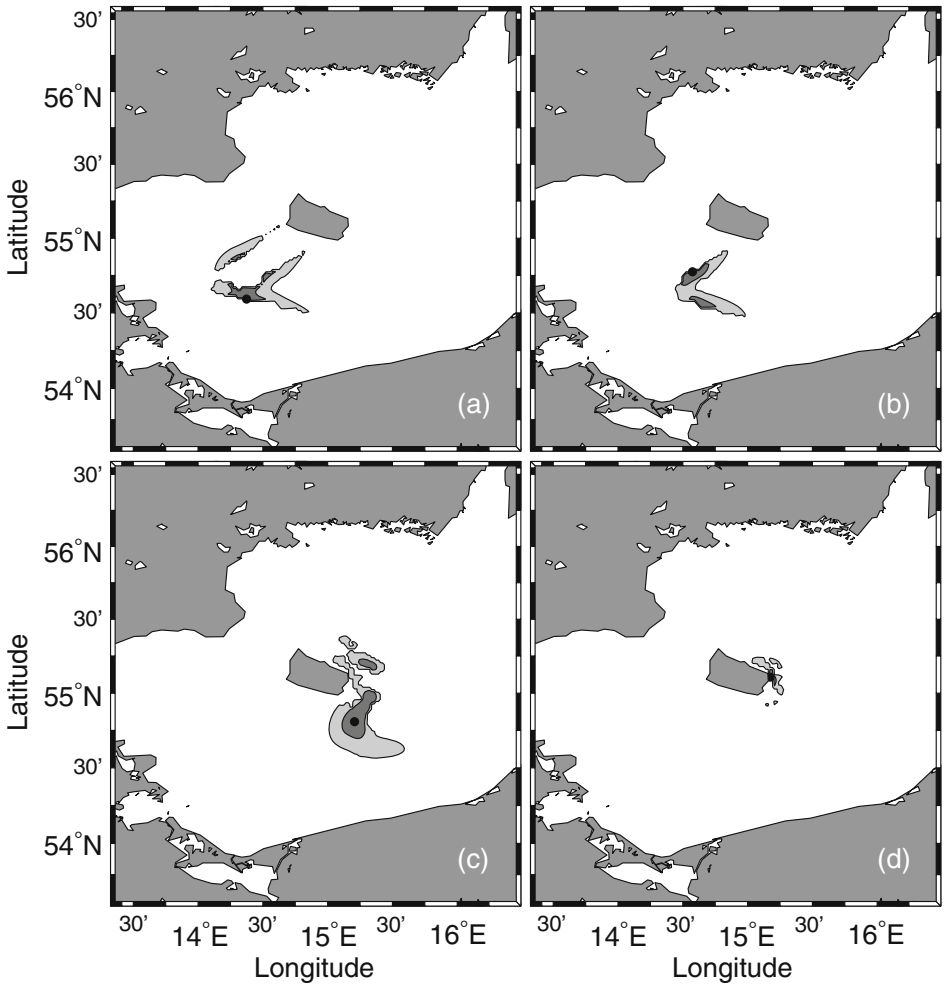


Fig. 3 The effect of the smoothing step at times t_{56} and t_{143}. *Light gray* is 95% confidence areas, *dark gray* is 50% confidence areas and the black dot is the mode of the distribution. (**a**) Contour of $\Phi(t_{56}|t_{56})$. (**b**) Contour of $\Phi(t_{56}|\infty)$. Note that apparent 'dead-ends' and outlying areas of probable positions are removed in the smoothing step due to the use of future observations. (**c**) Contour of $\Phi(t_{143}|t_{143})$. (**d**) Contour of $\Phi(t_{143}|\infty)$, note that the distribution is narrowed considerably and shifted following the smoothing step. Some parts of the 50% confidence region of $\Phi(t_{143}|\infty)$ were not even within the 95% confidence region of $\Phi(t_{143}|t_{143})$ which confirms the importance of smoothing the estimates

These $\Phi(t_k|\infty)$ are computed iteratively in a backward sweep, i.e. starting with t_N and ending with t_1.

In our case, the convolution kernel K is the mirror image of H, i.e. $K_{i,j}(t_k, t_{k+1}) = H_{-i,-j}(t_k, t_{k+1})$. This corresponds to flipping the matrix first along columns, then along rows. In general, when the mean swimming velocity u or the dispersal D varies in space, this convolution is replaced by numerical solution of Kolmogorov's backward equation.

The smoothing step can be explained as follows: At time t_k, let $\Lambda(t_k)$ be the likelihood function of all strictly future observations, viewing the position at time t_k as an unknown parameter. Then we can iterate Λ backwards in time as

$$\Lambda(t_k) = K(t_k, t_{k+1}) * (\Lambda(t_{k+1}) \times L(t_{k+1})).$$

Here, the multiplication with $L(t_{k+1})$ takes the measurement at t_{k+1} into account while the convolution with K propagates the result backwards in time, from t_{k+1} to t_k, in a discrete version of Kolmogorov's backward equation. Now, from these likelihoods we may generate the smoothed position estimates from Bayes' rule by combining the 'prior' probability $\Phi(t_k|t_k)$ with the likelihood $\Lambda(t_k)$:

$$\Phi(t_k|\infty) = \frac{\Phi(t_k|t_k) \times \Lambda(t_k)}{|\Phi(t_k|t_k) \times \Lambda(t_k)|}$$

The iteration (5) does exactly this, but eliminates Λ, which makes it slightly easier to implement.

Track Estimation

Although the posterior distributions $\Phi(t_k|\infty)$ of the position at time t_k are useful results, they do not contain all information. For example, from these distributions it is not possible to determine the probability that the fish ever entered a certain region of interest, or the probability that the fish passed a certain island to the North rather than to the South. Such questions concern the entire trajectory, not just the position at one fixed point of time, and therefore they cannot be answered from the conditional probability distributions Φ. To answer such questions, we describe in this section how to draw random samples from the joint posterior distribution, i.e. how to simulate random tracks. Given a number of such tracks, we may for example use the fraction of simulated tracks where the fish enters the region of interest as an estimate of the probability that this event happened.

A complementary problem is that the posterior distributions contain too much information; for one thing, they cannot all be displayed in a single plot panel! When the geolocation is sufficiently accurate, it is desirable to summarize the results in one single track. To this end, we describe how to compute the most probable track.

Sampling a Random Track

The movement model and the measurements in combination assign a posterior probability to each trajectory, i.e. the joint distribution of the position at all times. A random track is simulated from this joint distribution using a backward sweep:

1. Set $k = N$ and define $\Psi(t_N) = \Phi(t_N \mid \infty)$.
2. Sample a random position (I_k, J_k) from the distribution $\Psi(t_k)$.
3. Set $k := k - 1$
4. Compute the conditional distribution of the position at time k, where the conditioning is on previous data *and* on the sampled position at time $k + 1$:

$$\Psi(t_k) = \Phi(t_k|t_k) \times [K(t_k, t_{k+1}) * \{\delta(t_{k+1})/\Phi(t_{k+1}|t_k)\}].$$

   where

$$\delta_{ij}(t_{k+1}) = \begin{cases} 1 & \text{if } \ i = I_{k+1},\ j = J_{k+1} \\ 0 & \text{otherwise} \end{cases}$$

5. If $k > 1$ then go to step 2.

When the sweep concludes with $k = 1$, we have obtained one simulated trajectory, and we can then check if it enters a specific region, or if it passes a specific island to the North or to the South. To estimate the probability that this event happened, we simulate a number of trajectories, and use the fraction of trajectories for which the event occurred. The number of simulated trajectories should be chosen so that sufficient statistical accuracy is obtained. To assess this it may be used that the variance on the estimated probability is $p(1 - p)/M$ where p is the true probability and M is the number of simulated tracks.

The Most Probable Track

Our objective is now to find the track that maximizes the total probability, i.e. we find the mode of the joint posterior probability distribution of all positions (Fig. 4). Finding this track is a dynamic optimization problem; in the context of Hidden Markov Models the algorithm for solving such problems is known as the Viterbi algorithm (Viterbi, 2006).

First, consider a candidate track $(x_1, x_2, \ldots, x_N)$ where each position x_k is in the discrete two-dimensional grid, $x_k = (i_k, j_k)$, and x_1 is the known position at release. The joint probability of this track *and* the observations can be written

$$\prod_{k=1}^{N-1} P(X(t_{k+1}) = x_{k+1}|X(t_k) = x_k) L_{x_{k+1}}(t_{k+1})$$

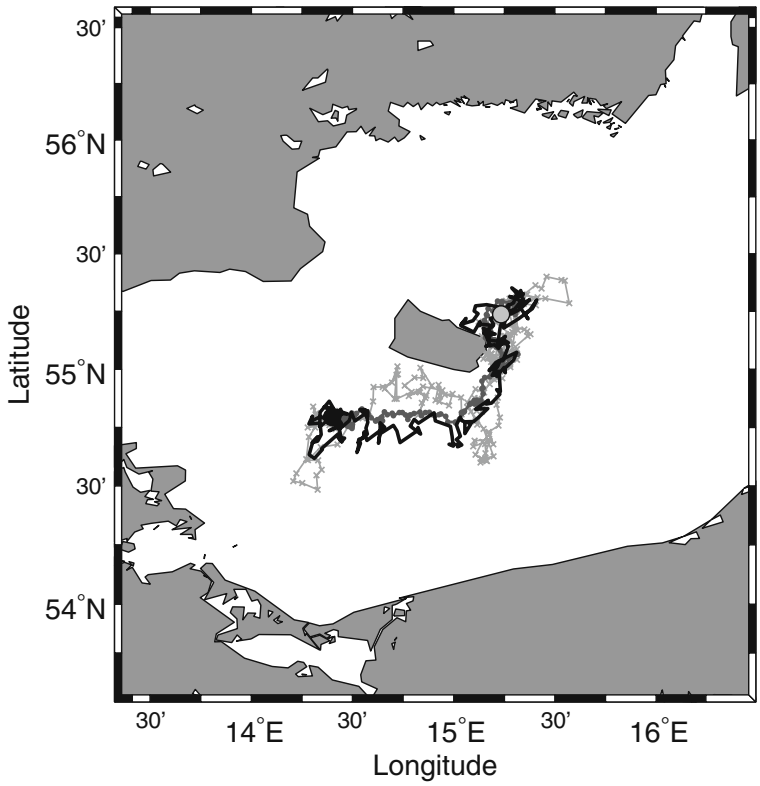


Fig. 4 True simulated track (*black solid, irregular*), most probable track (*dark grey, smooth*) and two typical tracks sampled at random from the posterior distribution (*light grey, irregular*). The most probable track is a smooth track which represents the overall behaviour of the fish whereas the sampled tracks show the likely variations of this general trend. This is particularly evident in the part where the fish moves south of the island. The slight spatial variation of the bathymetry in this area increases the uncertainty of the geolocation and allows the sampled tracks to display appreciable deviations from the most probable track

The first term corresponds to the state transitions, while the second term corresponds to the observations. Up to a normalization constant, this probability is identical to the conditional probability of the track given the observations. Take logarithms and define the *branch metric* as

$$B_{x,y}(k) = \log \mathrm{P}(X(t_k) = y | X(t_{k-1}) = x) + \log L_y(t_k).$$

Now, finding the most probable track means to solve the maximization problem

$$\max_{x_2,\ldots,x_N} \sum_{1=2}^{N-1} B_{x_k,x_{k+1}}(k+1)$$

Instrumental in the solution is the *state metric S*, defined as the log-probability of the first part of a track, $x_1, \ldots, x_n$, with $n \leq N$:

$$S_{ij}(n) = \max_{x_2,\ldots,x_{n-1}} \sum_{k=2}^{n} B_{x_{k-1},x_k}(k) \quad \text{with} \quad x_n = (i, j)$$

What allows us to solve the optimization problem is that this function satisfies Bellman's dynamic programming equation (Bertsekas, 1995):

$$S_{lm}(n+1) = \max_{ij}\{S_{ij}(n) + B_{ij,lm}(n)\}.$$

This leads to the following recursive algorithm: First, we initialize S_{ij} (1) to be 0 at the known position of release, and $-\infty$ elsewhere. Next, we let n iterate from $n = 1$ to $N - 1$, at each step computing $S(n + 1)$ according to (6). This yields the state metric $S(N)$ and from this we find the end point of the most probable track:

$$(I_N, J_N) = \arg\max_{i,j} S_{ij}(N)$$

Finally, we find the rest of the most probable track, letting n iterate downwards from $N - 1$ to 1, at each step finding the position at time t_n as

$$(I_n, J_n) = \arg\max_{ij}(S_{ij}(n) + B_{ij,I_{n+1}J_{n+1}}(n))$$

This iteration stops at $n = 1$ and yields the most probable track.

As is evident, the computation of the most probable track is somewhat involved, and it is therefore worthwhile to consider simpler alternatives. In the case of the linear Kalman filter, the most probable track at time t_k coincides with the conditionally expected position. However, since the original motivation for departing from the linear Kalman filter was harsh non-linearities, such as dry land, one should use caution when summarising the posterior distribution with its expectation. Indeed, multimodal distributions are often seen in real world applications. In Fig. 2d the fish is likely to be on the slopes around the ledge Southwest of the island of Bornholm (compare with the bathymetry in Fig. 1a), but the fish is too deep to be on the ledge itself. The expected position, which is at the middle of the ledge, is a bad representation of the most probable position. The same phenomenon is seen when using the tidal method to geolocate cod in the North Sea (Pedersen et al., 2008) when the fish migrates past shallow banks.

An alternative could be to use the mode of the distribution at each time step. But when connecting these modes, the resulting track can be very improbable. For example, it can display large jumps over short time intervals, when the posterior distribution has several competing local minima.

In summary, when the geolocation problem is reasonably close to linear, then plausible tracks can be obtained by combining either the posterior means or the posterior modes. These tracks should always be inspected carefully and treated with suspicion if they display residence in unlikely regions, or discontinuities.

Particular in such situations, but also in general when non-linearities effect the geolocation problem, it is recommendable to compute the most probable track as described here, and compare with simulated typical tracks to examine the accuracy of the geolocation.

Discussion

In this paper we have considered the reconstruction of fish tracks based on data storage tags. We have proposed a method based on direct numerical solution of the advection-diffusion equation by discretizing the two-dimensional space. The discretised problem is then solved using standard methodologies for Hidden Markov Models. The new contributions in a geolocation context are the computation of the probability distributions for the position of the animal at each point of time, in typical tracks sampled from the joint posterior distribution, and in the single most probable track as the mode of the joint distribution.

The resulting framework has several advantages when compared to previously used techniques. The non-parametric representation of the posterior distribution makes is particularly well suited to fish approaching coastal regions. In such situations the posterior distributions are not well approximated by Gaussians, as an extended Kalman filter would assume. Compared to a particle filter, it is an advantage to avoid the stochastic simulation, as such a simulation step makes it difficult to maximise the likelihood function. With a particle filter it is also not at all trivial to perform the steps of smoothing, sampling typical tracks, and identifying the most probable track. In turn, all these techniques do offer a consistent statistical assessment of the accuracy in geolocation, in contrast to early deterministic approaches.

In our simulation study we based the geolocation on depth records. We made this choice to stress that the method does not require 'raw' position estimates. Indeed, the method is able to include and combine information from different sources, such as light, temperature, tidal patterns, and presence/absence, similarly to other state-space methods (Patterson et al., 2008). The filtering method can also be used to remove outliers from otherwise accurate position measurements, such as obtained by GPS, a problem that has been considered by Jonsen et al. (2005).

The crucial part in our model framework is the Fokker-Planck equation (1). For conceptual and computational simplicity, we took the parameters in this model to be constant in space and time. This prior model of movements may seem oversimplistic as one would anticipate bias, spatial heterogeneity, temporal patterns, as well as temporal autocorrelation on multiple scales. Although this is a reasonable objection, it should be kept in mind that the actual behaviour of fish is far more complex than any model posed so far, and that this complexity cannot be estimated with confidence. Now, and in the foreseeable future, there is a trade-off between fidelity and identifiability of models.

If the advective velocity u, or the diffusivity D, in (1) had changed with space or time, then the time update (3) would consist of solving the advection-diffusion equation (1) numerically. Using the finite element method or the finite volume

method (Versteeg and Malalasekera, 1995) would allow for irregular gridding of space, making the computational mesh follow coast lines and be refined in specific areas of interest. However, it is difficult to parameterize the dependence of u and D with space and time with enough flexibility to fit real patterns while maintaining statistical feasibility. For this reason Pedersen et al. (2008) used vertical movements to distinguish between 'idle' and 'active' periods with different diffusivity, thus improving geolocation.

A particularly important question from an ecological and management point of view is how to synthesize the geolocation of several individuals. For example, are they synchronised in time, or in space, or neither? Which patterns are consistent over the population, and which are specific to the individual? One approach is to consider the individual movement parameters as a realisation of random variables related to the population to which they belong (Jonsen et al., 2003). This leads to a hierarchical random-effects model. A similar question is how to combine the posterior distributions of several individuals to draw conclusions on a population level. The obvious first step is to compute population densities by averaging individual probability distributions (Andersen et al., 2007): it is less clear how to move beyond this. In summary, it remains a formidable challenge to disentangle the many factors affecting a tagging experiment, and to draw conclusions about the species and the population. To meet this challenge requires efforts in both statistical design of experiments as well as analysis of results.

A fundamental challenge for fisheries oceanography is our limited knowledge of the actual movements and behaviour of marine fish. This impairs our ability to predict the response of marine ecosystems to changed environmental conditions, for example due to human exploitation or climate. With the technological development of data storage tags, and with the growth in numbers and size of tagging programmes, it is reasonable to predict that fisheries oceanography in the future will rely increasingly on knowledge obtained with data storage tags. This underlines the need for analysis tools which are statistically firm, yet operational. We believe that the framework proposed in this paper is one such tool.

References

Ådlandsvik, B., G. Huse, and K. Michaelsen (2007). Introducing a method for extracting horizontal migration patterns from data storage tags. *Hydrobiologia* 582: 187–197.

Andersen, K. H., A. Nielsen, U. H. Thygesen, H. H. Hinrichsen, and S. Neuenfeldt (2007). Using the particle filter to geolocate atlantic cod (*Gadus morhua*) in the Baltic Sea, with special emphasis on determining uncertainty. *Can. J. Fish. Aquat. Sci*. 64(4): 618–627.

Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control*, Vol. 1–2. Athena Scientific, Belmont, Mass.

Brockwell, P. and R. Davis (1987). *Time Series: Data Analysis and Theory*. New York.

Cappé, O., E. Moulines, and T. Ryden (2005). *Inference in Hidden Markov Models*. Springer.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hunter, E., J. N. Aldridge, J. D. Metcalfe, and G. P. Arnold (2003). Geolocation of free-ranging fish on the european continental shelf as determined from environmental variables. *Mar. Biol.* 142: 601–609.

Hunter, E., J. D. Metcalfe, B. H. Holford, and G. P. Arnold (2004). Geolocation of free-ranging fish on the european continental shelf as determined from environmental variables ii. reconstruction of plaice ground tracks. *Mar. Biol.* 144: 787–798.

Jonsen, I. D., J. M. Flemming, and R. A. Myers (2005). Robust state-space modeling of animal movement data. *Ecology* 86(11): 2874–2880.

Jonsen, I. D., R. A. Myers, and J. M. Flemming (2003). Meta-analysis of animal movement using state-space models. *Ecology* 84(11): 3055–3065.

Metcalfe, J. D. and G. P Arnold (1997). Tracking fish with electronic tags. *Nature* 387: 665–666.

Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology* 85(9): 2436–2445.

Nielsen, A. and J. R. Sibert (2007). State-space model for light-based tracking of marine animals. *Can. J. Fish. Aquat. Sci*. 64: 1055–1068.

Okubo, A. (1980). *Diffusion and Ecological Problems: Mathematical Models*. Springer.

Patterson, T.A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Mathhiopoulos (2008). State-space models of individual animal movement. *Tree* 23(2): 87–94.

Pedersen, M. W. (2007). Hidden markov models for geolocation of fish. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark.

Pedersen, M. W., D. Righton, U. H. Thygesen, K.H. Andersen, and H. Madsen (2008). Geolocation of north sea cod using hidden markov models and behavioural switching. *Can. J. Fish. Aquat. Sci*. 65: 2367–2377.

Righton, D., O. S. Kjesbu, and J. D. Metcalfe (2006). A field and experimental evaluation of the effect of data storage tags on the growth of cod. *J. Fish Biol.* 68: 385–400.

Ristic, B., S. Arulampalam, and N. Gordon (2004). *Beyond the Kalman Filter. Particle Filters for Tracking Applications*. Artech House.

Royer, F., J.-M. Fromentin, and P. Gaspar (2005). A state-space model to derive bluefin tuna movement and habitat from archival tags. *Oikos* 109(3): 473–484.

Shaffer, S. A., Y. Tremblay, J. A. Awkerman, R. W. Henry, S. L. H. Teo, D. J. Anderson, D. A. Croll, B. A. Block, and D. P. Costa (2005). Comparison of light-and sst-based geolocation with satellite telemetry in free-ranging albatrosses. *Mar. Biol.* 147(4): 833–843.

Sibert, J. R., J. Hampton, D. A. Fournier, and P. J. Bills (1999). An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna (*Katsuwonus pelamis*). *Can. J. Fish. Aquat. Sci*. 56: 925–938.

Sibert, J. R., M. K. Musyl, and R. W. Brill (2003). Horizontal movements of bigeye tuna (*Thunnus obesus*) near hawaii determined by Kalman filter analysis of archival tagging data. *Fish. Oceanogr.* 12: 141–151.

Thygesen, U. H. and A. Nielsen (2009). Lessons from a prototype geolocation problem. In J. L. Nielsen et al. (eds.), Tagging and Tracking of Marine Animals with Electronic Devices, Reviews: Methods and Technologies in Fish Biology and Fisheries 9, DOI 10.1007/978-1-4020-9640-2. Springer.

Versteeg, H. K. and W. Malalasekera (1995). *An Introduction to Computational Fluid Dynamics: The Finite Volume Method*. Prentice Hall, Harlow, England.

Viterbi, A. J. (2006). A personal history of the Viterbi Algorithm. *IEEE Signal Processing Magazine* 23(4): 120–142.

Appendix C

Nonlinear tracking in a diffusion process with a Bayesian filter and the finite element method

---

Authors:
M.W. Pedersen, U.H. Thygesen and H. Madsen.

Published in:
Computational Statistics and Data Analysis **55**: 280-290 (2011).

Computational Statistics and Data Analysis 55 (2011) 280–290

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Nonlinear tracking in a diffusion process with a Bayesian filter and the finite element method

M.W. Pedersen [a,*], U.H. Thygesen [b], H. Madsen [a]

[a] Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[b] National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark

ARTICLE INFO

Article history:
Received 17 February 2009
Received in revised form 18 March 2010
Accepted 25 April 2010
Available online 2 May 2010

Keywords:
Finite element method
Hidden Markov model
Nonlinear state estimation
Point-mass filter
Sequential Monte Carlo
Stochastic differential equation

ABSTRACT

A new approach to nonlinear state estimation and object tracking from indirect observations of a continuous time process is examined. Stochastic differential equations (SDEs) are employed to model the dynamics of the unobservable state. Tracking problems in the plane subject to boundaries on the state-space do not in general provide analytical solutions. A widely used numerical approach is the sequential Monte Carlo (SMC) method which relies on stochastic simulations to approximate state densities. For off-line analysis, however, accurate smoothed state density and parameter estimation can become complicated using SMC because Monte Carlo randomness is introduced. The finite element (FE) method solves the Kolmogorov equations of the SDE numerically on a triangular unstructured mesh for which boundary conditions to the state-space are simple to incorporate. The FE approach to nonlinear state estimation is suited for off-line data analysis because the computed smoothed state densities, maximum *a posteriori* parameter estimates and state sequence are deterministic conditional on the finite element mesh and the observations. The proposed method is conceptually similar to existing point-mass filtering methods, but is computationally more advanced and generally applicable. The performance of the FE estimators in relation to SMC and to the resolution of the spatial discretization is examined empirically through simulation. A real-data case study involving fish tracking is also analysed.

1. Introduction

Since the introduction of the Kalman filter (Kalman, 1960), state-space models (SSMs) have been widely used to solve object tracking problems and have undergone constant development. Initially, tracking problems were mostly related to radar observations (Pearson and Stear, 1974), but more recent applications now range from animal tracking, based on e.g. observations of daylight (Sibert et al., 2003), to object tracking in surveillance camera recordings (Comaniciu et al., 2003). The Kalman filter and its variants are suitable for linear and mildly nonlinear tracking problems owing to their computational simplicity. However, the majority of the Kalman filter variants build on a parametric representation of the probability density, thus limiting their use when problems become complex and non-Gaussian. Hence, for highly nonlinear problems, nonparametric Bayesian filtering techniques must be applied.

A common approach to filtering nonlinear SSMs is the use of simulation methods such as Markov chain Monte Carlo (Golightly and Wilkinson, 2008) and sequential Monte Carlo (SMC) (Creal, 2008; Ristic et al., 2004). The SMC methodology is intuitive and applies to virtually all SSMs regardless of order or type. One drawback of the method is that randomness is

* Corresponding author. Tel.: +45 4525 3095; fax: +45 4588 2673.
E-mail addresses: mwp@imm.dtu.dk, wpsgodd@gmail.com (M.W. Pedersen), uht@aqua.dtu.dk (U.H. Thygesen), hm@imm.dtu.dk (H. Madsen).

0167-9473/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2010.04.018

introduced into the filter, which can complicate matters of smoothing and parameter estimation for certain problems. For example, it can be difficult to obtain maximum likelihood estimates of model parameters because the randomness makes the likelihood function non-differentiable. Advanced algorithms have been devised to ameliorate this issue, although more work is still required in this respect (Doucet and Tadić, 2003). SMC methods are particularly appealing for high-dimensional problems because they suffer minimally from the curse of dimensionality when evaluating the integrals of Bayesian filters (Cappé et al., 2005).

Another approach to nonlinear state estimation is to deterministically discretize the continuous state-space and then solve the governing partial differential equations and related integrals numerically on this discretization (Bucy and Senne, 1971; Kitagawa, 1987). The method is feasible for SSMs with a low-dimensional state-space, up to dimension 3, say. This paper focuses on the case of a planar (two-dimensional) state-space, but the method is applicable to scalar and theoretically also to higher-dimensional situations although the required computational effort in such cases becomes substantial. Planar problems are common in many applications, for example when tracking individuals on the surface of the Earth, and therefore deserve special attention.

The discretized SSM can be solved by the framework of discrete hidden Markov models (HMMs) (Cappé et al., 2005). Historically, HMMs have been applied in cases where the state-space is discrete by nature, e.g. for digital signals, and therefore discretized continuous SSMs are rarely labeled as HMMs even though HMM theory applies. Other common names for methods that operate on the discretized problem are "point-mass" filters or direct Fokker–Planck-based methods (FPMs).

Previous studies that have considered discretized SSMs have opted for the simplest and most obvious discretization approach which is a uniform gridding of the state-space (Šimandl et al., 2006; Thygesen et al., 2009). As a result, the solution to the forward Kolmogorov equation or, synonymously, the Fokker–Planck equation can be obtained by finite differencing. The finite difference (FD) solution scheme is easy to implement yet gives powerful results. It comes, however, with some critical drawbacks such as lack of a standard procedure for imposing boundary conditions and non-conformity of the discretization to curvatures in the boundary geometry beyond one dimension. As a consequence of the rectangular uniform grid, the computational resources are distributed evenly over the state-space, which is rarely appropriate as some states will have close to zero probability. To this end there have been attempts to focus computer power to the relevant parts of the state-space by adaptively changing the discrete grid during the filtering process (Šimandl et al., 2006). This improvement reduces the computational burden, but severely complicates matters of smoothing and state sequence estimation and consequently limits the application of the method to filtering problems.

In this study we introduce an alternative non-uniform discretization of a continuous SSM and a method suited for obtaining smoothed state estimates and maximum *a posteriori* estimates of model parameters on this grid. In the two-dimensional case the discretization, that remains unchanged during computations, consists of a number of differently sized triangular shaped elements that are joined at the vertices to form a mesh. The lack of uniformity gives the mesh the ability to conform to curved boundary geometries that arise in many applications of geographical tracking, which is our primary focus. Also, the unstructured mesh allows for finer discretization of important parts of the state-space, thereby utilizing computer resources more efficiently. Within this framework the solutions to the Kolmogorov equations are provided by the finite element (FE) method, reviewed in Clough (1980). The performance in relation to SMC is illustrated by a simple Brownian bridge example and numerical properties are assessed through synthetic and real data sets.

The paper is composed such that Section 2 establishes the state-space model. Section 3 presents the details of the Bayesian filter while Section 4 details the discretization of the state-space. Section 5 examines parameter and state sequence estimation and Section 6 illustrates the application of the method through examples. Section 7 provides concluding remarks.

2. The stochastic model

The dynamic system that we are considering is described by a two-dimensional stochastic differential equation

$$\mathrm{d}\boldsymbol{X}_t = f(t, \boldsymbol{X}_t)\mathrm{d}t + g(t, \boldsymbol{X}_t)\mathrm{d}\boldsymbol{B}_t, \tag{1}$$

where $\boldsymbol{X}_t \in \mathbf{R}^2$ is the state at time t, $f : \mathbf{R}^{1+2} \to \mathbf{R}^2$, $\boldsymbol{B}_t$ is two-dimensional Brownian motion and $g : \mathbf{R}^{1+2} \to \mathbf{R}^{2\times 2}$. We assume $\boldsymbol{X}_t$ to be reflected at boundaries according to (10) in the Appendix. We have observations indexed by $k \in \{1, \ldots, N\}$ with the observation pertaining to the interval $[t_{k-1}, t_k]$ denoted by $\boldsymbol{z}_k \in \mathbf{R}^p$. We will refer to a set of observations as $\mathcal{Z}_k = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k)^T$. There are no restrictions on the type and nature of $\boldsymbol{z}_k$. The observation equation relates the noisy time-discrete observations to the continuous process

$$\boldsymbol{z}_k = h(t_k, \boldsymbol{X}_k, \boldsymbol{w}_k), \tag{2}$$

where $\boldsymbol{w}_k \in \mathbf{R}^q$ is a random perturbation and $\boldsymbol{X}_k$ is short for $\boldsymbol{X}_{t_k}$. No assumptions are made about the form of the possibly nonlinear mapping $h : \mathbf{R}^{1+2+q} \to \mathbf{R}^p$. The two-dimensional state-space can be written explicitly as $\boldsymbol{X}_k = (X_k, Y_k)^T$. The observation errors, $\boldsymbol{w}_k$, are assumed to be independent and identically distributed and can have any type of distribution.

3. The Bayesian filter and smoother

In this section we describe the filtering and smoothing recursions required for predicting, updating and smoothing the state estimates given data and the stochastic model. We assume here that the parameters of the model are known; in Section 5 we address estimation of these parameters.

3.1. Filtering

We apply a Bayesian filter to estimate the probability density of the state. The density conditional on $\mathcal{Z}_k$ is given by $\phi(t, \boldsymbol{x}_t|\mathcal{Z}_k)\mathrm{d}\boldsymbol{x}_t = P(\boldsymbol{X}_t \in \mathrm{d}\boldsymbol{x}_t|t, \mathcal{Z}_k)$ for $t \geq t_k$. The filter consists of two steps, the time update and the data update, that are performed recursively. The time update is related to the system Eq. (1), and has the purpose of predicting the evolution of $\phi(t, \boldsymbol{x}_t|\mathcal{Z}_k)$ throughout the elapsed time between observations. The time evolution of $\phi(t, \boldsymbol{x}_t|\mathcal{Z}_k)$ is described by the Kolmogorov forward equation (or Fokker–Planck equation) (Øksendal, 2007) in two dimensions

$$\dot{\phi} = -\nabla \cdot (u\phi - D\nabla\phi), \tag{3}$$

with $\phi = \phi(t, \boldsymbol{x}_t|\mathcal{Z}_k)$, $\dot{\phi}$ denoting $\frac{\partial\phi}{\partial t}$, $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$, and where $D = \frac{1}{2}gg^T$ and $u = f - \nabla D$ are the diffusion and advection parameters respectively. No-flux boundary conditions are imposed to ensure conservation of probability mass within the domain. The initial condition for (3) is $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_k)$ and the solution, $\phi(t_{k+1}, \boldsymbol{x}_{k+1}|\mathcal{Z}_k)$, is obtained by solving the equation over the time period $[t_k, t_{k+1}]$.

The data update step consists of applying the information in $\boldsymbol{z}_k$ to $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_{k-1})$ using Bayes' rule

$$\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_k) = \psi_k^{-1}\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_{k-1})L(\boldsymbol{z}_k|\boldsymbol{x}_k), \tag{4}$$

where $\psi_k = \int \phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_{k-1})L(\boldsymbol{z}_k|\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k$, i.e. a normalization constant. The term $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_{k-1})$ comes from solving (3) and the term $L(\boldsymbol{z}_k|\boldsymbol{x}_k)$ is the likelihood of $\boldsymbol{z}_k$ given $\boldsymbol{x}_k$. We will refer to $L(\boldsymbol{z}_k|\boldsymbol{x}_k)$ as the "data likelihood" to avoid confusion with the likelihood of model parameters introduced later.

By recursively solving Eqs. (3) and (4) the density $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_k)$ is computed for all $k \in \{1, \ldots, N\}$. The filter is initialized by computing $\phi(t_1, \boldsymbol{x}_1|\boldsymbol{z}_1)$ from (4) with a prior distribution that reflects information available about the initial state.

3.2. The smoother

The smoothing recursions run in reverse time using the results of the filter and give the smoothed estimates, $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_N)$, by assuming full knowledge of $\mathcal{Z}_N$ for all $k \in \{1, \ldots, N\}$. At t_k the unused observations in $\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_k)$ are $\mathcal{W}_{k+1} = (\boldsymbol{z}_{k+1}, \ldots, \boldsymbol{z}_N)^T$. The information contained in these observations is given by

$$\Lambda(t_{k+1}, \mathcal{W}_{k+1}|\boldsymbol{x}_{k+1}) = \frac{\phi(t_{k+1}, \boldsymbol{x}_{k+1}|\mathcal{Z}_N)}{\phi(t_{k+1}, \boldsymbol{x}_{k+1}|\mathcal{Z}_k)}. \tag{5}$$

Smoothing involves the use of the Kolmogorov backward equation (Øksendal, 2007) which describes the reverse time evolution of $\Lambda(t, \mathcal{W}_{k+1}|\boldsymbol{x}_{k+1})$ according to

$$-\dot{\Lambda} = u\nabla\Lambda + \nabla \cdot (D\nabla\Lambda), \tag{6}$$

where $\Lambda = \Lambda(t, \mathcal{W}_{k+1}|\boldsymbol{x}_{k+1})$ and with the Neumann boundary condition $n \cdot \nabla\Lambda = 0$, where n is a vector normal to the boundary of the state-space. The result of solving (6) with $\Lambda(t_{k+1}, \mathcal{W}_{k+1}|\boldsymbol{x}_{k+1})$ as the terminal condition over the time period $[t_k, t_{k+1}]$ we denote $\Lambda(t_k, \mathcal{W}_{k+1}|\boldsymbol{x}_k)$, which can be interpreted as the data likelihood of the observations $\mathcal{W}_{k+1}$ at t_k. This is perhaps clearer when it is realized that

$$\Lambda(t_k, \mathcal{W}_k|\boldsymbol{x}_k) = \Lambda(t_k, \mathcal{W}_{k+1}|\boldsymbol{x}_k)L(\boldsymbol{z}_k|\boldsymbol{x}_k) \tag{7}$$

and therefore that $\Lambda(t_k, \mathcal{W}_{k+1}|\boldsymbol{x}_k)$ for any k can be calculated from a recursion of (6) and (7). Finally the smoothed estimate satisfies

$$\phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_N) = \phi(t_k, \boldsymbol{x}_k|\mathcal{Z}_k)\Lambda(t_k, \mathcal{W}_{k+1}|\boldsymbol{x}_k).$$

The smoothing recursions are initialized with the final filter estimate $\phi(t_N, \boldsymbol{x}_N|\mathcal{Z}_N)$ which is also a smoothed estimate.

4. Discretization of the state-space

The primary computational burden of the smoothing problem is the need to solve (3) and (6). The partial differential equations can be solved numerically with the finite element (FE) method. The FE method uses an unstructured discretization of the state-space, referred to as the mesh, defined by $M < \infty$ nodes (vertices) with states (locations) $\boldsymbol{x}^{(j)}$ for $j \in \{1, \ldots, M\}$. Associated with the nodes are basis functions $v_j(\boldsymbol{x})$ that act as interpolants via

$$\phi(t, \boldsymbol{x}) \approx \sum_{j=1}^{M} \beta_j(t)v_j(\boldsymbol{x}), \tag{8}$$

where $\beta_j(t) = \phi(t, \boldsymbol{x}^{(j)})$. The shape and form of the basis function depend on the mesh element type as explained by Cook et al. (2001). We use a three-node triangular element which is the simplest two-dimensional element. The technical details of the FE algorithm for this application are reviewed in the Appendix. For a more in depth coverage see e.g. Cook et al. (2001).

The mesh can be created and modified by importing the geometry of the state-space, e.g. a landscape, into mesh generating software. We used the open-source meshing tool Triangle (Shewchuk, 1996). The initial mesh should be coarse with uniformly sized elements to achieve a quick but rough estimate of the posterior distribution. The mesh can then be refined in important regions and coarsened in unimportant regions according to the posterior distribution. In Section 6.2 we examine empirically how the mesh fineness influences the accuracy of results.

Solving (3) requires $T = (t_{k+1} - t_k)/\delta$ FE iterations, where δ is the FE time step. The computational complexity of (3) is therefore $O(TM^2)$ since the solution is obtained by a simple matrix vector multiplication. However, matrix sparsity algorithms can be employed that significantly reduce computation time and memory requirements. For a band matrix with bandwidth K the complexity becomes $O(TKM)$ where $TK \ll M$. For SMC methods (e.g. the particle filter of Cappé et al., 2007) the complexity is $O(M_*)$ for the time update, where M_* is the number of particles. The FE smoothing recursion has complexity $O(TKM)$ as for the filtering recursion. Marginal smoothing for SMC methods requires $O(M_*^2)$ per recursion (Doucet et al., 2000); hence for smoothing problems, when accurate density estimation is required, i.e. when M and M_* become large, SMC methods suffer from a significant increase in computation time compared to the FE approach.

5. Parameter and state estimation

In Section 3 the parameters of the model were assumed to be known. In practice this assumption rarely holds and parameters must therefore be estimated. In this section we describe how maximum likelihood (ML) or maximum *a posteriori* (MAP) parameter estimates can be obtained and we discuss estimation of the optimal state sequence given the posterior distribution.

5.1. Parameter estimation

We focus on off-line parameter estimation and let $\boldsymbol{\theta}$ denote the parameters that we wish to estimate. In filter-based likelihood inference the recursions of Section 3.1 executed with a given set of parameters provide the likelihood function evaluated at those parameters (Ljung, 1997). Specifically, the log-likelihood function of $\boldsymbol{\theta}$ is given by the joint density of the observations

$$l(\boldsymbol{\theta}|\mathcal{Z}_N) = \log\left\{\phi(z_1)\prod_{k=2}^{N}\phi(z_k|\mathcal{Z}_{k-1})\right\}.$$

By storing $\psi_k = \phi(z_k|\mathcal{Z}_{k-1})$ given by (4) for all $k \in \{1,\ldots,N\}$ we obtain a log-likelihood value of $\boldsymbol{\theta}$. Assuming that we have *a priori* information about $\boldsymbol{\theta}$ quantified by the prior density $\pi(\boldsymbol{\theta})$ the MAP estimate of $\boldsymbol{\theta}$ is obtained by maximizing the posterior density, i.e.

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\{l(\boldsymbol{\theta}|\mathcal{Z}_N)\pi(\boldsymbol{\theta})\}. \tag{9}$$

Note that (9) gives the ML estimate of $\boldsymbol{\theta}$ if $\pi(\boldsymbol{\theta})$ is uninformative. The log-likelihood function is a deterministic function of $\boldsymbol{\theta}$ conditional on $\mathcal{Z}_N$ and the mesh. In this study we use the fmincon function in the Matlab optimization toolbox to find $\widehat{\boldsymbol{\theta}}$. The optimizer evaluates the gradients numerically which can be computationally cumbersome, so optimization using analytically derived gradients is an important subject of future research.

5.2. State sequence estimation

For tracking problems we are typically more interested in a state sequence or synonymously a track, supplemented by the probability density of the state, rather than the density alone. Consider the restricted FE state-space where we have the discrete states $\boldsymbol{\xi}_k$ for $\boldsymbol{\xi}_k \in \{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(M)}\}$. We define a track on the mesh nodes $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_N)$ and express the posterior mean $\bar{\boldsymbol{\Xi}} = (\bar{\boldsymbol{\xi}}_1,\ldots,\bar{\boldsymbol{\xi}}_N)$ by its marginals

$$\bar{\boldsymbol{\xi}}_k = \sum_{\boldsymbol{\xi}} \boldsymbol{\xi}\phi(t_k,\boldsymbol{\xi}|\mathcal{Z}_N).$$

The posterior mean is a robust and L^2 optimal estimator of $\boldsymbol{\Xi}$ but cannot be guaranteed to lie within the state-space, e.g. when the state-space is not convex. An alternative is to calculate the posterior mode

$$\widehat{\boldsymbol{\Xi}} = (\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_N) = \arg\max_{\boldsymbol{\Xi}} L(\boldsymbol{\Xi}),$$

where the track likelihood is given by

$$L(\boldsymbol{\Xi}) = L(\boldsymbol{z}_1|\boldsymbol{\xi}_1)\prod_{k=2}^{N} P(\boldsymbol{X}_k = \boldsymbol{\xi}_k|\boldsymbol{X}_{k-1} = \boldsymbol{\xi}_{k-1})L(\boldsymbol{z}_k|\boldsymbol{\xi}_k).$$

The Viterbi algorithm (Viterbi, 2006) is an efficient way of computing the posterior mode. For skewed posterior distributions the mode does not necessarily capture the overall trends in the posterior distribution because it is based on

Table 1
Metrics describing the quality of the estimates of the Brownian bridge for the FE approach and SMC method respectively. The SMC results are calculated as the average of 50 runs with the standard deviation in parenthesis. Computing times T_{filt} and T_{smoo} have the unit of seconds.

$t = 0.5$	$e_{\mu}^{(x)} \times 10^{-4}$	$e_{\mu}^{(y)} \times 10^{-4}$	$e_{\sigma^2}^{(x)} \times 10^{-4}$	$e_{\sigma^2}^{(y)} \times 10^{-4}$	$D(\Phi \parallel \phi)$	$e_{L^1} \times 10^3$	T_{filt} (s)	T_{smoo} (s)
FE	3.0	−1.47	−1.45	0.40	103	4.2	3.3	2.3
SMC	−13.2 (24.1)	1.72 (19.3)	0.70 (1.81)	−0.30 (1.50)	3413 (792)	28.4 (4.3)	0.1 (0.0)	5.3 (0.9)

a single outcome. However, in contrast to the posterior mean case it always holds that $\widehat{\boldsymbol{\xi}}_k \in \{\boldsymbol{x}_k(1), \ldots, \boldsymbol{x}_k(M)\}$ for all k. The suitable choice of track depends on the specific application. For the tracking applications considered here we use the posterior mode to avoid invalid state estimates. We note that is also possible to draw random tracks from the posterior distribution, using a recursive algorithm similar to that of Thygesen et al. (2009).

6. Examples

In this section we illustrate the properties of the FE approach in comparison to an SMC method (Section 6.1), with respect to numerics (Section 6.2) and a real-data application (Section 6.3).

6.1. Comparison of FE with SMC

We set up a simple smoothing problem for tracking a particle in two state dimensions $\boldsymbol{X} = (X, Y)^T$. We assume the initial state at $t_1 = 0$ is known, $\boldsymbol{X}_1 = (0.6, 0.5)^T$, and that we have observed the state $\boldsymbol{z}_2 = (0.4, 0.5)^T$ of the particle without uncertainty at $t_2 = 1$. We also assume that $u = \mathbf{0}$, i.e. that the motion of the particle is Brownian. The aim is now to estimate the state probability density of the particle at $t = 0.5$ conditional on $\boldsymbol{z}_2$. The smoothing problem, known as a Brownian bridge, can be solved analytically and has a Gaussian posterior density with mean $\mu_x = 0.6 - 0.2t$, $\mu_y = 0.5$ and variance $\sigma^2 = 2Dt(1-t)$, where $D = 0.01$ is the diffusion coefficient of the Brownian motion which is assumed known and isotropic. The example resembles the situation where one wants to estimate the smoothed density between two observations which is often the case, in particular when analyzing sparse tracking data.

We compute the smoothed density with the FE approach as described in this text and compare it to the density estimated with an SMC method. Since resampling is not required we let the SMC implementation follow a sequential importance sampling scheme as in Cappé et al. (2007), their algorithm 2, using the prior kernel as the importance distribution and with marginal smoothing as proposed by Doucet et al. (2000). We set $M_* = M = 2665$, i.e. the number of SMC particles is equal to the number of nodes in the FE mesh. We construct a uniformly sized triangular FE mesh over the domain $\boldsymbol{x} \in \{0, \ldots, 1\}^2$. We define the FE time step $\delta = 0.014$ to obtain close to equal computation times for the two schemes. Observations without uncertainty are difficult to handle for both methods so the analyses are started at $t = 0.01$ and ended at $t = 0.99$ while adjusting the initial condition and the data likelihood to match the analytically known mean and variance at these times.

We compute a number of metrics that detail the comparison of the two methods. The deviation from the true mean and variance in the x direction are respectively

$$e_{\mu}^{(x)} = \mu_x - E(X), \qquad e_{\sigma^2}^{(x)} = \sigma^2 - V(X),$$

and similarly for the y direction we have

$$e_{\mu}^{(y)} = \mu_y - E(Y), \qquad e_{\sigma^2}^{(y)} = \sigma^2 - V(Y).$$

The Kullback–Leibler divergence, which can be regarded as an asymmetric distance between two probability densities, is

$$D(\Phi \parallel \phi) = \sum_{\boldsymbol{x}} \Phi(t, \boldsymbol{x}|z_2) \log \frac{\Phi(t, \boldsymbol{x}|\boldsymbol{z}_2)}{\phi(t, \boldsymbol{x}|\boldsymbol{z}_2)}.$$

The L^1 distance is

$$e_{L^1} = \sum_{\boldsymbol{x}} |\Phi(t, \boldsymbol{x}|\boldsymbol{z}_2) - \phi(t, \boldsymbol{x}|\boldsymbol{z}_2)|,$$

where $|\cdot|$ means absolute value.

For the FE approach we obtain $\phi(t, \boldsymbol{x}|\boldsymbol{z}_2)$ by interpolation of the nodal values via the linear basis functions to a rectangular grid $\boldsymbol{x} \in \{0, \ldots, 1\}^2$ with $n_x = n_y = 501$ grid points. For the SMC method we used kernel density estimation on the same grid with a two-dimensional Gaussian kernel with isotropic variance 0.0005. The kernel bandwidth was chosen to obtain the best approximation of the kernel density estimation to the true pdf (i.e. smallest L^1 and Kullback–Leibler divergence). The accuracy of the kernel density estimate is rather sensitive to the choice of bandwidth. It is therefore important in practice to employ some objective bandwidth selection protocol.

The results with computing times for filtering (T_{filt}) and smoothing (T_{smoo}) are summarized in Table 1. The SMC results are values averaged over 50 runs with corresponding empirical standard deviations. The FE solution gave slightly better

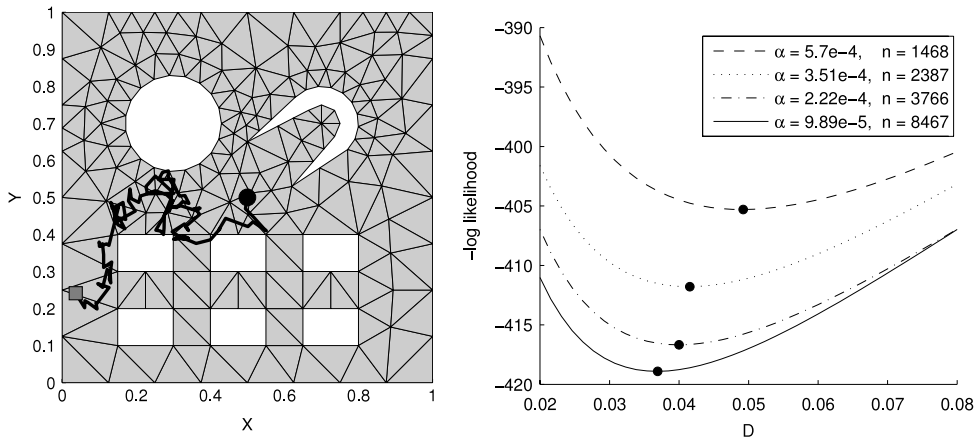


Fig. 1. Left: Coarsely meshed geometry of the artificial cityscape with the "true" simulated track of the test person. The solid circle represents the starting location; the shaded square represents the end location. White areas represent buildings, i.e. areas that are inaccessible to the test person. Right: Negative log-likelihood functions for D for meshes of different fineness indicated by α, the average element area, and n, the number of elements. It is noted that the ML estimates show convergence towards some value that is close to the true value of D.

estimates of the means while the SMC method was marginally better at estimating the variances. The density function itself was significantly better represented by the FE solution (by a factor of 33 for the Kullback–Leibler divergence and a factor of 7 for the L^1 distance). The SMC computing time was largely dominated by the smoothing step whereas the FE approach spent the time more evenly on the two steps. Owing to the $O(M_*^2)$ scaling, the computing effort of the SMC smoothing step may become prohibitively large for the SMC solution to reach a desired density accuracy. In such cases the FE approach is expected to deliver superior results. In fact, increasing M_* by a factor of 10 only led to an improvement in $D(\Phi \parallel \phi)$ by a factor of 2.5, and an improvement in e_{L^1} by a factor of 1.5 while increasing the total computing time of the SMC method by a factor of 100.

6.2. Synthetic data example

Now we address the behavior of the estimator in relation to the mesh fineness. For clarity we disregard the advection term, i.e. $u = \mathbf{0}$. We estimate the horizontal location $\boldsymbol{X}_k$ of a moving object, e.g. a human, for $k \in \{1, \ldots, N\}$, $N = 100$. Data consist of readings from an attached GPS device with constant time step. We thus have the observation equation

$$\boldsymbol{z}_k = \boldsymbol{X}_k + \boldsymbol{w}_k,$$

where $\boldsymbol{w}_k \sim N(\mathbf{0}, 0.01^2\boldsymbol{I})$. The movements of the object are Brownian with $D = 0.03\boldsymbol{I}$ and boundary behavior is implemented by disallowing steps outside of the model domain. Fig. 1 shows the artificial cityscape and the true simulated trajectory of the object.

We investigate how the posterior mode estimate and the likelihood function of D behave as functions of the fineness of the mesh quantified by the average element area α. We set $\boldsymbol{\theta} = D$, thus assuming that $\sigma_{\boldsymbol{w}}^2$ is known, and investigate the behavior of $l(\boldsymbol{\theta}|\mathcal{Z}_N)$ as a function of α; see Fig. 1. An indication of convergence towards a value close to the true value of D is observed as the mesh is refined. Furthermore, the likelihood of the parameter increases, an effect arising owing to the fact that the data likelihood is resolved more accurately. A coarse mesh results in overdispersion of the data likelihood which leads to a higher parameter estimate and a smaller likelihood value.

For each of the increasingly fine meshes we fix the value of D to the true value and compute the posterior mode. Define the track error metric

$$\varepsilon = 1/N \sum_{k=1}^{N} (\widehat{X}_k - X_k)^2,$$

where $\widehat{X}_k$ is the x coordinate of the posterior mode at t_k. The track error metric is computed analogously for the y coordinate. Fig. 2 shows how ε behaves as a function of α. Linear regression lines in the log–log domain fitted to the sloped part of the curves show an approximately linear order of convergence with visually estimated cut-offs that lie at $\alpha \approx 0.2 \cdot 10^{-4}$. The analytical variance of the posterior mode at t_k for Gaussian infinite domain problems (Thygesen and Nielsen, 2009) is given by

$$\sigma_{MPT}^2 = \sigma_{\boldsymbol{w}}^2 \frac{U(U + \sqrt{U^2 + 4})}{4 + U(U + \sqrt{U^2 + 4})} = 0.997 \cdot 10^{-4},$$

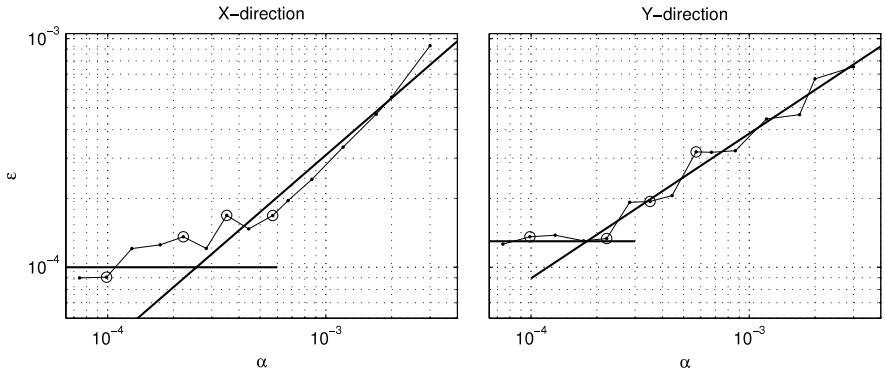


Fig. 2. Convergence plots of the track error metric of the posterior mode as a function of the resolution of the mesh (average element area, α) shown for the x and y coordinates. The four points highlighted with circles correspond to the likelihood curves in Fig. 1. With coarse grids ($\alpha > 2 \cdot 10^{-4}$), the track error is determined by the grid resolution. With fine grids, the error is largely independent of the grid and determined by the random measurement errors.

where $U = \sqrt{2Dh}/\sigma_{\boldsymbol{w}}$, with constant time step h and $\sigma^2_{\boldsymbol{w}}$ being the variance of the Gaussian noise, $\boldsymbol{w}_k$. The posterior distribution is not resolved accurately on the coarse mesh indicated by the values of ε becoming larger than their analytically expected value. As the mesh is refined this effect diminishes, as seen in Fig. 2, and ε converges to σ^2_{MPT}.

6.3. Real-data example

Finally we give a real-data application of the FE approach. We consider a geolocation problem which consists of estimating the large-scale horizontal movements of a fish based on observations from an electronic data storage tag (DST) attached to the fish while at liberty (Righton et al., 2006). The observations are readings of depth and salinity from which the horizontal movements of the fish can be inferred by comparing with databases that contain bathymetry and hydrographically modeled information for the salinity field. Here, a $N = 294$ data set from a cod in the Baltic Sea is analyzed. Fig. 3 depicts the FE mesh; note the highly complex boundary geometry which makes the governing partial equations difficult to solve for other methods. The mesh consists of 12 706 elements with a total of 7703 nodes and was refined recursively through two solutions of the problem. Again, we disregard the drift term, i.e. $u = \mathbf{0}$.

The data likelihood on a given day is computed by assuming that the observation $\boldsymbol{z}_k = (s_k, d_k)^T$, of salinity and depth respectively, is given by

$$\boldsymbol{z}_k = h(\boldsymbol{X}_k) + \boldsymbol{w}_k,$$

where $\boldsymbol{w}_k$ is Gaussian distributed with zero mean and covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{w}} = \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_d^2 \end{bmatrix}.$$

The values of σ_s^2 and σ_d^2 are considered as known and defined on the basis of the measurement uncertainty in the DST and the uncertainty within the databases. Including these in the estimation procedure is straightforward, although this leads to an increase in computation time.

The data likelihood for $j \in \{1, \ldots, M\}$ is

$$L(\boldsymbol{z}_k | \boldsymbol{x}_k^{(j)}) = \frac{1}{2\pi\sqrt{\det \boldsymbol{\Sigma}_{\boldsymbol{w}}}} \exp\left\{-\frac{1}{2}[\boldsymbol{z}_k - \widehat{\boldsymbol{z}}_k(\boldsymbol{x}_k^{(j)})]^T \boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}[\boldsymbol{z}_k - \widehat{\boldsymbol{z}}_k(\boldsymbol{x}_k^{(j)})]\right\},$$

where $\widehat{\boldsymbol{z}}_k(\boldsymbol{x}_k^{(j)})$ is the expected observation of salinity and depth at location $\boldsymbol{x}_k^{(j)}$ given by the databases. In the data likelihood at t_1 and t_N we include information about the known release and recapture locations respectively.

MAP estimation of D using a uninformative prior converged at

$$\widehat{D} = \begin{bmatrix} 21.1 & -2.1 \\ -2.1 & 48.3 \end{bmatrix} \text{ km}^2 \text{ day}^{-1},$$

with $\text{sd}(\widehat{D}_{xx}) = 8.33 \text{ km}^2 \text{ day}^{-1}$, $\text{sd}(\widehat{D}_{yy}) = 19.2 \text{ km}^2 \text{ day}^{-1}$ and $\text{sd}(\widehat{D}_{xy}) = 10.6 \text{ km}^2 \text{ day}^{-1}$ estimated by the inverse Hessian of the likelihood function, i.e. the observed Fisher information.

The posterior mode is depicted in Fig. 4. Note that the track, most prominently in the final part, at times crosses tounges of dry land. These steps seem unlikely and result from coarse temporal resolution which allows the fish to move around narrow tongues of land within the time span of two observations. A simple remedy is to increase the temporal resolution of the track estimation procedure. Still, the posterior mode gives an overall sensible impression of the movements of the fish.

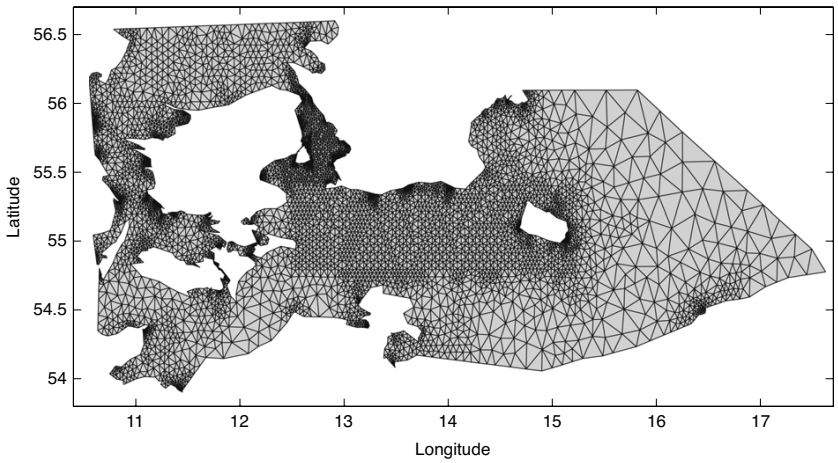


Fig. 3. The FE-triangulated mesh. Refinements to the mesh have been made in the central part of the Baltic Sea and around Zealand, particularly in Øresund, the sound between Zealand and Sweden. The mesh consists of 12 706 elements with a total of 7703 nodes.

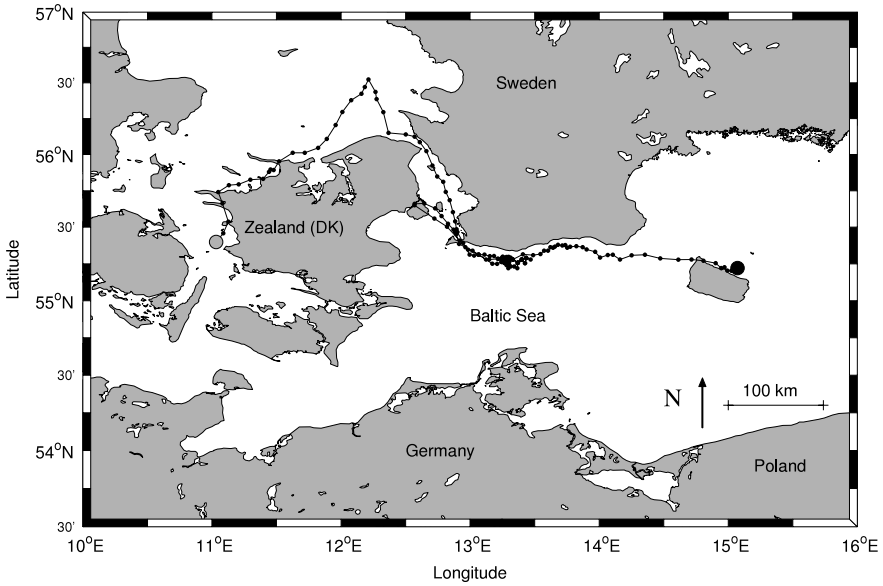


Fig. 4. Fish movements illustrated by the posterior mode estimated with the Viterbi algorithm suggesting that the fish, released at the solid circle at 55°15′N, 15°E and recaptured at the shaded circle at 55°30′N, 11°E, resided in the Baltic Sea at 13°E before migrating around Zealand via a northerly route.

A feature of this problem is that much information is contained in the recapture location at the end of the track, and thus the filtered locations differ markedly from the smoothed locations. As a result of this feature, an SMC method for this problem would require a large number of particles for smoothing.

The calculations were conducted on a laptop PC with a 1.4 GHz Centrino CPU and 2 GB RAM. Matlab 7.4 was used as computing environment leading to computation times on the scale of minutes for the smoothed posterior density, hours for parameter estimation and days for track estimation.

7. Conclusion

The FE approach to nonlinear state estimation problems presented is an alternative to existing SMC and point-mass filters. It is a deterministic method for solving the Kolmogorov equations with the ability to refine the numerical discretization at

complex boundary geometries and in areas where a more accurate solution is required. The discretization remains fixed throughout the computations, thus simplifying the estimation of smoothed state densities and MAP parameter values. The FE method differs markedly from SMC in terms of the effort required for implementation. While simple particle filters are straightforward to implement, greater effort is required for accurate smoothing and parameter estimation, in particular for problems such as the real-data example of Section 6.3. In contrast, with the FE method, the majority of the effort is spent at an early stage when implementing the solution scheme for the governing Kolmogorov equations. For some applications, for example oceanography which motivated this work, discretizing and solving partial differential equations is standard and optimized code may even be available. This would argue in favour of the FE method. On the other hand, for high state dimensions, the FE method becomes increasingly complex to implement and expensive to compute and is therefore best suited for problems on lower-dimensional state-spaces.

The two-dimensional Brownian bridge example showed that the SMC and FE smoothing methods performed equally well for moment estimation but for density estimation, the FE approach was superior by a factor of 33 with respect to the Kullback–Leibler divergence and a factor of 7 with respect to the L^1 distance for equal computation times. It is expected that the FE method will be preferable for smoothing problems, in particular when high density accuracy is required. The synthetic data example illustrated the convergence properties of the FE estimators where an approximately linear relation was observed for the mean square error of the posterior mode in relation to the average element area in the sloped part of the curve. Unlike SMC methods, the FE approach provides a deterministic likelihood function of the model parameters which, conditional on the fixed mesh, is simple to optimize numerically. On refining the mesh, the likelihood function converged towards the true value until plateauing when the optimally resolved density was reached. The real-data example showed how a simple implementation of the FE approach was used to solve a tracking problem involving complex state-space boundary geometry, parameter estimation and state density smoothing which would have been challenging to do with other available methods.

Acknowledgements

The authors thank the associate editor and two anonymous reviewers for their very helpful comments that significantly improved the quality of this text.

Appendix

This Appendix summarizes the use of the FE method for advection–diffusion partial differential equations (PDEs). To solve (3) with the FE method, we derive its weak formulation. For transparency we consider only the case where parameters are homogeneous in time and space. Comprehensive treatment of the FE method can be found in e.g. Cook et al. (2001) and Reddy and Gartling (2001).

For the PDEs (3) and (6) we assume no-flux boundary conditions, i.e. it holds for u that

$$(u\phi - D\nabla\phi) \cdot n = 0, \tag{10}$$

at any boundary point where n is a vector normal to the boundary $\partial\Omega$. The weak formulation of (3) is obtained by constructing the inner product with v_i:

$$\langle v_i, \dot{\phi}\rangle = \langle v_i, -\nabla \cdot (u\phi - D\nabla\phi)\rangle. \tag{11}$$

This equation is reduced by first considering the LHS of (11) and inserting (8), thereby obtaining after simplifications

$$\langle v_i, \dot{\phi}\rangle = \sum_{j=1}^{M} \dot{\beta}_j(t)\langle v_i, v_j\rangle.$$

The diffusion part of the RHS becomes

$$\begin{aligned}\langle v_i, \nabla \cdot (D\nabla\phi)\rangle &= \sum_{j=1}^{M} \beta_j(t)\langle v_i, \nabla \cdot (D\nabla v_j)\rangle \\ &= \sum_{j=1}^{M} \beta_j(t)\left[-\langle \nabla v_i, D\nabla v_j\rangle + \int_{\partial\Omega} v_i D\nabla v_j \mathrm{d}n\right].\end{aligned}$$

Due to the no-flux boundary condition, the term $\int_{\partial\Omega} v_i D\nabla v_j \mathrm{d}n$ vanishes. The advection part of the RHS becomes

$$\begin{aligned}\langle v_i, \nabla\cdot(u\phi)\rangle &= \sum_{j=1}^{M}\beta_j(t)\langle v_i, \nabla\cdot(uv_j)\rangle \\ &= \sum_{j=1}^{M}\beta_j(t)\left[-\langle\nabla v_i, uv_j\rangle + \int_{\partial\Omega} v_i u v_j \mathrm{d}n\right].\end{aligned}$$

Again, due to the no-flux boundary condition, the term $\int_{\partial\Omega} v_i u v_j \mathrm{d}n$ vanishes. Collecting the above terms we obtain

$$\sum_{j=1}^{M}\dot{\beta}_j(t)\langle v_i, v_j\rangle = \sum_{j=1}^{M}\beta_j(t)\left[\langle\nabla v_i, uv_j - D\nabla v_j\rangle\right], \tag{12}$$

which is a system of ordinary differential equations that can be solved for $\beta_j(t)$ with standard numerical methods. With

$$c_{ij} = \langle v_i, v_j\rangle$$

and

$$a_{ij} = \langle\nabla v_i, D\nabla v_j - uv_j\rangle,$$

we can express (12) in matrix notation as

$$\boldsymbol{C}\dot{\boldsymbol{\beta}}(t) = -\boldsymbol{A}\boldsymbol{\beta}(t), \tag{13}$$

with

$$\boldsymbol{\beta}(t) = \begin{bmatrix}\beta_1(t)\\ \vdots \\ \beta_j(t)\end{bmatrix}, \qquad \boldsymbol{C} = \begin{bmatrix}c_{11} & \cdots & c_{1j}\\ \vdots & \ddots & \vdots \\ c_{i1} & \cdots & c_{ij}\end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix}a_{11} & \cdots & a_{1j}\\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij}\end{bmatrix}.$$

There are various ways of solving (13) numerically. To avoid unstable, oscillating, and possibly negative solutions for the density, we utilize a simple implicit Euler scheme

$$\frac{1}{\delta}(\boldsymbol{\beta}_{t_k+(m+1)\delta} - \boldsymbol{\beta}_{t_k+m\delta}) = -\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{\beta}_{t_k+(m+1)\delta} \quad m\in\{0,\ldots,T-1\}, \tag{14}$$

where $\boldsymbol{\beta}_t = \boldsymbol{\beta}(t)$, δ is the FE time step and $T\delta = t_{k+1} - t_k$. The choice of δ and the fineness of the mesh determines the magnitude of the approximation error of the solution as discussed in Cook et al. (2001). The appropriate value of T depends on the required accuracy of the solution and on the computing speed. Reformulating and simplifying (14) gives

$$\boldsymbol{\beta}_{t_k+(m+1)\delta} = \boldsymbol{R}^{-1}\boldsymbol{C}\boldsymbol{\beta}_{t_k+m\delta}, \tag{15}$$

where $\boldsymbol{R} = \boldsymbol{C}+\delta\boldsymbol{A}$ is called the coefficient matrix. The matrices involved are sparse for large systems. It is therefore important to exploit this in the implementation to save memory and reduce the number of trivial computations.

In order to solve (6), close to identical derivations can be made.

References

Bucy, R.S., Senne, K.D., 1971. Digital synthesis of non-linear filters. Automatica 7, 287–298.

Cappé, O., Godsill, S.J., Moulines, E., 2007. An overview of existing methods and recent advances in sequential Monte Carlo. IEEE Proc. Signal Proc. 95 (5), 899–924.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York, NY.

Clough, R.W., 1980. The finite element method after twenty-five years: a personal view. Comput. Struct. 12 (4), 361–370.

Comaniciu, D., Ramesh, V., Meer, P., 2003. Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 564–577.

Cook, R.D., Malkus, D.S., Plesha, M.E., Witt, R.J., 2001. Concepts and Applications of Finite Element Analysis. John Wiley & Sons Inc., Hoboken, NJ.

Creal, D., 2008. Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models. Comput. Statist. Data Anal. 52 (6), 2863–2876.

Doucet, A., Godsill, S., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat. Comput. 10 (3), 197–208.

Doucet, A., Tadić, V., 2003. Parameter estimation in general state-space models using particle methods. Ann. Inst. Statist. Math. 55 (2), 409–422.

Golightly, A., Wilkinson, D., 2008. Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput. Statist. Data Anal. 52 (3), 1674–1693.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Trans. ASME 82 (1), 35–45.

Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. J. Amer. Statist. Assoc. 82 (400), 1032–1041.

Ljung, L., 1997. System Identification: Theory for the User, 2nd ed.. Prentice-Hall, Upper Saddle River, NJ.

Øksendal, B., 2007. Stochastic Differential Equations. Springer-Verlag, New York, NY.

Pearson, J., Stear, E., 1974. Kalman filter applications in airborne radar tracking. IEEE Trans. Aerosp. Electron. Syst. 319–329.

Reddy, J.N., Gartling, D.K., 2001. The Finite Element Method in Heat Transfer and Fluid Dynamics. CRC Press, New York, NY.

Righton, D., Kjesbu, O.S., Metcalfe, J.D., 2006. A field and experimental evaluation of the effect of data storage tags on the growth of cod. J. Fish Biol. 68, 385–400.

Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter. Particle Filters for Tracking Applications. Artech House Publishers, Boston, MA.

Shewchuk, J.R., 1996. Triangle: engineering a 2D quality mesh generator and Delaunay triangulator. In: Lin, M.C., Manocha, D. (Eds.), Applied Computational Geometry: Towards Geometric Engineering. In: Lecture Notes in Computer Science, vol. 1148. Springer-Verlag, pp. 203–222. from the First ACM Workshop on Applied Computational Geometry.

Sibert, J.R., Musyl, M.K., Brill, R.W., 2003. Horizontal movements of bigeye tuna (*Thunnus obesus*) near Hawaii determined by Kalman filter analysis of archival tagging data. Fish. Oceanogr. 12, 141–151.

Šimandl, M., Královec, J., Söderström, T., 2006. Advanced point-mass method for nonlinear state estimation. Automatica 42 (7), 1133–1145.

Thygesen, U., Nielsen, A., 2009. Lessons from a prototype geolocation problem. In: Nielsen, J., Arrizabalaga, H., Fragoso, N., Hobday, A., Lutcavage, M., Sibert, J. (Eds.), Tagging and Tracking of Marine Animals with Electronic Devices. In: Reviews: Methods and Technologies in Fish Biology and Fisheries, vol. 9. Springer, pp. 257–276.

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

Viterbi, A.J., 2006. A personal history of the Viterbi algorithm. IEEE Signal Process. Mag. 23 (4), 120–123.

Appendix D

Estimating animal behavior and residency from movement data

Authors:
M.W. Pedersen, T.A. Patterson, U.H. Thygesen and H. Madsen.

Submitted to *Oikos* June 2010, **accepted** January 2011

Estimating animal behavior and residency from movement data

M.W. Pedersen, Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, mwp@imm.dtu.dk, Corr. auth. Tel. +45 4525 3095

T.A. Patterson, Commonwealth Scientific and Industrial Research Organization (CSIRO), Wealth From Oceans Flagship and Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia, Toby.Patterson@csiro.au

U.H. Thygesen, National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark, uht@aqua.dtu.dk

H. Madsen, Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, hm@imm.dtu.dk

Abstract

We present a process-based approach to estimate residency and behavior from uncertain and temporally correlated movement data collected with electronic tags. The estimation problem is formulated as a hidden Markov model (HMM) on a spatial grid in continuous time, which allows straightforward implementation of barriers to movement. Using the grid to explicitly resolve space, location estimation can be supplemented by or based entirely on environmental data (e.g. temperature, daylight). The HMM method can therefore analyze any type of electronic tag data. The HMM computes the joint posterior probability distribution of location and behavior at each point in time. With this, the behavioral state of the animal can be associated to regions in space, thus revealing migration corridors and residence areas. We demonstrate the inferential potential of the method by analyzing satellite-linked archival tag data from a southern bluefin tuna (*Thunnus maccoyii*) where longitudinal coordinates inferred from daylight are supplemented by latitudinal information in recorded sea surface temperatures.

Keywords: animal movement analysis, behavior switching, electronic tagging, hidden Markov model, residency, southern bluefin tuna, state-space model.

1 Introduction

Movements of individual animals constitute important and highly complex processes which influence the outcome of many large-scale ecological processes. For many species, individual movements can now be assessed empirically using electronic tracking and logging techniques (Cooke et al., 2004). Such information is increasing our understanding of both individual species and ecosystems. However, several problems invariably arise in the resulting data which require a statistical solution. Namely, the need to correct for location uncertainty (Vincent et al., 2002), handle missing or irregular data (Johnson et al., 2008) and the incorporation of barriers to movement (Ovaskainen, 2004).

The most immediate problem facing empirical measurement of movement is noise in the observations of location. The noise is mainly a result of two factors: uncertainty inherent in the observation process, and the fact that observations are a discrete sub-sample of the underlying continuous movement process. This error structure necessitates statistical models that are able to separate the two noise contributions to estimate the most likely location of an animal at any point in time. State-space models (SSMs, Patterson et al., 2008b) have recently become the favored tool for this (e.g. Jonsen et al., 2006; Pedersen et al., 2008; Patterson et al., 2010). As an alternative to SSMs Sumner et al. (2009) suggests a Bayesian approach which merges an unconventional underlying movement model with a likelihood model for the observed data.

Recently, models have been investigated which incorporate different movement modes reflecting shifts in the underlying behavioral state of the animal (e.g. Morales et al., 2004). Behavioral states, being unobserved, are often vaguely defined. Commonly the labels attached to each state reflect predictions from optimal foraging theory. Thus, animals should search more intensively in productive habitats and minimize time in other areas. The labels used for the different behavior states include "migrating", "ballistic", or "extensive" for fast, directed movements and "diffusive", "foraging". or "intensive" for slow movements with many direction changes and increased probability of foraging. Such behaviors driving movement are typically hidden to the observer and may only be inferred from the movement data itself (Patterson et al., 2009).

Data from tracking technology is often non-spatial (e.g. data from a temperature logger) yet can be mobilized in a spatial context. As demonstrated below, data from a temperature sensor can be used to inform about spatial location if synoptic spatial coverage of similar data is available. Fortunately, spatial data (e.g. remote sensing data) is often available, and can provide exactly this. Such data have been used by Nielsen et al. (2006) to improve location estimates from an SSM.

Animal movements are often constrained by barriers or edges. For example, the sea is a barrier to terrestrial animals, as is land for aquatic animals. Such restrictions provide useful information in that certain movement trajectories can be ruled out. This is an aspect which has not been included in many SSMs, in particular those that rely on linear-Gaussian models which cannot incorporate hard constraints. With Monte Carlo simulation methods (Sumner et al., 2009) it is possible to implement barriers using rejection sampling, however this has a tendency to bias the location distribution near barriers because naively proposed movement paths encountering barriers are removed. Methods using reflective barriers (Pedersen et al., 2011) on the other hand allow obstructed movements to be reoriented and remain inside the model domain to avoid rejection bias.

This paper presents a method that combines all of the above mentioned features in an integrated

Bayesian state-space approach using so-called hidden Markov models (HMMs). The aim of any Bayesian state-space analysis is to estimate the posterior distribution of the state (in our case the state is location and behavior of the animal). Previous approaches to Bayesian analyses of tracking data have disregarded the state posterior distribution and restricted their attention to reconstructed movement trajectories (typically the posterior mean, e.g. Jonsen et al., 2005). A track representation, however, does not express the uncertainty of the estimated locations. The full posterior distribution on the other hand, provides this insight and is therefore instrumental in assessing which features of the estimated movement that can be trusted.

The paper is composed as follows. The next section contains the continuous-time formulation of the SSM comprising location and behavior and explains parameter estimation and model selection in the context of HMMs. By simultaneously estimating location and behavior we are able to use the posterior distribution to link certain behavior types to certain locations. In the section following we analyse satellite tracking data from a southern bluefin tuna (*Thunnus maccoyii*). We demonstrate how the posterior distribution can be used to reveal geographical areas of residency and migration while accounting for data uncertainty. The final section discusses the pros and cons of the method and its potential for estimation of residency.

2 Materials and methods

Using a state-space model (SSM) the animal tracking problem is governed by two parts. The system process describes the animal movement and behavior, and the observation model links the process (i.e. movements) to the data (Harvey, 1992). Inference regarding the unobservable system process can then be established via this link using statistical methodology (filtering) which updates location and behavior estimates with observed data. Table 1 includes a reference list for the mathematical symbols used in the paper.

2.1 Model formulation in continuous time

Since animals change their movement patterns through time as a response to changing environmental factors, prey abundance, habitats etc. (Morales and Ellner, 2002), it is necessary to regard the system as a hierarchy of two sub-processes: An underlying behavior process that controls the switching between a number of different movement states; and a derived process that describes the movement dynamics conditional on the behavioral state. Formally we model the behavior process as a continuous-time Markov chain, I_t, with a finite state-space, $I_t \in \{1, 2, \ldots, n\}$, where t denotes time. State switching of the behavior process is described by the generator matrix, $\mathbf{G}^b$ (superscript b for behavior), which contains the switching rates, λ_{ij}, of jumping from behavior state i to behavior state j (Ibe, 2009).

The movement of the animal in continuous time is a (biased) brownian motion in the longitudinal (x) and latitudinal (y) direction. Given the current behavior state I_t of the animal, the Brownian motion is decribed by a drift $\boldsymbol{u}_{I_t} = (u_x, u_y)^T_{I_t}$ with unit km $\cdot$ day^{-1}and a diffusivity matrix $\boldsymbol{D}_{I_t}$ with unit km$^2 \cdot$ day^{-1}, where superscript T means transpose. Diffusion processes of this type are well established for modeling animal movement, both within analysis of tagging data (Sibert et al., 1999; Pedersen et al., 2008) and in ecology in general (Okubo, 1980).

To proceed with the analysis of the joint process of movement and behavioral shifts, we introduce the probability density $\phi_i(x, y, t)$ which describes the probability that the animal at time t is located

Symbol	Description
i	Behavioral state index.
x	Longitudinal state index.
y	Latitudinal state index.
n	Number of behavioral states.
N	Number of observations.
t_k	k'th sample time.
Δ_k	Length of time interval $[t_k, t_{k+1}]$.
$\boldsymbol{z}_k$	Data observed at time k.
$\mathcal{Z}_k$	All observations available by t_k.
λ_{ij}	Rate of switching from behavior i to j.
$\boldsymbol{u}$	Advection parameter, unit: $\text{km} \cdot \text{day}^{-1}$.
$\boldsymbol{D}$	Diffusion parameter, unit: $\text{km}^2 \cdot \text{day}^{-1}$.
$\mathbf{G}^b$	Generator matrix for behavioral process.
$\mathbf{G}_i^m$	Generator matrix for movement process in behavior state i.
$\mathbf{P}_k$	Probability transition matrix related to Δ_k.
ϕ_i	Probability density of the animal's location in behavior state i.
$\boldsymbol{\phi}$	Vector containing state probabilities.
$\boldsymbol{\theta}$	Model parameter vector.

Table 1: Symbol overview.

at (x, y) and at the same time is in behavioral state i. In Okubo (1980) it is shown that the time evolution of the probability density of a particle performing Brownian motion follows a diffusion-advection equation, which is a partial differential equation (PDE). Therefore, by including behavior switching dynamics the PDE describing the time evolution of ϕ_i is a diffusion-advection equation augmented with a term representing the behavior switching dynamics of the animal:

$$\frac{\partial \phi_i}{\partial t} = -\nabla \cdot (\underbrace{\boldsymbol{u}_i \phi_i}_{\text{adv.}} - \underbrace{\boldsymbol{D}_i \nabla \phi_i}_{\text{diffusion}}) + \underbrace{\sum_j \lambda_{ji} \phi_j}_{\text{behav. switch}} , \tag{1}$$

where ∇ is the two-dimensional spatial gradient operator. The diffusion and advection terms describe the flow of probability between regions in space. The behavior switching term is a weighted sum over the switching rates that jump into state i, i.e. this term represents the net flow of probability into behavioral state i. Recall from theory of continuous-time Markov chains (Grimmett and Stirzaker, 2001) that λ_{ii} are always negative while $\lambda_{ji} \geq 0$ for $j \neq i$. Thus, in Eq. 1 the term $\lambda_{ii}\phi_i$ is negative and represents the probability that the animal jumps away from the behavioral state i while the terms $\lambda_{ji}\phi_j$ for $j \neq i$ represents jumps into the behavioral state i. Together, the n coupled PDEs in Eq. 1 describe the underlying dynamics (movement and behavior) of the system in continuous time and continuous space.

To solve Eq. 1 some form of numerical approximation is required. Our approach discretizes the continuous spatial state-space into a finite, albeit large, number of uniformly sized squares (states) (Thygesen et al., 2009). The size of a grid cell is denoted dx. On a discrete state-space, ϕ is no longer a probability density, but is instead represented by a vector $\boldsymbol{\phi}$ containing the state probabilities ϕ_α, where the state index $\alpha = (x, y, i)$ is composed of location in x and y and the behavior state

i. The discretized state-space allows us to derive the generator matrices, $\mathbf{G}_i^m$ (superscript m for movement), related to the movement processes $i \in \{1, 2, \ldots, n\}$ (see Appendix A.1 for derivation of $\mathbf{G}_i^m$).

It is simple to manipulate $\mathbf{G}_i^m$ to explicitly exclude locations from the state-space that are not accessible to the animal by setting the rate of jumping to these states to zeros. In PDE terminology this is a "reflecting" boundary condition, which is a simple but natural way to incorporate barriers to movement. The ecological interpretation of reflecting boundaries is that animals that encounter barriers reorient themselves and move on. Thus, steps into and steps through obstacles are avoided.

Observations are assumed to be generated through a function h of the true animal location $\boldsymbol{X}_t$ and a random perturbation or error $\boldsymbol{w}_t$ which is related to the uncertainty of the measurement process. Formally the observation equation is written

$$\boldsymbol{Z}_t = h(t, \boldsymbol{X}_t, \boldsymbol{w}_t) \tag{2}$$

where $\boldsymbol{Z}_t$ is a vector containing the observations available at t. Note that the behavior state i is not part of the observation equation and is therefore fully hidden. This formulation does not require h to be linear and there are no restrictions on the form of the distribution of $\boldsymbol{w}_t$. For example non-Gaussian errors on satellite telemetry location estimates (Jonsen et al., 2005) or animal locations derived from radio-tracking triangulation (Anderson-Sprecher, 1994) are often heavy tailed in distribution. This necessitates a non-Gaussian distribution of $\boldsymbol{w}_t$ such as the t-distribution to accommodate outliers and stabilize estimation. However, the non-linear form of h may also allow for more subtle relations between observations and state, e.g. linking observations of daylight intensity to location (Nielsen et al., 2006). For marine animals, the lack of constraints on h is particularly useful as observations are rarely of location itself but rather of light intensity, depth, temperature etc. and so h becomes strongly nonlinear (Pedersen et al., 2008).

With tracking data we have observations at N points in time, i.e. t_k is the time point of the k'th observation and the set of observations available by this time is $\mathcal{Z}_k = \{\boldsymbol{z}_{t_1}, \ldots, \boldsymbol{z}_{t_k}\}$. The length of a sampling interval is $\Delta_k = t_{k+1} - t_k$. For irregularly sampled data or data sets with missing observations these time intervals have different lengths. For a given time interval length Δ_k we can compute the probability transition matrices $\mathbf{P}_k$ of the combined behavior and movement process using the generator matrices $\mathbf{G}_i^m$ and $\mathbf{G}^b$ (see Appendix A.1). Visualizing the structure of a simple transition matrix illustrates the hierarchical dependency between the movement and the behavior processes (Fig. 1).

2.2 Estimation and model selection with HMM

A hidden Markov model (HMM) filter (Zucchini and MacDonald, 2009) provides the probability distribution of the states forward in time conditional on data, $\boldsymbol{\phi}(t_k|\mathcal{Z}_k)$ (hereafter termed "state estimates"). State estimates are calculated successively by alternating between so-called time and data updates of the current state. Time updates predict the state at the next time given the current state, while data-updates use the next observation available to correct the time-updates. Similar recursions are well known from other algorithms such as the Kalman filter or particle filter (Andersen et al., 2007) and are generally referred to as filtering recursions. In addition to the state estimates, the filter returns a likelihood measure which indicates how well the model fits the data. Thus, the likelihood function, $\mathcal{L}$, of the unknown parameters $\boldsymbol{\theta}$ (drift, diffusion, switching

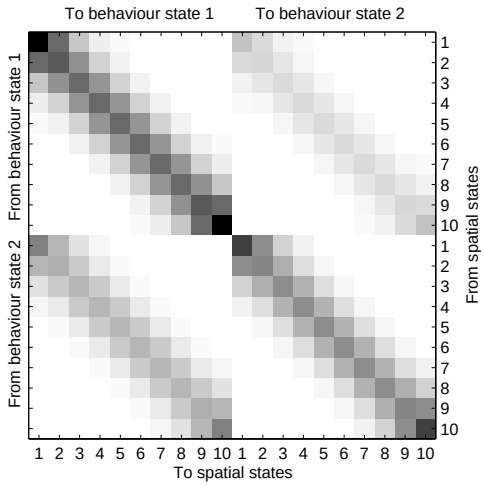


Figure 1: Probability transition matrix of a one-dimensional movement process with 10 spatial states combined with a two state behavior switching process leading to a 20×20 matrix. Darker colors represent state transitions with higher probability. The matrix consists of four 10×10 sub-matrices: the top two represent movement in behavior state one (e.g. foraging) and the bottom two represent movement in behavior state two (e.g. migration). The values of the sub-matrices are scaled by the probabilities of switching between the two behavior states.

rates) can be evaluated at, say, $\boldsymbol{\theta}_0$ by running the filter using the parameter values in $\boldsymbol{\theta}_0$ (Thygesen et al., 2009). Details of evaluating the likelihood function are given in Appendix A.2. Maximum likelihood (ML) estimation of model parameters is then straight forward:

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}_N). \tag{3}$$

This optimisation problem can be solved by most standard numerical optimizers which typically also provide an approximation to the Hessian matrix (i.e. curvature) of the likelihood function from which the uncertainty of $\widehat{\boldsymbol{\theta}}$ can be assessed (Pawitan, 2001). For the parameter estimation in this work the optimization toolbox included in Matlab (Mathworks, Natick USA) was used. In a Bayesian context it is common to introduce *a priori* information about the parameters through the prior density $\pi(\boldsymbol{\theta})$. The maximum *a posteriori* (MAP) estimate of the parameters is therefore the value of $\boldsymbol{\theta}$ which maximizes the posterior density $\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}_N)\pi(\boldsymbol{\theta})$. In practice, however, substantial prior information is rarely available (Jonsen et al., 2005) in which case the MAP and the ML estimates are close to identical. Furthermore, for model selection purposes the maximum value of the likelihood function is required and we therefore use the ML estimate in this study. Selection among alternative models in a Bayesian context would typically employ the Bayesian Information criterion (BIC). Unfortunately, calculating the BIC involves integrals without analytical solutions which therefore must be approximated (see Wasserman, 2000). The assumptions required by this approximation impose restrictions on the priors thus reducing the relevance of the BIC in context of the present problem. Instead we calculate Akaike's Information criterion, $AIC = -2\ell_{max} + 2M$ where ℓ_{max} is the maximum value of the log-likelihood function and M is the number of unknown model parameters. The model having the lower AIC is more likely and therefore ranked higher.

When parameters have been estimated only one step remains which is the so-called HMM

smoothing step (Thygesen et al., 2009). The recursions of the smoothing step work backwards in time using the filtered state estimates and all available data to determine the smoothed state estimates, $\boldsymbol{\phi}(t_k|\mathcal{Z}_N)$. The smoothed state estimates are more accurate and generally appear "smoother" than the filtering estimates because they exploit the full data set ($\mathcal{Z}_N$). When fitting an SSM in a Bayesian context the smoothing step provides the posterior distribution of the state. By posterior distribution we mean the probability distribution of all states at all times given all data. Obviously, this distribution has a high dimension and is quite complex. For post-processing purposes it is therefore common to use time marginals of the posterior distribution (i.e. probability distributions of all states at specific times) which, in fact, are the state estimates returned from the HMM smoothing algorithm.

See Appendix A.2 for the mathematical and algorithmic details regarding filtering, smoothing, and parameter estimation.

2.3 Visualizing results

The posterior distribution obtained from the HMM smoothing procedure allows detailed information about behavior and location to be extracted through time. For visualizing details of short-term animal movements we sum the posterior distribution over the behavioral state, i.e.

$$V(x,y,t) = \sum_{i=1}^{n} \phi_{(x,y,i)}(t|\mathcal{Z}_N), \tag{4}$$

which is the probability distribution of the location at time t. Viewing $V(x,y,t)$ in succession for increasing time t (i.e. as an animation) presents an illustrative description of how the location of the animal and its uncertainty changes in time on a day-to-day basis (see Video appendix VA.1 for an example). The animation gives particularly important insights when observation errors are non-Gaussian or indirect (e.g. of daylight) since in this case the variance of the location is no longer sufficient to describe the spatial correlation patterns.

It is often of interest to examine the amount of time spent by the animal within a spatial region (e.g. Walli et al., 2009). As a necessary simplification, previous approaches to calculating the time spent in a region of interest often ignore the fact that the observed location of the animal is uncertain (Aarts et al., 2008). However, Monte Carlo based alternatives incorporating observation uncertainty are available (Sumner et al., 2009). Within the HMM framework the time spent can be expressed as a statistical expectation. At first glance this is not a straightforward calculation because the true locations are always observed with error and, effectively, hidden. Fortunately, the posterior distribution can be used to give a reasonable estimate of the time spent at location (x,y) in the time interval τ. This time is calculated as the expectation given data and is computed as

$$R(x,y) = \sum_{l\in\tau} \sum_{i=1}^{n} \phi_{(x,y,i)}(t_l|\mathcal{Z}_N), \tag{5}$$

where l indexes time uniformly. This is to avoid that the possibly uneven sample intervals given by the index k lead to a bias in the expectation. Using l means that $\phi_{(x,y,i)}(t_l|\mathcal{Z}_N)$ must be computed at time points that have no related observation, however at these times the data-update step is simply omitted. So, by summing over the time and behavior indices of the posterior distribution (which incorporate the data induced spatial variability), we get $R(x,y)$ which is a distributional estimate of residence time.

We prefer to normalize $R(x, y)$ and view its cumulative distribution where grid cells are assigned a number between 0% and 100%, so that the 15% contour line, say, contains the smallest region where the animal was expected to spend 15% of its time. This "residency distribution" (RD) is conceptually similar to the utilization distribution ("UD: The name given to the distribution of an animal's position in the plane" cf. Worton, 1989). However, as noted by Royle and Dorazio (2008), this and other concepts such as home-ranges (Burt, 1943), activity centers etc. (Dixon and Chapman, 1980), are often vaguely defined. Despite being notionally similar, the quantity in Eq. 5 should not be directly interpreted as a UD in the usual sense. Nor should it necessarily be related to any sort of home-range, which, in any case, would not make sense for the highly migratory animals we consider here (see Fig. 3 bottom panel).

In general we may decide to sum over other variables and variable ranges of interest to obtain information about the movement or behavior over a specific time period or for a specific location. Behavior switching results may be visualized by summing over space

$$B(i,t) = \sum_{x,y} \phi_{(x,y,i)}(t|\mathcal{Z}_N) \tag{6}$$

to get the probability of each behavioral state at all time points (see Fig. 3 top panel, green line). Viewing $B(i,t)$ with the animation may reveal links between behavior and certain spatial regions (Video appendix VA.1, top panel). An alternative approach to relating behavior to space is the expected total time spent in a given region and behavioral state, i.e.

$$M(x,y,i) = \sum_{l \in \tau} \phi_{(x,y,i)}(t_l|\mathcal{Z}_N). \tag{7}$$

These distributions are useful for identifying e.g. migration corridors or residency hot spots while, at the same time, quantifying the spatial uncertainty for these different behaviors (see Fig. 3).

Track estimation is another way to visualize the posterior distribution. A track is an outcome of the posterior distribution and is in the context of this paper defined by a vector $\boldsymbol{a} = (\alpha_1^T, \ldots, \alpha_N^T)$. A track in the sense of $\boldsymbol{a}$ not only contains the estimated geographical coordinates of the animal, but also the most probable switching sequence through the behavior states (see Fig. 3 top panel, black line, for an example). Random tracks, conditional on data, can be simulated from the posterior distribution as described in Thygesen et al. (2009); this is useful for examining a range of possible tracks and for estimating statistics such as the probability that the individual enters certain regions. Similarly, the most probable track is the outcome that returns the highest value of the posterior distribution. In technical terms it is a maximum *a posteriori* estimate, i.e. the state sequence that maximizes the posterior distribution. The probability distribution and the most probable track estimate are different ways of decoding the HMM (cf. Zucchini and MacDonald, 2009) and may deviate at times when data are weak (Fig. 3 at the transition from the Tasman Sea to the Southern Ocean).

The algorithm (Viterbi, 2006) used to calculate the most probable track is detailed in Appendix A.3. The performance of the HMM approach with respect to state estimation, parameter estimation and model selection was validated in a simulation study which is in Appendix A.4.

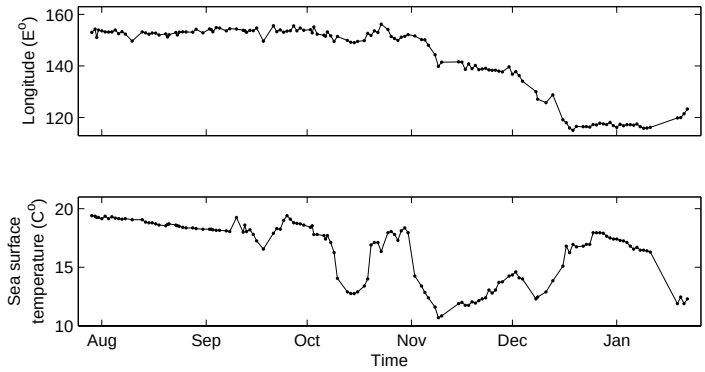


Figure 2: The dataset from the southern bluefin tuna as transmitted from the PSAT tag. Top: Longitude as estimated on-board the tag from observed daylight intensity. Bottom: Sea surface temperature measured by the tag when the tuna visited the surface. Notice that data is not sampled uniformly. This is most clear in the final part of the dataset (mid January).

2.4 Data analysis

To demonstrate the described framework, the model was applied to field data from PSATs attached to southern bluefin tuna (SBT, *Thunnus maccoyii*). Complete details of the data collection procedure are given in (Patterson et al., 2008a). The PSAT (Wildlife Computers PAT-3, Redmond, USA) was deployed on a 168cm/13 year old SBT captured off the east coast of Australia in the Tasman Sea in July 2003. The known start location was used to initialize the HMM filter. The PSAT detached from the SBT 177 days later, south of Western Australia in the Southern Ocean. Longitude estimates (Fig. 2, top) were generated from the PSAT data using the WC-GPE.1.02.0000 software (Wildlife Computers). Measurements of sea surface temperature (SST, Fig. 2, bottom) were taken from the temperature sensors on the tag. The PSAT was programmed to measure SST and longitude twice per day. However, it was not always the case that the SBT visited the surface in every sample interval. Thus, the returned data were sampled irregularly in time which necessitated a continuous-time analysis.

For the described observation scheme, Eq. 2 becomes

$$\begin{pmatrix} T_t \\ L_t \end{pmatrix} = h(t, \boldsymbol{X}_t) + \begin{pmatrix} \varepsilon_T \\ \varepsilon_L \end{pmatrix},$$

where h is a non-linear function that describes how SST (T_t) and longitude (L_t) inferred from daylight intensity vary as function of location and time. This relation is expressed by hydrographical SST prediction models (six day composite images of remotely sensed SST constructed from Advanced Very High Resolution Radiometer (Armstrong and Vazquez-Cuervo, 2001), CSIRO Marine and Atmospheric Research) and astronomical models of sunlight exposure (Hill and Braun, 2001). Both white noise terms, ε_T and ε_L, were assumed to be Gaussian distributed with standard deviations $\sigma_T = 0.71\,^{\circ}\mathrm{C}$ and $\sigma_L = 35\,\mathrm{km}$ estimated based on independent data sets (see Appendix A.5) and were therefore omitted from parameter estimation. For the final results we used a grid size of 111×201 grid cells equating to square cell size of $dx = 25.52$ km. This grid size was limited by computer memory requirements and to keep run times at feasible levels (estimation took 10-40 hours depending on the specific model; see below for model configurations).

Model acronym	Model parameters	No. of parameters
D	D_1	1
DA	D_1, u_x, u_y	3
SD	D_1, D_2, p_{11}, p_{22}	4
SDA	D_1, D_2, u_x, u_y, p_{11}, p_{22}	6

Table 2: The four models and their parameter configurations. Model acronyms mean D: diffusion, DA: diffusion-advection, SD: switching diffusion, SDA: switching diffusion-advection.

We considered four movement-behavior models (see Table 2) that were different parameter configurations of the SSM and analyzed their relative performance using AIC-based model selection. To maintain focus on the model's ability to estimate behavior we assume that the x and y components of the diffusion terms are uncorrelated. Thus $\boldsymbol{D}_i = \text{diag}\{[D_i,\, D_i]\}$.

Parameters were estimated for each of the four models listed in Table 2. To ease interpretation we converted the behavior switching rates estimated in continuous-time to transition probabilities for a fixed time step, $\Delta_k = 0.5$ day (12 hours). To summarize the movement rate of the SBT we calculated the square root of the expected squared distance moved in a time period of length $dt = 1$ day (24 hours):

$$\begin{aligned} S_i &= \sqrt{E(X_t^2) + E(Y_t^2)} \\ &= \sqrt{2\widehat{D}_i dt + (\widehat{u}_{xi} dt)^2 + 2\widehat{D}_i dt + (\widehat{u}_{yi} dt)^2}, \end{aligned}$$

which we refer to as the expected movement with unit km $\cdot$ day^{-1}. This formula comes from the definition of variance, i.e. that $E(A^2) = V(A) + [E(A)]^2$, where A is a random variable. The quantity S_i is a useful gauge of the level of activity in behavior state i.

3 Results

Model selection clearly favored switching models over non-switching models (see AIC values, Table 3), a difference which was also highlighted by a 297 km RMS discrepancy between estimated trajectory locations of the diffusion-advection model (DA) and the switching diffusion-advection model (SDA). The estimated values of the advection parameters u_x and u_y for models DA and SDA were of moderate size and their estimated variance relatively large indicating a reduced influence of these parameters on the tracks and a limited support for these parameters by the data. Similarly, pairwise comparison of the AIC for the pure diffusion models (D, SD) versus diffusion-advection models (DA, SDA) reported only a slight advantage when the advection parameters were included. However, the SDA model did have the lowest AIC and therefore showed the best fit to data. Estimates of the behavior switching transition rates (presented here as transition probabilities) were almost identical for the two switching models, again supporting the conclusion that the advection contribution to the migratory behavior state for this data was of minor importance. Also the RMS difference in trajectories between the two switching models was small (88 km) and only four locations were classified into different states between the two models. Estimated parameter values and associated uncertainties are given in Table 3.

Fig. 3 summarizes the movement and behavior estimation for the model with the lowest AIC,

	Model D			Model DA			Model SD			Model SDA		
AIC		2241.32			2239.04			2183.15			2180.27	
Param.	0.025	MLE	0.975	0.025	MLE	0.975	0.025	MLE	0.975	0.025	MLE	0.975
D_1	4923	6644	8365	4873	6739	8605	48	275	502	40	277	514
D_2	–	–	–	–	–	–	9519	15439	21360	9391	15577	21763
u_x	–	–	–	−39.6	−22.2	−4.8	–	–	–	−97.4	−53.3	−9.2
u_y	–	–	–	−11.8	5.7	23.1	–	–	–	−33.0	9.3	51.6
S_1	–	163	–	–	166	–	–	33	–	–	33	–
S_2	–	–	–	–	–	–	–	248	–	–	255	–
p_{11}	–	–	–	–	–	–	0.86	0.95	0.98	0.88	0.95	0.98
p_{22}	–	–	–	–	–	–	0.88	0.95	0.98	0.86	0.95	0.98

Table 3: Results of the data analysis. Maximum likelihood estimates (MLEs) of model parameter values with 95% confidence intervals of the four models. S_i is the expected movement per day. Unit for D_i is $\text{km}^2 \cdot \text{day}^{-1}$, unit for u_x and u_y is $\text{km} \cdot \text{day}^{-1}$and unit for S_i is km.

i.e. the SDA model. Initially, the SBT resided in the Tasman Sea, east of the Australian continent, for about two months after tag deployment before it moved south to a region northeast of Tasmania. From October and onwards an increased migration probability was apparent (see Fig. 3 top panel) as the fish made a westerly migration into the Southern Ocean. The activity level dropped in January as the SBT stayed resident off the Western Australian coast. The RD highlighted four primary areas of residency (see Fig. 3 areas A–D). While the RD shown in Fig. 3 is only from a single individual, these areas coincide with apparent residency areas for large SBT from other studies (Patterson et al., 2008a). In the Tasman sea (areas A, B) large SBT have long been targeted by Australian domestic fisheries (Caton, 1991). The RD also highlights an apparent residency phase in an area off the southern coast of Australia, to the Northwest of Tasmania (area C). This area is known as the "Bonnie Upwelling" (Schahinger, 1987) and has been characterized as a local hotspot for a range of predator species, presumably due to the large concentrations of prey species.

4 Discussion

We presented a hidden Markov model (HMM) as an advanced and versatile approach to state-space modeling. The method provides a unified solution to a number of important complications related to the analysis of movement data: the need to explicitly account for movement uncertainty and the entanglement of movement and behavior; accounting for barriers to movement; and accommodating multiple sources of non-spatial and possibly irregular data with non-Gaussian error structures. The method can, however, also be useful for mapping behavioral modes present in accurate location data e.g. as recorded by telemetry devices. Output from the model is calculated using the posterior distribution of the state of the animal. The results therefore have a form that allows detailed biological insights to be obtained which have not previously been available from tracking data. Additionally, the computation time and accuracy of the solution can be controlled by altering the grid resolution. Thus, coarse results can be obtained rapidly in the implementation phase while final results are computed with high accuracy using a fine grid with longer computation times.

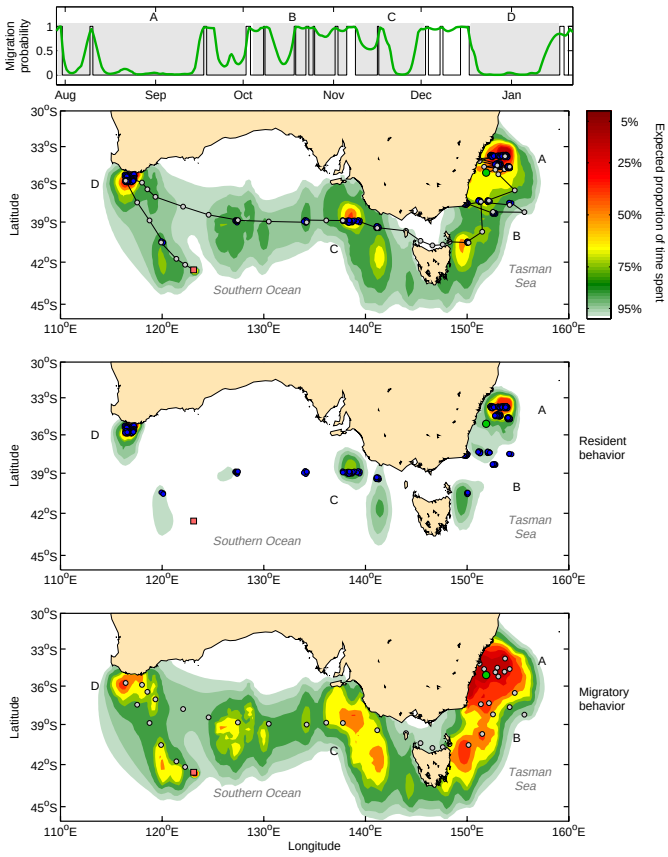


Figure 3: Top panel: the black line is the most probable behavior switching sequence and the green line is smoothed probability of the migratory behavioral state. Shaded areas relate the behavior switching to the corresponding spatial regions specified in the distribution plots below. Bottom panels: most probable trajectory of the switching diffusion-advection model for the southern bluefin tuna from its release 29 July 2003 to pop-up 22 January 2004. Shaded circles indicate migrating behavior, blue circles indicate resident behavior, green circle is release location and orange square is pop-up location. Underlaid are residency distributions (top: both behaviors, middle: resident, bottom: migratory) showing the expected proportion of time spent by the SBT within the contoured regions. Note that the trajectory deviates from the residency distribution at the migration from the Tasman Sea to the Southern Ocean (details in Methods and Discussion sections). Matlab's contourf function was used to plot the matrices containing the residency distributions.

4.1 Treatments of space and time

A key component of the HMM approach is the need to discretize space. At some level, this requirement may be seen as a limitation since predictions of locations are indeterminate at scales smaller than the model's spatial units. However, if the size of a grid cell is smaller than the uncertainty on the estimated position, this limitation is not critical. Also, in many situations discretization of space is actually required and spatially continuous location estimates would need to be discretized *post-hoc*. This applies for example when the objective is to determine residency in a reserve, a habitat patch, or a management unit.

A cogent point for the data considered here is that the precision of an inferred location is much higher than could be inferred from the data alone. Conversely, the scale of the spatial units in the model is much smaller than the scale of migrations made by the animal. Moreover, in this method, barriers to movement are easily included simply by setting the probability of moving to or through impossible locations to zero. Thus any loss of realism stemming from spatial discretization may be offset by ruling out impossible behaviors such as fish crossing land or terrestrial animals crossing large bodies of water. Therefore, the utility of this approach is not significantly diminished by spatial discretization and in fact may offer an integrated approach to aggregating location estimates up to larger spatial scales.

The primary factor influencing the computing time of the method is the grid resolution. Other important factors are the number of behavioral states, the number of parameters to be estimated, and the values of movement parameters. The parameter values are influential because they determine the sparsity of the probability transition matrices (larger values lead to denser matrices and therefore more computations). With the grid used for our final results the filter requires about one minute to run for the switching model with advection and reasonable parameter values. Total time required for estimation of parameters, tracks, and residency distributions is about a day. If parallel computing facilities are available more models can be estimated simultaneously thus avoiding extra computing time.

The switching HMMs presented here operate in continuous time. Electronic tag data is often subject to regular or irregular gaps in the data stream. As other authors have pointed out (Johnson et al., 2008; Patterson et al., 2010) continuous-time methods handle this seamlessly. Given that PSAT data is actually a regular time series (twice daily locations) with gaps, a discrete-time approach which handled missing data could equally have been applied. However, the continuous-time approach is more general.

4.2 Behavior models and model selection

Model selection for switching Brownian motion models is a challenging process. The simulation study (see Appendix A.4) confirmed that the correct model was selected if it were present in the candidate set. Predictably, for the analysis of a real data set the situation was not so clear cut. For the SBT track the most complex model (two-state with advection) was ranked highest. From an ecological standpoint, a constant advection term is unlikely to model movement behavior consistently. Future work should, therefore, consider some of the more advanced models that can be formulated within the state-space framework (see below). While this excludes mechanistic approaches for which a model likelihood cannot be computed, a viable future step could be to incorporate time-varying advection in the Brownian motion model. For example, models including

constant advection can be rejected when directed movements are present, yet do not exhibit an overall trend through time. In this case a constant advection term would most likely be estimated close to zero. Such a result would not however, entail the absence of advective processes in the tuna's motion but stems from positing sub-optimal models. Then, to incorporate the complexity of the observed movement, the diffusivity parameters D_i, could end up being spuriously large.

Potentially, model structure issues (such as the inclusion of advection terms or not) are important as they may influence inference of biologically relevant quantities, such as estimates of the percentage of time in each behavior mode. For instance, an advection-free model may need to place more locations in the "fast" movement mode to accommodate directed movement phases. A model with advection may be more flexible and thus able to move the animal faster between locations. However, this did not appear to be a significant factor in the data set we examined. The percentage of time in the migratory state was only slightly different between the model without advection (9.8% of days) and the model including advection (9.6% of days). Nonetheless, we suspect that further examination of the linkages between model structure, estimation, model selection and subsequent inference of biological quantities is required.

4.3 Alternative movement models

As model for individual movement we used variants of Brownian motion, which is the continuous-time equivalent to a random walk model. The correlated random walk (CRW) is an alternative model, which is able to capture short-term persistence in the animal's movement direction (Codling et al., 2008). Thus, the CRW is expected to provide more realistic uncertainty contours for the estimated locations as compared to the Brownian motion. Yet for relatively accurate data an estimated movement path is largely determined by the observations and relies to a smaller extent on the specific model for movement, while for inaccurate data it is not possible to reliably estimate small-scale correlations in movement. Therefore, for the type of data presented here it is unlikely that the estimated overall movement would change significantly if estimated using a CRW instead of the advection-diffusion model. Implementing a CRW in the HMM framework is theoretically possible, but requires gridding of four state dimensions (two-dimensional space and velocity), which entails a substantial total number of states. Even if the velocity can be coarsely discretised, memory requirements and calculation time of a CRW HMM will be immense and possibly impractical.

The Lévy walk (LW, random walk with Lévy distributed steps) is another movement model which has received much attention from ecologists (Sims et al., 2008). It has been argued that LWs in certain scenarios represent the optimal search strategy for animals (Viswanathan et al., 1999). However, theoretical studies have shown (e.g. Plank and Codling, 2009) that Lévy type movement patterns may arise by sub-sampling of composite random walks (similar to the switching model presented here) and vice versa. Similarly, theoretical results of another study (Thygesen and Nielsen, 2009) showed that even if the animal does follow a LW, estimation based on a simple random walk will give only marginally poorer estimation accuracy. Using real data, the state-space and model selection framework we have presented could in a future study be used to compare the estimation performance of switching models versus Lévy models while accounting for observation uncertainty. Such an assessment, while outside the scope of this study, would provide useful insights, for example into the ecological relevance of LWs through statistical tests at the individual animal level.

Our primary focus of this study was on behavior and residency estimation, and therefore we

employed simple isotropic diffusive schemes in the simulation and real data analyses. Naturally, this is a simplification since anisotropic diffusion is likely particularly in the presence of advective terms (Codling et al., 2010). Thus, using the HMM approach presented here, or alternative state-space modelling frameworks, to test the statistical significance of anisotropic diffusion with a larger movement dataset would be an interesting topic in future studies.

4.4 Spatial patterns

4.4.1 Calculation of most-probable tracks

The most probable track (MPT) is an example of "global decoding" (see Zucchini and MacDonald, 2009) of the HMM and seeks to find the most likely path given the entire data series taken simultaneously. So-called "local decoding" would involve taking the most likely state from the smoothed time marginal distributions at any particular point in time. Which is used depends on the goal for the inference. For calculating a track the MPT is the most useful, however it can produce some unexpected results. For example, in Fig. 3 the MPT at times deviated from the path that would result from simply choosing the maximum of the time marginal distributions of locations at each time. This can happen when the data for a particular time are uncertain. In this instance the MPT assigns more weight to the distributions from times before and after, and accordingly down-weights the uncertain intervening distribution. In such cases, the movement model dominates the MPT and minimizes the rate of movement (depending on the estimated values). The analysis shown here demonstrates that this phenomena may be particularly apparent during migration phases.

One further, useful aspect of the MPT for behavioral switching models is that it avoids the need for *ad hoc* thresholds for deciding on a most likely behavioral mode. For example, Jonsen et al. (2007) used 0.25 and 0.75 as threshold probabilities for foraging and searching states, respectively.

4.4.2 Mapping spatial uncertainty

Our estimation procedure is Bayesian and therefore gives an estimate of the posterior distribution i.e. the probability distribution of the animal's location and behavior. Having access to the posterior we can calculate the "residency distribution" (RD). As discussed previously, the RD is conceptually, and to a degree mathematically, similar to the ecological concept of a utilization distribution (UD). However, the RD, as specified here, gives the expected total time the animal visits a location. This is different to other measures such as the UD which assumes that the data provide a representative (and accurate) snapshot of the distribution of an animal in space. Our RD, on the other hand, accounts for the spatial uncertainty of an animal (see Fig. 3) and therefore has certain similarities to the time spent estimation of Sumner et al. (2009) although the computational approach is fundamentally different. Whilst the RD and UD are indeed different quantities, the RD has implications for making spatial inferences from uncertain spatial data. To obtain the UD, usage maps (Aarts et al., 2008; Matthiopoulos, 2003), or activity centers, from the model given here, one would use the MPT and treat this as a known set of locations without error. Then, kernel smoothing (e.g. Breed et al., 2006) or some other approach (e.g Gitzen et al., 2006) could be used for a given sample of MPTs from multiple animals. However, doing this would neglect the uncertainty in the animal's true locations. Instead, specialized methods for calculating UD (e.g. Benhamou and Cornélis, 2010) which also incorporate barriers to movement may be considered.

In calculating the RD, it may also be useful to marginalize over specific periods of the track.

For instance, a researcher may be interested in determining spatial residency of tagged animals over a particular month. Also, the RD may be calculated with respect to specific behavioral states in order to assess which areas are most important as either residency areas or migration corridors (see Fig. 3). Finally, by jointly considering the RD from multiple animals it is possible to assess the degree of overlap in their movement paths while simultaneously accounting for uncertainty. These sorts of approaches could be used to large tracking data sets and serve as an advanced alternative to common kernel density estimation methods (e.g. Walli et al., 2009).

4.5 Conclusion

This paper has demonstrated advances to state-space methods for behavioral estimation. The HMM approach can simultaneously estimate movement parameters, most likely behavioral state, the most-probable track and demonstrates how some basic model selection and inference may be carried out. Importantly, the paper provides a method for computing an index of residency which explicitly accounts for the uncertainty and auto-correlation in both location and behavior. This is an important and often neglected aspect of studies which examine the distribution of animals in space and time using telemetry and electronic tracking data.

4.6 Acknowledgments

Karen Evans provided valuable advice and support to this research. Funds were contributed by the Australian Fisheries Management authority.

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.

Andersen, K. H., A. Nielsen, U. H. Thygesen, H. H. Hinrichsen, and S. Neuenfeldt. 2007. Using the particle filter to geolocate Atlantic cod (*Gadus morhua*) in the Baltic Sea, with special emphasis on determining uncertainty. Canadian Journal of Fisheries and Aquatic Sciences **64**:618–627.

Anderson-Sprecher, R. 1994. Robust estimates of wildlife location using telemetry data. Biometrics **50**:406–416.

Armstrong, E. and J. Vazquez-Cuervo. 2001. A new global satellite-based sea surface temperature climatology. Geophysical Research Letters **28**:4199–4202.

Benhamou, S. and D. Cornélis. 2010. Incorporating movement behavior and barriers to improve kernel home range space use estimates. Journal of Wildlife Management **74**:1353–1360.

Breed, G. A., W. D. Bowen, J. I. McMillan, and M. L. Leonard. 2006. Sexual segregation of seasonal foraging habitats in a non-migratory marine mammal. Proceedings of the Royal Society B-Biological Sciences **273**:2319–2326.

Burt, W. 1943. Territoriality and home range concepts as applied to mammals. Journal of mammalogy **24**:346–352.

Caton, A. E. 1991. Review of aspects of southern bluefin tuna biology, population and fisheries. Inter-Am. Trop. Tuna Comm. Spec. Rep **7**:181350.

Codling, E., R. Bearon, and G. Thorn. 2010. Diffusion about the mean drift location in a biased random walk. Ecology **91**:3106–3113.

Codling, E., M. Plank, and S. Benhamou. 2008. Random walk models in biology. Journal of the Royal Society Interface **5**:813–834.

Cooke, S. J., S. G. Hinch, M. Wikelski, R. D. Andrews, L. J. Kuchel, T. G. Wolcott, and P. J. Butler. 2004. Biotelemetry: a mechanistic approach to ecology. Trends in Ecology & Evolution **19**:334–343.

Dixon, K. and J. Chapman. 1980. Harmonic mean measure of animal activity areas. Ecology **61**:1040–1044.

Gitzen, R., J. Millspaugh, and B. Kernohan. 2006. Bandwidth selection for fixed-kernel analysis of animal utilization distributions. The Journal of Wildlife Management **70**:1334–1344.

Grimmett, G. and D. Stirzaker. 2001. Probability and random processes. Oxford University Press, USA.

Harvey, A. 1992. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Hill, R. and M. Braun, 2001. Geolocation by light level. Page 315 in Electronic tagging and tracking in marine fisheries: proceedings of the Symposium on Tagging and Tracking Marine Fish with Electronic Devices, February 7-11, 2000, East-West Center, University of Hawaii. Kluwer Academic Pub.

Ibe, O. 2009. Markov processes for stochastic modeling. Academic Pr.

Johnson, D. S., J. M. London, M. A. Lea, and J. W. Durban. 2008. Continuous-time correlated random walk model for animal telemetry data. Ecology **89**:1208–1215.

Jonsen, I. D., J. M. Flemming, and R. A. Myers. 2005. Robust state-space modeling of animal movement data. Ecology **86**:2874–2880.

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.

Jonsen, I. D., R. A. Myers, and M. C. James. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology Progress Series **337**:255–264.

Matthiopoulos, J. 2003. Model-supervised kernel smoothing for the estimation of spatial usage. Oikos **102**:367–377.

Morales, J., D. Haydon, and F. J. e. al. 2004. Extracting more out of relocation data: building movement models as mixtures of random walks. Ecology **85**:2536–2445.

Morales, J. M. and S. P. Ellner. 2002. Scaling up animal movements in heterogeneous landscapes: The importance of behavior. Ecology **83**:2240–2247.

Nielsen, A., K. A. Bigelow, M. K. Musyl, and J. R. Sibert. 2006. Improving light-based geolocation by including sea surface temperature. Fisheries Oceanography **15**:314–325.

Okubo, A. 1980. Diffusion and ecological problems: Mathematical models. Springer-Verlag, Berlin Heidelberg.

Ovaskainen, O. 2004. Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. Ecology **85**:242–257.

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

Patterson, T., K. Evans, T. Carter, and J. Gunn. 2008a. Movement and behaviour of large southern bluefin tuna (*Thunnus maccoyii*) in the Australian region determined using pop-up satellite archival tags. Fisheries Oceanography **17**:352–367.

Patterson, T., B. McConnell, M. Fedak, M. Bravington, and M. Hindell. 2010. Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. Ecology **91**:273–285.

Patterson, T., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008b. State–space models of individual animal movement. Trends in Ecology & Evolution **23**:87–94.

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

Pedersen, M., U. Thygesen, and H. Madsen. 2011. Nonlinear tracking in a diffusion process with a Bayesian filter and the finite element method. Computational Statistics & Data Analysis **55**:280–290.

Pedersen, M. W., D. Righton, U. H. Thygesen, K. H. Andersen, and H. Madsen. 2008. Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching. Canadian Journal of Fisheries and Aquatic Sciences **65**:2367–2377.

Plank, M. and E. Codling. 2009. Sampling rate and misidentification of Lévy and non-Lévy movement paths. Ecology **90**:3546–3553.

Royle, J. and R. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press.

Schahinger, R. B. 1987. Structure of coastal upwelling events observed off the southeast coast of South-Australia during February 1983-April 1984. Australian Journal of Marine and Freshwater Research **38**:439–459.

Sibert, J., J. Hampton, D. Fournier, and P. Bills. 1999. An advection-diffusion-reaction model for the estimation of fish movement parameters from tagging data, with application to skipjack tuna *(Katsuwonus pelamis)*. Canadian Journal of Fisheries and Aquatic Sciences **56**:925–938.

Sims, D. W., E. J. Southall, N. E. Humphries, G. C. Hays, C. J. A. Bradshaw, J. W. Pitchford, A. James, M. Z. Ahmed, A. S. Brierley, M. A. Hindell, D. Morritt, M. K. Musyl, D. Righton, E. L. C. Shepard, V. J. Wearmouth, R. P. Wilson, M. J. Witt, and J. D. Metcalfe. 2008. Scaling laws of marine predator search behaviour. Nature **451**:1098–1102.

Sumner, M., S. Wotherspoon, and M. Hindell. 2009. Bayesian estimation of animal movement from archival and satellite tags. PloS one **4**:e7324.

Thygesen, U. H. and A. Nielsen, 2009. Lessons from a prototype geolocation problem. Pages 257–276 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.

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.

Vincent, C., B. J. McConnell, V. Ridoux, and M. A. Fedak. 2002. Assessment of Argos location accuracy from satellite tags deployed on captive gray seals. Marine Mammal Science **18**:156–166.

Viswanathan, G., S. Buldyrev, S. Havlin, M. Da Luz, E. Raposo, and H. Stanley. 1999. Optimizing the success of random searches. Nature **401**:911–914.

Viterbi, A. J. 2006. A personal history of the Viterbi Algorithm. IEEE Signal Processing Magazine **23**:120–142.

Walli, A., S. Teo, A. Boustany, C. Farwell, T. Williams, H. Dewar, E. Prince, and B. Block. 2009. Seasonal movements, aggregations and diving behavior of Atlantic bluefin tuna (*Thunnus thynnus*) revealed with archival tags. PLOS one .

Wasserman, L. 2000. Bayesian model selection and model averaging. Journal of Mathematical Psychology **44**:92–107.

Worton, B. 1989. Kernel methods for estimating the utilization distribution in home-range studies. Ecology **70**:164–168.

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

A Appendix

A.1 Computing the generator and the probability transition matrices

The movement dynamics of the animal is described by a Brownian motion. As shown in Okubo (1980) the time evolution of the probability density of the animal's location is described by the advection-diffusion equation

$$\frac{\partial \phi_i}{\partial t} = -\nabla \cdot (\boldsymbol{u}_i \phi_i - \boldsymbol{D}_i \nabla \phi_i). \tag{1}$$

To derive the entries of the movement generator matrix we write Eq. 1 on its finite difference form (Mitchell and Griffiths, 1980) in one dimension:

$$\phi(t_k, x) = (\mu - v)dt\, \phi(t_{k-1}, x + dx) + (1 - 2\mu dt)\, \phi(t_{k-1}, x) + (\mu + v)dt\, \phi(t_{k-1}, x - dx)$$

where $\mu = \frac{D}{dx^2}$, $v = \frac{u}{2dx}$ and $dt = t_k - t_{k-1}$. From this we gather that the probability of a step of length $\pm dx$ approaches $(\mu \pm v)dt$ as $dt \downarrow 0$. A stable finite difference scheme has the conditions that $dt < \frac{dx^2}{2D}$ and $|u| < \frac{2D}{dx}$ where $|\cdot|$ means absolute value.

This leads to the one dimensional generator of the movement process having the entries

$$g_{ij} = \begin{cases} \mu - v & \text{for} \quad j = i - 1 \\ \mu + v & \text{for} \quad j = i + 1 \\ -2\mu & \text{for} \quad j = i \\ 0 & \text{otherwise} \end{cases}.$$

This generalizes to two dimensions with the generator having two additional entries in each row except at boundary locations.

The dynamics of the behavior and movement in behavior state i are described in continuous time by their generators $\mathbf{G}^b$ and $\mathbf{G}_i^m$ respectively. For a given time interval Δ_k, probability transition matrices can be computed for the behavior process by $\mathbf{P}^b(t_k) = \exp(\mathbf{G}^b \Delta_k)$, where $\exp(\cdot)$ means the matrix-exponential operation (Grimmett and Stirzaker, 2001). Calculating transition matrices for large state-spaces requires a matrix exponential implementation that utilizes the so-called uniformization algorithm (Grassmann, 1977) which exploits the sparsity of the generator matrix.

We write

$$\mathbf{P}^b(t_k) = \begin{bmatrix} \mathbf{p}_1^b(t_k) \\ \vdots \\ \mathbf{p}_n^b(t_k) \end{bmatrix} \tag{2}$$

where $\mathbf{p}_i^b(t_k)$ are row vectors containing transition probabilities conditional on state i. The probability transition matrices of the movement processes are analogously given by $\mathbf{P}_i^m(t_k) = \exp(\mathbf{G}_i^m \Delta_k)$.

We can assemble $\mathbf{P}^b(t_k)$ and $\mathbf{P}_i^m(t_k)$ into a probability transition matrix that describes the joint process of behavior and movement

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{p}_1^b(t_k) \otimes \mathbf{P}_1^m(t_k) \\ \vdots \\ \mathbf{p}_n^b(t_k) \otimes \mathbf{P}_n^m(t_k) \end{bmatrix} \tag{3}$$

where $\otimes$ is the Kronecker product operator. The matrix $\mathbf{P}_k$ has a block structure which is illustrated for a simple case in Figure 1 in the main text.

A.2 State and parameter estimation

As described in Appendix A.1 the $\mathbf{P}_k$ matrices merge the two processes of movement and behavior into a single Markov process in which all spatial and behavioral dynamics are captured. For this process the symbol $\alpha = (x, y, i)$ is used to represent a state, where x and y refer to position in the two-dimensional space and i indexes the n behavioral states. We have partitioned the longitudinal and latitudinal directions into n_x and n_y cells and therefore the total number of spatial states is $n_{xy} = n_x n_y - n_u$, where n_u is the number of cells inside the grid that are inaccessible to the animal (such as land areas for marine animals). The probability distribution of the position and behavior states at time t_k is therefore a column vector $\boldsymbol{\phi}(t_k|\mathcal{Z}_k)$ of length $n_{xy}n$ since n vectors of length n_{xy} are concatenated. The vector $\boldsymbol{\phi}(t_k|\mathcal{Z}_k)$ has elements $\phi_\alpha(t_k|\mathcal{Z}_k)$.

A HMM filter recursion consists of a time and a data update step. The time update gives the predicted distribution $\boldsymbol{\phi}(t_{k+1}|\mathcal{Z}_k)$ and the data update gives the estimated distribution $\boldsymbol{\phi}(t_k|\mathcal{Z}_k)$. The time update of the probability distribution is a simple multiplication of the state probability vector with the transition matrix

$$\boldsymbol{\phi}(t_{k+1}|\mathcal{Z}_k)^T = \boldsymbol{\phi}(t_k|\mathcal{Z}_k)^T \mathbf{P}_k. \tag{4}$$

In the data update step the predicted distribution $\boldsymbol{\phi}(t_k|\mathcal{Z}_{k-1})$ is adapted to the observation $\boldsymbol{z}_k$ by applying Bayes' rule

$$\boldsymbol{\phi}(t_k|\mathcal{Z}_k) = \psi_k^{-1} \boldsymbol{\phi}(t_k|\mathcal{Z}_{k-1}) \odot \boldsymbol{L}(\boldsymbol{z}_k|\alpha), \tag{5}$$

where $\odot$ denotes elementwise multiplication, $\boldsymbol{L}(\boldsymbol{z}_k|\alpha)$ is the likelihood of $\boldsymbol{z}_k$ given the state and $\psi_k = \sum_\alpha \boldsymbol{\phi}(t_k|\mathcal{Z}_{k-1}) \odot \boldsymbol{L}(\boldsymbol{z}_k|\alpha)$. The data likelihood $\boldsymbol{L}(\boldsymbol{z}_k|\alpha)$ has a value for all α and is computed by comparing the observed data to the data expected to be generated in a given state. The way to compute $\boldsymbol{L}(\boldsymbol{z}_k|\alpha)$ depends on the form of the mapping function h in the observation equation (Eq. 2 in the main text). For example if $\boldsymbol{z}_k$ are noisy positions e.g. from satellite tags, h is linear which makes computations simple

$$\boldsymbol{L}(\boldsymbol{z}_k|\alpha) = N_{pdf}(\boldsymbol{z}_k, \alpha, \Sigma_w), \quad \text{for all } \alpha$$

where N_{pdf} is a Gaussian probability density function with mean α and covariance matrix Σ_w evaluated at $\boldsymbol{z}_k$. For outlier-prone observations such as Argos positions, a heavy tailed t-distribution may be applied instead of a Gaussian. Even more complex and non-linear links may be implemented if needed, see for example Pedersen et al. (2008).

The likelihood value of a given parameter set, $\boldsymbol{\theta}$, is a product of the one-step prediction errors

$$\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}_N) = \prod_{k=1}^{N} \psi_k. \tag{6}$$

Maximum likelihood estimates are obtained by optimizing $\mathcal{L}(\boldsymbol{\theta}|\mathcal{Z}_N)$ with respect to $\boldsymbol{\theta}$.

To incorporate all observations in each state estimate, i.e. to get $\boldsymbol{\phi}(t_k|\mathcal{Z}_N)$, the so-called smoothing step is required. The smoothed state estimates are therefore not only conditioned on data

observed by t_k but also on future measurements and are more accurate and appear "smoother" than $\boldsymbol{\phi}(t_k|\mathcal{Z}_k)$.

We state the smoothing recursions, but omit derivation details. For supplements on the smoothing step see Thygesen et al. (2009). The recursions are

1. Compute the vector

$$\boldsymbol{\Psi}(t_{k+1}) = \boldsymbol{\phi}(t_{k+1}|\mathcal{Z}_N) \oslash \boldsymbol{\phi}(t_{k+1}|\mathcal{Z}_k),$$

   where $\oslash$ is elementwise division.

2. Right multiply with the transition matrix to step backwards in time

$$\boldsymbol{\Lambda}(t_k) = \mathbf{P}_k \boldsymbol{\Psi}(t_{k+1}).$$

3. Get the smoothed estimate at t_k by

$$\boldsymbol{\phi}(t_k|\mathcal{Z}_N) = \boldsymbol{\phi}(t_k|\mathcal{Z}_k) \odot \boldsymbol{\Lambda}(t_k)$$

   where $\odot$ denotes elementwise multiplication.

The recursion is initiated with the last estimated distribution from the final iteration of the forward filter, $\boldsymbol{\phi}(t_N|\mathcal{Z}_N)$ which is also a smoothed estimate.

A.3 Finding the most probable track

The likelihood of a track (an outcome of the posterior distribution), $\boldsymbol{a} = (\alpha_1, \ldots, \alpha_N)$, can be computed as

$$L(\boldsymbol{a}) = L(\boldsymbol{z}_1|\alpha_1) \prod_{k=2}^{N} p_{\alpha_{k-1},\alpha_k} L(\boldsymbol{z}_k|\alpha_k),$$

where $p_{\alpha_{k-1},\alpha_k}$ is the entry in $\mathbf{P}_{k-1}$ corresponding to the transition from α_{k-1} to α_k. For summarizing movement and behavior switching we use the track that maximizes $L(\boldsymbol{a})$ denoted $\widehat{\boldsymbol{a}}$. For HMMs, estimating $\widehat{\boldsymbol{a}}$ is an often occurring problem that can be solved by the Viterbi algorithm (Viterbi, 2006) which relies on principles from dynamic programming and is proved to be a maximum likelihood estimator (Forney, 1973).

We define the branch metric

$$B_{\alpha_{k-1},\alpha_k}(t_k) = \log p_{\alpha_{k-1},\alpha_k} + \log L_{\alpha_k}(\boldsymbol{z}_k|\alpha_k).$$

An intermediate step in the maximization algorithm uses the state metric, $S_{\alpha_k}(t_k)$, which is the log-likelihood of the most likely of all possible tracks leading from the initial state to state α_k. The state metric is given by

$$S_{\alpha_k}(t_k) = \max_{\alpha_1,\ldots,\alpha_k} \left\{ \log L_{\alpha_1}(\boldsymbol{z}_1|\alpha_1) + \sum_{l=2}^{k} B_{\alpha_{l-1},\alpha_l}(t_l) \right\}.$$

This maximization problem can be solved recursively forward in time when it is realized that

$$S_{\alpha_k}(t_k) = \max_{\alpha_k} \left\{ S_{\alpha_{k-1}}(t_{k-1}) + B_{\alpha_{k-1},\alpha_k}(t_k) \right\}.$$

The procedure exploits the Markov property of the HMM to reject all but the most likely paths after each recursion. The final state of the overall most probable track is given by

$$\widehat{\alpha}_N = \arg\max_{\alpha_N} S_{\alpha_N}(t_N).$$

By continuously storing the most probable tracks for each iteration of the recursion the overall most probable track, $\widehat{\boldsymbol{a}}$, is simply given by extracting the track related to $\widehat{\alpha}_N$ when iterations are finalized.

A.4 Simulation study with results

To better understand the performance of the HMM approach with respect to estimation and model selection, we first applied the method to synthetic data sets. We examined if a relatively complex model could be reliably differentiated from simpler, candidate models. The synthetic data sets were generated with a two-state switching model comprised of a resident state with low diffusivity and no advection and a migratory state with a higher diffusivity and advection (i.e. the SDA model, see main text Table 1).

The simulation was intended to mimic the natural behavior of southern bluefin tuna (SBT) (*Thunnus maccoyii*). These fish make long distance migrations into the Indian Ocean from the Great Australian Bight (see Bestley et al., 2008). The parameter values of the data generating movement model were $(D_1, D_2, u_x, u_y, p_{11}, p_{22}) = (300, 1000, -50, 0, 0.95, 0.95)$ and the initial location of the fish (35.5°S, 126.7°E) was considered as known. The synthetic data of sea surface temperature (SST) and longitude were collected daily for 183 days from the simulated horizontal movements of the fish (see main text for details on error model).

The quality of the estimated most probable tracks was quantified using the root mean square (RMS) error of the residuals, σ_{MPT} (as compared to the true track). To evaluate the model's ability to correctly estimate the behavioral state, the average number of misclassified states $\text{avg}(n_{mis})$ was calculated. A total of 50 synthetic data sets were generated. For each data set maximum likelihood parameter estimation, model selection between the four models listed in the main text Table 1, state estimation and estimation of the most probable track was undertaken.

A.4.1 Results

AIC based model selection of the synthetic data sets resulted in 50 out of the 50 analyses arriving at the correct model (SDA) as the final model. The significant advection term made the data generating model easily distinguishable from the simpler switching model, SD, for which the average of the maximum likelihood estimates of D_2 was 2358 $\text{km}^2 \cdot \text{day}^{-1}$. It was clear that the estimation process compensated for the lack of advection by inflating the diffusivity estimate, thus decreasing the likelihood value of this model. Generally estimation of models D, DA and SD did not result in correct parameter values. This was expected since their model structures deviate from the data-generating model. The correct model however, did provide parameter estimates that were identifiable and consistent with the true parameters (Table 1).

The average numbers of misclassified states, $\text{avg}(n_{mis})$, showed that state switching was significantly better estimated by the SDA model as compared to the SD model, and in the optimal

		D_1	D_2	u_x	u_y	p_{11}	p_{22}	σ_{MPT}	n_{mis}
D	avg	1353	-	-	-	-		88.9	-
	sd	58.9	-	-	-	-		2.9	-
DA	avg	1004	-	-19.4	-0.1	-		84.1	-
	sd	31.1	-	1.15	0.37	-		2.8	-
SD	avg	276	2358	-	-	0.94	0.94	84.7	39.6
	sd	17.2	85.0	-	-	0.006	0.009	2.8	2.4
SDA	avg	307	1060	-48.4	-1.2	0.94	0.92	79.2	23.5
	sd	15.1	36.0	1.23	0.85	0.008	0.009	2.3	1.7

Table 1: Simulation results. Empirical averages (avg) and standard deviations of the averages (sd) of ML estimates and statistics from the 50 synthetic data sets. σ_{MPT} is the root mean square of the residuals of the estimated most probable track compared to the true track with unit km. n_{mis} is the average number of misclassified state estimates. Unit for D_i is $\mathrm{km}^2 \cdot \mathrm{day}^{-1}$, unit for u_x and u_y is $\mathrm{km} \cdot \mathrm{day}^{-1}$.

situation when the filter model is equal to the data generating model 87% of the behavior states were correctly estimated. The average value of σ_{MPT} for the SDA model was not significantly different to the three other models, which indicates that reasonable track estimates was still obtained even when the applied model differed from the data-generating model.

A.5 Estimation of data error variances

Using diagnostics data transmitted by the PSAT we were able to examine the relationship between the remotely-sensed surface temperature and that measured by the PSAT. While there were a few departures this relationship was strongly linear (Figure 1a). The residuals from the fit were approximately Gaussian distributed (Figure 1b) and the residual variance, $\sigma_T^2 = (0.71\,^\circ\mathrm{C})^2$, was therefore used as estimate for the temperature error variance in the model.

An average of the empirical root mean square estimates determined in Musyl et al. (2001) was used as longitude error variance, $\sigma_L^2 = (35\,\mathrm{km})^2$.

VA.1 Animation of time marginals

The animation is found via this link
http://www2.imm.dtu.dk/~mwp/VA1_animated_time_marginals.avi

References

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.

Forney, G. D. 1973. The viterbi algorithm. Proc. IEEE **61**:268–278.

Grassmann, W. 1977. Transient solutions in Markovian queueing systems. Computers and Operations Research **4**:47–53.

Grimmett, G. and D. Stirzaker. 2001. Probability and random processes. Oxford University Press, USA.

Mitchell, A. and D. Griffiths. 1980. The finite difference method in partial differential equations. Chichester, Sussex, England and New York, Wiley-Interscience, 1980. 281 p .

Musyl, M., R. Brill, D. Curran, J. Gunn, J. Hartog, R. Hill, D. Welch, J. Eveson, C. Boggs, and R. Brainard, 2001. Ability of archival tags to provide estimates of geographical position based on light intensity. Pages 343–367 *in* J. Sibert and J. Nielsen, editors. Electronic tagging and tracking in marine fisheries. Kluwer, Dordrecht.

Okubo, A. 1980. Diffusion and ecological problems: Mathematical models. Springer-Verlag, Berlin Heidelberg.

Pedersen, M. W., D. Righton, U. H. Thygesen, K. H. Andersen, and H. Madsen. 2008. Geolocation of North Sea cod (*Gadus morhua*) using hidden Markov models and behavioural switching. Canadian Journal of Fisheries and Aquatic Sciences **65**:2367–2377.

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.

Viterbi, A. J. 2006. A personal history of the Viterbi Algorithm. IEEE Signal Processing Magazine **23**:120–142.

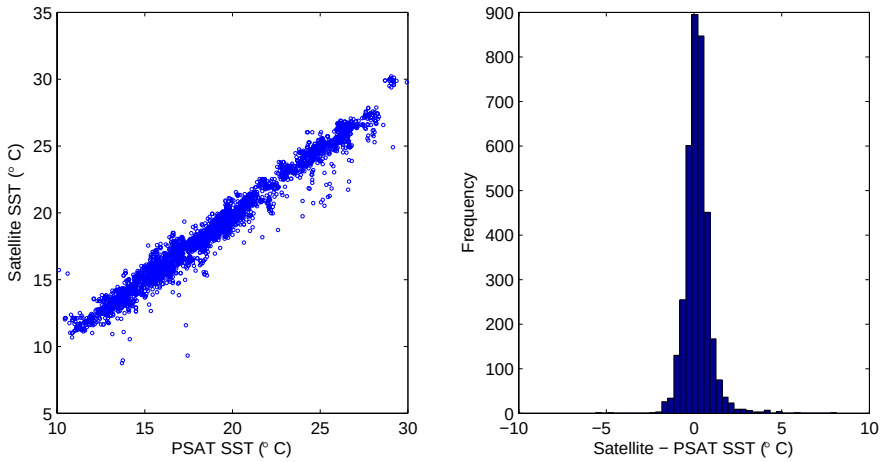


Figure 1: Left: Plot of SST as measured by PSAT versus remotely sensed SST by satellite. A strong linear relation is clear. Right: Histogram of the residuals of a linear regression model for the SST data in the left pane. The errors are approximately Gaussian distributed with zero mean.

Appendix E

Estimation methods for nonlinear state-space models in ecology

---

Authors:
M.W. Pedersen, C.W. Berg, U.H. Thygesen, A. Nielsen and H. Madsen.

Submitted to *Ecological Modelling* September 2010, accepted January 2011.

Estimation methods for nonlinear state-space models in ecology

M.W. Pedersen[*,a,1], C.W. Berg[a,b], U.H. Thygesen[b], A. Nielsen[b], H. Madsen[a]

[a] *Department for Informatics and Mathematical Modelling, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[b] *National Institute of Aquatic Resources, Technical University of Denmark, 2920 Charlottenlund, Denmark*

Key words: AD Model Builder, hidden Markov model, mixed model, Monte Carlo, theta logistic population model, WinBUGS

Abstract

The use of nonlinear state-space models for analyzing ecological systems is increasing. A wide range of estimation methods for such models are available to ecologists, however it is not always clear, which is the appropriate method to choose. To this end, three approaches to estimation in the theta logistic model for population dynamics were benchmarked by Wang (2007) Ecol. Model. 200, pp. 521–528. Similarly, we examine and compare the estimation performance of three alternative methods using simulated data. The first approach is to partition the state-space into a finite number of states and formulate the problem as a hidden Markov model (HMM). The second method uses the mixed effects modeling and fast numerical integration framework of the AD Model Builder (ADMB) open-source software. The third alternative is to use the popular Bayesian framework of BUGS. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

[*] Corresponding author
Email addresses: mwp@imm.dtu.dk (M.W. Pedersen), cbe@aqua.dtu.dk (C.W. Berg), uht@aqua.dtu.dk (U.H. Thygesen), an@aqua.dtu.dk (A. Nielsen), hm@imm.dtu.dk (H. Madsen)
[1] Tel.: +45 4525 3095, fax: +45 4588 2673

1. Introduction

State-space models (SSMs) have become the favored approach in modelling time varying ecological phenomena such as population dynamics (Wang, 2007; Gimenez et al., 2007), animal movement (Patterson et al., 2008) and animal behavior (Morales et al., 2004). SSMs come in a variety of classes depending on the problem type (linear or nonlinear) and the error structure of the data (Gaussian or non-Gaussian). In the linear and Gaussian case an exact solution to the SSM can be found using the Kalman filter (KF), which is the optimal estimator (Madsen, 2008). In case of minor departures from linearity, KF variants, such as the extended KF or unscented KF, can be employed. Both methods are reviewed and discussed by Wang (2007). In cases where the state-space equations are highly nonlinear, it is inappropriate to use any KF variant. For ecological problems Markov chain Monte Carlo (MCMC) is perhaps the most common approach to accommodate model nonlinearities owing to its flexibility and general applicability. In addition, free software for MCMC analysis is available, for example the widely used WinBUGS (Gimenez et al., 2008). An example of non-WinBUGS MCMC population modeling is explained by Wang (2007).

We address three powerful methods for the analysis of nonlinear state-space models, two of which have only gained moderate attention previously within the field of ecology compared to the third. The idea of the first method we present is to discretize the continuous state-space and then reformulate the SSM as a hidden Markov model (HMM) (see Zucchini and MacDonald, 2009). A similar approach was described by Kitagawa (1987). The second method we consider is implemented in the open-source software AD Model Builder (ADMB-project, 2009a). In ADMB the SSM is formulated as a statistical model with mixed effects. A major advantage of ADMB is that it makes efficient use of available computer resources by so-called automatic differentiation. Thirdly, we apply OpenBUGS, which is the open-source version of WinBUGS (Spiegelhalter et al., 1996). BUGS is flexible and therefore widely used in modeling ecological systems (Gimenez et al., 2008).

To broaden the perspective of this study we apply the three methods to simulated data from the theta logistic population model, which is a nonlinear SSM. The same example was analyzed by Wang (2007). The performance of the three methods is summarized with respect to a range of aspects: complexity of implementation, computing time, estimation accuracy, limiting assumptions, and algorithmic design. Algorithmic design refers to the amount of subjective tuning required before actual estimation can begin. Because of reduced subjective influence, methods with fewer tuning parameters are often preferable. Finally, we discuss some differences between Bayesian (BUGS) and frequentist (HMM, ADMB) methods.

2. Methods

A state-space model describes the dynamics of a latent state ($\mathbf{X}_t$) and how data (Y_t) relate to this state. An important feature of SSMs is their ability to model random variations in the latent state and in data. For $t \in \{1, \ldots, N\}$ the general system and observation equations of the SSM are respectively $\mathbf{X}_t = g(t, \mathbf{X}_{t-1}, \mathbf{e}_t)$, and $\mathbf{Y}_t = h(t, \mathbf{X}_t, \mathbf{u}_t)$, where $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{Q}_t)$ is the system error and $\mathbf{u}_t \sim N(0, \mathbf{R}_t)$ is the observation error. Here, "$\sim N(\cdot)$" means Gaussian distributed. Because of the possible nonlinearity of g and h, advanced filtering and smoothing methods must be employed to gain meaningful estimates of $\mathbf{X}_t$. In this respect, the extended Kalman filter, the unscented Kalman filter, and Bayesian filtering e.g. using Markov chain Monte Carlo (MCMC) sampling or BUGS are common approaches. Alternative methods for nonlinear state estimation are hidden Markov models (HMMs, Zucchini and MacDonald, 2009) and mixed effects models using the software AD Model Builder (ADMB). ADMB is freely available and open-source (ADMB-project, 2009a).

2.1. Benchmarking of estimation methods

The log-transformed theta logistic population growth model (Wang, 2007) was used as benchmark example for assessing the estimation performance of HMM, ADMB and BUGS. The system and observation equations for this model are

$$X_t = X_{t-1} + r_0\left(1 - \left(\frac{\exp(X_{t-1})}{K}\right)^{\theta}\right) + e_t, \tag{1}$$

$$Y_t = X_t + u_t, \tag{2}$$

where $e_t \sim N(0, Q)$ and $u_t \sim N(0, R)$.

Following Wang (2007), two different tests of the methods were carried out:

1. State estimation performance with known parameter values, i.e. the ability of the methods to estimate the population level x_t for all t. Obviously, this test is free of Bayesian prior assumptions on parameters.
2. Estimation of states and all five model parameters, $\boldsymbol{\lambda} = [\log(\theta), \log(r_0), K, \log(Q), \log(R)]$, simultaneously. This situation is common in practice if model parameters cannot be estimated from independent data. Notice that parameters that may yield estimates close to zero are log-transformed to avoid invalid parameter values.

Specifically for test 1, $T = 2000$ data replicates were simulated with $N = 200$, $K = 1000$, $Q = 0.01$, $R = 0.04$, and the initial state $x_0 = 3$ using twelve different sets of the θ and r_0 parameters (see Table 1). The performance of the methods was evaluated using an estimate of the state estimation error:

$$\text{RMSE} = \frac{1}{T}\sum_{i=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}(\widehat{x_{i,t}} - x_t)^2\right)^{\frac{1}{2}}, \tag{3}$$

where $\widehat{x_{i,t}}$ is the state estimate for replicate i at time t, and x_t is the true state at time t.

Specifically for test 2, two datasets were simulated using two other sets of parameter values : $\boldsymbol{\lambda}_1 = (\theta = 0.5, r_0 = 0.1, K = 900, Q = 0.01, R = 0.04)$ and $\boldsymbol{\lambda}_2 = (\theta = 1.5, r_0 = 0.1, K = 900, Q = 0.01, R = 0.04)$ with the number of data points $N = 200$. Parameter estimates for these data using the three methods were found similarly to the study of Wang (2007). We further used these two parameter configurations to generate plots of the joint profile likelihood surfaces for r_0 and θ, which were transformed to confidence contours via a χ^2-distribution as in Polansky et al. (2009). The simulated data sets for $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are available in the supplementary material to enable comparison of our results with future estimation methods. Additionally for test 2 we estimated all five model parameters along with 95% intervals using $T = 200$ of the data sets simulated for test 1. Inspired by Lambert et al. (2005), the purpose here was to evaluate the frequentist properties of the intervals provided by the three estimation methods.

2.2. Hidden Markov model with Matlab

The integrals involved in the prediction, filtering, and smoothing steps for nonlinear SSMs (see e.g. eq. 2.2, 2.3, 2.5 in Kitagawa, 1987) can, in general, not be solved analytically. However, by partitioning the continuous state-space uniformly into n parts the solution can be computed using hidden Markov models (HMMs) (Zucchini and MacDonald, 2009). See de Valpine and Hastings (2002) for an ecologically motivated study using a similar method. A state is denoted Ω_i, where $i \in \{1, 2, \ldots, n\}$. The probability distribution of the state given the observations $\mathcal{Y}_t$ is $P(\mathbf{X}_t \in \Omega_i|\mathcal{Y}_t) = p_t(i|\mathcal{Y}_t)$ which are collected in the row vector $\mathbf{p}_t(\mathcal{Y}_t) = \{p_t(i|\mathcal{Y}_t)\}$. The transition probability of jumping from Ω_i to Ω_j (see Figure 1) is

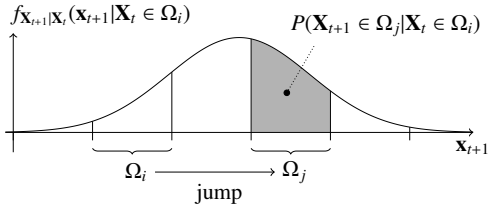


Figure 1: Probability of a jump (transition) from the state Ω_i to the state Ω_j in the time interval from t to $t+1$ in a HMM. The shaded area corresponds to the integral in (4).

$$\begin{aligned} p_t(i,j) &= P(\mathbf{X}_{t+1} \in \Omega_j | \mathbf{X}_t \in \Omega_i) \\ &= \int_{\Omega_j} f_{\mathbf{X}_{t+1}|\mathbf{X}_t}(\mathbf{x}_{t+1}|\mathbf{X}_t \in \Omega_i)\, d\mathbf{x}_{t+1}. \end{aligned} \tag{4}$$

For one-dimensional problems Ω_i are intervals on the line, in two dimensions Ω_i are areas, and analogously for higher dimensions. Note that the $n \times n$ probability transition matrix $\mathbf{P}_t = \{p_t(i,j)\}$ is not homogeneous, i.e. the transition probabilities may change as a function of time as indicated by (1). Now, the HMM prediction, filtering, and smoothing equations are respectively

$$\begin{aligned} \mathbf{p}_t(\mathcal{Y}_{t-1}) &= \mathbf{p}_{t-1}(\mathcal{Y}_{t-1})\mathbf{P}_{t-1}, \\ \mathbf{p}_t(\mathcal{Y}_t) &= \psi_t^{-1}\mathbf{p}_t(\mathcal{Y}_{t-1}) \odot \mathbf{L}(y_t), \\ \mathbf{p}_t(\mathcal{Y}_N) &= \mathbf{p}_t(\mathcal{Y}_t) \odot \left[\{\mathbf{p}_{t+1}(\mathcal{Y}_N) \oslash \mathbf{p}_{t+1}(\mathcal{Y}_t)\}\, \mathbf{P}_t^T\right] \end{aligned}$$

where '$\odot$' and '$\oslash$' are elementwise matrix multiplication and division, respectively. The likelihood of the observations $\mathbf{L}(y_t)$ is a row vector with elements $p_t(y_t|i)$ and $\psi_t = \mathbf{p}_t(\mathcal{Y}_{t-1}) \cdot \mathbf{L}(y_t)^T$ is a normalization constant with '$\cdot$' denoting dot product. The estimate of the state given all N observations is simply the mean of the distribution $\mathbf{p}_t(\mathcal{Y}_N)$.

Using the above scheme we can estimate the unknown parameters ($\boldsymbol{\lambda}$) of the SSM by maximizing the likelihood function

$$\mathcal{L}(\boldsymbol{\lambda}|\mathcal{Y}_N) = f_{\mathcal{Y}_N}(\mathcal{Y}_N|\boldsymbol{\lambda}) = [\mathbf{L}(y_1) \cdot \mathbf{1}] \prod_{t=2}^{N} \psi_t, \tag{5}$$

as in Kitagawa (1987), where $\mathbf{1}$ is a column vector of ones. The maximum likelihood (ML) estimate of the model parameters $\widehat{\boldsymbol{\lambda}}$ is found by optimizing (5) as a function of $\boldsymbol{\lambda}$. The covariance matrix of $\widehat{\boldsymbol{\lambda}}$ is approximated by the inverse Hessian of the likelihood function at the optimum $\widehat{\boldsymbol{\lambda}}$. This approximation is appropriate because the ML estimate is asymptotically Gaussian under certain regularity conditions (Cappé et al., 2005). Thus, confidence intervals can be constructed using the approximated covariance matrix. Under parameter transformations it is important to construct the confidence intervals in the transformed parameters and then reverse transform the computed confidence limits.

When analyzing the theta logistic model we set $n = 251$. The bounds of the discrete state-space are chosen such, that the probability of the true state falling outside the grid is negligible. That is, we use the observation model (2)

to determine bounds that envelope the true latent state with a probability close to 1. This approach is similar to the one used in de Valpine and Hastings (2002). Details on grid specification can be found in the supplementary material containing model code.

The HMM code provided in the supplementary material was written in Matlab, but the method is not language specific. Matlab was chosen because it is widely used and has a syntax which is relatively easy to understand even for non-Matlab users.

2.3. Mixed effects model with AD Model Builder

Hierarchical mixed effects models are an alternative framework for analyzing nonlinear SSMs. The states are the random effects of the model and are collectively referred to as $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. Here, as in Madsen and Thyregod (2010), we specify a model for the data conditional on the unobserved random effects, $f_{\mathcal{Y}_N|\mathcal{X}}(\mathcal{Y}_N|\mathcal{X}, \boldsymbol{\lambda}_a)$ which corresponds to (2). We also specify a model for the random effects, $f_{\mathcal{X}}(\mathcal{X}|\boldsymbol{\lambda}_b)$ which corresponds to (1). The joint density of random effects and observations conditional on the parameters is

$$f_{\mathcal{X},\mathcal{Y}_N}(\mathcal{X}, \mathcal{Y}_N|\boldsymbol{\lambda}) = f_{\mathcal{X}}(\mathcal{X}|\boldsymbol{\lambda}_b) f_{\mathcal{Y}_N|\mathcal{X}}(\mathcal{Y}_N|\mathcal{X}, \boldsymbol{\lambda}_a).$$

To obtain the marginal likelihood for estimating $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_a, \boldsymbol{\lambda}_b\}$ we integrate over the unobserved random effects

$$\mathcal{L}(\boldsymbol{\lambda}|\mathcal{Y}_N) = f_{\mathcal{Y}_N}(\mathcal{Y}_N|\boldsymbol{\lambda}) = \int_{\mathbb{R}^N} f_{\mathcal{X},\mathcal{Y}_N}(\mathcal{X}, \mathcal{Y}_N|\boldsymbol{\lambda}) d\mathcal{X}. \tag{6}$$

The N-dimensional integral in (6) is generally challenging to solve, and for nonlinear mixed models we must resort to numerical approximation methods for estimating the model parameters. An efficient and widely used method for this is the Laplace approximation (Wolfinger and Xihong, 1997), which replaces the integrand with a second order Taylor expansion around the optimum of the log-likelihood function. This allows for elimination of the integral, because the second-order Taylor expansion can be formulated as a known constant multiplied by a multivariate Gaussian density, which integrates to unity. For nonlinear models the distribution of the random effect may not be Gaussian. Then the Laplace approximation is not exact. In particular for multi modal distributions one should use the Laplace approximation with caution. Still, when analyzing nonlinear models with moderately skewed unimodal distributions good results can be obtained with the Laplace approximation (Vonesh, 1996; Mortensen, 2009). In any case it is important to investigate if the approximation is critically violated e.g. by Monte Carlo sampling from the random effects distribution.

Even with the Laplace approximation maximization of the marginal log-likelihood with respect to $\boldsymbol{\lambda}$ is challenging. A computationally efficient method is to combine the Laplace approximation with so-called automatic differentiation (AD, Skaug and Fournier, 2006). AD is a technique for finding the gradient of a function h (in our case the log-likelihood), provided that h can be expressed in computed code. Evaluating h using AD gives the function value along with the gradient of h at the point of evaluation. The gradient is computed using the chain rule of calculus on every operation in the code that contributes to the value of h. For efficient maximization of the Laplace approximation of the marginal log-likelihood with respect to $\boldsymbol{\lambda}$, up to third order partial derivatives must be found. Skaug and Fournier (2006) show how this can be accomplished by repeated use of AD.

The above procedure is implemented in AD Model Builder (ADMB), which we use to analyze the theta logistic model. ADMB is an open-source software package and programming language based on C++. It includes a function minimizer for ML parameter estimation and a random effects module, which utilizes the Laplace approximation for integration of random effects. Standard deviations for constructing confidence intervals are calculated using the delta method (Oehlert, 1992) and automatically reported on all estimated quantities. The covariance matrix for all

states in an SSM is a banded matrix (Skaug and Fournier, 2006). ADMB can exploit this property by using the SEPARABLE FUNCTION construct (ADMB-project, 2009b) to gain significant speed improvements. Other than this useful property ADMB has no tuning parameters as such.

2.4. Monte Carlo estimation with BUGS

Finally, we analyze the theta logistic model using the Bayesian modeling language BUGS, which is an MCMC estimation method (Spiegelhalter et al., 1996). BUGS is a popular tool in ecological modeling (e.g. Gimenez et al., 2007; Jonsen et al., 2005; Schofield et al., 2009). BUGS is best known in the WinBUGS form which has a graphical user interface. Here, however, we use the open-source alternative OpenBUGS, yet the BUGS code provided in the supplementary material is compatible with WinBUGS.

A Bayesian analysis requires that prior distributions are specified for the model parameters. The type of prior distributions and parameter values related to these distributions should reflect the *a priori* knowledge that is available about the model parameters. BUGS then uses Gibbs sampling (Casella and George, 1992) to explore the posterior distribution of the parameter and state-space by incorporating the information specified by the priors, the state-space model, and the observed data. The Gibbs algorithm exploits that sampling the posterior is sometimes simpler via its conditional distributions rather than directly from the joint distribution. This is the case for state-space models where direct sampling of the posterior for states and parameters is difficult. Instead, sampling model parameters from priors and then sampling X_t conditional on model parameters and remaining states $(X_1, \ldots, X_{t-1}, X_{t+1}, \ldots, X_N)$ for all t is simple using (1). The sampling algorithm applied by BUGS in specific cases depends on the form and type of the conditional distribution, and also on the composition of priors on model parameters (see Spiegelhalter et al., 1996, 2003, for details).

We consider the common practical situation where *a priori* knowledge is unavailable and estimation therefore relies entirely on information in data. How to specify vague (or uninformative) priors is a topic of on-going research (Gelman, 2006; Lambert et al., 2005), which is outside the scope of this study. One suggested vague prior is a uniform distribution with wide support (Spiegelhalter et al., 1996). So, we choose a uniform prior for K, and uniform priors for $\log\theta$ and $\log r_0$ that were much wider than the natural biological bounds for the parameter values. By log-transforming θ and r_0 biological meaningful (i.e. positive) parameter values are ensured. The state-space formulation implies that the variance parameters Q and R are non-zero and therefore also require prior distributions. It is common to assign vague inverse-gamma distributed priors to variance parameters (Spiegelhalter et al., 2003; Lambert et al., 2005). Gelman (2006), however, recommends using a uniform prior on the log-transformed standard deviation. Therefore, to asses the sensitivity of the estimation results to the choice of prior we perform BUGS estimation in two separate cases: BUGS1 using an inverse-gamma distribution for Q and R, and BUGS2 using a uniform distribution on the log-transformed standard deviation, i.e. $0.5\log(Q)$ and $0.5\log(R)$.

Estimation using BUGS involves a number of tuning parameters: The initial values for the sampling scheme can be found in the supplementary material online along with the specifics of the priors. The total number of generated samples was 100,000 with 50,000 used for burn-in. The appropriate number of samples was found iteratively by repeated application of Geweke Z score test for convergence (Geweke, 1992). The BUGS thinning rate was 50 (for reducing sample autocorrelation, which was apparent for θ and r_0 at lower thinning rates). With these values of the tuning parameters we get an effective sample size of 1000. For summarizing the estimation results the maximum *a posteriori* (MAP) parameter estimates along with 95% credible intervals are reported (where the lower bound equals the 2.5% quantile and the upper bound equals the 97.5% quantile of the posterior distribution).

3. Results

State estimation results for the three methods using known parameter values were practically identical (Table 1). ADMB was an order of magnitude faster than HMM, which, in turn, was an order of magnitude faster than BUGS (Table 2). State estimation using estimated parameter values also gave practically identical results for all three methods (Figure 2). Regarding ML parameter estimation and confidence intervals (CIs) for λ_1 and λ_2, HMM and ADMB performed almost identically (Table 3). Likewise, MAP estimates and credible intervals provided by BUGS1 and BUGS2 were overall similar in the λ_2 case. In the λ_1 case, however, BUGS1 MAP estimates of θ and r_0 were markedly lower and higher respectively than the estimates provided by HMM, ADMB, and BUGS2. Perhaps most surprisingly was the upper limit of the credible interval for K seemingly quite sensitive to the choice of prior employed by BUGS, and in both cases considerably higher than the HMM and ADMB CI upper limits. Some notable differences between CIs and credible intervals were present for θ, K, and r_0 in the λ_1 case (Table 3), with BUGS generally being more conservative and providing wider intervals (in the log domain). Inspection of the joint profile likelihood surfaces for θ and r_0 revealed that contour lines closely approximated elliptical shapes for λ_2 (Figure 4, panel B), thus indicating that the quadratic approximation used by HMM and ADMB was appropriate. For λ_1, on the other hand, the quadratic approximation was only appropriate until the 65% confidence limit where the contour shape started to diverge from the elliptical shape (Figure 4, panel A). If comparing the limits of the intervals provided by all three methods for the λ_1 case (Table 3) with the extents of the likelihood surface (Figure 4, panel A), it is clear that neither credible intervals nor CIs captured the actual range of plausible parameter values.

Visualizing the empirical distributions of the $T = 200$ parameter estimates (Figure 3) showed largely identical results for all three methods. For all parameters the average 95% CIs provided by HMM and ADMB closely approximated the 2.5% and 97.5% quantiles of the corresponding empirical distribution. Similar results were observed for BUGS1 and BUGS2 for parameters R and Q. Regarding the three remaining parameters θ, K, and r_0, on the other hand, the average credible intervals were markedly wider than the corresponding quantiles of their empirical distribution, and therefore also wider than their CI counterparts. The difference in results between the two vague priors (BUGS1 and BUGS2) was minimal except for the credible intervals for K where BUGS2 gave wider intervals than BUGS1. Since both priors have been regarded in the literature as vague their influence on the resulting intervals is surprising. Computing times for parameter estimation showed that ADMB, again, was significantly faster than HMM and BUGS (Table 2). Interestingly, BUGS1 was considerably (six times) faster than BUGS2. This results can most likely be ascribed to BUGS using different sampling algorithms in the two cases.

4. Discussion

Dynamical processes are prevalent in ecology. State-space models are commonly used in the analysis of such nonlinear processes because they join separate models of the ecological system and the observation process. This paper assessed the performance of three methods for estimation in nonlinear state-space models: an approach using hidden Markov models (HMM), the open-source AD Model Builder framework (ADMB), and the BUGS language. HMM and ADMB are frequentist (non-Bayesian) methods, while BUGS is Bayesian. To facilitate a transparent comparison among available estimation methods we considered the theta logistic population model, which Wang (2007) analyzed with three other methods (extended Kalman filter, the unscented Kalman filter and a Metropolis-Hastings approach). To increase accessibility, the computer code for our three methods can be found in the online supplementary material.

The state estimation root mean square errors (RMSEs) of HMM, ADMB, and BUGS (Table 1) were lower than those for the three methods presented by Wang (2007), his Table 1. The 95% intervals for the parameter estimates of θ

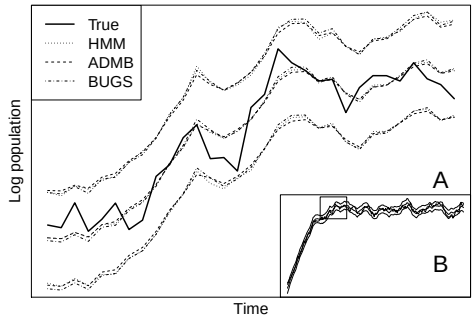


Figure 2: State estimation of the theta logistic model with 95% intervals using the estimated parameter values in Table 3. True states were generated using λ_2 ($\theta = 1.5$). Panel A is a zoom of a part of the full time series indicated by the small box in panel B. Clearly in this case, HMM, ADMB, and BUGS gave close to identical state estimation results.

Sim.	r_0	θ		RMSE	
no.			HMM	ADMB	BUGS
1	0.1	0.5	0.100	0.100	0.100
2	0.5	0.5	0.099	0.099	0.100
3	0.75	0.5	0.097	0.097	0.097
4	1.0	0.5	0.095	0.095	0.095
5	0.1	1.0	0.100	0.100	0.100
6	0.5	1.0	0.095	0.095	0.095
7	0.75	1.0	0.091	0.092	0.092
8	1.0	1.0	0.090	0.090	0.090
9	0.1	1.5	0.100	0.100	0.100
10	0.5	1.5	0.092	0.092	0.092
11	0.75	1.5	0.091	0.091	0.091
12	1.0	1.5	0.096	0.096	0.096

Table 1: Performance of state estimation as defined by Eq. (3) for HMM, ADMB, and BUGS.

	HMM	ADMB	BUGS1	BUGS2
State est.	6.12 s	0.49 s	58 s	58 s
Par. est.	225 s	2.5 s	118 s	614 s

Table 2: Computing times for HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations). All times are for a single dataset run on the same computer.

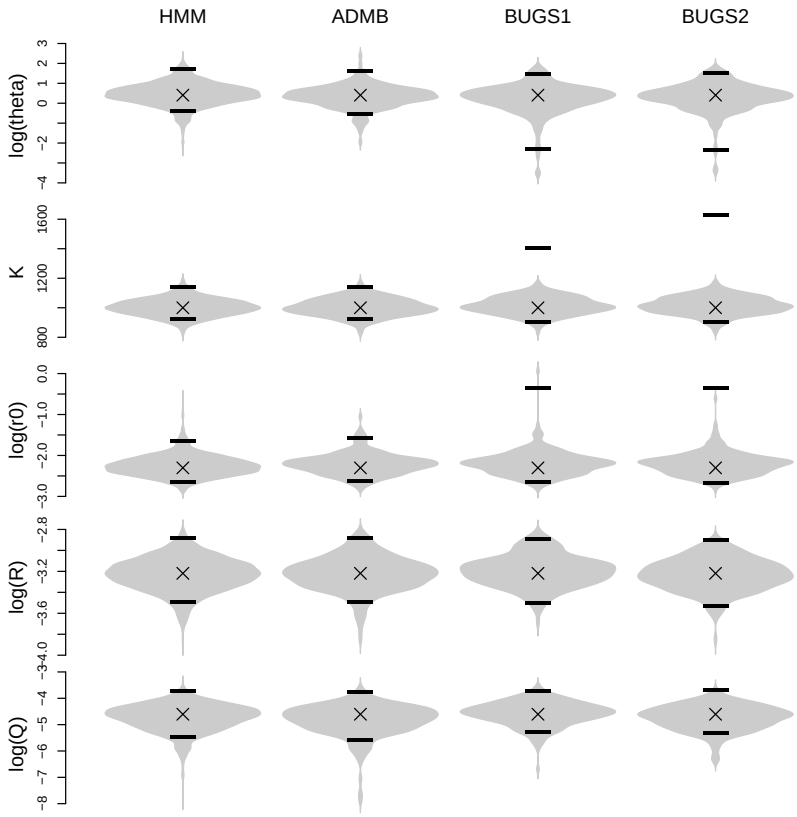


Figure 3: Violin plots showing the empirical distribution of $T = 200$ parameter estimates. Data used for estimation were simulated with the parameter configuration $\boldsymbol{\lambda} = (\theta = 1.5, r_0 = 0.1, K = 1000, Q = 0.01, R = 0.04)$. Crosses indicate the true parameter values, $\boldsymbol{\lambda}$. Horizontal lines indicate the average limits of the 200 individual 95% intervals.

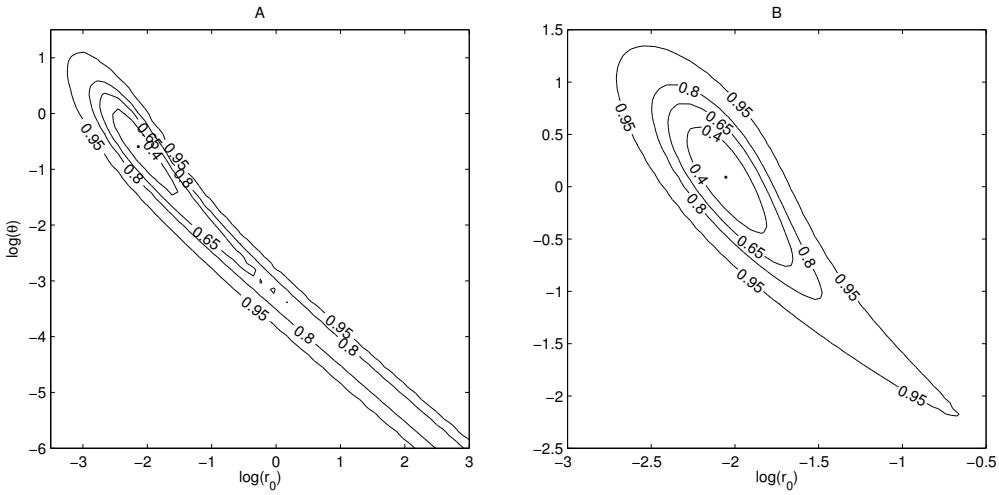


Figure 4: Joint profile likelihood surfaces for two simulated data sets of the theta-logistic model (see also Table 3). Panel A: Parameters used for simulation $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$. Panel B: $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$. Following Polansky et al. (2009) the joint profile log-likelihood surfaces have been transformed to confidence contours via a χ^2-distribution of the profiled models versus the model where all five parameters are estimated. Dots indicate the minima of the transformed surfaces equivalent to the maximum likelihood (ML) point. Both surfaces have elliptically shaped contours in proximity to the ML point in which case a quadratic approximation as used by HMM and ADMB is appropriate. While the surface for λ_2 (panel B) is close to quadratic even at the 95% level, the surface for λ_1 (panel A) departs from the quadratic shape at the 65% level.

	HMM		ADMB		BUGS1		BUGS2	
	ML est.	95% conf. intv.	ML est.	95% conf. intv.	MAP est.	95% cred. intv.	MAP est.	95% cred. intv.
λ_1								
θ	0.588	0.134 – 2.588	0.583	0.129 – 2.640	0.374	0.0210 – 1.446	0.538	0.020 – 1.496
K	829.3	643.3 – 1015	829.5	639.2 – 1020	860.0	629.2 – 1900	834.0	638.3 – 4957
r_0	0.116	0.046 – 0.298	0.117	0.045 – 0.305	0.135	0.053 – 1.667	0.118	0.045 – 1.666
R	0.041	0.032 – 0.053	0.041	0.032 – 0.053	0.042	0.031 – 0.054	0.041	0.031 – 0.052
Q	0.0092	0.0052 – 0.016	0.0092	0.0051 – 0.017	0.011	0.0055 – 0.017	0.0099	0.0060 – 0.018
λ_2								
θ	1.098	0.412 – 2.926	1.079	0.402 – 2.902	1.006	0.043 – 2.551	1.037	0.043 – 2.869
K	886.9	792.7 – 981.0	887.0	790.5 – 983.5	891.3	769.3 – 1121	910.0	774.9 – 1097
r_0	0.128	0.082 – 0.201	0.129	0.081 – 0.203	0.127	0.078 – 1.136	0.134	0.074 – 1.032
R	0.043	0.032 – 0.056	0.043	0.032 – 0.056	0.043	0.031 – 0.056	0.044	0.032 – 0.056
Q	0.0082	0.0038 – 0.018	0.0081	0.0045 – 0.015	0.0094	0.0041 – 0.018	0.0086	0.0043 – 0.019

Table 3: Parameter values estimated by HMM, ADMB, BUGS1 (inverse-Gamma prior on variances), and BUGS2 (uniform prior on log-standard deviations) with related 95% intervals. Data were simulated with the listed true parameter values: $\lambda_1 = (\theta = 0.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ and $\lambda_2 = (\theta = 1.5,\ r_0 = 0.1,\ K = 900,\ Q = 0.01,\ R = 0.04)$ of the theta logistic model.

provided by our three methods all included the true values (Table 3). Note that they also included $\theta = 1$, which means that the models could not distinguish between a concave and convex relation between population size and growth rate. This is in contrast with the credible intervals in Wang (2007), his Table 2, that excluded $\theta = 1$, however three out of six of his credible intervals also excluded the true parameter value, which is of some concern.

Recent studies have indicated that θ and r_0 of the theta logistic model (1) can be difficult to identify for certain data sets (Polansky et al., 2009). This is the case because given $\theta < 1$ similar model dynamics can be generated for different values of θ (Clark et al., 2010). Supporting this, a joint profile likelihood surface for $\log\theta$ and $\log r_0$ showed that combinations of different values for the two parameters may fit data equally well, i.e. result in practically identical model likelihoods (Figure 4, panel A, data generated with $\theta = 0.5$). Still parameters estimated by HMM and AMDB were reasonably accurate (Table 3, case $\boldsymbol{\lambda}_1$), however the confidence intervals (CIs) were too narrow when compared to the contours of the confidence regions in Figure 4, panel A. This result underlines the importance of validating the quadratic approximation to the log-likelihood function employed by HMM and ADMB before using it to construct CIs. The credible intervals from BUGS were wider and therefore more realistic than the CIs provided by HMM and ADMB, yet the interval bounds were narrower than the range of plausible models indicated by the profile likelihood surface. A possible explanation for this difference is that a substantial Monte Carlo sample size may be required to fully explore the posterior distribution when two parameters are highly correlated (Gamerman, 1997). If complications with parameter identifiability as illustrated in Figure 4, panel A, are encountered in practical situations it is recommended to switch to a simpler model with fewer parameters, e.g. by setting $\theta = 1$ (Clark et al., 2010).

For the data set generated with $\theta = 1.5$, the joint profile likelihood surface for $\log\theta$ and $\log r_0$ was well approximated by a quadratic function (Figure 4, panel B). Thus, log-transforming θ and r_0 in the theta-logistic model avoids a boomerang-shaped likelihood surface (see e.g. Fig. 2 in Polansky et al., 2009), which deviates considerably from a quadratic function. Thus, the CIs computed for HMM and ADMB in the log-transformed parameter space (Table 3, case $\boldsymbol{\lambda}_2$) corresponded well to the confidence contours in Figure 4, panel B. For BUGS credible intervals the conclusion was the same. Similarly to Lambert et al. (2005), the frequentist properties of the three estimation methods were evaluated. To this end we used so-called violin-plots (Figure 3), where the empirical distribution of 200 parameter estimates was compared with the average of the corresponding 200 95% interval bounds. In discussing our results it is important to stress that CIs provided by frequentist methods (HMM and ADMB) and credible intervals provided by Bayesian methods (BUGS) have fundamentally different interpretations. A 95% CI is an interval which contains the true parameter in 95% of a large number of repeated experiments. Conversely, a 95% credible interval is an interval which has a 95% posterior probability of containing the parameter for the experiment at hand. From Figure 3 it was evident that the CIs were consistent with corresponding quantiles of the empirical distributions. This further confirms the validity of the quadratic approximation of the log-likelihood function. The empirical distributions of the BUGS parameter estimates under vague prior assumptions were largely identical to their HMM and ADMB counterparts. However, Figure 3 showed that even when assigning vague priors it cannot be expected that credible intervals coincide with frequentist CIs, which by definition do not incorporate *a priori* knowledge. In addition, considerable differences in credible intervals were present between the two BUGS analyses using different vague priors (Figure 3). Thus, it is crucial, when employing Bayesian methods in the absence of *a priori* knowledge, to assess the sensitivity of credible intervals to the choice of distribution for the vague prior.

ADMB uses automatic differentiation to estimate the states and parameters of the model, which is the main reason for its computing time superiority (Table 2). This advantage will only increase further as models become more complex and the number of parameters grows. The main disadvantage of ADMB is, that the Laplace approximation for the density of the random effects (here equivalent to the latent states) must be reasonable. In our test cases the latent state estimation of ADMB was close to identical to the HMM and BUGS results (Figure 2), which justifies using

the Laplace approximation. If results from alternative methods are not available, the quality of the approximation can be assessed using the built-in importance sampling functionality (p. 35, ADMB-project, 2009b). Another possible complication of ADMB is that some programming experience in C++ is required. The HMM approach, on the other hand, has the advantage of being language independent, i.e. the method can be implemented in any programming language, for which a function optimizer is available. The programming background of the modeler is therefore of minor concern. The computing speed of the HMM approach is, at worst, proportional to the number of grid cells squared, a number which grows rapidly with increasing state dimension. Thus, HMMs are best suited for one or two-dimensional problems. BUGS depends less on state dimension because it is Monte Carlo based and it requires no density approximations nor differentiability. Consequently, BUGS is flexible and applicable to the widest variety of problems of the three methods we have examined. In addition, WinBUGS (Spiegelhalter et al., 2003) can be used to view and produce BUGS code graphically. This further increases the accessibility of the method.

BUGS and Monte Carlo based methods in general have tuning parameters that cannot be estimated from data and therefore require subjective input from the modeler. The tuning parameters include the number of samples, burn-in time, thinning rate, convergence assessment, and choice of prior distribution, all of which influence the estimation results significantly. This fact is underlined in the BUGS manual (Spiegelhalter et al., 1996, p. 1), and it is emphasized that the modeler using BUGS must have a sound understanding of the Gibbs sampler. Our results supported this in that computing times (Table 2) and interval estimation (Figure 3) were significantly influenced by the choice of prior. In contrast, ADMB has no tuning parameters as such, but it does have certain options that are more or less relevant depending on the type of problem, for example the SEPARABLE FUNCTION construct. HMM has two tuning parameters: the extent of the grid and the grid resolution. Limiting the state-space involves a risk of truncating the latent state path. To minimize this risk the approach of de Valpine and Hastings (2002) was followed, where bounds are chosen so wide that the probability of latent path truncation is negligible. Naturally, wider grid extents and higher grid resolution entail an increase in computation time. Thus, determining the value of these parameters is a tradeoff between computing speed and accuracy of results. Generally, if one is uncertain about the grid specifications, we recommend to start with a wide and coarse grid to get preliminary results, and then adapt extents and refine the grid accordingly if needed. If the conclusion is unchanged on the adapted grid there is strong evidence that the latent path is enclosed and properly resolved by the discretization.

5. Conclusion

In summary, the three methods considered in this paper are all powerful approaches to nonlinear state-space modelling of ecological systems. ADMB is by far the fastest method owing to its use of the Laplace approximation and automatic differentiation. This limits ADMB to problems where the state distributions are unimodal, which, however, is the case in the majority of practical examples. In contrast, HMM and BUGS are more general and are able to handle arbitrary state distributions. HMM requires specification of a spatial grid and is limited to problems with low state dimensions, say below four. BUGS has fewest model restrictions, but requires specification of prior information and other subjective input from the modeler in the form of algorithmic tuning parameters.

State-space methods provide a natural paradigm for ecosystem modeling. Thus, it is imperative that the ecological community is alert to progress in other scientific fields where state-space models are used and developed. This paper evaluated the performance, with respect to estimation accuracy and speed, of three advanced methods for state-space analysis. The study showed that state and parameter estimation performance for all three methods was largely identical, however with BUGS providing overall wider credible intervals for parameters than HMM and ADMB confidence intervals.

References

ADMB-project, 2009a. AD Model Builder: Automatic Differentiation Model Builder. Developed by David Fournier and freely available from admb-project.org.

ADMB-project, 2009b. Random effects in AD Model Builder: ADMB-RE user guide.

Cappé, O., Moulines, E., Ryden, T., 2005. Inference in hidden Markov models. Springer.

Casella, G., George, E., 1992. Explaining the Gibbs sampler. American Statistician 46 (3), 167–174.

Clark, F., Brook, B., Delean, S., Reşit Akçakaya, H., Bradshaw, C., 2010. The theta-logistic is unreliable for modelling most census data. Methods in Ecology and Evolution 1 (3), 253–262.

de Valpine, P., Hastings, A., 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs 72, 57–76.

Gamerman, D., 1997. Sampling from the posterior distribution in generalized linear mixed models. Statistics and Computing 7 (1), 57–68.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian analysis 1 (3), 515–533.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Barnardo, J., Berger, J., David, A., Smith, A. (Eds.), Bayesian statistics 4. Clarendon Press, Oxford, UK, pp. 169–193.

Gimenez, O., Bonner, S., King, R., Parker, R., Brooks, S., Jamieson, L., Grosbois, V., Morgan, B., Thomas, L., 2008. WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In: David L Thomson, E. G. C., Conroy, M. J. (Eds.), Modeling Demographic Processes In Marked Populations. Vol. 3 of Mathematics and statistics. Springer US, pp. 883–915.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J., Pradel, R., 2007. State-space modelling of data on marked individuals. Ecol. Model. 206 (3-4), 431–438.

Jonsen, I. D., Flemming, J. M., Myers, R. A., 2005. Robust state-space modeling of animal movement data. Ecology 86 (11), 2874–2880.

Kass, R., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Am. Stat. Assoc. 91 (435).

Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series. J. Am. Stat. Assoc. 82 (400), 1032–1041.

Lambert, P., Sutton, A., Burton, P., Abrams, K., Jones, D., 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine 24 (15), 2401–2428.

Madsen, H., 2008. Time series analysis. Chapman & Hall/CRC London, London.

Madsen, H., Thyregod, P., 2010. Introduction to general and generalized linear models. Chapman & Hall/CRC, London.

Morales, J., Haydon, D., Frair, J., Holsinger, K., Fryxell, J., 2004. Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology 85 (9), 2436–2445.

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

Oehlert, G. W., 1992. A note on the delta method. American Statistician 46 (1), 27–29.

Patterson, T., Thomas, L., Wilcox, C., Ovaskainen, O., Matthiopoulos, J., 2008. State–space models of individual animal movement. Trends Ecol. Evol. 23 (2), 87–94.

Polansky, L., De Valpine, P., Lloyd-Smith, J., Getz, W., 2009. Likelihood ridges and multimodality in population growth rate models. Ecology 90 (8), 2313–2320.

Schofield, M., Barker, R., MacKenzie, D., 2009. Flexible hierarchical mark-recapture modeling for open populations using WinBUGS. Environmental and Ecological Statistics 16 (3), 369–387.

Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-gaussian hierarchical models. Comput. Stat. Data An. 51 (2), 699–709.

Spiegelhalter, D., Thomas, A., Best, N., Gilks, W., 1996. Bayesian inference using Gibbs sampling. Version 0.5,(version ii), MRC Biostatistics Unit, Cambridge.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. WinBUGS user manual. Version 1.4. MRC Biostatistics Unit, Cambridge, UK.

Vonesh, E. F., 1996. A note on the use of laplace's approximation for nonlinear mixed-effects models. Biometrika 83 (2), 447–452.

Wang, G., 2007. On the latent state estimation of nonlinear population dynamics using Bayesian and non-Bayesian state-space models. Ecol. Model. 200 (3-4), 521–528.

Wolfinger, R., Xihong, L., 1997. Two Taylor-series approximation methods for nonlinear mixed models. Comput. Stat. Data An. 25 (4), 465–490.

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

Appendix F

Individual based population inference using tagging data

Authors:
M.W. Pedersen, U.H. Thygesen, H. Baktoft and H. Madsen.

Published in:
IMM Technical research report (September 2010).

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

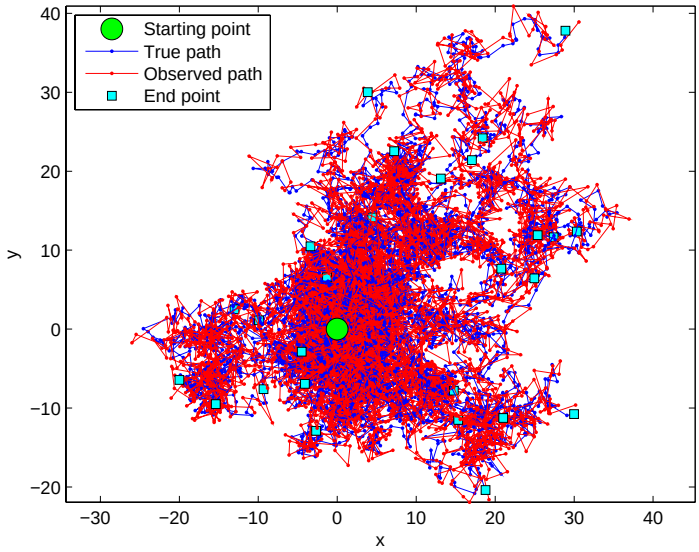


Figure 1: Simulated data from $M = 30$ individuals with a biased random walk in two dimensions.

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

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.

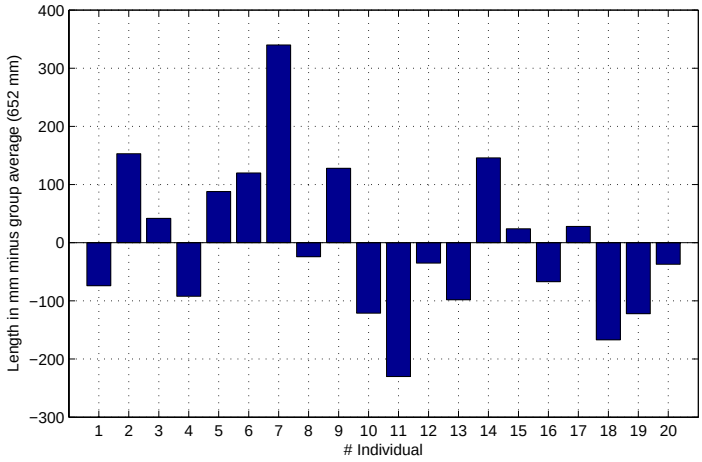


Figure 2: Length distribution of the tagged pike.

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

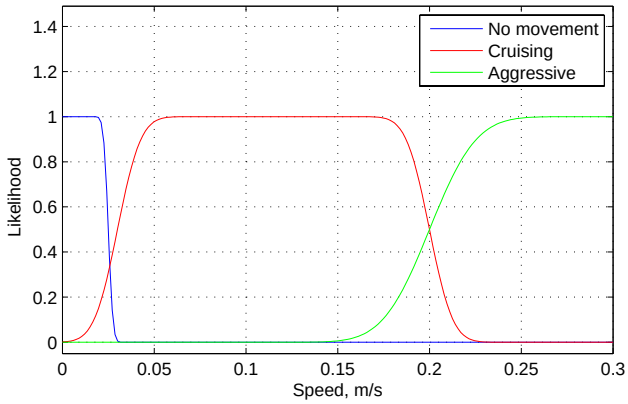


Figure 3: The likelihood of each of the three movement behaviours as a function of the observed speed.

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

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 *
```