UC Santa Cruz

Title

A continuous-time state-space model for rapid quality control of argos locations from animal-borne tags.

Permalink

https://escholarship.org/uc/item/03c5q2rn

Authors

Jonsen, Ian
Patterson, Toby
Costa, Daniel
et al.

Peer reviewed

METHODOLOGY ARTICLE

A continuous-time state-space model for rapid quality control of argos locations from animal-borne tags

Ian D. Jonsen[1]* ![ORCID], Toby A. Patterson[2], Daniel P. Costa[3], Philip D. Doherty[4], Brendan J. Godley[4], W. James Grecian[5], Christophe Guinet[6], Xavier Hoenner[2], Sarah S. Kienle[3], Patrick W. Robinson[3], Stephen C. Votier[4], Scott Whiting[7], Matthew J. Witt[4], Mark A. Hindell[8], Robert G. Harcourt[1] and Clive R. McMahon[9]

Abstract

Background: State-space models are important tools for quality control and analysis of error-prone animal movement data. The near real-time (within 24 h) capability of the Argos satellite system can aid dynamic ocean management of human activities by informing when animals enter wind farms, shipping lanes, and other intensive use zones. This capability also facilitates the use of ocean observations from animal-borne sensors in operational ocean forecasting models. Such near real-time data provision requires rapid, reliable quality control to deal with error-prone Argos locations.

Methods: We formulate a continuous-time state-space model to filter the three types of Argos location data (Least-Squares, Kalman filter, and Kalman smoother), accounting for irregular timing of observations. Our model is deliberately simple to ensure speed and reliability for automated, near real-time quality control of Argos location data. We validate the model by fitting to Argos locations collected from 61 individuals across 7 marine vertebrates and compare model-estimated locations to contemporaneous GPS locations. We then test assumptions that Argos Kalman filter/smoother error ellipses are unbiased, and that Argos Kalman smoother location accuracy cannot be improved by subsequent state-space modelling.

Results: Estimation accuracy varied among species with Root Mean Squared Errors usually <5 km and these decreased with increasing data sampling rate and precision of Argos locations. Including a model parameter to inflate Argos error ellipse sizes in the north - south direction resulted in more accurate location estimates. Finally, in some cases the model appreciably improved the accuracy of the Argos Kalman smoother locations, which should not be possible if the smoother is using all available information.

Conclusions: Our model provides quality-controlled locations from Argos Least-Squares or Kalman filter data with accuracy similar to or marginally better than Argos Kalman smoother data that are only available via fee-based reprocessing. Simplicity and ease of use make the model suitable both for automated quality control of near real-time Argos data and for manual use by researchers working with historical Argos data.

Keywords: Animal-borne sensors, Bio-telemetry, foieGras R package, Global Positioning System, Seabird, Pinniped, Sea turtle, Template Model Builder

---

*Correspondence: ian.jonsen@mq.edu.au
[1]Dept of Biological Sciences, Macquarie University, Sydney, Australia
Full list of author information is available at the end of the article

Jonsen *et al. Movement Ecology* (2020) 8:31

Page 2 of 13

Background

State-space models have emerged as important tools both for quality control and ecological analysis of error-prone animal movement data [1–5]. Analysis of these data with discrete-time models is simple in principle, breaking down animal movement into a series of discrete steps that occur on some fixed time interval (e.g., [1, 6]). Yet animal movement is a process that unfolds continuously through time, usually absent of clear breaks that could delineate discrete steps. We merely measure the movements from locations obtained over discrete, often irregular intervals in time. In this sense, a continuous-time model can more naturally handle temporally irregular observations while mimicking the true underlying continuous movement process(es) [2, 7].

In the marine realm, air-breathing animal locations are typically measured by satellite-linked electronic tags at irregular time intervals dictated by a combination of satellite availability and an animal's surface behaviour. The Argos satellite telemetry system is one of the most common platforms used to track animals at sea, with over 40,000 individuals tracked since 2007 (S. Baudel, pers. comm.). In this system, transmissions from electronic tags are received by one of several polar-orbiting satellites as they pass overhead, and the Doppler shift in transmission frequency along with other information is used to geolocate the tags [8]. The polar orbits of Argos satellites result in more dense coverage and potentially higher temporal , resolution data closer to the poles than at the equator. From inception in 1978 to 2011, CLS (Collecte Localisation Satellites) has used a Least-Squares algorithm to geolocate the tag transmissions. This approach does not quantify location uncertainty but rather provides location quality classes based on information including the number of transmissions received [8].

State-space models developed for Argos Least-Squares locations have relied on independent, ground-truth data (e.g., [9]) to quantify location uncertainty for each of the location quality classes [1, 2]. However, independently quantified uncertainties, based on a single or small number of data sets, are unlikely to be appropriate for all species in all locations. For example, modifications to assumed Least-Squares error variances can influence the accuracy of locations predicted by different state-space models [10].

In 2011, CLS replaced their Least-Squares algorithm with a state-space model, based on a multiple model Kalman filter algorithm, to estimate locations and their uncertainty [11]. This approach provides more location estimates, each with a corresponding estimated error ellipse, and with greater accuracy compared to the original Least-Squares method. These locations are provided in near real-time; here defined as within 24 h of occurrence. A fixed-interval Kalman smoother is also provided by CLS as an extra service to further improve location accuracy from the original Kalman filter-based location estimates [12]. Whereas the Kalman filter employs a one-step recursion to estimate locations based only on the current and previous observations, the Kalman smoother uses a two-pass approach, first employing the Kalman filter and then employing a backwards smooth of the data [13]. In this sense, the Kalman smoother uses information from the entire animal track to estimate locations and their uncertainty. This results in more accurate location estimates than the Kalman filter alone [12]. Such smoother-based location estimates are theoretically optimal given the available data, and it should not be possible to improve on them if uncertainty is characterised and propagated accurately (e.g., [14]). Currently, CLS does not provide Kalman smoother-based locations in near real-time, they can only be obtained with reprocessing, for an additional fee, after a tag deployment ends.

Traditional use of animal tracking data has required neither near real-time data provision nor rapid modelling tools for quality control or ecological analysis. However, real-time management of at-risk species' mortality from interactions with human activities such as offshore wind farms, fisheries and shipping increasingly relies on animal telemetry data [15–17]. Dynamic ocean management applied at high spatial and temporal resolutions can increase the efficiency and efficacy of measures to reduce mortality [18], placing an onus on rapidly available, high-resolution data. Similarly, the utility of animal-borne sensors for ocean observing [19, 20] as part of the Global Ocean Observing System has spurred coordinated animal telemetry programs, such as the Australian Integrated Marine Observing System's Animal Tracking Facility (IMOS ATF[1]) and the U.S. Integrated Ocean Observing System's Animal Telemetry Network (IOOS ATN[2]). These programs aim to provide near real-time ocean measurements via the World Meteorological Organization's Global Telecommunication System for assimilation in operational ocean and atmospheric forecast models. In all these cases, near real-time telemetry data provision requires rapid and therefore automated, reliable quality control processes, including the error-prone Argos location data that are essential for understanding animal movements and distribution, and for providing geospatial context to ocean measurements.

Here we present a continuous-time state-space model for rapid filtering of any Argos location data. This model is now used as part of the IMOS ATF's quality control/quality assurance process for animal-borne ocean observations. To facilitate fast automation, we trade off

---

[1] http://imos.org.au/facilities/animaltracking
[2] https://ioos.noaa.gov/project/atn

Jonsen *et al. Movement Ecology* (2020) 8:31

Page 3 of 13

realism - the ability to explain complex movement processes - for reliability by using a simple continuous-time random walk on velocity with a single variance parameter. We evaluate the model by: 1) comparing fits to all three Argos location types from the same individuals; 2) assessing accuracy of model-estimated locations against contemporaneous GPS locations; 3) assessing how a model assumption about Argos error ellipses influences estimation accuracy; 4) comparing the accuracy of modelled and un-modelled Kalman Smoother locations.

Methods

A continuous-time state-space model for animal telemetry data

We model animal movement as a continuous-time random walk on velocity $\mathbf{v}_t$ in two coordinate axes:

$$\mathbf{v}_t = \mathbf{v}_{t-\Delta} + \Sigma_\Delta \qquad (1)$$

where Δ is the time increment and Σ_Δ is a zero-mean, bi-variate Gaussian random variable with variance $2D\Delta$. The parameter D is a 1-d diffusion coefficient accounting for variability in velocity, which increases with the time interval Δ. Noting that locations $\mathbf{x}$ are the summed velocities, given a starting location, the following equation describes a simple process model subject to variable time increments:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{v}_i \Delta_i \qquad (2)$$

where the subscript i indexes time t_i, $\mathbf{x}_i$ is the true location of the animal at time t_i and $\mathbf{v}_i \Delta_i$ is the displacement (velocity x elapsed time) between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$. To simplify the model, we assume that the velocity random walk variances $2D\Delta_i$ are equal on the two axes but they could also be assumed to vary independently [2]. Correlation in movements arises from allowing the locations to be the sum of the velocities.

We couple this process model to a generally applicable measurement model that describes how the error-prone and possibly irregularly-timed observed locations $\mathbf{y}_i$ map onto the corresponding true location states $\mathbf{x}_i$:

$$\mathbf{y}_i = \mathbf{x}_i + \epsilon_i; \qquad \epsilon_i \sim \mathrm{N}(0, \Omega_i) \qquad (3)$$

where $\mathbf{y}_i$ the location observed at time t_i corresponding to $\mathbf{x}_i$, and Ω_i is the measurement error variance-covariance matrix that can be structured to suit different types of location data. Below, we focus on modifications to accommodate different Argos location types, but other location data (e.g., processed light-level geolocations) could also be considered in this framework.

Argos Least-Squares data

Locations measured using CLS' older Least-Squares (LS) approach [8] are associated with location quality class designations: 3, 2, 1 0, A, B, and Z. These classes are the only

contemporaneous information about location quality and provide only a relative index of measurement uncertainty [1]. We use the class information, along with independent estimates of their associated standard errors from Argos transmitters deployed on seals held captive at a known location [9], to construct the following variance-covariance matrix:

$$\Omega_i = \begin{bmatrix} \tau_x^2 K_{x,i}^2 & \rho \tau_x K_{x,i} \tau_y K_{y,i} \\ \rho \tau_x K_{x,i} \tau_y K_{y,i} & \tau_y^2 K_{y,i}^2 \end{bmatrix} \qquad (4)$$

where τ_x^2 and τ_y^2 are the overall measurement error variances on the two coordinate axes, $K_{x,i}$ and $K_{y,i}$ are error weighting factors that scale the $\boldsymbol{\tau}$'s appropriately for the Argos location quality class associated with the ith observation, and ρ is the correlation between $\tau_x K_{x,i}$ and $\tau_y K y, i$. The τ's are estimated during model fitting and the error weighting factors are the standard error ratios between the best quality class, 3, and each other class (2, 1, 0, A, B, Z).

Argos kalman filter and kalman smoother data

Locations measured using CLS' Kalman filter (KF) or Kalman smoother (KS) algorithms have their estimated uncertainties provided to users as error ellipses [11]. Ellipses are defined by three variables: semi-major axis, semi-minor axis and semi-major axis orientation from north. Building on McClintock et al. [14], the error variance-covariance matrix is:

$$\Omega_i = \begin{bmatrix} \tau_{x,i}^2 & \tau_{xy,i} \\ \tau_{xy,i} & \tau_{y,i}^2 \end{bmatrix} \qquad (5)$$

with the elements being derived from the Argos error ellipse components:

$$\tau_{x,i}^2 = \left(\frac{M_i}{\sqrt{2}}\right)^2 \sin^2 c_i + \left(\frac{m_i \psi}{\sqrt{2}}\right)^2 \cos^2 c_i \qquad (6)$$

$$\tau_{y,i}^2 = \left(\frac{M_i}{\sqrt{2}}\right)^2 \cos^2 c_i + \left(\frac{m_i \psi}{\sqrt{2}}\right)^2 \sin^2 c_i \qquad (7)$$

and

$$\tau_{xy,i} = \left(\frac{M_i^2 - m_i^2 \psi^2}{2}\right) \cos c_i \sin c_i \qquad (8)$$

where M_i is the ellipse semi-major axis length of the i-th observation, m_i is the semi-minor axis length and c_i is the semi-major axis orientation [11, 14].

McClintock et al. [14] used a bivariate t-distribution, with variance-covariance defined by the Argos error ellipses, in their measurement model to account for occasional outlier observations (i.e., where error ellipses underestimate the true measurement uncertainty). Here we chose to identify and remove outlier locations using a travel rate filter [21] prior to fitting the state-space model, as per [2, 22]. Additionally, we included the parameter ψ to account for possible consistent under estimation of

the Kalman filter (& smoother)-derived location uncertainty (Fig. 1). ψ re-scales all ellipse semi-minor axes m_i, where estimated values > 1 inflate the uncertainty region around measured locations by lengthening ellipse semi-minor axes.

In all cases, we project the $\mathbf{y}_i$'s from geographic coordinates (lon, lat) onto a Cartesian plane prior to modelling, using the WGS84 World Mercator projection (EPSG 3395). To facilitate optimization, all planar coordinates and their uncertainty estimates, where available, are converted from m to km.

Estimation

We used the R package TMB (Template Model Builder, [23]) to fit the state-space model, using maximum likelihood to estimate model parameters and the Laplace approximation to rapidly estimate the random effects - the unobserved location and velocity states, $\mathbf{x}$ and $\mathbf{v}$

[5, 24]. Using this estimation approach, uncertainty in $\mathbf{x}$ and $\mathbf{v}$ estimates are obtained using a generalised delta method (see [23] for details).

The model presented here and associated general data preparation code are available in the foieGras R package [25], available on the CRAN server (https://CRAN.R-project.org/package=foieGras). The latest version can also be downloaded from the lead author's GitHub site (https://github.com/ianjonsen/foieGras).

Data and pre-processing

We model all three types of Argos satellite location data: LS, KF, and KS. The data are comprised of four pinnipeds, one seabird and two sea turtle species (Table 1); with deployment locations ranging between polar, temperate, and tropical marine regions (Additional file 1: Fig S1). The number of individual data sets by species and data type range from 6 to 13 with all having locations measured by
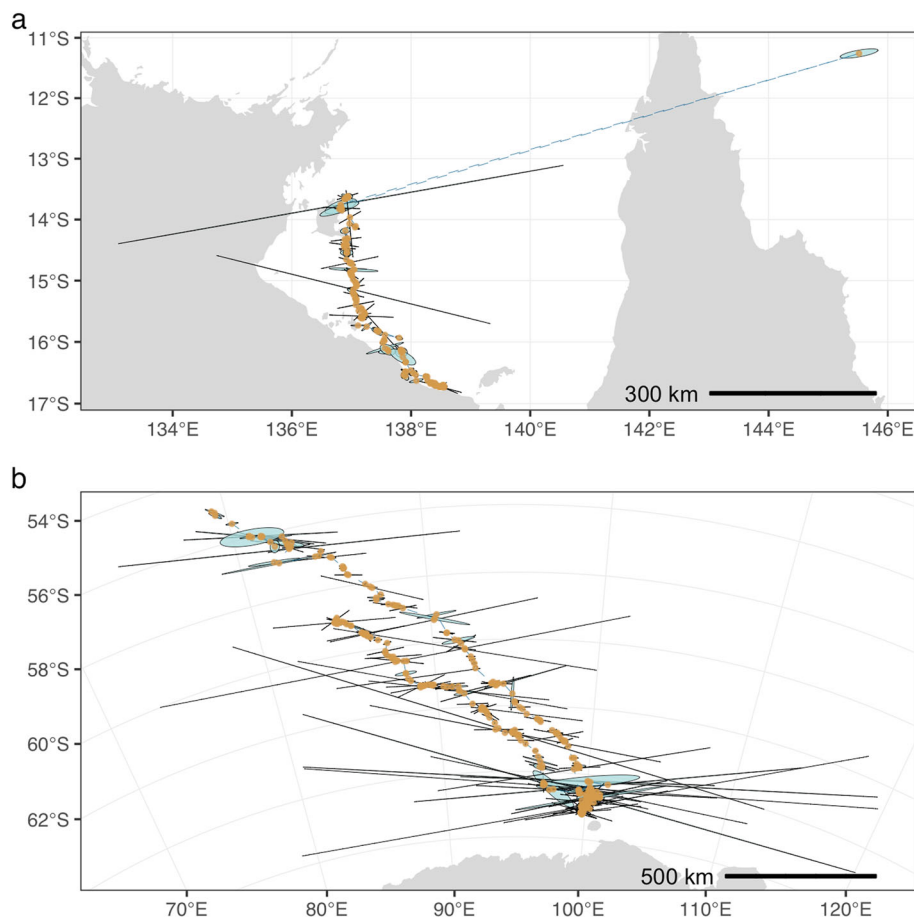

Fig. 1 Unprocessed Argos Kalman filter (KF) locations (gold points) and error ellipses (pale blue with black borders) for (**a**) hawksbill turtle (HBTU) and (**b**) southern elephant seal (SESE). Temporally sequential but spatially distant locations are connected by dashed blue lines. Scale bars provide an indication of the magnitude of errors. The extreme outlier location at top right (**a**) is approximately 650 km distant from the preceding and following locations, and has a vastly underestimated error ellipse

Jonsen *et al. Movement Ecology*   (2020) 8:31

Page 5 of 13

Table 1 Number of individual data sets by species and data type. Argos data types are: Least-Squares (LS); Kalman filter (KF); Kalman smoother (KS). Mean track durations were calculated from the data after removing periods of prolonged data gaps. Tag programming details were not available for all deployments, so Argos and GPS sampling rates were calculated from the unfiltered data

Species	Common name	Code	Deployment year(s)	Mean track duration (d)	Data type				GPS sample rate (min)	Fastloc GPS
					LS	KF	KS	GPS		
Zalophus californianus	California sea lion	CASL	2007	79	8	.	.	8	58	Y
Arctocephalus pusillus	Cape fur seal	CPFS	2007	25	6	.	.	6	49	Y
Dermochelys coriacea	leatherback turtle	LBTU	2008/12	92	13	.	.	13	195	Y
Eretmochelys imbricata	hawksbill turtle	HBTU	2009/10	24	6	6	6	6	124	Y
Hydrurga leptonyx	leopard seal	LESE	2018	171	8	8	8	8	47	Y
Mirounga leonina	southern elephant seal	SESE	2009/11/12/14	53	11	11	11	11	28	Y
Morus bassanus	northern gannet	NOGA	2010	23	9	9	9	9	60	N

GPS and at least one Argos type (Table 1). All data collected after 2008 were reprocessed by CLS to obtain the three Argos data types (4 species; Table 1).

We used an automated pre-filtering step to identify outlier observations to be ignored by the state-space model. This pre-filtering used the `argosfilter` R package [21] to identify locations implying travel rates >3 ms^{-1} for all pinnipeds and sea turtles and travel rates >17 ms^{-1} for northern gannets. These speed thresholds represent conservative upper limits of travel for these species and are intended to identify only the extreme outlier observations. This resulted in < 30% of Least-Squares, < 15% of Kalman filter, and < 10% of Kalman smoother data being removed. The proportion of data removed by pre-filtering is considerably less than those associated with optimal speed thresholds for other species (e.g., [22]).

Empirical validation

We examined the accuracy of model-predicted locations, assuming GPS data represent truth. Although GPS data have higher spatial accuracy and precision, and typically have higher sampling rates than Argos data, they are nonetheless discrete measurements of a continuous-time process. As a consequence, they are also likely to misrepresent animals' true movement paths but to a far smaller extent (10's of m; [26]) than Argos data.

For all validations presented, we compared GPS locations to model-fitted locations (hereafter model-estimated locations), which are location states estimated at the times of the Argos-measured locations. By focusing on model-estimated locations and not predicted locations that occur at regular time intervals, we reduce the degree to which model accuracy is confounded with data sampling rates that are known to vary across species and Argos data types (see Discussion).

We compared model-estimated locations from fits to all three Argos data types, where available, with GPS data. In all cases, the times of GPS observations do not match

the times of Argos observations or the corresponding model-estimated locations. To account for this mismatch, we initially considered three approaches for comparing between GPS and modelled locations. First, using a linear interpolation of GPS locations to model-estimated location times [27]. Second, using the temporally closest GPS observation if any occurred within ±10 min. Third, using the model to predict locations at the GPS observation times. In several cases, it was not feasible to predict model locations for each GPS observation time as the typically higher frequency of GPS observations resulted either in implausible artefacts in the model fits to the Argos data or in convergence failures of the optimiser used to fit the model. For these reasons, we chose not to consider this approach further.

Fitting the state-space model with a fixed 2-h prediction interval resulted in optimiser convergence for all individual tracks. For each individual track, we summarized the deviations between model-estimated locations and either the linearly interpolated GPS locations or the temporally matched GPS locations by taking the root mean of the squared distances (RMSD in km) between all pairs of locations and comparing distributions of individual RMSD values among species. We report results of comparisons with the linearly interpolated GPS locations here and comparisons with the temporally matched GPS locations in Supplementary Information. We discuss implications of using each of these approaches.

Total sequential processing time for validation using all 129 Argos data sets (Table 1) was 13.43 min, an average of 6.25 s per data set. This included both the pre-filter algorithm and state-space model estimation, running on a 2018 MacBook Pro 15" laptop with 2.9 GHz i9 processor, 32 GB RAM, with R 3.6.2.

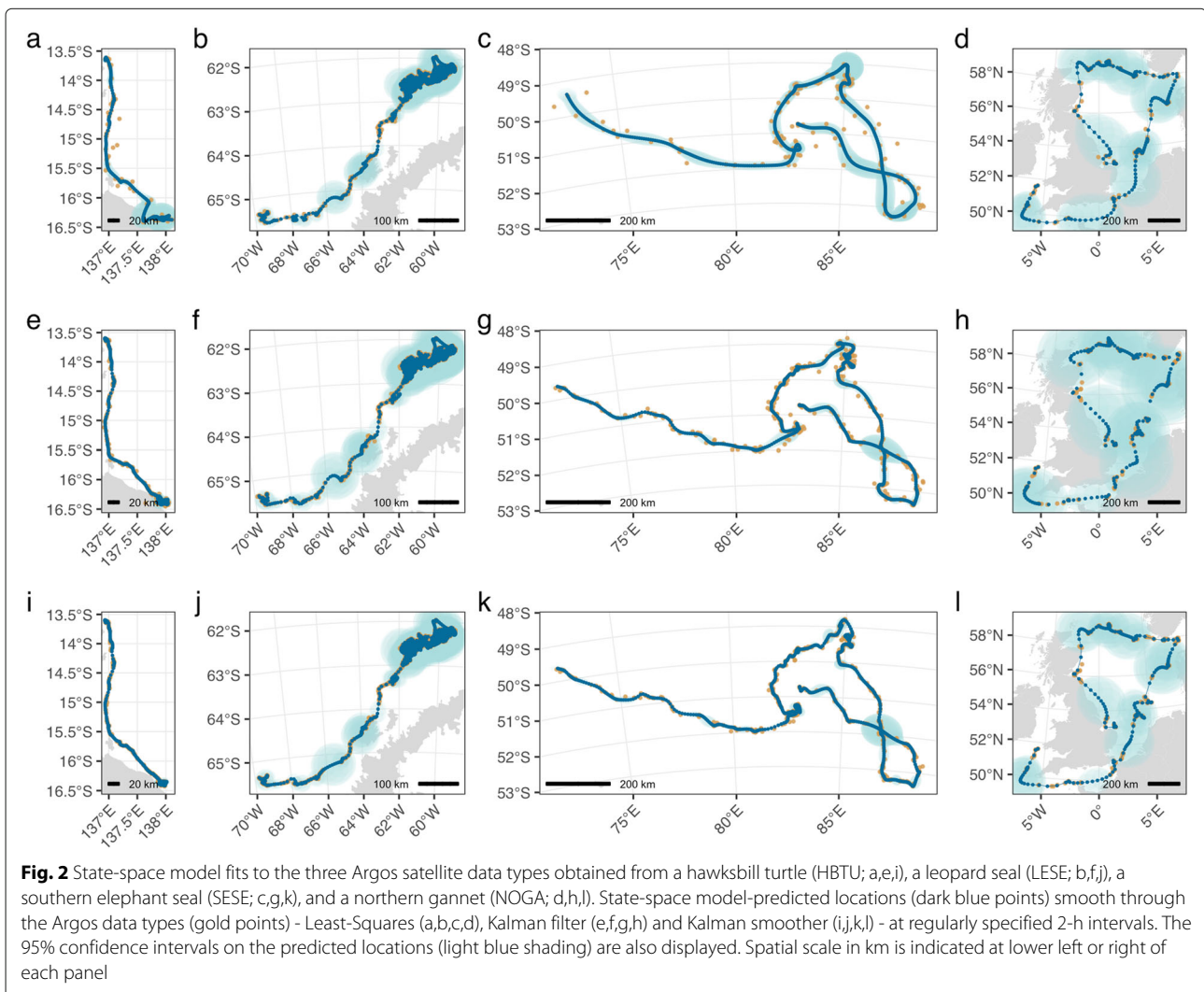***Potential under-representation of Argos KF/KS location uncertainty***

Our default model accounts for a perceived under-estimation of the size of CLS' Kalman filter and Kalman

Jonsen *et al. Movement Ecology*   (2020) 8:31

Page 6 of 13

smoother error ellipses (Figs. 1 and Additional file 1: S2-S5) by including the parameter ψ (Eqns. 6, 8). Although uncertainty is expected to be lower in the general North - South plane due to the polar orbits of the Argos satellites [11], the frequent compression of error ellipses in this plane (semi-minor axis; e.g., Fig. 1b) seems extreme. Values of $\psi > 1$ inflate the semi-minor axis, increasing the uncertainty region around Argos KF/KS observations and could allow the model to more appropriately smooth the data. It is unclear how much the parameter actually improves the accuracy of estimated tracks versus yielding a less accurate over-smoothing of the data. To assess this, we evaluated the influence of the ψ parameter on the accuracy of model-estimated locations by comparing RMSD values from models with and without the ψ parameter. To simplify the results, we pooled RMSD values across species and assessed the $\log_e$ difference in RMSD (denoted as $\log \Delta$ RMSD), which approximates % difference on the linear scale [28].

Argos KS location accuracy

The CLS Kalman smoother locations have greater spatial accuracy and precision than Least-Squares or Kalman filter data [12]. In principle, it should not be possible to improve the accuracy of KS-based locations with subsequent modelling because they are theoretically optimal estimates, using all available data. It does seem reasonable, however, to question whether this is actually the case. We evaluated this by comparing $\log \Delta$ RMSD derived from GPS and KS locations to those derived from GPS and estimates from the state-space model fit to the KS locations. In both cases, we apply the same pre-filtering to identify and remove outlier locations, though these outliers should not be present in KS-based locations.

Fig. 2 State-space model fits to the three Argos satellite data types obtained from a hawksbill turtle (HBTU; a,e,i), a leopard seal (LESE; b,f,j), a southern elephant seal (SESE; c,g,k), and a northern gannet (NOGA; d,h,l). State-space model-predicted locations (dark blue points) smooth through the Argos data types (gold points) - Least-Squares (a,b,c,d), Kalman filter (e,f,g,h) and Kalman smoother (i,j,k,l) - at regularly specified 2-h intervals. The 95% confidence intervals on the predicted locations (light blue shading) are also displayed. Spatial scale in km is indicated at lower left or right of each panel

Results

State-space model fits to the 3 Argos data types

We fit the state-space model to the four species with all three Argos data (Table 1), and present fits with a 2-h prediction time interval. Model fits to hawksbill turtle and southern elephant seal data show a consistent increase in spatial resolution and decrease in estimation uncertainty of the predicted tracks across the three Argos data types (top to bottom; (Fig. 2 a,e,i and b,f,j, respectively). This effect is due to an increase in the number of observations from least-squares to Kalman filter data, and to a shrinking of the error ellipses (measurement uncertainty), by nearly half, from Kalman filter to Kalman smoother data (Table 2). Model fits to leopard seal and northern gannet data do not show any clear differences in resolution or estimation uncertainty across the Argos data types (Fig. 2 c,g,k and d,h,l, respectively). This appears due to smaller differences in the number of observations for Least-Squares versus Kalman filter data, arising from lower proportions of class A and B locations, relative to hawksbill turtles and southern elephant seals (Table 2). The lower proportions of class A and B locations for leopard seals and northern gannets are likely due to the large amount of time they spend at or above the ocean surface. Additionally, northern gannets had, on average, far larger error ellipses than the other species (Table 2). The uncertainty of their state-space model-predicted locations was consequently larger, regardless of Argos data type (light blue 95% confidence ellipses in Fig. 2 d,h,l).

Validation with GPS data

Median distances between state-space model-estimated and interpolated GPS locations were within 8 km for all species and data types, with most species and data types having 95% of estimated locations within 12 km of GPS locations (Table 3). Northern gannets were an exception, with 95-th percentiles extending > 40 km for all Argos data types (Table 3). Importantly, the median accuracy of state-space model-estimated locations, regardless of Argos data type, were all smaller or comparable to those of pre-filtered but un-modelled KS locations (Table 3). Across species, the weighted average ($\pm$ se) improvement of state-space model-estimated location accuracy relative to un-modelled KS location accuracy was: LS = 0.21 $\pm$0.60 km; KF = 0.14 $\pm$0.07 km; KS = 0.34 $\pm$0.05 km.

Six of the 7 species' estimated tracks had median RMSD values under 5 km with all values under 10 km, regardless of Argos data type (Fig. 3). Northern gannet tracks had considerably higher and more variable RMSD's (between 13 and 31 km), across all Argos data types (Fig. 3). This is consistent with their considerably larger state-space model-predicted location uncertainty (Fig. 2). Both hawksbill turtle and southern elephant seal tracks had declining RMSD values as Argos data frequency and precision increased (Fig. 3), and this was consistent with the increasing resolution and precision of their state-space model-predicted tracks (Fig. 2). Conversely, leopard seal and northern gannet tracks showed no such pattern, which was consistent with the general lack of increasing resolution of both the observed and predicted tracks

Table 2 Argos track summary statistics by species and data type. prop'n A,B is the proportion of all locations that are in quality class A and B. Error ellipse shape is the ratio of semi-minor to semi-major axis length, with values closer to 1 indicating a more circular shape. Shape and area statistics were calculated from values pooled among individuals within species and Argos data type

Species code	Argos type	Total n locations	prop'n A,B	Error ellipse shape	Error ellipse area (km)	
					median	95th %-ile
HBTU	LS	603	0.71	.	.	.
HBTU	KF	1 038	0.84	0.24	17.31	120.29
HBTU	KS	1 038	0.84	0.27	8.44	60.31
LESE	LS	32 780	0.47	.	.	.
LESE	KF	41 367	0.57	0.13	1.73	69.23
LESE	KS	41 367	0.57	0.14	1.03	31.50
SESE	LS	3 152	0.82	.	.	.
SESE	KF	5 016	0.91	0.13	34.40	780.11
SESE	KS	5 016	0.91	0.14	16.26	338.18
NOGA	LS	1 568	0.36	.	.	.
NOGA	KF	2 066	0.52	0.18	25.43	5 573.62
NOGA	KS	2 066	0.52	0.19	16.73	2 104.87

Jonsen *et al. Movement Ecology*   (2020) 8:31

Page 8 of 13

Table 3 Accuracy in km of state-space model-estimated locations and pre-filtered KS locations (pf_KS), by species and Argos data type. The pf_KS locations had the pre-filter algorithm applied but not the state-space model. Median and 95th percentile statistics were calculated from distances to GPS locations, pooled among individuals within species and Argos data type

Species	LS		KF		KS		pf_KS	
	median	95th %-ile	median	95th %-ile	median	95th %-ile	median	95th %-ile
CASL	1.70	10.15	.	.	.	.	.	.
CPFS	1.87	8.03	.	.	.	.	.	.
LBTU	2.77	10.06	.	.	.	.	.	.
HBTU	1.60	8.11	1.69	7.08	1.47	6.05	1.62	5.91
LESE	1.88	11.11	2.05	11.61	1.89	10.90	2.25	12.62
SESE	4.40	15.03	3.24	12.06	2.97	9.50	3.29	11.02
NOGA	6.04	43.95	7.50	52.62	7.37	46.47	7.85	53.67

(Fig. 2). Results were similar, although with overall lower RMSD values, when comparing state-space model estimated locations to the temporally closest GPS location within ±10 min (Additional file 1: Fig. S6).

Effect of ψ parameter

Inclusion of the ψ parameter resulted in lower RMSD values, on average, implying that Argos error ellipses under-represent the true location uncertainty in the general north - south direction (Fig. 4). This result was less pronounced with fits to Argos Kalman smoother locations, with 81% of individuals having a log Δ RMSD <0 versus 90% of individuals for Argos Kalman filter locations (KF Δ RMSD: median = -0.57 km, range = -3.78,0.45; KS Δ RMSD: median = -0.27 km, range = -3.34, 0.85). Of the four species, predicted locations for hawksbill turtle

tracks were least likely to benefit from re-scaled error ellipses, with most individuals having log Δ RMSD values close to or >0 (Fig. 4). It is unclear whether this is due to: 1) their relatively low absolute RMSD values (Fig. 3); 2) their slightly more circular error ellipses (Table 2), where the ψ re-scaling effect would be less pronounced; or, 3) a combination of the two.

Argos KS accuracy

Argos Kalman smoother locations were less accurate by an average of 0.34 km without subsequent state-space model filtering (Table 3; compare KS and pf_KS values), although comparisons of log Δ RMSD were variable both within and among species (Fig. 5). The mean log Δ RMSD across species implied a average 6% increase in accuracy with subsequent state-space model filtering of Argos KS
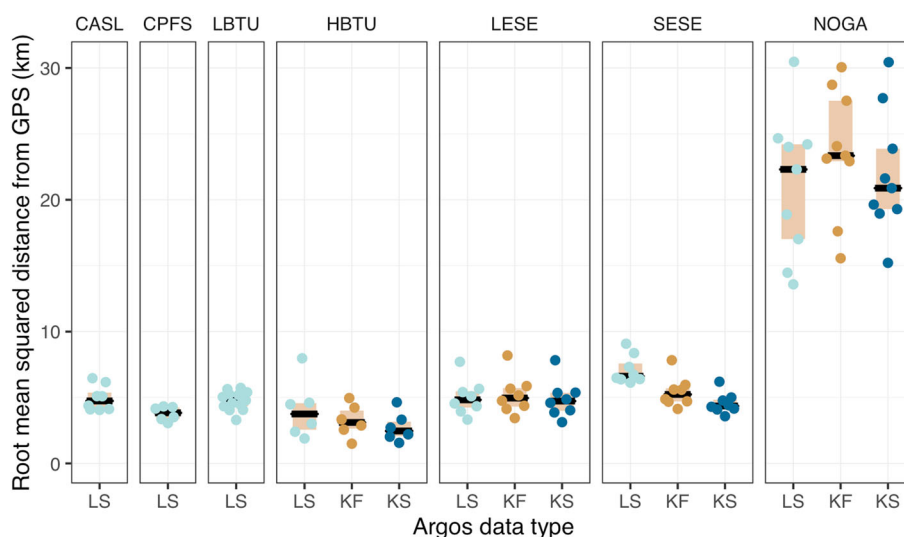

Fig. 3 Accuracy of state-space model-estimated locations across the three Argos location data types and by species. Accuracy is assessed using root mean squared distance (RMSD in km) from interpolated GPS locations. Individual RMSD values are filled circles, coloured by Argos data type, beige boxes are inner quartile ranges, and black bars are medians. Species codes: CASL - California sea lion; CPFS - Cape fur seal; LBTU - leatherback sea turtle; HBTU - hawksbill sea turtle; LESE - leopard seal; SESE - southern elephant seal; NOGA - northern gannet
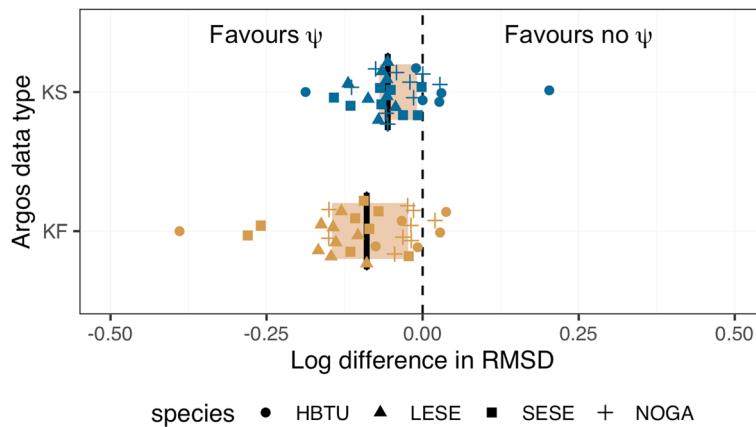
Fig. 4 Accuracy of estimated locations from models fit with and without a ψ parameter to re-scale Argos error ellipse sizes. The log difference in root mean squared distance (log Δ RMSD) from interpolated GPS locations was calculated for the two models fit to each individual animal. Negative values indicate the model with a ψ parameter is more accurate, whereas positive values indicated the model without a ψ parameter is more accurate. Results are pooled across species with individual log differences in RMSD displayed as species-specific shapes (gold, blue). The inner quartile ranges are beige boxes and medians are black bars

locations. However, results were equivocal for southern elephant seals and hawksbill turtle tracks were typically more accurate without any subsequent state-space filtering (Fig. 5).

Discussion

We presented a continuous-time model for animal movement, fit in a state-space framework that allows flexible handling of Argos satellite telemetry data. The model was initially intended for automated quality control of large Argos animal tracking data sets, but is broadly applicable for any Argos location data. Using Argos - GPS double tagged animals, we assessed the accuracy of

model-estimated locations, comparing across three types of Argos data where possible. Median accuracy was within 4 km for most species and data types, with state-space model-estimated locations being slightly more accurate (by 0.1 - 0.3 km on average) than the best quality CLS Kalman smoother locations. Median root mean squared deviations were typically at or under 5 km for 6 of the 7 species studied. In most cases, RMSD values were lowest when fitting to Argos Kalman smoother data and highest when fitting to Argos Least-Squares or Kalman filter data, although the within-species differences in RMSD between data types were typically small. Although the model was evaluated over a limited number of individuals
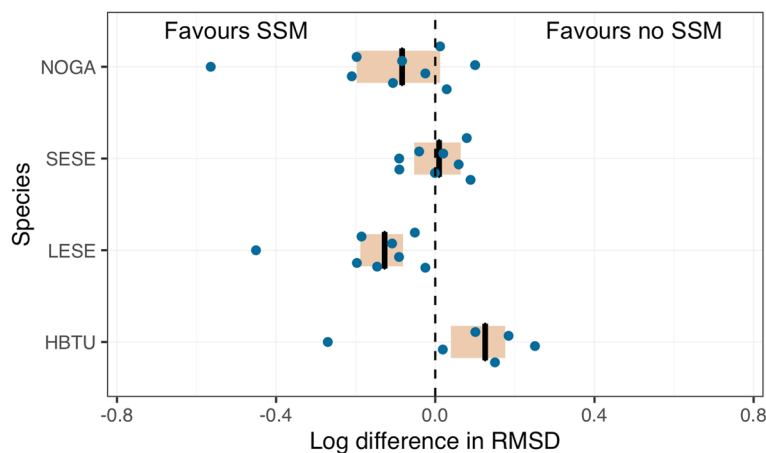
Fig. 5 Accuracy of Argos Kalman smoother locations with and without subsequent state-space model filtering. The log difference in root mean squared distance (log Δ RMSD) from interpolated GPS locations was calculated for each individual animal. In both cases, the Kalman smoother locations were subjected to the same travel-rate filtering to remove highly implausible observations. Negative values indicate the state-space model-estimated locations are more accurate. The inner quartile ranges are beige boxes and medians are black bars

and species, it is apparent that the accuracy and spatio-temporal resolution of inferred locations is situational.

Highlighting this situational aspect are the northern gannet results (Table 3; Figs. 2 & 3), which are clearly distinct from the other species. Accuracy of model-estimated locations was approximately 4-5 times worse than for other species, although absolute magnitude is subject to the approach used for matching model-estimated and GPS locations (compare Figures. 3 & Additional file 1: Fig S5). Unlike other species where median distances between model-estimated and GPS locations either declined consistently or were similar when comparing LS to KF and KF to KS data types, gannets had the lowest median distances for fits to LS data and had far broader distributions of distance across the 3 data types. We suspect this pattern may arise from the considerably faster mean travel rates of northern gannets (12 km h^{-1}, with cruising speeds up to 45 km h^{-1}) compared to the other species (approximately 0.7 - 3 km h^{-1}). Similarly, Lopez et al. [12] reported lower overall coverage probabilities of error ellipses estimated by their Kalman filter and Kalman smoother algorithms for two avian species analyzed in comparison to other platforms (terrestrial and marine mammals, sea turtles, ships and drifters). Combined, this implies that Argos error ellipses may be more strongly underestimated for species/platforms that travel faster and/or at higher altitude.

McClintock et al. [14] used a bivariate t-distribution, parameterised by the Argos error ellipse information, to model location measurement error. Their estimates of the t degrees of freedom parameter implied that the Argos error ellipses do not fully explain location measurement error. To avoid computational challenges associated with t-distribution parameter estimation, we used a two-step approach for dealing with location measurement error in Argos Kalman filter and Kalman smoother data. First, we identified and removed potentially large outliers using a travel-rate filter [21] prior to fitting the state-space model, as per [2, 22]. Although underestimation of location error was acknowledged by Lopez et al. [11, 12] and has been reported by others [14, 29], it is unclear why occasional, apparent hugely underestimated error ellipses are present in the Kalman filter and Kalman smoother data. Second, we accounted for potential Argos error ellipse underestimation by including the ψ parameter to inflate the semi-minor axis. We adopted this approach given the observation that Argos error ellipses often have semi-minor axes vastly smaller than corresponding semi-major axes, resulting in "squashed" error ellipses (Additional file 1: Figures S2-S4). We found that in most cases the ψ parameter contributed to more accurate location estimates, implying that the error ellipses commonly underestimate the true uncertainty in Argos-measured locations. This result is evident but less pronounced when fitting to

Kalman smoother versus Kalman filter data. Location estimates were more accurate for at least some individuals of all species, however, hawksbill turtles and northern gannets appeared least likely to benefit from the ψ re-scaling effect (see Fig. 4). Both of these species had somewhat more circular error ellipses, in comparison to the leopard and southern elephant seals, and thus any possible contribution of ψ would be reduced. Ultimately, we are unsure why Argos error ellipses appear to be so commonly biased low in the semi-minor axis direction (generally north - south).

Where possible, both Kalman filter and Kalman smoother data types were included in this study. We found, in most cases, that the model-estimated locations were most accurate when using the Kalman smoother data, but on average by less than 200 m compared with fits to Kalman filter data. Although the Kalman smoother data should represent optimal estimates of location because information along the entire movement track is used to update and smooth each location estimate, we show that fitting the state-space model to these estimates can further improve location accuracy in some cases (by an average reduction in error of approximately 6%). The Kalman smoother data are not provided in the default, near real-time service from CLS, rather they are only available with post-processing by CLS at an additional cost. There are two points to be made about this. First, the smoothing algorithm is a standard approach that can be implemented rapidly, with computing requirements no greater than the Kalman filter. It could be applied in near real-time. Second, a near real-time Kalman smoother would result in the best available location estimates changing as new data became available. This incremental improvement, due to information gain propagating backwards in time, would reduce as locations become less recent. This should be of little consequence to most wildlife users who typically do not use their data in near real-time, and users who do require near real-time data may see greater benefit in more accurate locations even if they are subject to change in retrospect.

Our state-space model produced location estimates with a median accuracy comparable to or greater than CLS' Kalman smoother locations, regardless of input Argos data type. This implies that users can obtain similar or better accuracy than CLS' Kalman Smoother locations by applying the state-space model to their Least-Squares or Kalman filter data. Therefore the method we describe is a viable alternative to the CLS' fee-based reprocessing service. The Laplace approximation approach employed in Template Model Builder models states (velocity and location) as unknown random effects, providing a most likely estimate of the current state from the posterior of it's location given all available data, both forward and backward in time. This is precisely what a Kalman smoother does. That

our model can improve on the CLS Kalman smoother's location estimates may imply that uncertainty is somehow not well-propagated from the raw Doppler shift data available to CLS through to the location estimates available to users. If this is indeed the case, it is unclear why this is so. The issue may be due to necessary trade-offs between accuracy and precision versus providing a near real-time location service for a multitude of moving platforms, of which wildlife are a small component.

Spatio-temporal resolution and spatial accuracy
It is important to note that when comparing GPS locations with those from models fitted to Argos-measured locations, accuracy is interlinked with the temporal resolution (sampling rate) of Argos relative to GPS locations. As GPS resolution is typically greater than Argos, comparisons to determine spatial accuracy of estimated locations are confounded by this difference. No model fit to Argos-measured locations alone can resolve all the nuances of a movement path that are present in higher resolution GPS data. This discrepancy will be reflected in measures of spatial accuracy, unless GPS data are suitably sub-sampled or interpolated.

We interpolated GPS locations to the times of the Argos-measured locations to which the state-space model was fitted. Our reasoning was that interpolation of the generally higher resolution GPS data should be less corrupted by spatial error than a similar interpolation of the lower resolution and irregularly occurring model-estimated locations. Sub-sampling GPS locations by matching them with the temporally closest model-estimated location, commonly used elsewhere [12, 30, 31], resulted in lower RMSD or greater (apparent) accuracy than comparison with the linearly interpolated GPS locations. These lower RMSD values, however, were based on fewer (n <10) temporally matched pairs of model-estimated and GPS locations for some species/individuals (Additional file 1: Fig S5); using a 20-min window. Although sample sizes could be increased by choosing a wider time window, the potential for biased comparisons would increase differently across species due to their different spatio-temporal scales of movement.

Fits to the three Argos location types from the same individuals showed that movement pathways can be predicted with increasing spatial resolution, i.e., resolve greater spatial detail despite the same prediction time interval (2 h), and precision as the number of Argos-measured locations increased (transition from Least-Squares to Kalman filter data) and as their uncertainty decreased (transition from Kalman filter to Kalman smoother data). One of the main advantages of Argos' Kalman filter over the older Least-Squares method is a gain in the number of location estimates, mostly by resolving locations from the single transmissions between tag and satellite that Least-Squares can not [11]. This increase in resolution and precision is case-dependent, however, as species with lower overall proportions of class A and B locations do not gain as many new locations when transitioning from Least-Squares to Kalman filter data. This case-dependency is likely tied to typical surface time intervals of diving species, and, for those species spending the majority of time in air, on the magnitude of their travel rates.

Caveat
Fitting a state-space model to animal location data is not a panacea. Many ecological analyses of animal tracking data consider remotely sensed or other environmental data at spatial resolutions (2 - 10 km; e.g., [32]) approaching the state-space model accuracy limits found here. This highlights the need for researchers to consider the appropriate resolution of their environmental data given their specific questions and the limitations of their location estimates. While spatio-temporal mismatches between location estimates and environmental data can sometimes be dealt with by specifying coarser or finer prediction time intervals, such an approach has implications both for spatial and temporal autocorrelation affecting inference from subsequent analyses and for uncertainty in the location estimates themselves. Researchers should consider carrying location uncertainty estimates provided by state-space models through to subsequent ecological analyses. For example, by repeatedly sampling from the location uncertainty, conducting the analysis, and pooling results (sensu [33]). This can be done either completely through the whole analysis or partially via subsequent sensitivity analysis. Failure to examine or directly account for potential influences of spatio-temporal autocorrelation and estimation uncertainty of locations in subsequent analyses risks biased inferences.

Conclusions
The state-space model developed and validated here can be used to obtain quality-controlled animal locations from Argos Least-Squares or Kalman filter data in near real-time, with median accuracy comparable to or marginally better than CLS' reprocessed Kalman smoother data. Our model also accounts for apparent north-south bias in Kalman filter- and Kalman smoother-derived error ellipses.

The model's near real-time capability provides the best estimates of location, given the available data, that can be continually updated as new data arrive via the Argos system. This rapid, continual quality control of animal tracking data is necessary as near real-time monitoring and forecasting of ocean states increasingly incorporates oceanographic data from animal-borne sensors, and as

Jonsen *et al. Movement Ecology*   (2020) 8:31

Page 12 of 13

the need for dynamic ocean management grows in our increasingly exploited and rapidly changing oceans.

Although the model was developed for fast, automated quality control processes, its simplicity and ease of use also make it suitable for manual use by researchers wishing to conduct quality control of historical or otherwise less immediate Argos data.

Supplementary information

Additional file 1: Supplementary figures.

Abbreviations

CLS: Collecte Localisation Satellites; EPSG: European Petroleum Survey Group; GPS: Global Positioning System; LS: Least-Squares; KF: Kalman filter; KS: Kalman smoother; IMOS ATF: Integrated Marine Observing System Animal Tracking Facility; IOOS ATN: Integrated Ocean Observing System Animal Telemetry Network; RMSD: Root Mean Squared Deviation; WGS84: World Geodetic System 1984

methodology is publicly available on the lead author's GitHub site: https://github.com/ianjonsen/foieGras and published as an R package on the CRAN server: https://cran.r-project.org/package=foieGras

Author details
[1]Dept of Biological Sciences, Macquarie University, Sydney, Australia. [2]CSIRO Oceans and Atmosphere, Hobart, Australia. [3]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, USA. [4]Environment and Sustainability Institute, University of Exeter, Penryn, UK. [6]Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, UK. [7]CNRS, Chize, France. [8]Department of Biodiversity, Conservation and Attractions, Government of Western Australia, Kensington, Australia. [9]Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia. [10]Sydney Institute of Marine Science, Mosman, Australia.

References

1. Jonsen I, Flemming J, Myers R. Robust state–space modeling of animal movement data. Ecology. 2005;86(11):2874–80.
2. Johnson DS, London JM, Lea M, Durban JW. Continuous-time correlated random walk model for animal telemetry data. Ecology. 2008;89(5):1208–15.
3. Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J. State-space models of individual animalmovement. Trends Ecol Evol. 2008;23:87–94.
4. Albertsen CM, Whoriskey K, Yurkowski D, Nielsen A, Mills Flemming J. Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder. Ecology. 2015;96:2598–604.
5. Auger-Méthé M, Albertsen CM, Jonsen ID, Derocher AE, Lidgard DC, Studholme KR, et al. Spatiotemporal modelling of marine movement data using Template Model Builder (TMB). Mar Ecol Prog Ser. 2017;565:237–49.
6. McClintock BT, King R, Thomas L, Matthiopoulos J, McConnel BJ, Morales JM. A general discrete–time modeling framework for animal movement using multistate random walks. Ecol Monogr. 2012;82:335–349.
7. McClintock BT, Johnson DS, Hooten MB, Ver Hoof JM, Morales JM. When to be discrete: the importance of time formulation in understanding animal movement. Mov Ecol. 2014;2:21.
8. Service Argos. Argos User's Manual. CLS. 2016. http://www.argos-system.org/manual. Accessed 07 July 2020.
9. Vincent C, McConnell BJ, Fedak MA, Ridoux V. Assessment of ARGOS location accuracy from satellite tags deployed on captive grey seals. Mar Mamm Sci. 2002;18:301–22.
10. Lowther AD, Lydersen C, Fedak MA, Lovell P, Kovacs KM. The Argos-CLS Kalman filter: error structures and state-space modelling relative to Fastloc GPS data. PLOS One. 2015;10:4.
11. Lopez R, Malardé J, Royer F, Gaspar P. Improving Argos Doppler location using multiple-model Kalman filtering. IEEE Trans Geosci Remote Sens. 2014;52:4744–55.
12. Lopez R, Malardé J, Danès P, Gaspar P. Improving Argos Doppler location using multiple-model smoothing. Anim Biotelemetry. 2015;3:32.
13. Rauch HE, Tung F, Striebel CT. Maximum likelihood estimates of linear dynamic systems. AIAA J. 1965;3:1445–50.
14. McClintock BT, London JM, Cameron MF, Boveng PL. Modelling animal movement using the Argos satellite telemetry location error ellipse. Methods Ecol Evol. 2015;6:266–77.
15. Maxwell SM, Hazen EL, Lewison RL, Dunn DC, Bailey H, Bograd SJ, et al. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. Mar Policy. 2015;58:42–50.

16. Hazen EL, Scales KL, Maxwell SM, Briscoe DK, Welch H, Bograd SJ, et al. A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. Sci Adv. 2018;4:eaar3001.

17. Pirotta V, Grech A, Jonsen ID, Laurance WF, Harcourt RG. Consequences of global shipping traffic for marine giants. Front Ecol Environ. 2019;17: 39–47.

18. Dunn DC, Maxwell SM, Boustany AM, Halpin PN. Dynamic ocean management increases the efficiency and efficacy of fisheries management. Proc Natl Acad Sci. 2016;113:668–73.

19. Treasure AM, Roquet F, Ansorge IJ, Bester MN, Boehme L, Bornemann H, et al. Marine mammals exploring the oceans pole to pole: A review of the MEOP consortium. Oceanography. 2017;30(2):132–8.

20. Harcourt R, Sequeira AMM, Zhang X, Roquet F, Komatsu K, Heupel M, et al. Animal-Borne Telemetry: An Integral Component of the Ocean Observing Toolkit. Front Mar Sci. 2019;6:326. https://www.frontiersin.org/article/10.3389/fmars.2019.00326. Accessed 07 July 2020.

21. Freitas C, Lydersen C, Fedak MA, Kovacs KM. A simple new algorithm to filter marine mammal Argos location. Mar Mamm Sci. 2008;24:315–25.

22. Patterson TA, McConnell BJ, Fedak MA, Bravington MV, Hindell MA. Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error. Ecology. 2010;91:273.

23. Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. TMB: Automatic Differentiation and Laplace Approximation. J Stat Softw. 2016;70:1–21.

24. Jonsen I, McMahon CR, Patterson TA, Auger-Méthé M, Harcourt R, Hindell MA, et al. Movement responses to environment: fast inference of variation among southern elephant seals with a mixed effects model. Ecology. 2019;100:e02566.

25. Jonsen I, Patterson TA. foieGras: Fit continuous-time state-space and latent variable models for filtering Argos satellite (and other) telemetry data and estimating movement behaviour. 2019. R package version 0.4.0. Available from: https://cran.r-project.org/package=foieGras. Accessed 07 July 2020.

26. Bryant E. 2D location accuracy statistics for Fastloc® cores running firmware versions 2.2 & 2.3. Redmond: Wildtrack Telemetry Systems Ltd.; 2007. Technical Report TR01. http://www.wildtracker.com/results_files/Technical%20Report%20TR01.pdf.

27. Silva MA, Jonsen I, Russell DJF, Prieto R, Thompson D, Baumgartner MF. Assessing performance of Bayesian state-space models fit to Argos satellite telemetry locations processed with Kalman filtering. PLOS One. 2014;9:e92277.

28. Tornqvist L, Vartia P, Vartia YO. How should relative changes be measured?. Am Stat. 1985;39:43–6.

29. Boyd JD, Brightsmith DJ. Error properties of Argos satellite telemetry locations using least squares and Kalman filtering. PLOS One. 2013;8: e63051.

30. Costa DP, Robinson PW, Arnould JPY, Harrison AL, Simmons SE, Hassrick JL, et al. Accuracy of ARGOS locations of pinnipeds at-sea estimated using Fastloc GPS. PLOS One. 2010;5:e8677.

31. Hoenner X, Whiting SD, Hindell MA, McMahon CR. Enhancing the use of Argos satellite data for home range and long distance migration studies of marine animals. PLOS One. 2012;7:e40713.

32. Hindell MA, Reisinger RR, Ropert-Coudert Y, et al. Tracking of marine predators to protect Southern Ocean ecosystems. Nature. 2020;580: 87–92.

33. McClintock BT. Incorporating telemetry error into hidden Markov models of animal movement using multiple imputation. J Agric Biol Environ Stat. 2017;22:249–69.

Publisher's Note