A comprehensive framework for handling location error in animal tracking data*

C. H. Fleming[1,2], J. Drescher-Lehman[1,3], M. J. Noonan[1,2,4], T. S. B. Akre[1], D. J. Brown[5,6], M. M. Cochrane[7], N. Dejid[8,9], V. DeNicola[10], C. S. DePerno[11], J. N. Dunlop[12], N. P. Gould[11], J. Hollins[13], H. Ishii[14], Y. Kaneko[14], R. Kays[11,16], S. S. Killen[13], B. Koeck[13], S. A. Lambertucci[17], S. D. LaPoint[18], E. P. Medici[19], B.-U. Meyburg[20], T. A. Miller[21], R. A. Moen[7], T. Mueller[8,9], T. Pfeiffer[22], K. N. Pike[23], A. Roulin[24], K. Safi[25,26], R. Séchaud[24], A. K. Scharf[25,26], J. M. Shephard[27], J. A. Stabach[1], K. Stein[28], C. M. Tonra[28], K. Yamazaki[29,14], W. F. Fagan[2], and J. M. Calabrese[30,31,32,1,2]

[1]Smithsonian Conservation Biology Institute, Front Royal, Virginia, USA
[2]Department of Biology, University of Maryland, College Park, Maryland, USA
[3]Department of Biology, George Mason University, Fairfax, Virginia, USA
[4]The Irving K. Barber School of Arts and Sciences, The University of British Columbia, Okanagan campus, Kelowna, BC, Canada
[5]School of Natural Resources, West Virginia University, Morgantown, West Virginia, USA
[6]Northern Research Station, US Forest Service, Parsons, West Virginia, USA
[7]Natural Resources Research Institute, Department of Biology, University of Minnesota-Duluth, Minnesota, USA
[8]Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Frankfurt, Germany
[9]Department of Biological Sciences, Goethe University, Frankfurt, Germany
[10]White Buffalo Inc., Moodus, Connecticut, USA
[11]Department of Forestry & Environmental Resources, North Carolina State University, Raleigh, North Carolina, USA
[12]Conservation Council of Western Australia, West Perth, Australia
[13]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom
[14]Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan
[16]North Carolina Museum of Natural Sciences, Raleigh, North Carolina, USA
[17]Grupo de Investigaciones en Biología de la Conservación, INIBIOMA (CONICET-Universidad Nacional del Comahue), Bariloche, Argentina
[18]Black Rock Forest, Cornwall, New York, USA
[19]Lowland Tapir Conservation Initiative, Instituto de Pesquisas Ecologicas, Campo Grande, Mato Grosso do Sul, Brazil
[20]BirdLife Germany (NABU), Berlin, Germany
[21]Conservation Science Global, Inc., West Cape May, New Jersey, USA
[22]Weimar, Germany
[23]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia
[24]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[25]Department of Migration, Max Planck Institute of Animal Behavior, Radolfzell, Germany
[26]Department of Biology, University of Konstanz, Konstanz, Germany
[27]College of Science, Health, Engineering and Education, Murdoch University, Murdoch, Australia
[28]School of Environment and Natural Resources, The Ohio State University, Columbus, Ohio, USA
[29]Ibaraki Nature Museum, Bando, Ibaraki, Japan
[30]Center for Advanced Systems Understanding (CASUS), Görlitz, Germany
[31]Helmholtz-Zentrum Dresden Rossendorf (HZDR), Dresden, Germany
[32]Department of Ecological Modelling, Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Abstract

*This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and pre-decisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the U. S. Geological Survey (USGS), it does not represent any official USGS finding or policy.

Animal tracking data are being collected more frequently, in greater detail, and on smaller taxa than ever before. These data hold the promise to increase the relevance of animal movement for understanding ecological processes, but this potential will only be fully realized if their accompanying location error is properly addressed. Historically, coarsely-sampled movement data have proved invaluable for understanding large scale processes (e.g., home range, habitat selection, etc.), but modern fine-scale data promise to unlock far more ecological information. While location error can often be ignored in coarsely sampled data, fine-scale data require much more care, and tools to do this have been lacking. Current approaches to dealing with location error largely fall into two categories—either discarding the least accurate location estimates prior to analysis or simultaneously fitting movement and error parameters in a hidden-state model. Unfortunately, both of these approaches have serious flaws. Here, we provide a general framework to account for location error in the analysis of animal tracking data, so that their potential can be unlocked. We apply our error-model-selection framework to 190 GPS, cellular, and acoustic devices representing 27 models from 14 manufacturers. Collectively, these devices are used to track a wide range of animal species comprising birds, fish, reptiles, and mammals of different sizes and with different behaviors, in urban, suburban, and wild settings. Then, using empirical data on tracked individuals from multiple species, we provide an overview of modern, error-informed movement analyses, including continuous-time path reconstruction, home-range distribution, home-range overlap, speed and distance estimation. Adding to these techniques, we introduce new error-informed estimators for outlier detection and autocorrelation visualization. We furthermore demonstrate how error-informed analyses on calibrated tracking data can be necessary to ensure that estimates are accurate and insensitive to location error, and allow researchers to use all of their data. Because error-induced biases depend on so many factors—sampling schedule, movement characteristics, tracking device, habitat, etc.—differential bias can easily confound biological inference and lead researchers to draw false conclusions.

Keywords: animal tracking, Argos Doppler-shift, DOP, GPS, location error, VHF.

1 Introduction

Technological advances have ushered movement ecology into a new frontier of high-resolution tracking data on a broad range of taxa (Kays et al., 2015). Animal movement data have become central to ecology and can shed light on previously unanswerable questions relating to behavior, population dynamics, and ecosystem function (Bestley et al., 2015; Kays et al., 2015; Noonan et al., 2015; Strandburg-Peshkin et al., 2015; Lennox et al., 2017; McKinnon and Love, 2018; Noonan et al., 2018; Abernathy et al., 2019). As the quality and quantity of tracking data continue to improve, the scope of inference they provide can continue to expand, so long as analytic methods keep pace. Already, modern movement data are informing on activity and energy budgets (Williams et al., 2014; Noonan et al., 2019), fine-scale habitat use and selection (Ewald et al., 2014), species interactions and encounter processes (Martinez-Garcia et al., 2020; Noonan et al., 2020*a*), but the ability to account for location error is an emerging bottleneck (Noonan et al., 2019). With tracking technology, an historical tradeoff exists between the size of a device and the precision of its location estimates, with larger Global Positioning System (GPS; Table 1) and other global navigation satellite system

(GNSS) location estimates typically having location errors on the order of meters, smaller Argos Doppler-shift location estimates having errors on the order of kilometers[1], and the smallest (<1g) light-level geolocator-derived location estimates having errors on the order of hundreds of kilometers (Kaplan and Hegarty, 2006; CLS, 2016; McKinnon and Love, 2018). However, many other factors can influence GPS precision, including satellite reception, the specific hardware and software used in a tracking device, and the time-lag between fixes, with more frequent fixes potentially leading to more accurate post-processed estimates (sub-meter with integrated Kálmán filtering, Kreye et al., 2004). Thus, a universal, one-size-fits-all estimate for the location error representing a single class of technology would neglect a large amount of variation. Here we highlight the fact that having accurate measures of location error can be essential in providing accurate measurements of animal movement, depending on the relative scales involved. This point has long been recognized by those using less accurate technologies such as Argos Doppler-shift tags, where state-space models have been used to account for location error when tracking animal movement (e.g., Jonsen et al., 2007; Johnson et al., 2008; McClintock et al., 2014). However, if not properly accounted for, we argue that typical location errors of GPS units can also result in mistaken conclusions. For example, Noonan et al. (2019) showed that simple measures of speed are particularly sensitive to location error as step-lengths are overtaken by error. Ross-Smith et al. (2016) and Péron et al. (2017) found it necessary to account for vertical errors in the estimation of flight-altitude distributions. Location error is also important when estimating home ranges from triangulated VHF (Gerber et al., 2018) and Argos Doppler-shift (Meckley et al., 2014) tracking data, and so GPS errors should be expected to impact the home-range estimates of animals with smaller home ranges. In all of these cases, biological inference is compromised when location error becomes comparable to the relevant movement scales, and the biases due to mishandling or ignoring location error can easily outstrip the truth (e.g., the more than 10-fold overestimation of wood turtle speed noted by Noonan et al., 2019). As we will elaborate on, these differential biases can cause researchers to draw mistaken conclusions, such as a species traveling faster and more tortuously under dense canopy than in an open habitat.

Most animal movement analyses simply ignore the issue of location error. Beyond that, conventional solutions to mitigating against location error involve either discarding the most erroneous locations according to a threshold (e.g., Bjørneraas et al., 2010; Poessel et al., 2018*a*,*b*) or simultaneously modeling the movement and error with a hidden state model (e.g., Ross-Smith et al., 2016; Péron et al., 2017). As we detail, the former solution is at best incomplete, while the latter cannot reliably distinguish between variance due to movement and variance due to location error. Related to these issues is the problem of outlier detection, which has, until now, relied on metrics assuming no location error, in accordance with the above principle of thresholding. Here we present comprehensive methods and guidelines for dealing with location error (Fig. 1), largely focused on GPS, but also including Argos Doppler-shift, Global System for Mobile Communications (GSM), acoustic trilateralization, and other tracking technologies that provide location estimates. First, we put forward the need and best practices for the error calibration of tags and encourage researchers to obtain

[1]By 'Argos Doppler-shift', we specifically refer to location estimates derived from communication with the Argos satellite system and not GPS location data that are uploaded to the Argos satellite system.

this information before they start collecting tracking data. Second, we provide a statistically efficient framework for quantifying location error, showing how to create and select among error models with 190 example calibration datasets. Finally, we demonstrate how calibrated tracking data then serve as ideal inputs for error-informed movement analyses, including path reconstruction, connectivity quantification, home-range estimation, and distance estimation.

Tracking data and their error information

Satellite location data are accompanied by variable degrees of location error—on the order of 1–10 meters for differential and tri-lane GPS location error, 10–100 meters for conventional GPS horizontal error, several times that for GPS vertical error, and 100–10,000 meters for Argos Doppler-shift horizontal error (Kaplan and Hegarty, 2006; Parkinson and Enge, 1996; CLS, 2016). Location errors are generally *heteroskedastic*—meaning that their variances can change in time, depending on the reception and spatial configuration of the satellites and tracking device. While older Argos Doppler-shift location data only come with error classes that must be calibrated (Vincent et al., 2002) or fit simultaneously with a movement model (e.g., Johnson et al., 2008), modern Argos data come with $\sqrt{2}$-standard-deviation error ellipses, estimated by a custom multiple-model Kálmán filter (Lopez et al., 2014). Error ellipses are more necessary with Argos Doppler-shift location estimates, because the polar orbits of Argos satellites provide better resolution of latitudes than longitudes. Argos Doppler-shift location data were also the first to be modeled via state-space models incorporating both continuous-time non-Markovian movement processes and heteroskedastic errors (Johnson et al., 2008).

GPS device manufacturers do not generally provide calibrated error circles of known quantile or standard deviation. In contrast to modern Argos Doppler-shift data, GPS location estimates typically come with unitless and device-specific "horizontal dilution of precision" (HDOP) and "vertical dilution of precision" (VDOP) values. Ideally, DOP values are designed to account for the order-of-magnitude heteroskedasticity stemming from satellite reception, geometry, latitude, and dynamics (Kaplan and Hegarty, 2006; Yahya and Kamarudin, 2008), which, for the purpose of animal tracking, includes the effects of local habitat, canopy, cover, burrowing, device orientation, speed, and most everything else short of space-weather events (Coster and Komjathy, 2008). DOP values provide the time-dependent transformation

$$\underbrace{\text{location error}(t)}_{\text{heteroskedastic}} = \text{DOP}(t) \times \underbrace{\text{UERE}(t)}_{\text{homoskedastic}} , \tag{1.1}$$

between heteroskedastic location error and homoskedastic "user equivalent range error" (UERE). UEREs can be thought of as partially standardized location errors, and can be assumed to be a mean-zero, stationary Gaussian process (Kaplan and Hegarty, 2006), which defines our null model. By quantifying the relatively simple distribution of UEREs that can often be described by a single scale parameter, we can back-transform with DOP values to quantify the time-dependent distribution of location errors. Traditionally, the UERE scale parameter estimated is the root-mean-square (RMS) UERE. We refer to the direct estimation of UERE parameters—and subsequent back-transformation via relation (1.1)—as error 'calibration'. However, GPS location data are sometimes missing relevant DOP information and may exhibit more complex error structures than present in the null model (1.1), which can necessitate error-model selection.

Triangulated VHF location errors tend to be intermediate to GPS and Argos Doppler-shift errors in magnitude, and their location estimates come equipped with error ellipses when using the method of Lenth (1981), such as from `R` package `sigloc` (Berg, 2015). (Also, see Gerber et al. (2018) for full posterior estimation of the location-error distributions.) Therefore, operationally speaking, triangulated VHF tracking data can be treated analogously to modern Argos Doppler-shift data.

Acoustic trilateralization is more popular in marine systems, where electromagnetic signals rapidly attenuate. These systems operate much like GPS, but via acoustic signals with receivers fixed underwater instead of electromagnetic signals with satellites in orbit. In particular, Vemco positioning system (VPS) location estimates come with "HPE" values, which Vemco documentation describes as being analogous to GPS HDOP, in that they are proportional to error-circle radii and must be calibrated on a per-network basis (Smith, 2013).

'Global System for Mobile Communications' (GSM) cellular networks provide a number of data sources that devices can use to estimate location, such as time-of-arrival, angle-of-arrival (azimuth), and signal strength with respect to multiple cellular towers (Martínez Hernández et al., 2019). Resulting location estimates can vary in quality, depending on the techniques employed and network coverage, and may return error-parameter estimates analogous to GPS, though error ellipses may be more desirable.

In summary and irrespective of how they are collected, tracking data that provide location estimates usually come with either error-circle (GPS, GSM, VPS) or error-ellipse (triangulated VHF, Argos Doppler-shift, geolocator) information. Sometimes this information is only proportional to the location error's variance (GPS, GSM, VPS), and further estimation is required to fully quantify the error distribution.

When can location error be reasonably ignored?

It can be reasonable and pragmatic to simply ignore location error in a focal study, if the scales of error are far smaller than all relevant scales of movement. For example, 10-100 meter GPS error can be considered insubstantial in comparison to a global migration sampled on a daily schedule, where individuals travel many kilometers between location fixes. In situations where this comparison is less clear cut, researchers can perform a sensitivity analysis to approximate the impact of ignoring location error via simulation (App. S1). Because these sensitivity estimates are approximate and require as much effort as more-exact, error-informed analysis, we only recommend this procedure in cases where there is no error-informed analysis at hand.

Conventional approaches to handling location error

The most common approach in the literature to deal with location error is simply to discard the most erroneous location estimates, either with an arbitrary HDOP threshold for GPS data or a location-class threshold for Argos Doppler-shift data (Bjørneraas et al., 2010). This threshold is especially arbitrary for GPS data, as HDOP values do not have the same meaning across device models, even when assuming the null model of location error = DOP $\times$ UERE (1.1), and complicated by the fact that most manufacturers do not provide UERE parameters for their wildlife telemetry devices. Consider, for example, that if the RMS UERE of device model A is twice that of device model B, then an HDOP value of 10 on device model A is equivalent to an HDOP value of 20 on device model B. As an improved

criterion, Meckley et al. (2014) argued for objective thresholding, based on calibrated error estimates (in meters), rather than unitless precision estimates that vary by device model. However, thresholding the data still involves a trade off, whereby reduced bias comes at the cost of a smaller sample size. At both extremes—large included location errors and small sample sizes—there can be substantial bias from too much location error or not enough data, and an intermediate threshold that can sufficiently mitigate both biases may not exist (Noonan et al., 2019).

Rather than discarding data, state-space models consider the location fix as the combination of both movement and location error, with unknown parameters for each model. State-space models constitute the second most common approach in the literature to account for location error. However, the movement and location-error parameters are almost always simultaneously estimated from tracking data (e.g., Kuhn et al., 2009; Swimmer et al., 2009; Howell et al., 2010; Sulikowski et al., 2010; Vermard et al., 2010; Matthews et al., 2011; Blanco et al., 2012; Carman et al., 2012; Hückstädt et al., 2014; Dalleau et al., 2014; Kennedy et al., 2014; Ross-Smith et al., 2016; Afonso et al., 2017; Péron et al., 2017). As has been pointed out by Auger-Méthé et al. (2016), state-space models can misappropriate variance due to movement and location error when the two sets of parameters are fit simultaneously to tracking data, as is common practice in ecology. Fundamentally, this misappropriation is due to a lack of statistical *identifiability*—if movement and error models can produce similar outputs, then their parameters (and influences) cannot be reliably teased apart (Hilborn and Mangel, 1997).

As we detail here, error calibration is a straightforward solution to the problem of statistical identifiability. However, there is an abundance of historical tracking data for which no calibration data exist, and the simultaneous fitting of movement and location-error parameters performs better in some situations than others. Therefore, it is still useful to understand under what circumstances simultaneous fitting can succeed or fail.

A comprehensive framework for addressing location error

Our proposed framework to account for location error in animal tracking data is a general approach with three basic steps—calibration-data collection, error-model selection, and error-informed analysis—summarized in Fig. 1 and briefly described below. Calibration-data collection is detailed more thoroughly in Sec. 2.1. Error-model selection is then covered in Sec. 2.2, with derivations given in App. S2–S3, and 27 error-model-selection examples provided in App. S4. Finally, Sec. 3 presents empirical examples of error-informed movement analyses on tracking data, where accounting for location error impacts inference.

1. Calibration-data collection First, we address the lack of statistical identifiability between movement and location-error parameters, when estimating the two simultaneously in a state-space framework. Instead of simultaneous estimation, we propose calibrating all tracking data before movement analysis. In the calibration step, researchers collect location data that have a known movement model, which we refer to as calibration data. Most simply, we record location data at fixed locations, so that all variance in the calibration data is solely due to location error. We then estimate error-model parameters from the calibration data, which addresses the identifiability issues pointed out by Auger-Méthé et al. (2016).

2. Error-model selection Second, we address the lack of standardization across GPS devices, the incomplete location-error information accompanying many GPS data, and the variable degree of performance provided by information on location error. The location-error predictors we discuss in App. S3 and apply in App. S4 include DOP values, the number of satellites, and various location-class proxies. Therefore, we introduce a new model-selection framework based on corrected Akaike information criterion (AIC_C) values (App. S2). AIC-based model selection achieves asymptotically optimal predictions (Yang, 2005), and is known to perform well when selecting among a small number of plausible candidate models, while 'corrected' AIC_C values are adjusted for bias (Burnham and Anderson, 2002). We also derive a goodness-of-fit statistic to objectively gauge the selected model's performance.

3. Error-informed analyses The third and final step of our proposed framework is to feed these calibrated data into error-informed—and preferably, continuous-time—analyses. Both the `crawl` (Johnson, 2008; Johnson et al., 2008) and `ctmm` (Fleming and Calabrese, 2015; Calabrese et al., 2016) `R` packages are capable of performing a large number of continuous-time movement analyses while accounting for (calibrated) heteroskedastic location errors in a state-space framework, including path reconstruction, conditional simulation, and the calculation of utilization distributions (Fleming et al., 2015, 2016, 2017). The family of Brownian-bridge methods within `R` package `move` (Kranstauber et al., 2012) can also incorporate calibrated location errors, as can the behavioral-segmentation methods in `R` packages `momentuHMM` (McClintock and Michelot, 2018) and `smoove` (Gurarie et al., 2017)—both by leveraging `crawl`.

Here, we expand on these methods by introducing error-informed statistics for outlier detection and variogram-based autocorrelation visualization. We provide a thoroughly tested and documented software implementation for all analyses in the CRAN-hosted `ctmm` `R` package, complete with both help files and long-form documentation via `vignette('error')`. Finally, we demonstrate the necessity and utility of these methods with several empirical examples, where accounting for location error makes the difference between accurate inference and qualitatively incorrect inference.

2 Calibration-data collection and error-model selection

2.1 How to collect calibration data

In the simplest case, 'calibration data' are location fixes obtained when a tracking device is not moving. Error-model parameters are best estimated when the true movement model is known, to avoid the potential for misspecification. The simplest movement model possible is one of no movement. Therefore, to collect calibration data, tracking devices should be left at fixed locations over an extended period of time—preferably for days. The true locations do not need to be known a priori; these can be estimated simultaneously with the UERE parameters, in a statistically efficient manner.

How many tracking devices should be deployed?

Ideally, only one device per model is needed to collect calibration data, as all devices of the same make and model should have similar UERE distributions. However, multiple devices of the same model can be deployed and the assumption of identical devices can be tested via

model selection. If UERE parameter estimates are consistent, then those calibration data can be integrated together to obtain more accurate estimates. If calibration data come at the cost of limited battery life that would otherwise be used to collect animal tracking data, we suggest a two-stage calibration deployment when device heterogeneity is a concern. In the first round of deployments, a small amount of pilot calibration data are collected for each device, such that the total number of sampled locations can meet the target accuracy, were the UERE distributions identical. Discrepancy among devices can then be tested for and, if found significant, further per-device calibration data can be collected in a second round of deployments. Otherwise, if UERE parameter estimates are found to be consistent, the single estimate is sufficient.

How much calibration data should be collected?

UERE parameters are estimated from data, and so increasing the sample size will increase their precision. At minimum, calibration data should be collected over the course of a day, to sample a wide range of satellite configurations and DOP values. Researchers may want to collect many days of calibration data if space weather events are of concern (Coster and Komjathy, 2008). More location fixes will provide more precise UERE statistics, but the sampling frequency should not be so high as to induce location-error autocorrelation not present in the tracking data. For GPS devices featuring on-board location-error filtering, the RMS UERE may shrink at higher sampling rates ($\sim$1 Hz) and, therefore, the sampling schedule of the calibration data should be the same as that of the tracking data. If the same device has been used at disparate sampling rates, then the presence of an on-board error filter can be tested with calibration data collected over the same range of values.

Relative error in a single mean-square (MS) UERE estimate will be of standard deviation $\sqrt{2/q(N-K)}$ (c.f., Eq. S2.7), where q is the number of spatial dimensions ($q = 2$ for horizontal UEREs and $q = 1$ for vertical UEREs), N is the total number of sampled locations, and K is the number of unknown locations where calibration data were collected. With only one calibration deployment per tracking device, K is simply the number of devices. Therefore, given some relative error tolerance $\mathrm{TOL} = \mathrm{STD}[\hat{\theta}]/\theta$ in the MS UERE estimate, the *total* number of calibration fixes necessary to meet that tolerance in the MS UERE estimate is

$$N_{\mathrm{TOL}} = K + \frac{2}{q\,\mathrm{TOL}^2}\,. \tag{2.1}$$

E.g., for 10 devices in two dimensions and deployed once each, the requisite number of calibration fixes *per device* is 41 (or 1001) for relative error tolerances of 5% (or 1%). If the 10 devices are found to have significantly different UERE parameters, then in the second round of deployments an additional 361 (or 9001) calibration fixes must be collected per device, to meet the same relative error tolerance in the individualized parameter estimates.

Where should calibration data be collected?

For newer tracking devices that estimate reliable DOP values and mitigate against multipath errors (Table 1), it should not be necessary to sample calibration data in multiple habitats or latitudes. The effect of variable signal reception across different habitat types should be accounted for in the DOP values. But, again, this assumption can be challenged, by first collecting a small amount calibration data in each habitat, testing for discrepancy, and

then collecting more calibration data if necessary. For devices that feature multiple 'location classes'—location estimates of substantially different precision, regardless of DOP value—it may be important to sample in areas of poorer satellite reception, because if only the best location classes are observed in the calibration data, then the worst location classes cannot be reliably calibrated. Furthermore, to test the veracity of a device's DOP values, less-than-ideal signal reception may be preferable to magnify their presumed impact.

Opportunistic calibration data

For studies where calibration data were not collected and cannot be collected after the fact, it still may be possible to assemble calibration data opportunistically. Segments of time when the individual was deceased, at a rest site, or when the collar was detached may be subsetted and treated as calibration data. Researchers should be especially careful that opportunistic calibration data are free of movement. Methods to test for the presence of movement in relocation data include correlogram analysis (Fig. S4.1) and attempting to fit an autocorrelated movement model (simultaneous with unknown UERE parameters) to the data. However, both of these approaches assume that the data are sampled finely enough that movement and error are easily distinguishable, yet coarsely enough that the location error itself is not autocorrelated, which will typically be the case for GPS location estimates sampled at least 1–2 minutes apart (Tao and Bonnifait, 2015).

2.2 An error-model-selection framework

After error calibration data have been collected, an appropriate error model can be selected. In the null model of GPS location error (1.1), DOP values contain all dependence on satellite geometry, including the number of satellites in communication. On the other hand, UEREs are generally assumed to be a mean-zero Gaussian process (Kaplan and Hegarty, 2006), characterized by its scale parameter (e.g., RMS UERE), even though the location errors may have a heavy-tailed distribution (Fig. 2). However, complications may arise, such as if DOP values are not recorded by the device or there are multiple location classes with different UERE distributions. Different error models may be appropriate in different situations. Therefore, in App. S2 we derive RMS UERE estimators, $\mathrm{AIC_C}$ values, and goodness-of-fit statistics for location-error models fit to calibration data. We have focused great attention on the statistical efficiency of our estimators, which have ideal performance for one location class and normally distributed UEREs, including for the null model (1.1), and better performance than maximum likelihood more broadly. Unbiased estimators are important when pooling small amounts of calibration data, and when calibrating multiple location classes where some fix types rarely occur (c.f. S4.11).

As tracking devices can vary in behavior and poorly specified location-error models can lead to heavy-tailed UERE distributions, error-model selection can be important for making the best use of tracking data. AIC values serve as a model selection criterion that balances goodness of fit against parsimony (Burnham and Anderson, 2002) and can provide asymptotically optimal predictions (Yang, 2005). However, in situations where sample sizes are small and debiased parameter estimates differ substantially from (biased) maximum-likelihood parameter estimates, 'corrected' $\mathrm{AIC_C}$ values can differ substantially from regular AIC values. Furthermore, it is not sufficient to adopt off-the-shelf $\mathrm{AIC_C}$ formulas, because they are both model and estimator dependent (Calabrese et al., 2018; Fleming et al., 2019), and the most

commonly used $\mathrm{AIC_C}$ formulas are derived specifically for linear models estimated via maximum likelihood. Therefore, our $\mathrm{AIC_C}$ model-selection criterion is based on newly derived formulas, specific to our model structures and debiased parameter estimators (App. S2.1.1 & S2.2.1).

While useful as a model selection criterion, $\mathrm{AIC_C}$ values scale with the amount of data and cannot be used to assess a model's goodness-of-fit for comparing performance across devices (Burnham and Anderson, 2002). Therefore, we develop an absolute goodness-of-fit statistic—reduced Z^2 or Z^2_{red}—that can be compared across animals, devices, and studies. Z^2_{red} is an error-model analog to the well-known reduced chi-squared statistic, χ^2_{red}, but based on Fisher's Z distribution (App. S2.1.2 & S2.1.2). As with χ^2_{red}, smaller values of Z^2_{red} denote better performance and knowing the true model results in $Z^2_{\mathrm{red}} = 1$ on average.

In App. S3 we introduce a number of error models for GPS devices with missing or incomplete DOP information. The predictors covered include the hierarchy of DOP values (HDOP, VDOP, PDOP and GDOP—Table 1), approximating DOP values with the number of satellites, and the possibility of GPS fix type and time-to-fix as modifiers. We then apply these error models to 190 GPS, GSM, and VPS devices representing 27 device models from 14 manufacturers in App. S4, using our error-model-selection framework with `R` implementation in the `ctmm` package (Fleming and Calabrese, 2015; Calabrese et al., 2016).

For most of the devices we analyze in App. S4, `ctmm`'s null location-error model—based on the advice of App. S3—is the selected error model, and the calibration step is as simple as the `R` assignment

```
uere(DATA) ← uere.fit(CALIBRATION) ,
```

where `DATA` refers to the tracking data and `CALIBRATION` refers to the calibration data. For pre-calibrated location data, such as Argos Doppler-shift and triangulated VHF location estimates, no calibration steps need to be taken, while problematic devices exhibiting a statistically inefficient null model will require model selection. A worked example is given in `vignette('error')` within `R` package `ctmm`.

2.3 Error or outlier?

A dual task to modeling location error is the classification of outliers, which are observations that defy the combined movement and error model (Fig. 3). For instance, a recorded flight altitude only slightly below ground with a large corresponding VDOP could be considered typical (p-value $\gg 0$), while a recorded flight altitude deep below ground with a small corresponding VDOP should be extraordinarily rare (p-value ≈ 0) under normal circumstances and therefore considered an outlier. Normatively erroneous data are not outliers and can still contain meaningful information that should be retained, as long as the applied statistical framework incorporates an appropriate error model. Discarding data that are not true outliers can bias inference (Péron et al., 2020).

The classification of outliers aside, the second problem is what to do with them. If the data contain many outliers, then the ideal solution is to improve the error model until said outliers can be considered normative (c.f., App. S4.5.2). However, if the number of outliers is small, then requisite error-model parameters may not be well resolved, and even so, the information gained by incorporating these data may be insubstantial. Therefore, removing outliers can be a pragmatic choice, especially in animal tracking data, which can be plagued

with baffling abnormalities. In any case, these decisions should always be reported with resulting analyses.

It is standard practice to classify outliers in animal tracking data using simple metrics such as speed and distance, so that location estimates implying implausible movements can be ruled out (Douglas et al., 2012; Safi and Kranstauber, 2021). However, traditionally these outlier metrics have not been informed by location error, and so they can confound normative errors with true outliers. For example, straight-line distance is commonly used as an estimate of the minimal path-length requirement between sequential location estimates, yet location errors cause this calculation to overestimate true distances (Ranacher et al., 2016; Noonan et al., 2019). In App. S5 we derive maximum-likelihood distance and speed estimators that account for location error and are minimally conditioned on the data. These estimators are implemented in the `ctmm` (Fleming and Calabrese, 2015; Calabrese et al., 2016) method `outlie()`, and demonstrated in Fig 3. While the example in Fig. 3 is visually striking, outliers in finely sampled data are not generally obvious, and their implausibility cannot be properly discriminated without accounting for the associated location errors.

Location-error outliers can occur in both tracking and calibration data. Identifying outliers in calibration data, therefore, may require some degree of iteration, whereby researchers first assume a plausible RMS UERE value and then calculate a more accurate RMS UERE value after removing all contingent outliers. The updated RMS UERE can then be checked for consistency by applying it to the initial data (with outliers intact) and determining if the same classification of outliers is recovered (Fig. 1).

3 Empirical examples

We present two sets of empirical analyses—one set on an expansive collection of calibration data and another set on individual tracked animals. First, we analyze the quality of error information provided by tracking devices in Sec. 3.1. Second, we demonstrate the impact of proper error modeling and error-informed methods on biological inference in Secs. 3.2-3.5.

3.1 Device calibrations

Here we summarize the application of our error-model-selection techniques to 190 GPS, cellular, and VPS devices representing 27 device models from 14 manufacturers. These tracking devices are used on a number of species, including American black bear (*Ursus americanus*), Andean condor (*Vultur gryphus*), bald eagle (*Haliaeetus leucocephalus*), barn owl (*Tyto alba*), black-crowned night-heron (*Nycticorax nycticorax*), European perch (*Perca fluviatilis*), eastern and western Santa Cruz Island Galapagos giant tortoises (*Chelonoidis donfaustoi* and *Chelonoidis porteri*), eastern coyote (*Canis oriens*), fisher (*Pekania pennanti*), golden eagle (*Aquila chrysaetos*), northern pike (*Esox lucius*), raccoon (*Procyon lotor*), white-tailed deer (*Odocoileus virginianus*), and wood turtle (*Glyptemys insculpta*). For each device model, the selected error models and their goodness-of-fit statistics are reported in table 2, while the details of each individual model-selection procedure are provided in App. S4. With the following statistics on these 27 device models, we also report 95% binomial confidence intervals to make population-level inferences on devices of this era.

In 56% (39%–75%) of device models, the reported data on location error was trustworthy enough that model selection was not necessary to obtain the best performing error model, if

following the guidelines suggested in App. S3. In 77% (63%–91%) of device models, one set of calibration parameters could safely be assumed valid for all devices of the same make and model. Only 7% (1%–24%) of device models had a selected error model that was nontrivial and required substantial modeling effort. The Sirtrack Pinnacle Solar G5C275F had location errors that shrank by a factor of four after 32 seconds of "time on", while the ATS G2110e had location errors that decreased smoothly with number of satellites and increased smoothly with time-to-fix. Both of these device models' reported HDOP values that were either not informative or were marginally informative.

In terms of horizontal errors, an estimated 22% (9–42%) of device models had DOP values that were *misinformative*, in that they provided worse information on location error than a model of homoskedastic errors. All other properties held fixed, a larger Z^2_{red} statistic indicates a heavier tailed UERE distribution. In several examples, the DOP values were so poor that their Z^2_{red} were comparable to that of a Cauchy distribution, which does not even have a finite variance. Furthermore, an estimated 0% (0%–52%) of device models had particularly informative VDOP values, in that they provided the best information on vertical location error, as they are expected to.

Finally, for each device we estimated its (horizontal) RMS location error under ideal conditions, which we denote as RMSE_{min}. Ideal conditions typically amount to HDOP=1, in which case RMSE_{min} is equivalent to the RMS UERE calibration parameter, but some devices return DOP values below 1 or error estimates in meters rather than unitless DOP values. RMSE_{min} estimates ranged an order of magnitude, from 1.7 meters to 21 meters, with the more precise devices likely relying on newer dual-frequency GPS receivers or on-board Kálmán filtering. Similarly, RMS UERE calibration parameters also varied by an order of magnitude.

3.2 Path reconstruction of a GPS-tracked common noddy

Methods

To demonstrate the impact of heteroskedastic location errors on the simultaneous fitting of movement and error parameters, we considered 4 days of common noddy (*Anous stolidus*) GPS data, collected at 20-minute intervals with corresponding position-DOP (PDOP) values. Locations were clipped from around the nest to avoid non-stationarity due to the switching between foraging and nesting behaviors. To this dataset we performed a standard continuous-time movement-model-selection procedure (Calabrese et al., 2016) with three different error models—a simultaneously fit homoskedastic location-error model, a simultaneously fit PDOP-informed error model, and a calibrated PDOP-informed error model. For each selected model we inspected the location-error estimates and corresponding occurrence distributions. Occurrence distributions quantify where the individual was located during the sampling period by probabilistically reconstructing the movement path (Fleming et al., 2016). The most common example of an occurrence distribution is the Brownian bridge (Horne et al., 2007), which assumes a movement model of Brownian motion. In contrast, here we selected a movement model via AIC.

Results

When simultaneously fitting movement parameters with a homoskedastic error model, the RMS location error was estimated to be 500 meters (95% CI: 390–610 m), which is implausibly large for GPS-quality data. Simultaneously fitting movement parameters with a PDOP-informed error model improved the RMS location-error estimate by an order of magnitude to 44 meters (21–68 m) at PDOP=1, which is still several times larger than expected. By simultaneously fitting error and movement model parameters, these results correspond to the conventional application of state-space models. Finally, when using a small sample of calibration data, we found the RMS location error to actually be only 3–4 meters at PDOP=1, which is consistent with a GPS device with a dual-frequency GPS receiver or an on-board Kálmán filter.

Occurrence distributions corresponding to the selected movement and error models were calculated and are shown in Fig. 4. These distributions correspond to the probability of where the animal was during the sampling period, and their coverage area quantifies our ignorance of where the animal was located throughout that time (Fleming et al., 2016). Occurrence distributions can essentially be thought of as posteriors for the unknown trajectory. Compared to the most accurate location-error model (Fig. 4C), the simultaneously fit homoskedastic error model produced a 95% occurrence area that was 140% larger (Fig. 4A). In contrast, the model with simultaneously fit PDOP-informed error model only produced a 95% occurrence area that was 4% larger (Fig. 4B). Importantly, occurrence distributions are dependent on both the underlying movement and location-error processes. And while there was an expected trade-off between estimated variance due to location error and due to movement, in this example the movement parameter estimates only varied by $\sim 15\%$, even though the location-error parameter estimates varied by orders of magnitude. Therefore, larger differences here in the occurrence distributions can mostly be attributed to differences in the location-error models.

3.3 Home range of an Argos-tracked night-heron

Methods

We considered 6 months of data on a black-crowned night heron (*Nycticorax nycticorax*), tracked with an Argos Doppler-shift tag in ∼60-minute intervals during its range-resident period—wherein the heron was not migrating and had a well established home range. These Argos Doppler-shift data were of the older variety that do not come pre-calibrated. Instead, each location estimate was accompanied with a location class—3,2,1,0,A,B,Z—for which we used the calibration results of Vincent et al. (2002) to estimate the error ellipses. We applied two home-range estimators, a conventional kernel density estimator (KDE) and an error-informed autocorrelated kernel density (AKDE) estimator with weights optimized to mitigate against irregular sampling (Fleming et al., 2015; Fleming and Calabrese, 2017; Fleming et al., 2018). In the latter method, both the autocorrelation estimator and bandwidth optimization tease apart variance due to movement and error, and the data are first Kriged to reduce location error before placing the kernels (Fleming et al., 2016, 2017). The estimator represents the default output of the `akde()` method in the `ctmm R` package when conditioning on an error-informed movement model. Moreover, for these data, AKDE without a location-error model also results in conventional KDE, because these data appear independent when

ignoring error. Therefore, our comparison effectively includes both error-informed and non-informed AKDE.

With each estimator, we manipulated the quality of input data, from only including the best location class (3) to including all of the data (3–B). Researchers with Argos Doppler-shift data typically will threshold their data class-wise like this based on quality before applying methods that do not account for error. Therefore, our analysis tests the sensitivity to this choice of cutoff, in addition to biases that stem from not accounting for location error.

Results

The conventional KDE home-range estimates were highly sensitive to the location-class threshold, and larger than all error-informed estimates even when restricted to the highest quality data (Fig. 5). In contrast, error-informed AKDE produced consistent home-range estimates with respect to included location class, though the sample size dropped precipitously when only including the highest quality fixes.

3.4 Home-range overlap of an Argos-tracked night-heron

Methods

To further demonstrate the impact of location error on home-range estimates we considered the degree of home-range overlap for a night-heron across two summers (2017, 2018), also tracked with an Argos Doppler-shift tag, but complete with pre-calibrated error-ellipse information. We applied both conventional KDE and error-informed AKDE, and compared overlap (Winner et al., 2018) between summer ranges as a measure of year-to-year site fidelity. Again, these data appear independent when ignoring location error, and so AKDE without an error model results in conventional KDE. Therefore, our comparison effectively includes both error-informed and non-informed AKDE.

Results

According to the conventional KDEs, which is biased by location error to have inflated variance and thus inflated overlap, the 2017 and 2018 summer ranges were largely similar at 82% (78%–86%) overlap, whereas according to the error-informed AKDEs the 2017 and 2018 summer ranges were adjacent, but largely distinct, at 33% (22%–46%) overlap (Fig. 6). Therefore, in this case, the estimator used has a direct impact on biological inference, because the Argos Doppler-shift location errors are large relative to the home-range area of the night-heron.

3.5 Speed of GPS-tracked wood turtle

Methods

We considered 6.3 months of wood turtle (*Glyptemys insculpta*) data, tracked with a Lotek GPS tag in 1-hour intervals, for which we apply the location-error model selected in Sec. S4.5.2. We applied two mean-speed estimators: a conventional straight-line-distance estimator and a continuous-time speed and distance estimator that accounts for both location error and tortuosity (Noonan et al., 2019). For each estimator, we manipulated the input location error by only considering location fixes below a threshold HDOP value.

Results

While the straight-line distance estimates were extremely sensitive to the HDOP cutoff, the error-informed estimates were largely consistent across threshold levels (Fig. 7). With the conventional straight-line distance estimator, two primary sources of bias exist—positive bias from ignoring location error (Ranacher et al., 2016; Noonan et al., 2019) and negative bias from ignoring tortuosity (Rowcliffe et al., 2012; Noonan et al., 2019). In contrast to home-range estimation, discarding erroneous observations to mitigate against the first bias actually increases the second bias. Therefore, while in Fig. 7 the mean-speed estimate is improved by discarding erroneous data, in other cases with faster species the same manipulation can produce an increasingly net negative bias. Importantly, the "Goldilocks"—not too much, not too little—HDOP threshold required to balance these two biases cannot be known a priori, as it depends sensitively on the scales of movement, scales of location error, and sampling schedule.

4 Discussion

We are currently in a golden age of biotelemetry, where more individuals and more taxa are being tracked than ever before, and with no end in sight (Kays et al., 2015). Tracking data have become highly desirable for a variety of ecological, evolution and conservation purposes. That is why, when working with animal tracking data, it is critical to include an appropriate error model in any movement analysis where the scales of location error are substantial in comparison to the relevant scales of movement (Noonan et al., 2019). The widespread practice of discarding location fixes based on a DOP-value or location-class threshold can only serve as a partial mitigation against location error, at the cost of diminished sample size. Using empirical data, we have demonstrated that error-naive estimates can be very sensitive to the chosen threshold, and cannot be expected to converge to error-informed estimates, even when most of the data has been discarded (Figs 5-7). Without an appropriate error model, common movement metrics can easily be biased by an order of magnitude (also see Noonan et al., 2019). We have also demonstrated that device-specific errors can vary by an order of magnitude (table 2), which means that a DOP value of 1 on one model device can be equivalent to a DOP value of 10 on another model device. On the other hand, simultaneously estimating both the movement and location-error parameters, which is also a common practice, is known to be problematic because the variances due to movement and due to error cannot be reliably distinguished (Auger-Méthé et al., 2016; Noonan et al., 2019).

Motivated by these challenges, we have introduced a comprehensive guide and framework for addressing location error in animal tracking data. Our framework consists of a three-step solution to account for location error in movement analysis (Fig. 1) starting with the collection of calibration data and the application of formal model-selection techniques to construct a statistically efficient location-error model. Finally, with error-informed movement analyses, researchers can use all of their data to produce the best possible estimates. We have also added to the existing number of error-informed analyses with statistics for outlier detection and autocorrelation estimation.

Error-informed analysis on calibrated tracking data can be necessary for meaningful biological inference, as the amount of bias present in conventional estimates is a function of the sampling schedule, scales of movement, and scales of error. When making comparisons

across time, across individuals, or across taxa, observed differences in estimated behavior may simply be the result of differential biases, when using traditional methods. For instance, if estimating energy expenditure via distance traveled for conspecifics at two different study sites, presumed differences could be the result of environmental factors driving hypothesized biological mechanisms, or there could a be non-biological explanation related to inaccuracy. Differences in distance estimates could be the result of two different model tracking devices with substantially different calibration parameters having been deployed at the two study sites, so that the individuals tracked with less precise receivers appear to move more tortuously and travel for longer distances. Differences in distance estimates could also be the result of the study sites corresponding to different habitats with inconsistent satellite reception, such as an open grassland versus a canopied woodland, and so the individuals that spend more time in woodland habitat appear to move more than their grassland counterparts. Unfortunately, there is no sure way to discriminate these possibilities with commonly used methods. Furthermore, because these biases also depend on the animals' movement characteristics, it not generally sufficient to make error-naive comparisons even if the tracking device, sampling schedule, and habitat are all identical (e.g., Noonan et al., 2019, Fig. 4c). Differential bias is especially likely to happen in repeated studies over time, as older devices are replaced by newer device models, and in studies where different researchers collected data in different habitats and with different tracking devices. As movement data continue to amass and larger, collaborative studies become increasingly possible, this issue will only increase in importance. For instance, Morato et al. (2016) compared the movement behaviors of 44 jaguars across 5 biomes with 5 different model GPS tracking devices, which was "the largest collection of jaguar movement data analyzed to date." Combinations of GPS and Argos Doppler-shift tracking data are also common in marine systems (e.g., Phalan et al., 2007; Raymond et al., 2015; Citta et al., 2018).

As location data may be applied to a variety of purposes in ecology, we suggest that the collection and analysis of GPS calibration data needs to become standard practice for researchers and managers in the short term. The error-model-selection framework we have introduced here consists of statistically efficient parameter estimators, $\mathrm{AIC_C}$ formulas, and a goodness-of-fit statistic (Z^2_{red})—all custom-derived for working with error-laden location data. As we have shown in App. S4, GPS tracking devices output a variety of data related to location error and can exhibit a variety of error structures. At this time, model selection is often necessary to ensure that location error is accurately described, but consumer pressure on manufacturers could ease this burden considerably, as we detail below.

For existing and near-future tracking data, it would also be useful to have a repository of calibration data, sorted by device make and model, for researchers to draw from. For many historical datasets, the original tracking devices no longer exist or are no longer functional, and even today the choice to collect adequate calibration data can sometimes come at the cost of tracking fewer individuals. Our curated calibration dataset (Fleming et al., 2020), featuring 190 devices comprising 27 models from 14 manufacturers, is a first start toward this goal.

Manufacturer Recommendations

We suggest that the task of calibration and error-model selection (Fig. 1) be performed by GPS device manufacturers, as is the case with Argos Doppler-shift horizontal errors, by

reporting $\sqrt{2}$-standard-deviation error circles or equivalent. At a minimum, all error information from the GPS module should be returned and calibration data should be provided—preferably per device and at different sampling rates if on-board error filtering varies. Moreover, complete version and build information are also necessary to know which calibration parameters can be expected to be applicable across devices.

On a lower level, we note that Argos offers both Kálmán-filtered Doppler-shift location estimates and unfiltered (least-squares) location estimates. This kind of choice would also be appealing to GPS consumers. Researchers that process their own data with software packages such as `crawl` and `ctmm` can potentially obtain better precision with unfiltered data, instead of processing the same data twice over. Error ellipses would also be beneficial for GPS tracking data in some cases and GSM tracking data in most cases.

Device performance

It is straightforward to answer the question of which GPS devices have the smallest errors in a given environment. However, such a comparison would be limited to the specified environment. Our Z^2_{red} goodness-of-fit statistic provides a different comparison that addresses which device has the more predictable distribution of location errors. In other words, a small Z^2_{red} indicates that the device gives fair warning of large errors, with few false alarms. Moreover, Z^2_{red} does not require that environmental conditions be controlled for, and so these statistics can be compared more generally. Smaller values of Z^2_{red} indicated better error-model (e.g., DOP-value) performance, and $Z^2_{\text{red}} = 1$ results when the location errors are properly calibrated and the resulting UEREs are normally distributed, which is not an unreasonable assumption (Kaplan and Hegarty, 2006). In general, less informative DOP values will fail to explain heteroskedasticity and produce heavier tailed UERE distributions and larger values of Z^2_{red}. For the 27 GPS, cellular, and VPS device models we analyzed, the AIC_C-best error models produced Z^2_{red} statistics ranging from 1.3–4.2, with a median value of 2.3 (Table 2).

Beyond the Z^2_{red} statistic, at least two other measures of error performance are necessary for a comprehensive comparison—the RMS location error under ideal conditions (RMSE_{min}) and a standardized measure of DOP-value response to satellite-signal attenuation. Though the calibration data we analyzed were not collected for this task, we can report RMSE_{min} values between 1.7 and 21 meters (Table 2), with the smaller location errors likely being the result of dual-frequency GPS receivers and multiple location fixes processed by an on-board Kálmán filter. Specifically, devices may condition the current location estimate on past data, under an assumed movement and location-error model, possibly also leveraging Doppler-shift information. Some devices filter all location estimates, keeping the constituent data unreported at coarse sampling rates, while other devices only apply filtering when the sampling rate reaches a specific threshold ($\sim$1 Hz). On this point we also caution researchers that devices with on-board filtering may have smaller RMS UERE parameters at higher fix rates, and, as such devices warm up, these parameters can continue to shrink until they reach an asymptote.

For 22% (9–42%, population CI) of tracking device models, their recorded DOP values were misinformative, in that they provided worse information on location error than a simple model of homoskedastic errors, and so using them could actually have a deleterious effect on analysis. Because informative DOP values are necessary to account for the many sources

of heteroskedasticity in tracking data, they can be important in many situations, such as if animals traverse through habitats of differing satellite reception. For instance, animals that burrow or rest under trees might appear to be scattering about ∼100 meters when not moving, without informative DOP values to explain the increase in variance. In these cases, researchers should attempt to collect calibration data in different environments, as an approximate correction, but such calibration data are unlikely to exactly reproduce the same quality location estimates as the animal, and habitats may differ between the location estimate and the true location.

Finally, we note that *none* (0%–52%) of the 25 GPS tracking devices we analyzed had particularly informative VDOP values, where the null model based on VDOP outperformed models using HDOP values or our number-of-satellites model (S3.1) as proxies for vertical error, or even a homoskedastic model. Informative VDOP values are necessary for an accurate assessment of vertical errors, and 12 of the devices we tested were GPS tags, rather than collars, which would be applied more often to aerial or arboreal species. These data are increasingly needed in good quality, considering human impacts in the airspace, such as wind farms, buildings, planes, power lines, etc. (Lambertucci et al., 2015; Ross-Smith et al., 2016; Péron et al., 2017; Poessel et al., 2018*a*).

Error model performance

Particularly with older Argos Doppler-shift location data, which only feature location class information (3,2,1,0,A,B,Z), several analyses have employed t-distributed location errors, as the error distributions within location classes are heavy tailed (Jonsen et al., 2005; Hoenner et al., 2012; Albertsen et al., 2015; Brost et al., 2015). Error models that fail to capture the heteroskedasticity of satellite location data will generally be heavier tailed (c.f., Fig. S4.2). This holds for both historical Argos Doppler-shift location data and GPS location data with inadequate error-model performance (e.g., data that lack informative DOP values). While t-distributed errors are readily available in `R` package Template Model Builder (`TMB`; Albertsen et al., 2015; Péron et al., 2017; Auger-Méthé et al., 2019), heavy-tailed error distributions eventually need to be incorporated into the more user-friendly `R` packages, such as `crawl` (Johnson et al., 2008; Johnson, 2008) and `ctmm` (Fleming and Calabrese, 2015; Calabrese et al., 2016). As pointed out by Albertsen et al. (2015), there are straightforward and reliable approximations for heavy tailed error distributions implemented in `TMB`, which can also be implemented in similar frameworks like `crawl` and `ctmm`.

On the other hand, technological advances, such as solar-powered GPS tags, now allow very high sampling rates to be sustained for long periods of time (App. S4.3.2). These new data introduce several complications when it is necessary to account for location errors. First, as the sampling interval falls below 1–2 minutes, GPS location errors become autocorrelated (Tao and Bonnifait, 2015). Second, many GPS devices employ on-board filters to post-process high-frequency location estimates and shrink their errors, which comes at the cost of increased location-error autocorrelation. As these data become increasingly common, the calibration models and estimators introduced here will need to be upgraded from independently sampled errors to autocorrelated errors. The accommodation of location-error autocorrelation may also be necessary for VPS data (App. S4.14.1), though further testing is warranted. Generalizing the methods introduced here to incorporate autocorrelated location errors is a largely straightforward, though tedious, computational task.

When using the independent and identically distributed (IID) UERE methods presented here on high-resolution GPS tracking data, we can approximate the resulting biases as follows. If Δt is the sampling interval, then effective sample sizes are overestimated by an inflation factor on the order of $I \equiv (1 \text{ min})/\Delta t$. AIC_C differences will be overestimated by a factor on the order of I, which will overstate differences in model performance if not accounted for. UERE variance parameters will be relatively underestimated by a factor on the order of I/N, where N is the nominal sample size assuming independence. However, this bias is quite minor in calibration datasets that sample for an hour or more. In any case, we recommend collecting calibration data over the course of a day, if not many days.

Simultaneous estimation as a last resort

The simulation example considered by Auger-Méthé et al. (2016) represents a more extreme case of lack of identifiability between relatively featureless, discrete-time movement and error models, though model misspecification can produce worse outcomes. If estimating movement and location error simultaneously, it is important to enforce as much distinguishing structure as possible on the two processes, to increase identifiability in their parameters. One example of distinguishing structure is the heteroskedasticity of location error, as captured by informative DOP values. It is unlikely that an individual's finescale displacements will scale with DOP values as well as location error, and therefore homoskedastic errors, as simulated in Auger-Méthé et al. (2016), are much more vulnerable to issues of identifiability than heteroskedastic errors (Fig. 4). Another feature that distinguishes the movement and error processes are their contrasting autocorrelation structures. GPS location errors are typically not substantially autocorrelated at sampling intervals > 1-2 minutes (Tao and Bonnifait, 2015), while the movement processes of large mammals are typically correlated for days to weeks (McNay et al., 1994; Rooney et al., 1998; Boyce et al., 2010; Fleming et al., 2014*a*,*b*; Morato et al., 2016). Because of this separation of scales, irregular sampling is helpful for teasing apart movement from error, given an appropriate continuous-time movement model. In contrast, shoehorning irregularly sampled data into a discrete-time framework will result in a loss of this useful information.

When forced to fit movement- and error-model parameters simultaneously, we suggest two checks for validity. First, the error-parameter estimates should be plausible. For example, the RMS UERE should be around 10 meters for GPS telemetry. Second, the autocorrelation structure of the data should be adequately explained by the combined model of movement and error. Variograms are particularly useful for this task, because they provide unbiased estimators of autocorrelation (Cressie, 1993; Fleming et al., 2014*a*). As the autocorrelation of tracking data consists of both movement and location-error autocorrelation, and as location errors are typically heteroskedastic, we derive error-informed variogram algorithms in App. S6 (Fig. S6.1).

Error-informed analyses

The number of movement analyses based on continuous-time state-space models, which can allow for both irregular sampling and location error, is steadily expanding (e.g., Breed et al., 2017; Michelot and Blackwell, 2019). However, there remain several analytic tasks that have yet to be formulated in a way that accounts for location error. Primary among these tasks are resource selection and home-range estimation. Resource selection with uncertain

location data is a straightforward challenge that can be addressed by incorporating an observation model. Non-parametric home-range estimation is less obvious. The more general problem of kernel density estimation with measurement error is conventionally approached by 'deconvolving' or de-smoothing the kernels according to the spread of the error (Stefanski and Carroll, 1990). However, it is straightforward to show that this approach will not generally produce an asymptotically consistent estimator, meaning that the estimate will not converge to the truth in the limit of infinite data, as the individual kernels remain biased in their spread. The method we applied here, which involves Gaussian smoothing of the data (Fleming et al., 2016) and error-informed bandwidth optimization, is relatively ad hoc but is ensured to produce a density estimate of appropriate scale. Future home-range estimators will need to both optimally weight the data, in the sense of Fleming et al. (2018), so that more erroneous location estimates are treated as less informative, and smooth the data conditional on the kernel density estimate, rather than on a normal approximation of the density function. The latter is needed to shrink the estimate in towards the data, rather than, simply, the mean of the data. Finally, we note that the error-informed statistics for outlier detection introduced in App. S5 assume circular error distributions and could be improved for Argos Doppler-shift, triangulated VHF, and light-level geolocator-derived data.

User-friendly `R` software packages that can incorporate the kind of heteroskedastic location errors we model here include `crawl` (Johnson et al., 2008; Johnson, 2008), `ctmm` (Fleming and Calabrese, 2015; Calabrese et al., 2016), `move` (Kranstauber et al., 2012), `momentuHMM` (McClintock and Michelot, 2018), and `smoove` (Gurarie et al., 2017). The methods and models we have introduced are implemented in the `R` package `ctmm`, complete with long-form documentation `vignette("error")`. Based on the calibration results presented here (Apps. S2-S4), error model selection is obviated in `ctmm` for Argos Doppler-shift, e-obs, Telonics, Vemco, and many other device models. Going forward, we envision both error modeling and error-informed animal movement analysis becoming more necessary, more user friendly, and more accurate.

Final remarks

GPS location error has historically been ignored because the magnitude of these errors have been negligible, relative to the scales of movement. However, improvements in the temporal resolution of tracking data, device miniaturization for smaller taxa (Kays et al., 2015), and an increase in multi-study analyses (Tucker et al., 2018, 2019; Noonan et al., 2020*b*) all escalate the need for reliable statistical methods that can account for location error. We have presented a major step towards achieving this goal, by deriving a comprehensive framework—from experimental design to error-model selection, error-model performance evaluation, and error-informed analyses—with the capability to pin down the observation model of animal tracking data and tame its heteroskedastic errors. Our framework allows tracking data to more directly and more accurately inform ecological analyses, now and going forward. This work should provide strong guidelines for both researchers and manufacturers in wildlife ecology and conservation biology, as well as other fields where data represent processes of interest and error processes that are difficult to tease apart.

Acknowledgments

We thank T. E. Katzner for his many comments on and contributions to this manuscript. CHF, JMC, MJN, and WFF were supported by NSF ABI 1458748. CHF, JMC, and WFF were supported by NSF IIBR 1915347. MMC and RAM were supported by MN DNR SWG F14AP00028. NPG and CSD were supported by the Pittman-Robertson Federal Aid to Wildlife Restoration Grant, NCWRC and the NCSU Fisheries, Wildlife, and Conservation Biology program. RK was supported by NSF IIBR 1914928. KNP was supported by the Winifred Violet Scott Charitable Trust. AR and RS were supported by SNSF grant 173178. KS and CMT were supported by Laura Kearns, the Ottawa National Wildlife Refuge, the Winous Point Marsh Conservancy, and the ODNR Division of Wildlife. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Abernathy, H. N., D. A. Crawford, E. P. Garrison, R. B. Chandler, M. L. Conner, K. V. Miller, and M. J. Cherry. 2019. Deer movement and resource selection during hurricane irma: implications for extreme climatic events and wildlife. Proceedings of the Royal Society B: Biological Sciences 286:20192230.

Afonso, A. S., R. Garla, and F. H. V. Hazin. 2017. Tiger sharks can connect equatorial habitats and fisheries across the Atlantic Ocean basin. PLOS ONE 12:1–15.

Albertsen, C. M., K. Whoriskey, D. Yurkowski, A. Nielsen, and J. M. Flemming. 2015. Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder. Ecology 96:2598–2604.

Auger-Méthé, M., C. Field, C. M. Albertsen, A. E. Derocher, M. A. Lewis, I. D. Jonsen, and J. M. Flemming. 2016. State-space models' dirty little secrets: even simple linear Gaussian models can have estimation problems. Scientific Reports 6.

Auger-Méthé, M., C. M. Albertsen, I. D. Jonsen, A. E. Derocher, D. C. Lidgard, K. R. Studholme, W. D. Bowen, G. T. Crossin, and J. M. Flemming. 2019. Spatiotemporal modelling of marine movement data using Template Model Builder (TMB). Marine Ecology Progress Series 565:237–249.

Berg, S. S. 2015. The package "sigloc" for the R software: A tool for triangulating transmitter locations in ground-based telemetry studies of wildlife populations. The Bulletin of the Ecological Society of America 96:500–507.

Bestley, S., I. D. Jonsen, M. A. Hindell, R. G. Harcourt, and N. J. Gales. 2015. Taking animal tracking to new depths: synthesizing horizontal–vertical movement relationships for four marine predators. Ecology 96:417–427.

Bjørneraas, K., B. Van Moorter, C. M. Rolandsen, and I. Herfindal. 2010. Screening global positioning system location data for errors using animal movement characteristics. The Journal of Wildlife Management 74:1361–1366.

Blanco, G. S., S. J. Morreale, H. Bailey, J. A. Seminoff, F. V. Paladino, and J. R. Spotila. 2012. Post-nesting movements and feeding grounds of a resident East Pacific green turtle Chelonia mydas population from Costa Rica. Endangered Species Research 18:233–245.

Boyce, M. S., J. Pitt, J. M. Northrup, A. T. Morehouse, K. H. Knopff, B. Cristescu, and G. B. Stenhouse. 2010. Temporal autocorrelation functions for movement rates from global

positioning system radiotelemetry data. Philosophical Transactions of the Royal Society B: Biological Sciences 365:2213–2219.

Breed, G. A., E. A. Golson, and M. T. Tinker. 2017. Predicting animal home-range structure and transitions using a multistate Ornstein-Uhlenbeck biased random walk. Ecology 98:32–47.

Brost, B. M., M. B. Hooten, E. M. Hanks, and R. J. Small. 2015. Animal movement constraints improve resource selection inference in the presence of telemetry error. Ecology 96:2590–2597.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference. 2nd ed. Springer, New York.

Calabrese, J. M., C. H. Fleming, W. F. Fagan, M. Rimmler, P. Kaczensky, S. Bewick, P. Leimgruber, and T. Mueller. 2018. Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. Philosophical Transactions of the Royal Society of London B: Biological Sciences 373.

Calabrese, J. M., C. H. Fleming, and E. Gurarie. 2016. ctmm: An R package for analyzing animal relocation data as a continuous-time stochastic process. Methods in Ecology and Evolution 7:1124–1132.

Carman, V. G., V. Falabella, S. Maxwell, D. Albareda, C. Campagna, and H. Mianzan. 2012. Revisiting the ontogenetic shift paradigm: The case of juvenile green turtles in the SW Atlantic. Journal of Experimental Marine Biology and Ecology 429:64–72.

Citta, J. J., L. F. Lowry, L. T. Quakenbush, B. P. Kelly, A. S. Fischbach, J. M. London, C. V. Jay, K. J. Frost, G. O'Corry-Crowe, J. A. Crawford, P. L. Boveng, M. Cameron, A. L. V. Duyke, M. Nelson, L. A. Harwood, P. Richard, R. Suydam, M. P. Heide-Jørgensen, R. C. Hobbs, D. I. Litovka, M. Marcoux, A. Whiting, A. S. Kennedy, J. C. George, J. Orr, and T. Gray. 2018. A multi-species synthesis of satellite telemetry data in the Pacific Arctic (1987–2015): Overlap of marine mammal distributions and core use areas. Deep Sea Research Part II: Topical Studies in Oceanography 152:132–153.

CLS. 2016. Argos User's Manual: Worldwide tracking and environmental monitoring by satellite. http://http://www.argos-system.org/manual/.

Coster, A., and A. Komjathy. 2008. Space weather and the Global Positioning System. Space Weather 6:S06D04.

Cressie, N. 1993. Statistics for spatial data. revised ed. Wiley, New York.

Dalleau, M., S. Benhamou, J. Sudre, S. Ciccione, and J. Bourjea. 2014. The spatial ecology of juvenile loggerhead turtles (Caretta caretta) in the Indian Ocean sheds light on the "lost years" mystery. Marine Biology 161:1835–1849.

Douglas, D. C., R. Weinzierl, S. C. Davidson, R. Kays, M. Wikelski, and G. Bohrer. 2012. Moderating argos location errors in animal tracking data. Methods in Ecology and Evolution 3:999–1007.

Ewald, M., C. Dupke, M. Heurich, J. Müller, and R. B. 2014. LiDAR remote sensing of forest structure and GPS telemetry data provide insights on winter habitat selection of european roe deer. Forests 5:1374–1390.

Fleming, C. H., T. Akre, D. J. Brown, M. M. Cochrane, N. Dejid, V. DeNicola, C. S. DePerno, J. Dresher-Lehman, N. P. Gould, J. Hollins, H. Ishii, Y. Kaneko, T. E. Katzner, S. Killen, B. Koeck, S. A. Lambertucci, S. D. LaPoint, E. P. Medici, B.-U. Meyburg, T. A. Miller, R. A. Moen, T. Mueller, T. Pfeiffer, A. Roulin, K. Safi, A. K. Scharf, R. Séchaud,

J. A. Stabach, and K. Yamazaki. 2020. GPS calibration data (global). Movebank ID: 1092737859.

Fleming, C. H., and J. M. Calabrese. 2015. ctmm: Continuous-Time Movement Modeling. https://CRAN.R-project.org/package=ctmm.

———. 2017. A new kernel-density estimator for accurate home-range and species-range area estimation. Methods in Ecology and Evolution 8:571–579.

Fleming, C. H., J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan. 2014*a*. From fine-scale foraging to home ranges: A semi-variance approach to identifying movement modes across spatiotemporal scales. The American Naturalist 183:E154–E167.

———. 2014*b*. Non-Markovian maximum likelihood estimation of autocorrelated movement processes. Methods in Ecology and Evolution 5:462–472.

Fleming, C. H., W. F. Fagan, T. Mueller, K. A. Olson, P. Leimgruber, and J. M. Calabrese. 2015. Rigorous home-range estimation with movement data: A new autocorrelated kernel-density estimator. Ecology 96:1182–1188.

———. 2016. Estimating where and how animals travel: An optimal framework for path reconstruction from autocorrelated tracking data. Ecology 97.

Fleming, C. H., M. J. Noonan, E. P. Medici, and J. M. Calabrese. 2019. Overcoming the chalenge of small effective sample sizes in home-range estimation. Methods in Ecology and Evolution 10:1679–1689.

Fleming, C. H., D. Sheldon, W. F. Fagan, P. Leimgruber, T. Mueller, D. Nandintsetseg, M. J. Noonan, K. A. Olson, E. Setyawan, A. Sianipar, and J. M. Calabrese. 2018. Correcting for missing and irregular data in home-range estimation. Ecological Applications 28:1003–1010.

Fleming, C. H., D. Sheldon, E. Gurarie, W. F. Fagan, and J. M. Calabrese. 2017. Kálmán filters for continuous-time movement models. Ecological Informatics 40:8–21.

Gerber, B. D., M. B. Hooten, C. P. Peck, M. B. Rice, J. H. Gammonley, A. D. Apa, and A. J. Davis. 2018. Accounting for location uncertainty in azimuthal telemetry data improves ecological inference. Movement Ecology 6.

Gurarie, E., C. H. Fleming, K. L. Laidre, J. Hernandez-Pliego, W. F. Fagan, and O. Ovaskainen. 2017. Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. Movement Ecology 5.

Hückstädt, L. A., R. A. Quiñones, M. Sepúlveda, and D. P. Costa. 2014. Movement and diving patterns of juvenile male South American sea lions off the coast of central Chile. Marine Mammal Science 30:1175–1183.

Hilborn, R., and M. Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press, Princeton, New Jersey, USA.

Hoenner, X., S. D. Whiting, M. A. Hindell, and C. R. McMahon. 2012. Enhancing the use of Argos satellite data for home range and long distance migration studies of marine animals. PLOS ONE 7:e40713.

Horne, J. S., E. O. Garton, S. M. Krone, and J. S. Lewis. 2007. Analyzing animal movements using Brownian bridges. Ecology 88:2354–63.

Howell, E. A., P. H. Dutton, J. J. Polovina, H. Bailey, D. M. Parker, and G. H. Balazs. 2010. Oceanographic influences on the dive behavior of juvenile loggerhead turtles (Caretta caretta) in the North Pacific Ocean. Marine Biology 157:1011–1026.

Ishii, H., K. Yamazaki, M. J. Noonan, C. D. Buesching, C. Newman, and Y. Kaneko. 2019.

Testing cellular phone-enhanced GPS tracking technology for urban carnivores. Animal Biotelemetry 7.

Johnson, D. S. 2008. crawl: Fit Continuous-Time Correlated Random Walk Models to Animal Movement Data. https://CRAN.R-project.org/package=crawl.

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. 2007. Identifying leatherback turtle foraging behaviour from satellite telemetry using a switching state-space model. Marine Ecology Progress Series 337:255–264.

Kaplan, E. D., and C. J. Hegarty. 2006. Understanding GPS: Principles and applications. 2nd ed. Artech House, Norwood, MA, USA.

Kays, R., M. C. Crofoot, W. Jetz, and M. Wikelski. 2015. Terrestrial animal tracking as an eye on life and planet. Science 348.

Kennedy, A. S., A. N. Zerbini, B. K. Rone, and P. J. Clapham. 2014. Individual variation in movements of satellite-tracked humpback whales Megaptera novaeangliae in the eastern Aleutian Islands and Bering Sea. Endangered Species Research 23:187–195.

Kranstauber, B., M. Smolla, and A. K. Scharf. 2012. move: Visualizing and Analyzing Animal Track Data. https://CRAN.R-project.org/package=move.

Kreye, C., B. Eissfeller, and G. Ameres. 2004. Architectures of GNSS/INS integrations: Theoretical approach and practical tests. Symposium on Gyro Technology pages 1–16.

Kuhn, C. E., D. S. Johnson, R. R. Ream, and T. S. Gelatt. 2009. Advances in the tracking of marine species: using GPS locations to evaluate satellite track data and a continuous-time movement model. Marine Ecology Progress Series 393:97–109.

Lambertucci, S. A., E. L. C. Shepard, and R. P. Wilson. 2015. Human-wildlife conflicts in a crowded airspace. Science 348:502–504.

Lehmann, E. L., and H. Scheffé. 1950. Completeness, similar regions, and unbiased estimation: Part i. Sankhyā: The Indian Journal of Statistics (1933-1960) 10:305–340.

———. 1955. Completeness, similar regions, and unbiased estimation: Part ii. Sankhyā: The Indian Journal of Statistics (1933-1960) 15:219–236.

Lennox, R. J., K. Aarestrup, S. J. Cooke, P. D. Cowley, Z. D. Deng, A. T. Fisk, R. G. Harcourt, M. Heupel, S. G. Hinch, K. N. Holland, N. E. Hussey, S. J. Iverson, S. T. Kessel, J. F. Kocik, M. C. Lucas, J. M. Flemming, V. M. Nguyen, M. J. W. Stokesbury, S. Vagle, D. L. VanderZwaag, F. G. Whoriskey, and N. Young. 2017. Envisioning the Future of Aquatic Animal Tracking: Technology, Science, and Application. BioScience 67:884–896.

Lenth, R. V. 1981. On finding the source of a signal. Technometrics 23:149–154.

Lopez, R., J.-P. Malardé, F. Royer, and P. Gaspar. 2014. Improving argos doppler location using multiple-model kalman filtering. IEEE transactions on geoscience and remote sensing 52:4744–4755.

Marcotte, D. 1996. Fast variogram computation with FFT. Computers & Geosciences 22:1175–1186.

Martinez-Garcia, R., C. H. Fleming, R. Seppelt, W. F. Fagan, and J. M. Calabrese. 2020. How range residency and long-range perception change encounter rates. Journal of Theo-

retical Biology page 110267.

Martínez Hernández, L. A., S. Pérez Arteaga, G. Sánchez Pérez, A. L. Sandoval Orozco, and L. J. García Villalba. 2019. Outdoor location of mobile devices using trilateration algorithms for emergency services. IEEE Access 7:52052–52059.

Matthews, C. J. D., S. P. Luque, S. D. Petersen, R. D. Andrews, and S. H. Ferguson. 2011. Satellite tracking of a killer whale (Orcinus orca) in the eastern Canadian Arctic documents ice avoidance and rapid, long-distance movement into the North Atlantic. Polar Biology 34:1091–1096.

McClintock, B. T., J. M. London, M. F. Cameron, P. L. Boveng, and O. Gimenez. 2014. Modelling animal movement using the Argos satellite telemetry location error ellipse. Methods in Ecology and Evolution 6:266–277.

McClintock, B. T., and T. Michelot. 2018. momentuHMM: R package for generalized hidden Markov models of animal movement. Methods in Ecology and Evolution 9:1518–1530.

McKinnon, E. A., and O. P. Love. 2018. Ten years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging. The Auk 135:834–856.

McNay, R. S., J. A. Morgan, and F. L. Bunnell. 1994. Characterizing independence of observations in movements of Columbian black-tailed deer. The Journal of Wildlife Management 58:422–429.

Meckley, T. D., C. M. Holbrook, C. M. Wagner, and T. R. Binder. 2014. An approach for filtering hyperbolically positioned underwater acoustic telemetry data with position precision estimates. Animal Biotelemetry 2.

Michelot, T., and P. G. Blackwell. 2019. State-switching continuous-time correlated random walks. Methods in Ecology and Evolution 10:637–649.

Morato, R. G., J. A. Stabach, C. H. Fleming, J. M. Calabrese, R. C. De Paula, K. M. P. M. Ferraz, D. L. Z. Kantek, S. S. Miyazaki, T. D. C. Pereira, G. R. Araujo, A. Paviolo, C. De Angelo, M. S. Di Bitetti, P. Cruz, F. Lima, L. Cullen, D. A. Sana, E. E. Ramalho, M. M. Carvalho, F. H. S. Soares, B. Zimbres, M. X. Silva, M. D. F. Moraes, A. Vogliotti, J. A. May, Jr., M. Haberfeld, L. Rampim, L. Sartorello, M. C. Ribeiro, and P. Leimgruber. 2016. Space use and movement of a neotropical top predator: The endangered jaguar. PLOS ONE 11:1–17.

Noonan, M. J., C. H. Fleming, T. Akre, J. Drescher-Lehman, E. Gurarie, A.-L. Harrison, R. Kays, and J. M. Calabrese. 2019. Scale-insensitive estimation of speed and distance traveled from animal tracking data. Movement Ecology 7.

Noonan, M. J., C. H. Fleming, R. Martinez-Garcia, M. C. Crofoot, G. Davis, R. Kays, M. Michelangeli, E. Payne, A. Sih, O. Spiegel, W. F. Fagan, and J. M. Calabrese. 2020*a*. Predicting encounter locations from tracking data *in preparation*.

Noonan, M. J., C. H. Fleming, M. A. Tucker, R. Kays, A.-L. Harrison, M. C. Crofoot, B. Abrahmsg, S. C. Albertsh, A. H. Ali, J. Altmann, P. C. Antunes, N. Attias, J. L. Belant, D. E. Beyer Jr., L. R. Bidner, N. Blaum, R. B. Boone, D. Caillaud, R. C. de Paula, J. A. de la Torre, J. Dekker, C. S. DePerno, M. Farhadinia, J. Fennessy, C. Fichtel, C. Fischer, A. Ford, J. R. Goheen, R. W. Havmøller, B. T. Hirsch, C. Hurtado, L. A. Isbell, R. Janssen, F. Jeltsch, P. Kaczensky, Y. Kaneko, P. Kappeler, A. Katna, M. Kauffman, F. Koch, A. Kulkarni, S. LaPoint, P. Leimgruber, D. W. Macdonald, A. C. Markham, L. McMahon, K. Mertes, C. E. Moorman, R. G. Morato, A. M. Moßbrucker, G. Mourão, D. O'Connor, L. G. R. Oliveira-Santos, J. Pastorini, B. D. Patterson, J. Rachlow, D. H. Ranglack,

N. Reid, D. M. Scantlebury, D. M. Scott, N. Selva, A. Sergiel, M. Songer, N. Songsasen, J. A. Stabach, J. Stacy-Dawes, M. B. Swingen, J. J. Thompson, W. Ullmann, A. T. Vanak, M. Thaker, J. W. Wilson, K. Yamazaki, R. W. Yarnell, F. Zięba, T. Zwijacz-Kozica, W. F. Fagan, T. Mueller, and J. M. Calabrese. 2020*b*. Body-size-dependent underestimation of mammalian area requirements. Conservation Biology *in press*.

Noonan, M. J., A. Markham, C. Newman, N. Trigoni, C. D. Buesching, S. A. Ellwood, and D. W. Macdonald. 2015. A new magneto-inductive tracking technique to uncover subterranean activity: what do animals do underground? Methods in Ecology and Evolution 6:510–520.

Noonan, M. J., C. Newman, A. Markham, K. Bilham, C. D. Buesching, and D. W. Macdonald. 2018. In situ behavioral plasticity as compensation for weather variability: implications for future climate change. Climatic Change 149:457–471.

Parkinson, B. W., and P. K. Enge. 1996. Differential GPS. Pages 3–50 *in* Global positioning system: Theory and applications. Vol. 2. American Institute of Aeronautics and Astronautics, Inc., Washington, DC.

Péron, G., J. M. Calabrese, O. Duriez, C. H. Fleming, A. Johnston, S. Lambertucci, K. Safi, and E. L. C. Shepard. 2020. The challenges of estimating the distribution of flight heights from telemetry or altimetry data. Animal Biotelemetry 8.

Péron, G., C. H. Fleming, R. C. de Paula, and J. M. Calabrese. 2016. Uncovering periodic patterns of space use in animal tracking data with periodograms, including a new algorithm for the Lomb-Scargle periodogram and improved randomization tests. Movement Ecology 4.

Péron, G., C. H. Fleming, O. Duriez, J. Fluhr, C. Itty, S. Lambertucci, K. Safi, E. L. C. Shepard, J. M. Calabrese, and S. Bauer. 2017. The energy landscape predicts flight height and wind turbine collision hazard in three species of large soaring raptor. Journal of Applied Ecology 54:1895–1906.

Phalan, B., R. A. Phillips, J. R. D. Silk, V. Afanasyev, A. Fukuda, J. Fox, P. Catry, H. Higuchi, and J. P. Croxall. 2007. Foraging behaviour of four albatross species by night and day. Marine Ecology Progress Series 340.

Poessel, S. A., J. Brandt, L. Mendenhall, M. A. Braham, M. J. Lanzone, A. J. McGann, and T. E. Katzner. 2018*a*. Flight response to spatial and temporal correlates informs risk from wind turbines to the California Condor. The Condor 120:330–342.

Poessel, S. A., A. E. Duerr, J. C. Hall, M. A. Braham, and T. E. Katzner. 2018*b*. Improving estimation of flight altitude in wildlife telemetry studies. Journal of Applied Ecology 55:2064–2070.

Ranacher, P., R. Brunauer, W. Trutschnig, S. V. der Spek, and S. Reich. 2016. Why GPS makes distances bigger than they are. International Journal of Geographical Information Science 30:316–333.

Raymond, B., M.-A. Lea, T. Patterson, V. Andrews-Goff, R. Sharples, J.-B. Charrassin, M. Cottin, L. Emmerson, N. Gales, R. Gales, S. D. Goldsworthy, R. Harcourt, A. Kato, R. Kirkwood, K. Lawton, Y. Ropert-Coudert, C. Southwell, J. van den Hoff, B. Wienecke, E. J. Woehler, S. Wotherspoon, and M. A. Hindell. 2015. Important marine habitat off east antarctica revealed by two decades of multi-species predator tracking. Ecography 38:121–129.

Rooney, S. M., A. Wolfe, and T. J. Hayden. 1998. Autocorrelated data in telemetry studies:

Time to independence and the problem of behavioural effects. Mammal Review 28:89–98.
Ross-Smith, V. H., C. B. Thaxter, E. A. Masden, J. Shamoun-Baranes, N. H. K. Burton, L. J. Wright, M. M. Rehfisch, A. Johnston, and D. Thompson. 2016. Modelling flight heights of lesser black-backed gulls and great skuas from gps: a bayesian approach. Journal of Applied Ecology 53:1676–1685.
Rowcliffe, J. M., C. Carbone, R. Kays, B. Kranstauber, and P. A. Jansen. 2012. Bias in estimating animal travel distance: The effect of sampling frequency. Methods in Ecology and Evolution 3:653–662.
Safi, K., and B. Kranstauber. 2021. Analysis and Mapping of Animal Movement in R. Chapman & Hall. *in press*.
Smith, F. 2013. Understanind HPE in the VEMCO Positioning System (VPS) .
Stefanski, L. A., and R. J. Carroll. 1990. Deconvolving kernel density estimators. Statistics 21:169–184.
Strandburg-Peshkin, A., D. R. Farine, I. D. Couzin, and M. C. Crofoot. 2015. Shared decision-making drives collective movement in wild baboons. Science 348:1358–1361.
Sulikowski, J. A., B. Galuardi, W. Bubley, N. B. Furey, W. B. I. Driggers, G. W. J. Ingram, and P. C. W. Tsang. 2010. Use of satellite tags to reveal the movements of spiny dogfish Squalus acanthias in the western North Atlantic Ocean. Marine Ecology Progress Series 418:249–254.
Swimmer, Y., L. McNaughton, D. Foley, L. Moxey, and A. Nielsen. 2009. Movements of olive ridley sea turtles Lepidochelys olivacea and associated oceanographic features as determined by improved light-based geolocation. Endangered Species Research 10:245–254.
Tao, Z., and P. Bonnifait. 2015. Modeling L1-GPS errors for an enhanced data fusion with lane marking maps for road automated vehicles. *In* European Navigation Conference (ENC 2015). Bordeaux, France.
Thompson, W. R. 1935. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. The Annals of Mathematical Statistics 6:214–219.
Tucker, M. A., O. Alexandrou, R. O. Bierregaard Jr., K. L. Bildstein, K. Böhning-Gaese, C. Bracis, J. N. Brzorad, E. R. Buechley, D. Cabot, J. M. Calabrese, C. Carrapato, A. Chiaradia, L. C. Davenport, S. C. Davidson, M. Desholm, C. R. DeSorbo, R. Domenech, P. Enggist, W. F. Fagan, N. Farwig, W. Fiedler, C. H. Fleming, A. Franke, J. M. Fryxell, C. García-Ripollés, D. Grémillet, L. R. Griffin, R. Harel, A. Kane, R. Kays, E. Kleyheeg, A. E. Lacy, S. LaPoint, R. Limiñana, P. López-López, A. D. Maccarone, U. Mellone, E. K. Mojica, R. Nathan, S. H. Newman, M. J. Noonan, S. Oppel, M. Prostor, E. C. Rees, Y. Ropert-Coudert, S. Rösner, N. Sapir, D. Schabo, M. Schmidt, H. Schulz, M. Shariati, A. Shreading, J. Paulo Silva, H. Skov, O. Spiegel, J. Y. Takekawa, C. S. Teitelbaum, M. L. van Toor, V. Urios, J. Vidal-Mateo, Q. Wang, B. D. Watts, M. Wikelski, K. Wolter, R. Žydelis, and T. Mueller. 2019. Large birds travel farther in homogeneous environments. Global Ecology and Biogeography 28:576–587.
Tucker, M. A., K. Böhning-Gaese, W. F. Fagan, J. M. Fryxell, B. Van Moorter, S. C. Alberts, A. H. Ali, A. M. Allen, N. Attias, T. Avgar, H. Bartlam-Brooks, B. Bayarbaatar, J. L. Belant, A. Bertassoni, D. Beyer, L. Bidner, F. M. van Beest, S. Blake, N. Blaum, C. Bracis, D. Brown, P. J. N. de Bruyn, F. Cagnacci, J. M. Calabrese, C. Camilo-Alves,

S. Chamaillé-Jammes, A. Chiaradia, S. C. Davidson, T. Dennis, S. DeStefano, D. Diefenbach, I. Douglas-Hamilton, J. Fennessy, C. Fichtel, W. Fiedler, C. Fischer, I. Fischhoff, C. H. Fleming, A. T. Ford, S. A. Fritz, B. Gehr, J. R. Goheen, E. Gurarie, M. Hebblewhite, M. Heurich, A. J. M. Hewison, C. Hof, E. Hurme, L. A. Isbell, R. Janssen, F. Jeltsch, P. Kaczensky, A. Kane, P. M. Kappeler, M. Kauffman, R. Kays, D. Kimuyu, F. Koch, B. Kranstauber, S. LaPoint, P. Leimgruber, J. D. C. Linnell, P. López-López, A. C. Markham, J. Mattisson, E. P. Medici, U. Mellone, E. Merrill, G. de Miranda Mourão, R. G. Morato, N. Morellet, T. A. Morrison, S. L. Díaz-Muñoz, A. Mysterud, D. Nandintsetseg, R. Nathan, A. Niamir, J. Odden, R. B. O'Hara, L. G. R. Oliveira-Santos, K. A. Olson, B. D. Patterson, R. Cunha de Paula, L. Pedrotti, B. Reineking, M. Rimmler, T. L. Rogers, C. M. Rolandsen, C. S. Rosenberry, D. I. Rubenstein, K. Safi, S. Saïd, N. Sapir, H. Sawyer, N. M. Schmidt, N. Selva, A. Sergiel, E. Shiilegdamba, J. P. Silva, N. Singh, E. J. Solberg, O. Spiegel, O. Strand, S. Sundaresan, W. Ullmann, U. Voigt, J. Wall, D. Wattles, M. Wikelski, C. C. Wilmers, J. W. Wilson, G. Wittemyer, F. Zięba, T. Zwijacz-Kozica, and T. Mueller. 2018. Moving in the anthropocene: Global reductions in terrestrial mammalian movements. Science 359:466–469.

Vermard, Y., E. Rivot, S. Mahévas, P. Marchal, and D. Gascuel. 2010. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian hidden markov models. Ecological Modelling 221:1757–1769.

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.

Williams, T. M., L. Wolfe, T. Davis, T. Kendall, B. Richter, Y. Wang, C. Bryce, G. H. Elkaim, and C. C. Wilmers. 2014. Instantaneous energetics of puma kills reveal advantage of felid sneak attacks. Science 346:81–85.

Winner, K., M. J. Noonan, C. H. Fleming, K. Olson, T. Mueller, D. Sheldon, and J. M. Calabrese. 2018. Statistical inference for home range overlap. Methods in Ecology and Evolution 9:1679–1691.

Yahya, M. H., and M. N. Kamarudin. 2008. Analysis of GPS visibility and satellite-receiver geometry over different latitudinal regions. *In* International Symposium on Geoinformation.

Yang, Y. 2005. Can the strengths of AIC and BIC be shared? a conflict between model indentification and regression estimation. Biometrika 92:937–950.

Argos	Satellite system for geolocation, based on the Doppler effect with satellites in polar orbits. This system is often used to upload GPS location data.
DOP	'Dilution of precision', a relative measure of RMS location error.
GDOP	'Geometric dilution of precision', an aggregate of PDOP and TDOP.
GNSS	'Global navigation satellite system', a catch-all term for satellite navigation systems, including America's GPS, Russia's GLONASS, China's BDS, and India's NAVIC.
GPS	'Global Positioning System' for geolocation, based on trilateration with satellites in medium earth orbits.
GSM	'Global System for Mobile Communications' standard for cellular networks that provides information useful for location estimation. This system is often used to upload GPS location data.
HDOP	'Horizontal dilution of precision', a relative measure of RMS horizontal location error.
HPE	'Horizontal position error', which is only provided as a relative measure of RMS horizontal location error in VPS location data.
Kálmán filter	A state-space method for conditioning on past data to improve the current location estimate. This method can be used in real time and is sometimes employed on GPS and GSM devices.
Kálmán smoother	A state-space method for conditioning on all data to improve location estimates.
multipath error	Location-estimate errors caused by indirect signals as the result of reflection or diffraction.
PDOP	'Position dilution of precision', an aggregate of HDOP and VDOP.
RMS	'Root mean square' statistic, which is equivalent to the standard deviation in one dimension.
TDOP	'Time dilution of precision', a relative measure of RMS temporal error.
UERE	'User equivalent range error', which is location error partially standardized by DOP value.
VDOP	'Vertical dilution of precision', a relative measure of RMS vertical location error.
VPS	'Vemco positioning system', an acoustic telemetry system based on trilateration with stationary receivers.

Table 1: Glossary of terms

Device	#	**Best model**	Z^2_{red} *Joint*	Z^2_{red} *Ind.*	RMSE_{min} (m)
ATS G2110e	19	$\widehat{\text{HDOP}}(N_{\text{sat}}) \times \text{TTF}^2$	2.9	2.8	2.4–2.5
ATS G5-2D	4	homoskedastic†	1.5	1.5	3.5–3.6
ATS G10 UltraLITE	15	$\widehat{\text{HDOP}}(N_{\text{sat}})^\dagger$	2.9	2.7	21.0–21.3
CTT 1090	3	HDOP	4.6	4.3	8.7–10.3
CTT ES400	15	$\widehat{\text{HDOP}}(N_{\text{sat}})^\dagger$	2.4	2.3	3.9–4.0
CTT BT3	8	HDOP	4.3	4.0	9.0–9.3
e-obs generation-1 (1059–1078)	17	"horizontal accuracy"	2.1	2.0	5.8–6.1
e-obs generation-3 (5081–6375)	15	"horizontal accuracy"	1.3	1.3	1.70–1.72‡
e-obs generation-3 (6577–6752)	6	"horizontal accuracy"	2.5	1.9	2.6–2.7
e-obs generation-3 (7095–7106)	5	"horizontal accuracy"	1.8	1.7	2.5–2.6
e-obs generation-3 (7401–7405)	3	"horizontal accuracy"	2.5	2.4	2.8–2.9
GlobalTop PA6H	1	homoskedastic	NA	1.5	2.7–2.9‡
Lotek Lifecycle 330	12	homoskedastic	2.3	1.2	NA
Lotek PinPoint 240	2	DOP & is(timeout)	2.5	2.4	8.0–8.7
Lotek WildCell GPS-GSM	5	is(validated)	1.8	1.5	NA
NTT Decomo CTG-001G	3	"error"	1.9	1.9	3.1–3.4
Ornitela 20 gram (182902–182928)	6	homoskedastic	2.2	1.6	NA
Ornitela 25-gram (191771–191779)	6	homoskedastic	3.4	2.1	NA
Sirtrack Pinnacle Solar G5C275F	3	TTF $\geq$ 32 sec	4.2	4.2	1.63–1.68
Technosmart GiPSy 5	2	$\widehat{\text{HDOP}}(N_{\text{sat}})$	2.1	2.1	10.1–10.5
TS Quantum 4k Enhanced	2	HDOP & is(2D)	2.4	2.4	6.5–7.4
Telonics Gen4 GPS-VHF	3	HDOP & is(QFP)	1.9	1.8	4.8–5.1
Telonics Gen4 GPS-Iridium	3	HDOP & is(QFP)	1.7	1.7	2.6–3.0
UniKN Logger	3	$\widehat{\text{HDOP}}(N_{\text{sat}})$	2.1	2.1	9.3–10.1
Vectronic GPS Plus	6	"error"	1.5	1.4	3.7–4.9
Vectronic Vertex Lite-3D	18	DOP	2.5	2.4	16.3–16.9
Vemco HR2-V9 VPS	6	HPE	2.1	1.4	NA

Table 2: Selected horizontal error models, detailed in App. S4, with corresponding goodness-of-fit statistics (Z^2_{red}) and root-mean-square location error given ideal reception (RMSE_{min}) for 27 GPS and VPS device models, where '#' denotes the number of devices sampled. The Z^2_{red} statistic assesses the performance of a device's error model, where a lower value indicates more informative DOP values and a shorter tailed UERE distribution. Z^2_{red} statistics are given for both the joint error model, assuming all devices of the same type are equivalent, and the individualized error model, assuming all devices of the same type have unrelated error parameters. If these two statistics are close, then error calibration can be assumed to be transferable among devices of this type. †Performance of the selected model was not significantly better than the null HDOP model. ‡This performance was obtained in 1-Hz data.

Figure 1: Proposed workflow to account for location error in animal movement analysis, with the three most critical steps highlighted. We note that most of these steps are avoidable. In particular, if the tracking data are supplied pre-calibrated, then researchers can proceed immediately to error-informed movement analysis.

Figure 2: Demonstration of how ignored heteroskedasticity (time-varying variance) can lead to heavy-tailed distributions. In subpanel **A**, the location-error distribution is given for the most common variance, as a representative example. This distribution is Gaussian with 10-meter RMS error, and is light tailed. In subpanel **B**, 10 random HDOP values are drawn with resulting location-error distributions overlayed. Each individual distribution is light tailed, but the scale of location error varies dramatically among them. In subpanel **C**, the average location error is considered, which is proportionally t-distributed. This final distribution is what is considered when ignoring heteroskedasticity, and is heavy tailed.

Figure 3: Simulated animal tracks with location error (**A**-**B**) and a visual analysis of outliers (**C**-**F**). In the first column, an outlier is introduced—colored blue in the first and third rows—where the corresponding horizontal dilution of precision (HDOP) value is underestimated by a factor of 50. The second column contains the same location estimates as the first column, but where all HDOP values are accurate. Scatter plots are provided in the first row, with 95% error circles. Error-informed speed and distance estimates are calculated and used to color the data in row two, where fast transits are more blue and remote locations are more red. Finally, the same speed and distance point estimates from the second row are given confidence intervals and plotted in the third row. 'Minimum speed' refers to the minimum speed capable of explaining the data, while 'core deviation' refers to distance from the median location (see App. S5 for precise explanation of these quantities). The obvious outlier in panels **A** & **E** is not an outlier in panels panels **B** & **F**.

Figure 4: GPS data with 95% error circles (red) and occurrence distributions (blue) for a common noddy (*Anous stolidus*), as calculated from the AIC-best continuous-time movement model fit (**A**) simultaneously with a homoskedastic location-error model, (**B**) simultaneously with a PDOP-informed error model, and (**C**) with a calibrated PDOP error model. Subpanels **A**-**B** represent the conventional application of state-space models, which simultaneously fit movement and error model parameters, while subpanel **C** represents the calibration approach we advocate. Compared to the best location-error model (**C**), the simultaneously fit error model without PDOP information (**A**) overestimates the fix-location uncertainty by a factor of 970 and the overall occurrence area is larger by a factor of 2.4. By better distinguishing the covariance structure of the movement and error models, simultaneously fitting the PDOP-informed error model (**B**) reduces these factors of 160 and 1.04 respectively.

Figure 5: 95% home-range estimates with 95% CIs for an Argos-tracked night-heron versus the worst location class included in the data, with less-accurate location classes censored for that estimate, where the black points correspond to conventional kernel density estimates (KDE) and the blue points correspond to error-informed autocorrelated kernel density estimates (AKDE). The Argos location classes 3,2,1,A,0,B are sorted by accuracy, and so the first estimates at '3' only include data from the most accurate location class, and therefore have the smallest sample size, while the final estimates at 'B' include data from location classes 3–B, and are the largest when not accounting for location errors. Only the error-informed estimates are consistent across included location classes, though the effective sample size drops precipitously when only including location class 3. Note the logarithmic scale on the y-axis.

Figure 6: 2017 (red) and 2018 (blue) summer ranges of a night heron, as given by (**A**) Argos Doppler-shift error ellipses zoomed out to their largest location-error scale, (**B**) the same Argos error ellipses zoomed in to the inferred home-range scale, (**C**) conventional KDE home ranges, and (**D**) error-informed AKDE home range. All contours depict the 95% coverage areas. The conventional KDE estimate is inflated from location error, which makes the two summer ranges appear more similar than they are.

Figure 7: Mean speed estimates with 95% CIs (where available) for a GPS-tracked wood turtle versus the worst horizontal dilution of precision (HDOP) value included in the data. The black points correspond to conventional straight-line distance (SLD) estimates and the blue points correspond to an error-informed continuous-time speed and distance (CTSD) estimator based on time-series Kriging (Noonan et al., 2019). The first estimate only includes location fixes with HDOP ≤ 4, while the final estimate includes all of the data. Only the error-informed estimates are consistent across all included HDOP values, though the lowest HDOP locations will be biased towards open habitats.

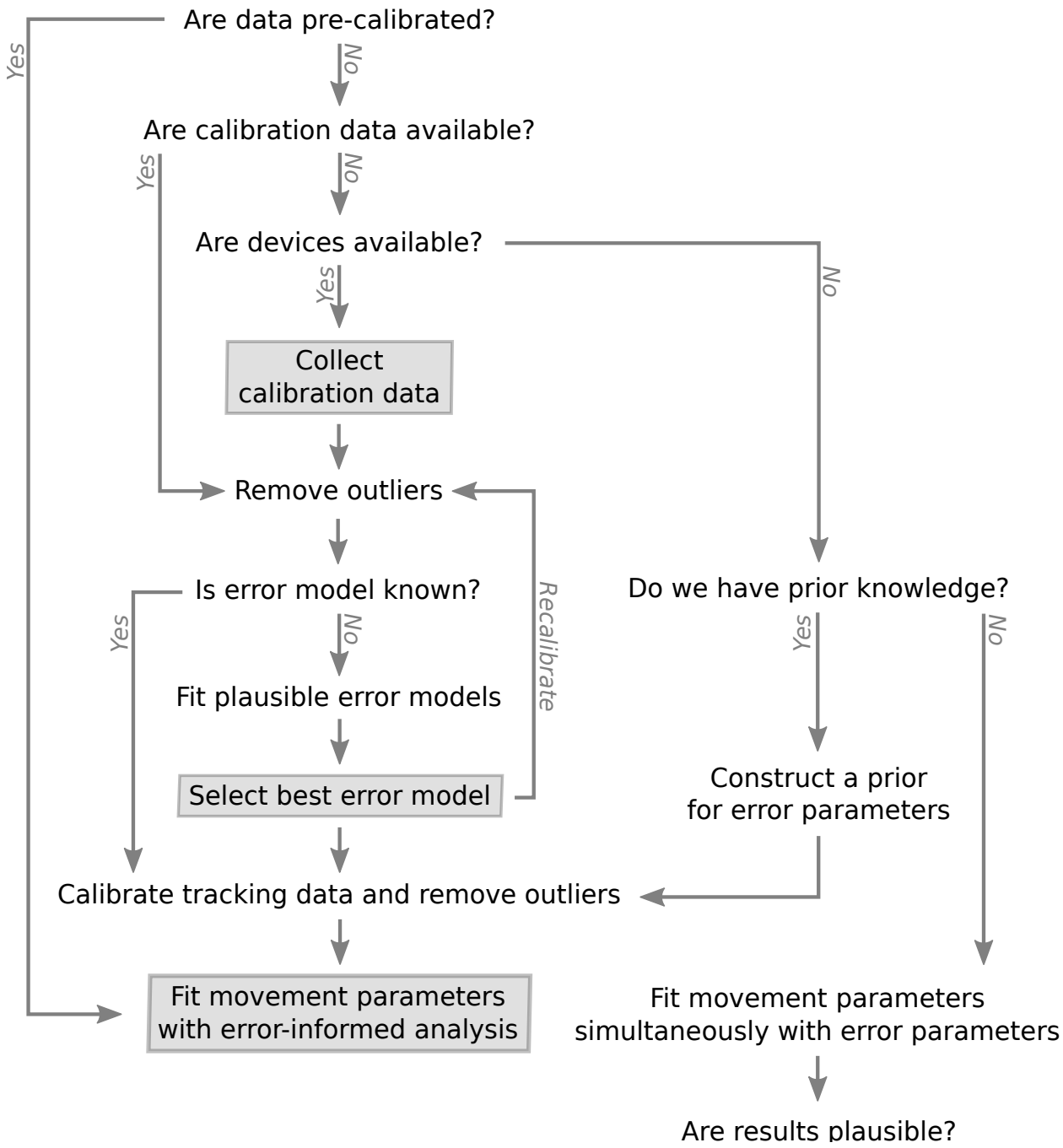
Are data pre-calibrated?
Yes
No
Are calibration data available?
Yes
No
Are devices available?
Yes
No
Collect calibration data
Remove outliers
Is error model known?
Yes
No
Recalibrate
Fit plausible error models
Select best error model
Do we have prior knowledge?
Yes
No
Construct a prior for error parameters
Calibrate tracking data and remove outliers
Fit movement parameters with error-informed analysis
Fit movement parameters simultaneously with error parameters
Are results plausible?

Figure 1

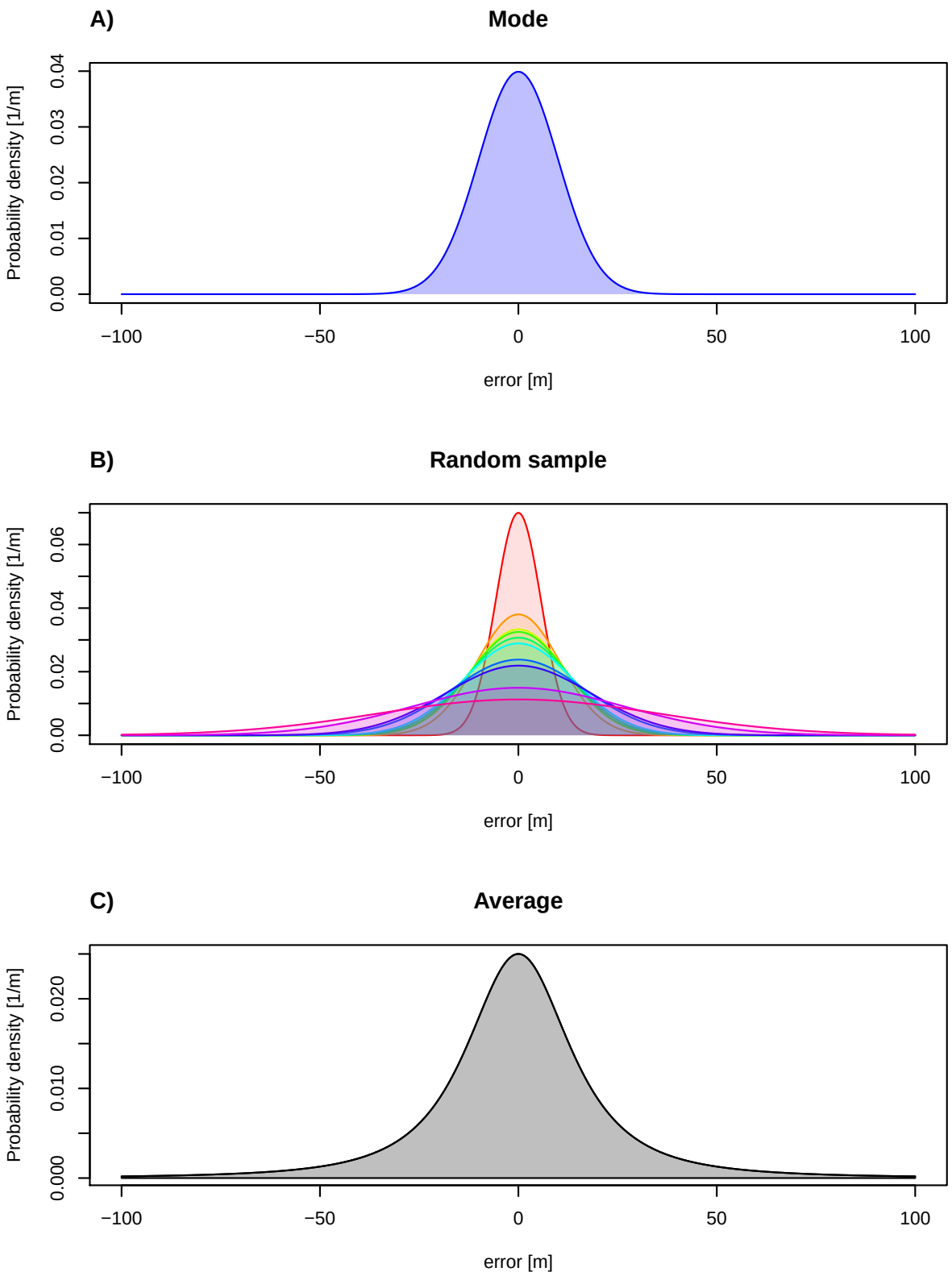
A)
Mode
Probability density [1/m]
0.00
0.01
0.02
0.03
0.04
−100
−50
0
50
100
error [m]
B)
Random sample
Probability density [1/m]
0.00
0.02
0.04
0.06
−100
−50
0
50
100
error [m]
C)
Average
Probability density [1/m]
0.000
0.010
0.020
−100
−50
0
50
100
error [m]

Figure 2

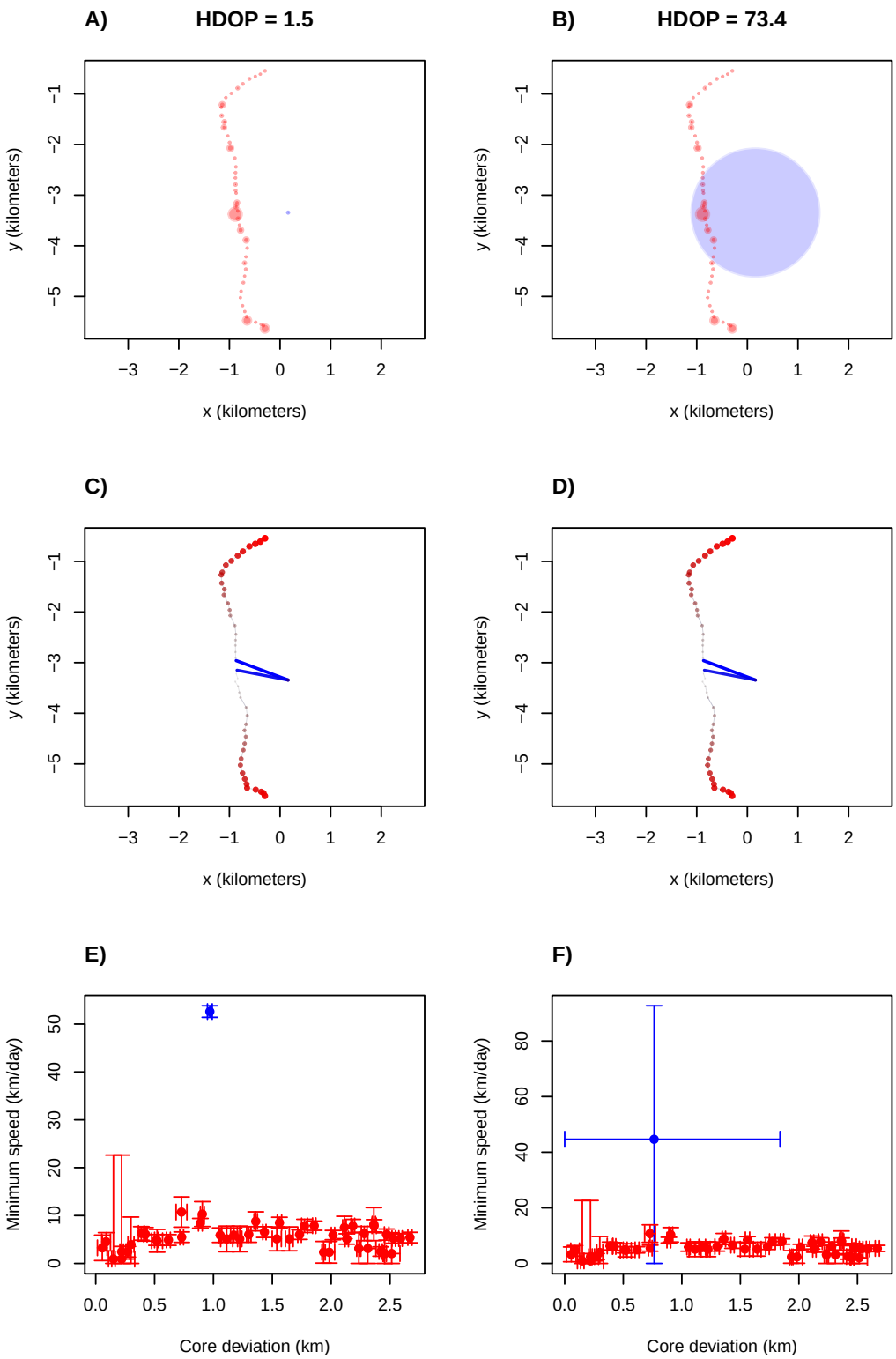
A)
HDOP = 1.5
B)
HDOP = 73.4
C)
D)
E)
F)
y (kilometers)
x (kilometers)
Minimum speed (km/day)
Core deviation (km)

Figure 3

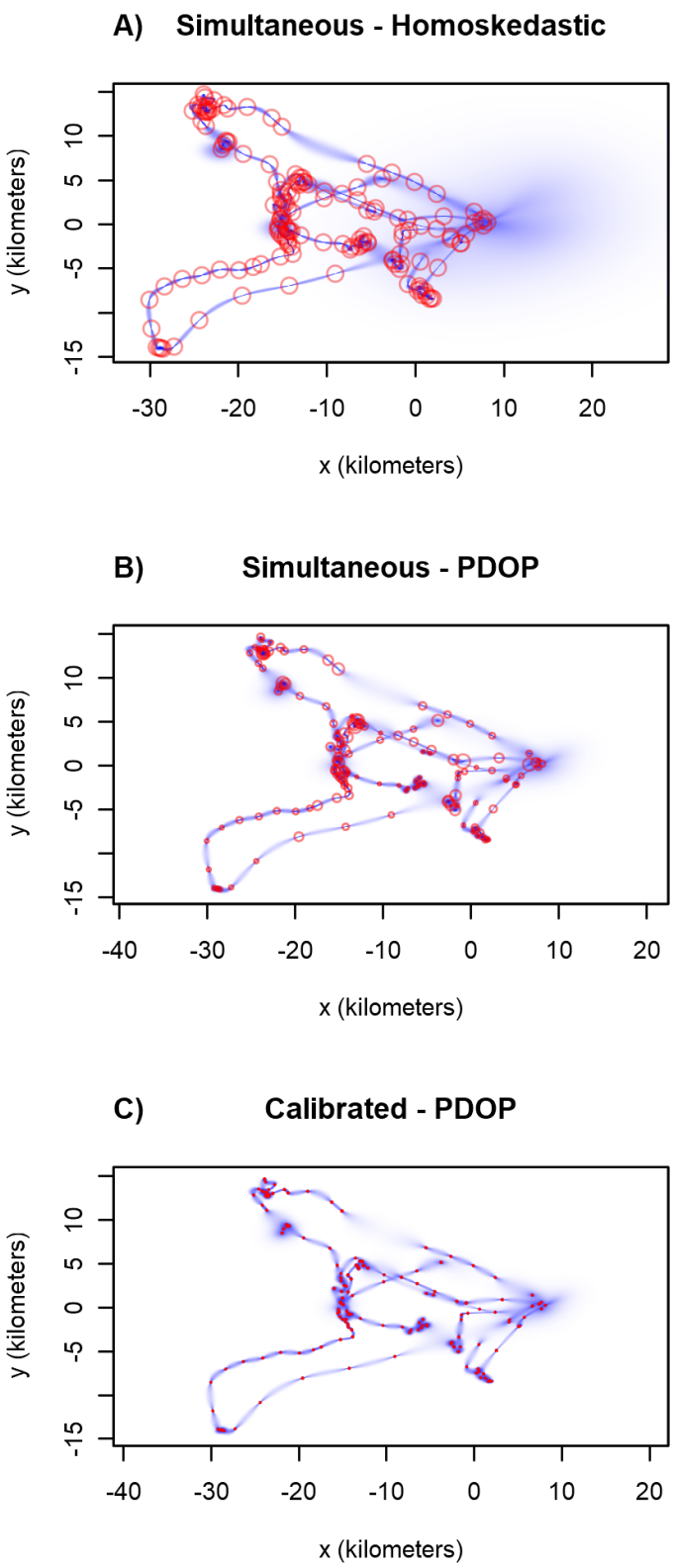
A) Simultaneous - Homoskedastic
y (kilometers)
x (kilometers)
B) Simultaneous - PDOP
y (kilometers)
x (kilometers)
C) Calibrated - PDOP
y (kilometers)
x (kilometers)

Figure 4

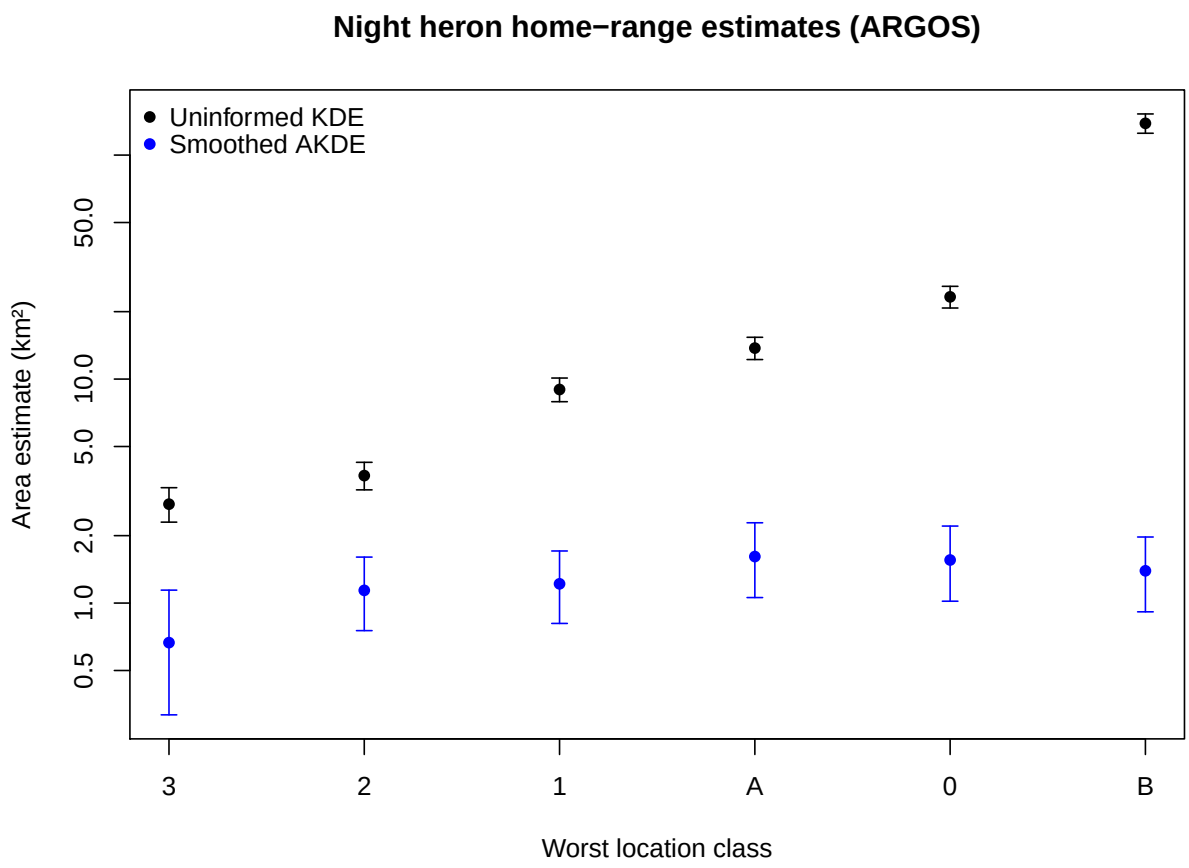
Night heron home−range estimates (ARGOS)
Uninformed KDE
Smoothed AKDE
Area estimate (km²)
0.5
1.0
2.0
5.0
10.0
50.0
3
2
1
A
0
B
Worst location class

Figure 5

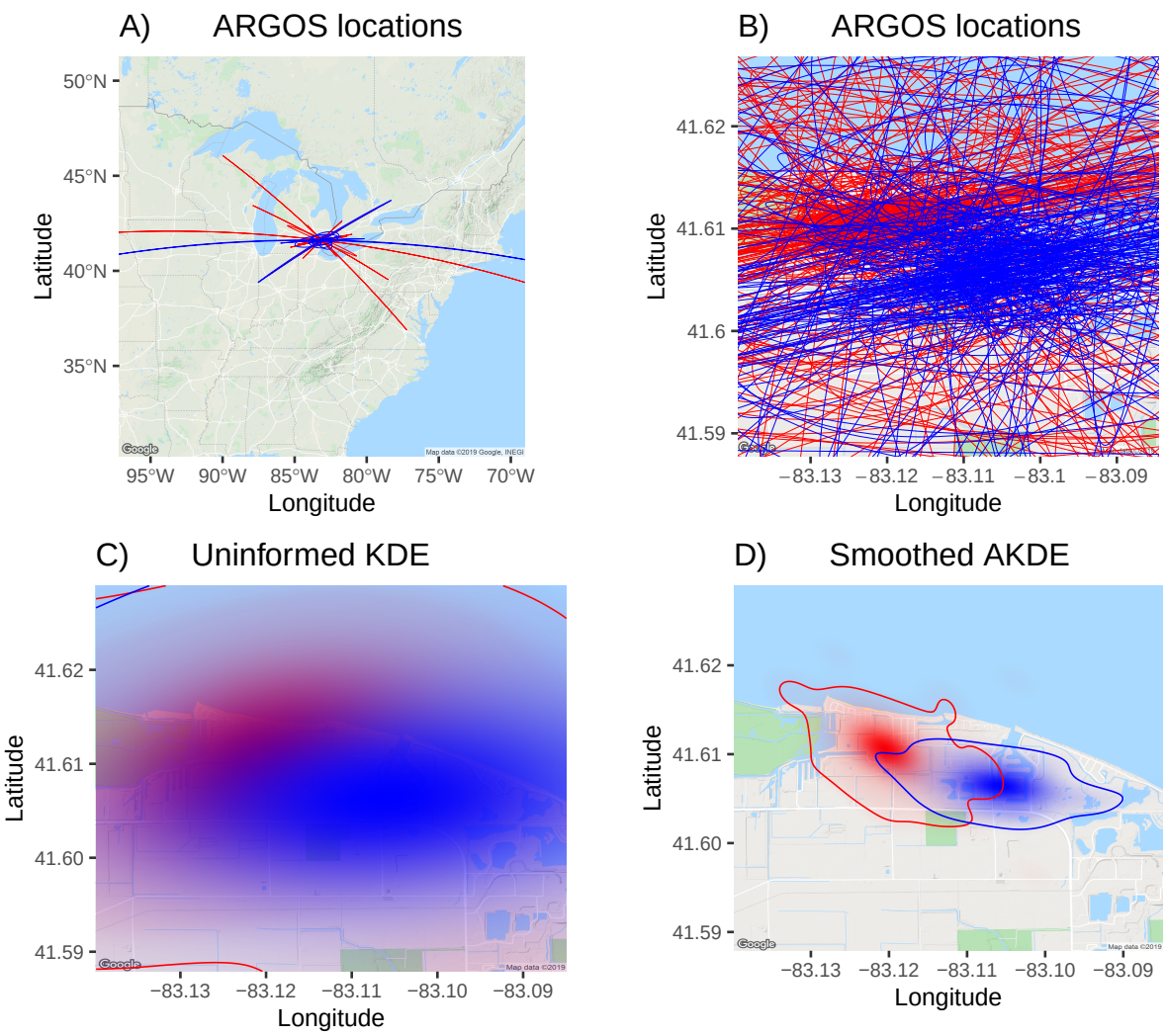
A) ARGOS locations
50°N
45°N
40°N
35°N
Latitude
95°W
90°W
85°W
80°W
75°W
70°W
Longitude
B) ARGOS locations
41.62
41.61
41.6
41.59
Latitude
−83.13
−83.12
−83.11
−83.1
−83.09
Longitude
C) Uninformed KDE
41.62
41.61
41.60
41.59
Latitude
−83.13
−83.12
−83.11
−83.10
−83.09
Longitude
D) Smoothed AKDE
41.62
41.61
41.60
41.59
Latitude
−83.13
−83.12
−83.11
−83.10
−83.09
Longitude

Figure 6

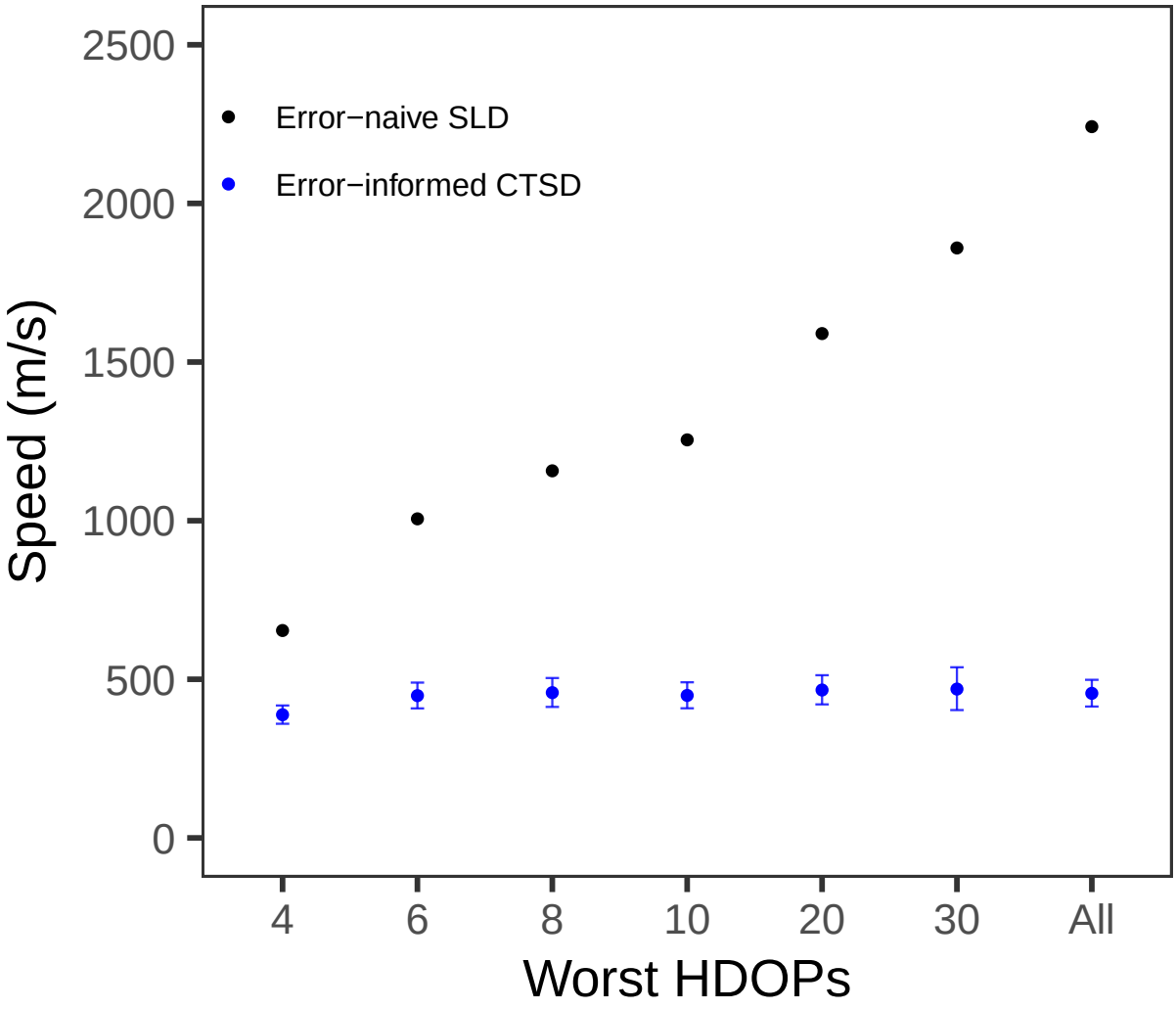
Error−naive SLD
Error−informed CTSD
Speed (m/s)
2500
2000
1500
1000
500
0
4
6
8
10
20
30
All
Worst HDOPs

Figure 7

S1 Sensitivity analysis

S1.1 Sensitivity analysis via simulation of errors

If $\hat{\theta}(\text{data})$ is the output of some analysis conditioned on the data that does not account for location error, and $\hat{\theta}(\text{data} + \text{error}_{\text{sim}})$ is the output of the same analysis, but with additional simulated location error added to the original data (atop its unknown location error), then the second-order variance induced by location error is simply the variance of the estimates $\hat{\theta}(\text{data} + \text{error}_{\text{sim}})$, with respect to the ensemble of simulations, and is additive. The second-order bias induced by location error is given by the difference in the average simulated analysis and the unaltered analysis

$$\left\langle \hat{\theta}(\text{data}) - \hat{\theta}(\text{truth}) \right\rangle_{\text{error}} = \left\langle \hat{\theta}(\text{data} + \text{error}_{\text{sim}}) - \hat{\theta}(\text{data}) \right\rangle_{\text{sim}} + \mathcal{O}\big(\text{error}^4\big)\,, \tag{S1.1}$$

where $\langle \cdots \rangle_X$ denotes the expectation value with respect to the process X, data = truth + error, and $\mathcal{O}(\text{error}^4)$ denotes higher-order corrections, including $\text{VAR}[\text{error}]^2$, which require an error-informed analysis to determine. The only assumptions made here are that $\hat{\theta}(\text{data})$ is analytic in the neighborhood of $\hat{\theta}(\text{truth})$, and that the distribution of errors is symmetric, which includes the errors being of mean zero.

S1.2 Sensitivity analysis via simulation of trajectories

As a slight improvement, here we upgrade the previous analysis from estimates based on $\hat{\theta}(\text{data})$ to estimates based on $\hat{\theta}(\text{prediction}|\text{data})$, where 'prediction' refers to the Kriged location estimates, which are conditioned on the data and fitted movement model (Fleming et al., 2016). The second-order variance induced by location error is the variance of the estimates $\hat{\theta}(\text{trajectory}|\text{data})$, where 'trajectory' refer to simulated locations conditioned on the data and fitted movement model, which are centered on the conditional predictions (Fleming et al., 2017). The second-order bias induced by location error is given by the difference in the average simulated analysis and the analysis based on the Kriged estimates

$$\begin{aligned} &\left\langle \hat{\theta}(\text{prediction}|\text{data}) - \hat{\theta}(\text{truth}) \right\rangle_{\text{error}} \\ &= \left\langle \hat{\theta}(\text{trajectory}|\text{data}) - \hat{\theta}(\text{prediction}|\text{data}) \right\rangle_{\text{trajectory}} + \mathcal{O}\big(\text{error}^4\big)\,. \end{aligned} \tag{S1.2}$$

S2 Error-model statistics

S2.1 Single-UERE estimation

Here we derive the minimum variance unbiased (MVU) estimator for the mean-square UERE, $\langle \mathrm{UERE}^2 \rangle$, given calibration data where the GPS tag/collar is at rest. Lack of bias and low variance is important so that multiple (possibly opportunistic) calibration datasets can be safely pooled regardless of size. Achieving the MVU property requires an assumption of normality in the distribution UEREs, which can still allow for a heavy-tailed distribution of location errors. For more general UERE distributions, these estimators are still unbiased.

In two dimensions, the RMS UERE, $\sqrt{\langle \mathrm{UERE}^2 \rangle}$, serves as the proportionality constant relating RMS location error and HDOP value

$$\sqrt{\mathrm{VAR}[x(t)] + \mathrm{VAR}[y(t)]} = \mathrm{HDOP}(t) \cdot \sqrt{\langle \mathrm{UERE}_{\mathrm{H}}^2 \rangle}\,, \tag{S2.1}$$

$$\mathrm{VAR}[x(t)] + \mathrm{VAR}[y(t)] = \mathrm{HDOP}(t)^2 \cdot \langle \mathrm{UERE}_{\mathrm{H}}^2 \rangle\,, \tag{S2.2}$$

where $\mathrm{UERE}_{\mathrm{H}}$ denotes horizontal UERE values. As GPS satellite coverage is relatively uniform, GPS location errors will generally be circular on average, in contrast to Argos Doppler-shift errors where there are only a few satellites in polar orbits. There are exceptions to the assumption of circular errors (Fig. S2.1), but GPS tracking devices do not record error-ellipse information, so there is little that can be done in these cases.

Given circular location errors, we have

$$\mathrm{VAR}[x(t)] = \mathrm{VAR}[y(t)] = \frac{1}{2}\,\mathrm{HDOP}(t)^2 \cdot \langle \mathrm{UERE}_{\mathrm{H}}^2 \rangle\,, \tag{S2.3}$$

and similarly for the vertical dimension we have

$$\mathrm{VAR}[z(t)] = \mathrm{VDOP}(t)^2 \cdot \langle \mathrm{UERE}_{\mathrm{V}}^2 \rangle\,, \tag{S2.4}$$

where in practice the horizontal and vertical RMS UERE values are not equal.

Let k index the K calibration datasets, with each corresponding to identical tags/collars at fixed locations $\boldsymbol{\mu}_k$. The MVU mean-location estimators are then given by the weighted average

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{W_k} \sum_{i=1}^{n_k} w_k(t_i)\, \mathbf{r}_k(t_i)\,, \qquad w_k(t) = \frac{q}{\mathrm{DOP}_k(t)^2}\,, \qquad W_k = \sum_{i=1}^{n_k} w_k(t_i)\,, \tag{S2.5}$$

where $\mathbf{r}_k(t) = (x_k(t), y_k(t))$ in two dimensions. From hereon we keep our relations general with $q = \dim(\mathbf{r})$ denoting the number of spatial dimensions ($q = 2$ for horizontal and $q = 1$ for vertical. The MVU mean-square UERE estimator is

$$\langle \hat{\mathrm{UERE}}^2 \rangle = \frac{1}{q} \frac{1}{N-K} \sum_{k=1}^{K} \sum_{i=1}^{n_k} w_k(t_i)\, |\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2\,, \qquad N = \sum_{k=1}^{K} n_k\,, \tag{S2.6}$$

given N total sampled times for the K calibration datasets. The sampling distributions are given by

$$\hat{\boldsymbol{\mu}}_k \sim N(\boldsymbol{\mu}_k, W_k^{-1} \langle \mathrm{UERE}^2 \rangle)\,, \qquad q\,(N-K) \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{\langle \mathrm{UERE}^2 \rangle} \sim \chi^2_{q(N-K)}\,, \tag{S2.7}$$

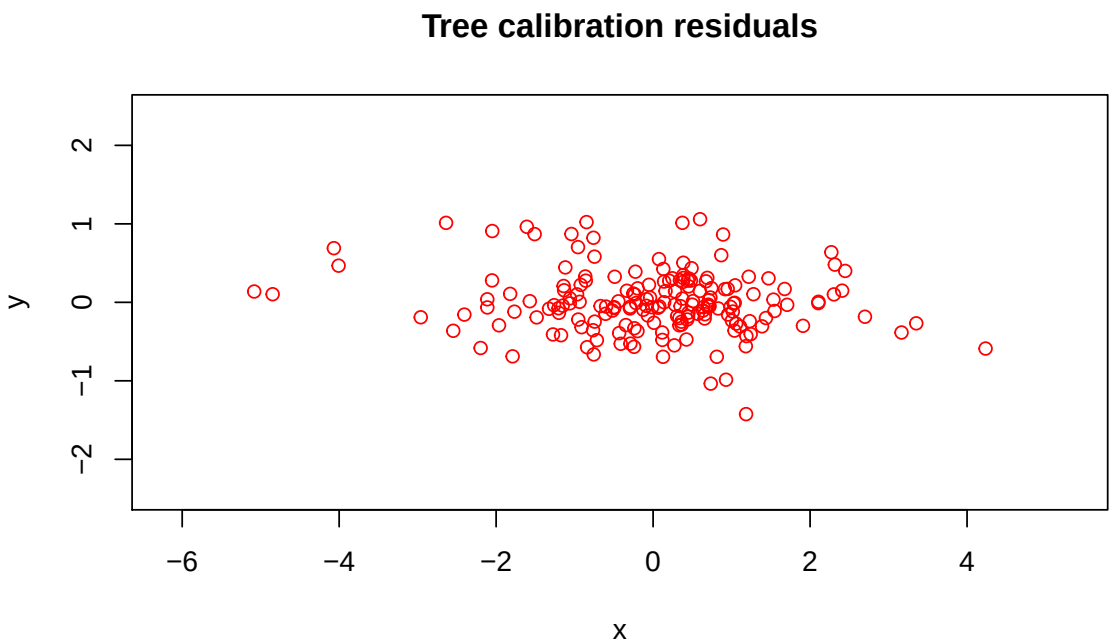

Figure S2.1: Error-model residuals from GPS calibration data where the tag was placed on the side of a tree. As the tree blocks out half of the sky, the spread in satellites is narrower along the axis perpendicular to the tree, which results in larger errors along that axis. Other situations where GPS location errors are elongated include when obtaining location fixes alongside cliffs or buildings.

and the standardized or "studentized" residuals are given by

$$\mathbf{z}_k(t) = \sqrt{\frac{w_k(t)}{\langle \hat{\mathrm{UERE}}^2 \rangle}}\,(\mathbf{r}_k(t) - \hat{\boldsymbol{\mu}}_k)\,, \tag{S2.8}$$

which can be checked for substantial deviance from normality or autocorrelation, as in Fig. S4.1.

S2.1.1 Single-UERE model selection

There are several situations where model selection is useful when calibrating data. First, while DOP values are usually informative, we do not always know the best proxy for the location-error variance (c.f., Sec. S3). Second, we may be unsure as to whether different devices share the same UERE parameters. Third, we may be unsure as to whether or not different location classes share the same UERE parameters. Therefore, it is useful to have $\mathrm{AIC_C}$ values to facilitate model selection. AIC values estimate the Kullback-Leibler divergence between the fitted model and true model (Burnham and Anderson, 2002). In situations where sample sizes are small and debiased parameter estimates differ substantially from maximum-likelihood (ML) parameter estimates, then $\mathrm{AIC_C}$ values can differ substantially from regular AIC values. Standard $\mathrm{AIC_C}$ formulas are specific to maximum-likelihood estimation with a simple error model, whereas here we do better than ML estimation with more complex error models.

Following analogous calculations in (Calabrese et al., 2018; Fleming et al., 2019), unbiased $\mathrm{AIC_C}$ values must satisfy the cross-validated log-likelihood, or double expectation $\langle\!\langle \cdots \rangle\!\rangle$ (over both the data and estimates calculated from independent data) of the log-likelihood function, regardless of whether or not the estimates are derived from maximizing said likelihood function.

$$\langle \mathrm{AIC_C} \rangle = \left\langle\!\!\left\langle -2\,\ell\Big(\langle \hat{\mathrm{UERE}}^2 \rangle, \hat{\boldsymbol{\mu}} \,\Big|\, \mathbf{r} \Big) \right\rangle\!\!\right\rangle\,, \tag{S2.9}$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(q \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{w_k(t_i)} \right\rangle + \left\langle \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}^2 \rangle} \right\rangle \left\langle\!\!\left\langle |\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2 \right\rangle\!\!\right\rangle \right)\,, \tag{S2.10}$$

$$= \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(q \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{w_k(t_i)} \right\rangle + \left\langle \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}^2 \rangle} \right\rangle q \left(\frac{\langle \mathrm{UERE}^2 \rangle}{w_k(t_i)} + \frac{\langle \mathrm{UERE}^2 \rangle}{W_k} \right) \right)\,, \tag{S2.11}$$

$$= q \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{w_k(t_i)} \right\rangle + q\,(N+K) \left\langle \frac{\langle \mathrm{UERE}^2 \rangle}{\langle \hat{\mathrm{UERE}}^2 \rangle} \right\rangle, \tag{S2.12}$$

$$= q \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{w_k(t_i)} \right\rangle + \frac{q^2 (N+K)(N-K)}{q\,(N-K) - 2}, \tag{S2.13}$$

and so by the Lehmann-Scheffé theorem (Lehmann and Scheffé, 1950, 1955), the MVU $\mathrm{AIC_C}$ values must be given by

$$\mathrm{AIC_C} = q \sum_{k=1}^{K} \sum_{i=1}^{n_k} \log 2\pi \frac{\langle \hat{\mathrm{UERE}}^2 \rangle}{w_k(t_i)} + \frac{q^2 (N+K)(N-K)}{q\,(N-K) - 2}, \tag{S2.14}$$

which one can check for asymptotic consistency with the standard AIC formula.

S2.1.2 Single-UERE goodness of fit

$\mathrm{AIC_C}$ differences allow us to rank our models' expected predictive capability, but these differences scale with sample size and so they do not indicate how well the models perform in an absolute sense. If our variance model was fixed and our comparison was among trend models, $\boldsymbol{\mu}$, then we might consider the reduced χ^2 statistic to index goodness of fit. However, our trend model is fixed and we are comparing variance models, and so we derive a reduced Z^2 statistic that is analogous to the familiar reduced χ^2 statistic.

We start by considering the "internally" studentized residuals, as goodness of fit can be measured by their closeness to unity.

$$u_{ki}^2 = \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}^2 \rangle} \frac{|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2}{q}\,, \tag{S2.15}$$

These square residuals are not exactly F-distributed because the numerators and denominators are not independent. Next we apply a trick due to Thompson (1935), decomposing all terms into independent variates.

$$|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2 = \alpha_{ki}^2\, |\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_{ki}|^2\,, \tag{S2.16}$$

$$\langle \hat{\mathrm{UERE}}^2 \rangle = \beta\, \langle \check{\mathrm{UERE}}_{ki}^2 \rangle + \frac{\alpha_{ki}}{N-K}\, w_k(t_i) \frac{|\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_{ki}|^2}{q}, \tag{S2.17}$$

$$\alpha_{ki} = \frac{W_k - w_k(t_i)}{W_k} \leq 1\,, \tag{S2.18}$$

$$\beta = \frac{N-K-1}{N-K} \leq 1\,, \tag{S2.19}$$

where all $\check{\theta}_{ki}$-estimates are calculated leaving out the ki^{th} observation, and then the "externally" studentized residuals

$$t_{ki}^2 = \frac{w_k(t_i)}{\langle \check{\mathrm{UERE}}_{ki}^2 \rangle} \frac{|\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_k|^2}{q} = \frac{\beta\, u_{ki}^2}{\alpha_{ki}^2 - \frac{\alpha_{ki}}{N-K} u_{ki}^2} \sim F_{q(N-K-1)}^{q}\,, \tag{S2.20}$$

are exactly F-distributed. Importantly, there is a one-to-one correspondence between the internally studentized u^2 and externally studentized t^2 statistics, and so it does not matter which statistic we base our index on. Furthermore, relation (S2.49) is particularly useful because direct calculation of all t_{ki}^2 statistics via the leave-one-out $\check{\theta}$-estimators has a computational cost of $\mathcal{O}(N^2)$, whereas calculation here via u_{ki} has a computational cost of only $\mathcal{O}(N)$.

Again, goodness of fit can be measured by t^2 being close to unity. Both small values of t^2, where the model variance is too large, and large values of t^2, where model variance is too small, are equally bad and so we proceed to Fisher's Z-distribution

$$Z_{ki} = \frac{1}{2} \log t_{ki}^2\,, \tag{S2.21}$$

which places extreme values of t^2 on the same relative scale, with a variance model that is a factor too small being just as bad as a variance model that is a factor too large. Goodness of fit is now determined by the proximity of Z to zero, and lack of performance can be inferred from the magnitude of Z^2, which for all ik has an expectation value being only a function of the degrees of freedom. Our reduced Z^2 statistic is given by

$$Z^2_{\text{red}} = \frac{1}{N\,\langle Z^2\rangle} \sum_{k=1}^{K} \sum_{i=1}^{n_k} Z^2_{ki}\,, \qquad \langle Z^2_{\text{red}}\rangle = 1\,, \tag{S2.22}$$

where, $Z^2_{\text{red}} \ll 1$ indicates an over-performing model, $Z^2_{\text{red}} \approx 1$ indicates a well-performing model, and $Z^2_{\text{red}} \gg 1$ indicates an under-performing model, all relative to a true Gaussian model. If the true UERE distribution is not Gaussian, the reduced Z^2 statistic can still be used to compare performance across error models, only the value of 1 no longer separates over-performance from under-performance. In this context, over-performance indicates that the square errors are exceptionally proportional to the model variances, while under-performance indicates a weaker relationship between the two quantities. For example, bivariate Laplace errors with correctly modeled variance will produce $Z^2_{\text{red}} \sim 1.75$ when using the Gaussian formula for $\langle Z^2\rangle$. Even more extreme, bivariate Cauchy errors—which do not admit a finite variance—limit to $Z^2_{\text{red}} \sim 5.8$ with large N. All else being equal, a more heavy-tailed distribution will give rise to a larger value of Z^2_{red} under the normal assumption.

Finally, for any $Z = \frac{1}{2}\log t^2$, where $t^2 \sim F^{d_1}_{d_2}$ as above, the desired second moment $\langle Z^2\rangle$ is straightforwardly calculated to be

$$\langle Z^2\rangle = \left\langle \left(\frac{1}{2}\log\frac{\chi^2_{d_1}}{d_1}\right)^2 \right\rangle + \left\langle \left(\frac{1}{2}\log\frac{\chi^2_{d_2}}{d_2}\right)^2 \right\rangle - 2\left\langle \frac{1}{2}\log\frac{\chi^2_{d_1}}{d_1}\right\rangle \left\langle \frac{1}{2}\log\frac{\chi^2_{d_2}}{d_2}\right\rangle, \tag{S2.23}$$

in terms of the normalized log-χ^2 moments

$$\left\langle \frac{1}{2}\log\frac{\chi^2_d}{d}\right\rangle = -\frac{1}{2}\left(\log\left(\frac{d}{2}\right) - \psi\left(\frac{d}{2}\right)\right), \tag{S2.24}$$

$$\left\langle \left(\frac{1}{2}\log\frac{\chi^2_d}{d}\right)^2 \right\rangle = \left\langle \frac{1}{2}\log\frac{\chi^2_d}{d}\right\rangle^2 + \frac{1}{4}\,\psi'\left(\frac{d}{2}\right), \tag{S2.25}$$

where $\psi(z) = \frac{d}{dz}\log\Gamma(z)$ is the digamma function and $\Gamma(z) = (z-1)!$ is the gamma function.

S2.1.3 Mean-zero processes

The simplifications to the previous formulas for mean-zero process calibration, such as velocity, are

$$K = 0\,, \qquad \alpha = 1\,. \tag{S2.26}$$

S2.2 Multi-UERE estimation

There are several situations where multiple RMS UEREs need to be estimated. Many GPS modules record different quality locations, including 2D, 3D, "quick fixes" resolved and unresolved, but no HDOP information or their HDOP values derive from different methods

and are not on the same scale. In some cases, unresolved location quality can only be inferred from missing attributes like speed and altitude, or from the fix time reaching its maximum value, which requires one RMS UERE value for fix times below the threshold value and a larger RMS UERE value for timed-out fixes.

When there are $C > 1$ location classes for which we need to estimate C RMS UERE parameters, we cannot derive MVU estimators, but we can derive residual maximum likelihood (REML) estimators that reduce to MVU when $C = 1$. Let c index the C location classes, and define the indicator function

$$\delta_{ck}(t) = \begin{cases} 1\,, & \text{location class of tag } k \text{ at time } t \text{ is } c \\ 0\,, & \text{location class of tag } k \text{ at time } t \text{ is } \textbf{not } c \end{cases}\,, \tag{S2.27}$$

then the REML log-likelihood function is

$$\begin{aligned} \ell = &-\frac{1}{2}\sum_{c=1}^{C}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\delta_{ck}(t_i)\left(q\log 2\pi\frac{\langle \mathrm{UERE}_c^2\rangle}{w_k(t_i)} + \frac{w_k(t_i)}{\langle \mathrm{UERE}_c^2\rangle}\left|\mathbf{r}_k(t_i)-\hat{\boldsymbol{\mu}}_k\right|^2\right) \\ &-\frac{q}{2}\sum_{k=1}^{K}\log 2\pi\sum_{c=1}^{C}\sum_{i=1}^{n_k}\delta_{ck}(t_i)\frac{w_k(t_i)}{\langle \mathrm{UERE}_c^2\rangle}\,, \end{aligned} \tag{S2.28}$$

where the last term on the right hand side is the REML correction to the regular log-likelihood function.

The mean-square UERE and mean location estimating equations are given by

$$\langle \hat{\mathrm{UERE}}_c^2\rangle = \frac{1}{q\,\hat{\mathrm{DOF}}_c}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\delta_{ck}(t_i)\,w_k(t_i)\left|\mathbf{r}_k(t_i)-\hat{\boldsymbol{\mu}}_k\right|^2\,, \tag{S2.29}$$

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{\hat{P}_k}\sum_{c=1}^{C}\sum_{i=1}^{n_k}\delta_{ck}(t_i)\,\frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2\rangle}\,\mathbf{r}_k(t_i)\,, \tag{S2.30}$$

given the DOF and precision-weight estimates

$$\hat{\mathrm{DOF}}_c = \overbrace{\sum_{k=1}^{K}\sum_{i=1}^{n_k}\delta_{ck}(t_i)}^{N_c} - \overbrace{\sum_{k=1}^{K}\frac{\hat{P}_{ck}}{\hat{P}_k}}^{\hat{K}_c}\,, \tag{S2.31}$$

$$\hat{P}_k = \sum_{c=1}^{C}\hat{P}_{ck}\,, \tag{S2.32}$$

$$\hat{P}_{ck} = \sum_{i=1}^{n_k}\delta_{ck}(t_i)\,\frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2\rangle}\,, \tag{S2.33}$$

where the second term on the right-hand-side of (S2.31) is the REML correction to the MLE. We solve the estimating equations iteratively, by evaluating the above relations in reverse order, starting with an initial guess for the RMS UEREs. In the case of one location class ($C = 1$), this algorithm converges in one iteration. For $C > 1$, convergence is at worst quadratic and machine precision is typically achieved in a few iterations.

Effectively, the K lost degrees of freedom are distributed across the C RMS UERE estimates (K_c), according to the precision weights P_{ck}, as it is straightforward to show that

$$\left\langle \sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ck}(t_i) \frac{w_k(t_i)}{\langle \mathrm{UERE}_c^2 \rangle} \left|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k\right|^2 \right\rangle = \mathrm{DOF}_c \,, \tag{S2.34}$$

at the true RMS UERE values, and the total number of degrees of freedom are indeed $N-K$.

The residuals are given by

$$\mathbf{z}_k(t) = \sqrt{\frac{w_k(t)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}} \left(\mathbf{r}_k(t) - \hat{\boldsymbol{\mu}}_k\right) \quad \text{where} \quad \delta_{ck}(t) = 1 \,. \tag{S2.35}$$

From the Hessian of the REML log-likelihood function, the sampling distributions of interest are asymptotically given by

$$\hat{\boldsymbol{\mu}}_k \sim N(\boldsymbol{\mu}_k, P_k^{-1}) \,, \qquad\qquad q\,\mathrm{dof}_c \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{\langle \mathrm{UERE}_c^2 \rangle} \mathrel{\dot{\sim}} \chi^2_{q\,\mathrm{dof}_c} \,, \tag{S2.36}$$

in terms of the degrees of freedom

$$\mathrm{dof}_c = N_c - \sum_{k=1}^{K} \left(\frac{P_{ck}}{P_k}\right)^2 \geq \mathrm{DOF}_c \,, \tag{S2.37}$$

which is larger than the degrees of freedom in the estimating equations that decompose the variance of the residuals (S2.31), though still smaller than the biased maximum likelihood value. Both dof_c and DOF_c reduce to $N-K$ when $C=1$.

S2.2.1 Multi-UERE model selection

Generalizing the results of Sec. S2.1.1, we again evaluate the cross-validated log-likelihood. However, instead of an exact calculation, we must make asymptotic approximations ignoring the correlation between the RMS UERE estimates and mean location estimates. These approximations result in an improved, though not fully unbiased, estimator of the Kullback-Leibler divergence (AIC).

$$\langle \mathrm{AIC_C} \rangle = \left\langle\!\!\left\langle -2\,\ell\Big(\langle \hat{\mathrm{UERE}}^2 \rangle, \hat{\boldsymbol{\mu}} \,\Big|\, \mathbf{r}\Big) \right\rangle\!\!\right\rangle \,, \tag{S2.38}$$

$$= \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ck}(t_i) \left(q \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{w_k(t_i)} \right\rangle + \left\langle \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle} \right\rangle \left\langle\!\!\left\langle \left|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k\right|^2 \right\rangle\!\!\right\rangle \right) , \tag{S2.39}$$

$$= \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ck}(t_i) \left(q \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{w_k(t_i)} \right\rangle + \left\langle \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}^2 \rangle} \right\rangle q \left(\frac{\langle \mathrm{UERE}_c^2 \rangle}{w_k(t_i)} + \frac{1}{P_k} \right) \right) , \tag{S2.40}$$

$$= q \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{w_k(t_i)} \right\rangle + q \sum_{c=1}^{C} (N_c + K_c) \left\langle \frac{\langle \mathrm{UERE}_c^2 \rangle}{\langle \hat{\mathrm{UERE}}_c^2 \rangle} \right\rangle , \tag{S2.41}$$

$$= q \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left\langle \log 2\pi \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{w_k(t_i)} \right\rangle + \sum_{c=1}^{C} \frac{q^2 (N_c + K_c)\,\mathrm{dof}_c}{q\,\mathrm{dof}_c - 2} , \tag{S2.42}$$

and so our second-order $\mathrm{AIC_C}$ values are given by

$$\mathrm{AIC_C} = q \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \log 2\pi \frac{\langle \hat{\mathrm{UERE}}_c^2 \rangle}{w_k(t_i)} + \sum_{c=1}^{C} \frac{q^2 (N_c + K_c)\, \mathrm{dof}_c}{q\, \mathrm{dof}_c - 2}. \tag{S2.43}$$

S2.2.2 Multi-UERE goodness of fit

Again we start with the internally studentized residuals

$$u_{ki}^2 = \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle} \frac{|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2}{q}, \qquad \text{where} \quad \delta_{ck}(t_i) = 1, \tag{S2.44}$$

and apply the method of Thompson (1935)

$$|\mathbf{r}_k(t_i) - \hat{\boldsymbol{\mu}}_k|^2 = \hat{\alpha}_{ki}^2 \, |\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_k|^2, \tag{S2.45}$$

$$\langle \hat{\mathrm{UERE}}_c{}^2 \rangle = \hat{\beta}_{ki} \, \langle \check{\mathrm{UERE}}_{ki}^2 \rangle + \frac{\alpha_{ki}}{\hat{\mathrm{DOF}}_c} \, w_k(t_i) \frac{|\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_k|^2}{q}, \tag{S2.46}$$

$$\hat{\alpha}_{ki} = \frac{\hat{P}_{ck} - \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}}{\hat{P}_{ck}} \leq 1, \tag{S2.47}$$

$$\hat{\beta}_{ki} = \frac{\check{\mathrm{DOF}}_{ki}}{\hat{\mathrm{DOF}}_c} = \frac{N_c - 1 - \sum_{k'=1}^{K} \frac{\hat{P}_{ck'} - \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}}{\hat{P}_{k'} - \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}}}{\hat{\mathrm{DOF}}_c} \leq 1, \tag{S2.48}$$

to transform into the externally studentized residuals

$$t_{ki}^2 = \frac{w_k(t_i)}{\langle \check{\mathrm{UERE}}_{ki}^2 \rangle} \frac{|\mathbf{r}_k(t_i) - \check{\boldsymbol{\mu}}_k|^2}{q} = \frac{\hat{\beta}_{ki} \, u_{ki}^2}{\hat{\alpha}_{ki}^2 - \frac{\hat{\alpha}_{ki}}{\hat{\mathrm{DOF}}_c} u_{ki}^2} \mathrel{\dot{\sim}} F^{q}_{q\,\check{\mathrm{dof}}_{ki}}, \tag{S2.49}$$

$$\check{\mathrm{dof}}_{ki} = N_c - 1 - \sum_{k'=1}^{K} \left(\frac{\hat{P}_{ck'} - \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}}{\hat{P}_{k'} - \frac{w_k(t_i)}{\langle \hat{\mathrm{UERE}}_c^2 \rangle}} \right)^2. \tag{S2.50}$$

The relationship between the internally and externally studentized residuals is no longer as clear, but this remains a computationally efficient way of calculating the externally studentized residuals, for which we have a better approximation to the statistic's distribution.

Fisher's Z-statistic is then given by

$$Z_{ki} = \frac{1}{2} \log t_{ki}^2, \tag{S2.51}$$

and our reduced Z^2 statistic is given by

$$Z_{\mathrm{red}}^2 = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \frac{Z_{ki}^2}{\langle Z_{ki}^2 \rangle}, \qquad \langle Z_{\mathrm{red}}^2 \rangle = 1, \tag{S2.52}$$

with the necessary formulas for $\langle Z_{ki}^2 \rangle$ calculated in relation (S2.23).

S2.2.3 Mean-zero processes

The simplifications to the previous formulas for mean-zero process calibration, such as velocity, are

$$\hat{\mathrm{DOF}}_c = \hat{\mathrm{dof}}_c = N_c\,, \qquad \alpha = 1\,, \qquad \beta_c = \frac{N_c - 1}{N_c}\,, \qquad \check{\mathrm{dof}}_c = N_c - 1\,. \tag{S2.53}$$

S3 Precision models in practice

Here we detail precision models beyond null model (1.1), for GPS devices that do not supply informative DOP values or exhibit complications. In App. S4, these models are applied to a wide range of devices, demonstrating their coverage and utility.

S3.1 DOP proxies

Co-opting DOP values

In lieu of HDOP values, some GPS devices only record "position DOP" (PDOP), "geometric DOP" (GDOP), or ambiguous DOP values. PDOP values combine HDOP and VDOP; GDOP values combine PDOP and "time DOP" (TDOP) values. As HDOP and VDOP values are both a function of satellite number and spread, they are correlated. Therefore, it is reasonable to co-opt or appropriate related DOP values in the absence of proper HDOP and VDOP values. Model selection can then determine whether alternative or ambiguous DOP values are more predictive than nothing (homoskedastic location errors).

By default, when importing data in to `ctmm` via the `as.telemetry()` function, and assuming the data are not of a special variety (Argos Doppler-shift or e-obs), then the `as.telemetry()` will first look for an HDOP value. If HDOP values are not found, then `as.telemetry()` will look for ambiguous DOP values, followed by PDOP and then GDOP values. If no DOP values are found, then `as.telemetry()` will look for the reported number of satellites, and apply model (S3.1) to approximate DOP values.

Number of Satellites

Some GPS data lack DOP values but do record the number of satellites used when triangulating location estimates. Therefore, as proxy for DOP, a positive, monotonically decreasing function of the number of satellites, N_{sat}, can be used. The simplest error variance relation would decay with the density of the GPS satellite constellation, which is proportional to $1/N_{\text{sat}}^2$. Furthermore, triangulation is not possible with $N_{\text{sat}} \leq 2$. Therefore, we can improve our simplest model to

$$\text{VAR}(N_{\text{sat}}) \propto \text{HDOP}(N_{\text{sat}})^2 = \left(\frac{N_{\text{max}}-2}{N_{\text{sat}}-2}\right)^2, \tag{S3.1}$$

where $N_{\text{max}} = 12$, which fixes the minimum HDOP value to 1 for convenience. Fig. S3.1 depicts the relationship between HDOP and N_{sat}, along with our best-fit model. GPS data was taken from 35 blue wildebeests (*Gorgon taurinus albojubatus*) and HDOP values were aggregated into RMS HDOP values and their variance, per value of N_{sat}, as non-linear least-squares regressions would not converge when using the full dataset (likely due to truncation error in the DOP values). To approximate an exact regression on the full dataset, we applied a weighted non-linear least squares regression to the RMS HDOP values under a log link. We considered all combinations of models with a power of 2 or unknown and a singularity at N_{sat} values of 0, 2, or unknown. In terms of both AIC and BIC, our simple model (S3.1) outperformed all others, save for the most complex model with both unknown power and unknown singularity. However, the more complex model could not maintain performance under different modeling choices—including HDOP summary statistic, regression weights, and link function—and its parameter estimates were not meaningful.

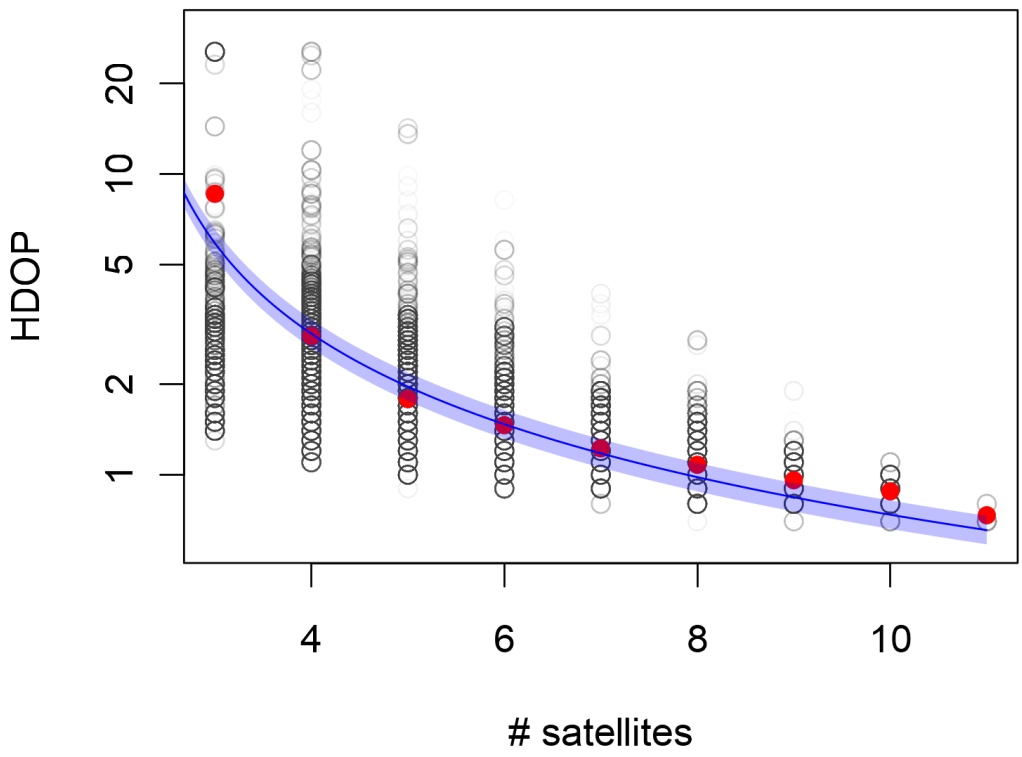

Figure S3.1: Horizontal dilution of precision (HDOP) versus number of satellites from 35 GPS collared wildebeests, with the RMS HDOP in red and approximate HDOP model (S3.1) in blue.

S3.2 Precision modifiers

Fix Type

Some GPS devices specify a categorical "fix type" that distinguishes location estimates obtained via different algorithms or with different characteristics. The most common fix-type levels are 2D versus 3D, where the 2D location estimates are calculated with only 3 satellites and lack corresponding altitude estimates. In principle, HDOP values are supposed to account for the diminished accuracy of 2D location estimates, however, some GPS devices require distinct 2D/3D RMS UERE values (App. S4). Furthermore, some GPS devices require distinct 2D/3D parameters, yet do not provide an explicit "GPS fix-type" column. In this case, a fix-type column can be calculated from the logical test $N_{\text{sat}} > 3$.

By default, `ctmm`'s `as.telemetry()` function will assume that any GPS fix-type column is meaningful, and create a corresponding location `class` column in the output `telemetry` data object. Model selection can then determine whether or not extra location classes are supported by the calibration data. `as.telemetry()` also checks if some location estimates lack corresponding speed, altitude, and/or DOP values, in which case location classes will be created to distinguish the level of missingness, as this often corresponds to different location estimation algorithms.

Time to fix

GPS devices take some amount of time to acquire satellite signals and corresponding satellite orbital information before estimating a location. Location estimates can either be calculated with the full orbital information from each satellite in communication, or approximate orbital information from fewer satellites. In some GPS devices, this change in location estimate quality occurs at a maximum recorded "time to fix" value, when the device terminates its attempt to obtain full orbital information and proceeds to an approximate calculation. If this approximation is not accounted for in the device's HDOP values, then separate "in time" and "timed out" RMS UERE values may be required (App. S4). Furthermore, as the level of approximation can vary considerably in "timed out" fixes, their UERE distribution may be heavy tailed.

In `ctmm`'s `as.telemetry()` function, the `timeout` argument can be used to create an additional location class for TTF values greater than or equal to the timeout value.

S4 Error-model selection examples

S4.1 Advanced Telemetry Systems

S4.1.1 ATS G2110e

We analyzed calibration data on nineteen Advanced Telemetry Systems (ATS) G2110e GPS/Iridium collars with two deployments each, where data were collected in 10 and 60 minute intervals for an average of 2.5 days. Before calibration, we removed all gross outliers, which constituted 0.3% of the total records. In our first round of analysis, we tested the veracity of the recorded HDOP values against our number-of-satellites model (S3.1) and homoskedastic location errors. We found the provided HDOP values to be more informative than nothing (homoskedastic errors), but inferior to our simple number-of-satellites model, which should not be the case. Performance of the initially best model, $\hat{\text{HDOP}}(N_{\text{sat}})$, was poor ($Z^2_{\text{red}} = 3.9$), with a large variation among devices. Upon further investigation into other data provided by the GPS devices, we notice a smooth, positive trend between the calibration residuals and time to fix. Therefore we also included powers of $(\text{TTF}/\text{TTF}_{\text{min}})$ in our HDOP model, with the selected model being a product of our number-of-satellites model and square TTF (Table S4.1), and where the expected RMS location error of this device is 2.4–2.5 meters under ideal conditions ($N_{\text{sat}} = 12$; $\text{TTF} = \text{TTF}_{\text{min}}$), which is small enough to indicate a higher quality dual-frequency GPS receiver or on-board Kálmán filtering.

To assess variability among GPS devices, we compared our best joint model to the same class of model, but with calibration individualized by GPS device and by deployment. In either case, the increase in performance was not substantial.

Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
$\hat{\text{HDOP}}(N_{\text{sat}}) \times \text{TTF}^2$	0.0	2.9
$\hat{\text{HDOP}}(N_{\text{sat}}) \times \text{TTF}^3$	816.4	3.3
$\text{HDOP} \times \text{TTF}^2$	1,432.7	3.2
$\hat{\text{HDOP}}(N_{\text{sat}}) \times \text{TTF}^1$	1,716.0	3.0
$\text{HDOP} \times \text{TTF}^3$	2,403.2	3.7
$\text{HDOP} \times \text{TTF}^1$	2,945.7	3.3
TTF^2	4,403.0	3.6
TTF^3	5,111.6	4.0
TTF^1	6,942.7	4.0
$\hat{\text{HDOP}}(N_{\text{sat}})$	7,285.0	3.9
HDOP	8,060.7	4.1
homoskedastic	1,4863.6	5.3

Table S4.1: AIC_C differences and reduced Z^2 values for error models fit to calibration data from nineteen ATS G2110e GPS collars with two deployments each.

S4.1.2 ATS G5-2D

We analyzed calibration data from four ATS G5-2D GPS/Iridium collars, with data collected in 10-minute intervals for 10-12 days. We pitted the reported HDOP values against our

number-of-satellites model (S3.1) and a model of homoskedastic errors. Performance of all three error models was very similar (Table S4.2), though HDOP values only ranged from 0.6 to 2.1, which implies a consistently good signal during calibration. RMS error under ideal conditions (HDOP=0.6) was estimated to be 3.5-3.6 meters. Variation among devices was not substantial.

Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
homoskedastic	0.0	1.5
HDOP	34.7	1.5
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	39.6	1.5

Table S4.2: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from four ATS G5-2D GPS collars.

In cases such as this, where the calibration data had consistently good reception yet the tracking data might not, we suggest performing error-model selection (but not parameter estimation) again with the tracking data. It could very well be the case that the HDOP or $\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$ error models prove superior when faced with poor satellite reception.

S4.1.3 ATS G10 UltraLITE

We analyzed calibration data from 15 ATS G10 UltraLITE GPS tags deployed in 4 environments—deciduous forest, evergreen forest, forested wetland, and clearing—with data collected in 10-minute intervals for 2-15 days. For both horizontal and vertical errors we tested the provided HDOP, PDOP, and VDOP values against our number-of-satellites model (S3.1) and a model of homoskedastic errors (Table S4.3). Horizontal errors were comparably modeled by either HDOP value or our $\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$ model, with the RMS error estimated to be 21.0–21.3 meters under ideal conditions ($N_{\mathrm{sat}} = 12$). However, VDOP values were slightly worse than all other predictors for vertical errors. There was some variability in horizontal calibration parameters, with Z^2_{red} dropping from 2.9 to 2.7 upon individual calibration. Differences in performance among habitats was not as significant.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	0.0	2.9	homoskedastic	0.0	2.4
HDOP	10.5	2.9	HDOP	27.6	2.4
PDOP	5,993.9	3.4	PDOP	176.6	2.4
homoskedastic	6,174.9	3.4	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	343.6	2.4
VDOP	16,133.3	4.3	VDOP	2,331.3	2.5

Table S4.3: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 15 ATS G10 UltraLITE GPS tags.

S4.2 Cellular Tracking Technologies

S4.2.1 CTT 1090

We analyzed test data from 3 Cellular Tracking Technologies (CTT) 1090 GPS-GSM tags, with data collected every 15 minutes for up to 36 hours. For both horizontal and vertical errors, we compared the performance of reported HDOP and VDOP values against a model of homoskedastic location errors after removing 3 gross outliers (Table S4.4). RMS horizontal location error was estimated to be 8.7–10.3 meters under ideal conditions (HDOP=1). Individualized calibration had some impact on performance, decreasing Z^2_{red} from 4.6 to 4.3.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}	Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
HDOP	0.0	4.6	HDOP	0.0	2.1
VDOP	175.2	6.9	homoskedastic	69.6	1.6
homoskedastic	191.8	6.4	VDOP	70.0	1.8

Table S4.4: AIC_C differences and reduced Z^2 values for error models fit to calibration data from three CTT 1090 GPS tags.

S4.2.2 CTT BT3

We analyzed calibration data from 8 Cellular Tracking Technologies (CTT) BT3 GPS tags, with data collected in 10 and 15 minute intervals for up to one month. For both horizontal and vertical errors, we compared the performance of the reported HDOP and VDOP values to our $\hat{\text{HDOP}}(N_{\text{sat}})$ model (S3.1) and a model of homoskedastic location errors (Table S4.5). There was some variability in horizontal calibration parameters, with Z^2_{red} dropping from 4.3 to 4.0 upon individual calibration.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}	Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
HDOP	0.0	4.3	$\hat{\text{HDOP}}(N_{\text{sat}})$	0.0	2.3
$\hat{\text{HDOP}}(N_{\text{sat}})$	1,528.0	4.7	HDOP	827.4	2.4
VDOP	3,685.0	5.1	VDOP	913.2	2.3
homoskedastic	4,300.5	5.0	homoskedastic	1,744.6	2.3

Table S4.5: AIC_C differences and reduced Z^2 values for error models fit to calibration data from five CTT BT3 GPS tags.

S4.2.3 CTT ES400

We analyzed calibration data from 15 Cellular Tracking Technologies (CTT) ES400 (4th Gen) GPS-GSM solar powered telemetry units, with data collected in 10- and 15-minute intervals for two weeks. For both horizontal and vertical errors, we compared the performance of the reported HDOP and VDOP values to our $\hat{\text{HDOP}}(N_{\text{sat}})$ model (S3.1) and a model of homoskedastic location errors (Table S4.6). Location errors in both dimensions were best modeled by our $\hat{\text{HDOP}}(N_{\text{sat}})$ model, though HDOP was a close second. Curiously, the

VDOP-based model was outperformed by both $\hat{\text{HDOP}}(N_{\text{sat}})$ and HDOP for vertical errors. RMS horizontal error was estimated to be 3.9–4.0 meters under ideal conditions (N_{sat}=12). There was slight variability among devices, with Z^2_{red} dropping from 2.4 to 2.3 upon individual horizontal calibration.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}	Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
$\hat{\text{HDOP}}(N_{\text{sat}})$	0.0	2.4	$\hat{\text{HDOP}}(N_{\text{sat}})$	0.0	1.3
HDOP	1,527.8	2.4	HDOP	732.9	1.3
homoskedastic	10,895.3	2.9	VDOP	4,059.6	1.4
VDOP	16,075.8	3.2	homoskedastic	4,138.7	1.5

Table S4.6: AIC_C differences and reduced Z^2 values for error models fit to calibration data from five CTT ES400 GPS collars.

S4.3 e-obs

e-obs devices generally do not provide DOP values, but a "horizontal accuracy estimate", which, according to e-obs documentation is an *in*accuracy estimate measured in meters. From personal communications with e-obs, the statistical interpretation of this information—in terms of a specific number of standard deviations or quantile—is unknown. Based on the following results, `ctmm`'s `as.telemetry` function imports e-obs horizontal accuracy estimates as HDOP values.

S4.3.1 e-obs generation-1 GPS collars (1059–1078)

We analyzed calibration data collected from 17 first-generation e-obs GPS collars encircled around a tree and sampled at 10 minute intervals over the course of a day. To confirm the veracity of the e-obs error estimates we performed a null hypothesis test, with a null model of homoskedastic location errors of unknown variance, and an alternative model with error standard deviations proportional to the e-obs error estimate by treating the e-obs error estimates as dimensionful HDOP values. We found the e-obs error estimates to be highly informative (table S4.7), and from these data the RMS "UERE" value was estimated to be 1.67 (1.63–1.70). RMS horizontal error was estimated to be 5.8–6.1 meters under ideal conditions. There was slight variability in the devices, with Z^2_{red} decreasing from 2.1 to 2.0 when using individualized horizontal calibration, and where collars with a higher percentage of missing data yielded smaller RMS UERE estimates.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}	Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
horizontal-accuracy	0.0	2.1	horizontal-accuracy	0.0	1.6
homoskedastic	2,442.9	3.1	homoskedastic	160.6	1.6

Table S4.7: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 17 e-obs GPS tags.

S4.3.2 e-obs generation-3 GPS tags (5081–6375)

We analyzed calibration data from 15 third-generation e-obs solar tags placed at a fixed location and sampled for 10-20 days with 5-20 minute sampling (depending on battery status) and 1-second sampling in direct sunlight for a minimum of 10 minutes. With these data, and using the same error model, the e-obs RMS UERE value was estimated to be 1.673 (1.672–1.674), consistent with our first-generation GPS collars and highly informative (Table S4.8). Variation among devices was not substantial.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}^\dagger$	Z^2_red	Model	$\Delta\text{AIC}_\text{C}^\dagger$	Z^2_red
horizontal-accuracy	0.0	1.3	horizontal-accuracy	0.0	1.1
homoskedastic	6,167,049	3.2	homoskedastic	2,430,763	1.8

Table S4.8: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 15 e-obs GPS tags. †Due to autocorrelation, the AIC_C differences are exaggerated by a factor on the order of a hundred.

Due to the presence of substantial autocorrelation in these data, which we did not account for in estimating the RMS UERE, both the point estimate and standard error are downwardly biased. We roughly approximate the influence of autocorrelation by inspection of the correlogram (Fig. S4.1), where we note that most of the autocorrelation is gone within 2 minutes of time lag. Following the analysis of Fleming et al. (2019), 10 days of sampling at 2-minute intervals would yield an effective sample size on the order of 7200 (10 days / 2 minutes), and a resulting bias of 1/7200, which is not substantial. The standard errors, however, are on the order of 11 ($\sqrt{2 \text{ minutes} / 1 \text{ second}}$) times too small. Therefore, a corrected RMS UERE estimate would be closer to 1.67 (1.66–1.68), which is still highly consistent with our first calibration results. Similarly, the RMS horizontal error would then be 1.70–1.72 meters under ideal conditions, which includes the data being collected at 1 hertz.

Inspection of the 1-second solar-tag error residuals and error estimates versus time since the unit was "cold" (not collecting 1-second fixes) revealed RMS location errors that shrink over time, from 30 meters initially to 3 meters after two minutes, which is evidence of on-board error-filtering of the location estimates. Likely related to the e-obs error filter, we found two further issues in the 1-second solar tag data not present in the 10-minute collar data. First, the e-obs error estimates were more substantially autocorrelated for a more prolonged period of time (Fig. S4.1). Second, the first few seconds of cold fixes had substantially underestimated location errors ($\sim$25%).

S4.3.3 e-obs generation-3 24-gram GPS tags (6577–6752)

We analyzed calibration data from 6 third-generation e-obs 24-gram GPS tags, with location fixes collected every 2 minutes for 3 days. Three of the tags were placed under forest canopy and three were placed on a roof. We compared the e-obs "horizontal accuracy estimate" to a model of homoskedastic errors. The e-obs horizontal accuracy estimate proved to be informative (table S4.9), but, in contrast to older e-obs models, the RMS UERE was estimated to be 0.94 (0.93–0.95). RMS horizontal error was estimated to be 2.6–2.7 meters under ideal conditions, which suggests a dual-frequency GPS receiver or on-board Kálmán

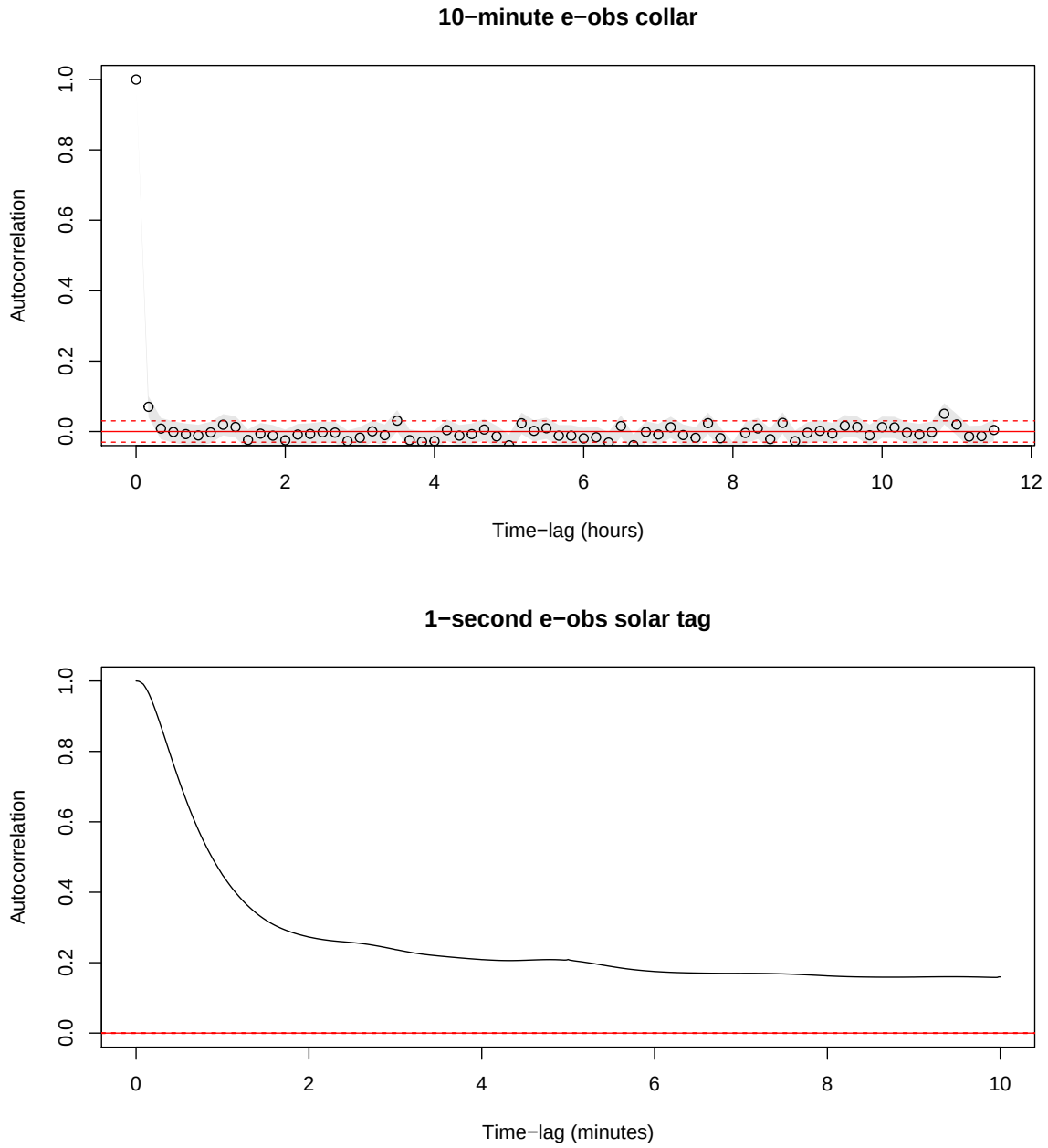

Figure S4.1: Joint correlograms of standardized calibration residuals from (**A**) 17 e-obs collars sampled at 10-minute intervals and (**B**) 15 e-obs solar tags sampled at 1-second intervals. Autocorrelation is not very substantial in the 10-minute calibration residuals ($< 10\%$), though it is significant at the first lag. In contrast, the 1-second e-obs data are substantially autocorrelated even after 10 minutes of time lag.

filter. There was substantial heterogeneity among devices, with $\mathrm{AIC_C}$ decreasing by 4,903.9 and Z^2_{red} dropping from 2.5 to 1.9 when using individiualized horizontal calibration. This heterogeneity could not be explained by habitat ($\Delta\mathrm{AIC_C}$ = 4,846.9).

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
horizontal-accuracy	0.0	2.5	homoskedastic	0.0	1.8
homoskedastic	14,234.7	3.8	horizontal-accuracy	614.9	1.8

Table S4.9: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 6 third-generation e-obs GPS tags.

S4.3.4 e-obs generation-3 42-gram GPS tag (7095–7106)

We analyzed calibration data from 5 third-generation e-obs 42-gram GPS tags, with location fixes collected every 2 minutes for 3 days. Three of the tags were placed under forest canopy and two were placed on a roof. We compared the e-obs "horizontal accuracy estimate" to recorded HDOP values, our number-of-satellites error model (N_{sat}; S3.1), and a model of homoskedastic location errors (Table S4.10). The e-obs "horizontal accuracy estimate" proved to be informative, with the RMS UERE estimated to be 0.99 (0.98–1.00). The RMS UERE value of ∼1 is close to some other third-generation e-obs calibration values and consistent with the interpretation of the e-obs "horizontal accuracy estimate" being an estimate of the root-mean-square location error. RMS horizontal error was estimated to be 2.5–2.6 meters under ideal conditions, which suggests a dual-frequency GPS receiver or on-board Kálmán filter. There was slight heterogeneity among devices, with Z^2_{red} dropping from 1.8 to 1.7 when using individualized calibration.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
horizontal-accuracy	0.0	1.8	horizontal-accuracy	0.0	1.6
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	12,123.4	3.3	HDOP	635.3	1.8
homoskedastic	12,166.1	3.2	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	1,119.8	1.8
HDOP	12,630.7	3.5	homoskedastic	1,520.5	1.8

Table S4.10: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 5 third-generation e-obs GPS tags.

S4.3.5 e-obs generation-3 GPS tag (7401–7405)

We analyzed calibration data from 3 third-generation e-obs GPS tags, with location fixes collected every half hour or hour for 3-6 days in four different environments—on an open deck, beneath a tree, indoors, and beneath a house. For both horizontal and vertical errors, we compared the e-obs "horizontal accuracy estimate" to ambiguous DOP values, our number-of-satellites error model (N_{sat}; S3.1), and a model of homoskedastic location errors (Table S4.11). Again, the e-obs "horizontal accuracy estimate" proved to be the most informative predictor of horizontal errors, but with the RMS UERE estimated to be 0.86

(0.84–0.88), which is smaller than that of previous e-obs devices. There was no substantial difference in performance when using per-device calibration, however there were substantial differences across habitat, with a clear trend of the RMS UERE increasing with signal quality, from 0.35–0.39 beneath the house to 0.95–1.00 in an open setting, which is similar to the trend we noticed in generation-1 e-obs GPS collars (App. S4.3.1). It appears that, while the e-obs error estimates are proportional to the location-error variance with good satellite reception, they overestimate variance with poor satellite reception. We leave further examination to future studies.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
horizontal-accuracy	0.0	2.5	DOP	0.0	1.5
DOP	675.3	2.8	homoskedastic	29.7	1.5
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	748.1	2.9	DOP	183.1	1.7
homoskedastic	924.1	3.0	horizontal-accuracy	214.9	1.8

Table S4.11: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 3 third-generation e-obs GPS tags.

S4.4 GlobalTop

S4.4.1 GlobalTop PA6H

We analyzed calibration data from a GlobalTop PA6H GPS module, with location fixes taken every second, overnight. The PA6H GPS module returned all relevant DOP values and the number of satellites, which we compared to a model of homoskedastic location errors (Table S4.12). In terms of goodness of fit, homoskedastic errors were the best performing and VDOP values were worst performing, even for vertical errors.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC}^{\dagger}_{\mathrm{C}}$	Z^2_{red}	Model	$\Delta\mathrm{AIC}^{\dagger}_{\mathrm{C}}$	Z^2_{red}
HDOP	0.0	1.5	homoskedastic	0.0	0.6
homoskedastic	4,897.2	1.3	HDOP	155.7	0.6
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	7,143.4	1.5	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	1,273.5	0.7
PDOP	11,333.7	1.5	PDOP	3,484.8	0.6
VDOP	17,776.3	1.5	VDOP	6,395.1	0.7

Table S4.12: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from a GlobalTop PA6H GPS module. Error models either consist of a single RMS UERE value, or a different RMS UERE for in-time and timed-out fixes. †Due to autocorrelation, the $\mathrm{AIC_C}$ differences are exaggerated by a factor on the order of a hundred.

S4.5 Lotek

S4.5.1 Lotek Lifecycle 330

We recovered opportunistic calibration data for Lotek Lifecycle 330 GPS collars from twelve deceased Mongolian gazelles, all sampled in 23-hour intervals for 5 days to 8 months after death. This model Lotek collar records a fix type of either 3D or 3D-V, but insufficient 3D fixes were obtained for calibrating both location classes. The only other location-error information included was an ambiguous DOP value, which we tested against a model of homoskedastic location errors. The homoskedastic error model proved to be more predictive than the DOP values (Table S4.13), but there was substantial variation among GPS collars, with Z^2_{red} dropping from 2.3 to 1.2 when using individualized horizontal calibration.

Model	$\Delta\text{AIC}_\text{C}$	Z^2_{red}
homoskedastic	0.0	2.3
DOP	50.9	2.5

Table S4.13: AIC_C differences and reduced Z^2 values for error models fit to opportunistic calibration data from 12 Lotek Lifecycle 330 GPS collars.

S4.5.2 Lotek PinPoint 240

In addition to ambiguous 'DOP' values, Lotek PinPoint 240 GPS tags record a "duration" column, corresponding to the number of seconds lapsed before the GPS fix is completed or "time to fix" (TTF). When using these tags to track wood turtles, a large number of outliers were observed wherein recorded locations implied biologically implausible speeds, even when an using error-informed speed estimator on calibrated data (c.f., Sec. 2.3). Upon further inspection it was noticed that most of these outliers had a maximized fix duration of 70 seconds, which motivated our choice of candidate models.

We performed model selection using calibration data from two tags left under a tree for a day, with 10-minute sampling intervals. In the candidate models we compared homoskedastic location errors, DOP values, and our number-of-satellites precision model (N_{sat}; Sec. S3.1). Furthermore, we also test for a larger RMS UERE value in the timed-out fixes. Our results are summarized in Table S4.14.

For the horizontal data, the combination of DOP value and distinct in-time/timed-out UERE parameters was the clear winner. The horizontal RMS UERE of the timed-out fixes was four times larger than that of the regular fixes, transforming most of the previously labeled outliers into normative data (Fig. S4.2). Results were slightly different for the vertical data, with our number-of-satellites precision model (N_{sat}; S3.1) outperforming the DOP model. This suggests that the ambiguous DOP value is likely an HDOP value, and demonstrates the utility of our N_{sat} model—not only is it capable of outperforming a homoskedastic error model, but it can also outperform co-opted DOP values. RMS horizontal error was estimated to be 8.0–8.7 meters under ideal conditions (HDOP=1 & in-time). For the horizontal errors, there was slight heterogeneity among devices, with Z^2_{red} dropping from 2.5 to 2.4 when using individualized calibration.

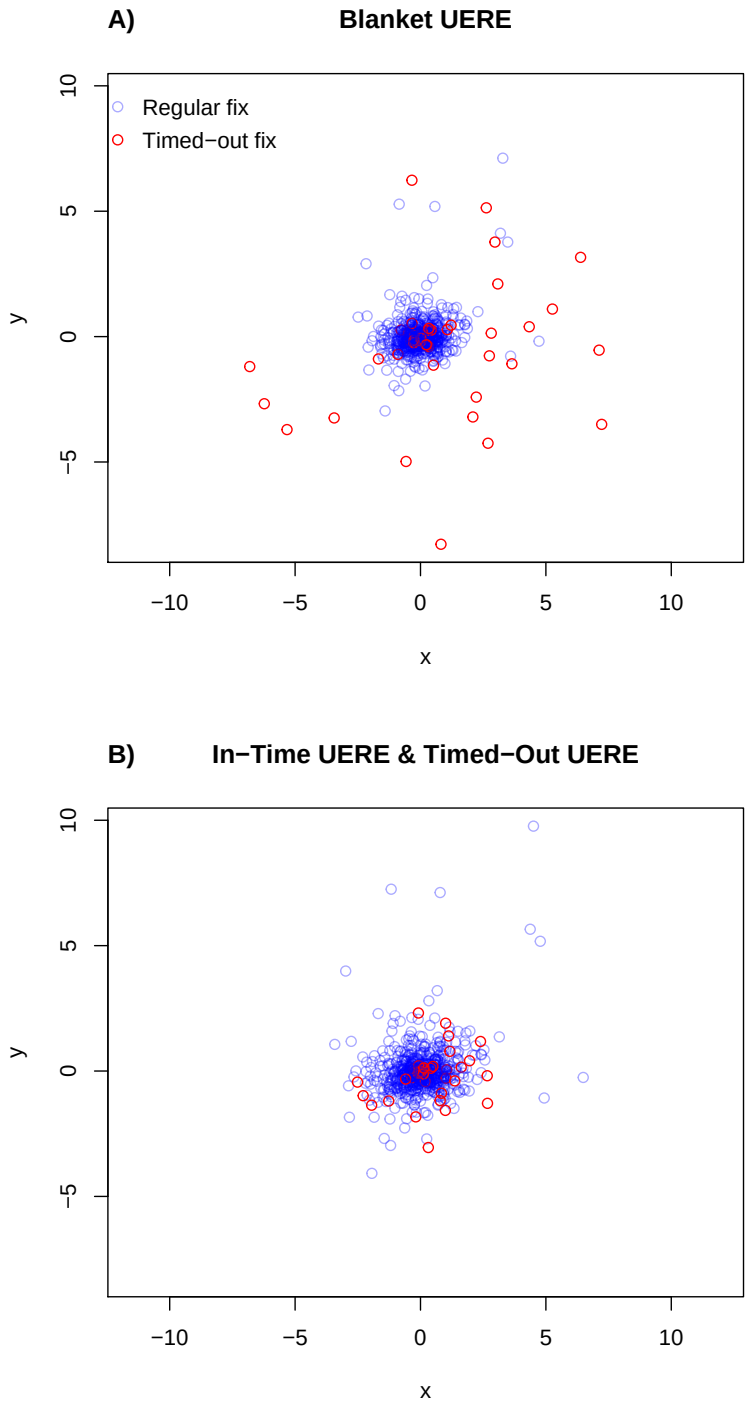

Figure S4.2: Residuals of calibration data for Lotek tags when **A**) considering all location fixes to share the same RMS UERE value and **B**) considering regular location fixes and timed-out location fixes as having distinct RMS UERE values. A shorter tailed residual distribution indicates a better performing error model.

Horizontal			Vertical		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
DOP & is(timeout)	0.0	2.5	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$ & is(timeout)	0.0	1.7
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$ & is(timeout)	571.0	3.2	DOP & is(timeout)	57.4	2.0
DOP	602.2	3.3	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	140.0	1.9
is(timeout)	1,416.6	5.1	is(timeout)	279.7	2.2
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	1,434.0	4.6	DOP	452.0	2.4
homoskedastic	3,724.0	9.6	homoskedastic	773.1	2.3

Table S4.14: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 2 Lotek PinPoint 240 GPS tags. Error models either consist of a single RMS UERE value, or a different RMS UERE for in-time and timed-out fixes.

S4.5.3 Lotek WildCell GPS-GSM

We collected over three days worth of hourly calibration data from five Lotek WildCell GPS-GSM collars. These data recorded an ambiguous DOP value and whether or not the location fix was "validated". The "validated" column proved to be marginally informative, with validated location fixes having 8.5–9.4 meter RMS horizontal error and non-validated fixes having 1–3 times that amount, but the DOP values were not found to be informative (Table S4.15. There was moderate heterogeneity among devices, with the horizontal Z^2_{red} dropping from 1.8 to 1.5 upon individualized calibration.

Horizontal			Vertical		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
is(validated)	0.0	1.8	DOP	0.0	1.3
homoskedastic	8.0	1.8	homoskedastic	1.4	1.3
DOP & is(validated)	91.5	2.1	DOP & is(validated)	2.0	1.4
DOP	102.6	2.1	is(validated)	3.2	1.3

Table S4.15: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 5 Lotek WildCell GPS-GSM tags. Error models either consist of a single RMS UERE value, or a different RMS UERE for validated and non-validated fixes.

S4.6 NTT Docomo

S4.6.1 NTT Docomo CTG-001G

We collected a weeks' worth of hourly calibration data from three NTT Docomo CTG-001G units. The CTG-001G is a cellular-assisted GPS device that can triangulate with cellular towers, using Japan's GSM equivalent—'Freedom of Mobile Multimedia Access' (FOMO) (Ishii et al., 2019). This device provided an error estimate in meters, which we tested against a model of homoskedastic location errors (Table S4.16). A location class was also provided, but the low-accuracy classes were too sparsely populated for analysis. We found the CTG-001G error estimates to be highly informative, and there was no substantial variation in device performance. The RMS location error was estimated to be 3.1–3.4 meters under ideal

conditions.

Model	$\Delta\text{AIC}_\text{C}$	Z^2_red
"error"	0.0	1.9
homoskedastic	1,835.8	5.7

Table S4.16: AIC_C differences and reduced Z^2 values for error models fit to opportunistic calibration data from three NTT Docomo CTG-001G GPS device.

S4.7 Ornitela

S4.7.1 Ornitela 20-gram tag (182902–182928)

We placed three 20-gram Ornitela GPS tags on a roof and three in a forest for 2 days, collecting data in 2-minute intervals. We tested joint error models using the reported HDOP values and number of satellites, but both of these models failed to outperform one of homoskedastic location errors (Table S4.17). Because the HDOP values and number of satellites were not informative, and because habitats differed among devices, individual calibrations were highly heterogeneous, with AIC_C decreasing 4,116.2 and Z^2_red dropping from 2.2 to 1.6. 99% of the AIC_C difference was solely due to habitat, with forested location errors being twice that of rooftop errors. In principle, HDOP values can account for signal loss due to a wooded environment, but, again, the stated HDOP values were not informative.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_red	Model	$\Delta\text{AIC}_\text{C}$	Z^2_red
homoskedastic	0.0	2.2	homoskedastic	0.0	1.6
HDOP	763.7	2.4	HDOP	374.6	1.5
$\hat{\text{HDOP}}(N_\text{sat})$	1,717.8	2.5	$\hat{\text{HDOP}}(N_\text{sat})$	526.2	1.7

Table S4.17: AIC_C differences and reduced Z^2 values for error models fit to calibration data from six 20-gram Ornitela GPS tags.

S4.7.2 Ornitela 25-gram tag (191771–191779)

We placed three 25-gram Ornitela GPS tags on a roof and three in a forest for 2 days, collecting data in 2-minute intervals. We tested joint error models using the reported HDOP values and number of satellites, but both of these models failed to outperform one of homoskedastic location errors (Table S4.18). The tags were very heterogeneous, as individualized calibration caused Z^2_red to drop from 3.4 to 2.1. Habitat alone could not explain these differences ($\Delta\text{AIC}_\text{C}$=2,357.8).

S4.8 Sirtrack

S4.8.1 Sirtrack Pinnacle Solar G5C275F

We recovered opportunistic calibration data for Sirtrack Pinnacle Solar G5C275F GPS-Iridium collars from three deceased Mongolian gazelles, all sampled hourly for 3-12 months after death. After a first round of calibration, pitting HDOP versus our number-of-satellites

Horizontal			Vertical		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
homoskedastic	0.0	3.4	homoskedastic	0.0	1.7
HDOP	334.0	3.6	HDOP	578.6	1.8
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	1,028.8	3.5	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	2,310.3	2.1

Table S4.18: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from six 25-gram Ornitela GPS tags.

model (N_{sat}; S3.1) and homoskedastic location errors, we noticed a precipitous drop in the RMS error at precisely 32 seconds of "time on", which we assume to be the 'time to fix' (TTF). Therefore, we included among our candidate models variants with separate location-estimate classes for TTF<32 seconds and TTF≥32 seconds (Table S4.19). TTF proved to be

Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
is(TTF<32 sec)	0.0	4.2
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$ & is(TTF<32 sec)	206.8	4.1
HDOP & is(TTF<32 sec)	1,707.6	4.3
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	11,343.1	4.8
homoskedastic	13,681.6	5.5
HDOP	13,810.6	5.0

Table S4.19: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to opportunistic calibration data from 3 Sirtrack GPS collars. Error models either consist of a single RMS UERE value or separate RMS UERE values for location estimates resolved before or after 32 seconds.

more informative than HDOP, which was not predictive in the Mongolian grasslands where the collars were used, and the median HDOP value was 1.2. The best-fit RMS UERE was 1.66 meters (1.63–1.68 meters) for location estimates with 32 seconds or more of "time on", and 6.56 meters (6.48–6.64 meters) for location estimates with less than 32 seconds of "time on". The sub-dekameter scale of the former location errors suggest on-board processing of the location estimates with sufficient "time on", even though the scheduled sampling interval was only one hour. Most likely, this device aggregates multiple location fixes into one reported estimate. Negligible heterogeneity was observed among devices.

S4.9 Technosmart

S4.9.1 Technosmart GiPSy 5

We collected calibration data from two Technosmart GiPSy 5 tags, both sampled in 10 second intervals for 30 minutes and 9 hours. Although this device came with informative HDOP values, they were outperformed by our number-of-satellites model (N_{sat}; S3.1), as summarized in table S4.20. RMS horizontal error was estimated to be 10.1–10.5 meters under ideal conditions (N_{sat}=12). Differences in tag calibration were not substantial.

Horizontal			Vertical		
Model	$\Delta\mathrm{AIC}_\mathrm{C}^\dagger$	Z^2_red	Model	$\Delta\mathrm{AIC}_\mathrm{C}^\dagger$	Z^2_red
$\hat{\mathrm{HDOP}}(N_\mathrm{sat})$	0.0	2.1	$\hat{\mathrm{HDOP}}(N_\mathrm{sat})$	0.0	1.5
HDOP	491.5	2.3	HDOP	656.8	2.0
homoskedastic	966.9	2.3	homoskedastic	953.4	1.8

Table S4.20: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 2 Technosmart GiPSy GPS tags. †Due to autocorrelation, the AIC_C differences are exaggerated by a factor on the order of ten.

S4.10 Telemetry Solutions

S4.10.1 TS Quantum 4000 Enhanced

Model Quantum 4000 Enhanced Telemetry Solutions (TS) GPS tags report an HDOP, number of satellites, and location fix type, which is either 2D or 3D depending on whether or not three or more more satellites were in reception. Both the 3D and 2D fixes are accompanied with HDOP values, and the 2D HDOP values were on-average larger than the 3D HDOP values, which is expected to be the case if the 3D and 2D HDOP values are on the same scale.

We performed model selection on calibration data from two tags left under a tree for a month, with 2-hour sampling intervals. We considered models with 1-2 RMS UERE values for the 3D/2D fix types, and we also pitted the HDOP values against homoskedastic location errors and our number-of-satellites precision model (N_sat; S3.1). Our results are summarized in Table S4.21. The highest performing model featured HDOP-informed errors, but different RMS UERE values for the 3D and 2D fixes. The 2D RMS UERE was approximately four times larger than the 3D RMS UERE. RMS horizontal error was estimated to be 6.5–7.4 meters under ideal conditions (HDOP=1 & 3D). There was no significant heterogeneity between the two devices.

Model	$\Delta\mathrm{AIC}_\mathrm{C}$	Z^2_red
HDOP & is(2D)	0.0	2.4
$\hat{\mathrm{HDOP}}(N_\mathrm{sat})$ & is(2D)	225.4	3.3
$\hat{\mathrm{HDOP}}(N_\mathrm{sat})$	248.6	3.6
HDOP	286.6	3.9
is(2D)	418.5	3.7
homoskedastic	637.3	4.7

Table S4.21: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 2 Telemetry Solutions GPS tags. Error models either consist of a single RMS UERE value or separate RMS UERE values for 2D and 3D location estimates.

S4.11 Telonics

"Gen4" Telonics data come standard with both an HDOP value and a "horizontal error" estimate in meters, with the two values not proportional and instead corresponding to er-

ror estimates of differing quality. Telonics documentation describes the new horizontal error estimate as a 100% confidence radius, but this is unlikely to be the case. In personal communications with Telonics, the specific quantile or number of standard deviations was unknown. Given that the Gen4 horizontal error estimates are measured in meters, we considered a model where they are proportional to the RMS horizontal error. Furthermore, Gen4 Telonics data contain two distinct location classes—regular GPS location fixes and "quick-fix points" (QFPs). The quick fixes do not come with a corresponding error estimate and Telonics documentation explicitly states that their RMS UERE value is larger than that of the ordinary fixes. Based on the following results, `ctmm`'s `as.telemetry` function ignores the Telonics Gen4 error estimate and instead uses HDOP values. Location classes are also designated, based on whether or not the location fix is a QFP.

S4.11.1 Telonics Gen4 GPS-VHF

To determine the calibration parameters of Telonics Gen4 GPS-VHF collars, we used opportunistic calibration data collected from one deceased lowland tapir (*Tapirus terrestris*) and two collars that detached from lowland tapir, all sampled hourly in the Pantanal for 2 to 25 days. We pitted the Telonics Gen4 "horizontal error" estimate against HDOP values, our number-of-satellites model (N_{sat}; S3.1), and homoskedastic errors (Table S4.22). The HDOP values proved to be the most predictive, with an RMS UERE of 6.2 meters (6.0–6.4 meters). However, the HDOP model was not substantially different from a model of homoskedastic location errors, and, surprisingly, the "horizontal error" estimates did not appear proportional to RMS location errors. RMS horizontal error was estimated to be 4.8–5.1 meters under ideal conditions (HDOP=0.8). Variation among devices was not substantial.

Only one of the three devices recorded QFPs. By fitting an error model with two RMS UERE values—one for regular location estimates and another for QFPs—we found the QPF RMS UERE to be about twice that of the regular RMS UERE, which is consistent with Telonics documentation.

Horizontal			**Vertical**		
Model	$\Delta\text{AIC}_\text{C}$	Z^2_red	Model	$\Delta\text{AIC}_\text{C}$	Z^2_red
HDOP	0.0	1.9	homoskedastic	0.0	1.3
$\hat{\text{HDOP}}(N_\text{sat})$	77.9	2.0	HDOP	52.9	1.3
homoskedastic	79.5	1.8	$\hat{\text{HDOP}}(N_\text{sat})$	58.0	1.3
horizontal-error	437.8	2.8	horizontal-error	418.0	1.8

Table S4.22: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 3 Telonics Gen4 VHF GPS collars.

S4.11.2 Telonics Gen4 GPS-Iridium

We performed a follow-up analysis with newer Telonics Gen4 GPS-Iridium collars, using opportunistic calibration data from three detached collars, all sampled hourly in the Cerrado for 2-3 days. Our model comparison resulted in the same model ranking for horizontal errors (Table S4.23), with the selected model based on HDOP, though the GPS-Iridium RMS UERE value of 6.9 meters (6.4–7.4 meters) was slightly larger than that of the GPS-VHF

model (6.0–6.4 meters). RMS horizontal error was estimated to be 2.6–3.0 meters under ideal conditions (HDOP=0.4). Variation among devices was not substantial.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
HDOP	0.0	1.7	HDOP	0.0	1.5
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	23.8	1.9	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	6.1	1.2
homoskedastic	49.7	2.0	homoskedastic	18.0	1.2
horizontal-error	171.7	2.6	horizontal-error	71.1	1.4

Table S4.23: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 3 Telonics Gen4 Iridium GPS collars.

S4.12 UniKN

S4.12.1 UniKN Logger GPS tag

We placed three UniKN Logger solar tags on a roof for two weeks, where under sunlight conditions they were able to obtain location fixes approximately every 16 minutes. This model UniKN Logger tag did not report an HDOP value, but it did report the number of satellites, with which we applied number-of-satellites precision model (S3.1). After removing one gross outlier, the best performing model was our number-of-satellites model (Table S4.24). RMS horizontal error was estimated to be 9.3–10.1 meters under ideal conditions (N_{sat}=12). There was no significant variability among devices.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	0	2.1	$\hat{\mathrm{HDOP}}(N_{\mathrm{sat}})$	0	1.1
homoskedastic	66.9	2.3	homoskedastic	57.0	1.2

Table S4.24: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 3 UniKN Logger GPS tags.

S4.13 Vectronic

S4.13.1 Vectronic GPS Plus

We collected testing data on 6 Vectronic GPS Plus collars, with fixes taken in 15-minute intervals for 2 hours. With less statistically efficient methods this might have been too little data to analyze, but with our minimum-variance unbiased (MVU) parameter estimators and $\mathrm{AIC_C}$ values, and the ability to pool the 6 datasets into a joint estimate, this was enough data to provide 50 degrees of freedom in the RMS UERE estimates. In addition to ambiguous DOP values, the devices also recorded an error estimate in meters, which was described in Vectronic's documentation as "the difference [m] between the real position and the transmitted position". We compared (ambiguous) DOP values and recorded error estimates (in meters) to a model of homoskedastic location errors, and found the on-board error estimates to be the most predictive (Table S4.25). Differences in performance among

devices were not very significant. The estimated RMS UERE for the recorded error estimates was 0.80–1.06, which is consistent with these being estimates of the RMS location error.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
"error"	0	1.5	"error"	0	0.9
DOP	31.4	1.7	DOP	39.6	1.2
homoskedastic	49.3	2.8	homoskedastic	46.4	2.0

Table S4.25: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 6 Vectronic GPS Plus collars.

S4.13.2 Vectronic Vertex Lite-3D GPS-Iridium

We collected calibration data on 18 Vectronic Vertex Lite-3D GPS-Iridium collars, with fixes taken in 15-minute and 4-hour intervals for 11 days. We compared the reported (ambiguous) DOP values to a model of homoskedastic location error, and found them to be only mildly informative (Table S4.26). RMS horizontal error was estimated to be 16.3–16.9 meters under ideal conditions (HDOP=0.8). There was slight variability among devices, with Z^2_{red} dropping from 2.5 to 2.4 upon individualized calibration.

Horizontal			**Vertical**		
Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}	Model	$\Delta\mathrm{AIC_C}$	Z^2_{red}
DOP	0	2.5	DOP	0	1.5
homoskedastic	513.0	2.6	homoskedastic	281.6	1.6

Table S4.26: $\mathrm{AIC_C}$ differences and reduced Z^2 values for error models fit to calibration data from 18 Vectronic Vertex Lite GPS collars.

S4.14 Vemco

Vemco positioning system (VPS) devices are analogous to GPS, but with acoustic communication to an array of fixed receivers instead of electromagnetic communication to a constellation of satellites, so that they can function underwater where GPS signals rapidly attenuate. VPS location estimates come with an 'HPE' value, which Vemco documentation describes as being analogous to GPS HDOP and requiring site specific calibration.

S4.14.1 Vemco HR2-V9 VPS

We attached 6 V9 180-kHz VPS tags to static moorings (metal poles fixed in cement anchors) and collected data over the span of 3 months in a Vemco HR2 high-residence-receiver network. The median sampling interval was 28 seconds, but some location estimates were obtained within a fraction of a second. Inspection of the residual correlogram (when using the HPE null model) revealed that approximately half of the autocorrelation decayed within minutes, which is similar in behavior to GPS. The remaining autocorrelation took on the order of a day to decay, which could be due to movement in the calibration tags, filtering in the VPS location estimates, or some other property of VPS fixes.

We considered three candidate error models: a model of homoskedastic location errors, a null model of HDOP = HPE, and a model of HDOP = RMSE, where RMSE was an unknown column in the data that we hoped might correspond to an RMS location-error estimate. We found HPE to provide the best error model (Table S4.27), with an RMS UERE of 5.8 meters. There was substantial variability among locations, with RMS UEREs ranging 3–9 meters and the reduced Z^2 statistic shrinking from 2.1 to 1.4 upon individualized calibration, but there was no significant dependence on depth.

Model	$\Delta\mathrm{AIC}_\mathrm{C}^\dagger$	Z^2_red
HPE	0.0	2.1
homoskedastic	7.5×10^6	5.3
RMSE	5.4×10^7	64.0

Table S4.27: AIC_C differences and reduced Z^2 values for error models fit to calibration data from 6 VPS tags. †Due to autocorrelation, the AIC_C differences are exaggerated by a factor on the order of ten.

S5 Error-informed statistics for outlier estimation

S5.1 Distance statistics for outlier estimation

First we consider a location sample $\mathbf{r} = (x, y)$ and corresponding displacement vector $\mathbf{d} = \mathbf{r} - \boldsymbol{\mu}_0$, as measured from a reference point $\boldsymbol{\mu}_0$. For the reference point $\boldsymbol{\mu}_0$ we consider the geometric median, which can be robustly estimated, and we ignore its relatively small $\mathcal{O}\big(n^{-1/2}\big)$ error, but other choices are also possible. In considering a normal distribution for the sampled location

$$\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\,\mathbf{I})\,, \tag{S5.1}$$

with unknown true location $\boldsymbol{\mu}$ and known location error variance σ^2, which we do not assume to be stationary in time, the distribution of the displacement vector $\mathbf{d}$ is then distributed according to

$$\mathbf{d} \sim \mathcal{N}(\boldsymbol{\Delta}, \sigma^2\,\mathbf{I})\,, \qquad\qquad \boldsymbol{\Delta} = \boldsymbol{\mu} - \boldsymbol{\mu}_0\,, \tag{S5.2}$$

where $\boldsymbol{\Delta}$ is the true displacement vector. Choosing a coordinate system with the x-axis parallel to $\boldsymbol{\Delta}$, our distribution can be represented in polar coordinates with $\mathbf{d} = (d\cos\theta, d\sin\theta)$ as

$$p(d,\theta) = \frac{e^{-\frac{(d\cos\theta-\Delta)^2+(d\sin\theta)^2}{2\sigma^2}}}{2\pi\sigma^2}\,, \tag{S5.3}$$

$$= \frac{e^{-\frac{d^2+\Delta^2}{2\sigma^2}}}{2\pi\sigma^2} e^{+\frac{d\Delta}{\sigma^2}\cos\theta}\,, \tag{S5.4}$$

where the circular error assumption is critical to obtaining this tractable relation. The marginal distribution of d is then

$$p(d|\Delta) = \int_{-\pi}^{+\pi} p(d,\theta)\,d\theta = \frac{e^{-\frac{d^2+\Delta^2}{2\sigma^2}}}{2\pi\sigma^2}\,I_0\!\left(\frac{d\Delta}{\sigma^2}\right)\,, \tag{S5.5}$$

where I_n is the modified Bessel function of the first kind. The log-likelihood function is then given by

$$\ell(\Delta|d) = \log(2\pi\sigma^2) - \frac{d^2}{2\sigma^2} - \frac{\Delta^2}{2\sigma^2} + \log I_0\!\left(\frac{d\Delta}{\sigma^2}\right)\,, \tag{S5.6}$$

and so by differentiation the maximum likelihood estimate $\hat{\Delta}$ of the true distance Δ satisfies the transcendental equation

$$\frac{d\hat{\Delta}}{\sigma^2}\,I_0\!\left(\frac{d\hat{\Delta}}{\sigma^2}\right) = \frac{d^2}{\sigma^2}\,I_1\!\left(\frac{d\hat{\Delta}}{\sigma^2}\right)\,, \tag{S5.7}$$

or equivalently

$$z\,I_0(z) = w\,I_1(z)\,, \tag{S5.8}$$

where $z = d\hat{\Delta}/\sigma^2$ and $w = d^2/\sigma^2$. We solve this equation via Newton-Raphson iteration, expanding the Bessel functions around the current best estimate. It can be shown that $\hat{\Delta} < d$, as should result from the inclusion of a location-error model, and $\lim_{d \gg \sigma} \hat{\Delta} = d$, which also must be the case. Interestingly, for all $d^2 < 2\sigma^2$, $\hat{\Delta} = 0$.

Finally, from the Hessian of the log-likelihood function, and after some algebraic simplification using (S5.7), we have the uncertainty estimate

$$\hat{\mathrm{VAR}}[\hat{\Delta}] = \frac{\sigma^2}{2 - \frac{d^2 - \Delta^2}{\sigma^2}}\,, \tag{S5.9}$$

which becomes larger than $\sigma^2/2$ as the improved distance estimate $\hat{\Delta}$ becomes smaller than the naive distance estimate, d.

S5.2 Speed statistics for outlier estimation

For speed-based outlier detection, we want to avoid accidentally ruling out realistically fast displacements during which motion would be ballistic. Therefore, it is sufficient to consider speed as distance $\Delta = |\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1|$ divided by time, ignoring tortuosity. To condition on as few locations as possible, we only consider sequential pairs. For the distance estimate, we use the results of the previous section, where the variance is now the sum of the variances of the two locations, or $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Otherwise, all relations there are the same.

In dividing distance by the sampled time $\hat{t} = \hat{t}_2 - \hat{t}_1$, there are situations where roundoff or truncation errors in t result in large values of the distance/time estimate, with otherwise usable data (permitting a location-error model). To accommodate truncated times, it is easier to model the imparted location error, when considering the rounded time as true, than it is to directly consider the temporal error caused by rounding. In the `outlie` method, we use a greatest common divisor routine to attempt to automatically detect the temporal sampling resolution δ, which is usually 1 second, but frequently a minute, hour, or day. Then we count the maximum $M = \max M_i$ of the number of location fixes M_i per truncated time $\hat{t}_i$. The minimum time to fix is then estimated by δ/M, which we subsequently use in place of zero for our speed estimates. Note that none of these crude estimates are used in movement analysis, but only for outlier detection.

Thus far, we have estimated speeds over the intervals, such as (t_1, t_2) and not specifically at the sampled times. To clean the data, we assign the highest speeds to the most suspect times. We proceed in ordered fashion from the smallest speed estimate to the largest speed estimate and assign each speed estimate to the adjacent time with the larger adjacent speeds, with the newer (larger) speed estimates overwriting the older (smaller) speed estimates. At the endpoints we compare to both the first (adjacent) times and the second times out. For convenience, we precede this loop by a reverse-ordered loop with reverse (low-speed) assignment to ensure the low-speed times are populated as well. If any outliers are removed, the speeds can be re-estimated to test for consecutive outliers. This speed assignment heuristic has substantial advantages over simpler heuristics, such as assigning the average adjacent speed estimate, which can result in false positives, or assigning the minimum adjacent speed speed estimate, which can result in false negatives.

S6 Errors in variograms

S6.1 Fast variogram of errors

For evenly sampled data $z(t)$, the empirical variogram is given by

$$\hat{\gamma}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_2 - t_1 = \tau} |z(t_2) - z(t_1)|^2 \,, \tag{S6.1}$$

here in one dimension and where $n(\tau) = \sum_{t_2-t_1=\tau}$ denotes the number of location pairs with time-lag τ between them. If location errors are additive, mean zero and uncorrelated, then the "observed" empirical variogram decomposes into the semi-variance function of the true movement plus the semi-variance of the location error.

$$\gamma_{\text{obs}}(\tau) = \gamma_{\text{move}}(\tau) + \gamma_{\text{error}}(\tau) \,, \tag{S6.2}$$

The data provides us with $\gamma_{\text{obs}}(\tau)$, and the fitted movement model provides us with $\gamma_{\text{move}}(\tau)$. Therefore, it is useful to have an estimate of $\gamma_{\text{error}}(\tau)$ to finalize our comparison between the model and data.

The error variogram is given by

$$\hat{\gamma}_{\text{error}}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_2 - t_1 = \tau} |\text{error}(t_2) - \text{error}(t_1)|^2 \,, \tag{S6.3}$$

where the location errors are unknown, but can safely be considered as independent if the data are not sampled at exceptionally high frequencies (i.e., $\geq 1/\text{min}$). Therefore, for positive time lags we have simply

$$\hat{\gamma}_{\text{error}}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_2 - t_1 = \tau} (\text{VAR}[\text{error}(t_2)] + \text{VAR}[\text{error}(t_1)]) \qquad (\tau > 0), \tag{S6.4}$$

which we can estimate from DOP and RMS UERE values.

Naively, relation (S6.4) involves a sum for each lag, which would imply an $\mathcal{O}(n^2)$ computational cost for n sampled locations. In (Marcotte, 1996), an $\mathcal{O}(n \log n)$ algorithm is given for the ordinary variogram (S6.1), by leveraging the fast Fourier transform (FFT). Here, we provide an $\mathcal{O}(n \log n)$ algorithm FFT algorithm for the error variogram, under the assumption that we have the timeseries VAR[error(t)]. The trick begins by including an indicator function $\chi(t)$ that evaluates to 1 if time t is sampled and 0 otherwise.

$$\hat{\gamma}_{\text{error}}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_2 - t_1 = \tau} (\chi(t_1)\text{VAR}[\text{error}(t_2)] + \chi(t_2)\text{VAR}[\text{error}(t_1)]) \qquad (\tau > 0). \tag{S6.5}$$

Then we may express the constrained sum with a Kronecker delta function as

$$\hat{\gamma}_{\text{error}}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_1, t_2} (\chi(t_1)\text{VAR}[\text{error}(t_2)] + \chi(t_2)\text{VAR}[\text{error}(t_1)])\, \delta_{t_2 - t_1 = \tau} \quad (\tau > 0). \tag{S6.6}$$

Next, by expanding the Kronecker delta functions in a suitable harmonic basis (Péron et al., 2016), we have for positive lags

$$\hat{\gamma}_{\text{error}}(\tau) = \frac{1}{2\,n(\tau)} \sum_{t_1,t_2} \left(\chi(t_1)\text{VAR}[\text{error}(t_2)] + \chi(t_2)\text{VAR}[\text{error}(t_1)]\right) \frac{1}{N} \sum_f e^{2\pi\imath f(t_2-t_1-\tau)}\,, \tag{S6.7}$$

$$= \frac{1}{2\,n(\tau)} \frac{1}{N} \sum_f \left(\sum_{t_1} \chi(t_1)\, e^{-2\pi\imath f t_1} \sum_{t_2} \text{VAR}[\text{error}(t_2)]\, e^{+2\pi\imath f t_2} + \text{conj.} \right) e^{-2\pi\imath f \tau}\,, \tag{S6.8}$$

$$= \frac{1}{n(\tau)} \text{DFT}^{-1}\{\text{Re}[\text{DFT}\{\chi\}^* \,\text{DFT}\{\text{VAR}[\text{error}]\}]\}\,, \tag{S6.9}$$

in terms of discrete Fourier transform operations.

S6.2 Error-adjusted variogram

Here, instead of estimating the influence of error on the conventional variogram, we provide a semi-variance estimator that accounts for error. Because our error adjustment cannot be implemented with fast Fourier transforms, we consider the weighted variogram

$$\hat{\gamma}(\tau) = \frac{\sum_{t_i - t_j = \tau} w_{ij}\, |z(t_i) - z(t_j)|^2}{2 \sum_{t'_i - t'_j = \tau} w_{i'j'}}\,, \tag{S6.10}$$

which is more robust to irregular sampling (Fleming et al., 2014*a*; Fleming and Calabrese, 2015). It is straightforward to show that this estimator maximizes the χ^2 log-likelihood function

$$\ell(\gamma(\tau)) = -\frac{1}{2} \sum_{t_i - t_j = \tau} w_{ij} \left(\log 2\pi\gamma(\tau) + \frac{\frac{1}{2}\,|z(t_i) - z(t_j)|^2}{\gamma(\tau)} \right). \tag{S6.11}$$

Therefore, we can generalize this model to erroneous observations $\hat{z}$ with observation error.

$$\ell(\gamma(\tau)) = -\frac{1}{2} \sum_{t_i - t_j = \tau} w_{ij} \left(\log 2\pi\,(\gamma(\tau) + E_{ij}) + \frac{\frac{1}{2}\,|\hat{z}(t_i) - \hat{z}(t_j)|^2}{\gamma(\tau) + E_{ij}} \right), \tag{S6.12}$$

$$E_{ij} = \frac{1}{2}\left(\text{VAR}[\text{error}(t_i)] + \text{VAR}[\text{error}(t_j)]\right). \tag{S6.13}$$

The maximum-likelihood estimate then satisfies the relation

$$\hat{\gamma}(\tau) = \frac{\sum_{t_i - t_j = \tau} \frac{w_{ij}}{\left(1 + \frac{E_{ij}}{\hat{\gamma}(\tau)}\right)^2} \frac{1}{2}\,|\hat{z}(t_i) - \hat{z}(t_j)|^2}{\sum_{t_i - t_j = \tau} \frac{w_{ij}}{1 + \frac{E_{ij}}{\hat{\gamma}(\tau)}}}\,, \tag{S6.14}$$

which can be used to solve for $\hat{\gamma}(\tau)$ iteratively. An example calculation is given in Fig S6.1.

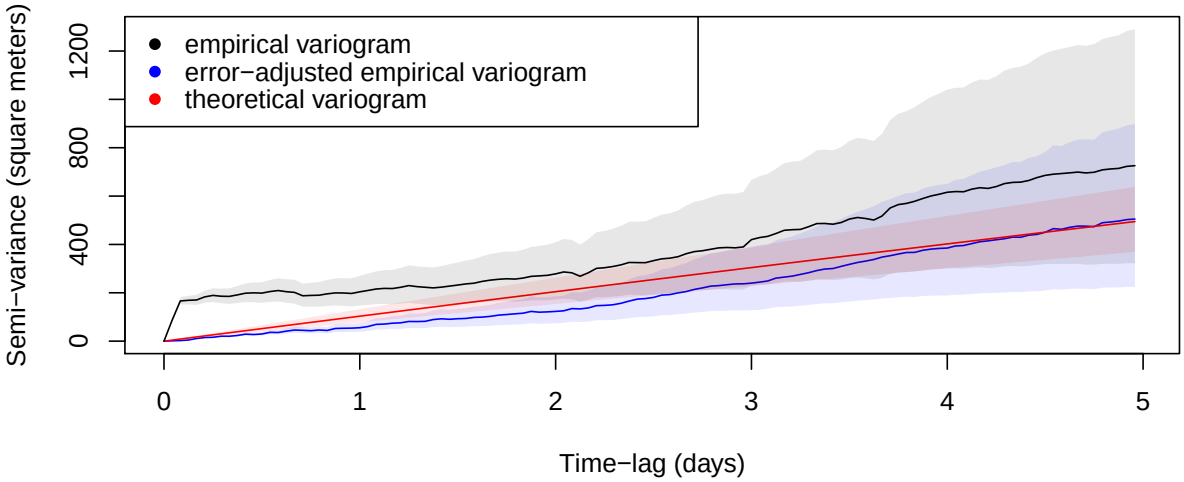

Figure S6.1: Conventional (black) and error-adjusted (blue) empirical variograms and AIC-best theoretical semivariance function (red) for a GPS-tracked wood turtle. The variogram is an unbiased measure of autocorrelation structure that estimates the average square displacement (y-axis) between two locations that are some time lag apart (x-axis), modulo a factor of $1/4$ in two dimensions. In the conventional variogram (black) this square displacement is a combination of movement and location error, which causes an initial discontinuity or 'nugget effect'. In this case, two locations sampled closely together in time will be almost $\sqrt{4 \times 200 \text{ m}^2} \approx 28$ m apart (RMS average), because of substantial location error from underbrush and canopy. The error-adjusted variogram, directly based on the data, and theoretical semivariance function, based on the AIC-best movement model, both agree that it actually takes the wood turtle ∼3 days, on average, to be found this far from the previous location.

S7 Delta approximation for fixed error estimates

Here we discuss an alternative method of integrating calibration and movement data. If we trust the movement model enough to allow our tracking data to update both the movement parameter estimates $\boldsymbol{\theta}$ and the error parameter estimates $\boldsymbol{\phi}$, then the joint likelihood is simply the product of the calibration and tracking data likelihoods

$$\ell_{\text{joint}}(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{X}_{\text{track}}, \mathbf{X}_{\text{cal}}) = \ell_{\text{track}}(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{X}_{\text{track}}) + \ell_{\text{cal}}(\boldsymbol{\phi}|\mathbf{X}_{\text{cal}})\,. \tag{S7.1}$$

However, if we do not trust our movement model enough to allow it to update the error model, then we can instead draw error parameters from the sampling distribution associated with ℓ_{cal} and treat them as fixed values within ℓ_{track}. The log-likelihood at a given value of $\hat{\boldsymbol{\phi}}$ is asymptotically given by

$$\ell_{\text{track}}\big(\boldsymbol{\theta}\big|\boldsymbol{\phi}\big) = \ell_{\text{track}}\Big(\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})\big|\boldsymbol{\phi}\Big) + \frac{1}{2}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})\right)^{\mathrm{T}} \boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})\big|\boldsymbol{\phi}\Big)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})\right), \tag{S7.2}$$

which is maximized at $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$, but we do not know the true parameters $\boldsymbol{\phi}$. Next considering the random variable $\hat{\boldsymbol{\phi}}$, we have the asymptotic expansion

$$\begin{aligned}\ell_{\text{track}}\big(\boldsymbol{\theta}\big|\boldsymbol{\phi}\big) = \ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big) &+ \left(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\right)^{\mathrm{T}} \boldsymbol{\nabla}_{\phi}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big) \\ &+ \left(\boldsymbol{\phi}-\hat{\boldsymbol{\phi}}\right)^{\mathrm{T}} \boldsymbol{\nabla}_{\phi}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\right) \\ &+ \frac{1}{2}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\right)^{\mathrm{T}} \boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\right),\end{aligned} \tag{S7.3}$$

which upon maximization gives us the asymptotic relation

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\phi}) = \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}}) + \left[\boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\right]^{-1}\left[\boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\phi}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\right]\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right). \tag{S7.4}$$

The uncertainty in $\hat{\boldsymbol{\phi}}$ contributes unbiased error to the final estimate. Therefore, our final point estimate is the same as when considering our point estimate $\hat{\boldsymbol{\phi}}$ to be the true value, but the covariance is increased according to

$$\begin{aligned}&\mathrm{COV}\Big[\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})\Big] = \mathrm{COV}\Big[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\Big] + \\ &\mathrm{COV}\Big[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\Big]\left[\boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\phi}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\right]\mathrm{COV}\Big[\hat{\boldsymbol{\phi}}\Big]\left[\boldsymbol{\nabla}_{\phi}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\right]\mathrm{COV}\Big[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\Big],\end{aligned} \tag{S7.5}$$

where the naive covariance is given by

$$\mathrm{COV}\Big[\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\phi}})\Big] = -\left[\boldsymbol{\nabla}_{\theta}\boldsymbol{\nabla}_{\theta}^{\mathrm{T}}\,\ell_{\text{track}}\Big(\boldsymbol{\theta}(\hat{\boldsymbol{\phi}})\big|\hat{\boldsymbol{\phi}}\Big)\right]^{-1}, \tag{S7.6}$$

which treats the error estimate as the truth. In conclusion, we can estimate our movement parameters under the assumption that the error parameters are known, calculate the cross derivatives of the log-likelihood function with respect to both movement and error parameters, and with this information account for error uncertainty in the movement parameter estimates without updating them. This is asymptotically equivalent to drawing values of $\hat{\boldsymbol{\phi}}$ from the calibration sampling distribution, and then repeatedly estimating movement parameters conditional on those values.