A framework to classify error in animal-borne technologies

Burns, Z.T.; van Loon, E.E.

Published in:
Frontiers in Ecology and Evolution

DOI:
10.3389/fevo.2015.00055

Citation for published version (APA):
A framework to classify error in animal-borne technologies

Zackory T. Burns¹ and E. Emiel van Loon¹, ²*

¹ Department of Zoology, University of Oxford, Oxford, UK, ² Computational Geo-Ecology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands

The deployment of novel, innovative, and increasingly miniaturized devices on fauna to collect data has increased. Yet, every animal-borne technology has its shortcomings, such as limitations in its precision or accuracy. These shortcomings, here labeled as “error,” are not yet studied systematically and a framework to identify and classify error does not exist. Here, we propose a classification scheme to synthesize error across technologies, discussing basic physical properties used by a technology to collect data, conversion of raw data into useful variables, and subjectivity in the parameters chosen. In addition, we outline a four-step framework to quantify error in animal-borne devices: to know, to identify, to evaluate, and to store. Both the classification scheme and framework are theoretical in nature. However, since mitigating error is essential to answer many biological questions, we believe they will be operationalized and facilitate future work to determine and quantify error in animal-borne technologies (ABT). Moreover, increasing the transparency of error will ensure the technique used to collect data moderates the biological questions and conclusions.

Keywords: animal-borne device, bio-logging, calibration, error, observation model

The Necessity of Animal-borne Technologies

Animal-borne technologies (ABT) collect high-resolution datasets to investigate a wide-range of functional and mechanistic biological questions. From quantifying energy expenditure in migrating geese (Branta leucopsis; Butler et al., 1998) to measuring the sociality of wild tool-using crows (Rutz et al., 2012), ABT are resolving questions at a steadily increasing resolution (c.f. Shillinger et al., 2012). In addition, ABT enable measurements on difficult to observe fauna, such as examining sea turtle feeding and breathing at sea (Okuyama et al., 2012), exploring diving behaviour of American mink (Hays et al., 2007) or observing the conservation and social dynamics of reintroduced wolves in Yellowstone National Park (Canis lupis; Fritts et al., 1997). Furthermore, ABT have mapped the spatial associations and exchange of tuberculosis between cattle (Bos primigenius) and badgers (Meles meles), demonstrating the importance of ABT to inform public policy (Böhm et al., 2009).

As quantification progresses in an increasing number of species, and as we collect more complete and robust data, the cycle of ABT stimulating new biological questions alongside biological questions driving innovative ABT will increase. Compared to the mid-twentieth century, ABT have weaved their way into the core of animal behavior, ecology, and conservation (Brown et al., 2013). We believe this broadening will continue to expand the use of ABT in biological research, leading to a strong reliance on ABT to resolve prevalent questions. As more studies utilize ABT and a considerable amount of effort is invested in storing and combining various kinds of data, ensuring...
data collection protocols consider limitations, and take into account measurement error becomes imperative (c.f. Krause et al., 2013).

**Error in Animal-borne Technologies**

Understanding and quantifying error is essential for discovery. Random, or systematic, error (see Table 1 for definitions) obscures the relationship between two variables, making it difficult, or even impossible, to collect sufficient evidence to reject a null hypothesis. Biases can, when unrecognized, lead to the acceptance of incorrect hypotheses and/or the identification of erroneous models. And even when error is recognized and properly modeled, lack of precision decreases the power of a study and increases the likelihood of missing a significant outcome. Thus, identifying sources of error *a priori* is essential to maximize the outcome of a study. While some sources of error have been investigated, such as the environmental effects on Argos (Patterson et al., 2010) or acoustic tags (Ehrenberg and Steig, 2002), the impact of the physical size of devices on Penguins’ swimming performance (e.g., speed recorders on *Spheniscus demersus* by Wilson et al., 1986; time-depth recorders on *Eudyptes schlegeli* by Hull, 1997), or the variation of GPS location accuracy and fix rate with vegetation and antenna position (Jiang et al., 2008), we believe and concur with other studies (see Frair et al., 2010) that most sources of possible error have neither been identified nor quantified.

We believe that the lack of error quantification in ABT studies is due to four factors. First, the study of device error will not directly lead to finding exciting biological phenomena: in fact, researching error comes at the cost of “real” discoveries in terms of time and resources. Moreover, studies at the interface of technology and biology could not be easily published due to the dearth of methodological journals. Secondly, progress in observation frequency and accuracy of ABT are so big relative to conventional methodologies (e.g., observations with binoculars) that device error is not the primary concern. Rather, sample size, automated processing, and correct analytical techniques for ecological interpretation were given a higher priority. Thirdly, computing power and statistical methods have only become available in the last decade. Finally, a framework to explore and quantify possible errors in ABT has been lacking.

While we believe that the first two causes for the limited attention are disappearing as new interdisciplinary and methodological journals have been founded (e.g., Journal of the Royal Society Interface or Methods in Ecology and Evolution), a general framework to discuss and analyze error is still lacking, though some previous studies focus specifically on error within a single device, such as GPS (Frair et al., 2010) or accelerometers (Shamoun-Baranes et al., 2012).

### A New Framework for Error

*Prima facie*, each technology has a unique set of pathways by which error is introduced; we believe that these errors can be categorized by the processes that introduce the variation. We propose that there are three fundamental concepts that, when used to classify error, enable a comprehensive evaluation of error, regardless of technology: (1) the basic physical properties underlying data collection; (2) the conversion of raw data into useful variables; and (3) the subjectivity in the selected settings. We believe the use of this classification system will facilitate future work to recognize additional sources of error and, subsequently, to propose statistical methods to mitigate and/or quantify error. Here, we outline and discuss this classification scheme; we do not attempt to review the literature exhaustively, as we discuss numerous technologies to demonstrate the breadth of technologies that would benefit from this categorization.

<table>
<thead>
<tr>
<th><strong>TABLE 1</strong></th>
<th>Definitions.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acoustic device</strong>: a device commonly used in aquatic environments, working through the transmission of a sound signal that is received subsequently by an array of hydrophone receivers. The location of the device can be estimated from the relative arrival times to the receivers. The sound signal is unique for each device so that each device can be uniquely identified.</td>
<td></td>
</tr>
<tr>
<td><strong>Encounter mapping devices (EMD)</strong>: a range of devices, such as GPS, proximity, or acoustic, used to map associations either directly or indirectly between objects.</td>
<td></td>
</tr>
<tr>
<td><strong>Random error</strong>: unpredictable fluctuations in observations due to the precision limitations of the measurement device, variations in the measurement procedure that is applied and/or heterogeneous environmental conditions. Random errors are on average zero. The variance of the fluctuations can be estimated but it cannot be corrected for.</td>
<td></td>
</tr>
<tr>
<td><strong>Systematic error</strong>: systematic deviations of the observation from the quantity that is meant to be observed due to the functioning of the measurement device, the measurement procedure that is applied and/or the environmental conditions. Systematic error can, when understood, be corrected for by, e.g., a calibration equation.</td>
<td></td>
</tr>
<tr>
<td><strong>Accelerometer</strong>: a device that measures the acceleration experienced by a test mass at rest in a reference frame in a single or multiple dimensions. Multi-axis acceleration is able to detect magnitude and direction of the acceleration as a vector quantity and can be used to sense orientation. An accelerometer measures change in its own motion (i.e., not in the location of the device within the environment (as done by devices based on remote sensing).</td>
<td></td>
</tr>
<tr>
<td><strong>Global Positioning System (GPS)</strong>: a satellite navigation system developed and managed by the United States military, consisting of 24 satellites above the earth and transmitting ultra-high frequency radio signals to receivers. The receivers compute longitude, latitude, and velocity by calculating the difference in the time signals that are received from at least four different satellites.</td>
<td></td>
</tr>
<tr>
<td><strong>Radio frequency identification (RFID) system</strong>: an automatic identification and data capture system comprised of one or more readers and one or more transponders (tags) in which data transfer is achieved by means of suitably modulated inductive or radiating electromagnetic carriers. An RFID reader transmits an encoded radio signal to interrogate the tag. The RFID tag receives the message and then responds with its identification and other information.</td>
<td></td>
</tr>
<tr>
<td><strong>Proximity sensor</strong>: an electronic device that detects and signals the presence of a selected object. When used in association with a radio frequency identification system the sensor is set up to detect the presence of a tagged or transponder carrying object when it enters the vicinity of the reader/interrogator so that the reader can then be activated to effect a read.</td>
<td></td>
</tr>
</tbody>
</table>
Basic Physical Properties Underlying Data Collection

All ABT utilize physical properties to function: encounter mapping devices, RFID tags, and GPS devices all transmit and/or receive data through radio waves (varying from low to ultra-high frequencies), geo-locators detect fluctuations in light intensity, and accelerometers record changes in capacitance (see Table 1). While each technology measures a unique property or transmits data through a unique method, each physical property has generalizable characteristics, uncertainty in their measurement, and identifiable limitations.

For example, encounter mapping devices use radio waves to transmit unique animal IDs between tags (see Rutz et al., 2012). As radio waves spread in an environment, the energy contained within the radio wave decreases at a certain rate, and extrinsic factors, such as height above ground or the composition of the habitat, can alter this rate. Thus, a calibration to determine the rate at which radio waves degrade due to the behavior of the animal is essential.

Similarly, GPS devices transmit and receive ultra-high frequency radio wave signals from satellites. If any medium, like a vegetation canopy, is obstructing the travel path of the radio wave toward the sensor, it results in a lower fix rate and lower number satellites that can be detected—hence a lower location accuracy. Therefore also GPS devices require a calibration (as done in Jiang et al., 2008 and Williams et al., 2012; see Frair et al., 2010 for a more complete overview) to know the difference in performance under different types of vegetation so that, e.g., seemingly different intensities in habitat utilization by an organism can be interpreted correctly.

Radio waves degrade at a specified rate (different rates for different frequency of radio waves), and certain extrinsic factors can distort the signal, such as trees, the placement of the tag on the animal, and the ground (Rutz et al., 2015). Studies have quantified the degradation and variation of the signal due to external factors on signal strength and/or reliability. Devices utilizing radio waves, such as proximity devices, have significant variation in signal strength, suggesting that using these devices to differentiate distances may be problematic (Rutz et al., 2015).

Conversion of Raw Data into Useful Variables

When using ABT, the biological outcome is only as good as the conversion of raw data into useful variables. Three steps are generally taken during this conversion: (1) understanding the structure of the collected data, (2) determining the relationship between the raw data and the variable of interest, and (3) quantifying the error associated with the conversion. While many studies undertake steps 1 and 2, few studies have concentrated on quantifying error due to the conversion of raw data into useful variables.

Light-based geo-location sensors, “geolocators”, combine a photoreceptor that converts incoming light levels into electrical energy with reference to an internal clock/calendar. To convert the light level and date-time information, use is made from a deterministic model (relating the time of local noon and midnight to the latitude, and day length to longitude). Usually the incoming light is limited to the 450 ±50 nm wavelength (a light-blue filter), because blue light predominates at the sun angles most appropriate for geo-location, just before sunrise and just after sunset (e.g., Lisovski et al., 2012).

On top of this well understood deterministic part of light based geo-location, there are large uncertainties involved due to factors like clouds, atmospheric refraction, atmospheric dust loading; foliage and feathers for terrestrial animals; depth and temperature effects for aquatic animals. Some of these factors can be corrected by measuring additional variables (such as pressure and temperature for aquatic; or cloudiness as observed at weather stations), whereas others have to be factored in as sources of noise. Several studies have been conducted to establish the uncertainty due to these factors, and also to provide a good protocol to calibrate sensors prior to deployment (Ekstrom, 2004; Fudickar et al., 2012; Bridge et al., 2013).

Subjectivity in the Settings Chosen

Nearly all ABT necessitate a priori selection of certain parameters. Generally, selecting parameters involves balancing the trade-offs between the highest resolution possible and the technological limits of the ABT, such as in the accessible memory and/or the capacity of the battery. Throughout the last decade, we have seen steady improvements in on-board memory and battery lifespans (Rutz and Hays, 2009), and may eventually see these limitations become non-existent.

One key parameter chosen is the frequency at which data is recorded. The rate of data collection determines the level of analysis data can be quantified, with higher rates of data collection allowing more a posteriori grouping of data for analysis, an indispensable step (at least for now) to examine data statistically. It is important to note, that we should maximize as much as possible the rate at which we collect data, as analytical methods allow for the comparison of higher-resolution data with older, less resolved data.

Another parameter commonly chosen is the power at which devices operate. In systems with passive RFID tags, for instance, the power of the receiver determines the distance (and rate) at which tags are being recognized. Since the detection range can, as a function of power, vary over an order of magnitude, it is crucial to know the exact power setting and its impact on this range in order to interpret the detection data (Nikitin and Rao, 2006; Finkenzeller, 2010). Hence the standardization of procedures in this context, or the availability of general calibration curves are required to compare results from different studies or interpret the results from a focal study at all (e.g., Pinter-Wollman et al., 2013).

A Four Step Workflow to Implement the Framework

The procedure we envision to implement the framework we outline in the previous section is based on 4 steps, which we labeled “Know”, “Identify”, “Evaluate”, and “Store”. Knowing consists of: (1) the physical principles of the technology, (2) the limitations of the technology, (3) the specific device and the context (i.e., the environment in which the device will be applied), and (4) experiences from others using the same device.
under comparable conditions. Knowing about the physics, sensor specification and context will enable a researcher to assess the likely influences in the specific location(s) the technology will be used, which is a required input to the next step: identify.

To “identify”, a model is constructed that describes the relationship between the variable of interest, relevant environmental factor(s), and the variable that is actually recorded and stored by the device. We call this model the observation model, since it is not a model for the biological system that we are studying, but only a way to translate the signal from the technology into the real-world variable that we are interested. Regardless, whether or not we have or need to build an existing observation model: experimental data or observational data with sufficient variability and data redundancy is required to evaluate model performance. In the ideal case, an appropriate observation model is already available and only needs to be evaluated (or calibrated and subsequently evaluated), and then the model can be used. If an observation model is not readily available, modeling from first principles (physical laws) or experimentation are the two remaining options. First, where possible, as much of the observation model should be defined on the basis of first principles (e.g., the effect temperature has on the functioning of a certain circuit, or the attenuation of a wave through a homogeneous medium at a given temperature). If the resulting model performs adequately, this could already form an endpoint. Otherwise, in addition, further parameterization with empirical relations may be required. For this task, experiments are required. In any case, an identified and calibrated observation model would need to be evaluated on a data set to which it was not calibrated.

Next, extra work is required to establish the robustness of the model in heterogeneous conditions. This is done in step 3, “Evaluate”, through a sensitivity and uncertainty analysis. For a sensitivity analysis, no real observations are required. Parameters or inputs are varied to evaluate how sensitive it is to perturbations. A sensitivity analysis is useful to find the parameters or inputs for which the observation model is particularly sensitive; these may subsequently be investigated in an uncertainty analysis. An uncertainty analysis applies realistic variability to a model to investigate how uncertainty propagates in the model and how it affects the results. If the resulting uncertainty is acceptable, the model is ready for use. In many

FIGURE 1 | Illustration of the proposed framework as well the as four-step workflow to handle error in ABT. We start by defining a phenomenon of interest and the part of it that is observable (e.g., not a total population but only a small sample can be observed; only larger individuals can be tagged—note that this partial observation may lead to biases). The observable phenomenon can be measured by a sensor. When doing this, a number of factors influence the ultimate error of the observation model—these are contained in the central green box of the figure. The consequences from physical principles, sensor settings, environmental variability and animal behavior can often be known on the basis of deduction or previous studies (hence this part is contained in the purple zone (A), step 1 in the workflow). But when the effect of these factors on a sensor is not known, these should be part of experiments to identify this error. In addition to the aforementioned factors, the rate of sampling and actual storage of these observations as well as the choice of an error metric to quantify the adequacy of an observation are of great importance to identify, calibrate and evaluate an observation model (see the right part of the central green box). In the ideal situation, a “gold standard measurement” is used which is insensitive to the factors that impact the sensor of interest (as illustrated by the dashed arrows). By comparing the sensor of interest to a gold standard the development of an observation model goes through two main cycles: identification of the model structure and best possible parameter values (calibration is a final step in this) and evaluation of the derived model on independent data. Identification and Evaluation are steps two and three in the 4-step workflow and are highlighted in the blue area (B) of the figure. The processes of identification, calibration and evaluation typically require several experiments, hence the dotted arrows link back into the green box. After an observation model passes the evaluation, it can be applied to predict an observable phenomenon (with associated uncertainty) on the basis of sensor data. The final step in the workflow is then to store the result of the identification and evaluation experiments (C). The fact that this box encloses all the components of the figure emphasizes that this storage should include all relevant contextual information, modeling choices, sensor settings (etc.) along with the observation model, and should also link to application studies where this particular observation model has been used. The three encircled numbers in the model-application part at the right highlight the three main concepts to classify ABT error: (1) the basic physical properties underlying data collection; (2) the conversion of raw data into useful variables; and (3) the subjectivity in the selected settings. These concepts can only be adequately quantified if the entire workflow is rigorously applied.
situations it may turn out that the model will work within certain boundaries of input parameters. In these cases it is useful to specify the conditions in which the observation model performs or doesn’t perform well. In addition, some questions require error estimation rather than quantification, requiring the evaluation of whether the error was estimated robustly.

Finally, we must properly store the observation model. With a simple parametric model one can store parameters, and parameter uncertainty. But with non-parametric models one should store the complete calibration dataset and algorithm; in that case, model uncertainty would need to be expressed in terms of predictive uncertainty. For any model, the assumptions underlying the model and the domain at which it is believed to be applicable. We strongly believe storing the model with uncertainty is essential to facilitate more transparency and use of the significant process just undertaken. Furthermore, the storage of raw data will aid error calculations in meta-analyses and facilitate more accurate evaluative criteria for inclusion.

An overview of both the framework and the workflow to implement is given in Figure 1.

Is There a Universal Protocol to Quantify Error?
No, but there are ways to think about error systematically, which help greatly to achieve error quantification. We propose a framework for this purpose. Each technology has its own unique limitations, necessitating different analytical or methodological approaches to mitigate error due to technological constraints. Yet, these confines exist in numerous technologies, and increasing the transparency through a common scheme will facilitate more rigorous studies and ensure the technological limitations refine and restrict the biological outcomes. Even in circumstances without opportunities to quantify different errors in ABT, the proper attribution of this error and the notion that it is irreducible will help to focus further developments (either in experimental work or development of new technologies).

Overall, organizing the process by which we think and calculate error is essential to ensure biological results are meaningful. To know, identify, evaluate and store, while recognizing the three processes introducing error into our data collections: physical properties, conversion from raw data into useful variables, and subjectivity in the settings. Specifically, recognizing similarities and differences between these devices will spur further integration of ideas from other devices to resolve the challenges of new devices, or old problems within a currently deployed device. Moreover, we believe the adoption of this scheme will facilitate feedback to designers, a necessary step to minimize certain types of error.

References


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Burns and van Loon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.