Beiträge zur Jagd- und Wildforschung, Bd. 40 (2015) 89–95

Beiträge zur
JAGD
&WILD
forschung · 40

Johannes Signer, Niko Balkenhol, Göttingen

Error detection and handling in GPS data

Schlagworte/key words: Error detection, GPS data, *Cervus elaphus*

Introduction

Data collected with GPS (Global Positioning System) telemetry are widely used to study wild animals and provide insights into their behavior (Cagnacci et al. 2010; Hebblewhite et al. 2010). Questions regarding habitat selection of animals (e.g., Thurfjell et al. 2014), behavioral states (e.g., Gurarie et al. 2009; Gurarie et al. 2015), space use (e.g., Laver & Kelly 2008), inter- and intraspecific interactions (e.g., Long et al. 2013) and movement in general (e.g., Avgar et al. 2013) can be addressed with telemetry data. With technological advances researchers are increasingly faced with larger data sets that open opportunities to investigate new questions, but are also accompanied by challenges of handling data adequately.

Dealing with GPS relocation data can be overwhelming and tools for managing data are needed (Urbano et al. 2010). Initiatives to harmonize and store data are available through data providers, such as EuroDeer or Movebank. Such database systems organize GPS telemetry data and can perform outlier detection. However, oftentimes GPS telemetry data are not stored in such databases and are only available through deliminator separated text files. Here, we focused on methods for situations where no database system with a sophisticated data model is available. Several other studies have introduced data models. For example, Calenge et al. (2009) introduced two different trajectory types (distinguishing trajectories of ordered data with and without time stamps). Urbano et al. (2010) developed a powerful database system based on PostGIS and PostgreSQL that is implemented with EuroDeer. Kranstauber et al. (2011) developed a data model for Movebank that is also tightly coupled to the move package for package R (R Core Team 2014; Kranstauber & Smolla 2015). Finally, Pebesma (2012) introduced a set of data models in R for handling spatio-temporal data implemented in the spacetime package. Ideally, a data model for animal tracking data would (1) integrate different types of trajectories (sensu Calenge et al. 2009); (2) accommodate attribute data of relocations, such as the habitat or time of the day when a relocation was recorded; (3) provide methods to manage tracking data and interact with other (environmental) covariates and (4) be implemented in a widely used and freely available software solution.

Once an appropriate data model is applied to tracking data, it is widely recognized that it is important to check data quality (Frair et al. 2010; Urbano et al. 2010; Bjørneraas et al. 2010) and detect outliers. Ideally, the analytical method accounts for erroneous observations (Patterson et al. 2008) and no cleaning of the data is necessary prior to the analysis. However, statistical methods that are able to account for observation errors (i.e., Bayesian state space models) are often non trivial to fit and usually require custom-written code. Since we often-

times use analytical approaches that do not account for an observation model (e.g., non state space models), preprocessing of the data is required. This involves removing erroneous relocation that are beyond the study region, where the GPS failed, that do not fit a given sampling regime, do not have sufficient accuracy (often measured through the number of satellites used to obtain a relocation) or lie to far apart (i.e., distances that the animal was not able to move).
In this article we start with introducing a data model for animal tracking data that builds on previously defined models. After a detailed description of the data model we show how this data model can be used to manage tracking data and to detect erroneous GPS locations based on space and time. Finally, we illustrate the implemented methods with relocation data from a red deer (*Cervus elaphus*) population from northern Germany and discuss the implementation, constraints and further plans for extensions.

Methods

Data Model and Implementation

A trajectory is characterized by a set of observations where the animal of interest was directly or indirectly observed. Each observation is characterized by an x and y coordinate that uniquely defines its position in space at a given point in time, and an ordering attribute (often a time stamp). In addition each relocation can have 0 to many additional attributes (e.g., habitat type, temperature, number of satellites used to obtain an observation). We distinguish three types of trajectories (following Calenge et al. 2009) based on whether time is known and on the regularity of the spacing that are represented in three different classes. The simplest trajectory type consists only of ordered relocations, but no time stamp is available.
The second trajectory type consists of coordinates with associated time stamps. Finally, the third trajectory type consists of coordinates with regular time stamps (i.e., two relocations are always separated by exactly the same amount of time). This type of trajectory can in most cases only be obtained through methods that regularize the trajectory. We distinguish two spatial components for trajectories: the relocations (these are the points where an animal was observed) and segments (this are the segments between two consecutive observations as linear interpolations between the start and end points). Each component can optionally have attribute data. By default a set of attributes are calculated for the segment attributes (e.g., time difference, length or turning angle; see also Calenge et al. 2009).
We implemented the highlighted data model in Program R (R Core Team 2014) and the rhr package (Signer & Balkenhol 2015). Spatial positions of animals are implemented using spatial classes for R that are available through the sp package (Pebesma & Bivand 2005). Spatio-temporal data are represented using the spacetime package in R (Pebesma 2012). The spacetime package represents spatial data using sp classes and time using the xts package (Ryan & Ulrich 2014). Beside the implementation of the classes themselves, we provide methods to access and assign spatial components and their attributes of trajectories using R's standard methods for manipulating objects. We also provide a set of methods that make already existing methods available to interact with other spatial data (i.e., raster layers with environmental information). Further methods to split a trajectory, calculate basic summary statistics (e.g., number of relocations, time span, bounding box) and mean squared displacement are available.

Burstifying trajectories

When tracking animals there are often periods of relocations followed by gaps with no observations. Such periods of continuous observations from the same animal are often referred to as bursts. In other words, a burst splits a trajectory of an animal in one or more complete sub-trajectories (i.e., there are no gaps). We deliberately did not implements bursts as part of the our data model, but think they are useful in the sense of sub-trajectories. Instead we always work on trajectory objects. R provides lists as a very flexible data structure that can accommodate various subsets of trajectories. We provide methods to (1) regularize a trajectory (Figure 1) and (2) burstify or splitting trajectories which results in a list of trajectories.

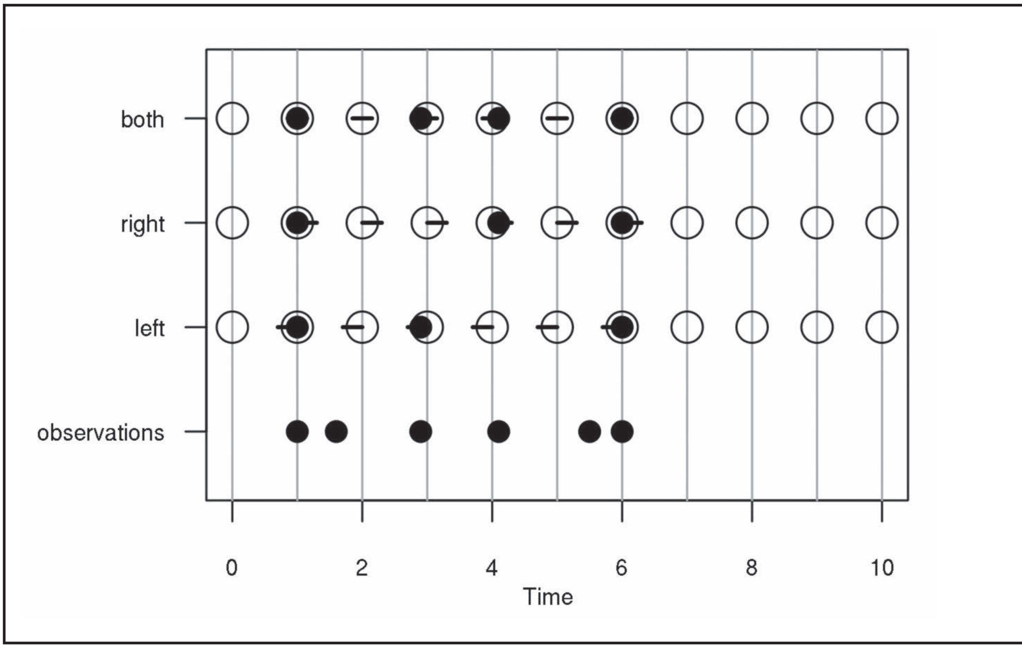

Fig. 1 Conceptual illustration on how a trajectory is regularized. The bottom row represents the actual observation. The top three rows illustrate the new trajectories (open circles) at regular time intervals with difference one. Black bars indicate the search radius that is either before, after or to both sides of the empirical observations. If a new empirical observations is within the search radius, it is considered in the new trajectory (black points).

To regularize a trajectory, observations of an existing trajectory are aligned, within a time window, to a new trajectory (Figure 1). New observations that are too far apart from any existing observation become empty observations, in the sense that they only contain time stamps and no relocations or segments. When regularizing trajectories the user can choose whether observations before, after or to both sides of a given new observations are considered (Figure 1). The old observation that is closest in time to the new observation is then chosen. Once a trajectory is regular, we can apply the concept of bursts. Either the trajectory is split by some covariate (e.g., by day or year) or the specially designed method to burstify the trajectory is applied. Burstify will split a trajectory in sub-trajectories after a prespecified number of missing observation in a regular trajectory. This two step approach of regularizing, and splitting or bursting a trajectory will achieve similar results to building bursts into the data model, but retains more flexibility.

Data Quality

When working with GPS relocation data, errors can occur with regard to the actual measurement of the relocation (Graves et al. 2006; Bjørneraas et al. 2010) and the study design. In the first case, missing or erroneous spatial data are recorded. In the second case, spatial data are recorded but outside the planned deployment period of the sensor (e.g., a sensor was not on the animal, or remained on the animal after the study terminated). Relocations with missing spatial data are relatively easy to detect and eliminate.

Relocation with erroneous GPS data can be detected through attribute data of fixes (e.g., number of satellites used to record a fix or the DOP of the relocation recording). With an appropriate data model (like the one we suggested above) it should also be easier to filter relocations for certain periods (i.e., the exact duration a collar was deployed) or a specific times or areas of interest.

Case Study

To demonstrate the data we used 107468 relocations from a telemetry study from Northern Germany (data are described in Reinecke et al. (2014)). For this illustration, we were interested to prepare the data set in such a way to use the year with most relocations available and to have two relocations per day (one at midnight and one at noon).

Results

Data model and implementation

We implemented the data model within the package rhr (Signer & Balkenhol 2015) for Program R (R Core Team 2014). The data model is recognized by all functions within the rhr package. Hence it is easy to prepare data (e.g., regularize or burstify trajectories) prior to the actual analysis (example code is available from the package website: rhr.spamwell.net).

Case study

We read the data from a separator delimited text file. In the first step, we created a trajectory with space and time. We than visually determined that for the year 2010 most relocations are available (Figure 2). Inspecting the distribution of the time of the day when relocations were recorded revealed that the intended interval of 6 hours is detectable, but significant noise is present (Figure 3). Finally, we regularized the trajectory to only a maximum of two relocations per day: the ones closest to noon and midnight (Figure 4).

Discussion

When working with GPS telemetry data, a solid data model is essential. A good data model can greatly improve efficiency and help to avoid errors during the analysis and detect errors within the data. We have extended previous works and implemented a data model for telemetry data

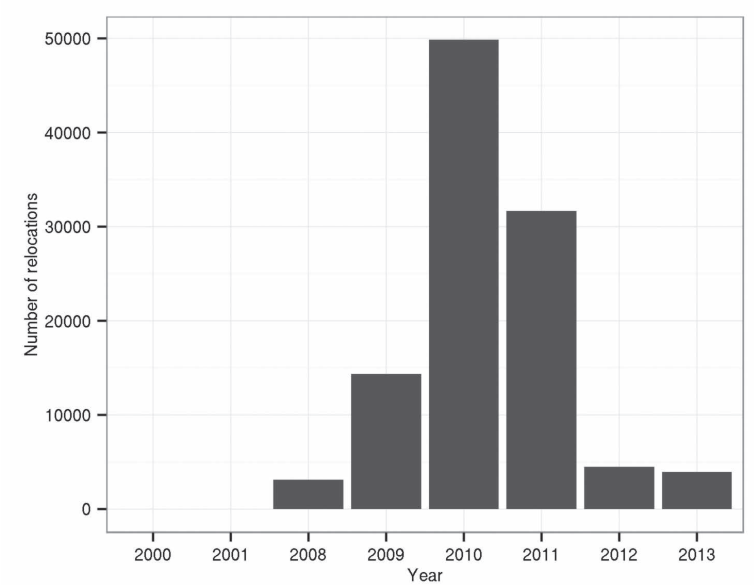

Fig. 2 Distribution of the times of the day when relocations where recorded by year. Relocations are unevenly distributed across years.

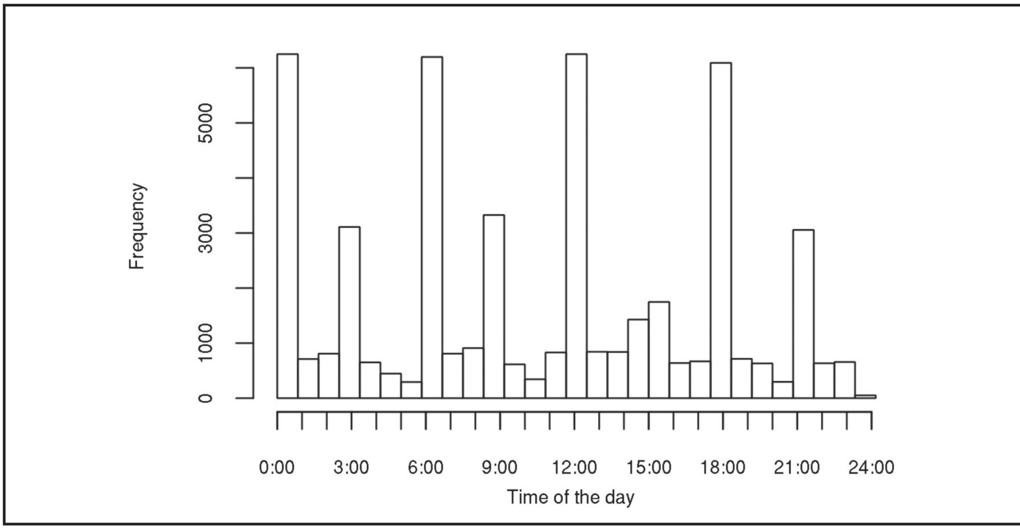

Fig. 3 Distribution of the time when fixes where taken. While all GPS collars where programmed to take a fix every 6 hours, substantial noise occurs.

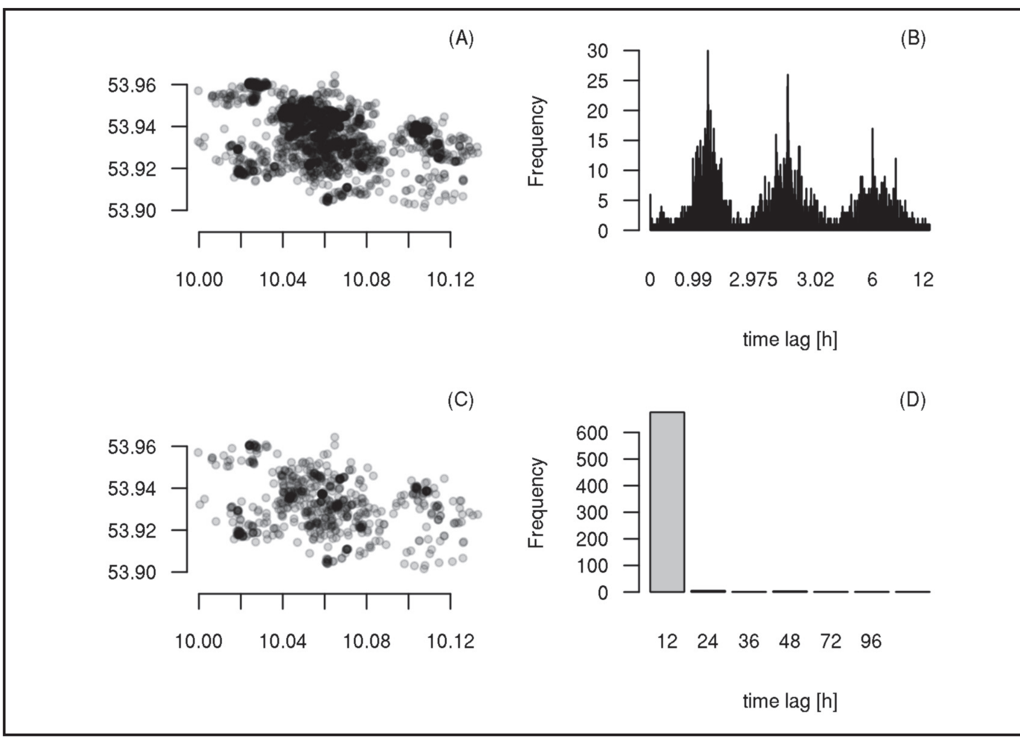

Fig. 4 Illustration of data preparing process for one animal. Starting with the full track for 2010 (panel A) and the distribution of time lags between two relocations (panel B; note that lags above 7 hours occur but are omitted here). We regularized the track to consider only two relocations a day (panel C). We considered relocations closest to midnight and noon with a search window of 5 hours before and after the actual relocation. Some days with missing data persist (panel D).

for the rhr package for the analysis of telemetry data within Program R.

The data model we implemented has the capabilities to represent GPS telemetry data (optionally also other telemetry data) and provides methods to query and manage telemetry data. Among others, methods are a available to regularize a path (i.e., ensure that the time interval between two relocations is always identical), split a trajectory into two or more sub-trajectories based on some criteria or into bursts, if gaps between relocation periods exists. Furthermore, the data model provides an infrastructure to save attribute information for the relocations themselves, and also for the segments between two relocations.

We demonstrate the usefulness of such a data model with a data set of a red deer population from Northern Germany. We were interested to correct the trajectory to obtain two relocations per day, one at midnight and one at noon. Using the newly implemented methods, it was a quite simple task to create such a trajectory.

Further extensions of this data model could include methods to detect interactions between trajectories (animals), more sophisticated error detection mechanisms (e.g., routines suggest by Bjørneraas et al. 2010) and to move to the next step of analysis telemetry data using path segmentation and/or step selection function.

Summary

Wild animals are by their nature often difficult to observe and study. Hence, wildlife biologists oftentimes rely on remote data collecting devices such Global Position Systems (GPS). Location data from GPS-collars have become very popular for studying the behavior, resource or space use of wild animals. Many analytical methods (e.g., habitat selection, home range analysis) that are used to answer interesting biological questions and guide management decisions heavily depend on GPS borne location data. In order to obtain correct results, it is fundamental to screen GPS data prior to any analysis for potential errors.

We give a short conceptual overview of the importance of a data model and the kind of errors that potentially occur when working with GPS data. We then discuss how wildlife professionals can handle these errors to improve the accuracy of location data and illustrate this with a data set from a red deer (*Cervus elaphus*) population from Northern Germany.

Zusammenfassung

Erkennung und Handhabung von Fehlern bei GPS Lokalisierungen

Wildbiologen greifen oft auf Telemetriedaten zurück um wildbiologische Fragestellunen zu untersuchen. Daten die mittels GPS Telemetrie gewonnen wurden kommen dabei immer häufiger zum Einsatz um Fragestellungen der Raumnutzung, Habitatselektion oder Bewegungsökologie zu bearbeiten. Dabei ist ein gutes Datenmodell eine wichtige Arbeitsgrundlage um effektiv mit den Daten zu arbeiten und fehlerhafte Ortungen zu entfernen. In diesem Artikel heben wir die zentrale Rolle eines geeigneten Datenmodelles für Telemetriedaten hervor und zeigen anhand eines Beispieles einer Rotwild (*Cervus elaphus*) Rotwild-Population aus Norddeutschland, wie das vorgestellte Datenmodell angewendet werden und in einem typischen Arbeitsablauf zur Analyse von Telemetrie integriert werden kann.

References

Avgar, T.; Mosser, A.; Brown, G.S.; Fryxell, J.M. (2013): Environmental and individual drivers of animal movement patterns across a wide geographical gradient. – Journal of animal ecology **82** (1): 96–106.

Bjørneraas, K.; Moorter, B.; Rolandsen, C.M.; Herfindal, I. (2010): Screening global positioning system location data for errors using animal movement characteristics. – The Journal of Wildlife Management **74** (6): 1361–1366.

Cagnacci, F.; Boitani, L.; Powell, R.A.; Boyce, M.S. (2010): Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. – Philosophical Transactions of the Royal Society B: Biological Sciences **365** (1550): 2157–2162.

Calenge, C.; Dray, S.; Royer-Carenzi, M. (2009): The concept of animals' trajectories from a data analysis perspective. – Ecological informatics **4** (1): 34–41.

Frair, J.L.; Fieberg, J.; Hebblewhite, M.; Cagnacci, F.; DeCesare, N. J.; Pedrotti, L. (2010): Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. – Philosophical

Transactions of the Royal Society of London B: Biological Sciences **365** (1550): 2187–2200.

Graves, T.A.; Waller, J.S. (2006): Understanding the causes of missed global positioning system telemetry fixes. – Journal of Wildlife Management **70** (3): 844–851.

Gurarie, E.; Andrews, R.D.; Laidre, K.L. (2009): A novel method for identifying behavioural changes in animal movement data. – Ecology letters **12** (5): 395–408.

Gurarie, E.; Bracis, C.; Delgado, M.; Meckley, T.D.; Kojola, I.; Wagner, C.M. (2015): What is the animal doing? Tools for exploring behavioral structure in animal movements. – Journal of Animal Ecology, *in press*.

Hebblewhite, M.; Haydon, D.T. (2010): Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. – Philosophical Transactions of the Royal Society of London B: Biological Sciences **365** (1550): 2303–2312.

Kranstauber, B.; Cameron, A.; Weinzerl, R.; Fountain, T.; Tilak, S.; Wikelski, M.; Kays, R.; (2011): The Movebank data model for animal tracking. – Environmental Modelling & Software **26** (6): 834-835.

Kranstauber, B.; Smolla, M. (2015): move: Visualizing and Analyzing Animal Track Data. R package version 1.4.496. http://CRAN.R-project.org/package=move.

Laver, P.N.; Kelly, M.J. (2008): A critical review of home range studies. – The Journal of Wildlife Management **72** (1): 290–298.

Long, J.A.; Nelson, T.A. (2013): Measuring dynamic interaction in movement data. – Transactions in GIS **17** (1): 62–77.

Patterson, T.A.; Thomas, L.; Wilcox, C.; Ovaskainen, O.; Matthiopoulos, J. (2008): State-space models of individual animal movement. – Trends in ecology & evolution **23** (2): 87–94.

Pebesma, E.J.; Bivand, R.S. (2005): Classes and methods for spatial data in R. R News **5** (2), http://cran.r-project.org/doc/Rnews/.

Pebesma, E. (2012): spacetime: Spatio-temporal data in r. Journal of Statistical Software **51** (7): 1–30.

R Core Team (2014): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Reinecke, H.; Leinen, L.; Thissen, I.; Meissner, M.; Herzog, S.; Schütz, S.; Kiffner, C. (2014): Home range size estimates of red deer in Germany: environmental, individual and methodological correlates. European journal of wildlife research **60** (2): 237–247.

Ryan, J.A.; Ulrich, J.M. (2014): xts: eXtensible Time Series. R package version 0.9-7. http://CRAN.R-project.org/package=xts.

Signer, J.; Balkenhol, N. (2015): Reproducible home ranges (rhr): A new, user-friendly R package for analyses of wildlife telemetry data. – Wildlife Society Bulletin, *in press*.

Thurfjell, H.; Ciuti, S.; Boyce, M.S. (2014): Applications of step-selection functions in ecology and conservation. – Mov. Ecol **2** (4).

Urbano, F.; Cagnacci, F.; Calenge, C.; Dettki, H.; Cameron, A.; Neteler, M. (2010): Wildlife tracking data management: a new vision. Philosophical Transactions of the Royal Society B: Biological Sciences **365** (1550): 2177–2185.

Anschrift des Verfassers:

Johannes Signer
Abteilung Wildtierwissenschaften
Georg-August-Universität Göttingen
Büsgenweg 3
D-37077 Göttingen
E-Mail: jsigner@gwdg.de