Results of the break-out group: Benchmarking

Group discussion with Emiel van Loon, Ross Purves, and Rob Weibel

An increasing diversity and quantity of movement data is becoming available while, on the other hand, an increasing amount of algorithms to analyse these data. Much of the data and algorithms seem to have a potential for answering theoretical as well as applied questions in domains like behavioural ecology, (spatial) economy, physiology and cognitive sciences.

However, many basic properties of movement data and derived products from these data are not clearly defined. In addition, analysis algorithms vary with respect to input as well as output data. As a result, it is difficult to evaluate the suitability of different algorithms for application to a given type of data and question. We think there is a need to define clear tests or experiments for this purpose. The activity to define and conduct these tests is called ‘benchmarking’.

This working group has discussed the possibilities to start developing benchmarking tools. It is clear that for this benchmarking, eventually a collection of representative (and annotated) data sets are required, a number of tests or standard ways by which algorithms can be compared, and a collection of analysis algorithms (ranging from pre- to post-processing steps). Also, accepted definitions (which will be domain specific) of important concepts will prove very important. Examples are terms like ‘trip duration’, ‘movement speed’ (relative to a fixed location on the earth surface, or relative to a fluid medium), ‘convergence’.

We decided that it would be best to start with concrete data and a concrete application field, and work from there. The available gull flight data are sufficiently general and offer many features for this purpose. In addition, the upcoming Lorentz workshop (june 2011) seems to be a good opportunity to test how far we can get with this approach. We designed a data challenge for this workshop, to generate a lot of input for a ‘benchmarking toolkit’.

We made a choice to provide some real movement data set as currently stored in the data base of the gull-movement project. The data will contain the attributes as shown in Table 1. The participants in the data challenge are provided with a description of that data set and are requested to analyse the data and report about it during the workshop to address the following basic issues:

1. Identify each trip and establish for each of these:
    a. trip duration in hours
    b. total distance traveled in meters
    c. maximum distance from nest in meters

Dagstuhl Seminar Proceedings 10491
Representation, Analysis and Visualization of Moving Objects
http://drops.dagstuhl.de/opus/volltexte/2011/2988

2. What is the mean, minimum, maximum and standard deviation of trip duration (in hours), total distance traveled (in meters)and maximum distance from nest (in meters):
    a. per individual
    b. for the entire sample
3. What is the mean, minimum, maximum and standard deviation of flight speed in m/s:
    a. per trip
    b. per individual
    c. for the entire sample
4. What is the time spent flying and time not spent flying
    a. per trip
    b. per individual
    c. for the entire sample

A spreadsheet will be prepared, in which the results for each of the above questions should be entered, so that the results will be comparable. In addition, a template for reporting the data pre-processing and analysis steps has been made. The results of the various participants to the challenge will be compared and discussed during the workshop.

Table 1. Explanation of the attributes in the gull movement data set, to be used in the data challenge.

attribute	**explanation**
device_info_serial	id of the tracking device (and the individual bird)
date_time	GMT
day	day – taken out of the date_time field
month	month – taken out of the date_time field
year	year – taken out of the date_time field
hour	hour – taken out of the date_time field
minute	minute – taken out of the date_time field
second	second – taken out of the date_time field
latitude	lat. location of the bird decimal degrees
longitude	lon. location of the bird decimal degrees
Y	Latitude projected in the RD-new coordinate system (Dutch system in m)
X	Longitude projected in the RD-new coordinate system (Dutch system in m)

altitude	height of the bird above sea level – m
temperature	Temperature of tracking device
speed	Instantaneous speed (m/s) measured by GPS
gps_fixstatus	indication whether an altitude measurement was possible (3D) or not (2D)