THIRD OPEN SOURCE GEOSPATIAL RESEARCH & EDUCATION SYMPOSIUM (OGRS)
*Espoo, Finland, 10-13 June 2014*

# Geographic knowledge discovery from sparse GPS-data with R – spatio-temporal examination of Amazonian river transports

Henrikki Tenkanen[1], Maria Salonen[1], Matti Lattu[2], and Tuuli Toivonen[1,3]

[1]Department of Geosciences and Geography, University of Helsinki, Finland
[2]Department of Teacher Education, University of Helsinki, Finland
[3]Department of Biosciences, University of Helsinki, Finland

---

**Abstract:** Technological developments have made it viable to collect movement data of humans, vehicles and animals. A low-cost GPS-based monitoring system was developed to collect movement data of riverboats in Western Amazonia. The pilot monitoring system provides real-time location and time data that indicate where and when the collaborating vessels are navigating. We developed a tool to identify individual trips from the data and compute motion parameters such as speed, change of speed and the direction of movement. The pilot monitoring system was tested to analyze different spatio-temporal transportation characteristics of the local riverboats in Western Amazonia during the year 2012. The aim of this case study was to explore the seasonal and directional variation in the transportation characteristics along the rivers. Furthermore, we analyzed the relationship between the water height of the rivers and the travel speed of individual journeys. Results of the analysis show that navigation along the rivers indeed has seasonal and directional variation, and that the river morphology seems to affect the movement patterns of the vessels. Our case study also proved that our tools to collect and analyze the movement data functioned as planned.

**Keywords:** Data mining, GPS, transportation, Western Amazonia, R

---

# 1 Introduction

A vast amount of spatio-temporal data has become available with the fast development of information technology and different monitoring systems over the last two decades. Position-aware devices are one of the most dominant sources for collecting movement data of various targets such as vehicles and people (Pelekis et al. 2012). Spatio-temporal information that is derived from the tracking devices enable to build movement patterns from the targets, and to calculate measurable motion parameters such as speed, change of speed or the direction of movement (Laube & Imfeld 2002, Dodge et al. 2009).

For this study a specific low-cost GPS-based monitoring system called Amazonian Riverboat Observation System (AROS) was developed to collect movement data of the local riverboats in the departments of Loreto and Ucayali in Peruvian Amazonia. AROS uses SPOT Personal Trackers[1] to provides real-time GPS-data with coordinates and timestamp that indicate where and when the collaborating vessels are navigating. Also a specific tool called Trajectory Reconstruction and Analysis Tool (TRAT) was developed to analyze the mobility data. The main purpose of TRAT is to identify individual trips from the data and calculate different movement parameters from the GPS-data that was obtained with AROS.

To test the use of AROS and TRAT, we analyzed different spatio-temporal transportation characteristics based on data from year 2012. The aim of the study was 1) to technically test AROS monitoring system for capturing the spatio-temporal characteristics of riverine transportation and 2) to test TRAT for detecting individual journeys from AROS data and calculate different movement parameters. The third aim was to study the transportation patterns in the study area based on developed tools. Here, our aim was to find out if there is seasonal and directional variation in transportation characteristics along the rivers, and if there can be find connection between water height of the rivers and travel speed of individual journeys.

Monitoring the movements of vessels along Amazonian rivers for long period of time enable us to evaluate how much the river dynamics (i.e. high water/low water) effect on transportation patterns. Such information is particularly useful, as the entire transportation system in the Peruvian Amazonia is highly dependent on river transportations (Rodriguez Achung 1994). There are practically no roads in the area of the size of Germany and the rivers form the main transportation network for the people living in the area. Local economies and interaction between populations depend on public transportation loosely organized with long distance river launches.

[1] http://www.findmespot.eu/en/index.php?cid=101 [accessed 12.5.2014]

# 2 Materials and methods

## 2.1 Data collection system

Data used in this study was collected with an experimental data collection and monitoring system AROS that resembles the automatic identification system (AIS) that is used to track professional maritime vessels (see e.g. Demsar & Virrantaus 2010). AROS is based on low cost satellite messengers (approx. 100€ each) that enable to determine the location of the devices by GPS satellite system and to send their location information via communication satellites to a database (see Figure 1).

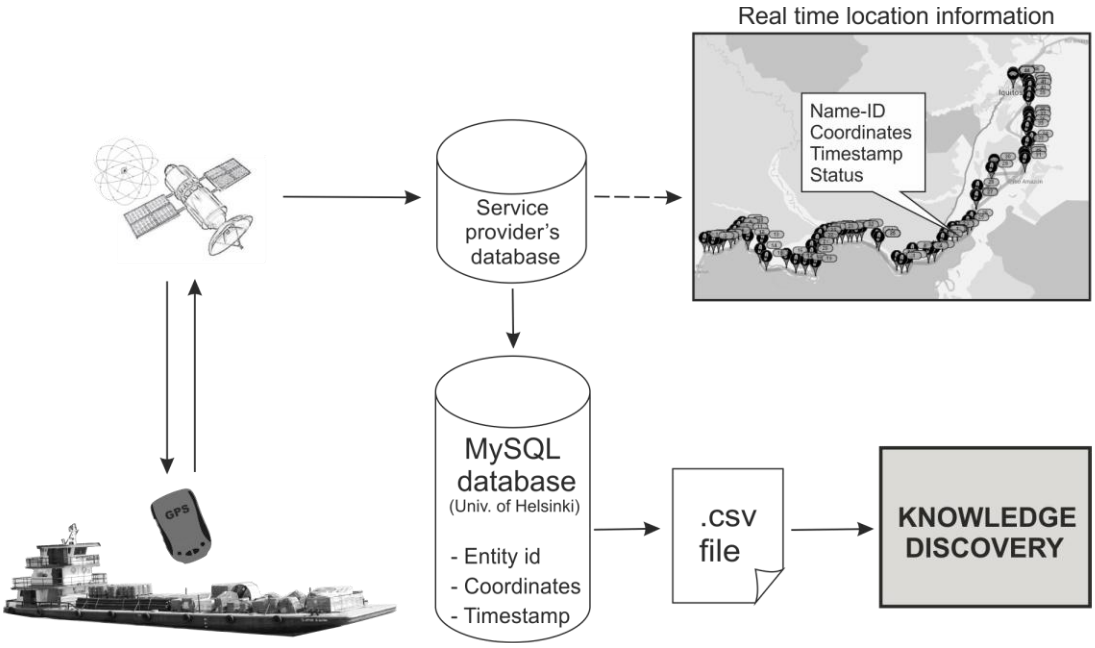


Figure 1: Concept of the pilot data collection system that utilizes GPS satellite messenger that is able to both receive and send location information

Altogether five satellite messengers were used to collect GPS-data of different boats. The messenger devices send their location information every 10 minutes to a service provider's database from where the data is automatically retrieved via XML-feed to a MySQL database at the University of Helsinki for permanent storage. Data consists of seven attributes: id, latitude and longitude coordinates, timestamp, Unix time and message type (test/tracking). The dataset used in this case study consisted of 11 572 GPS observations collected with AROS.

## 2.2 Data analysis

Analyzing mobility data and reconstructing trajectories from GPS-points is not a straightforward procedure (see e.g. Marketos et al. 2008). For detecting individual journeys (i.e. single trip from origin harbor to destination harbor) and extracting knowledge from such trajectories, an analytical tool (TRAT[2]) was developed to fully automatize the knowledge discovery and visualization processes. TRAT is distributed as open-source code and it is developed for R. It uses geospatial packages available for R, and a package called Muste (Sund et al. 2012) that is used as user interface and tool for data management.

GPS-data from AROS is relatively sparse, and therefore it was necessary to develop an analytical approach that utilizes training dataset for calculating different motion parameters. This was done by enriching the original GPS dataset with ancillary data that contains information about network distance from the origin harbor, sinuosity index of river etc. (see Figure 2). Each AROS observation was spatially joined to the closest reference point which enabled us to calculate network distance and speed between consecutive points along the meandering rivers, determine the navigation direction and detect individual journeys from the data. Detecting an individual journey from the data mass is based on a specific algorithm that determines if an observation belongs to the same trajectory as earlier one or not. This classification is intrinsically based on following criteria: direction of movement, stationary time, temporal gap between observations, and spatial gap between observations.

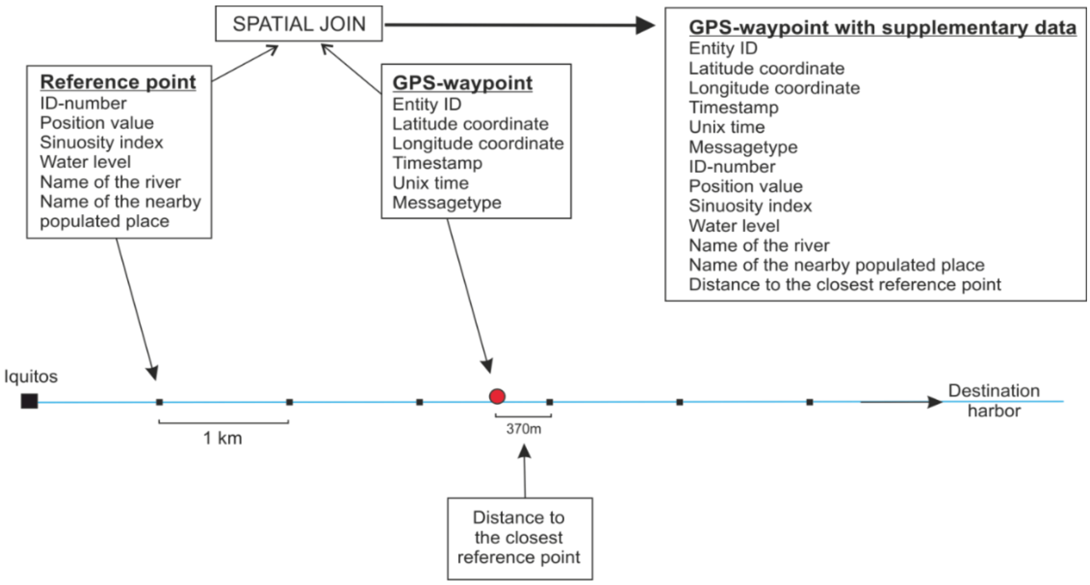


Figure 2: The final dataset is enriched by spatially joining each GPS-waypoint to the closest reference point

[2] http://www.github.com/UH-Accessibility/TRAT [accessed 12.5.2014]

# 3 Results

Here we present few key findings from our case study and give examples how mobility data can be visualized. Figure 3 represents individual journeys that were detected from AROS data (year 2012) and classified by main navigation direction. Interactive 3D visualization of trajectories in space-time cube, where x- and y-axis represent space and t-axis represents time, enabled us to assess the quality of TRAT for detecting individual journeys.

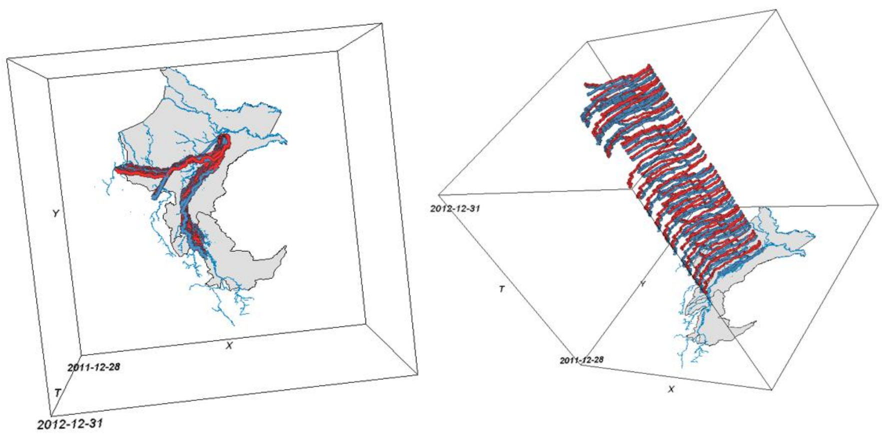


Figure 3: Individual journeys identified with TRAT and represented in space-time cube. Data was collected with AROS in 2012

Figure 4 represents an example of movement profile of the journeys which shows the travel speeds of the journeys (y-axis) along the vessel's navigation path (x-axis). Results show that navigation along the river has clear directional variation since navigating downstream is over 40% faster compared to upstream navigation during high water and intermediate. However, during low water there is not difference between navigation directions. Results show also clear seasonal variation in travel speeds, and hence navigation was over 30% faster during high water compared to low water (navigating downstream). Navigating upstream was fastest during low water but seasonal differences were considerably lower compared to downstream navigation.

Figure 5 represents the average speed of individual journeys (red lines) along circular axis that represents time. Black lines inside the red ones represent the percentage of the journey (from total route length) that was captured by AROS and thus can be used to assess the quality of the data. Blue polygon in the middle of the radar plots represent daily water levels of the river where vessels are navigating.

Fitting simple regression model between average travel speed of the journeys and water levels of the river revealed that there seems to be strong connection between travel speed and river height when travelled downstream ($R^2$=0.73) along highly meandering river. On other cases that were studied, the results suggest that there is no connection between travel speed characteristics and river height.

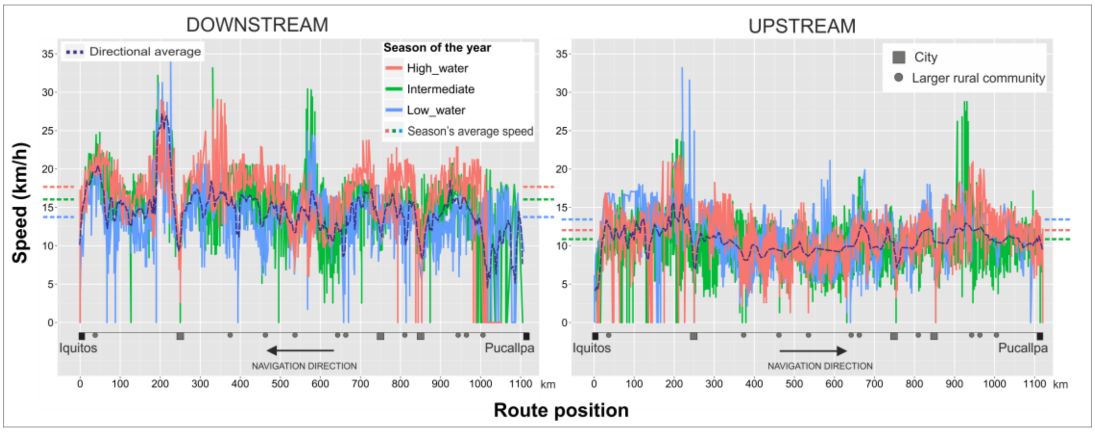


Figure 4: Movement profile representing the travel speeds of the vessels along the navigation route. Journeys were classified by season and navigation direction.

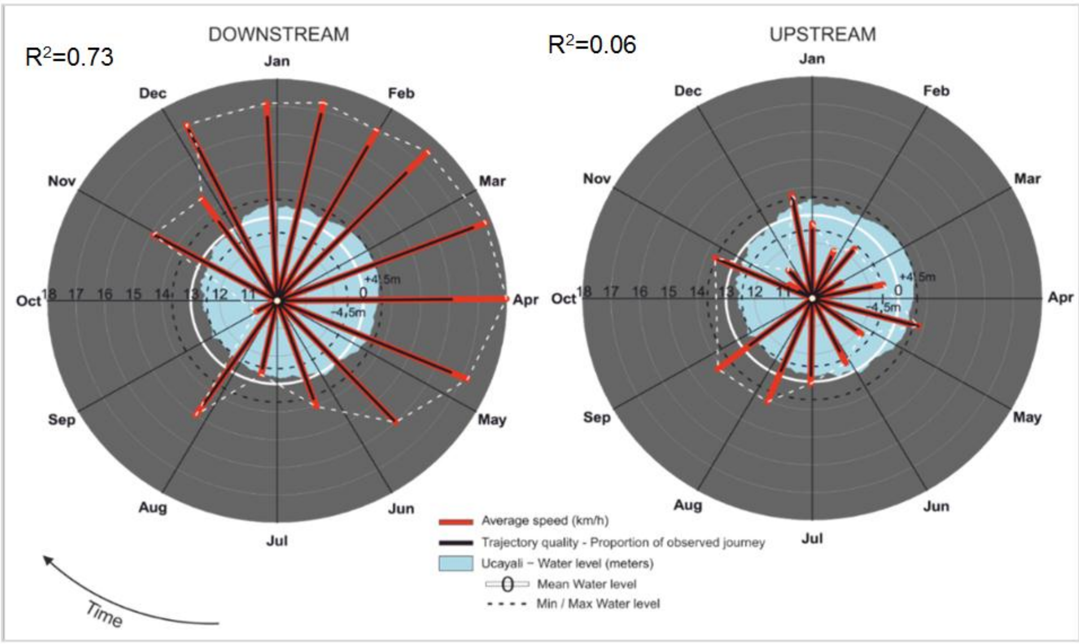


Figure 5: Radar plot representing the average speeds of individual journeys vs. water level. The $R^2$ values from linear regression (y=speed, β=water level) were 0.73 and 0.06 for downstream and upstream data respectively.

# 4 Discussion

Monitoring the movements of Amazonian riverboats with low-cost monitoring system AROS proved to work well as means for capturing the seasonal variation in movement characteristics along the rivers. Even though the sample frequency of AROS is relatively low (1 obs. / 10 minutes), it was high enough for our purposes because the travel speed of the vessels along the rivers is relatively slow (max 25 km/h). The satellite communication of the system was also functioning well and was a prerequisite for the system as there is no mobile network available in the area.

Travel speed calculations of the data analysis tool TRAT proved to be mainly accurate. The only exceptions were few cases where travel speed was higher than expected because of erroneous spatial join to the training dataset (i.e. river network). Assessing the accuracy of TRAT to detect individual journeys from the data proved that TRAT functioned as desired, thus 100% of the individual journeys and main direction of movement (upstream/downstream) were correctly detected from the data. Also the results of the case study seem to be in accordance with previous studies (cf. Salonen et al. 2012) that also imply that AROS and TRAT function as planned.

Transportation plays a key role when trying to find the factors affecting on development of a certain location. It, together with the overall accessibility of a location (i.e. how well the location can be reached) has a significant role in local livelihoods, land use change, deforestation, and conservation effectiveness in the region. Results of this study provide more accurate data for studies focusing on previously mentioned topics in the study area. Active development of openly available analytical tools is of paramount importance especially in areas such as rural Peru where local scientists can confront difficulties in getting hold of commercial GIS software that can analyze novel data sources such as mobility data. AROS and TRAT offer an excellent low cost solution to study global accessibility issues. By making a few modifications to the tool they could be used in other contexts such as studying global transportation patterns of professional vessels. R proved to be great platform for developing tools to analyze AROS data since it provides large collection of available packages that can be directly used for developing GIS tools that can handle GPS-data and visualize the results in an understandable manner.

# Acknowledgments

This study was done as part of accessibility research project in the Department of Geosciences and Geography, University of Helsinki. Study was funded by University of Helsinki, Finland.

## References

Demsar, U. & Virrantaus, K. (2010), Space–time density of trajectories: exploring spatio-temporal patterns in movement data. *International Journal of Geographical Information Science*, 24(10), 1527–1542.

Dodge, S., Weibel, R. & Forootan, E. (2009), Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems* 33, 419–434.

Laube, P. & Imfeld, S. (2002), Analyzing relative motion within groups of trackable moving point objects, *in* Egenhofer, M. & D. Mark (eds.), Geographic Information Science 2478, Springer, Berlin-Heidelberg, Germany, pp. 132–144.

Marketos, G., Frentzos, E., Ntoutsi, I., Pelekis, N., Raffaetá A. & Theodoridis, Y. (2008), Building Real-world Trajectory Warehouses, *in* 'MobiDE '08: Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access', ACM Press. New York, NY, USA, pp. 8-15.

Rodriguez Achung, M. (1994), Crecimiento urbano de Iquitos: condicionamientos estructurales en la decaca del ´70 y sus perspectivas. Documento técnico No 08, IIAP, Iquitos, Loreto, Peru, 103 pp.

Pelekis, N., Andrienko, G., Andrienko, N., Kopanakis, I., Marketos, G. & Theodoridis, Y. (2012), Visually exploring movement data via similarity-based analysis. *Journal of Intelligent Information Systems* 38, 343-391.