Identifying Requirements for Centralized Service for Movement and Biodiversity Data Analysis

Ivana Nižetić Kosović, Boris Milašinović, Krešimir Fertalj
Department of Applied Computing
Faculty of Electrical Engineering and Computing, University of Zagreb
Zagreb, Croatia
e-mails: {ivana.nizetic, boris.milasinovic, kresimir.fertalj}@fer.hr

***Abstract*—Positioning technology is lately widely used in many scientific fields to collect movement and biodiversity data for further analysis. That generates enormous amount of positions and tracking data and impose the need for developing new algorithms for analysis and prediction, which are managed in ever growing partial and incomplete software solutions. In this paper, we made a roadmap for development of centralized service-oriented software in which one could manage data about moving objects or plants, as well as spatial layers, contextual information and perform complex algorithms. We identified a set of interfaces for communication with users allowing data and algorithm manipulation. Furthermore, we proposed meta-model for storing data and algorithms in order to achieve adaptiveness and wide applicability. The proposed model establishes a baseline for a concrete implementation.**

Keywords-movement data; biodiversity; tracking; service-oriented software; meta-model.

I. Introduction

Widespread use of Global Positioning System (GPS) devices, smart phones and wireless communication devices induced the expansion of research on moving objects [1]-[3]. There is an increasing number of applications in which mobility plays an important role. Vehicles are monitored and analysed in the field of traffic management and control to predict driver's intentions or traffic congestions [4]-[7]. Mobile users' movement is analysed to assure fast access point availability [8][9]. In the field of behavioural ecology, wild animals are tracked in order to predict their migrations and predator-prey behaviour [10]-[12]. There is also a variety of location-based services provided to smart phone users such as museum or touristic attractions applications for tourists, hospital plan information for doctors and nurses, hotels, location-aware games and so on.

Ubiquitous tracking technology and various applications generate enormous amount of tracking data and impose the need for developing new algorithms for analysis and prediction. Furthermore, in the last decade, the need to enrich object's movement data with geographical and semantic information is recognized since raw trajectory data (positions and timestamps) are not sufficient to obtain meaningful movement patterns [3][13][14]. Inclusion of heterogeneous data (contextual information, environmental data) makes analysis and prediction even more complex and the need for explanation of movement and spatial data more indispensable. Similar problem exists not only in the field of moving objects but also in some other fields like botany [15]. Although plants do not move, the algorithms for analysis and prediction are still complex and yield different results.

Still, there is no unique software in which one could manage data about moving objects, contextual information and perform complex algorithms. Visualisation and algorithms are currently managed by partial or incomplete platform-dependent software which cannot fully satisfy researchers' needs. Moreover, existing applications are not adaptable to new (custom) algorithms. Furthermore, since mobile applications cannot perform complex calculations and manage voluminous data locally, the need for standardized service-oriented system is essential.

These problems are further elaborated in Section II in order to formulate a problem followed by the identification of typical algorithms and their inputs and outputs. Once identified, it would be shown that there is a common intersection between inputs and outputs which can lead to a solution presented in the Section III. After centralized service-oriented solution concept description, a set of interfaces is enumerated, followed by the meta-model for storing data and algorithms. The paper ends with a conclusion of a presented work and future work guidelines.

II. Problem formulation

In the field of moving objects, an increasing number of analysis and prediction algorithms are developing: data mining techniques to extract behavioural patterns from moving objects data [16][17], clustering algorithms to detect important places [13][18], prediction techniques to model and predict moving object's future location, such as neural networks, Markov models, and specific types of dynamic Bayesian networks, like Hidden Markov Models (HMM) or Kalman filter [5][19]-[22]. In recent years, considerable research has been devoted to mapping the flora distribution, spatial analysis, biodiversity analysis and prediction of

 ISBN: 978-1-61208-367-4

occurrence. Bedia et al. [23] presented variety of different algorithms applied to a particular geographic area. However, it is not guaranteed that a particular algorithm can equally be used for different regions of the world. Although there are significant differences between moving objects and (static) plants, some similarities exist that led us to identify main ideas and enumerate problems in order to propose an integrated software solution. Table I shows representative categories of algorithms and their inputs and outputs with some examples where results depend on algorithm being used.

Although aforementioned algorithms and their implementations are valuable, we listed their main drawbacks considering not only their performance accuracy but also their ease of use and adaptiveness. As it can be seen from Table I, each algorithm uses spatial data and a set of coordinates that represents object positions or species findings. Testing different algorithms is usually a work intensive task as algorithms use different input data format and produce output in different format. Moreover, findings data must be exported to the appropriate format prior to the use of an algorithm and output must be transformed back into user's format repeating the similar task for many algorithms. Moreover, existent software is:

- Partial, incomplete – e.g., ArcGIS can be used for calculation of probabilities distribution of species, Weka and IntelligentMiner for clustering or basic HMM modelling
- Local – desktop apps, rarely applets or web services
- Closed – with no possibility for adding/customizing algorithms, and inaccessible to ordinary people, e.g., mobile users, non-experts in certain field
- Too specialized – not general enough to encompass various needs and heterogeneous data, e.g., HMM is applied to many classes of moving objects but each model is specialized only for that class although they have common structure and performance
- Technically determined – platform limited, e.g., applets require additional software installations, Geographic Information System (GIS) software requires high capacity and performance,
- Lack of (customized) visualisation of results

Consequently, there is a need for centralized, interoperable, opened, adaptive (to data and to algorithms) and widely useful application that could be extended with additional algorithms and additional data necessary for a particular algorithm.

TABLE I. EXAMPLES OF COMMONLY USED ALGORITHMS IN MOVEMENT AND BIODIVERSITY DATA ANALYSIS

Algorithm Category	**Input**	**Output**	**Results depend on algorithm?**
Pattern discovery - construction	Sets of marked sequential positions (positions, timestamps and corresponding pattern), Algorithm parameters (list of patterns, initial probabilities)	HMM (states, transitions, transition and emission probabilities)	Yes (HMM, State-space model, Artificial neural network - ANN, ...)
Pattern discovery - usage	A set of sequential positions (positions and timestamps)	A set of marked sequential positions (positions, timestamps and corresponding pattern)	Yes (HMM, State-space model, ANN, ...)
Positions clustering	Positions Algorithm parameters (Eps, MinPt)	Sets of positions (clusters) and a set of noise positions	Yes (Density-based spatial clustering of applications with noise - DBSCAN, …)
Species distribution	Spatial layers, species findings	Matrix of species findings per spatial layer attribute	No
Species distribution prediction	Spatial layers, species findings	Matrix of probabilities corresponding to spatial shapes	Yes (Distance alg., Maximum Entropy Modelling (MAXENT), Multiple Logistic Regression - MLR, ANN,...)
Ecological profile	Spatial layers, species findings	Matrix of species occurrence per spatial layer attribute or spatial unit	No
Biodiversity analysis (alpha and research intensity)	Spatial layers, species findings	Collection of matrices (alpha and research intensity per layer attribute, other species data - ecological indices per spatial layer attribute, basic species data...) Extension of a spatial layer with data from matrices but for a unit instead of attribute	May depend, when data uncertainty exists [15]
Movement prediction	Positions, spatial layers, contextual data	Positions	Yes (ANN, HMM, ...)

Recently, some notable ideas of centralized solutions for managing moving objects' data and performing algorithms were proposed. Xu and Guting [24] proposed a generic data model for moving objects that can apply in more than one environment and applied it to transportation model. A conceptual data model for representing semantic trajectories applied to tourism and animal movement is shown by Bogorny et al. [25], giving the baseline for future research on semantic trajectory.

Considering wild life research there is a vision of centralized solution for wildlife data management in [26]. We have also presented a generic model and proposed the conceptual data model for analysis and movement prediction independent of application area and moving object type [27].

In [15], we proposed an object model for biodiversity analysis that can avoid the problem of data export. Input and output are modelled using interfaces thus making the model available in various usage scenarios as a web service or a layer in an application. As it only defines structure of input data, the model is independent of concrete data storage and the service is implemented in such way that it should be independent from data retrieval as long as the data follows some biological patterns. Data retrieval is done by implementing proposed interfaces and merging them with the core service implementation service using one of dependency injection techniques when the service is exposed as a web service.

An idea of an integrated solution is also presented by Ames et al. [28]. The authors have developed web services-based software for hydrologic data discovery, download, visualization, and analysis using extensible plug-ins for searching, viewing and exporting data.

Although from a different field of study, cited papers yielded an idea of using web services, data interfaces and plug-ins for algorithms. Using service interfaces and various plug-ins it would be easier to try a different type of analysis or analyze the same thing using different algorithms without need to do it manually or to convert input and output data between different formats.

III. A PROPOSED SOLUTION

A. *Concept of centralized service-oriented solution and its interfaces*

Figure 1 presents a conceptual view of centralized service-oriented solution we propose. The application consists of data stored permanently or temporary on a server, a set of algorithms that use data to produce result to be returned to a user and a set of interfaces for communication with users allowing data and algorithm manipulation. Each of the arrows from the figure presents usage of an interface, which we would only enumerate in the paper providing input and output descriptively leaving format to be standardized in the future work.

Positions are main input data to perform any algorithm. They can be collected via mobile device or uploaded using a personal computer but also can be stored by an on-premise server. As mentioned previously, one of disadvantages of existing software is that a user must export his/her data, adjust it to required formats and upload it again to a proprietary service/application. In the proposed system, the user can keep data on his/her server and enable a service on his/her own data server according to one of formats that would be proposed as a standard. Instead of concrete data, the user could enter service location and optional filters and data would be retrieved later as needed.

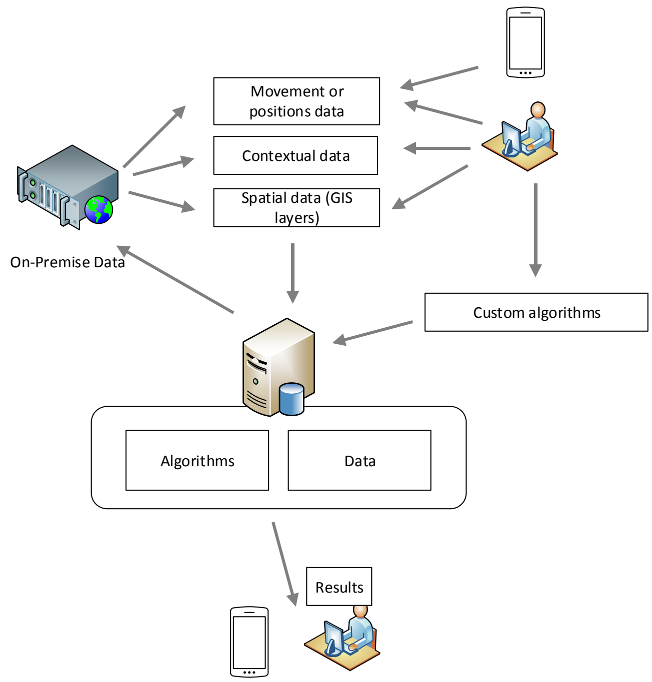

Figure 1. Concept of centralized service-oriented solution

The same principle applies also to contextual data and spatial data (GIS layers). Contextual data are usually environmental data, data about certain object, e.g., species etc. These data are collected from different sources: other data providing services, for example weather services [29][30] or they could be sent by user as well (for example, specific layer considering certain species). Contextual data are not limited in scope and additional information is facultative or obligatory depending on an algorithm requirements. As shown in Table I, some of analysis depends on used algorithm and a particular algorithm may need some additional data. Therefore these data can be of any type and format of data that can be recognized and used by a particular algorithm.

The idea is that algorithms are categorized and that there exists default algorithm per category but the application enables upload of a user's own algorithm. A solution must provide an API for access to movement and spatial data and any custom data from an input. An algorithm should implement an interface from a category it belongs to. Details, like language, format, runtime environment, etc., are left to implementation phase.

 ISBN: 978-1-61208-367-4

The results of algorithms can further be analysed and deployed to the server as well. Either retrieved on a temporary base or stored permanently, data stored on the server consists of data and algorithms which are both meta-modelled.

Summarizing all above, the set of interfaces used in Figure 1 can be described as in Table II. For each category of data, there are three possible interfaces: for data upload, for setting URL of a service that host user's data and for returning on-premise data once that centralized solution need users data.

B. *Meta-model*

To achieve adaptiveness and wide applicability, data and algorithms should be meta-modelled (Figure 2). Usage of a meta-model enables addition of algorithms, algorithms categories, domains and attributes that does not exist in the present moment but would occur in the future. Furthermore it enables integration of many different types of users under one, integrated model.

1) *Object and its characteristics*

Central entity of movement and biodiversity data is an *object* which represents either an animal or a plant. An object has recursive relationship to itself, thus allowing to model taxonomy (biological or zoological hierarchy of species) at any level - from unique instance of certain species to kingdom. For example, an object could be a certain animal (identified by name or a collar identifier), which is of species grey wolf (*canis lupus*), which is of genus *canis*, or it can be a species, e.g., *abies alba*, genus *abies* etc. Any of taxonomic ranks can have attributes which can comprise values from a certain domain. Values of defined attributes can be assigned to object itself. For example, species wolf can have attribute *social order*, while a particular wolf can have value of that attribute *alpha male*.

2) *Findings (positions)*

An object can be spotted at certain coordinate at certain timestamp. The attribute *source* represents type (source) of finding, for example GPS collar, terrain research, photo, literature citation etc. For example, a GPS collar carried by a wolf retrieves a recorded position of the wolf at a certain timestamp. Also, a researcher could see a wolf at a terrain and also record time and coordinates of finding. Another example is logging a plant's position by a researcher at the terrain or entering coordinates and timestamp of the plant from literature or a photo taken at the terrain.

Similarly to *time* and *source* other contextual data for a finding could be stored in the model. Optionally, if necessary, these contextual data can be related to attributes and domains.

3) *Spatial data*

Since spatial data is present in almost every kind of static or moving objects' positions analysis, they are meta-modelled as well. A layer is consisted of spatial elements (shapes) and layer attributes. By the term layer, we mean a GIS layer, such as a content of ESRI *shapefile* [31]. The shapes (usually polygons) belong to a certain layer. Values of defined attributes, which are valid in specific period of time, are assigned to certain shape. A coordinate (at which the object is spotted) belongs to certain shapes.

TABLE II. LIST OF INTERFACES

Category	**Purpose(s)**	**Input**	**Output**
Movement and positions	Data upload	List of n-tuples containing coordinates, species or object identifier and identifier of additional context data	
	Provides URL of a service hosted on a user's server (on-premise data)	URL of a service and custom data that should be send as-is to the service	On-premise server produce output same as used for input when data uploaded directly
Contextual data	Upload of species/objects data	List of n-tuples containing species/object identifier and additional parameters of any type (e.g., byte array) that must be interpreted by an algorithm	
	Upload of contextual information, i.e., those that are related to positions	List of n-tuples containing context identifier and additional parameters of any type (e.g., byte array) that must be interpreted by an algorithm	
	Provides URL of a service hosted on a user's server (on-premise data)	URL of a service for contextual data	On-premise server produce output same as used for input when data uploaded directly
Spatial data (GIS layers)	Data upload	Zipped folder containing one of supported GIS formats (e.g., ESRI shp+dbf+shx)	
	Provides URL of a services hosted on a user's server (on-premise data)	URL of a service that serves a GIS layer (e.g., WMS server)	Zipped folder or WMS server
Custom algorithms	Custom algorithm upload	Algorithm code	
Results	Results of an algorithm		One or more matrices that summarize algorithm results and optional output from custom algorithm that must be interpreted by a user
Analysis request	Initiate analysis	Type of analysis, chosen algorithm, optional filter on context and data	

4) *Algorithms*

Several classes of algorithms used in movement and biodiversity analysis are summarized in Table II. In each category, input and output types can be identified and these input and output parameters can be meta-modelled using entity *Attribute* thus describing interfaces for all algorithms in the category. In order to enable custom extension of parameters, the last parameter should always be in free form (e.g., XML or array of bytes) for custom data.

One or more algorithms can exist for each category and one of them is default one for the category. An algorithm can have additional parameters needed for a concrete implementation. Each algorithm instance produces results using data from findings, coordinates and shapes which do not have to be modelled as a relation in a model, but should be available using some form of API in implementation. Results from an algorithm instance are stored according to output parameters attributes.

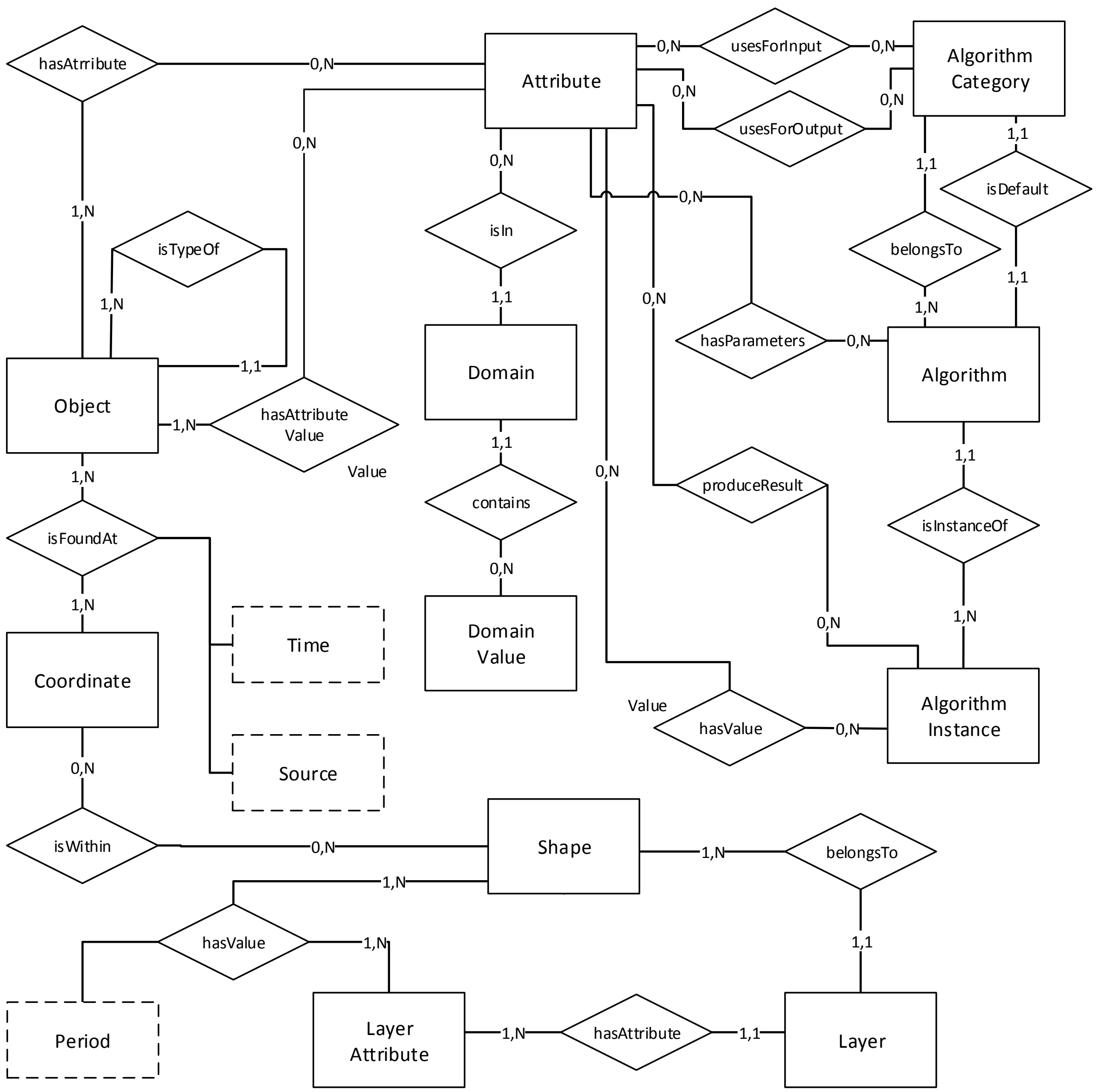

Figure 2. Meta-model for movement and biodiversity data and algorithms

IV. CONCLUSION AND FUTURE WORK

The proposed solution and the designed model establish a baseline for a concrete implementation. To achieve interoperability and openness, the proposed solution should be exposed as (web) services with clearly defined standards of input/output data. Future work should be related to the standardization of input and output formats. Furthermore, an appropriate solution for writing algorithms must be identified (in place compilation of a code in some specific language or custom Domain Specific Language) and APIs for data access must be defined.

REFERENCES

[1] S. Dodge, R. Weibel, and A.-K. Lautensch, "Towards a taxonomy of movement patterns," Information Visualization, vol. 7, 2008, pp. 240-252

[2] J. Drummond, R. Billen, E. Joao, and D. Forrest, Dynamic and Mobile GIS: Investigating Changes in Space and Time, CRC Press, Taylor & Francis Inc, 2006.

[3] C. Parent, et al., "Semantic trajectories modeling and analysis," ACM Comput. Surv., vol. 45, 2013, pp. 1-32

[4] W. Wang and G. Wets, Computational Intelligence for Traffic and Mobility: Atlantis Publishing Corporation, 2013.

[5] C. Barrios and Y. Motai, "Improving Estimation of Vehicle's Trajectory Using the Latest Global Positioning System With Kalman Filtering," IEEE T. Instrumentation and Measurement, vol. 60, 2011, pp. 3747-3755

[6] J. Krumm and E. Horvitz, "Predestination: Inferring Destinations from Partial Trajectories," UbiComp 2006: Ubiquitous Computing, vol. 4206, 2006, pp. 243-260

[7] J. Froehlich and J. Krumm, "Route Prediction from Trip Observations," presented at the Society of Automotive Engineers (SAE) 2008 World Congress, 2008.

[8] J.-M. François, G. Leduc, and S. Martin, "Learning movement patterns in mobile networks: a generic method," presented at the European Wireless 2004, Barcelona, Spain, 2004.

[9] G. Yavas, D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos, "A data mining approach for location prediction in mobile environments," Data Knowl. Eng., vol. 54, 2005, pp. 121-146

[10] A. Franke, T. Caelli, G. Kuzyk, and R. J. Hudson, "Prediction of wolf (Canis lupus) kill-sites using hidden Markov models," Ecological Modelling, vol. 197, Aug 10 2006, pp. 237-246

[11] N. F. Webb, M. Hebblewhite, and E. H. Merrill, "Statistical methods for identifying wolf kill sites using global positioning system locations," Journal of Wildlife Management, vol. 72, Apr 2008, pp. 798-807

[12] A. D. Middleton, et al., Linking anti-predator behaviour to prey demography reveals limited risk effects of an actively hunting large carnivore, 2013, pp. 1023–1030

[13] L. O. Alvares, et al., "A model for enriching trajectories with semantic geographical information," presented at the Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, Seattle, Washington, 2007.

[14] B. Guc, M. May, Y. Saygin, and C. Körner, "Semantic Annotation of GPS Trajectories " presented at the 11th AGILE International Conference on Geographic Information Science 2008, University of Girona, Spain, 2008.

[15] B. Milašinović, T. Nikolić, and K. Fertalj, "Biodiversity analysis supporting species-subspecies uncertainty in findings data," International journal of biology and biomedical engineering, vol. 7, 2013, pp. 127-134

[16] Y. Ishikawa, Data Mining for Moving Object Databases. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010.

[17] S. Elnekave, M. Last, and O. Maimon, "Measuring Similarity Between Trajectories of Mobile Objects," in Applied Pattern Recognition. vol. 91, H. Bunke, et al., Eds., ed: Springer Berlin / Heidelberg, 2008, pp. 101-128.

[18] I. Nižetić Kosović, K. Fertalj, and J. Kusak, "An expert system for discovering biogeographically interesting locations from animal movement data," presented at the European Computing Conference, Paris, France, 2011, pp. 348-353

[19] S. Akoush and A. Sameh, "Movement Prediction Using Bayesian Learning for Neural Networks," presented at the International conference on Wireless communications and mobile computing, Honolulu, Hawaii, USA, 2007, pp. 191-196

[20] R. Joo, S. Bertrand, J. Tam, and R. Fablet, "Hidden Markov Models: The Best Models for Forager Movements?," PLoS ONE, vol. 8 , 2013,

[21] P. E. Smouse, et al., "Stochastic modelling of animal movement," Philosophical Transactions of the Royal Society B-Biological Sciences, vol. 365, Jul 27 2010, pp. 2201-2211

[22] I. Nižetić, K. Fertalj, and D. Kalpić, "A Prototype for the Short-term Prediction of Moving Object's Movement using Markov Chains," presented at the 31st International Conference on Information Technology Interfaces, Cavtat, Croatia, 2009, pp. 559 - 564

[23] J. Bedia, J. Busqué, and J. M. Gutiérrez, "Predicting plant species distribution across an alpine rangeland in northern Spain. A comparison of probabilistic methods," Applied Vegetation Science, vol. 14, 2011, pp. 415-432

[24] J. Xu and R. H. Guting, "A generic data model for moving objects," Geoinformatica, vol. 17, 2013, pp. 125-172

[25] V. Bogorny, C. Renso, A. R. d. Aquino, F. d. L. Siqueira, and L. O. Alvares, "CONSTAnT – A Conceptual Data Model for Semantic Trajectories of Moving Objects," Transactions in GIS, vol. 18, 2014, pp. 66-88

[26] F. Urbano, et al., "Wildlife tracking data management: a new vision," Philosophical Transactions of the Royal Society B-Biological Sciences, vol. 365, Jul 27 2010, pp. 2177-2185

[27] I. Nižetić and K. Fertalj, "Automation of the Moving Objects Movement Prediction Process Independent of the Application Area," Computer Science and Information Systems, vol. 7, 2010, pp. 931-945

[28] D. P. Ames, et al., "HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis," Environmental Modelling & Software, vol. 37, 2012, , pp. 146-156

[29] M. Kanamitsu, et al., "NCEP-DEO AMIP-II Reanalysis (R-2)," Bulletin of the American Meteorological Society, vol. 83, 2002, pp. 1631-1643

[30] NOAA. (2011, 01.06.2011). Earth System Research Laboratory. Available: http://www.esrl.noaa.gov/

[31] ESRI. *ESRI Shapefile Technical Description. An ESRI White Paper—July 1998* [Online]. Available from: http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf 2014.08.15

 ISBN: 978-1-61208-367-4