Towards a semantic framework for wildlife modeling

Ioannis N. Athanasiadis[1], Ferdinando Villa[2], Georgina Examiliotou[1], Yorgos Iliopoulos[3], Yorgos Mertzanis[3]

Abstract

In this paper we present work in progress for developing a semantic modeling system for wildlife monitoring, management and conservation. Based on a Greek NGO experience in large carnivores conservation in the mountain ecosystems of northern Greece, we present a generic architecture for wildlife information fusion, sharing and reuse. Our framework employs ontologies for representing the key domain concepts and their relationships, and applies them for integrating sensory information from GPS/GSM animal tracking devices, along with other field data and habitat suitability models.

1. Introduction

Integrated environmental modeling deals with modern environmental problems, decisions, and policies. Computer tools can serve as enablers, allowing the environment to be represented in a holistic way [1]. Among the major challenges for integrated environmental modeling and decision support tools, Laniak *et al* [1] identify their capacity to exhibit (a) adaptive decision making, i.e. to be able to "*learn as we go*", and (b) seamless access to environmental data, with an increased capacity of interoperability at a conceptual level. All this is in line with the vision of semantic environmental modeling, introduced by Villa *et al* in [2], and hints to a *knowledge-driven approach* to environmental modeling that employs ontologies for expressing model statements.

The work presented here introduces an adaptive and interoperable decision support system for modeling large carnivore habitat that employs ontologies for integrating animal tracking data with ecological niche modeling. Employing semantics for representing knowledge encapsulated in the domain, either as evidence (data) or as models interpreting the system behaviour, the framework may improve decision in wildlife management, and uses sensory information to adapt to ever-changing environmental conditions, while at the same time maximizing the capacity for data and model reuse.

The paper is structured as follows. Section 2 presents some background on the case study, the project supporting this work and its challenges. Section 3 presents system goals and users, and in Section 4 system components are introduced. The paper concudes with a short discussion in Section 5.

2. Case study: From animal tracking data to adaptive decision support

Callisto is a Non-Governmental Organization operating in Northern Greece, dedicated to the study, conservation and management of populations and habitats of large carnivores and other endangered fauna species. Over the past 10 years, Callisto has been active in the Pindos, Rodopi and Gramos

[1] Democritus University of Thrace, Xanthi, Greece, ioannis@athanasiadis.info
[2] Basque Center for Climate Change, Bilbao, Spain
[3] Callisto, Thessaloniki, Greece

mountains, supporting conservation of the brown bear *(Ursus arctos)* population which is the last stronghold of the species at its southernmost range in Europe, with 350-400 individuals in Greece. During all these years of activity, Callisto has collected an invaluable amount of monitoring information on large carnivores (bears, wolves) that have been used in several projects co-funded by the European Commission [3]. Callisto uses GPS/GSM collars, IR photo-traps and/or transects to identify animal activity and presence.

The ALPINE project, co-funded by the Greek Secretariat of Research and Technology, aims at enhancing this approach by developing low-cost, low-power sensor network solutions that can be used by Callisto for wildlife monitoring. While details on the sensor architecture are outside of the scope of this article, this has been seen as an opportunity to investigate animal monitoring data fusion methods, and develop new tools that enable interoperability and adaptive decision support.

One of the ALPINE project case studies deals with the human-bear interaction in the case of the Egnatia highway, which crosses the Pindos mountain range in northern Greece, an important habitat for brown bears. The Egnatia highway construction has affected bears in several ways: population structure, movement, foraging and breeding patterns have changed due to the destruction of natural habitat. Additional pressure was created by disturbance along the highway, which forced bears to re-adapt to a human-shaped landscape within their original habitat. Furthermore, traffic fatalities involving bearstrying to cross the highway have increased public concern about both bear conservation and human safety. As a result, a bear-proof fence has been installed along 130 km of the highway.

Our case study provides a challenging problem for integrated modeling that calls for both regular re-assessments and adaptive decision support. We aim to dynamically integrate animal tracking data (originating from sensor networks) with state-of-the-art ecological models to assess human-bear interactions. Our system aims to make effective use of semantics for annotating animal tracking streams, along with background data and domain models to analyze them, and eventually produce maps and reports. Such a scientific workflow will be run as required when new data streams arrive from sensors. Capturing the semantics of environmental information in the wildlife domain is one of the key challenges of this work, in order to exploit existing and new resources, to annotate information connected with movement patterns, ecological requirements of wildlife species (bears, wolves), to represent knowledge that can be interpretable by both humans and computer applications, and to simplify the use of automated reasoning in data management tasks.

3. System goals and users

The *ALPINE system for wildlife modelling* (hereafter, ALPINE for short) aims to demonstrate how live streaming data from animal tracking sensors can be effectively combined with geo-statistical analysis models, in order to assess habitat suitability, and to quantify the risks of wildlife interaction with man-made infrastructures. The overall system architecture is presented below in Figure 1, and consists of three layers. First is the *data layer* that incorporates animal tracking data, with eco-geographical field data and infrastructure networks. Second, the *integrated modelling layer* employs statistical, geospatial and Bayesian models for ecological niche factor analysis. Third, the *presentation layer* generates maps and reports with the system results.

The system is intended for scientists who aim to answer questions related to habitat suitability and wildlife-human interactions. Indirect users include policy makers, the industry and other societal sectors who are interested in such studies. The ALPINE system enables scientists to hook up sensor data streams coming live from sensors with geographical information and build scientific workflows that enable integrated modeling studies. User requirements are:

a. minimizing human involvement in data preprocessing and manipulation, especially as new data arrive from sensors;

b. easing model re-runs, as new data arrive; and

c. providing tools for exporting results in different formats.

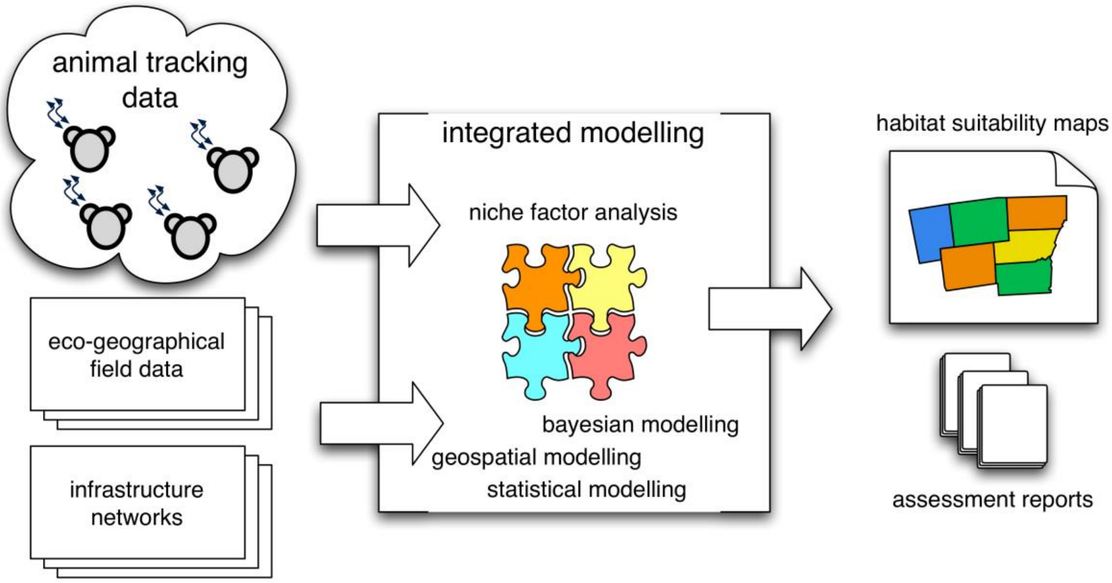

Figure 1: The ALPINE architecture for wildlife monitoring

4. System components

4.1. Evidence from sensor networks

Coupling models with data is a common problem that most scientists face, and e-science tools for integrating models and data came as a remedy that enables the mechanics of integration. Integrated modelling frameworks and scientific workflow engines, as Kepler (Ludäscher *et al*. 2006), OpenMI (Knapen *et al*. 2013), OMS (David *et al*. 2013), or Bioma (Donatelli *et al*. 2014) help scientists with the technical task of linking models to create compositions of different kinds, i.e. workflows that may involve numerical integration, parameter estimation, sensitivity analysis, execution on computer grids, or visual analysis tools. Integrating sensor data with integrated modeling frameworks and/or scientific workflow systems "*presents technical challenges that are difficult for scientists to overcome*" [9], due to the heterogeneity of the data streams and the technologies involved. Sensor data may arrive directly as streams, or become available from archives. In all cases, simplifying access via reusable data protocols and making data discoverable through rich annotations remains a challenge. To this end, the Open Geospatial Consortium provides standards for sharing sensor data such as SensorML or Sensor Observation Service [10], and INSPIRE [21] for datasets of geographical nature (i.e. elevation, land cover, water bodies, etc). New observational (field) data, collected by ALPINE sensors (i.e. from GPS/GSM collars), can be made available in the system using a service-oriented philosophy [12] that follows the Sensor Observation Service or Open Archives Initiative [11] protocols.

4.2. Interpretation via modelling

Sensor information alone is not sufficient for answering research questions. Evidence needs to be coupled with models in order to address issues of habitat suitability and estimating risks from human-wildlife interactions. Environmental models provide such interpretations, as they embody explanations or estimations of system behaviour. In the ALPINE system, we identified a need for three kinds for models. First, geo-spatial models that allow operational interpretations of spatial sources are typically used for creating derived information from original data. These include buffering functions, and density analysis. Secondly, Bayesian models should build probabilistic models in order to incorporate causal associations from evidence. For a more detailed discussion on Bayesian modelling for ecological risk assessments see [13]. Third, Ecological Niche Factor Analysis (ENFA) is a statistical procedure that uses only presence data, suitable to compare distributions among spaces that a population has a reasonable probability to occur using eco-geographical variables and the global space [14]. The ALPINE wildlife monitoring system aims to integrate these three components in order to enable scientists to perform their assessments.

4.3. Result presentation and communication

The last component of the system deals with presentation of results. Typically scientists spend adequate amount of time in order to analyse their results and post-process them. In the ALPINE system we aim to incorporate such aspects in the workflow, so that maps and reports are generated as new data arrive in the system and assessments are updated. For this we employ reusable templates that will incorporate model results.

4.4. Enabling semantic modelling and implementation

Ontologies that define the domain concepts and their interactions remain the heart of the system that will be built upon *Thinklab*, a semantic meta-modelling platform developed initially in the ARIES project [4], and currently under development as an open source software project [15]. Thinklab provides with the main infrastructure for a semantic modelling, as it offers:

(a) a reusable set of core ontologies, for defining domain concepts as observations, and quantify their qualities;

(b) a domain specific language, to define semantic models and prescribe behaviour; and

(c) a collaborative e-science space where modellers may share ontologies, data and models.

Thinklab remain outside of the scope of this paper and the reader is encouraged to visit its website [16]. We consider domain specific programming languages for environmental modelling to be key for advancing the notion of semantic modelling, where all concepts used to model natural systems are explicitly defined by ontologies [17]. Towards this end, we directed our developments for extending Thinklab, in several fronts in order to achieve ALPINE functionality.

First, we are in a process of extending the core Thinklab ontologies in order to accommodate for wildlife trajectory semantics. Based on previous research on semantic trajectories [18], we are investigating how to formally represent animal movements and accompany them with qualities that enable end-users to derive useful knowledge for their studies. A collaborative approach with scientists from Callisto and other organizations is in the process for enabling and validating this approach. The key element of our view is that an animal trajectory needs to encapsulate the species circadian rhythm and express the interactions of an individual with its habitats.

The second challenge is to incorporate sensor stream data sources in Thinklab. By extending the input functionality of the platform we will make available standard protocols for sharing sensor information. Currently, we plan to make sensor data available via simple protocols, like GeoRSS [19] and GeoJSON [20], and more sophisticated ones as OGC Sensor Observation Services [10]. In all cases, our Thinklab extensions will enable annotating sensor streams with semantics, so that they can be used transparently within a semantic modelling workflow. This goes a step ahead of just reading sensor data streams, as it requires associating observational semantics to the streams. As an example, consider the animal tracking data that we will gather from large carnivores collars in the Callisto case. Thinklab will interpret them based on the trajectory ontology under development, and ultimately make them available as such for subsequent use.

Third, we will incorporate the Ecological Niche Factor Analysis (ENFA) in the Thinklab platform. As Bayesian modeling and spatial analysis modeling are already available in the Thinklab platform, we will further extend it with ENFA in order to be able to perform statistical modeling on habitat suitability.

4.5. Callisto user story

For the Callisto case, we envisage the following user story. *Yorgos* an ecologist wants to perform a habitat suitability assessment of human-bear interaction in certain parts of the Egnatia highway, where car accidents occur often and involve bear fatalities. The first step is to collect field data for the specific area of study from various sources, including the local government, the ministry of the environment, the company managing the highway, bibliographic information and in-house information. Yorgos, using Thinklab, can reuse field data from previous studies, while he may annotate new public datasets made available with standard protocols. The second step is to capture and install collars on certain individuals, which will eventually start producing animal tracks. Yorgos makes the sensor data stream available in the platform by associating to it with appropriate semantics, and bear trajectories are identified as they arrive. With such information available Yorgos easily hooks up bear trajectories and ecological field data to set up a niche factor analysis model. The result of the model run is a habitat suitability map produced in Thinklab which is directly exported for his report. The workflow remains available in Thinklab and Yorgos may reuse it as new data arrive or in other areas of interest.

5. Discussion

In this paper we presented work in progress for developing a semantic framework for wildlife modeling. The system presented employs semantic meta-modeling and ontologies for representing domain knowledge, which includes both data and models. The system under development builds upon Thinklab, an existing open-source software, and extends it with an animal trajectory ontology, tools for reading sensor information from standard protocols and an implementation of the niche factor analysis models. Our implementation will be evaluated for a case study in Greece for assessing habitat suitability, and to quantify the risks of brown bear interaction with man-made infrastructures.

Acknowledgements

Research in this study has received funding by European and national funds from NSRF 2007-2013, OP Competitiveness and Entrepreneurship, Cooperation 2011, in the context of the ALPINE project (grant 11SYN-6-411).

References

1. Laniak, G. F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., ... & Hughes, A. (2013). Integrated environmental modeling: a vision and roadmap for the future. *Environmental Modelling & Software*, *39*, 3-23. doi:10.1016/j.envsoft.2012.09.006.
2. Villa, F., Athanasiadis, I. N., & Rizzoli, A. E. (2009). Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. *Environmental Modelling & Software*, *24*(5), 577-587. doi:10.1016/j.envsoft.2008.09.009
3. Kaczensky, P *et al* (eds) (2012), Status, management and distribution of large carnivores – bear, lynx, wolf & wolverine – in Europe. European commission, available online: http://ec.europa.eu/environment/nature/conservation/ species/carnivores/conservation_status.htm
4. Villa F, Bagstad KJ, Voigt B, Johnson GW, Portela R, et al. (2014) A Methodology for Adaptable and Robust Ecosystem Services Assessment. PLoS ONE 9(3): e91001. doi:10.1371/journal.pone.0091001
5. Ludäscher, B.; Altintas, I.; Berkley, C.; Higgins, D.; Jaeger-Frank, E.; Jones, M.; Lee, E.; Taoand, J. & Zhao, Y. Scientific Workflow Management and the Kepler System. Concurrency and Computation: Practice and Experience, 2006, 18, 1039-1065
6. David, O.; Ascough II, J.; Lloyd, W.; Green, T.; Rojas, K.; Leavesley, G. & Ahuja, L. A software engineering perspective on environmental modeling framework design: The Object Modeling System. Environmental Modelling & Software , 2013, 39, 201 – 213
7. Knapen, R.; Janssen, S.; Roosenschoon, O.; Verweij, P.; de Winter, W.; Uiterwijk, M. & Wien, J.-E. Evaluating OpenMI as a model integration platform across disciplines. Environmental Modelling & Software , 2013, 39, 274 – 282
8. Donatelli, M.; Bregaglio, S.; Confalonieri, R.; De Mascellis, R. & Acutis, M. A generic framework for evaluating hybrid models by reuse and composition - A case study on soil temperature simulation. Environmental Modelling & Software, Elsevier, 2014.
9. Barseghian, D.; Altintas, I.; Jones, M. B.; Crawl, D.; Potter, N.; Gallagher, J.; Cornillon, P.; Schildhauer, M.; Borer, E. T.; Seabloom, E. W. & Hosseini, P. R. Workflows and extensions to the Kepler scientific workflow system to support environmental sensor data access and analysis. Ecological Informatics, Elsevier, 2010, 5, 42-50.
10. OGC (2007). Sensor Observation Service, Open Geospatial Consortium Standard. Available online: http://www.opengeospatial.org/standards/sos.
11. OAI (2002). Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Available online: http://openarchives.org.
12. Samourkasidis, A. & Athanasiadis, I. N. Towards a low-cost, full-service air quality data archival system. Proc. 7th Intl. Congress on Environmental Modelling and Software, International Environmental Modelling and Software Society (iEMSs), 2014.
13. Pollino, C. A.; Woodberry, O.; Nicholson, A.; Korb, K. & Hart, B. T. Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment Environmental Modelling & Software, 2007, 22, 1140-1152.
14. Hirzel, A. H.; Hausser, J.; Chessel, D. & Perrin, N. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data. Ecology, 2002, 83, 2027-2036.
15. Villa, F. and others (2013) Thinklab software repository. Online: bitbucket.org/ariesteam/thinklab.
16. Integrated Modelling community. Online at: http://www.integratedmodelling.org
17. Athanasiadis, I. N. & Villa, F. A roadmap to domain specific programming languages for environmental modeling: key requirements and concepts Proceedings of the 2013 ACM workshop on Domain-specific modeling, 2013, 27-32.
18. Spaccapietra, S.; Parent, C.; Damiani, M. L.; de Macedo, J. A. F.; Porto, F. & Vangenot, C. A conceptual view on trajectories. Data and Knowledge Engineering, 2008, 65, 126-146.
19. GeoRSS Specification. Online: http://www.georss.org
20. GeoJSON Specification. Online: http://www.geojson.org
21. INSPIRE: Infrasructure for Spatial Infromation in the European Community. Online: http://inspire.ec.europa.eu/