Visualizing Uncertainty In Environmental Work-flows And Sensor Streams

Karthikeyan Bollu Ganesh and Patrick Maué

Institute for Geoinformatics (IFGI), University of Muenster,
D-48151 Muenster, Germany
{karthikeyan,pajoma}@uni-muenster.de

Abstract. Environmental data and models are uncertain by nature. The lack of knowledge about, for example, the magnitude of potential measurement errors may lead to unforeseen consequences. This makes it difficult to assess the data's or model's usefulness for critical applications. We present an approach for the visualization of uncertainty coming from in-situ environmental sensors. The visualization component is part of a Web-enabled environmental modelling platform which also supports the specification of processing workflows. The concept of uncertainty and means for its encoding as part of the environmental data are introduced. The individual components in the processing workflows propagate and update the uncertainty information. We also explain how the uncertainty in the original sensor data has been identified, and how the visualization component has been implemented.

Keywords: Uncertainty,Uncertainty Visualization, Environmental Workflows

1 Introduction

The on-going implementation of the European INSPIRE directive facilitated the access to geographic information on the Web. Embedded in Spatial Data Infrastructures (SDI), geospatial Web services provide means to access, process, and visualize spatial data. Interoperability between different Web services is to a certain degree ensured by the standards set by the Open Geospatial Consortium. The standards also enable the direct integration into local geospatial applications. But there has been only limited uptake by the environmental modelling community. The concept of a Shared Environmental Information System (SEIS) has been recently introduced to address this issue [2]. The published SEIS principles aim to facilitate access to environmental information for public authorities and the general public. This calls for new ways of how to publish environmental services, and how to present the resulting information to also non-ICT skilled end-users. Investigating the requirements for environmental services infrastructures, including the execution of environmental services via standard runtime engines for service work-flows and the Web-enabled visualization through environmental portals, is subject of the research project ENVISION[1] [6]. ENVISION aims to provide an environmental services infrastructure with ontologies which investigates the distributed execution, semantic discovery and annotation of environmental services. Results from our research regarding the encoding, propagation, and visualization of uncertainty are presented in this paper.

2 Environmental Models As Workflows

Environmental information is traditionally resulting from environmental computer models. Simple computer models could perform atomic computations, such as interpolating the elevation within the DEM (Digital Elevation Model). To solve complex problems such as predicting the impact of an oil spill on the local wildlife, the chaining of individual simple models might be required. In a first step, a weather forecasting service encapsulated as an OGC Web Processing Service takes real-time weather data from an OGC Sensor Observation Service to predict future weather conditions. The result is taken as input for the oil drift model, which computes the dispersion and weathering of an oil slick for the forecasting period. As final step, this distribution may be used to assess the impact of the oil slick on the local wildlife (e.g. using ESI[2] maps). Each of these models are implemented as WPS, the coupling with other Spatial Data Services (SDS) serving the input is done with the Business Process Modelling Language (BPEL). Standard BPEL engines can then execute these workflows.

3 Uncertainty In Environmental Data

Results from the workflows help domain experts in their decision-making tasks, e.g. for assessing response measures for an oil spill. Oil spills could be fought by distributing detergents, skimming, or even burning. Which method to apply depends on the results of the oil spill model. Taking the wrong decisions here could potentially have devastating effects on the environment; having correct and highly certain results is therefore crucial. Much of the data that forms the input for the computer models contains some sort of uncertainty. Environmental information is uncertain by nature. "You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. Uncertainty is everywhere and you cannot escape from it" [5].

3.1 Uncertainty in Work Flows

Uncertainty is an expression of confidence about our knowledge [4]. Uncertainty results from imprecise measurements of environmental phenomena. It is processed using geospatial algorithms, which generalise, infer, or merge data to

[1] More information available at http://www.envision-project.eu
[2] ESI stands for "Environmental Sensitivity Index"

generate new data. Uncertainty is part - and therefore an important aspect - of GI throughout the complete services work flow, starting with the creation of the data until its visualisation on a map [3]. The following two types of uncertainty should be explicitly quantified to enable users to assess the quality of the end product of environmental service chains. Input errors result from imprecise measurements, either due to human error or insufficient sensing technology. The human factor or wrongly calibrated sensors are hard to quantify, while errors coming from sensors not sensitive enough can be known beforehand. Model errors emerge during the processing of the data, e.g. the interpolation of unknown values from sparse input data. The uncertainty of the resulting product depends on both, input and model error. That requires a solution to add uncertainty information to the data, and to keep and update uncertainty information while processing the underlying data.

Processing uncertain environmental data propagates the uncertainty often unpredictably [8]. Environmental models may introduce their own model errors, which influence the original uncertainty of the input observational data. The introduced error depends on the selected algorithms. This work is focussing on the visualization of uncertainty coming from work flows of environmental data and processing services. Discussing how to compute the uncertainty coming from the models is out of scope of this paper. This is addressed by the research currently performed in the UncertWeb[3] project; here we have to assume that the processing components in the work flow are able to propagate and update the uncertainty parameters. Finding adequate visualization techniques for uncertainty is an active research topic. Examples for such techniques are portraying uncertainty in graphs by showing normal distribution and confidence intervals, colour models, or time-series charts. The following implementation presents a tool showing the uncertainty in environmental data as charts.

4 Implementation

The implementation of this project is divided into the following sub-tasks: (1) Identification of existence data quality parameters, (2) extending the standard data format for sensor data with uncertainty information, and (3) visualizing uncertainty in a graph.

4.1 Identification Of Existing Uncertainty Parameters

Information about data quality can be either qualitative (e.g. adding information about the data provider to address issues such as trust) or, in most cases, quantitative (e.g. completeness, accuracy, scale, and more). Data quality has been acknowledged to play an important role for geographic information, and OGC and ISO published standards for representing data quality parameters. This work is focussing on the accuracy of sensor measurements. A piezometric

[3] More information available at: http://www.uncertweb.org/

sensor system is used for ENVISION to monitor underground water levels. The sensed data can be accessed through an OGC SOS[4] interface. The accuracy of the measured data from these piezometers is found to be around 1% of the measured value[5].

4.2 Uncertainty Extensions For O&M 2.0

Sensed data always comes with some sort of error. From a conceptual perspective all data should be considered to be uncertain. Even though most data lacks information about uncertainty, some data sets may have descriptive information about it in its metadata (following the ISO 19013 standard for data quality metadata). This global definition of uncertainty is many cases insufficient for the visualization and assessment of the data set's usefulness for critical application. The Uncertainty Markup Language (UncertML) has been introduced as extension to the OGC Geography Markup Language to address this issue. It is focussed on an XML encoding for the transport and storage of uncertainty information [8]. UncertML includes means to express simple summary statistics (e.g., mean and variance) as well as complex representations such as parametric, multivariate distributions at each point of a regular grid [7]. Uncertainty can be encoded either in form of (1) statistics, e.g. values for probability or the quantile, (2) distributions, e.g. a normal distribution, or as (3) realisations. UncertML 2.0 relies on the OGC Observation & Measurement (O&M) standard to encode uncertainty [1]. The following listing for the piezometer observations encodes Uncertainty as a Gaussian distribution. The mean value is to be understood to be the actually sensed value, while the variance reflects the precision of the sensor measurement.

```
<om:resultQuality>
  <gmd:DQ_QuantitativeAttributeAccuracy>
    <gmd:result>
      <gmd:DQ_UncertaintyResult>
        <gmd:value>
          <un:GaussianDistribution>
            <un:mean>0.89 0.87 </un:mean>
            <un:variance>0.01 0.01</un:variance>
          </un:GaussianDistribution>
        </gmd:value>
      </gmd:DQ_UncertaintyResult>
    </gmd:result>
  </gmd:DQ_QuantitativeAttributeAccuracy>
</om:resultQuality>
```

Listing 1 - Example of uncertainty-enabled O&M 2.0

[4] http://swe.brgm.fr/pleiade-core-service-ades-om2-0.0.1-recette/REST/sos? Request parameters and examples are available at: http://sosades.brgm.fr/

[5] As reported by the service providers BRGM, the French geological survey

4.3 Visualization

The implementation adopts the visualization through the charts. The component expects the element Gaussian Distribution and computes the according uncertainty intervals. A time-series chart displays the maximum, minimum value of the uncertainty and the mean value of the incoming observations. The following screen-shots includes the chart viewer on the left side, and the map showing the according sensor positions on the right side. They belong to a set of components developed in the ENVISION project. The individual modules are implemented as Portlets (compliant to the Java Portlet Specification 286[6]), which can be best described as pluggable user interface components for the Web. By simply selecting one of the sensors displayed in the map, the according time series can be visualized in the chart. The red line represents the actually observed values (the mean), while the blue lines represent the according boundaries of the uncertainty intervals. The chart is based on a JavaScript library[7] which supports rich interaction with the graph. The user can hover over the chart to see the individual values at different points in time.

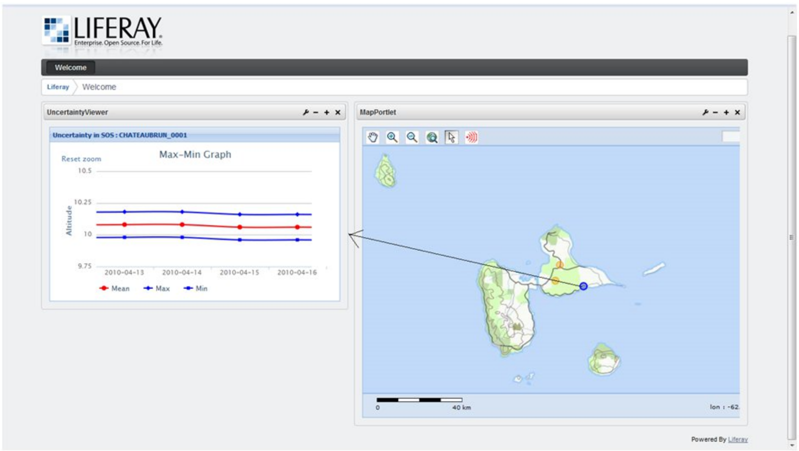

Fig. 1. Uncertainty Viewer in the Envision Infrastructure

5 Conclusion

The focus of this work is on the visualization of uncertainty resulting from measurement errors and the processing of the environmental data. It relies on the encoding of the uncertainty using the UncertML standard. How to come up with

[6] More information available at: http://www.jcp.org/en/jsr/detail?id=286
[7] More information available at: http://www.highcharts.com/

this uncertainty information - and how to propagate uncertainty in geospatial work flows - has not been subject of this research. This is investigated in the research project UncertWeb. Hence, future work will focus on integrating the other features supported by UncertML in the visualization components. This includes research on the usability, i.e. how can we best communicate uncertainty to support the end-users in the decision making process. In ENVISION the execution of the geospatial work-flow is handled by a distributed execution infrastructure. It includes techniques for the optimization of the work-flows, and the ad-hoc adaptation of execution paths according to certain context parameters. Future work will also investigate how uncertainty information may contribute to this adaptation process.

6 Acknowledgements

This work has been funded by the European research project ENVISION (FP7-249170, see `http://www.envision-project.eu`)

References

1. Cox S.: Observations and Measurements. Technical report 05-087, OGC. (2006)
2. European Commission. COM 46 final: Towards a Shared Environmental Information System (SEIS). Communication from the Commission. (2008)
3. Grønmo, R., Roman, D., Llaves, A., & Maué, P.: ENVISION D3.1: MaaS Composition Portal - Architecture specification. Available at `http://www.envision-project.eu` (2010)
4. Heuvelink, G. B. M., Brown, J. D., & Loon, E. E. van.: A probabilistic framework for representing and simulating uncertain environmental variables. International Journal of Geographical Information Science, 21(5), 497-513. (2007)
5. Lindley, D. V.: Understanding Uncertainty (p. 272). Wiley-Interscience (2006)
6. Roman D., S. Schade, A. Berre, N. Bodsberg, and J. Langlois.: Environmental Services Infrastructure with Ontologies - A Decision Support Framework. In Proceedings of EnviroInfo Conference 2009, Berlin, Germany. (2009)
7. Williams, M., Cornford, D., Bastin, L., & Ingram, B.: UncertML : an XML schema for exchanging uncertainty . Proceedings of GISRUK, Manchester, UK. (2008)
8. Williams, M., Cornford D., Bastin L., & Pebesma E.: OGC Discussion Paper 08-122r: Uncertainty Markup Language (UncertML). Available at: `http://portal.opengeospatial.org/files/?artifact_id=33234` (2009)