International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Uncertainty propagation between web services – a case study using the eHabitat WPS to identify unique ecosystems

Michael Schulz[a], Jon Olav Skøien[a], Lydia Gerharz[b], Grégoire Dubois[a], Edzer Pebesma[b],

[a] *European Commission - Joint Research Centre (JRC), Via Fermi 2749,TP 440, I-21027 Ispra (VA), (michael.schulz,jon.skoien,gregoire.dubois)@jrc.ec.europa.eu*
[b] *Institute for Geoinformatics, University of Muenster, Weseler Strasse 253, D-48151 Muenster,(Lydia.Gerharz,edzer.pebesma)@uni-muenster.de*

Abstract: Protected areas (PAs) are designed to protect ecosystems and their associated species against anthropogenic threats. When assessing the importance of the current network of PAs, and when considering new areas which should be protected, one of the criteria is the uniqueness of the ecosystems found inside the existing or planned PA when compared to other parks. As a helping tool for park managers and potential funders, eHabitat has been designed using the Open Geospatial Consortiums (OGC) Web Processing Service (WPS) interface specification. It allows end-users to compute, using different data and models, the likelihood of finding ecosystems with similar properties, as well as the potential changes these areas are exposed to according to different climate change scenarios. The most important input parameters, typically thematic geospatial "indicator layers" characterizing the ecosystem, are provided to the WPS as references using standardised web services or catalogues. This allows for a virtually infinite number of combinations to describe these ecosystems. However, the layers used can range from geophysical data captured through remote sensing to socio-economical indicators. eHabitat is therefore exposed to a broad range of different types and levels of uncertainties. Assessing these uncertainties, and as an additional component further propagating them when potentially included in a chain of model services, is a key aspect in the context of the Model Web. The use of the Uncertainty Markup Language (UncertML) as developed within the UncertWeb project to promote interoperability between data and models with quantified uncertainty and different approaches for encoding and visualising uncertainty information will be presented. Retrieving feedback of intermediate processing results like input layer histograms or variability using supplied uncertainty information will be discussed.

***Keywords**:* Habitat modelling; OGC web processing service; Uncertainty propagation

1 INTRODUCTION

According to the Aichi Biodiversity targets of the Convention on Biological Diversity (CBD), 17% of the terrestrial and 10% of the coastal and marine area of the earth's surface of special value to biodiversity and ecosystem services shall be conserved at the latest in 2020, e.g. through a well connected network of protected areas. As a comparison, 12.7% and 1.17% of the terrestrial and marine environment, respectively, are protected. To achieve these targets the values of existing protected areas have to be examined and new areas potentially suited for conservation have to be assessed and located. One building block of this assessment can be a modelling and prioritization of habitats and ecosystems by their uniqueness regarding a combination of relevant environmental and socioeconomic parameters. We show in this case-study a general approach to this, focusing on the possibilities to use and access existing datasets and to provide the modelling capacity through OGC web services. The generalization and simplification of a model usually goes together with an increase of data uncertainty (Heuvelink [1998]). We want to address this issue by taking into account various sources of uncertainties in the model input parameters and by propagating and visualising model-driven uncertainty using the UncertML model designed for encapsulating and communicating probabilistic uncertainties (Williams et al., [2009]).

2 MAPPING UNIQUE ECOSYSTEMS

2.1 Mahalanobis distances for multivariate modelling

Defining habitat or ecosystem values is a complex task. One approach adopted by Hartley et al. [2006] in their assessment of African protected areas was to use Mahalanobis distances to calculate similarities between ecosystems found in protected areas within the same ecoregion (see e.g. Clark et al.[1993]). For this purpose, Hartley et al. selected a set of 9 key ecological variables like e.g. the percentage of tree cover, the annual average photosynthetic activity or the elevation, to be used as input data when computing the Mahalanobis distances. In this multivariate modelling process, the relative composition of the values taken by each variable within a given protected area is compared with the values found outside this area. Regions with high similarities would indicate the presence of similar ecosystems while cases where only small regions have high similarities to a protected area would point to habitats with unique characteristics and potentially unique species. Figure 1 shows an example of such a probability map showing the small area where ecosystems are similar to those found in the reference area (here a game reserve in Botswana).

2.2 eHabitat Web Processing Service

The computation of Mahalanobis distances is a process that is relatively straightforward and which can be easily implemented as a Web Processing Service (WPS). The benefits in setting up such a service are potentially numerous. Besides the benefits of web modelling services in general (see e.g. Morozov et al., [2006]), end-users might want in this particular case to use different thematic layers for other purposes like assessing climate shifts (Skøien et al. [2011]) or more simply for working on marine habitats rather than on terrestrial ecosystems as in use cases discussed here. The number of available and accessible dataset is indeed increasing dramatically as shown at the latest plenary meeting of the Group on Earth Observations (GEO) in November 2011 in Istanbul. The number of data services in their catalogue increased from less than 200 in 2010 to over 1500 in November 2011 as indicated by the GEO [2011]. Designing a WPS capable to consume datasets through standard OGC web services was therefore a minimum but fundamental requirement. Figure 2 shows a simple scheme of eHabitat, a WPS designed mainly for computing the likelihood of finding ecosystems with equal properties (Dubois et al. [2012]).

2.3 eHabitat in the Model Web

Ideally, the WPS should also be easily integrated in more complex chains of models as envisaged in the Model Web (Geller and Turner [2007]), an environment

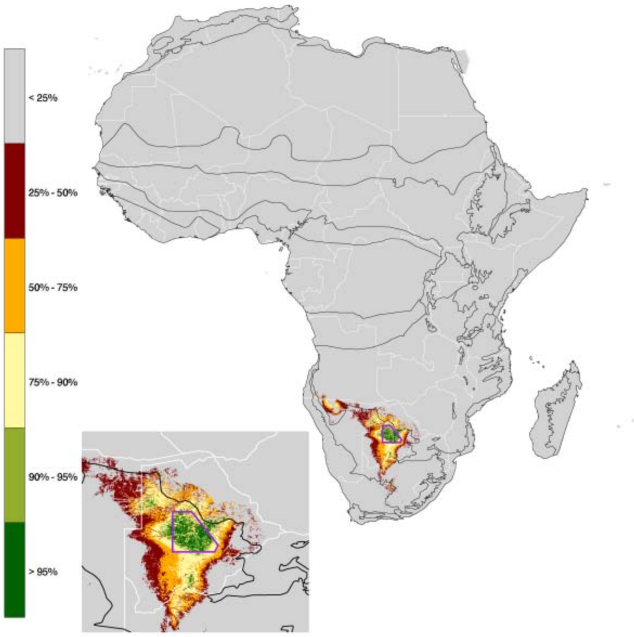

Figure 1. Map of probabilities (scale on the left) of finding ecosystems in the Kalahari-Highveld ecoregion that are similar to those found in the Central Kalahari Game Reserve (purple polygon) in Botswana.

of interacting models allowing for more complex operations like, in our case, assessing connectivity between ecosystems or for computing ecological niches for example (see e.g. Clarck et al. [1993]). Additionally, restricting the use of eHabitat to defined entities like the borders of protected areas is an unnecessary limitation considering the potential uses in other scenarios. We therefore adopted OGC's WPS interface to guarantee standardized access and discovery. The WPS is built using different Open Source libraries, the R package[1] for the mathematical functionalities and the statistical backend, PyWPS as the WPS implementation (Cepicky and Becchi, [2007]). The actual processes are written in Python and use the Geospatial Data Abstraction Layer (GDAL[2]) API to load and process the raster data. Results are returned referenced as OGC web map or coverage services (WMS, WCS), as raw data using GeoTIFF or NetCDF (Network common data format) and as rendered images in PNG format.

3 UNCERTAINTIES PROPAGATION AND MODELLING

The variability of the data potentially used in eHabitat requires a proper handling of the associated uncertainties as well as of their propagation in the modelling process. Sources of uncertainties in eHabitat are manifold:

- Observation uncertainty
- Processing uncertainty of remote sensing variables
- Uncertainty due to resampling and aggregation of rasters
- Interpolation uncertainty when rasters are derived from point observations
- Positional uncertainty of reference geometry, such as the PA
- Model and conceptual uncertainty

[1] http://www.r-project.org/
[2] http://www.gdal.org/

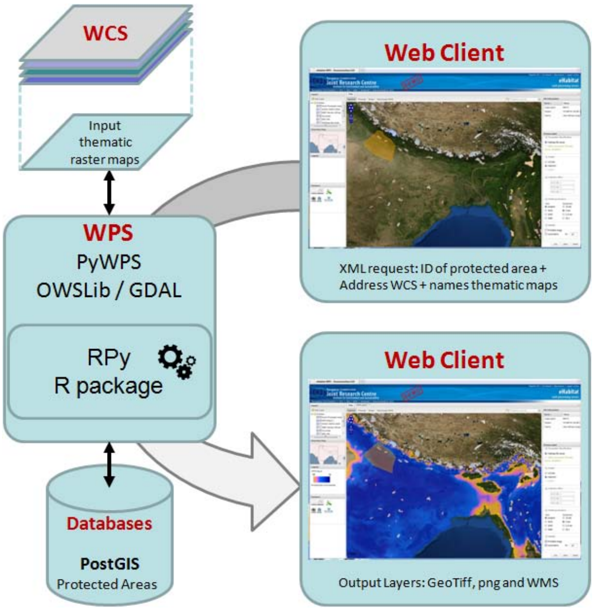

Figure 2. Schematic representation of the architecture of the eHabitat WPS.

We are not directly addressing model and conceptual uncertainty in this paper, but other models for similar purposes exist (see e.g. Segurado and Araujo [2004] for a review) and can easily be interchanged, at least conceptually and technically, with our function computing the Mahalanobis distances.

From the above, a clear framework for assessing and documenting the uncertainties in the input data as well as their propagation in the results is required, a topic that is at the heart of the UncertWeb research project[3]. We will discuss hereafter how these uncertainties can documented, tracked and visualised using the eHabitat use case.

3.1 Encoding uncertainty using UncertML

Skøien et al. [2011b] have shown how to assess possible uncertainties in eHabitat using geostatistics and Monte Carlo simulations when these uncertainties have not been documented in the input data, a case that is the most frequently encountered. The use of statistical simulations is computationally intensive and greatly benefits from an interoperable framework for exchanging uncertainties based on UncertML, a conceptual model and XML encoding for encapsulating probabilistic uncertainties (Williams et al. [2009]). This allows uncertainty to be documented and propagated through processing chains. Transposed to the use case of eHabitat, the use of UncertML and the quantification of the propagating uncertainties can be summarised as in Figure 3.

3.2 Encoding uncertainty using NetCDF

When mainly dealing with gridded datasets, propagating uncertainty using XML encoding in the WPS response has some drawbacks. The connection between the model result and propagated uncertainties is only referenced in the WPS response document. Propagating multiple model results with correlated uncertainties is complicated due to the simple structure of the WPS response.

[3] http://www.uncertweb.org/

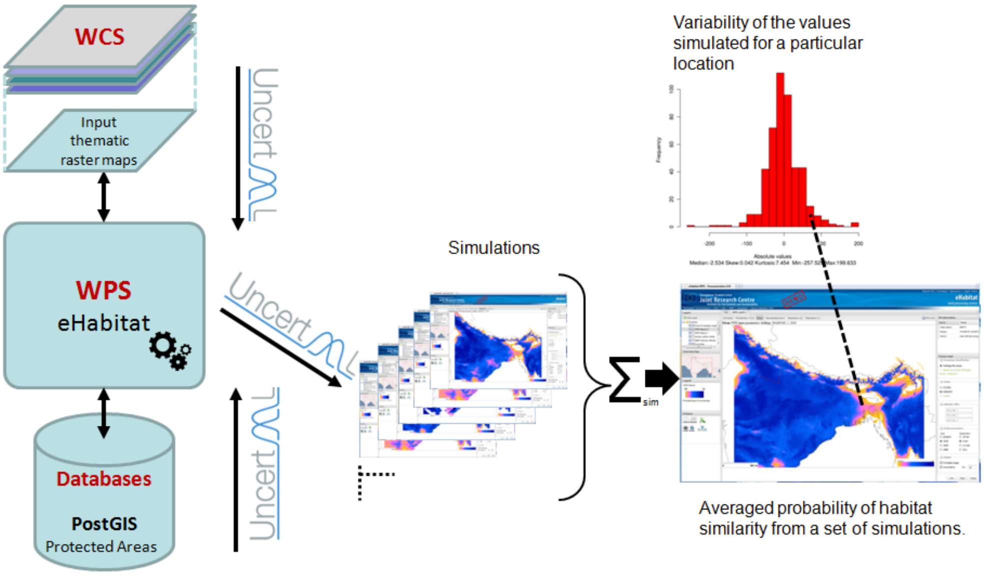

Figure 3. Uncertainty propagation and modelling in eHabitat. Uncertainties are documented through the whole modelling process using the UncertML scheme. Simulations are further used to describe local statistics distributions of simulated data, highlighting high uncertainties where standard deviation is highest.

The UncertWEB project proposes NetCDF-U as uncertainty conventions for NetCDF (Bigagli and Nativi [2011]). The convention is compatible with the climate& forecast convention (CF) and allows for referencing variables that represent uncertainty information inside the NetCDF dataset as ancillary variables to primary variables, the model results. These ancillary variables encode the type of uncertainty using URIs (uniform resource identifiers) to the UncertML dictionary.

4 COMMUNICATING UNCERTAINTIES

To allow effective interpretation of model results, uncertainties intrinsic to the model as well as uncertainties already contained in the input parameters and propagated through the model process must be included in the model output. This is especially important if the model can be used in chained workflows of models as envisioned in the Model Web approach. If model results are directly accessed by end-users for interpretation and possibly decision-making, means for an effective communication of the uncertainty information will be required. Uncertainty analysis and interpretation being an issue in itself, the UncertWeb project decided to develop an independent web client which could be reused with other web-based modelling services. Beside the benefits of reusing a web client for multiple services, this strategy has the advantage that existing web clients focusing on a given thematic (i.e. the identification of unique ecosystems in eHabitat) would not need to be crowded with the elements required to analyse the data uncertainties.

4.1 A Web-based client for visualising uncertainties

Following a review of possible approaches for visualising the various types of uncertainties, Senaratne et al. [2012] concluded that interactive visualisation tools appeared to be well accepted but that simple methods are required to allow also novice users to comprehend the uncertainties in the results. A web-based visualisation client was then developed by Gerharz et al. [2012] to communicate model results with the propagated uncertainties and allowing seamless integration into the Model Web infrastructure. The visualisation client is based on the OpenLayers JavaScript library[4] which provides standard mapping methods and control tools. Supported data types were extended to cope with uncertain data encoded in UncertML, and vector or raster formats. As NetCDF-U and GeoTIFF, like used by the eHabitat WPS, are binary formats which cannot be directly parsed by the OpenLayers client, an additional service component was added to extract the uncertain raster information from the provided data. This also allows additional visualisation methods, e.g. creating new raster layers by computing exceedance probabilities or expected values from a set of realisations.

Besides visualisation of model outputs, the communication of intermediate results from the model is important as well. For the case of eHabitat, the spatial pattern of the similarity is directly linked to the data distribution within the protected area. Examining intermediate results can help in understanding the final result, particularly in cases where it differs from the expectations. Input data, intermediate results and final results from eHabitat can be visualised by providing a reference to the data directly in the URL of the visualisation client. Figure 4 shows the visualisation of the mean and standard deviation estimated from 10 simulations eHabitat realisations, as well as probabilities to exceed a given threshold.

4.2 Exchange of data uncertainties with the Uncertainty Web Client

In a technical point of view, the data shown above can be either provided as NetCDF-U or as UncertML in a Java Script Object Notation (JSON) encoding with references to the GeoTIFF layers which can directly be parsed by the JavaScript libraries. Using the NetCDF-U convention to encode uncertain raster data has the advantage that the data is provided in a standardised format which can be processed by different existing clients and services. As the uncertainty information is provided as ancillary information, the NetCDF-U files could also be used by software unaware of uncertainties and UncertML. The GeoTIFF in contrast is encoded as references within an UncertML document which cannot be processed by existing clients without UncertML support.

5 DISCUSSION AND CONCLUSIONS

This paper has presented the combined use of a WPS for ecological modelling (eHabitat) accessing datasets using OGC WCS, propagating results as OGC web services, enhanced with uncertainty information encoded in XML or NetCDF-U using UncertML together with an independent web client designed to help end-users to analyse the data uncertainties in the model results.

Visualization in a separate client is beneficial for several reasons:

- End users get access to the best of two worlds – up-to-date modelling tools and flexible visualisation tools
- The developers of the thematic WPS do not have to spend time on developing the visualisation tool.
- New visualisation methods are immediately accessible for the thematic WPS
- Both intermediate results and final results can be visualised with the same layout for chained Web Services, instead of each WPS showing their results with different layout, colour scheme etc.

[4] http://openlayers.org

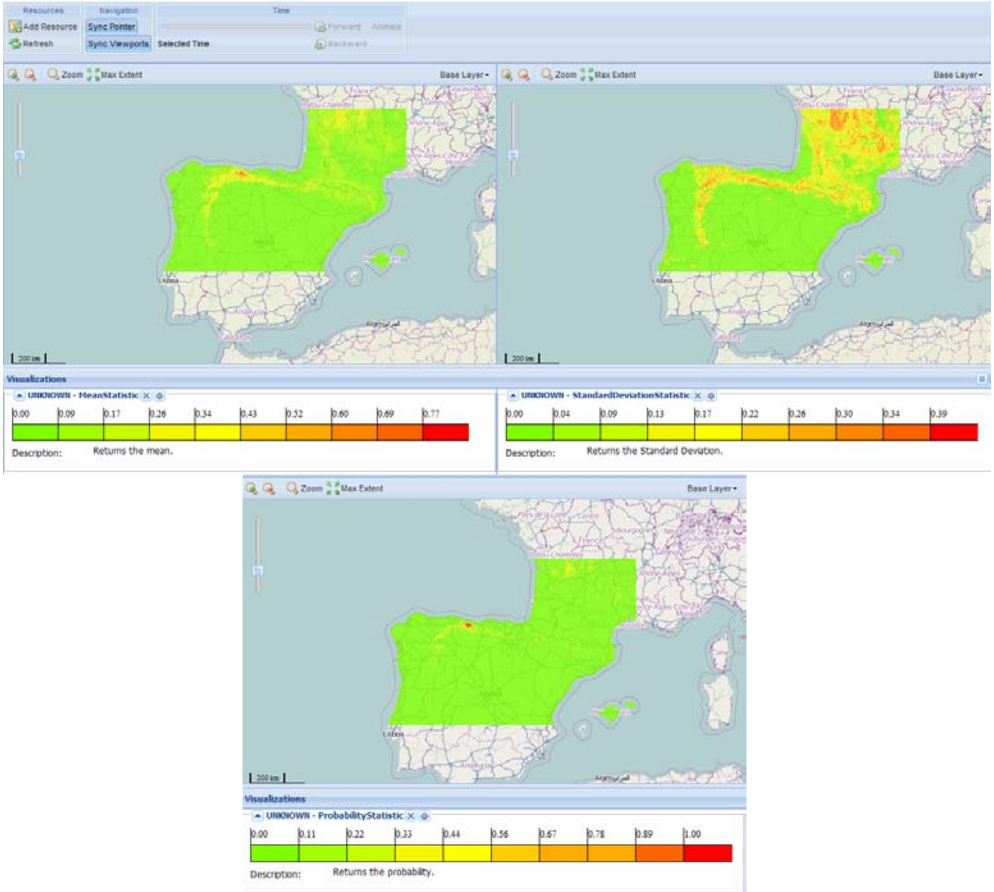

Figure 4. Top: Adjacent maps visualisation of mean and standard deviation from 10 realisations of the eHabitat model in the UncertWeb visualisation client. Bottom: Probabilities to exceed a given threshold, here, probabilities that a pixel exceed a similarity of 50%.

The adoption of NetCDF-U to tightly couple the actual model results with correlated uncertainties eases propagation, enables reuse even in not UncertML aware clients. It works for spatially similar information, e.g. the mean of simulations on the complete study area. If it is possible to encode information on a different spatial premise, e.g. variability of input data only inside the PA has to be investigated.

Communication of uncertainties is important not only for the final model outputs but also for input data and intermediate results. By using a web-based thin client for visualisation, this can be easily realised for each processing step by providing the data reference through the client URL. This concept can also be applied when using eHabitat in a chain of web services.

ACKNOWLEDGEMENTS

This work is partly supported by the European Commission, under the 7th Framework Programme, by the UncertWeb project funded by the DG INFSO. The views expressed herein are those of the authors and are not necessarily those of the European Commission.

More information about eHabitat and UncertWeb can be found on the Internet, see http://ehabitat.jrc.ec.europa.eu/ and http://www.uncertweb.org/, respectively.

REFERENCES

Bigagli, L., and S. Nativi, Managing uncertainty in data – the NetCDF uncertainty conventions, 79th OGC Technical Committee, Brussels Belgium, URL: http://www.uncertweb.org/documents/presentations/managing-uncertainty-in-data-the-netcdf-uncertainty-conventions/download (accessed 2 March 2012), 30 November 2011.

Clark, J.D., J.E. Dunn, and K.G. Smith, A multivariate model of female black bear habitat use for a geographical information system, *Journal of Wildlife Management*, 57:519-526, 1993.

Cepicky, J., and L. Becchi, Geospatial processing via Internet on remote servers – PyWPS, *OSGeo Journal* 1, 5pp., May 2007.

Dubois, G., J. Skøien, M. Schulz, S. Nativi, M, Santoro, L. Bastin and S. Peedell, eHabitat, a multi-purpose Web Processing Service for ecological modelling, *Environmental Modelling & Software* (Submitted) 2012.

Geller, G. N., and W. Turner, The model web: a concept for ecological forecasting, *Geoscience and Remote Sensing Symposium*, IGARSS 2007. IEEE International, 2469-2472, 23-28 July, 2007.

Gerharz, L., H. Senaratne, C. Autermann, P. Truong, G. Heuvelink, M. Williams et al., Tools for communicating and visualising uncertainties. Technical Report, Deliverable D3.3 of the UncertWeb project,University of Münster, Germany, http://www.uncertweb.org/uploads/deliverables/a2ddab1d6f728688f90a2c8b37601daaa248f54e.pdf (accessed 26 February 2012). 2012.

Group on Earth Observations. GEO VIII, Report of the Architecture and Data Committee (ADC). Document 19, 2011.

Hartley, A., A. Nelson, P. Mayaux, and J.-M. Grégoire, *The assessment of African protected areas.* JRC Scientific and Technical Reports, EUR 22780 EN, Office for Official Publications of the European Communities, 70 pp., Luxembourg, 2007.

Heuvelink, G.B.M., *Error Propagation in Environmental Modelling with GIS* Research Monographs in GIS. CRC Press, 144 pp., London, 1998.

Holdridge, L.R., Determination of world plant formations from simple climatic data, *Science*, 105, 367-368,1947.

Morozov, I., B. Reilkoff and G. Chubak, A generalized web service model for geophysical data processing and modelling. *Computers & Geosciences,* 32: 1403-1410, 2006.

Schade, S., N. Ostländer, C. Granell, M. Schulz, D. McInerney, G. Dubois et al., Which Service Interfaces fit the Model Web? Paper presented at GEO-Processing 2012. 4th Int. Conf. on Advanced Geographic Information Systems, Applications, and Services", Valencia, Spain, January 30 - February 4, 2012.

Segurado, P., and M.B. Araujo, An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31:1555-1568, 2004.

Senaratne, H.V., L.E. Gerharz, E. Pebesma, and A. Schwering, Usability of Spatio Temporal Uncertainty Visualisation Methods. The 15th AGILE International conference on Geographic Information Science. April 24-27, 2012, Avignon, France. 2012.

Skøien, J., G. Dubois and J. De Jesus, Forecasting biomes of protected areas. *Procedia Environmental Sciences*, 7: 44–49, 2011a.

Skøien, J., M. Schulz, G. Dubois, R. Jones, G.B.M. Heuvelink, and D. Cornford, Uncertainty propagation in chained web based modeling services: the case of eHabitat. In: "*Innovation in sharing environmental observations and information. Proceedings of EnviroInfo 2011, 25th International Conference Environmental Informatics*". W. Pillmann, S. Schade and P. Smits (Eds), pp: 46-58, Ispra, Italy 5-7 October 2011b.

Williams, M., D. Cornford, L. Bastin, and E. Pebesma, Uncertainty markup language (UncertML). OpenGIS Discussion Paper 08-122r2, Open Geospatial Consortium Inc. URL: http://portal.opengeospatial.org/files/?artifact_id=33234, (accessed 23 January 2012). 2009.