Evaluation and Visualization of Hydrological Sensor Network: an Integrated Approach Using MODIS Images

Jagannath Aryal
School of Geography & Environmental Studies, UTAS, Hobart, Australia
jagannath.aryal@utas.edu.au

Ritaban Dutta
Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, Australia
Ritaban.Dutta@csiro.au

Ahsan Morshed
Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, Australia
Ahsan.Morshed@csiro.au

Abstract

Advancement in sensor networks and their availability enhanced the spatial and temporal resolution of the data. High spatial and temporal resolution data are very useful in mapping and monitoring the environmental phenomena and in decision making processes. However, the reliability and efficiency of sensor networks and associated uncertainty of the data play a crucial role in understanding the phenomena. This paper describes an automatic evaluation method of the local sensor network using globally available MODIS images. The method uses the real time temperature time series data acquired from the 17 different locations of the sensor web and from MODIS satellite images. Complemented sensor web data are represented as 3D gridded interpolated surface. An integrated spatiotemporal performance map is developed to visualize the real time performance of the overall sensor network. Finally network's spatiotemporal performance map and related meta data are published on the Linked Open Data cloud environment as a global recommendation system for the end users of the hydrological sensor network.

Keywords: Sensor Web, Time Series Catalogue, Interpolated 3D Surface, MODIS, Spatiotemporal Performance Map, Linked Open Data.

1 Introduction

With the development of land based, airborne and satellite sensors and the public availability of data produced from such sensors it is now possible to acquire the data and use for intended purposes. These data are geographic and essentially observations about geographic features or phenomena referred to as "geographic reality". However, representing reality by means of data indeed is an approximation of the real world phenomena. Therefore, a fundamental discrepancy – uncertainty – exists between geographic data and the reality that they are intending to represent. This shows that reliability and efficiency of a sensor web is negatively correlated with its uncertainty [1]. Fundamentally, it is a challenge to increase reliability and efficiency that leads to minimize the associated uncertainty. From an application perspective, it is important to provide associated uncertainty information to users to ensure that the data is fit for its intended purpose [11]. For instance, a system for agricultural water resource management, weather forecasting and crop management (including irrigation scheduling) may produce erroneous results, if supporting data sources are not fit for purpose. It is evident that there is a need to capture and integrate knowledge from different sources in a sensor web, which includes sensor networks, simulation models and historical data. A tool to analyse and visualise the uncertainty of data sources is required to determine whether the sources are complementary for the purpose of integration.

Uncertainty is commonly associated with the limited availability of data (spatially and temporally) and/or the poor quality of the available data. The practical causes of sensor uncertainty include their operation under extreme conditions, calibration drift and bio-fouling. Furthermore, sensor nodes are subject to communication and electronic failures, in addition to energy depletion. Ultimate challenge in environmental forecasting and decision support systems is to overcome uncertainty associated with the data quality, to cross validate the knowledge automatically and to make the decision making process efficiently. Uncertainty factors in the environmental monitoring processes are more evident than before due to the most recent advanced sensor networks and communication technologies [7, 12, 15].

In this paper, we propose a knowledge integration platform and machine learning analysis based recommendation framework to represent the knowledge in a more meaningful way. We use this framework to integrate environmental data available from Commonwealth Scientific and Industrial Research Organisation (CSIRO) based in Australia and NASA's MODIS (MODerate resolution Imaging Spectroradiometer) satellite images. The environmental data are acquired from the CSIRO's South Esk hydrological sensor web available via a standard web service interface (Sensor Observation Service) developed by the Open Geospatial Consortium (OGC) [4]. We make a complementary integration of the data and provide uncertainty visualization with the spatiotemporal performance map in linked open data cloud environment.

This paper is organized into 7 parts. In the 2nd part we present background and the problem, in the 3rd part we introduce data and study area, in the 4th part we describe the applied methods, in the 5th part we present the results, in the 6th part we provide a discussion and we conclude our work in 7th part.

2 Background

Wireless sensor network survey is described in [13]. Sensor data uncertainty and its management using different computing techniques are proposed in [14, 2]. Sensor network system with distributed data processing approach [6] and adaptive service provisioning in wireless sensor network [5] are discussed. Uncertainty visualizations have been proposed in different ways [9]. However, visualizing uncertainty by complimenting a local sensor web with global data has not been studied and is a novel approach that we pursue in this research. Further, publishing the spatio-temporal performance map in the web for the use of wider community is an asset and it adds value to the GIS community.

In this research, we ask number of questions such as: How can we evaluate the hydrological sensor network? What are the evaluation criterions? Can the network in a local scale be complimented by a globally available datasets like MODIS images? How can we visualize the performance map of local and global system? Can we present and publish such performance map in linked open data cloud environment for the use of wider community?

In answering above questions, we take this as an opportunity for research with great potential in a wider applications including environmental decision making. We design and implement dynamic data processing methodology to process multiple environmental data sources simultaneously for data quality assurance. For given geographical locations described by latitude longitude combination and time frame (spatio-temporal aspect), data from different sources are downloaded, pre-processed, complemented, integrated and cross validated to provide high quality data assurance with known uncertainty limits.

Complemented sensor web data are represented as 3D gridded interpolated surface. A novel integrated spatiotemporal performance map is created to visualize the real time performance of the overall sensor network using multiple features calculated from the complemented data.

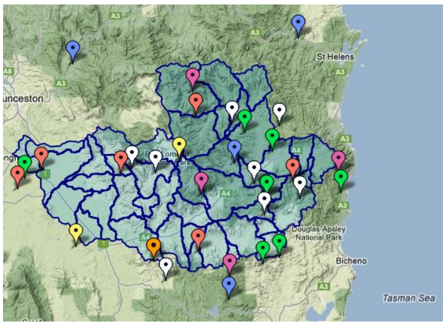
Next, MODIS based method is developed for the data quality assurance of the sensor network. Finally, the network performance indicators are integrated with pre-processed data and Meta data in a unified resource structure based knowledge representation using Resource Description Framework (RDF). Recommended knowledge is published on Linked Open Data cloud environment using RDFs for global accessibility.

3 Data and Study area

3.1 Hydrological Sensor Web

The Sensor Web is an advanced spatial data infrastructure in which different sensor assets can be combined to create a macro-instrument of sensing capability. This macro-instrument can be instantiated in many ways to achieve multi-modal observations across different spatial and temporal scales. CSIRO is investigating how emerging standards and specifications for Sensor Web Enablement can be applied to the hydrological domain. To this end, CSIRO has implemented a Hydrological Sensor Web in the South Esk river catchment in NE Tasmania (Figure 1). The South Esk sensor web covers an approximately 95 km × 220 km rectangular region. It covers a latitude range between 40.5°S and 43.5°S and a longitude range between 145°E and 148.5°E.

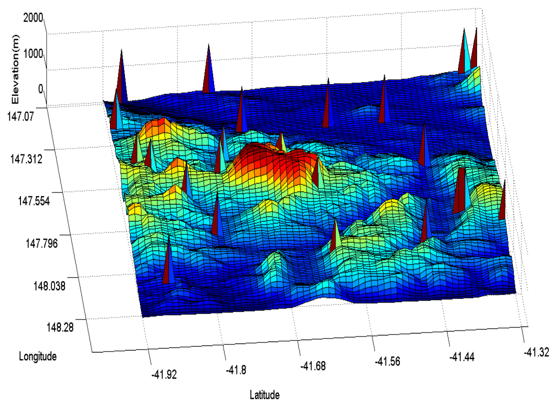
Figure 1: The Google Maps™ pane presents a federated view of near real-time sensor data from the different sensor networks operating in the South Esk catchment, Tasmania, Australia.



The whole region was mapped as a gridded rectangle where each cell represents a 5km × 5 km region. In Figure 2, the whole South Esk is depicted as a 3D surface based on patched elevation data. Brown spikes represent physical locations of the weather station sensor nodes on the 3D surface.

The South Esk river catchment was chosen because of its size (3350km2, large enough to show up differences in catchment response to rainfall events), spatial variability in climate (there is an 800mm range in average annual rainfall across the catchment), fickle nature of seasonal flow, and relatively high-level of instrumentation. This is made possible by re-publishing near real-time sensor data provided by the Bureau of Meteorology (BoM), Hydro Tasmania, Tasmania Department of Primary Industries, Parks, Wildlife and Environment (DPIPWE), Forestry Tasmania and CSIRO via a standard web service interface developed by OGC and 52 North [4].

Figure 2: Elevation based 3D map of the South Esk and associated coordinate system. Brown spikes represent distributed sensor nodes.



3.2 MODIS Data

MODIS is a multi-disciplinary, keystone instrument on Aqua and Terra spacecraft, providing a wide array of multispectral, daily observations of land, ocean, and atmosphere features at spatial resolutions between 250m and 1000m. The images are freely available in NASA website for different products. For this study, we have downloaded the MODIS/Terra Land Surface Temperature and Emissivity (LST/E) 8-day L3 Global 0.05Deg CMG product (MOD11C2) version 5 for 17 point locations in Tasmania (Table 1) during the period 2009/02/18 – 2011/12/29.

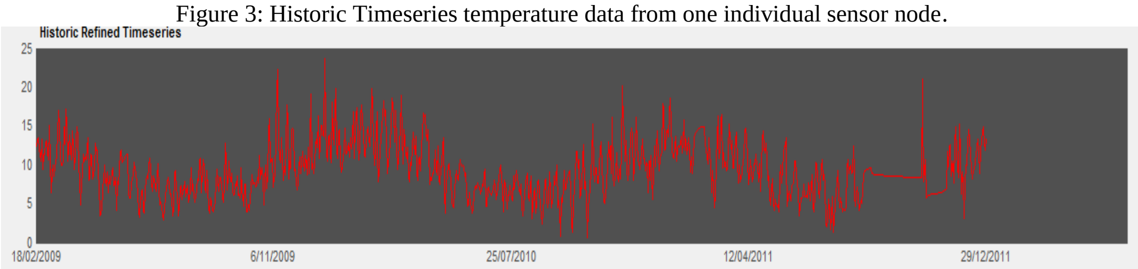
This product provides per-pixel temperature and emissivity values in a sequence of swath-based global products. MOD11C2 product comprises the Science Data Set (SDS) layers for day time and night time observations: LSTs, quality control assessments, observation times, view zenith angles, clear sky coverage, and emissivity for bands i.e. 20, 22, 23, 29, 31, and 32. This was explicitly mentioned in the Land Processes Distributed Active Archive Center (LP DAAC) portal of NASA that MOD11C2 products are ready for use in science applications. Time series are created from these downloaded image data files based on image processing techniques.

4 Methods

4.1 Timeseries Catalogs and Experimental Data

The Timeseries Catalog was a search facility available through the developed web service interface. This service provides near real-time status updates for the individual sensor nodes of the network. For this visualization study, one environmental variable (temperature) was acquired during the period 2009/02/18 – 2011/12/29 from the 17 different sensor nodes available in the South Esk hydrological sensor web.

Figure 3: Historic Timeseries temperature data from one individual sensor node.



4.2 Node Data Conditioning

Data was available from the time of deployment. As can be expected in real world networks, each of the available time series had periods with missing values. Initially a filter was designed to remove all of the ±Infinite values, and replace them with a 'Not a Number' string to keep the filtering statistically insignificant and the original time frame unaltered. A sensor measuring a particular environmental parameter should operate within a well-defined range. Hence, any value recorded outside of the operational range was treated as invalid data and replaced with a 'Not a Number' string.

Table 1: 17 different sensor nodes were used for this study.

Location Name	Latitude	Longitude
Doodys Hill	-43.14	146.94
MacGregors Peak	-42.97	147.93
Mount Dazzler	-41.23	146.71
Storys Creek	-41.63	147.74
Snow Hill Farm	-41.86	147.84
Mt Arthur	-41.16	147.17
Mt Horror	-41.63	147.74
Platts Lookout	-41.21	148.08
Snow Hill	-41.85	147.84
Tower Hill	-41.35	147.54
Tylers Hill	-43.37	146.99
Abbotts Lookout	-42.78	146.65
Bradys Sugarloaf	-42.26	146.5
Ben Ridge Road	-41.35	147.7
Watts Lookout	-41.36	146.08
Ben Lomond	-41.54	147.66
Tom Gibson Nature Reserve	-41.77	147.3

4.3 MODIS Data Processing

MOD11C2 products were downloaded from NASA website with the launch of Reverb / ECHO [8]. The hierarchical data format (HDF) was downloaded and processed using add on MODIS conversion toolkit in ENVI Image processing software. With the processing, the land surface temperature was extracted. In extracting the temperature following procedure was followed:

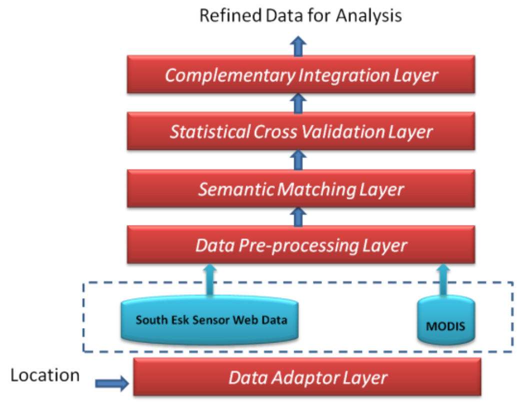
- *Download MODIS/Terra from Reverb / ECHO by defining geographical location and get hdf file*
- *Use MODOS Reproject Tool (MRT) to convert hdf file to TIF*
- *Stack the land surface temperature and emissivity files into one.*
- *Use IDL code to convert the integer value to geophysical values.*
- *Get the temperature value.*

5 Results

5.1 Complementary Data Integration

A metric of data availability was computed as the ratio of the total number of days since a particular sensor was deployed and total number of days since a valid data point was produced. Data availability varied between 0% and 100%. It was evident that realistic performance of a network is not always reliable and poor level of data availability increased the uncertainty.

Figure 4: Architecture for complementary data integration.



Complementary data integration was a novel way to tackle this problem. Figure 4 shows the data processing architecture developed and used for data processing and integration.

Equivalent (for the same time and location) temperature Time series extracted from the MODIS images were used to impute the missing values of the sensor web's Time series. Sensor nodes with less than 95% data availability were processed using this architecture. Figure 6 shows an example of the comparative time series visualization.

5.2 Temperature Surface Map

The concept behind this 3D surface visualization was to provide an environmental gridded surface based on complemented and cross validated South Esk temperature data using MODIS data.

Figure 5 Representation of the temperature surface map.

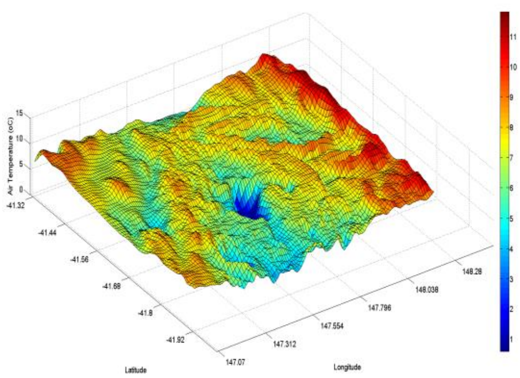


Figure 5 shows an example of the daily interpolated 3D spatiotemporal temperature surface for the entire South Esk sensor web.

5.3 Spatiotemporal Performance Map

On the basis of MODIS data a dynamic time series annotation system was developed to provide recommendations about the South Esk sensor web data. MODIS time series extracted from the satellite images was considered as ground truth information to evaluate and annotate the performance of the network and data quality. The node based comparative study between sensor web data and MODIS data were conducted using some statistical features, namely the monthly maximum value and the corresponding date, the monthly standard deviation, number of days 'NaN' was produced in a month, the monthly minimum value and its date, the longest length of the missing value segments during the whole time period with corresponding dates and maximum number of consecutive days with a constant data. Features from the individual time series were computed in the context of the equivalent MODIS time series. Multiple feature based performance scoring techniques were used to score all of the 17 sensor nodes between 0-100 percent. The purpose of the system was to process time series dynamically, annotate and then provide a general data usability recommendation for users of the network. Individual time series (representing individual sensor node) was labeled as one of the categories, namely, "Very Good Node (>=90%)", "Good Node (<90% and >=80%)", "Average Node (<80% and >=65%)", "Poor Node (<65% and >=50%)" and "Damaged Node (<50%)" depending on the values of the performance scores. Processed time series were stored in a data structure along with recommendations. A performance colourmap was created for the network based around the computed performance scores. Higher score represented better performing nodes so in other words less data uncertainty associated with that node. This colour map based visual representation of the network was called performance map. Figure 7 depicts the average temperature performance map of the South Esk sensor web for the period 2009/02/18 – 2011/12/29. This performance map was a dynamic annotation system which is capable of reflecting daily data quality variance [10].

5.4 Performance Map on LOD

Publishing the network's performance map on Linked Open Data (LOD) cloud was motivated by the fact that ultimately the data customers and environmental application designers require a web based advanced recommendation system about the performance of the network [3]. The South Esk sensor network's performance map was published on LOD using URIs and resource description frameworks (RDF) as it is the best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web. A user interface layer (see Figure 7) was developed on top of these triples to provide great flexibility to the application designers.

Figure 6: Comparison between recorded Timeseries with missing values (in the upper one) and MODIS based complemented Timeseries (in the lower one)

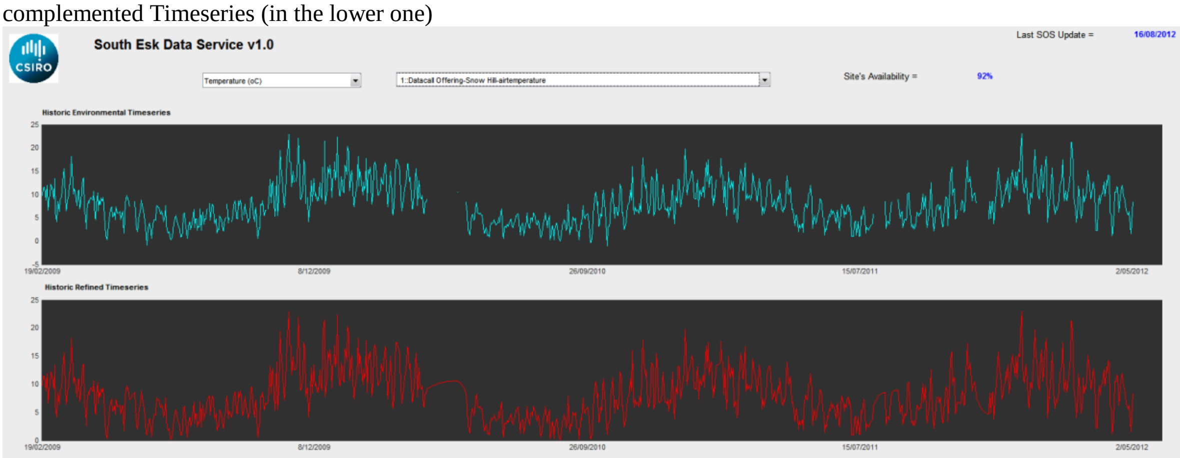
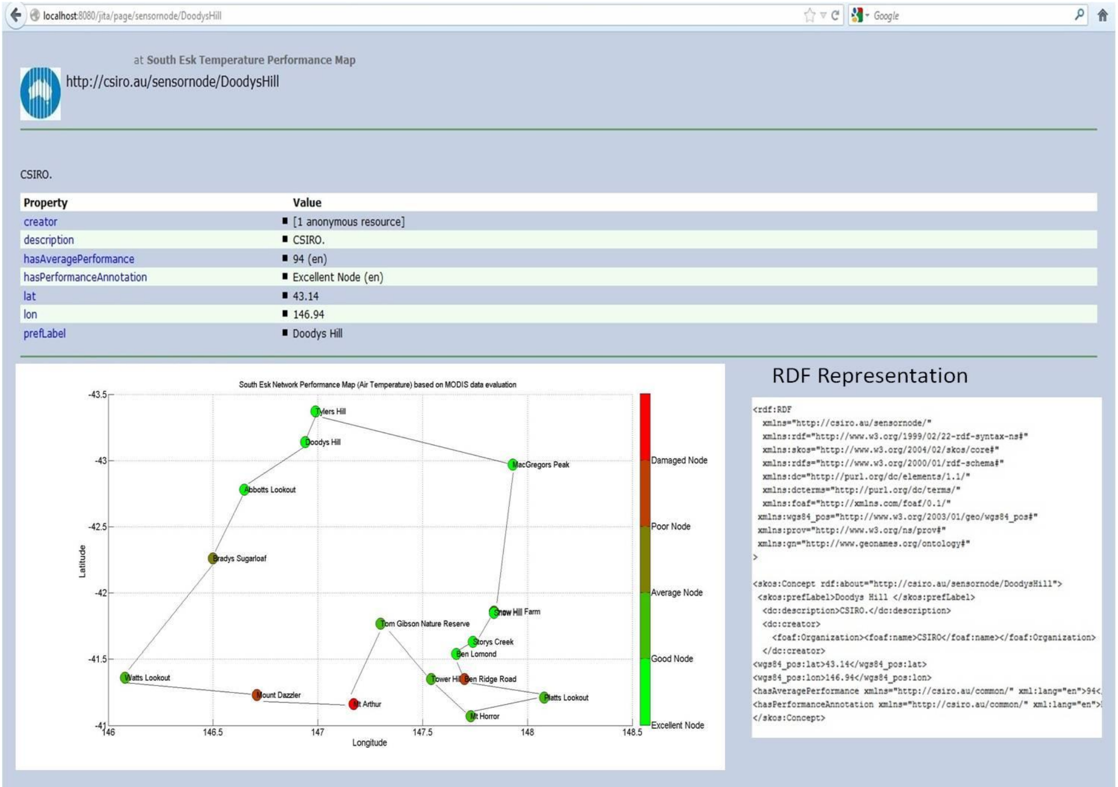


Figure 7: Representative Spatiotemporal Performance Map of South Esk Sensor Web and



6 Discussion

Development of the sensor network's performance map based on MODIS is the main achievement of this work. Complementary data integration to improve data quality of the network is another outcome of this work. This new performance map provided vital data quality information. South Esk hydrological sensor web was developed on a very difficult terrain in Tasmania (including Ben Lomond mountain peak) so it was expected that data acquisition and delivery would be extremely difficulty from some parts of the network. This study shows the importance of the statistical data annotation, data recommendation and sensor web visualization.

7 Conclusion and Future work

We conclude that this visualization model could be adapted for any sensor network in the environmental domain.

We are aware that minimizing uncertainty is a challenging task and we pursue our future research in the direction of uncertainty minimization and its visualization by incorporating various test and validation environmental data sets.

References

[1] Balazinska, Magdalena, Amol Deshpande, Michael J. Franklin, Philip B. Gibbons, Jim Gray, Suman Nath, Mark Hansen, Michael Liebhold, Alexander Szalay, and Vincent Tao. Data management in the worldwide sensor web. *Pervasive Computing, IEEE*, 6 (2): 30-40, 2007.

[2] Bastin, L., D. Cornford, R. Jones; G.B.M. Heuvelink; E. Pebesma; C. Stasch; S. Nativi; P. Mazzetti; and M. Williams.. Managing uncertainty in integrated environmental modelling: The UncertWeb framework. *Environmental Modelling & Software* 39(0): 116-134, 2013, http://dx.doi.org/10.1016/j.envsoft.2012.02.008

[3] Berners-Lee, T. Linked Data. http://www.w3.org/DesignIssues/LinkedData.html. [Retrieved April 10, 2011,]

[4] CSIRO, Sensor Web. http://www.csiro.au/sensorweb/au.csiro.OgcThinClient/OgcThinClient.html [Retrieved 10 Jan 2013)

[5] Fok, C.-L., Roman, G.-C., and Lu, C. Adaptive service provisioning for enhanced energy efficiency and flexibility in wireless sensor networks. *Science of Computer Programming* 78(2): 195-217, 2013.

[6] Fukatsu, T., T. Kiura, and Hirafuji, M. A web-based sensor network system with distributed data processing approach via web application. *Computer Standards & Interfaces* 33(6): 565-573, 2011.

[7] Hart, J. K. and K. Martinez. Environmental Sensor Networks: A revolution in the earth system science? *Earth-Science Reviews* 78(3–4): 177-191, 2006.

[8] NASA, Reverb /ECHO http://reverb.echo.nasa.gov/reverb/#utf8=%E2%9C%93&spatial_map=satellite&spatial_type=location&spatial=-42.000%2C%20147.000&place_name=Tasmania%2C%20Tasmania%2C%20Australia&keywords=MODIS%2011%20C2&temporal_start=2009-02-18%2000%3A00%3A00&temporal_end=2011-12-29%2023%3A59%3A59&processing_levels=Level%202 [Retrieved Dec 2, 2012).

[9] Pang, Alex T., Craig M. Wittenbrink, and Suresh K. Lodha. Approaches to uncertainty visualization. *The Visual Computer* 13(8): 370-390, 1997.

[10] Sheth, A, Henson C., Sahoo S., Semantic sensor web. *Internet Computing, IEEE* 12(4): 78-83, 2008.

[11] Szalay, Alexander, and Jim Gray. 2020 Computing: Science in an exponential world. *Nature*, 440 (7083):413-414, 2006.

[12] Upal Mahfuz, M. and K. M. Ahmed. A review of micro-nano-scale wireless sensor networks for environmental protection: Prospects and challenges. *Science and Technology of Advanced Materials*, 6(3 - 4): 302-306, 2005.

[13] Yick, J., B. Mukherjee, D. Ghosal. Wireless sensor network survey. Computer Networks 52(12): 2292-2330, 2008.

[14] Yu, B., R. Sen, and Jeong, D.H. An integrated framework for managing sensor data uncertainty using cloud computing. *Information Systems* (0), 2012, http://dx.doi.org/10.1016/j.is.2011.12.003.

[15] Zerger, A., R. A. Viscarra Rossel, D.L. Swain, T. Wark, R.N. Handcock, V.A.J. Doerr, G.J. Bishop-Hurley, E.D. Doerr, P.G. Gibbons and C. Lobsey. Environmental sensor networks for vegetation, animal and soil sciences. *International Journal of Applied Earth Observation and Geoinformation*. 12(5): 303-316, 2010.