Dynamic Annotation and Visualisation of the South Esk Hydrological Sensor Web

Ritaban Dutta [*1], Daniel Smith [1], Greg Timms [1]

[1] *Intelligent Sensing and Systems Laboratory,*
Commonwealth Scientific and Industrial Research Organisation (CSIRO),
Hobart, Australia 7001

Ritaban.Dutta@csiro.au

Abstract—**This research study focused on automatic sensor data annotation and visualisation of dynamic weather data acquired from a large sensor network. The aim was to develop a data visualisation method for CSIRO's South Esk hydrological sensor web to evaluate the overall network performance and provide visual data quality assessment. The visual data quality technique developed in this study could be applied to quality assurance of any sensor network.**

Index Terms—Dynamic Visualisation, Sensor Web. Weather Station, Time Series Catalogue, Interpolated 3D Surface.

I. Introduction

The development of wireless sensor network technology has led to an explosion in data enhancing the temporal and spatial scales at which empirical science can be performed. Given the data-rich environment, digital services have been adopted in conjunction with the Internet. These services have altered the traditional role of data stewardship, given we are now moving into an age where data storage is distributed, data is publicly accessible and annotated to promote re-use for a variety of purposes [1].

Along with the growing quantity of accessible data, the complexity of data required by particular applications is increasing. Next generation environmental monitoring, natural resource management and related forecast-based decision support systems are becoming increasingly dependent on web-based data integration of large scale sensor networks. This integration requires different forms of pre-processing, including accumulation and harmonization. Consequently, maintaining data quality in a large-scale sensor web is a difficult challenge. Due to an improvement in the transparency of recent sensor network and communication technologies, the uncertainty associated with environmental sensor webs is becoming increasingly evident. This 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. In addition, sensor nodes are subject to communication and electronic failures, in addition to energy depletion.

The reliability and efficiency of a sensor web is negatively correlated with its uncertainty. Consequently, it is important to provide users with this uncertainty information to ensure the data is fit for its intended purpose. 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 [2].

II. Science Objective

The science objective was to design and develop an adaptive tool for analysis and visualisation of the South Esk hydrological sensor web around a general theme of geographical location information. Near real-time analysis of sensors was performed in relation to the quality and availability of its data. Data cleansing and imputation techniques were also applied according to the weather station's sensory ontology. Finally this study predicted and visualized a dynamic 3D surface map of the South Esk region based upon available environment data.

III. South Esk Sensor Web

A. The Sensor Network

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 (Fig. 1). The South Esk river catchment was chosen because of its size (3350km^2, 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 (Sensor Observation Service) developed by the Open Geospatial Consortium (OGC). The Sensor Observation Service implementation that CSIRO is using was developed by 52North. Exposing sensor data via standard web service interfaces provides a much richer picture of what is going on in the catchment (enhanced situation awareness) [3].

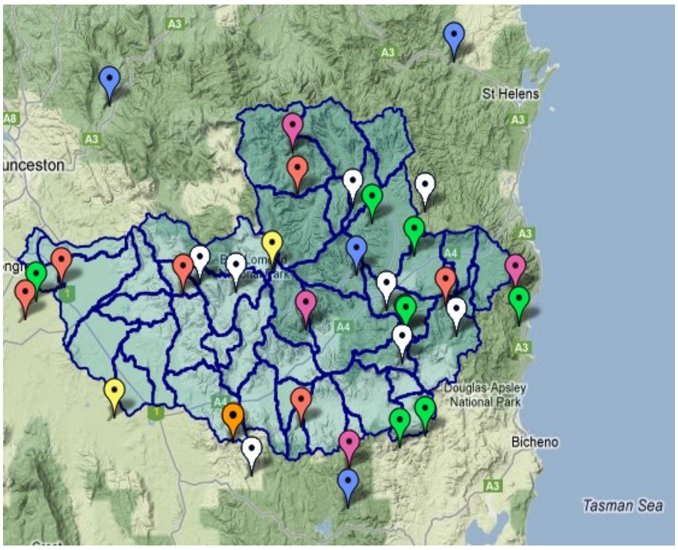

Fig. 1: The Google Maps™ pane presents a federated view of near real-time sensor data from the different sensor networks operating in the South Esk catchment (different colours correspond to sensors from different agencies).

B. Coordinate System

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.

The whole region was mapped as a gridded rectangle where each cell represents a 5km × 5 km region. In Fig 2, the whole South Esk is depicted as a 3D surface based on patched elevation data. Blue dots represent physical locations of the weather station sensor nodes on the 3D surface. This visualisation was completed as part of the South Esk Data Service Tool developed for this study.

C. Weather Station Sensor Systems

Vaisala automatic weather stations were used for most of the sensor network nodes; although a number of Siphon tipping bucket rain gauges were also used. The Vaisala weather transmitter (WXT520) measured barometric pressure, humidity, precipitation, temperature, and wind speed. Vaisala WINDCAP sensors (WMT52 and WM30) were used to measure horizontal wind speed using ultrasound technology.

Vaisala RAINCAP sensors were used to measure acoustic precipitation through the impact of individual rain drops. The RIMCO 7499 tipping bucket rain gauge was used to measure rainfall in a few of the BoM weather stations that form part of the South Esk sensor web.

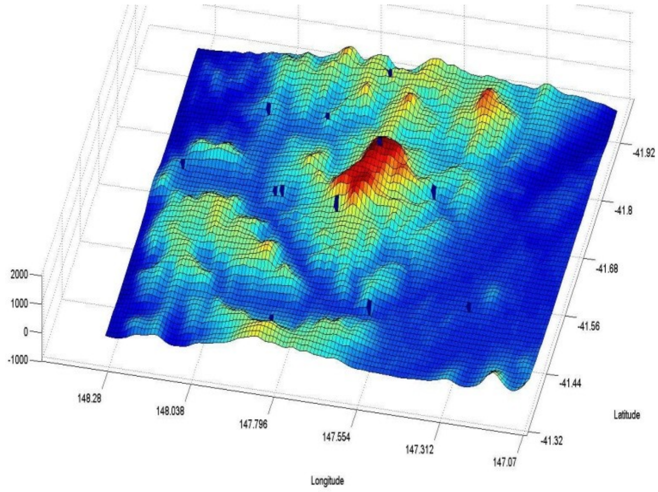

Fig. 2: Elevation based 3D map of the South Esk and associated coordinate system. Blue dots represent distributed sensor nodes.

High accuracy relative humidity and temperature were measured using Vaisala HUMICAP probes, namely the HMP45A/D. REMCO Middleton EP16 16/E Pyrano-Albedometer was a dual head version of the EP08 First Class Pyranometer and was used for the measurement of the ratio of upward to downward solar radiation flux density on a plane surface. Middleton SK08 was also used to measure solar radiation which was a solar radiometer for the accurate measurement of global irradiance on a plane surface. All of the sensor systems used in South Esk sensor web along with valid data ranges are listed in Table I. These valid measuring ranges were used later for dynamic data filtering purposes.

TABLE I
SENSORY SYSTEMS USED IN SOUTH ESK SENSOR WEB

Sensor	Measured Variable	Data Validity Range
WMT52	Wind Speed	0...60m/s
WM30	Wind Speed	0.5...60m/s
RAINCAP	Precipitation	0...200mm/h
RIMCO 7499	Precipitation	0.2mm bucket
HMP45A	Humidity	-40...+60 °C
HMP45A	Temperature	0.8 to 100% RH
EP 16 and EP 16-E	Upward/Downward Solar Radiation	0-1400 W/m^2
EQ08 and EQ08-E	Total Global Radiation	0-2000 W/m^2

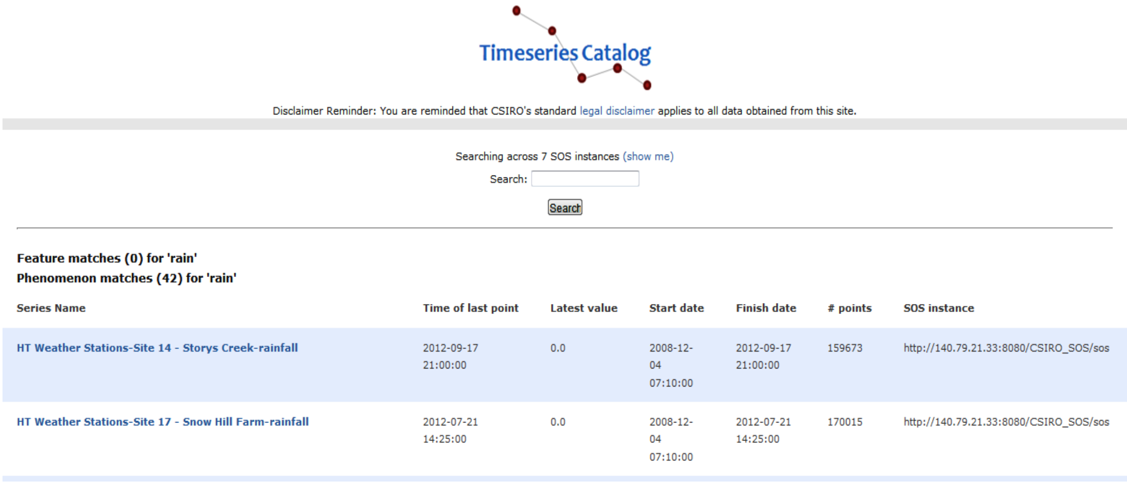

Fig. 3: Screen shot of the Web portal for Timeseries Catalog [4]. Data were accessed through this web interface using hypertext transfer protocol.

D. *Timeseries Catalog*

The Timeseries Catalog (Fig.3) 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 visualisation study, five different environmental variables (temperature, relative humidity, rainfall, solar radiation and wind speed) were acquired from estimates of 40 different sensor nodes available in the South Esk hydrological sensor web. A data service was developed to search, extract and download time series using the Timeseries Catalog hypertext transfer protocol [4].

IV. Sensor Web Visualisation

A. *South Esk Data Service Tool*

The data visualisation tool developed for this study, the South Esk Data Service Tool GUI interface, is shown in Fig. 4. Initially a coloured marker was placed on the exact location of the selected site, projected on the 2D coordinate system (magenta dot on the environmental surface view window in Fig.4). A display of pre-processed time series, data availability from sites and other environmental variables d and 3D surface visualisation (based on available patched point sensor node data) were the main features of this tool. Time series visualisation and 3D gridded environmental surface visualisation was based on average daily values [5].

Fig. 4: South Esk Data Service and dynamic recommendation Tool.

with a 'Not a Number' string. Filtered data was stored in a structured array.

B. *Node Data Filtering*

The pre-processing of downloaded time series was an important feature due to the uncertainty associated with data availability. Individual time series were identified according to the name of the selected site and environmental variable. 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. For some sensor nodes, there were a number of ±Infinite 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 [6].

In the next stage of data pre-processing, context based filtering was applied. Individual parametric filters were designed according to the sensor system ontology and valid operation range of a particular sensor. 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. Filtered data was stored in a structured array.

C. *Data Availability Visualisation*

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%. Fig. 5 shows the distribution of data availability for the South Esk sensor web. Darker circles represent higher data availability.

In the case that 70% or more of the data was available, nearest neighbour interpolation was applied to fill in the missing values of the time series. If the data availability was less than 70%, interpolation was not applied and the time series was presented with gaps. It was not essential to use time series with imputations for future analysis, but it was an important step in visualising data during the design of an application. This visualisation was quite helpful to compare original time series and interpolated time series in a statistically valid way. Fig. 6 shows an example of the comparative time series visualisation.

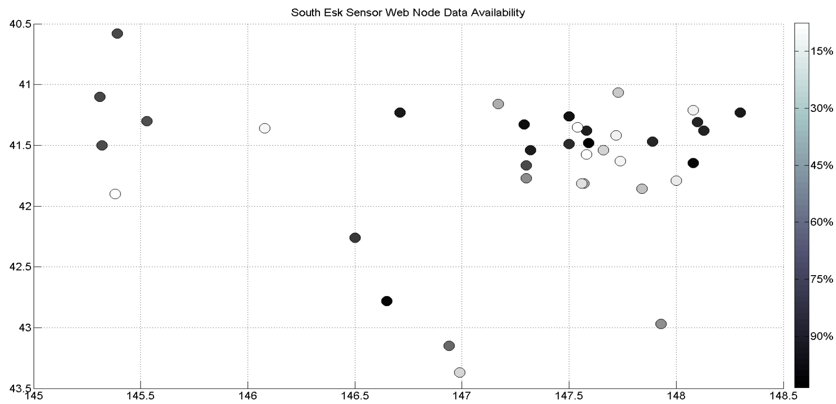

Fig. 5: Node based data availability for the South Esk sensor web.

Fig. 6: Recorded time series and interpolated time series with missing values for complete visualisation.

D. 3D Mesh Surface Visualisation

The concept behind this 3D surface visualisation was to provide an environmental gridded surface solely based upon the South Esk data. Dynamic data from 40 different sensor nodes were considered to create a 3D weather surface based upon cubic interpolation. The natural cubic spline interpolation is a form of interpolation where the interplant is a special type of piecewise polynomial called a spline. Spline interpolation is preferred over polynomial interpolation because the interpolation error can be made small even when using low degree polynomials for the spline. Spline interpolation avoids the problem of Runge's phenomenon which occurs when interpolating between equidistant points with high degree polynomials [7].

Individual surfaces were created for each environmental attribute. A daily surface was created based upon daily average data. Fig. 7 shows the interpolated 3D gridded mesh relative humidity (%) surface for the entire South Esk sensor web. Fig. 8 shows the 2D view of the same visualisation. Red dots indicate the unavailable sensor nodes for that particular day whereas green dots indicate the nodes that provided valid data. Surface was created based on available sensor nodes provided valid data.

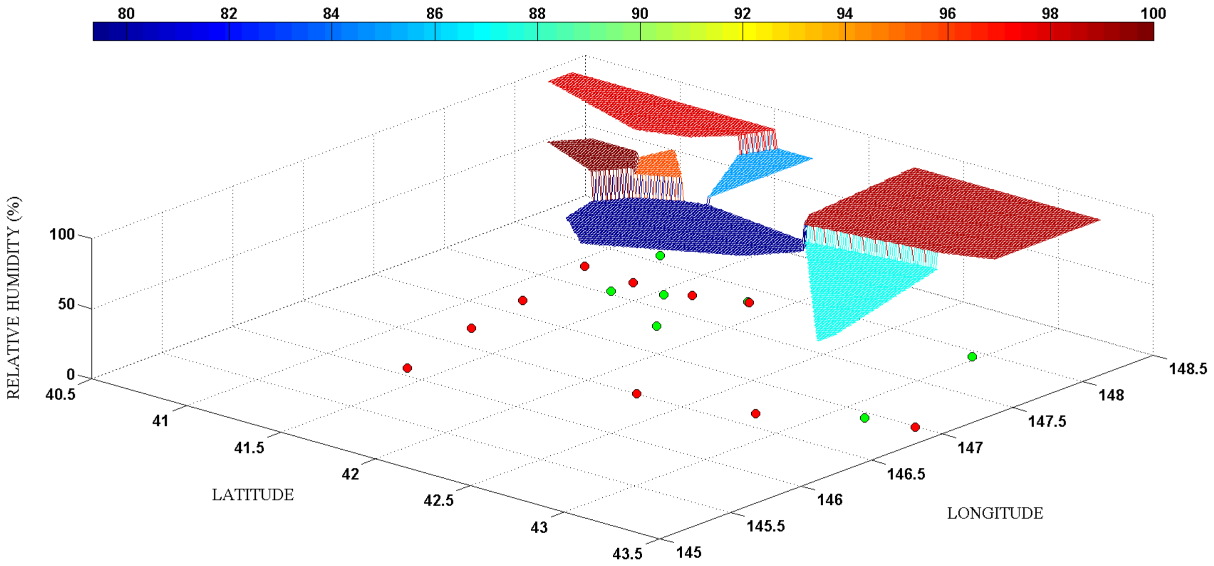

Fig. 7: 3D Gridded Mesh Relative Humidity Surface Visualisation on 17/08/2011

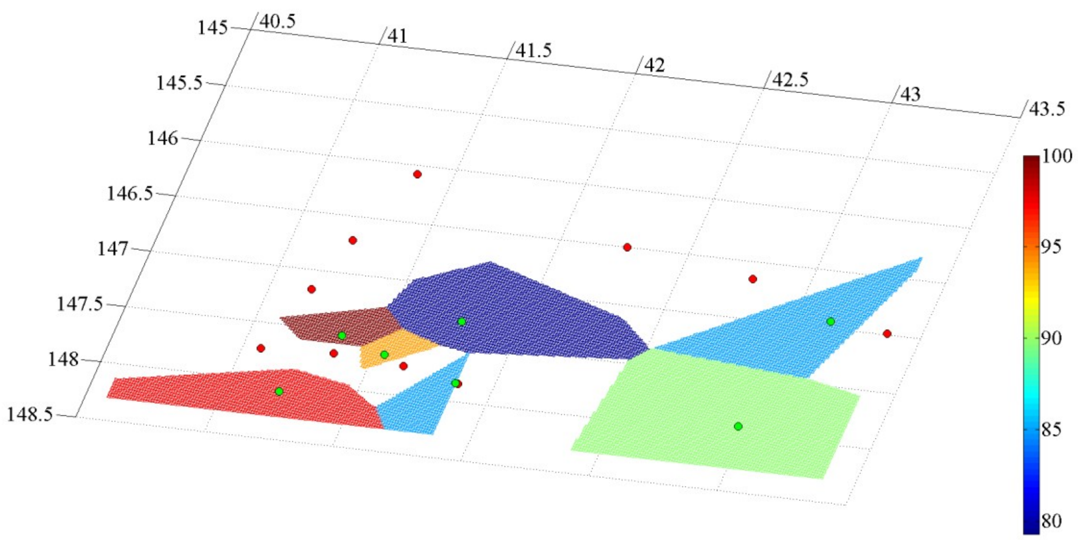

Fig. 8: 2D Gridded Mesh Relative Humidity Surface Visualisation on 17/08/2011

E. Dynamic Annotation and Recommendation

On the basis of data pre-processing, availability and interpolation results, a dynamic time series annotation system was developed to provide recommendations about the South Esk sensor web data. Individual time series were labelled as “Very Good”, “Good”, “Useful”, and “Poor”. Processed time series were stored in a data structure along with recommendations. The processed data incorporated additional statistical features: the maximum value of an event and its date, the minimum value of an event and its date, the longest missing value segment with corresponding dates and maximum number of consecutive days with the least data variance. All of this processed information was part of the dynamic data annotation system. 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. Statistical data annotation system’s recommendation helped the researchers to optimize the usage of data and increase the overall performance of any designed application significantly. This data visualisation based recommendation service was quite unique on its own merit but this also identified serious issues around sensor network data quality and data delivery. 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. Designing any hydrological application based solely on this network was almost impossible without proper visualisation and dynamic data annotation for the South Esk sensor web. This study shows the importance of the statistical data annotation, data recommendation and sensor web visualisation. This visualisation model could be adapted for any sensor network in the world.

V. Conclusion

Sensor webs are macro-instruments for sensing that can be used to integrate knowledge for enhanced situational awareness in decision support applications. One of the main issues with using large scale sensor webs in decision support is ensuring complementary data sources are fit for their intended purpose. The South Esk Data Service Tool was developed to analyze and visualize the uncertainty of sensor nodes belonging to the South Esk sensor web. Near real-time analysis of data quality and sensor availability was visualized and annotated to provide general recommendations. Sensor web assets were interpolated to produce a gridded three dimensional surface of climatic variability across the South Esk. Future work will involve developing new types of analysis for assessing the uncertainty of data using historical distributions. Furthermore, sophisticated data imputation methods will be developed using the spatial context provided by correlated sensor web assets and machine learning methods.

Acknowledgment

The CSIRO Intelligent Sensing and Systems Laboratory and the Tasmanian node of the Australian Centre for Broadband Innovation are assisted by a grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts. Authors would like to acknowledge Andrew Terhorst for providing access to the South Esk sensor web. Data were delivered through the CSIRO data service system developed by Pete Taylor, Chris Peters and Brad Lee.

References

[1] A. Nasipuri, and K. Subramanian, “Development of a Wireless Sensor Network for Monitoring a Bioreactor Landfill”, GeoCongress 2006. http://www.ece.uncc.edu/~anasipur/ pubs/geo06.pdf

[2] M. Bartsch, T. Weiland, and M. Witting, "Generation of 3D isosurfaces by means of the marching cube algorithm”, Magnetics, IEEE Transactions on Volume 32, Issue 3, Part 1, pp.1469 – 1472, May 1996

[3] The South Esk website. [Online] (2011). Available: http://www.csiro.au/sensorweb/au.csiro.OgcThinClient/OgcThinClient.html

[4] The Time series Catalog website. [Online] (2010). Available: http://www.ieee.org/http://www.csiro.au/sensorweb2/search .

[5] P. Buonadonna, D. Gay, J. M. Hellerstein, W. Hong, and S. Madden, “TASK: sensor network in a box”, Proceedings of the Second European Workshop on Wireless Sensor Networks, pp 133 -144, 2005.

[6] H. Yuxi, L. Deshi, H. Xueqin, S. Tao, H. Yanyan, “The Implementation of Wireless Sensor Network Visualization Platform Based on Wetland Monitoring”, Second International Conference on Intelligent Networks and Intelligent Systems, pp 224- 227, Nov 2009.

[7] B. Horling, R. Vincent, R. Mailler, J. Shen, R. Becker, K. Rawlins, V. Lesser, “Distributed sensor network for real time tracking”, Second International Conference on Intelligent Networks and Intelligent Systems, pp224- 227, Nov 2009.