

Dynamic Evaluation and Visualisation of the Quality and Reliability of Sensor Data Sources

Ritaban Dutta, Claire D'Este, Ahsan Morshed, Daniel Smith
Intelligent Sensing and Systems Laboratory
CSIRO, Hobart, Australia

Aruneema Das, Jagannath Aryal
University of Tasmania
Hobart, Australia

Abstract—Before using remote data sources, or those from external organisations, it is important to establish if the source is fit for purpose. We have developed an approach to automatic sensor data annotation and visualisation that evaluates overall sensor network performance and data quality. The CSIRO's South Esk hydrological sensor web combines data related to water management from five different organisations, which provides a suitable platform to explore the issues of reliability and uncertainty. An environmental gridded surface is generated based on the observations and evaluations of quality and reliability of the sensor node provider.

Keywords—*Dynamic Visualisation; Sensor Web; Weather Station; Time Series Catalogue; Interpolated 3D Surface.*

I. INTRODUCTION

Given the high cost of hardware, technical overhead, and significant maintenance required by environmental sensor networks there has been a shift towards sharing of data to distribute the load on organisations. Sensor data must be provided through means that promote re-use [1]. Standards such as Sensor Web Enablement from the Open Geospatial Consortium (OGC) [2] encourage sensor data interoperability. Web and cloud services assist with distributed data storage and public accessibility. However, these approaches do not provide assurances of the reliability and data quality of a sensor web.

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 sensors are often deployed unattended in harsh operational environments. The external causes of sensor uncertainty include their operation under extreme conditions, calibration drift, and bio-fouling. In addition, sensor nodes are subject to communication, software, electronic, and battery failures.

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. Automating the verification of the quality of the individual sources is essential to build trust in the users of these systems. 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 [3].

We have designed and developed 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 manufacturer's sensor specifications and attribute value range validations using well defined hydrological knowledgebase. Finally this study estimated and visualized a dynamic 3D surface map of the South Esk region based upon available environment data.

II. RELATED WORK

Automated assessment of sensor data has been explored in the marine domain to provide quality flags for data consumers [4].

There has been some attempt to quantify the reliability of a wireless sensor network. Most often the reliability analysis is performed with a probability graph [5] and include measures of factors such as fault diagnoses, analysis and recovery. Purohit et al. [6] modeled the hardware and software modules of a wireless sensor network as a series-parallel structure using a reliability block diagram approach. The problem can also be decomposed into sub-problems using a physical model to align with the dynamic nature of a sensor network [7].

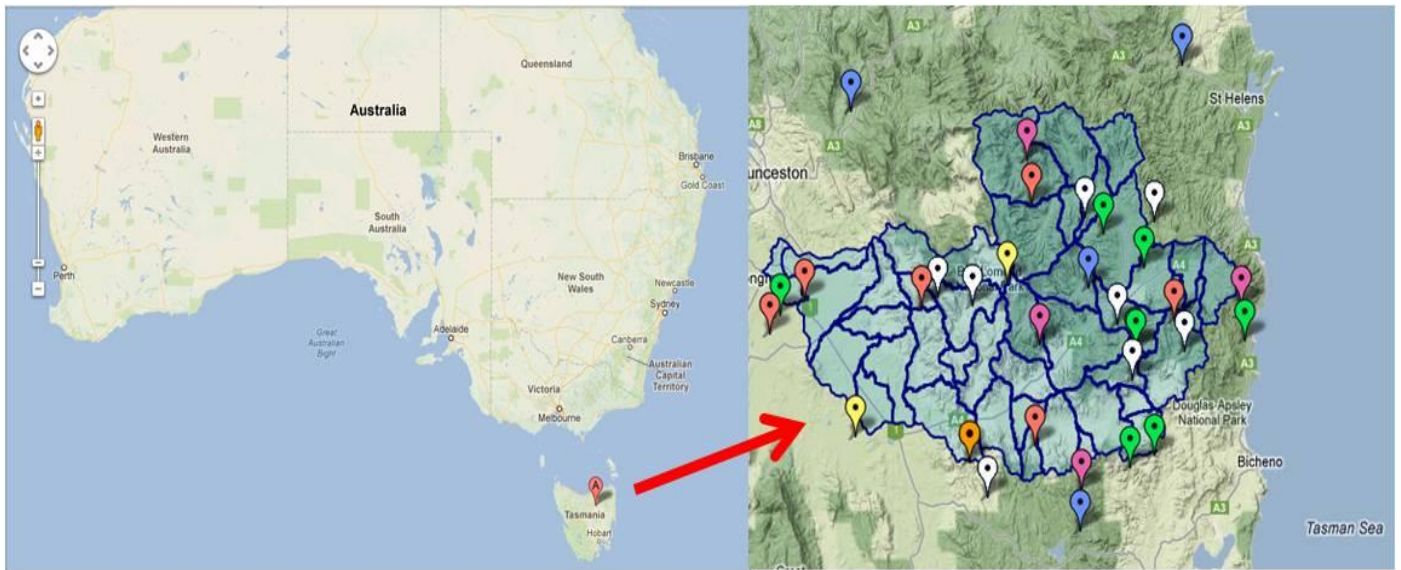


Fig. 1. The Google Maps™ pane presents a federated view of near real-time sensor data from the different sensor networks operating in the South Esk catchment (different colours correspond to sensors from different agencies).

III. THE SOUTH ESK HYDROLOGICAL 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 sensing macro-instrument. 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 river catchment was chosen because of its size (3350 square kilometers, large

enough to show up differences in catchment response to rainfall events), spatial variability in climate (an approximate 800mm range in average annual rainfall across the catchment), variable 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 specific SOS implementation was developed by the 52° North Initiative. Enhanced situational awareness of the catchment is gained by exposing sensor data via standard web service interfaces [8].

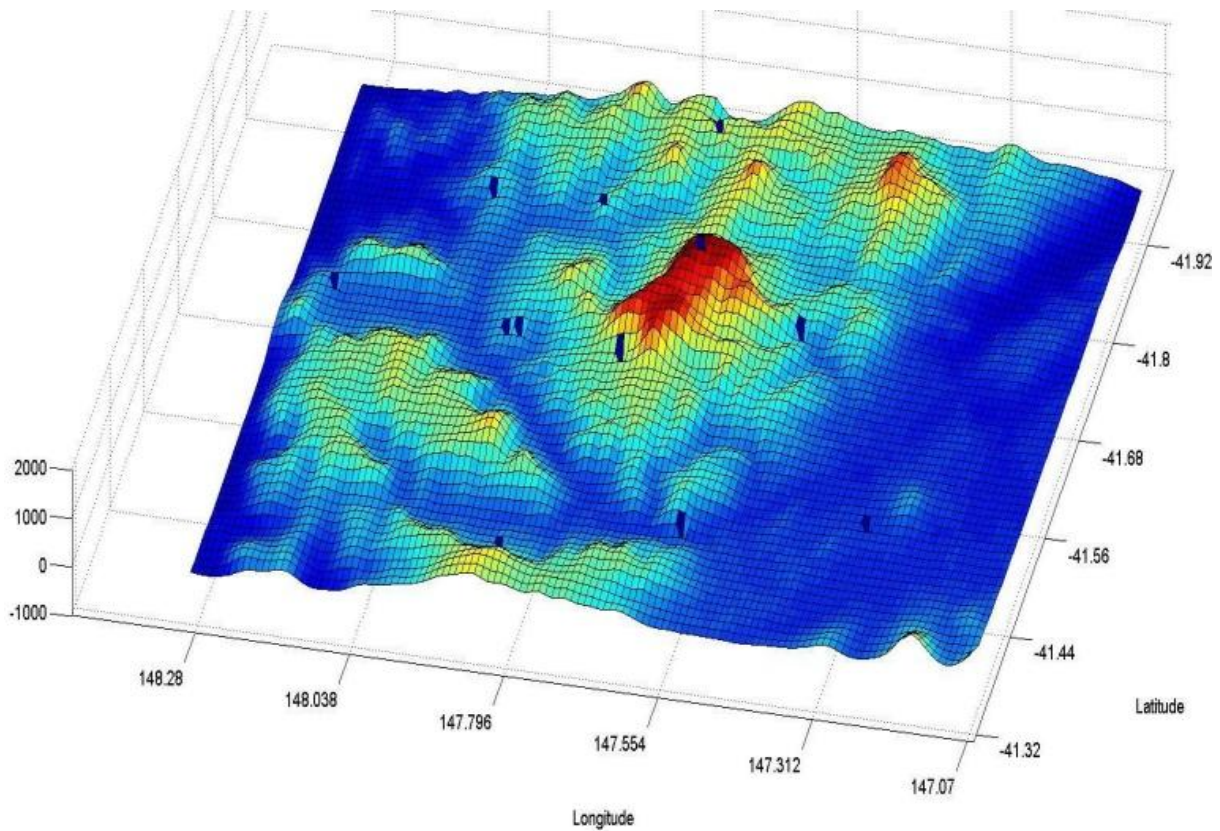


Fig. 2. Three dimensional map based on South Esk catchment topography data. Distributed sensor nodes locations are represented in dark blue.

B. Coordinate System

The South Esk sensor web covers an approximately 95 km \times 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.

Part of the study involved developing a South Esk Data Service Tool, which generates visualizations such as Figure 2. In this figure, the South Esk catchment is depicted as a 3D surface based on patched elevation data. The whole region was mapped as a gridded rectangle where each cell represents a 5km \times 5 km region. The physical locations of the weather station sensor nodes are visible as blue marks on the 3D surface.

C. Weather Stations

A small number of RIMCO 7499 tipping bucket rain gauges are operated by BoM in their weather stations.

Vaisala automatic weather stations made up the majority of the sensor network.

TABLE I. METEOROLOGICAL SENSORS IN THE SOUTH ESK SENSOR WEB

Sensor	Phenomenon	Method	Data Validity Range
Vaisala WMT52	Wind speed	Ultrasound	0 – 60 metres per second
Vaisala WM30	Wind speed	Ultrasound	0.5 – 60 metres per second
Vaisala RAINCAP	Precipitation	Acoustic impact of rain drops	0 – 200 millimetres per hour
RIMCO 7499	Precipitation	Tipping bucket	0.2 millimetre bucket
Vaisala HMP45A	Humidity	Glass substrate	-40 - +60 degrees Celcius
Vaisala HMP45A	Temperature	Passive thermocouple	0.8 – 100% relative humidity
RIMCO EQ08 & EQ08E	Total Global Radiation	Pyrano-Albedometer	0 – 2000 watts per square metre
RIMCO EP 16 & EP 16E	Upward/Downward Solar Radiation	Dual head (as above)	0 – 1400 watts per square metre
RIMCO Middleton SK08	Solar Radiation	Pyranometer plane surface	0 – 20000 watts per square metre

The Vaisala weather transmitter (WXT520) measures phenomena including barometric

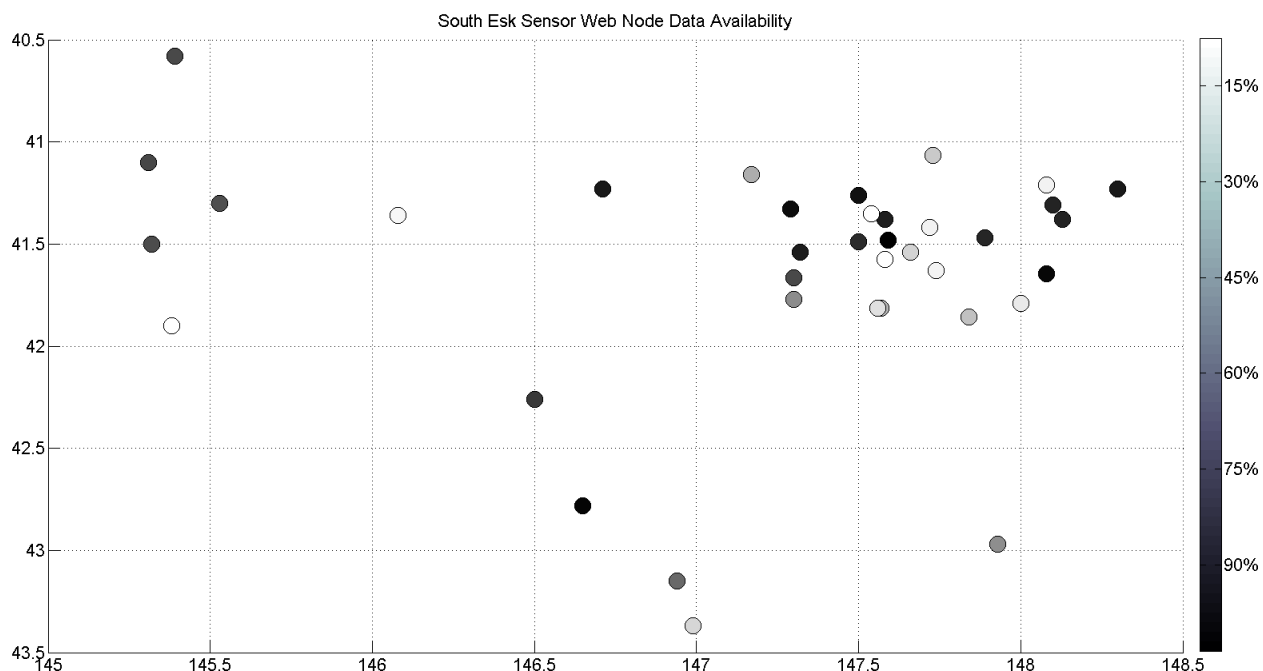


Fig. 3. Node based data availability for the South Esk sensor web. Darker node represents higher data availability with low uncertainty.

pressure, humidity, precipitation, temperature and wind speed.

Table 1 lists the sensors available in the South Esk including their valid data ranges. The valid measuring ranges were used later for dynamic data filtering purposes.

IV. SENSOR WEB VISUALISATION

A facility to query near real-time status updates from individual nodes in the sensor network was developed. The resulting system, the Timeseries Catalog, uses the available web services for this purpose.

For the visualisation study, we focused on five phenomena:

- temperature
- relative humidity
- rainfall
- solar radiation
- wind speed

The observations of these phenomena were acquired from the 40 sensor nodes 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 [9].

A. South Esk Data Service Tool

Figure 4 shows a sample from the data visualisation tool developed for this study, the South Esk Data Service Tool graphical user interface. Initially a coloured marker was placed on the exact location of the selected site, projected on the two dimensional coordinate system. The main features of this tool include display of:

- data availability from sites
- environmental observations
- pre-processed time series
- 3D surface visualisation

Time series visualisation and 3D gridded environmental surface visualisation was based on average daily values [10]. Three-dimensional surface visualisation was based on available patched point sensor node data. In reality some the sensor nodes not providing valid data at any given time were marked as red dot on the surface visualisation.

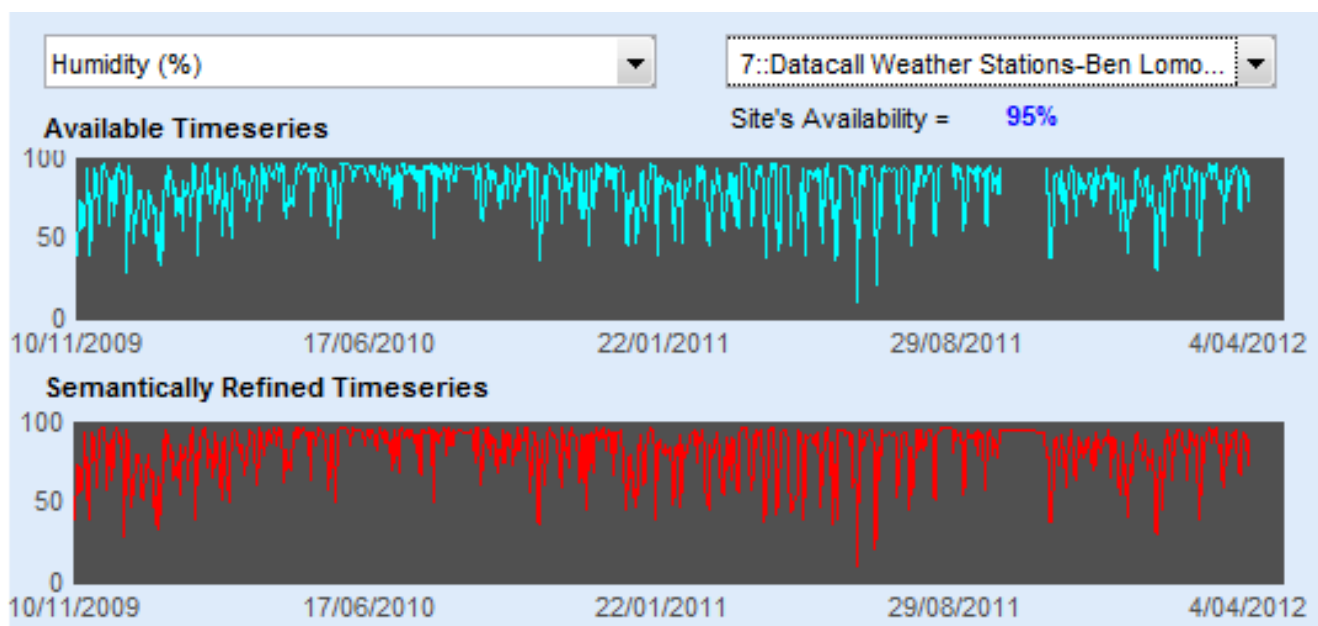


Fig. 4. Data visualisation tool developed for this study. Recorded time series and interpolated time series with missing values for complete visualisation.

B. Node Data Filtering

Pre-processing the 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 phenomena. The full time series were available since the beginning of deployment. Missing values are an unfortunate reality of nodes operating in remote, harsh operational environments and were present in these time series. For some sensor nodes, there were a number of \pm Infinite values. Initially a filter was designed to remove all of the \pm Infinite values, and replace them with a 'Not a Number' string to avoid introducing error and maintain the full time series length [11].

In the next stage of data pre-processing, context based filtering was applied. The valid operational ranges provided by the sensor manufacturers were used to design individual parametric filters. 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%. Figure 3 shows the distribution of data availability for the South Esk sensor web while representing all nodes historically provided humidity data. Darker node means more historical data availability from that node.

A threshold of at least 70% or more of available data was applied. For the time series above the threshold, nearest neighbour interpolation filled in missing values. If the data availability was less than 70%, interpolation was not applied and the time series was presented with gaps. Future analysis is still possible with incomplete time series, but imputed time series are advantageous for the visualisation. This visualisation assisted in the comparison of original time series and semantic feature based refined and interpolated time series in a statistically valid way. Figure 4 contains an example comparison between raw and processed time series data recorded from a single node.

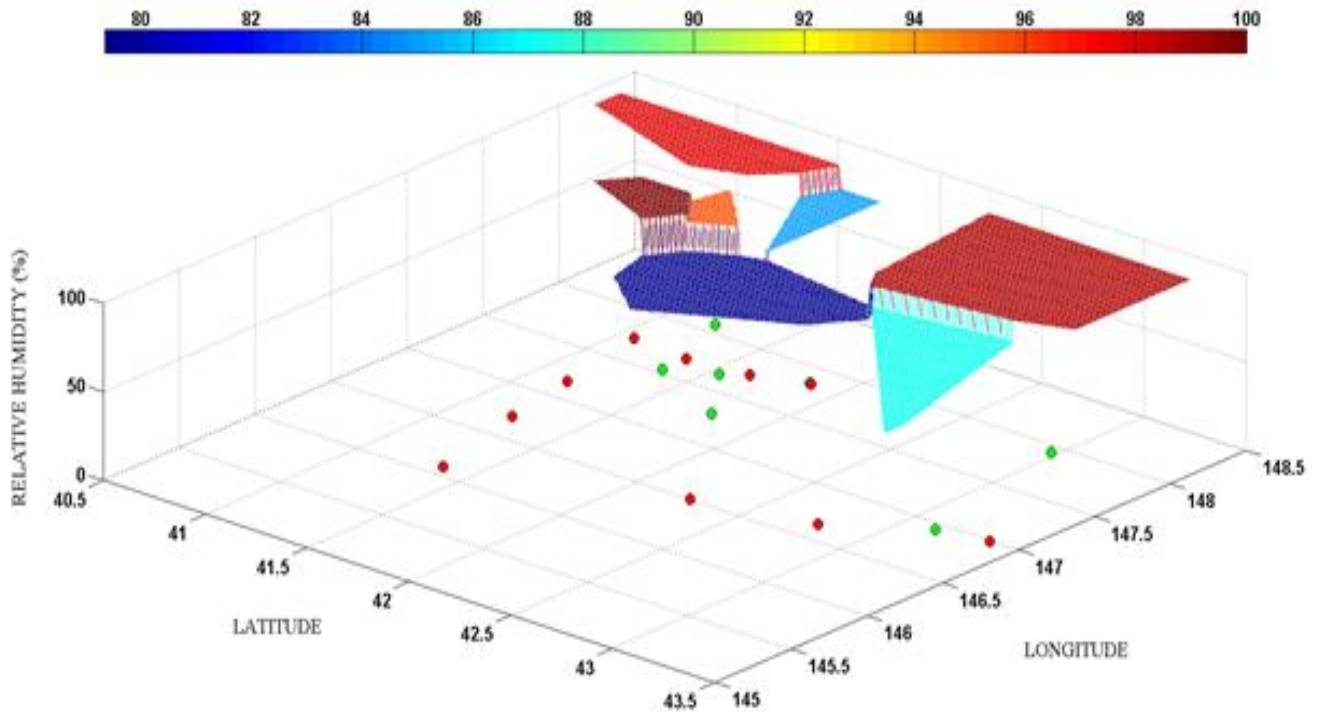


Fig. 5. 3D Gridded Mesh Relative Humidity Surface Visualisation on 19/12/2011.

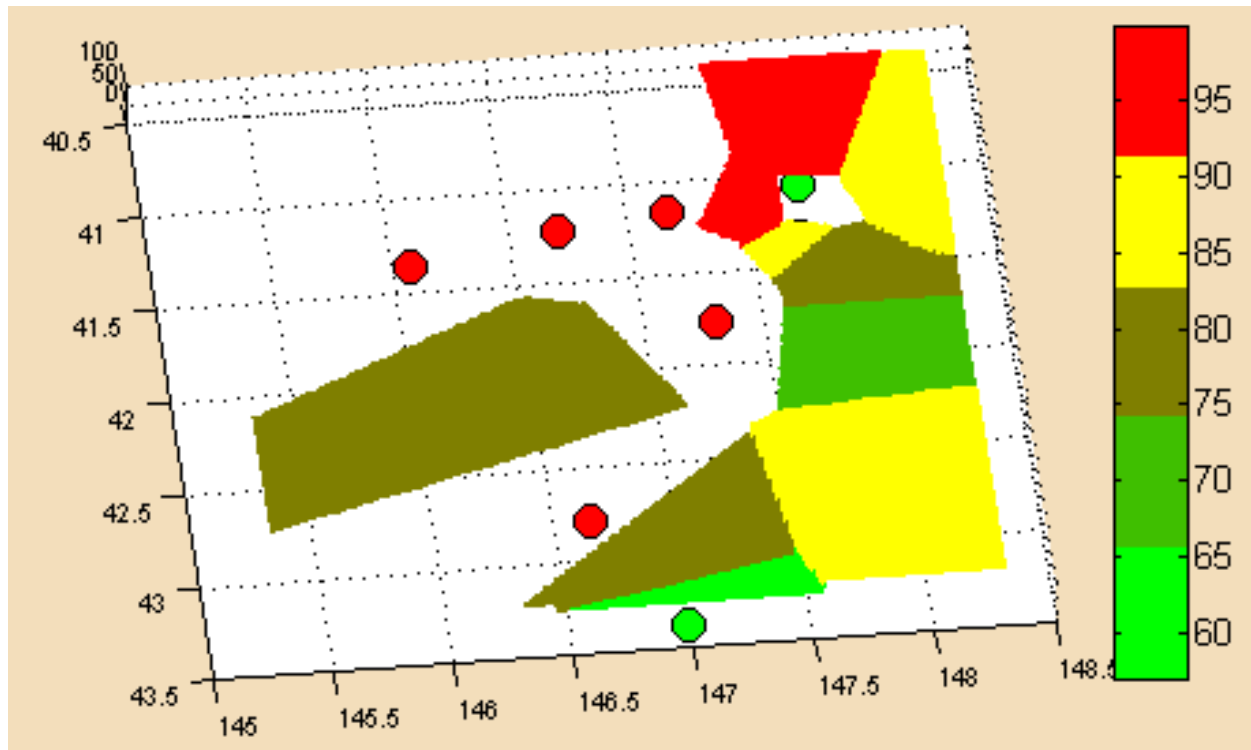


Fig. 6. 2D Gridded Mesh Relative Humidity Surface Visualisation on 19/12/2011.

D. 3D Mesh Surface Visualisation

The 3D surface visualisation was developed to provide an environmental gridded surface from the South Esk sensor web

data alone. Dynamic data from 40 sensor nodes was combined to create a 3D weather surface from cubic interpolation. The natural cubic spline is a form of interpolation that uses a piecewise polynomial interpolant called a spline. The benefit of

spline interpolation over polynomial interpolation is that 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 [12].

For each of the environmental phenomena an individual surface was created with a daily surface generated from daily averages. Figure 5 shows the interpolated 3D gridded mesh relative humidity (%) surface for the entire South Esk sensor web. Figure 6 shows the two-dimensional (2D) view of the same visualisation. The round markers in red indicate the unavailable sensor nodes for that day and green dots markers represent nodes that provided valid data. The final surface visualisation was created using only available sensor nodes that provided valid data.

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 data quality labels, namely {'Excellent Node', 'Good Node', 'Average Node', 'Poor Node', 'Damaged Node'}. Processed time series were stored in a data structure along with recommendations.

Additional statistical features were included in the processed data, including the:

- maximum value of an event and its date
- minimum value of an event and its date
- longest missing value segment with corresponding dates
- maximum number of consecutive days with the least data variance.

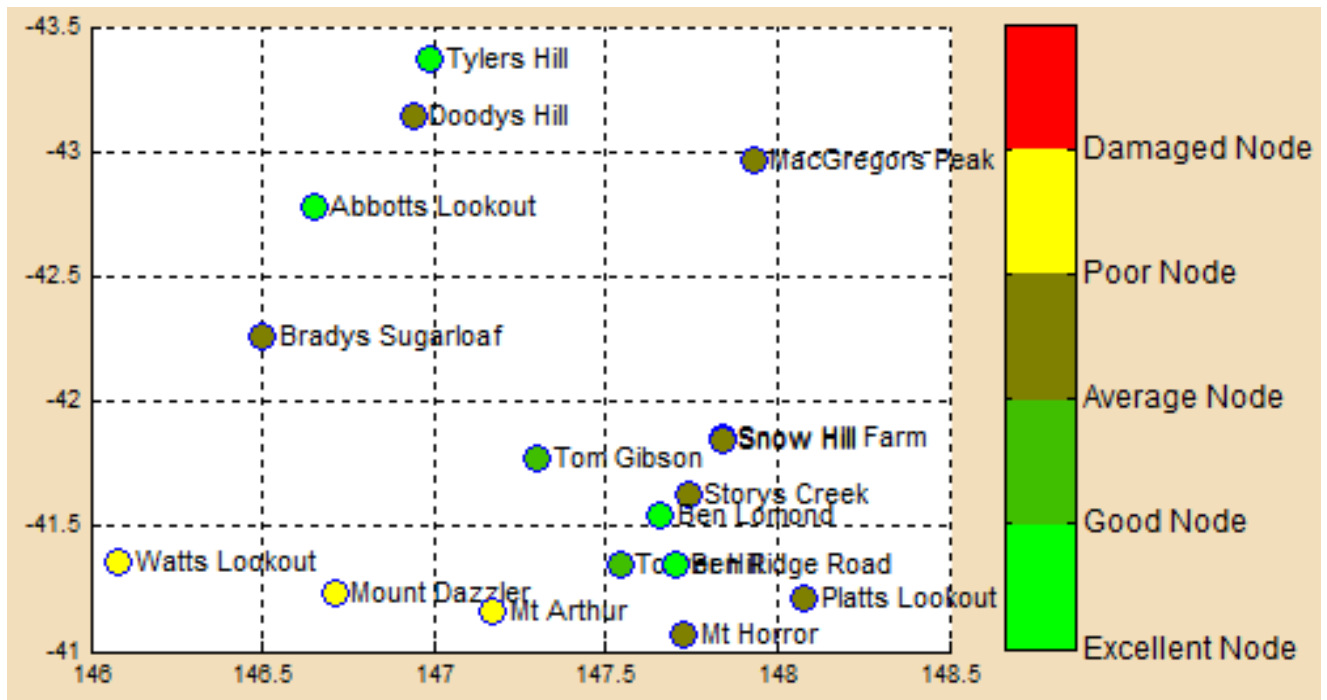


Fig. 7. Dynamic Annotation and Recommendation about the sensor network's node based data quality.

All of this processed information becomes part of the dynamic data annotation system. The purpose of this system is to process time series dynamically, annotate, and then provide a general data usability recommendation for users of the network. The recommendation of the statistical data annotation system can then assist researchers to optimize the usage of data and significantly increase the overall performance of any designed application.

Individual time series (representing individual sensor node) was labeled as one of the categories, namely, "Very Good Node ($\geq 90\%$)", "Good Node ($< 90\%$ and $\geq 80\%$)", "Average

Node ($< 80\%$ and $\geq 65\%$)", "Poor Node ($< 65\%$ and $\geq 50\%$)" and "Damaged Node ($< 50\%$)" depending on the values of the performance scores. These thresholds were defined according to the sensor specification and practical data quality experiences. Figure 7 shows the dynamic annotation and recommendation about the sensor network's node based data quality.

This data visualisation based recommendation provides a unique service, which also identifies serious issues around sensor network data quality and data delivery. The South Esk hydrological sensor web is a harsh operating environment

involving very difficult terrain (including a greater than 1500 metre mountain peak), which adversely affects data acquisition and delivery from many areas of the network. Any hydrological application based upon data from this is near impossible without evaluations of reliability and data quality.

V. CONCLUSION

Sensor webs are macro-instruments for sensing that can be used to integrate knowledge for enhanced situational awareness in decision support applications. The integration of data from large-scale sensor webs into decision support systems requires assurances that the complementary data sources are fit for their intended purpose. We have developed the South Esk Data Service Tool to analyse and visualise the uncertainty of sensor nodes in the South Esk sensor web.

We have presented a recommendation system that provides real-time analysis of data quality and sensor reliability that can be visualised and annotated. The sensor web data can be interpolated to produce a gridded three dimensional surface of climatic variability across the South Esk. In future work we are developing new types of analysis for assessing the uncertainty of data using historical distributions. We are continuing to investigate alternative data imputation methods using the spatial context provided by correlated sensor web assets and machine learning methods.

Although this study uses the South Esk sensor web as a use case the statistical data annotation, data recommendation and sensor web visualisation could be adapted for any sensor network in the world.

ACKNOWLEDGMENT

The authors wish to thank the Intelligent Sensing and Systems Laboratory and the Tasmanian node of the Australian Centre for Broadband Innovation. A grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts helped to conduct this research. Authors would like to acknowledge Andrew Terhorst for providing access to the South Esk sensor web. Data was 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. Botts, G. Percival, C. Reed, and J. Davidson. "OGC Sensor Web Enablement: Overview and High Level Architecture". GSN 2006, Lecture Notes in Computer Science, Volume 4540. pp. 175–190, 2008.
- [3] 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
- [4] D. Smith, G. Timms, P. de Souza, and C. D'Este, "A Bayesian Framework for the Automated Online Assessment of Sensor Data Quality", Sensors, Volume 12, Issue 7, pp. 9476-9501, 2012
- [5] H.M.F. AlboElFotoh, E.S. ElMallah, and H.S. Hassanein, "On the reliability of wireless sensor networks", Proceedings of the IEEE International Conference on Communications, pp. 3455-3460, 2006
- [6] H.L. Feng and S.Y. Liu, "Reliability analysis of a wireless sensor network based on a physical model", Journal of the Chinese Institute of Industrial Engineers, pp. 22-27, 2010
- [7] N. Purohit, P. Varadwaj, and S. Tokekar, "Reliability analysis of wireless sensor network," 16th IEEE International Conference on Networks, ICON 2008, pp.1-6, Dec 2008.
- [8] The South Esk website. [Online] (2011). Available: <http://www.csiro.au/sensorweb/au.csiro.OgcThinClient/OgcThinClient.html>
- [9] The Time series Catalog website. [Online] (2010). Available: <http://www.ieee.org/http://www.csiro.au/sensorweb2/search>
- [10] 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.
- [11] 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.
- [12] 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.
- [13] R. Dutta, D. Smith, G. Timms, "Dynamic Annotation and Visualisation of the South Esk Hydrological Sensor Web", pp 105-110, IEEE ISSNIP, Melbourne, Australia, April 2013.
- [14] R. Dutta, A. Morshed, "Performance Evaluation of South Esk Hydrological Sensor Web: Using Machine Learning and Semantic Linked Data Approach, Accepted in Special Issue of IEEE Sensor Journal on IoT, May 2013.