Intelligent Environmental Knowledge System for Sustainable Water Resource Management Solution

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

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

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

Abstract

Water shortage is a severe issue in many areas of Australia, particularly in the agricultural sector which is responsible for 65% of the total water usage. Water usage for irrigation and associated electricity costs are extremely high in Australia. So, accurate and timely decision support regarding efficient water usage is essential. Thus, a short-term low cost solution is needed to provide reliable advices on irrigation planning for the farmers such that the water wastage is optimized. Development of an integrated environmental knowledge recommendation system based on large scale dynamic web data mining and contextual knowledge integration to provide an expert water resource management solution was the main purpose of this work. Five different environmental data sources namely SILO, AWAP, ASRIS, CosmOz, and MODIS were integrated to develop and test the proposed knowledge recommendation framework called i-EKbase (intelligent Environmental Knowledgebase). Data driven supervised machine learning techniques namely Sugano type Adaptive Neuro Fuzzy Inference System, Multilayer perceptron network, Probabilistic neural network and Radial basis function network were used to learn and predict agricultural water balance for a specific location and given time period. This newly proposed predictive water resource estimation method based on large multi scale knowledge integration could potentially make the irrigation decision support systems more robust and efficient.

Keywords: Data Integration, Supervised Neural Network, Water Balance Equation, Water Resource Management, Resource Description Framework.

1 Introduction

In Australia, water shortage is a severe issue in many areas, particularly in the agricultural sector which is responsible for 65% of the total water usage. Much of the environmental degradation including salinisation is associated with the changes in the near-surface water balance induced by massive clearing of native vegetation and deforestation. These artificial changes have led to significant increases in groundwater re-charge, which in turn have led to rising water tables and salinisation. Other important aspect of the hydrological system is vegetation. Vegetation influences the hydrological cycle through the exchange of energy, water, carbon and other substances and is therefore critical for many hydrological processes, in particular transpiration, infiltration and runoff. The movement of water through the hydrological cycle varies significantly in both time and space. Australia, the driest continent, has the highest variability in rainfall and runoff and is therefore a difficult system to model [12, 13].

This presents a serious problem for sustainable agricultural development in the region because there is no reliable surface water for irrigation. So, the Australian farmers are always facing the big question about how much water they could buy and use for their potential future irrigation and crop growth. Water usage for irrigation and associated electricity costs are extremely high in Australia. Efficient and timely decision support regarding the sustainable management of water usage is essential. One way to overcome the problem is to combine field experiments with the conventional water balance modelling. But field experiments are expensive and only a limited number of land-use options could be trialled. So generalisation of the water balance estimation is near impossible just depending only on field experiments.

Even though regular weather data is available to the farmers but ultimately still farmers use their experience to make the water management decision as it’s extremely hard to interpret multivariate data. There is genuine need for multi source sensor and model data and knowledge integration to tackle this problem and to provide better decision support for the farmers based on strong scientific foundation rather than intuition. So there is a genuine need for multi-source sensor and model data and knowledge integration to tackle this problem.

In this paper novel knowledge integration and machine learning analysis based water usage recommendation system has been investigated and proposed. The proposed recommendation framework is called *i-EKbase*. This framework has the capabilities of on demand complementary knowledge integration from multiple data sources and automatic interpretation of the knowledge. In this paper *i-EKbase* has been applied for recommending and providing decision support regarding the sustainable management of water usage in Australian irrigation sector [2, 14, 15].

2 Science Challenge

Main focus of this paper was to develop an integrated multi source environmental knowledge framework to provide large scale availability of relevant sensor-model database to investigate unsupervised machine learning based data driven approach for hydrological application.

Second aspect of this paper was to use the integrated knowledge to calculate a water balance indicator using the conventional water balance equation over two years period. Idea behind this exercise was to develop a historic relationship matrix among the available environmental attributes and desired water balance estimation.

Third aspect of this research study was to apply machine learning based supervised algorithms on the historic relationship matrix data to develop a water usage recommendation and decision support system. Trained algorithm would then be used to provide water usage recommendation from the newly available environmental attributes.

3 *i-EKbase* Development

Five different environmental data sources were considered for the development of i-EKbase, namely, SILO [8], AWAP [6], CosmOz [7], ASRIS [5], and MODIS [9] (see Figure 1). Long Paddock SILO database is operated by the Queensland Climate Change Centre of Excellence (QCCCE) within the Department of Science, Information Technology, Innovation and the Arts (DSITIA). Australian Water Availability Project (AWAP) database is developed to monitor the state and trend of the terrestrial water balance of the Australian continent, using model-data fusion methods to combine both measurements and modelling. The Australian Soil Resource Information System (ASRIS) database provides online access to the best publicly available information on soil and land resources in a consistent format across Australia. The Australian Cosmic Ray Sensor Network (CosmOz) database is a near-real time soil moisture measurement network providing neutron counts related to bulk soil moisture. MODIS (MODerate resolution Imaging Spectroradiometer) database which includes data from Terra and Aqua satellites - viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups is available via NASA website. Individual web data adaptors were created to access and download data and integrate automatically based on semantic metadata matching mechanism. *i-EKbase* has been developed as an automatically adaptable dynamic knowledgebase.

3.1 Knowledge Integration

Based on any given location information (latitude and longitude) nearest available Bureau of Meteorology (BOM) weather station was selected based on geographical distance. SILO data file was downloaded and processed for that station. Nearest available CosmOz Data was also downloaded for the selected station. AWAP database was connected through a secured FTP server and grid files were downloaded locally. ASRIS database was downloaded from publicly available ASRIS website. For the same location a pixel position was derived on the daily continental AWAP gridded data and time series were extracted for individual variable for a given time frame. Similarly a pixel position was also calculated from The Australian Soil Resource Information System (ASRIS) data to extract soil resource information for the same latitude and longitude. MODIS images were downloaded and processed for that same location to extract time series data [9, 11].

3.2 Rain MODIS Data Processing

Tropical Rainfall Measuring Mission (TRMM) 3B42 satellite based precipitation products were constructed from the post real-time TRMM Multi-Satellite Precipitation Analysis (TMPA) product, 3B42 [9]. The purpose of the 3B42 algorithm is to produce TRMM-adjusted merged-infrared (IR) precipitation and root-mean-square (RMS) precipitation-error estimates. The algorithm consists of two separate steps. The first step uses the TRMM VIRS and TMI orbit data (TRMM products 1B01 and 2A12) and the monthly TMI/TRMM Combined Instrument (TCI) calibration parameters (from TRMM product 3B31) to produce monthly IR calibration parameters [4]. The second step uses these derived monthly IR calibration parameters to adjust the merged-IR precipitation data. These gridded estimates are on a 3-hour temporal resolution and a 0.25-degree by 0.25-degree spatial resolution which provided the adjusted merged-IR precipitation (mm/hr) and RMS precipitation-error estimates.

3.3 RDF Model Integrations

A unified knowledge integration and representation model was developed using the unified Resource Description Framework (RDF) model for *i-EKbase*. Unified knowledge RDFs were created for all the data sources based on pre-processed data, available meta data, and original provenance information. All RDFs were integrated in the next stage. Main purpose of this RDF based approach was to store the *i-EKbase* on a triplestore framework. A triplestore is a framework used for storing and querying RDF data. It provides a mechanism for persistent storage and access of RDF graphs. Recently, there has been a major development initiative in query processing, access protocols and triple-store technologies. *i-EKbase* was developed using a triple called "Sesame triplestore". Sesame is an open source framework for storage inference and querying of RDF data [10].

3.4 Experimental Data

Data for two different locations namely Daly River {131.4, -14.2}, and Tullochgorum {147.9, -41.7} were integrated and processed using i-EKbase for this study. These locations were selected to induce significant data variance in the generalization experiments as geographically they are quite different in nature. Daly River was a tropical savannah, whereas Tullochgorum was an improved pasture land. Environmental variables which were acquired and integrated for the time period 01/01/2011 – 31/12/2012 are listed in Figure 1.

4 Water Balance Model

Figure 1: Environmental variables extracted, processed and integrated from SILO, CosmOz, AWAP, ASRIS and MODIS to develop *i-EKbase*.

Index	Variable	Description	Unit
0	n-count COSMOZ	Atmospheric Pressure Corrected n-count from COSMOZ	N/A
1	maxT SILO	maximum temperature from SILO	degC
2	minT SILO	minimum temperature SILO	degC
3	rain SILO	Daily rainfall from SILO	mm
4	Evap SILO	evaporation from SILO	mm
5	Rad SILO	solar radiation from SILO	MJ/m2
6	VP SILO	vapour pressure from SILO	hPa
7	RH@Tmax SILO	Humidity at maximum temperature from SILO	%
8	RH@Tmin SILO	Humidity at minimum temperature from SILO	%
9	FAO56 SILO	FAO56 Potential Evapotranspiration from SILO	mm
10	Mlake SILO	Morton evaporation over shallow lakes from SILO	mm
11	Mpotential SILO	Morton potential evapotranspiration over land from SILO	mm
12	Mactual SILO	Morton actual evapotranspiration over land fromSILO	mm
13	Mwet SILO	Morton wet environment areal evapotranspiration over land fromSILO	mm
14	Span SILO	a comparison between measure class A pan evaporation and sythetic pan evaporation from SILO	mm
15	EvSp SILO	class A evaporation (used post 1970) followed by sythetic pan evaporation (pre 1970) from SILO	mm
16	MSLPres SILO	Mean Sea Level Pressure from SILO	hPa
17	SolarMJ AWAP	solar radiation from AWAP	MJ/m^2
18	maxT AWAP	maximum temperature from AWAP	degC
19	minT AWAP	minimum temperature from AWAP	degC
20	rain AWAP	Daily rainfall from AWAP	mm
21	MsoilUL AWAP	soil moisture (upper layer) from AWAP	Fraction (0-1)
22	MsoilULagg AWAP	soil moisture (upper layer)at end of aggregation period from AWAP	Fraction (0-1)
23	MsoilLL AWAP	soil moisture (lower layer) from AWAP	Fraction (0-1)
24	MsoilLLagg AWAP	soil moisture (lower layer) at end of aggregation period from AWAP	Fraction (0-1)
25	Evap(soil+veg) AWAP	evaporation (soil+vegetation) from AWAP	mm
26	totalTRANSP AWAP	Total Transpiration from AWAP	mm
27	Evap(soil) AWAP	Soil Evaporation from AWAP	mm
28	PotEvap AWAP	Potential Evaporation from AWAP	mm
29	LocDis (Run + Drain) AWAP	Local Discharge (Runoff+Drainage) from AWAP	mm
30	SurfRun AWAP	Surface Runoff from AWAP	mm
31	OpenWaEvap AWAP	Open Water Evaporation ('pan' equiv) from AWAP	mm
32	DeepDrain AWAP	Deep Drainage from AWAP	mm
33	SeniHeatFlux AWAP	Daily Sensible Heat Flux from AWAP	W/m^2
34	LatHeatFlux AWAP	Daily Latent Heat Flux from AWAP	W/m^2
35	% Clay ASRIS	Percentage Clay content	%
36	% Bulk Density ASRIS	Percentage Bilk Density	%
37	% Plant Available Water Capacity ASRIS	Percentage Plant Available Water Capacity	%
38	Temperature MODIS	Pixel temperature from MODIS	degC
39	Rain MODIS	Daily rainfall from MODIS	mm
40	Humidity MODIS	Daily Humidity from MODIS	%

In hydrology, a water balance equation can be used to describe the flow of water in and out of a system (Figure 2). A system can be one of several hydrological domains, such as a column of soil or a drainage basin. Water balance can also refer to the ways in which an organism maintains water in dry or hot conditions. It is often discussed in reference to plants or arthropods, which have a variety of water retention mechanisms, including a lipid waxy coating that has limited permeability [11-14].

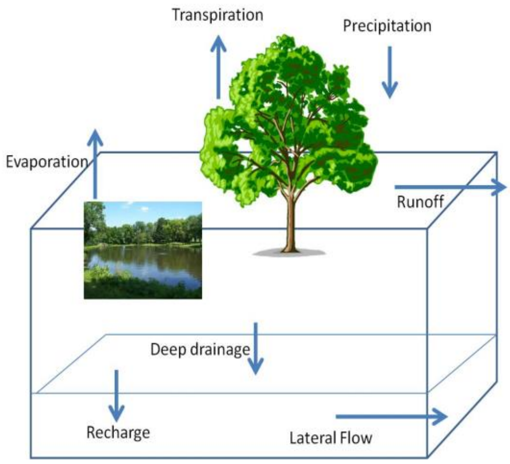
Water balance is based on the law of conservation of mass: any change in the water content of a given soil volume during a specified period must equal the difference between the amount of water added to the soil volume and the amount of water withdrawn from it.

The root zone water balance is usually expressed as:

$$\Delta S = P - E - T - RO - DD \qquad (1)$$

Where $\boldsymbol{\Delta S}$ is the change in root zone soil water storage over the time period of interest, $\boldsymbol{P}$ is precipitation, $\boldsymbol{E}$ is direct evaporation from the soil and water body surface, $\boldsymbol{T}$ is

Figure 2: Hydrological cycle schematic to form the water balance equation.



transpiration by plants and grass, $\boldsymbol{RO}$ is surface runoff or overland flow, and $\boldsymbol{DD}$ is deep drainage out of the root zone. All attributes are expressed in terms of volume of water per unit land area or equivalent depth of water over the period considered.

The recharge to the groundwater system can be calculated as:

$$RO = DD - SSF \qquad (2)$$

Where SSF is the lateral subsurface flow calculated according to Darcy's law. When the control volume is the entire catchment represented by given latitude and longitude information, the surface water balance equation can be expressed as:

$$\Delta < S > = < P > - < ET > - < Q > - < R > \quad (3)$$

Where $\Delta < S >$ is the change in spatially averaged catchment water storage, $< P >$ is spatially averaged precipitation, $< ET >$ is the spatially averaged catchment evapotranspiration, $< Q >$ is the spatially averaged catchment surface runoff, and $< R >$ is the spatially averaged catchment recharge. Equation 3 was used for *i-EKbase* based historic water balance calculation.

4.1 Irrigation Water Requirement Indicator

As described in the Figure 3, based on water balance equation and *i-EKbase* knowledge an irrigation water requirement indicator was calculated. This indicator was calculated for the whole duration of this research experiment. Historically there were only two possible water management decisions that one farmer could take on a day - either they did required to buy extra water for crop irrigation or there was enough water in soil that they did not required to buy extra water. Positive results from i-EKbase based water balance calculation indicated enough water in the soil (represented by Class 1 or '1' values) where as negative results indicated the necessity of buying water (represented by Class 2 or '2' values). A new time series was created to represent the variance of irrigation water requirement indicator over two complete years. Figure 4(a) and Figure 4(b) show the irrigation indicator profile for the Daly River and Tullochgorum locations as described in section 3.3.

5 Supervised Machine Learning (SML)

In a SML engine, a set of known samples (or known data) are systematically introduced to the learning algorithm, which then get trained, updates associated weight vectors and internally classifies data according to the known training targets or classes held in a knowledge base. In the second stage for identification, an unknown sample is tested against the knowledge base and then the membership class is predicted. Unknown samples are analyzed using relationships found in the initial calibration, learning or training stage. The idea of testing using unknown response vectors is a well-established technique and often referred to as cross-validation. Four supervised estimator, namely Sugano type ANFIS, Multilayer perceptron network (MLPN), Probabilistic neural network (PNN) and Radial basis function network (RBFN) were trained and tested independently for this study [1, 11].

Figure 3: Historic water balance indicator estimation based on integrated *i-EKbase*.

d<S> = Soil Moisture
= change in spatially averaged catchment water storage
= was calculated using i-EKbase variable set {**MsoilUL AWAP**, **MsoilULagg AWAP**, **MsoilLL AWA MsoilLLagg AWAP** and soil moisture represented by **neutron count COSMOZ**}

<P> = spatially averaged precipitation
= was calculated using i-EKbase variable set {**rain SILO**, **rain AWAP**, and **Rain MODIS**}

<ET> = spatially averaged evapotranspiration
= was calculated using i-EKbase variable set {**FAO56 SILO**, **Mlake SILO**, **Mpotential SIL**O, **Mactua SILO**, **Mwet SILO**, **Span SILO**, **EvSp SILO**, **Evap(soil+veg) AWAP**, **totalTRANSP AWAP**, **Evap(soil) AWAP**, **PotEvap AWAP**}

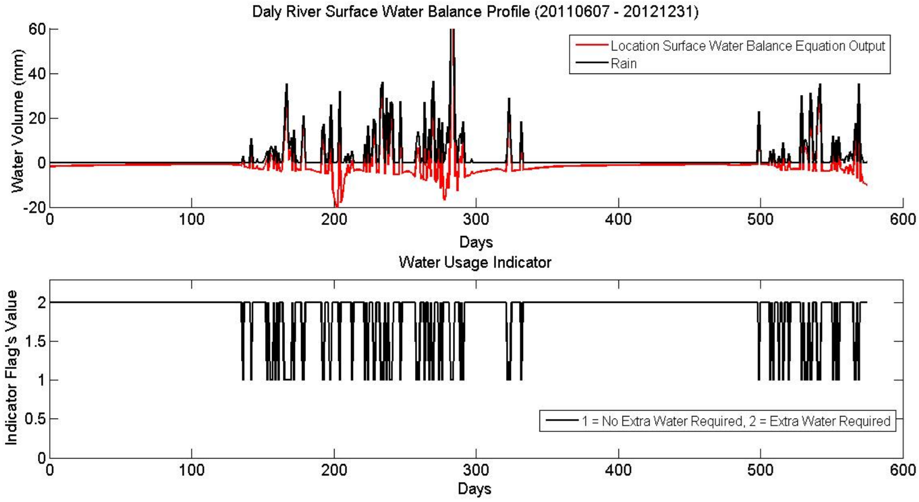
<Q> = spatially averaged catchment surface runoff
= was calculated using i-EKbase variable set {**SurfRun AWAP**, **OpenWaEvap AWAP**}

<R> = spatially averaged catchment recharge
= was calculated using i-EKbase variable set {**LocDis (Run + Drain) AWAP** and **DeepDrain AWAP**} while Sub-surface flow is considered zero.

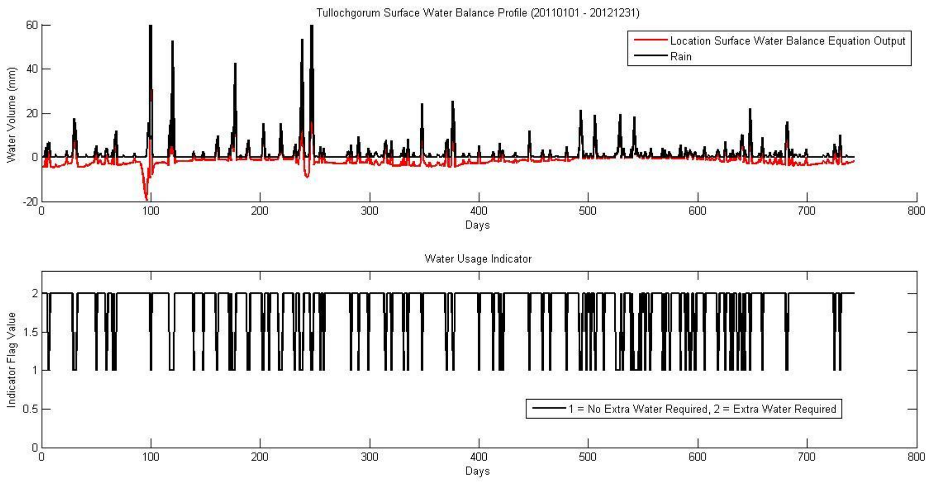
Conditional Rule:

d<Irrigation Water Requirement Indicator>
= <P> - <ET> - <Q> - <R> - d<S>
= positive if above catchment water storage capacity = ***Class 1***
= negative if below catchment water storage capacity = ***Class 2***
= ZERO if ideal balanced situation

Figure 4: Irrigation water requirement indicator and *i-EKbase* based estimated water balance profile for (a) Daly River location, (b) water balance profile for Tullochgorum location



(a)



(b)

The irrigation requirement indicator's time series profile was created to build a historic record of the most probable water management decision which was taken on a daily basis at a particular location. Any data driven machine learning approach requires training inputs and training targets. In this case 40 different environmental attributes of *i-EKbase* were training inputs (40 different time series representing daily data for same time period) where as irrigation indicator (one time series with same time length) based on water balance model was the training target. So, once the machine learning algorithm would be fully trained, any new daily data combination including all 40 attributes would then be processed as testing input by the trained algorithm and it would predict the probable water balance situation as Class 1 or Class 2. That class prediction would then be considered and interpreted as vital water management decision [1, 3, 11].

5.1 S-ANFIS Estimator

S-ANFIS is a kind of neural network that is based on the Takagi–Sugeno fuzzy inference system. Since it integrates both neural networks and fuzzy logic principles, it has the potential to capture the benefits of both in a single framework. The name ANFIS refers to one specific realization of such a system, architecture for a self-adaptive fuzzy system, taking a fuzzy inference system and tuning it with the back-propagation algorithm based on some collection of input-output data.

5.2 MLP Estimator

A MLP is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except both neural networks and fuzzy logic principles, it has the potential to capture the benefits of both in a single framework. The name ANFIS refers to one specific realization of such a system, architecture for a self-adaptive fuzzy system, taking a fuzzy inference system and tuning it with the back-propagation algorithm based on some collection of input-output data.

5.3 PNN Estimator

A PNN is a feed forward neural network, which was derived from the Bayesian network and a statistical algorithm called Kernel Fisher discriminant analysis. In a PNN, the operations are organized into a multilayered feed forward network with four layers, namely, Input layer, Hidden layer, Pattern layer/Summation layer, and Output layer.

5.4 RBFN Estimator

A RBFN is an artificial neural network that uses radial basis functions as activation functions. It is a linear combination of radial basis functions. They are used in function approximation, time series prediction, and control.

5.5 Experimental results

Three different evaluation data sets, namely, DATA SET 1: {Training 90%- Testing10%}, DATA SET 2: {Training 60% - Testing 40%} and DATA SET 3: {Training 49%- Testing 51%} were used for the training and testing of the ANFIS and three neural networks to estimate the generalization capability of these SML estimators in the current application context. Table 1 and Table 2 summarize all the generalization results (in terms of correct prediction percentages using all four SML estimators) for the Daly River and Tullochgorum respectively. Results are presented for using all three data sets {Data Set1, Data Set2, and Data Set3}. S-ANFIS was the best performer compared to the other three estimators while lesser amounts of data were used from training. Best result for Daly River location was 89.3% whereas for Tullochgorum it was 93.8% correct prediction. Both the best results were achieved using an S-ANFIS and DATA SET 2 where only 60% data were used for training. So achieved level of generalization was very encouraging in the context of predicting water balance and irrigation water requirement. Overall precision for Daly River data experiment was 90% where as for Tullochgorum was 92%.

Table 1: SML Evaluation (%) for Daly River.

	S-ANFIS	MLPN	PNN	RBFN
DATA SET 1	86.3	71.8	69.2	75.2
DATA SET 2	**89.8**	73.7	64.1	78.6
DATA SET 3	74.5	60.7	65.3	68.9

Table 2: SML Evaluation (%) for Tullochgorum.

	S-ANFIS	MLPN	PNN	RBFN
DATA SET 1	91.5	66.5	73.9	68.1
DATA SET 2	**93.8**	61.7	77.5	81.5
DATA SET 3	82.7	58.9	69.1	76.8

6 Conclusion

This paper has three main achievements. Firstly a multi -source environmental knowledge framework was developed to provide large scale availability of relevant sensor-model database for any environmental application. Next this integrated knowledgebase was applied to estimate historic surface water balance for two locations with significant geographical difference. Finally supervised machine learning paradigms were experimented to explore generalization capability and prediction accuracy of this proposed water resource management solution based on multi sensor – model integration. S-ANFIS based 93.8% accuracy performance proved that newly proposed predictive water resource estimation method based on large multi scale knowledge integration could potentially make the irrigation decision support systems more robust and efficient.

References

[1] C. T. Lin and C. S. G. Lee, “Neural Fuzzy Systems A Neuro-Fuzzy Synergism to Intelligent Systems”, Prentice Hall, Upper Saddle River, NJ, USA, 1996.

[2] Dawes, W.R. and Short, D.L. The efficient numerical solution of differential equations for coupled water and solute dynamics: the WAVES model. Canberra, CSIRO Division of Water Resources, Technical Memorandum 93/18, 1993.

[3] R Dutta, A Das, NG Stocks, D Morgan, Stochastic Resonance-based electronic nose: A novel way to classify bacteria Sensors and Actuators B: Chemical 115 (1), 17-27, 2006.

[4] Huffman, George J., David T. Bolvin, Eric J. Nelkin, David B. Wolff, Robert F. Adler, Guojun Gu, Yang Hong, Kenneth P. Bowman, and Erich F. Stocker. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology* 8 (1):38-55, 2007.

[5] http://www.asris.csiro.au/index_other.html. Accessed 12th November, 2012.

[6] http://www.eoc.csiro.au/awap/. Accessed 12th November, 2012.

[7] http://www.ermt.csiro.au/html/cosmoz.html. Accessed 12th November, 2012.

[8] http://www.longpaddock.qld.gov.au/silo/. Accessed 12th November, 2012.

[9] http://modis.gsfc.nasa.gov/. Accessed 12th November.

[10] http://www.w3.org/TR/REC-rdf-syntax/. Accessed 12th November, 2012.

[11] Folland, R; E. Hines, R. Dutta, P. Boilot, and D. Morgan. Comparison of neural network predictors in the classification of tracheal-bronchial breath sounds by respiratory auscultation, *Artificial intelligence in medicine*, 31 (3), 211- 220.

[12] White, R.E., Helyar, K.R., Ridley, A.M., Chen, D., Heng, L.K., Evans, J., Fisher, R., Hirth, J.R., Mele, P.M., Morrison, G.R.,Cresswell, H.P., Paydar, Z., Dunin, F.X., Dove, H. and Simpson, R.J. Soil factors affecting the sustainability and productivity of perennial and annual pastures in the high rainfall zone of south-eastern Australia. *Australian Journal of Experimental Agriculture*, 40(2),: 267–283, 2000.

[13] Zhang, L., I. H. Hume, M. G. O'Connell, D. C. Mitchell, P. L. Milthorpe, M. Yee, W. R. Dawes, and T. J. Hatton. Estimating episodic recharge under different crop/pasture rotations in the Mallee region. Part 1. Experiments and model calibration. *Agricultural Water Management*, 42 (2): 219–235, 1999.

[14] Zhang, L., Dawes, W. R., & Hatton, T. J. Modelling hydrological processes using a biophysically based model—application of WAVES to FIFE and HAPEX-MOBILHY, *Journal of Hydrology*, 185 (1): 147–169, 1996.

[15] Zhang, L., Dawes, W.R., Hatton, T.J., Hume, I.H., O'Connell, M.G., Mitchell, D.C., Milthorpe, P.L. and Yee, M. Estimating episodic recharge under different crop/pasture rotations in the Mallee region. Part 2. Recharge control by agronomic practices. *Agricultural Water Management*, 42(2): 237–249, 1999.