



Australian
National
University

Research School of Engineering

College of Engineering and Computer Science

ASSIGNMENT COVER SHEET

THIS SHEET SHOULD BE THE FIRST PAGE OF YOUR ASSIGNMENT WHEN IT IS SUBMITTED.

STUDENT ID NUMBER:u4646865.....

COURSE NAME: INDIVIDUAL PROJECT

COURSE CODE: ENGN4200

DUE DATE:14 Feb 2014.....

Submission of this assignment constitutes a declaration that:

- No part of this work has been copied from any other person's work except where due acknowledgement is made in the text; and
- No part of this work has been written by any other person except where such collaboration has been authorised by the course lecturer concerned; and
- No part of this work has been submitted for assessment in another course in this or another part of the University except where authorized by the lecturer(s) concerned.

Date: ____ 13 Feb 2014 ____

Signature: _____

Continental Scale Heterogeneous Knowledge Integration for Expert Agricultural Decision Support System

Project Midterm Review

Student: Cecil Li, u4646865

Primary supervisor: Dr. Ritaban Dutta, CSIRO

University supervisor: Dr. Jon Kim, ANU



ACKNOWLEDGMENT

I wish to thank my supervisor Dr. Ritaban Dutta for continuously supervising my work and providing insightful guidance. The CSIRO Computational Informatics, Hobart and the Tasmanian node of the Australian Centre for Broadband Innovation is providing the full support for my individual project. The cosmic ray sensor data was made available through the CosmOz Australian soil moisture monitoring network funded by CSIRO and deployed by Aaron Hawdon and Rex Keen of CSIRO Land and Water, Townsville, Australia. Soil samples for sensor calibration were collected by Bill Cotching of Tasmanian Institute of Agriculture. Data were delivered through CSIRO data service system developed by Peter Taylor. I would like to thank Peter Briggs and Simon Lovell for providing access to the AWAP database and SILO database. I would also like to acknowledge the use of FIRMS data and imagery from the Land Atmosphere Near-real time Capability for EOS (LANCE) system operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ

Abstract

Australia, despite being the 6th largest country of total area in the world, its water scarcity still a heated issue unresolved. Primarily in the agricultural sector, which accounts for 60% of the total water consumption nationwide. This study aims to develop an intelligent decision support system assisting in irrigation planning for the farming industry. The system integrates multiple data from different sources including but not limit to Bureau of Meteorology of Australia and National Aeronautics and Space Administration of the United States. In this study, various machine learning techniques are applied and tested to achieve optimal results, which then being delivered to the users in a visual manner via state-of-the-art mobile applications based on the Android OS, of which the work has been published and presented in an IEEE Conference in Baltimore, MD, USA in November 2013(Li et al 2013[1]). A more sophisticated and novel analysis is also discussed in this study, to explore the viability of using data-driven approach to model the impact of environmental change of any arbitrary region.

CONTENTS

1	Introduction	4
2	The Water Balance Model	4
3	The Datasets	6
4	Point-based Analysis& Application	7
5	Area-wise Analysis& Application(Work in Progress)	8
6	Further Development and Conclusion	9
	Appendix	11
	References	11

1 INTRODUCTION

Australia, despite being the 6th largest country in the world, area-wise speaking. Its water resources are heavily limited in the regions where the population and agricultural activities are concentrated. In the most recent Water Account Report released by Australian Bureau of Statistics (2013[2]), it is found that a total amount of around 75,000 GL of water was extracted and used within the Australian economy. Among which, the actual total consumption was 16,000 GL, a 20% increase from the previous annum. The agricultural sector alone accounted for 9,418 GL, or 59% of total water consumption, mainly for irrigation usage.

From the farming industry's point of view, proper planning of irrigation is also a significant issue to solve, due to the scarcity of water supply and highly priced electricity for operating irrigation systems, as well as the fact that excessive water in the soil could cause fungal infection in the plants[3], which results in reduced yields. Thus, a low-cost solution is needed to provide reliable advices on irrigation planning for the farmers in Australia so that wastage and costs can both be minimized.

One solution to the problem is to combine field experiments with conventional water balance modeling. However, field experiments are relatively expensive to conduct and has its limitations on specific land conditions, which means that it would be impractical to generalise the water balance estimation entirely on field experiment results. There is a need for on demand complementary knowledge integration from multiple data sources to increase efficiency and automatic interpretation of the knowledge.

This thesis explores the possibility of producing most probable high resolution prediction regarding the water balance in any region within Australia by implementing an intelligent system that can integrate multiple spatio-temporal data from various independent sensors and models, based on scientific foundation rather than human intuition.

In this thesis, a point-based approach is first implemented. Of which, a novel knowledge integration and machine learning analysis based water usage recommendation system has been investigated and proposed based on the CSIRO Sensor Cloud. To demonstrate the outcome, an Android OS based recommendation framework is developed for the purpose.

After validating the viability of such system, an area-wise approach is further developed. In this study, the novelty lies in where multiple points in a certain area are used to develop the Machine Learning model so that the limitation of conventional water balance model can be overcome by a data driven approach.

2 THE WATER BALANCE MODEL

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.

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.

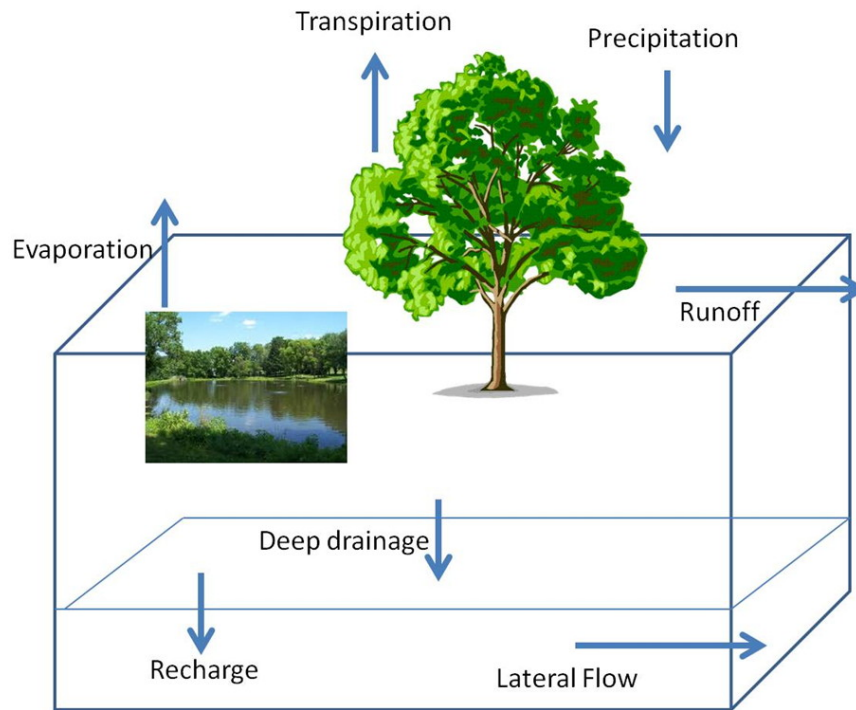


Fig. 1: Illustration of the Water Balance Model

The root zone water balance is usually expressed as:

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

Where ΔS is the change in root zone soil water storage over the time period of interest, P is precipitation that get to the soil, E is direct evaporation from the soil and water body surface, T is transpiration by plants and grass, RO is surface runoff or overland flow, and DD is deep drainage out of the root zone.

When the control volume is the entire catchment represented by given latitude and longitude information, the surface water balance equation can be expressed as:

$$\Delta \langle S \rangle = \langle P \rangle - \langle ET \rangle - \langle Q \rangle - \langle R \rangle \quad (2)$$

Where $\Delta \langle S \rangle$ is the change in spatially averaged catchment water storage, $\langle P \rangle$ is spatially averaged precipitation, $\langle ET \rangle$ is the spatially averaged catchment evapotranspiration, $\langle Q \rangle$ is the spatially averaged catchment surface runoff, and $\langle R \rangle$ is the spatially averaged catchment recharge. Equation 2 was used for the historic water balance calculation. A multi data source based "Water Balance Ontology" is created to represent this model in a machine readable structure.

```

1 Simplified Model for Catchment
2 Catchment Surface Water Balance Equation (Control Volume is the entire
  catchment)
3
4 d<S> = <P> - <ET> - <Q> - <R>
5       = SOIL MOISTURE = {MsoilUL AWAP // MsoilULagg AWAP // MsoillL AWAP //
  MsoillLagg AWAP}
6
7 d<S> = change in spatially averaged catchment water storage = ZERO Ideally
  when balance is perfect

```

```

8 <P> = spatially averaged precipitation {rain AWAP}
9 <ET> = spatially averaged evapotranspiration {Evap(soil+veg) AWAP +
    OpenWaEvap AWAP}
10 <Q> = spatially averaged catchment surface runoff {SurfRun AWAP}
11 <R> = spatially averaged catchment recharge {DeepDrain AWAP} while Sub-surface
    flow is considered zero.

```

Additionally, an irrigation water usage indicator is calculated for a clear representation of soil moisture level.

```

1 Conditional Rule:
2 d<Irrigation Water Requirement Indicator>
3   = <P> - <ET> - <Q> - <R> - d<S>
4   = positive if above catchment water storage capacity = Class 1
5   = negative if below catchment water storage capacity = Class 2
6   = Zero if ideal balanced situation

```

3 THE DATASETS

The implemented system integrates multiple data sources of various types, as listed in Table 1

Name	Abbreviation	Provider	Spatial Resolution	Temporal Resolution	Notes
Australian Water Availability Project	AWAP	The Bureau of Meteorology	5km	Weekly	A derived product providing 18 variables including Soil Moisture, Rainfall, Runoff etc.[4]
Australian National Cosmic Ray Soil Moisture Monitoring Facility	CosmOz	CSIRO Land and Water	12 Locations Nation-wide	60min	It adopts novel probes use cosmic rays originating from outer space to measure average soil moisture over an area of about 40 hectares to a depth up to 90 cm.
SILO climate data	SILO	The Science Delivery Division of DSITIA	5km	daily	Provides 15 variables including solar radiation, temprature, evaporation etc.
Australian Soil Resource Information System	ASRIS	CSIRO Land and Water	Varies	Varies	ASRIS provides online access to the best publicly available information on soil and land resources in a consistent format across Australia. It provides information at seven different scales.
Landsat	Landsat	National Aeronautics and Space Administration (NASA)	up to 15m	16 days with overlapping	Provides high resolution multi-band historic remote sensing data from 1970s. [5]

Moderate Resolution Imaging Spectroradiometer	MODIS	NASA	up to 250m	16 days	Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths. MODIS core mission, standard VI products include the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) to effectively characterize biophysical/ biochemical states and processes from vegetated surfaces. [6]
Digital Elevation Data	DED	Geoscience Australia, Australian Government	up to 30m	n/a	Provides digital elevation data which describes Australia's landforms and seabed is crucial for addressing issues relating to the impacts of climate change, disaster management, water security, environmental management, urban planning and infrastructure design.

TABLE 1: Data Sets used in the system

Based on any given location information (latitude and longitude), the data from previous sources listed in Table 1 is downloaded or extracted from pre-downloaded data storage, which will then be used for analysis and visualisation, as explained in the following sections.

4 POINT-BASED ANALYSIS& APPLICATION

As shown in Figure 2, an integrated Cloud-based decision support system is developed for conducting point-based analysis of the data. Essentially, the actual farmers, as the user, is able to enter and store his target location on a mobile application, which retrieves daily-updated water balance indicator from CSIRO Intelligent Environmental Knowledge Base (iEKbase) Cloud.

On the cloud side, where the servers are deployed, a few processes are involved. First of all, the Data Retrieval Module updates the database from multiple data sources, which then conveys an integrated data in a format of 36 columns time series array to the Machine Learning Process Module.

Principal component analysis(PCA) method is used to reduce the dimension of the variables. And *Linear Discriminant Analysis(LDA)* method is used to produce a trained supervised model for prediction. The most recent entry of the time series data is used as testing input to the model which returns a prediction indicator of water balance. A web interface is also constructed using the *Play 2.0* framework such that the mobile application can access to the up-to-date indicator data at anytime.

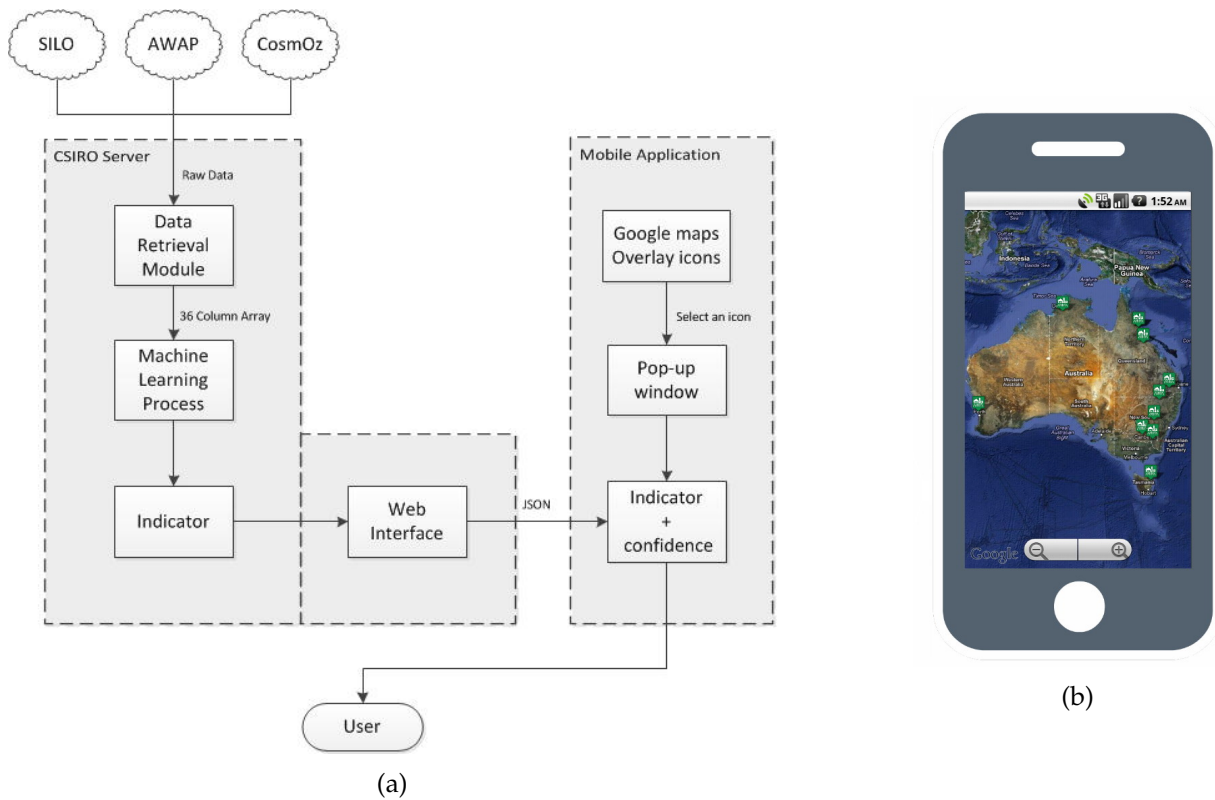


Fig. 2: (2a) Schematic diagram of point-based analysis and visualisation system
(2b) Screenshot of client-end mobile application

5 AREA-WISE ANALYSIS& APPLICATION(WORK IN PROGRESS)

As shown in Figure 3, an additional data *MODIS VI* provided by NASA, which has no dependency with AWAP data is added to the system. The target in the area-wise analysis is to train a Machine Learning model that can predict the water balance of any given location using historic and spatial data of that region.

This section of the project is still in-progress. The preprocessing and storage facility have been completed, which is essentially the preparation steps. To conduct the analysis, a variety of Machine learning algorithms is going to be implemented and tested, which includes but not limited to *Fuzzy-C-M*, *K-Means*, *self-organizing map(SOM) network*, *LDA* and *some ensemble neural network framework etc.*, as well as various water balance models that are in place. The historic data is going to be randomised in order to cross validate the precision of the model's output. The algorithm presents the highest probable results will be used for further development. The Houston salad farm located in Tasmania will be used as a case study for the analysis.

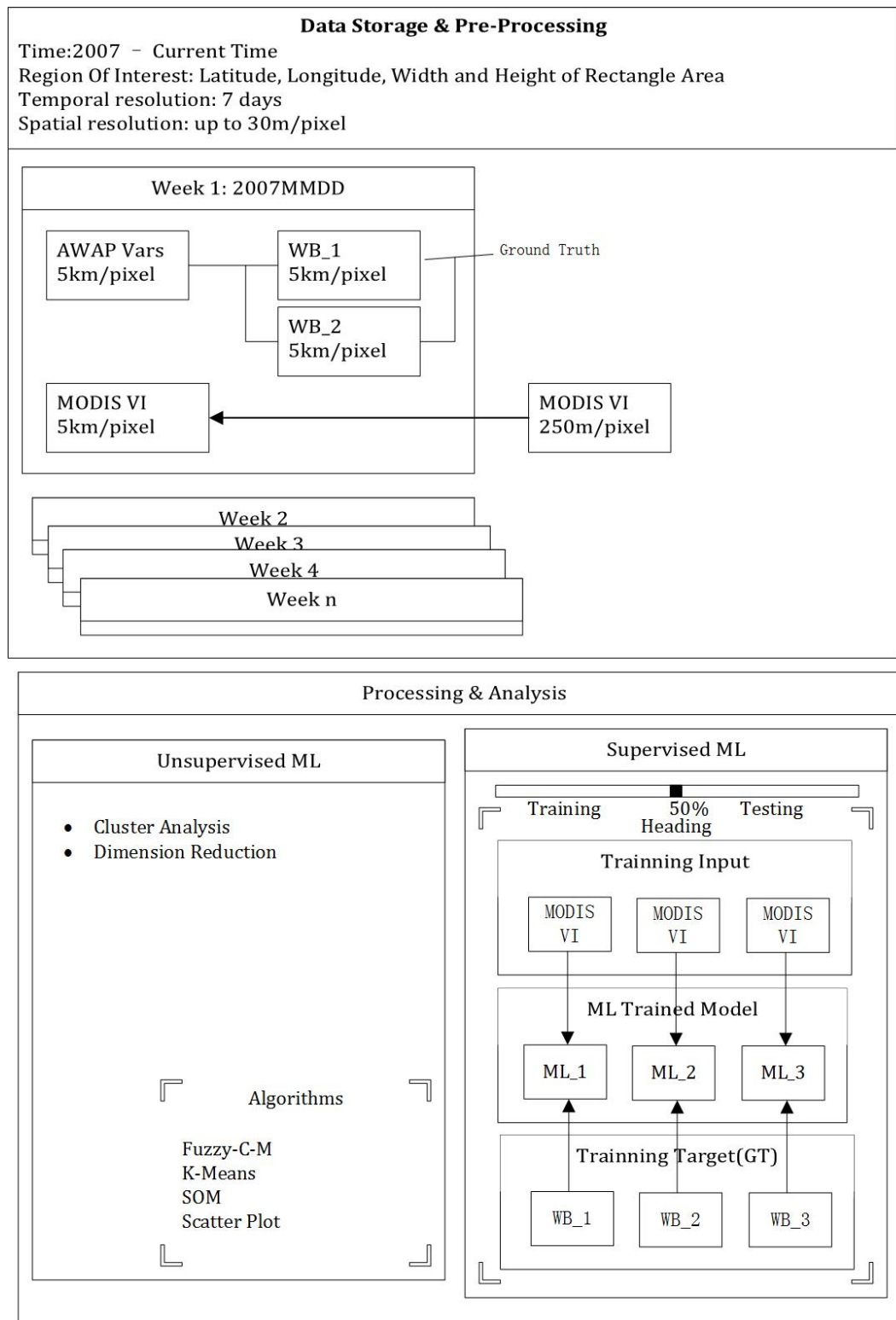


Fig. 3: Schematic diagram of area-wise analysis and visualisation system. Where WB_n is the various water balance model, ML_n is the machine learning framework

6 FURTHER DEVELOPMENT AND CONCLUSION

At this stage, the single point-based analysis and the mobile application has been finalised and presented in peer reviewed conference. The emphasis of the work is placed on conducting and

verify the viability of high resolution spatio-temporal analysis of any arbitrary location using a data driven approach. Most of the fundamental components for the final stage analytics have been completed. The analysis will trial as many potential algorithms and models as the time constrain allows.

APPENDIX

REFERENCES

- [1] Cecil Li, Ritaban Dutta, Corne Kloppers, Claire D'Este, Ahsan Morshed, Auro Almeida, Aruneema Das, and Jagannath Aryal. Mobile application based sustainable irrigation water usage decision support system: An intelligent sensor CLOUD approach. *2013 IEEE Sensors*, pages 1–4, November 2013.
- [2] Australian Bureau of Statistics. Water Account Australia 201112. Technical report, Australian Bureau of Statistics, 2013.
- [3] DH Wong, MJ Barbetti, and K Sivasithamparam. Effects of soil temperature and moisture on the pathogenicity of fungi associated with root rot of subterranean clover. *Australian Journal of Agricultural Research*, 35(5):675, 1984.
- [4] M R Raupach, P R Briggs, V Haverd, E A King, M Paget, and C M Trudinger. *Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component : Final Report for Phase 3*. Number 013. 2009.
- [5] LDC Mission. Landsat: A Global Land-Imaging Mission. *pubs.usgs.gov*, (May), 2013.
- [6] National Aeronautics. MODIS Vegetation Indices (MOD 13). volume 2, pages 173–175. 2000.