water e-journal
Online Journal of the Australian Water Association
ISSN 2206-1991
Volume 3 No 2 2018
doi.org/10.21139/wej.2018.010

A NOVEL DATA DRIVEN APPROACH TO ENSURING WATER SECURITY, VIA DATA ANALYTICS AND VISUALISATION

Exploring social-hydrology with advanced analytics to enhance long term regional water security

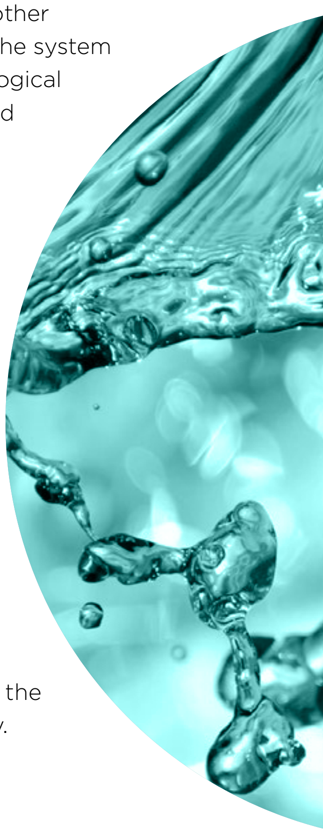
D Baker, M Blumenstein, J Li, J Sayer, S Gosh, R Mahendran, Y Zheng, S Fogelman, C Ferguson

INTRODUCTION

The NSW government has a strong position on data driven leadership and this also applies to managing water security and its associated infrastructure [1]. To align with the NSW Government objectives to become a World Leader in whole of government data analytics and insights, the NSW Department of Industries (DoI) Water has embarked on a big data project to provide the NSW Government with a more comprehensive picture of water demand requirements[1]. The focus of the project is to address the complex problem of being able to ensure water security and the quality of supply to meet the immediate and future long-term health, environmental and economic prosperity of communities and metropolitan areas across the state, whilst adapting to changes in water demand and availability based on climate and population variability[2, 3]. This also includes ensuring available water supplies for high security and productive economic sectors of agriculture and mining which contribute over $33 billion to the NSW economy.

The project aimed to develop a smart decision support tool based on data analytics and data mining that collectively incorporates the datasets currently maintained by DoI-Water, in addition to other government departments. This enabled the system to rapidly process factors such as hydrological information, future population growth and industry information, social, health and economic related information to identify issues and risks to each water source and Local Government Authority (LGA), through a Catchment Needs Assessment Framework (CNAF) [2, 4].

The CNAF Digitised System will enable government to look across the state and identify regions that would collectively benefit from infrastructure upgrades and therefore rationalise the economic costs associated with infrastructure building for regions. It would also enable DoI-Water to identify what socio-hydrological and environmental issues are affecting water security supply and demand, in order for the State to achieve long term water security.



The CNAF Digitised System will enable government to look across the state and identify regions that would collectively benefit from infrastructure upgrades and therefore rationalise the economic costs associated with infrastructure building for regions. It would also enable DoI Water to identify what socio-hydrological and environmental issues are affecting water security supply and demand, in order for the State to achieve long term water security. Through this process DoI Water and Infrastructure NSW are then able to ascertain the environmental, economic and social risk to the community if water demands are not satisfied and pre-emptively mitigate this occurring.

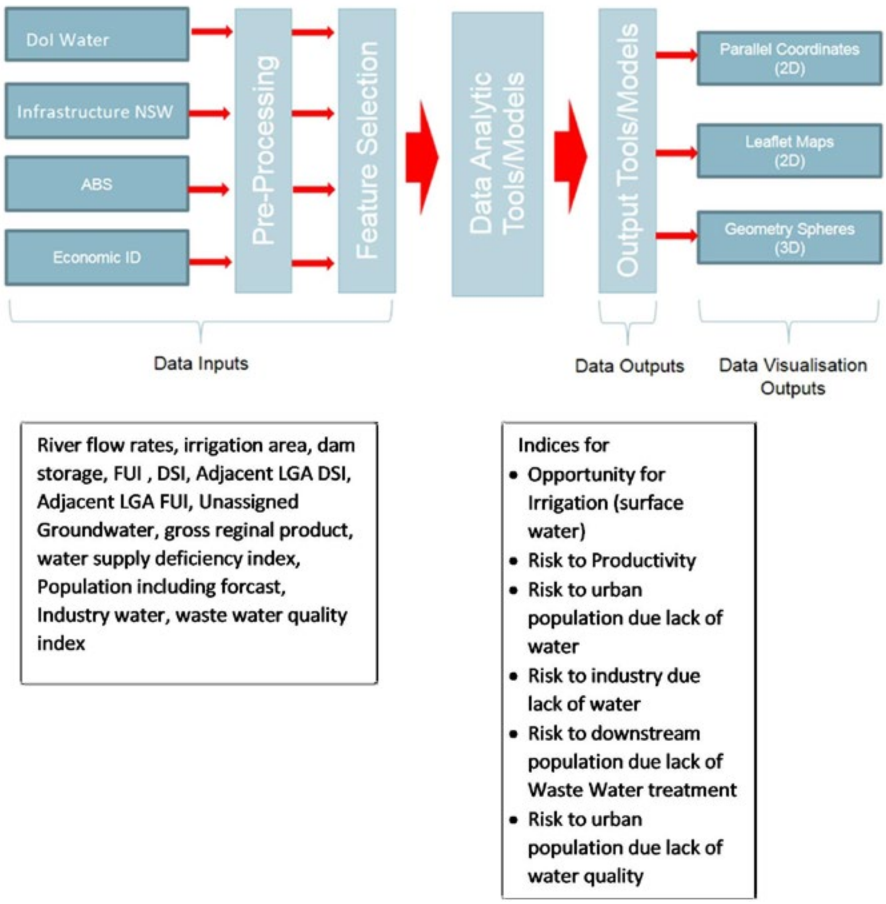


Figure 1. Shows a schematic overview of how the digitised CNAF project was developed including data visualisation outputs.

HIGHLIGHTS

- A digitised CNAF Decision Support Tool for Regulated and Unregulated River catchments was developed on the basis of the NSW government's key principles relating to data innovation.
- The Decision Support Tool included a visual analytic system which facilitates enhanced situational awareness to explore granularity CNAF including the ability to rapidly identify water security hot spots across the region.
- Water report cards for each priority valley and Local Government Area (LGA) can be produced using this tool.
- The output from this project will lead to increased operational efficiencies, infrastructure and cost optimization and risk mitigation for a variety of big water projects.

METHODOLOGY/ PROCESS

The CNAF was built around four primary qualitative indices [5] that were expanded to capture additional pertinent information around the social, environmental and economic resilience of a region:

- Irrigation Drought Security Index (DSI)
- Flow Utilization Index
- Flood Management Index
- Delivery Efficiency Index

The predictive methods and optimisation models considered an array of inputs including variables related to ecosystem, river networks, agriculture, natural resources and mines and health.

The CNAF model for regulated and unregulated catchments was developed using advanced machine learning methods (e.g. random forest [6], and multiple variable optimisation techniques [7] with expert opinion of priority to establish a robust CNAF index, which can map a wide range of variables with optimising targets such as water supply hot spots to develop a water security report card for each local government authority.

The CNAF was quantified for each of the indices between zero and one, where zero indicates no need to avoid risk and one indicates there is a significant need to mitigate risk.

The project used advanced data mining and machine learning techniques to identify the causes for hot spots in each LGA, based on raw data inputs. The Data Mining involved pre-processing, contextual understandings, feature building, data preparation and model design and subsequent sequential pattern mining. The outputs from the data analytic tools and pattern mining applied, were visually presented, so it was easier for a end-user to identify patterns or risks to catchment, from data that would otherwise be displayed in an excel spreadsheet.

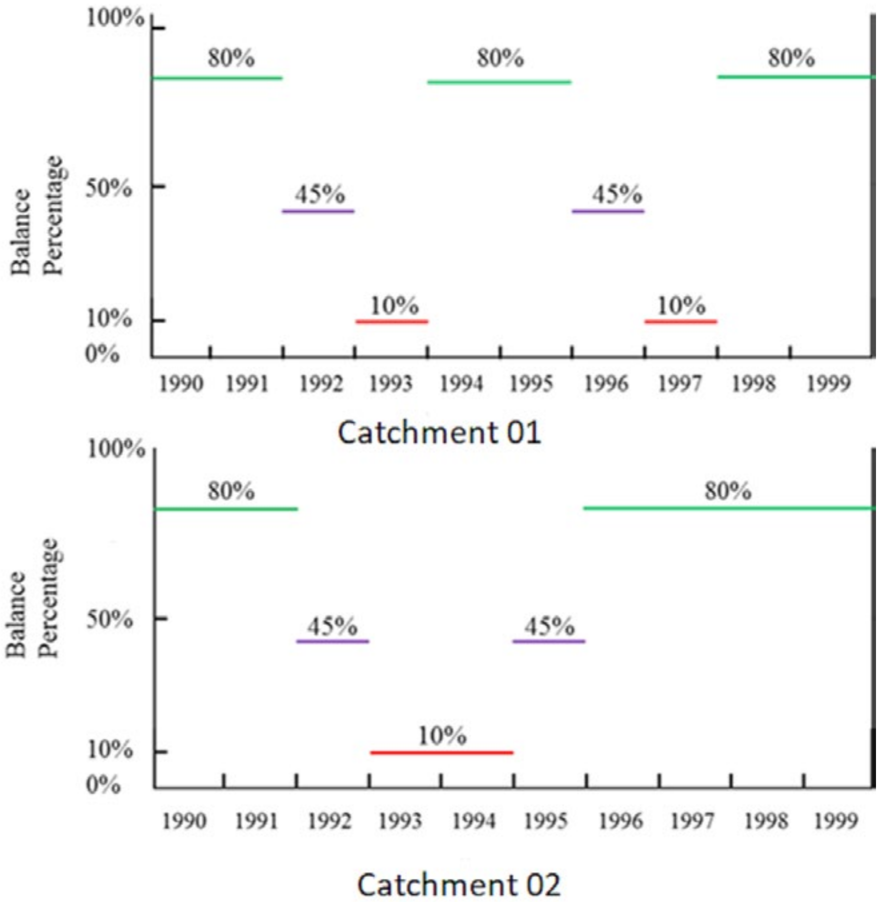


Figure 2. Shows an overview of 2 Catchments Drought Sequence Patterns and how new Drought severity risk can be used to characterise and quantify drought resilience patterns.

RESULTS/ OUTCOMES

A digitised CNAF decision support tool was built to provide the NSW Government with data analytics and insight to provide rapid evidenced-based decision making, of complex issues associated with ensuring immediate and long-term water security information. The CNAF was comprised of regulated and unregulated priority catchments. An unregulated CNAF was able to be developed based on inconsistent and limited data across the State, to develop a qualitative index set to determine water security issues. This has never been done before. The CNAF decision support tool was sufficiently able to compare and perform projected current and future forecasts to identify issues and types of risks to each water source. It has also provided enhanced insights to improve water security for a region through a rigorous and defensible assessment method. A comprehensive visual analytic system facilitates DoI-Water to explore the granularity of attributes with visual panel filters. This makes it easier for the end-user to identify areas of concern or opportunity.

Based on the indices we were able to quantify risks to each regulated catchment and explore the risks at a regional level. It gives an overall report card of the region and identifies the 'hot spot' focus areas in each catchment which should be further investigated.

It was evident from the analysis of the data that some catchments had similar Drought Security Index(DSI), despite having different historical drought sequences patterns (Figure 2).

The traditional IDSI is not able to characterise the impacts that sequential droughts have on a region. Therefore, we incorporated a risk and consequence score, to assess a regions ability to recover easily from periods of significant drought. Since the traditional DSI is not able to characterise the impacts that sequential droughts have on a region.

A scenario of balance percentage sequences of two catchments are visualized in Figure 2. The x-axis and y-axis represent the time and balance percentage respectively which is derived using the standard calculation formula [5]. Very similar DSI for two catchments, namely 38%. Although the two catchments both have six years at 80%, two years at 45% and two years at 10% in sequences, they occur over different time periods. For example, the first catchment 1 has two years at 80% between the first two drought years (1992 and 1993) and the second two drought years (1996. and 1997), while for catchment 2 there are four continuous drought years (1992-1995) where the balance percentage ≤ 50%. Therefore, actual drought situation of Catchment 2 has long periods of continuous drought years.

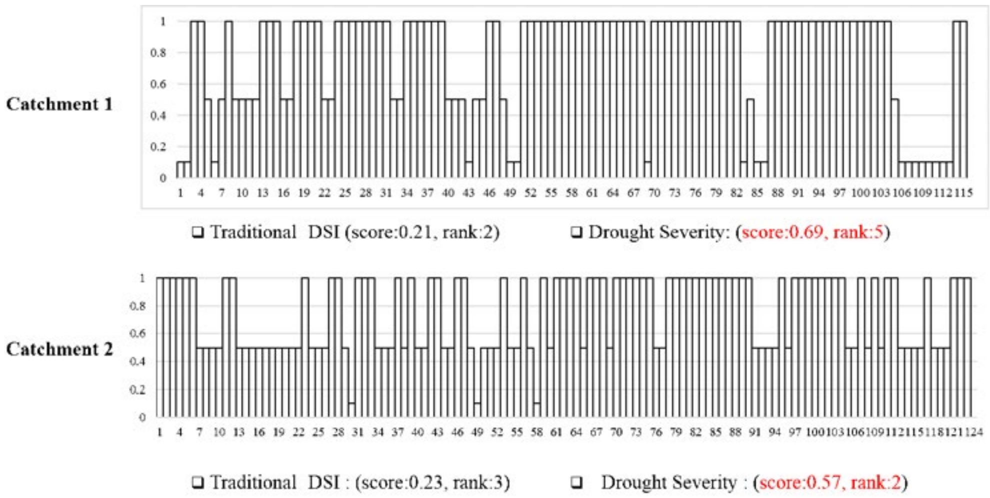


Figure 3. Shows an overview of 2 Catchments Drought Sequence Patterns comparing Traditional Drought Security Index and New Drought Severity index

Figure 3 shows how the new Drought Severity Risk (DSR) can be used to characterise and quantify drought resilience patterns of a region. It accounts for the length of time a catchment is in drought and the catchment's ability to recover. This was included in current rankings of regulated catchments to identify regions of priority and it did change the ranking of catchment 1 and catchment 2 (as shown below).

Parallel coordinates and corresponding leaflet mapping for the regulated CNAF were used to visually identify "hot spots" in the region (Figure 4 a and b) and provide a report card currently at the LGA level, so government end-users can understand the impacts of water security on local communities and business. Whereas the parallel coordinates in Figure 4 c can demonstrate how it can be used at a catchment level. For example, it indicated there were only two secure water catchments in terms of satisfying the 4 overall water security indexes. This indicated areas where there could be further enhancements to meet the long-term needs regions. Figure 4 c enabled the exploration of the causes and linked with figure 4 a and 4 b explore the risks to the LGAs in those catchments, based on water requirements.

UNREGULATED CATCHMENTS

The unregulated catchments decision support tool shown in Figure 5 provides DoI-Water and Department of Infrastructure NSW with an interactive snap shot of how each catchment at a LGA level is being affected by climatic, agricultural, social economic conditions using parallel coordinates. As previously mentioned this type of mapping had never been done before, as it was based on inconsistent or limited data sets, compared to the regulated catchment.

From the data we can visualise patterns that are affecting the region (including identifying 'hot spot' areas or where there are opportunities for infrastructure investments to support the regional economies and communities. The benefits of the tools is that it can easily identify regions with low flow utilisation that could be used to support nearby regions with low drought security. It can also be used to identify areas of potential economic opportunity that have been underutilised.

For example, if we explore the parallel coordinate named "Weighted Index" (shown in Figure 5 a and b) for the unregulated catchments, we can easily select the top LGAs that have low FUI. From this we can explore its drought security and total entitlements patterns (based on industry and populations in the regions), as well as availability of water with a high availability of irrigable land.

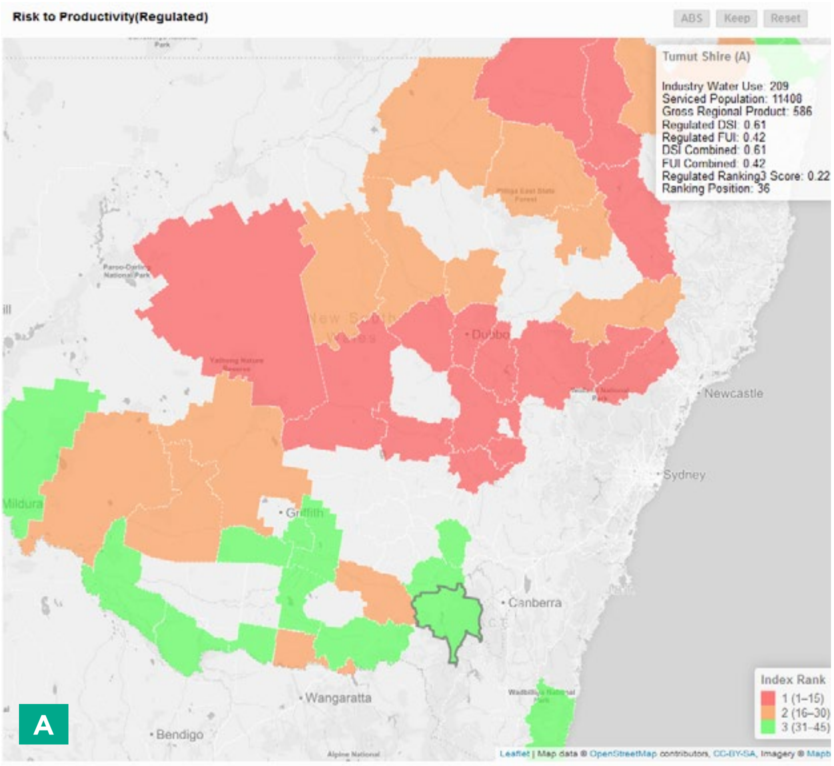
Data Analytics and Visualisation

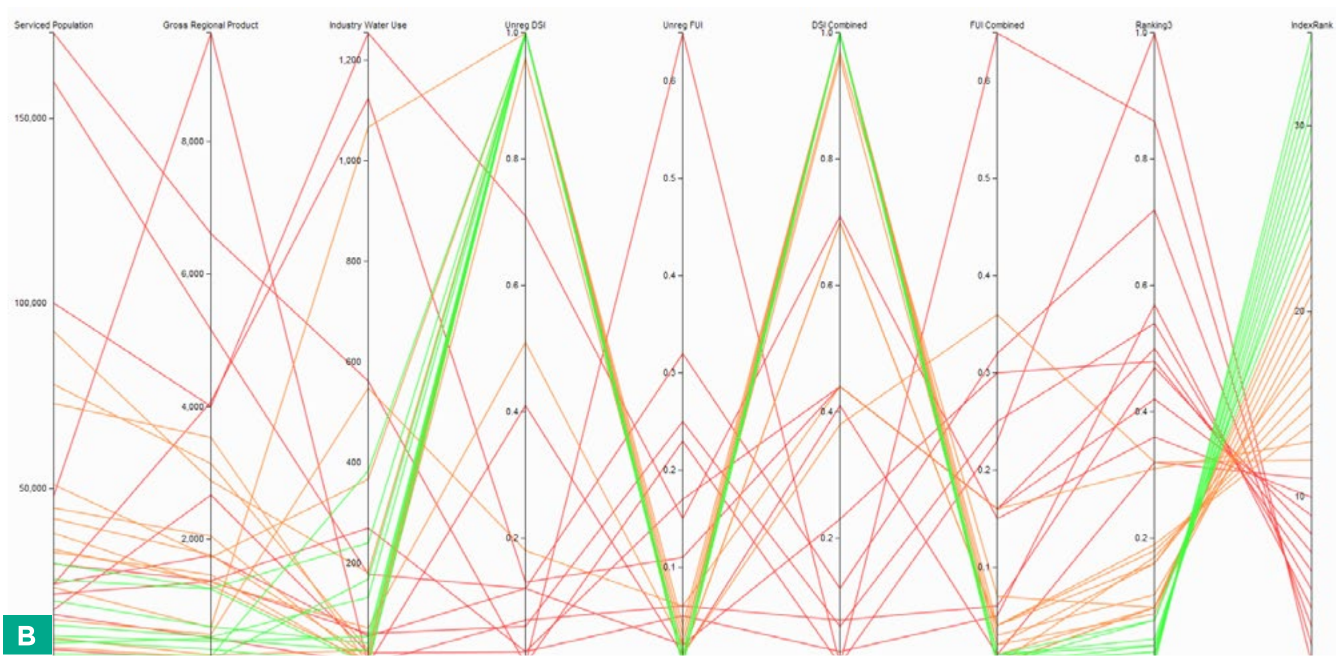
This information could be used to identify where there is a surplus of water and with adequate infrastructure could improve the gross regional product of that area or provide water security support to a neighbouring region with low drought security.

CONCLUSION

A novel approach to determine water security hot spots for the NSW region was developed via data analytics and data visualisation, to support evidence based decision making for water security. It enables DoI-Water to rapidly explore large volumes of information pertaining to water security, in a visual and understandable manner. This provides additional benefits when comparing and forecasting scenarios at a state, regional and local government level, to identify how to better ensure future water security.

The analysis was carried out in LGA level and catchment level. Further analysis at the sub-LGA scale is needed to plan at the project scale.





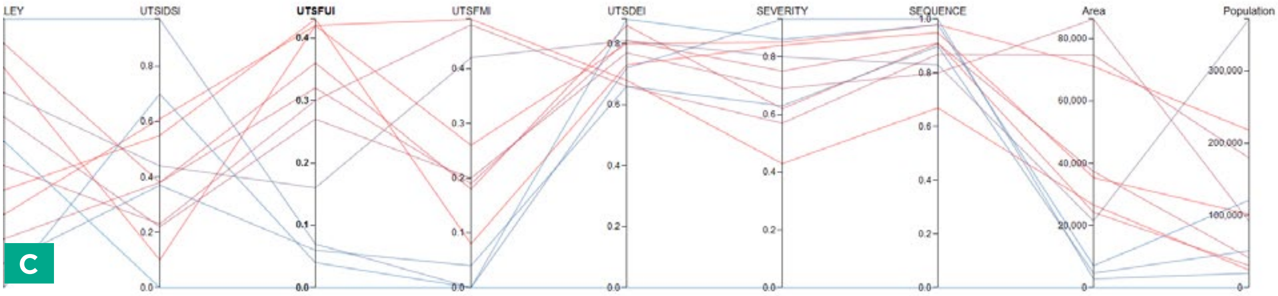


Figure 4. a and b Shows an overview of all the regulated catchments based on climatic, agricultural and social economic conditions visually at the LGA level using parallel coordinates and leaflet mapping and c) using parallel coordinates at the catchment level to ascertain the factors affecting catchment water security. Note: Catchments/LGAs have been De- identified for publication purposes

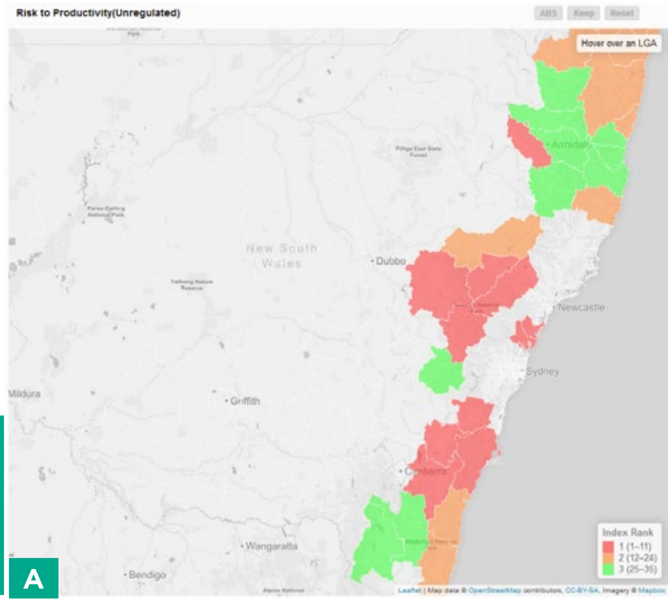
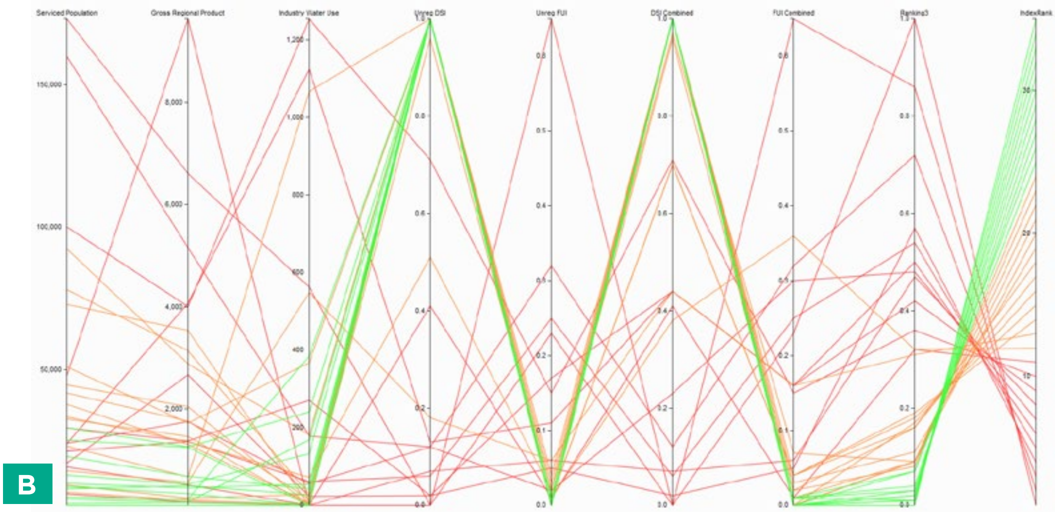


Figure 5. Figure 5 a) shows the information at a LGA level and b) Shows an overview of all the unregulated supplies based on climatic, agricultural and social economic conditions using parallel coordinates.

Note: Catchments/LGAs have been De- identified for publication purposes



REFERENCES

State of New South Wales (2015), DIGITAL+2016, NSW GOVERNMENT ICT STRATEGY FINAL

UPDATE, Department of Finance, Services and Innovation

Hirabayashi, Y, Mahendran,R et al (2013) Global flood risk under climate change. Nature Clim. Change. 3(9): p. 816-821.

Dong, G. and J. Li (1999) Efficient mining of emerging patterns: Discovering trends and differences. in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.

Candelieri, A., D. Conti, and F. Archetti (2014) Improving Analytics in Urban Water Management: A Spectral Clustering-based Approach for Leakage Localization. Procedia - Social and Behavioral Sciences. 108: p. 235-248.

Department of infrastructure, 2017, The opportunity Chapter 6 Water, avauilable at http://www.infrastructure.nsw.gov.au/media/1096/inf_j14_871_sis_report_ch06_web.pdf

Liaw, a, and M Wiener. 2002. "Classification and Regression by randomForest." R News 2 (December): 18–22. doi:10.1177/154405910408300516.

Pohekar, S. D., and M. Ramachandran. 2004. "Application of Multi-Criteria Decision Making to Sustainable Energy Planning - A Review." Renewable and Sustainable Energy Reviews 8 (4): 365–81. doi:10.1016/j.rser.2003.12.007.

THE AUTHORS

Danielle Baker | Danielle is a Principal water resource professional with almost 20 years of experience in the water industry. Danielle currently works at DoI Water as the Director of Water Analytics where she leads a team with expertise across engineering hydrology, data management and spatial science to deliver water resource modelling and management for NSW. Danielle is a technical specialist in the areas of water resource management and planning including bayesian modelling, water quality, risk, auditing and compliance. Danielle has delivered many specialist water resource projects across water, wastewater and reuse for utilities and agencies Australia wide and currently serves on the Australian Water Association Catchment Management Specialist Network Committee.

Professor Michael Blumenstein | Professor Blumenstein is currently the Associate Dean Research Strategy and Management and Head of School, Software in the Faculty of Engineering & IT at UTS. He comes from Griffith University in Queensland where he has accumulated over a decade of experience in leadership roles including portfolio Dean (Research) of the Sciences Group and Head of the School of ICT. Professor Blumenstein has been involved in a number of digital transformation research projects at UTS for banks, the water industry and businesses in the area of artificial intelligence and data science. Professor Blumenstein has also been invited to serve on the ARC College of Experts and as a Ministerial advisor to both the LNP and ALP Queensland State Governments in the area of information and communication technology, as well as many other national and international committees.

Prof Jinyan Li | Professor Li is the Bioinformatics Program Leader at the Advanced Analytics Institute and Centre for Health Technologies, Faculty of Engineering and IT, UTS. Prof Li has a Bachelor degree of Science (Applied Mathematics), a Master's degree of Engineering (Computer Engineering) from Hebei University of Technology (China), and a PhD degree (Computer Science) from the University of Melbourne (Australia). He joined UTS in March of 2011 after ten years of research and teaching work in Singapore (Institute for Infocomm Research, Nanyang Technological University, and National University of Singapore). He is an expert in data mining algorithms and new machine learning methods. He has published 90 journal articles and 80 conference papers, of which many are highly cited.

Jon Sayers | Jon has worked for over 20 years informing water planning processes, by providing analysis of demands for water in NSW rivers, both consumptive water demands, and ecological water requirements. Initially working in coastal unregulated rivers in the Hunter, Central Coast and Lower North Coast regions, where urban water demand are particularly significant.

Currently he is now working within the science team in NSW DoI Water, as a Senior Ecohydrologist, focused on the evaluation of ecological effectiveness of flows in major and minor rivers across the state, including working to develop tools that predict the ecological responses to streamflow management options.

Data Analytics and Visualisation

Dr Christobel Ferguson | Dr Ferguson is an Industry Advisory Board Member of the Faculty of Engineering and Information Technology UTS. Dr Ferguson is a graduate of UTS and has had over 30 years' experience leading industry and government departments in the area of water resource management, water quality, microbiology, technology transfer and business innovation. She initiated and led the DoI Water digital transformation in the area of water analytics. This includes UTS partnership with DoI Water in the area of water analytics. Dr Ferguson is a member of the Australian Water Association, a board member of the International Water Association Health Related Water Microbology Group and graduate of the Australian Institute of Company Directors.

Dr Shoshana Fogelman | Dr Fogelman is an Industry Project Manager for the Faculty of Engineering and Information Technology. Her key role at UTS is to drive and build strategic research partnerships with industry, government and community organisations, that will enable them to access skills, world leading expertise of UTS researchers (including facilities) to develop innovative solutions to solve real-world problems. Dr Fogelman has worked with NSW Crown Water & Lands, and other industry partners (such as banks, start-ups, water utilities) to develop successful project outcomes in digital transformation. Dr Fogelman is currently elected to the Australian Water Association (AWA) NSW Branch Committee. She previously set up the AWA National Mentoring Program that was launched in 2006 and co-founded the AWA YWP movement in Queensland. In recognition of her expertise Dr Fogelman has been invited to serve on numerous international and national conferences/committees in the water and technology area.

Roobavannan Mahendran | Mahendran Roobavannan holds a BSc in Civil Engineering (University of Peradeniya, Sri Lanka 2010), M.Eng in Water Resources Engineering (University of Tokyo, Japan 2013). Currently, he is a Ph.D. candidate in Centre for Technology in Water and Wastewater (CTWW), University of Technology Sydney. In his master thesis, He analyzed the future changes of flood risk in global scale using 11 Global Circulation Models(GCM). His master research was published in Nature climate change. Currently, he is studying co-evolution of society and water based on the case study centered Murrumbidgee River basin. This will facilitate the sustainable water management of river basin while society changes their value and preference on water usage. His research interests include (i) data analysis to understand the dynamics, (ii) numerical modelling (iii) hydrology and water resource management, (iv) socio-hydrology, (v) uncertainty analysis and future projection of water stress.

Mr Yi Zhang | Mr Zheng holds a BSc and Master in Computer Science and Technology. He is currently finalising his PhD at UTS, in the area of data analytics. Mr Zheng has a solid IT background and research focuses on data mining, which is to discover hidden useful patterns and rules from different kinds of data (including society network data, workforce data, medical data).

Dr Shameek Gosh | Dr Gosh is a Data Scientist with a Bsc (Computer Science) and is currently finalising his PhD at UTS. He has over 3 years of experience in research and development of machine learning and predictive technologies in the area of data analytics. He also has over 4 years of experience in software and independent test development, leading research projects, code reviews, verification, validation of modules and project/ research documentation. His expertise focus on implementation and design of optimisation in health care, bioinformatics and extraction of symbolic episodes as patterns from time series data and employing machine learning algorithms for forecasting and automated prediction of medical events.