18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Statistical challenges in water use accounting and data interpretation

Quanxi Shao [1], **Albert van Dijk** [2], **Simon Barry** [3] **and Graham Turner** [4]

[1] *CSIRO Mathematical and Information Sciences, Private Pag 5, Wembley, WA 6913, Australia.*
Email: Quanxi.Shao@csiro.au
[2] *CSIRO Land and Water, GPO Box 1666 Canberra ACT 2601, Australia.*
[3] *CSIRO Mathematical and Information Sciences, GPO Box 664 Canberra ACT 2601, Australia.*
[4] *CSIRO Sustainable Ecosystems, GPO Box 284 Canberra ACT 2601 Australia*

Abstract: Water resources accounting was defined by the Water Accounting Development Committee (2007) as "the application of a consistent and structured approach to identifying, measuring, recording and reporting information about water." Discussion of the desired outcomes of water accounting makes it clear that there is focus on water extraction and consumptive use for domestic, economic or public benefits. This aspect of water resources accounting is described here as water use accounting. Water use is not always assessed efficiently or accurately by holistic consideration of the entire water balance. For example, urban water systems may represent a relatively small component of a regional water balance, but an intensively and accurately metered one. In such cases, water use accounting may be a well posed problem that only requires appropriate aggregation and reporting of metering data. In rural systems, extractions of groundwater and river water and the water use associated with runoff intercepting activities are often only partially metered or not metered at all. However, they are often within the uncertainty ranges of common hydrological water balance measurements and models, and therefore cannot always be inferred from those. There is a need for purpose-designed water use metering and estimation strategies that address statistical issues such as sampling design and uncertainty estimation.

In this paper, we discuss the statistical challenges associated with the water use accounting and assessment. Possible methodological developments are discussed for specified water use components including intercepting activities (farm dams in particular), river water extractions and urban water use. Different water use components share common research needs. Uncertainty is the major issue amongst all the components. The key statistical techniques include (1) sparse/missing data problem in the estimation of water use; (2) sampling strategies in determining sample area and its consequences for estimation; (3) model/variable selection in the use of ancillary information (covariates, attributes) for spatial estimation of water use; and (4) uncertainty and sensitivity in regulated river system and urban water system which are also subject to water balance constraint.

The Australian Water Resources Information System (AWRIS) will contain a vast amount of spatial and temporal information. Useful and meaningful interpretation of these data is another important aspect of water accounting and assessment. Statistical techniques will be needed to summarise these data and allow interpretive statements to be made and provide further information gain in the sense of improving the current hydrological modelling practice. Techniques that can assist in interpreting and reporting water information are discussed, including trend analysis, quantile regression, inequality of water use distribution, Empirical Orthogonal Function and Principal Components Analysis techniques for spatial and temporal water data. Issues and recommendations in the current data interpreting techniques are also presented.

Keywords: *Data interpretation, Trend analysis, Water Accounting and Assessment, Water use accounting, Uncertainty.*

1. INTRODUCTION

Water resources accounting was defined by the Water Accounting Development Committee (2007) as "the application of a consistent and structured approach to identifying, measuring, recording and reporting information about water." [1] Water accounting is a relatively recent term in Australian water resources management. Although the Australian Bureau of Statistics developed and released national water accounts since 2000, the term attained sudden prominence with the National Water Initiative (NWI) in June 2004. The NWI can be described as the blueprint for national water reform in Australia. Among several objectives and agreed actions was the need for water resources accounting, which has a stated outcome: "*to ensure that adequate measurement, monitoring and reporting systems are in place in all jurisdictions, to support public and investor confidence in the amount of water being traded, extracted for consumptive use, and recovered and managed for environmental and other public benefit outcomes*." (NWI, 2004)[2]. This makes it clear that there is focus on water extraction and consumptive use for domestic, economic or public benefits. This aspect of water resources accounting is described here as water use accounting.

The latest Water Account (ABS, 2006) presents information on the supply and use of water in the Australian economy, compiled in accordance with the System of Integrated Environmental and Economic Accounting for Water (SEEAW) (UN, 2006). The report releases a statement providing a summary of some of the challenges in water use accounting: "*Calculating water use by industries is not straightforward. Water use can include self-extracted water, distributed water, or reuse water, and sometimes a combination of all three sources are used. Calculating water use estimates for an industry or business is made more complicated when water is also supplied to other users, or when water is used in-stream. As such, simply adding self-extracted water, distributed water, and reuse water to derive a figure for total water use can be misleading. In the Water Account, volumes of water used and supplied by each industry have been balanced to derive 'water consumption'. This figure takes into account the different characteristics of water supply and use of industries and is a way of standardising water use, allowing for comparisons between industries.*" The ABS also notes the development of water accounting required under the NWI and considers the national water accounts to be consistent with the requirements of the NWI.

This paper discusses some statistical challenges associated the water use accounting. Firstly, some water uses (such as farm dams) are only partially metered or not metered at all and are often within the uncertainty of common hydrological water balance measurements and models. Statistics is needed to guide the sampling strategy and uncertainty estimation. Secondly, although each component in a water system (such as urban water system and regulated river system) may be estimated or observed separately, as a system, the final water accounting needs to be considered simultaneously. Thirdly, as the Australian Water Resources Information System will contain a vast amount of spatial and temporal information, statistical techniques are also needed to summarise these data and allow interpretive statements to be made for the purpose of further information gain in the sense of improving the current hydrological modelling practice.

2. WATER USE ACCOUNTING

2.1. Accounting for Intercepting Activities

Clause 55 of the National Water Initiative agreement states that "*The Parties recognise that a number of land use change activities have potential to intercept significant volumes of surface and/or ground water now and in the future. Examples of such activities that are of concern, many of which are currently undertaken without a water access entitlement, include: (i) farm dams and bores; (ii) intercepting and storing of overland flows; and (iii) large-scale plantation forestry*." Other activities may include changes in the area under native vegetation and other forms of land use change or land use practices. All these activities change the amount of water used in a landscape and change the natural hydrological processes for both surface water and groundwater. This leads to the need to account for the water use of these activities.

As an illustration, we discuss the farm dams in detail as they play an important role in Australian agriculture, for stock water supply and irrigation (DLWC, 1999). Dam sizes vary from less than 1ML to over 100ML of capacity, also depending on the purpose of its usage. Farm dams as an intercepting activity in the NWI refers to the category of dams that is established on a hillside and directly captures overland flow to store it for later use. Growth in the number of hillside farm dams (defined here as dams directly intercepting surface runoff) have been recognised as a considerable risk to surface water resources in the Murray-Darling Basin, and

---

[1] http://www.wadc.gov.au/national-water-accounting-project/what-is-water-accounting.php

[2] http://www.nwc.gov.au/nwi/index.cfm#accounting

probably the most significant risk associated with land and water use (as opposed to climate factors; Van Dijk, et al. 2006).

Accounting for dam water use is a difficult task. The number of farm dams and the sizes and spatial locations are often unknown. While the total volume of farm dams in a region is important, the spatial location is also important. Aerial photographs, satellite imagery, and topographic maps are most commonly used to estimate the number and location of farm dams. Figure 1 is an example of farm dam mapping by CSIRO (G. Byrne, pers. Comm.) using high resolution remote sensing data (e.g. SPOT) and automated mage recognition technology. Geosciences Australia is currently producing maps of farm dam location (and for large farm dams also the maximum extent) for part of Eastern Australia. However, this mapping is currently done manually, and to repeat it annually is currently cost prohibitive. In the short to medium term there is a need for statistical methods to upscale sample mapping. A further complication is that the area of dams is only a moderately accurate estimator of their storage capacity and primary use (which are important factors in their water losses).

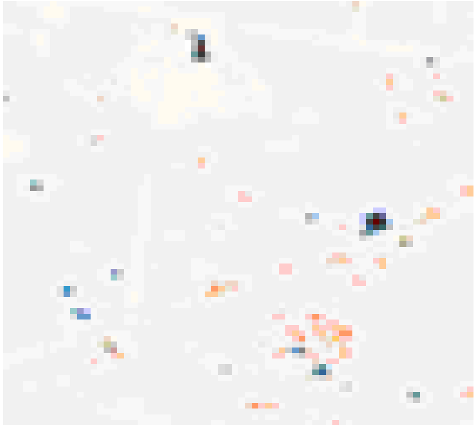

Figure 1. The preliminary results (Right) of automated farm dam mapping from SPOT satellite scene using object-oriented image recognition software. Black: mapped water bodies, blue: areas associated with water bodies (i.e. bare soil), red: bare soil areas classified as not associated with farm dams.

There are some challenges in practice as the dams are not evenly distributed and the sizes vary according to their purpose and geographic environment. To provide a better estimation of the sizes and locations, geographic information and land use practice can be considered in modelling. For example, AGRECON (2005) assessed the farm dams in 27 different landscape classes. A general statistical model is given as

$$y = f(X, \theta) + \varepsilon ,$$

where y is the variable of interest (eg. dam size), X a collection of potential variables affecting y (eg. land use, rainfall pattern etc.), θ the model parameter and ε the random error (which can be spatially correlated).

Two questions must be answered to use this model: the form of functional relationship f and the variables X which are important for determining y. Non-/semi-parametric statistics and variable/model selection offer solutions here. The results can be used to estimate the number of dams and sizes for a given area. Note that the variable X can include continuous, categorical and ordered components. A statistical method is needed to handle all these data types. The model can guide sample strategy and design and provide the estimation in the whole region of interest.

- Challenge 1. Develop a methodology to upscale localised high resolution mapping of farm dams on selected areas to large areas. This will need to consider issues such as sampling strategy and design, interpolation and inference using ancillary data.

Bores as a runoff intercepting activity in the NWI refer to the extraction of groundwater that would otherwise have discharged into a river system, as opposed to the extraction of groundwater from large aquifers that usually occur in alluvial lowlands. Situations were groundwater discharges significant water volumes into the stream are usually (though not necessarily always) found in areas with more relief where bore water is used for domestic purposes, stock drenching or small-scale irrigation (e.g. fodder crops, vineyards or orchards). Bore licenses are required and usually are tied to one of these purposes. Hydrological modelling methods cannot normally assist in estimating water use in these cases, as the availability of hydrometric data (groundwater level, streamflow) is insufficient and the magnitude of water extraction often too small in any case at the scale of measurement. Only irrigation bores (if any) are metered and therefore there is a statistical sampling problem similar to that associated with estimating hillside farm dam interception, requiring knowledge of the number and type of bores found in a given area and estimating the overall extraction from those.

- Challenge 2. Develop a methodology to estimate bore water use from sparsely metered bores. This includes spatial mapping of bore density with consideration of the purpose of bore usage, relationship of volume and purpose.

Similar challenges are faced in other intercepting activities.

2.2. Integrated Accounting for Regulated River Systems

If all components for water accounting in a regulated water system are estimated for individual river reaches, a water account for the whole river system can be developed. There are two types of data: observed data and modelled data. The observed data have measurement uncertainty, while the modelled data have uncertainty due to model uncertainty as well as the measurement uncertainty in the observations used as input data for that model. The water balance accounting reported as part of the Murray Darling asin Sustainable Yields (MDBSY) (Kirby et al., 2008) is taken as a reflection of the current state of river water balance accounting. Van Dijk et al. (2008) discuss different uncertainties in river modelling across the Murray-Darling Basin.

A promising avenue of research is to use the linear constraints that are explicitly defined in the water balance equations to further constrain and infer the values of the underlying population quantities. Markov Chain Monte Carlo approach is a way to deal with the constraint problem. To illustrate this approach consider the following situation: we have variables x_1, x_2, x_3 and x_4 with constraint $x_1+x_2+x_3=x_4$. For each of these variables we have a prior distribution expressing our beliefs about the value of the quantity based on the prior beliefs and the observed data and model used to construct it. Assume that all the variables follow the standard normal distribution. The correlation can be clearly seen from the simulation (Figure 2).

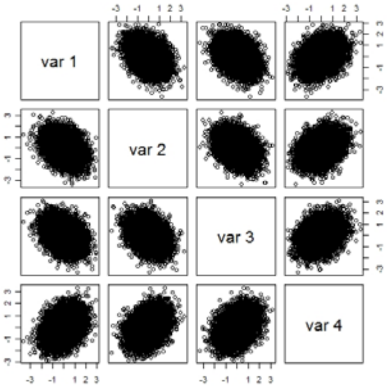

Figure 2. scatter plot of variables with constraint $\mathbf{x_1+x_2+x_3=x_4}$.

In practice, it is possible that there are more than one constraints and/or that one term involves more than one constraint. While there are some unmeasured or unknown terms in the constraints, more efforts need to be made on uncertainty reduction, for example by including historical or spatial knowledge and information.

- Challenge 3. Develop a methodology to cope with constraints in integrated water accounting for regulated river systems.

2.3. Accounting for System-wide Water Use and Availability

One of the applications of the water use account is to improve the understanding of the complete water system in term of strategic decision, and this needs to consider accounting for water use and availability in a broader system than regulated river systems only. The system can include all the water components in this paper. Even further, the system can include rainfall and evaporation to describe a complete water cycle.

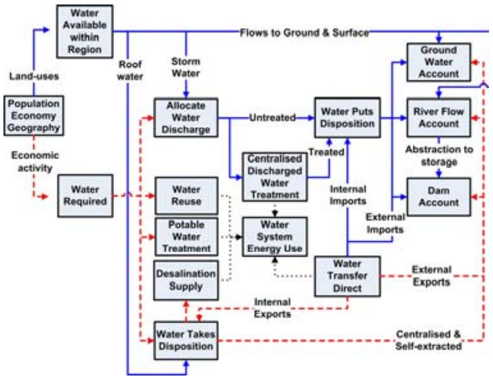

Figure 3. Structure of the system-wide WAS. Arrows show connections of data flows; blue represent data on water availability, red represent data on water requirements, black represent data on energy requirements of the water system.

A water accounting system (WAS) has been developed in CSIRO as an innovative new tool for strategic long-term water management (Turner et al., 2007 and Turner et al., 2009). The WAS incorporates both disaggregated water use and availability (i.e., storage stocks and river and groundwater flows), provides a comprehensive and consistent historical database, and can integrate climate and hydrological model outputs for historical accounts and the exploration of scenarios. The WAS is implemented using stocks and flows dynamics, in whatIf software[3] and can be used as a stand-alone facility, but has also been designed to be used in combination with other stocks and flows frameworks that provide data on key drivers such as demography, land-use and electricity generation. A key advantage of the WAS follows from the system-wide scope, which covers both use and availability of water, as well as the relationship with economic activity, if desired.

The assembly of a water use account in WAS faces the considerable issue of dealing with sparse data, at least until suitable metering and monitoring activities are well established. Previous calibration of the WAS (and other Stocks and Flows Frameworks) has relied on the substantial information that is embodied in the multiple relationships of a system-wide account, in order to impute a variety of unknowns given the poor state of historical data on the water system. Earlier exploration of other techniques for estimating unknown variables (e.g., Cross-Entropy and Kalman Filters) identified that these techniques were either not suitable or that further resources or data would be required to achieve worthwhile benefits. Statistical challenges discussed before are also applicable here. Further challenges are the sensitivity analysis to guide further investment.

- Challenge 4. Develop sensitivity and uncertainty analysis to assess which parts of the water accounts are particularly uncertain and important, and may be priority areas for further investment in metering.

3. INTERPRETING WATER DATA

After data are collected and collated, it is important to extract and report information in meaningful and useful ways. This is particularly true if the volume of data is large (and this is the case in water accounts). Different reports are required for different users. For example, to understand the current status of water accounting at national and state levels, the traditional tabulated balance sheet reports (as is used by ABS' water accounting reports) would be sufficient. However, for water resources planning and policy management, the trend of change in both spatial and time domains would be much useful. Researchers too need information to help assess potential driving forces to guide their investigations.

With this in the mind, below we address a number of issues that should be addressed in reporting water resources assessment and accounting information. It is assumed here that the basic information collected is from nested individual spatial units (catchments, regions, basins, drainage divisions, etc.) and at different time scales (daily, monthly, annually, etc).

3.1. Objectives to be Reported

Obviously one needs to know what is to be reported. There are many statistics available. The following statistics could be useful for different purposes and at different levels of reporting.

Mean and Total: Basic statistics and used frequently in official statistics.

Uncertainty. Standard Deviation is frequently used to assess the uncertainty of individual entries. It is not a straightforward task to assess the uncertainty here as there are some constraints associated such as water balance equations. Other quantities for uncertainty are quartile, range of the data.

Median and Quantile. If a distribution is skewed (not symmetrically distributed), the median may be a useful or even better statistic to be reported. A more general statistic is the quantile. For example, it might be important to see the impact of land use change on various percentiles of flow duration curve.

Inequality Measurement of Water Use. As is in the social statistics for income distribution in population, how the water distributed in the society may also be important (e.g. farmland water use). The Gini index or Gini Coefficient (Gini, 1921) is used in social statistics as a measure of inequality of income distribution or inequality of wealth distribution. A similar index could be useful for both management and scientific research.

Trend Analysis and Spatial Pattern Analysis. All the above statistics are static reporting tools. As we are living in a changing world and would like to understand the change and its driving forces, for planning and scientific research purpose, more dynamic reporting techniques are very important. Trend analysis and spatial pattern analysis can help to address this, and we discuss these in the next sections.

---

[3] http://www.whatiftechnologies.com

3.2. Trend Analysis

Trend analysis is not only a valuable reporting tool in scientific community but also provides insight information for resources planning and investment and policy management. ABS has started to report simple trend plots in its water accounting report.

When the time series is irregularly sampled or subject to large variation, summary statistics (such as the rate of change and change points) can provide a better summary statistics than the time series plot itself. Once the objective (e.g. mean and median), is determined, the selection of trend analysis tools will be important. There are several questions that should be addressed in trend analysis.

Before conducting any trend analysis, one must be clear about what the type of trend anticipated. A trend can be short term (intra-annual variation), medium term (iter-annual) or long term (eg., climate change). In general, one may be interested in slow structural change, secular variation, drift, tendency, evolution. Climatologists are interested in very long-term trends (eg., more than 100 years) while meteorologists are keen to understand relatively short-term trend. Trend can further represent smooth change or jump, direction or variation. Shao and Campbell (2002) defined three types of change points (see Figure 4): break point where the general trend can change freely, knot point where the periodic trend changes, and join point where the general trend must change continuously. These three types of trends can be combined to form different types of change point.

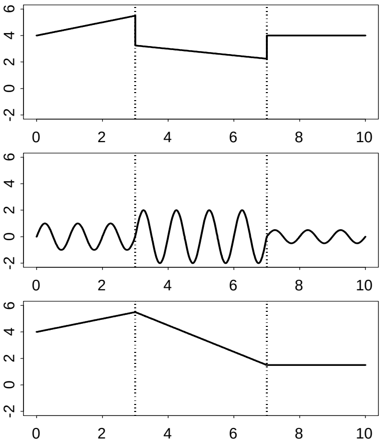

Figure 4. An illustration of different types of trend. Top: break point; Middle: knot point and Bottom: Join point.

The Handbook of Hydrology has a whole chapter about trend analysis and related issues (Salas, 1993). The CRC for Catchment Hydrology developed a software to conduct trend analysis which is available through the Catchment Modelling Toolkit[4]. Two independent CSIRO reviews were conducted from different perspectives (Shao and Li, 2008 and Henderson and Morton, 2008). The analysis tools used in literature can be classified into three groups: (1) rank-based approaches (Kendall's τ and Mann-Kendall, Spearson's ρ, Sequential Mann-Kendall, Wilcoxon-Mann-Whitney test, Pettitt test and CUMSUM), (2) parametric regression approaches (Linear regression and correlation coefficient, Segmented linear regression, Periodic regression) and (3) non-parametric regressions (Penalised spline, Regression spline). Unfortunately, the current practice in trend analysis does not typically take account of data dependence and seasonality, which are crucial for the use of these statistical tools.

- Challenge 5. Develop appropriate and useful trend analysis techniques with consideration of data dependence and seasonality as well as different types of change points.

Environmental changes are not always characterized by averages only. For example, climate change may cause only moderate or no change in annual mean river flow but more frequent flood and drought events. The impact of afforestation and deforestation on streamflow can be different for different quantiles of streamflow, depending on climate patterns and catchment characteristics. New techniques are required to take account of not only the mean but also various percentiles to characterise changes in extreme events.

- Challenge 6. Develop appropriate and useful objectives for trend analysis techniques which allow more flexibility in describing the trend function.

It is widely known that many hydroclimatic variables are closely related. One might understand that simultaneous analysis can potentially reduce the uncertainty causing by random errors from sampling. From our point of view, the multivariate trend analysis will be a next priority in hydroclimatic research. Any significant statistical research should generate serious impact and will be productive research area.

- Challenge 7. Develop multivariate trend analysis techniques which allow simultaneous consideration of correlated variables.

Mapping is frequently used to present spatial data. For a large dataset, the Empirical Orthogonal Functions (EOF) technique aims to find a new set of variables that capture most of the observed variance from the data

---

[4] http://www.toolkit.net.au/cgi-bin/WebObjects/toolkit.woa/wa/productDetails?productID=1000022

through a linear combination of the original variables and has been widely used in climatology and hydrology for various variables. However, it still remains a challenge to summarize data in both space and time domain in a manner that is easily interpreted.

- Challenge 8. Investigate better statistical visualization tools to summarize data in both space and time domains.

Further more, the purpose of trend analysis should be addressed. It is noted that the trend analysis is often abused in literature as a short-cut to answer in-depth scientific questions. For instance, the impact of land use cover change on streamflow is frequently investigated by comparing the trends of streamflow and the trend of land use. Time series analysis and general regression can deal with these issues more directly.

- Challenge 9. Investigate trend analysis techniques which allow simultaneous consideration of potential variables affecting the trend.

4. DISCUSSION AND CONCLUSIONS

In this paper we discuss some statistical challenges in water use accounting and data interpretation. Although many aspects in water use accounting are not discussed here, it is clear that the statistical challenges are not trivial and involve many active research activities in statistical science, such as model/variable selection, sparse/incomplete/missing data problem, data dependence, trend analysis and data visualization.

ACKNEWLEDGEMENT. This research is supported by the Water Information Research and Development Alliance (WIRADA) between the Australian Bureau of Meteorology (BoM) Water Division and the CSIRO Water for a Healthy Country Flagship Program.

REFERENCES

ABS (2006), Water Account Australia 2004-05. Catalogue Number 4610.0. Australia Bureau of Statistics, Canberra, Australia.

AGRECON (2005), MDBC Hillside Farm Dams Investigation. Murray-Darling Basin Commission Project 04/4677DO.

Department of Land and Water Conservation (DLWC) (1999), Farm Dam Assessment Guide. NSW DLWC Publication HO/13/999, NSW Department of Land and Water Conservation, Sydney.

Gini, C. (1921), Measurement of Inequality of Incomes. *The Economic Journal* 31, 124–126.

Henderson, B. and Morton, R. (2008), Strategies for the Estimation and Detection of Trends in Water Quality. CMIS Technical Report 08/70, CSIRO, Canberra, Australia.

Kirby M., Van Dijk A.I.J.M., Mainuddin, M., Peña-Arancibia, J., Guerschman, J.-P., Liu, Y., Marvanek, S., McJannet, D.L., Paydar, Z., McVicar, T.R., Van Niel, T.G. and Li, L.T. (2008), River Water Balance Accounts Across the MDB, 1990-2006. A report to the Australian Government from the CCSIRO Murray-Darling Basin Sustainable Yields project, CSIRO Australia, Canberra. 28pp.

Salas, J.D. (1993), Analysis and modeling of hydrologic time series, in *Handbook of Hydrology* by Maidment DR (ed). McGraw-Hill: New York. Chapter 19.

Shao, Q., Campbell, N.A. (2002), Modeling trends in groundwater levels. A*ustralian & New Zealand Journal of Statistics*, 44, 129-141.

Shao, Q. and Li, M. (2008), Statistical Review of Trend Analysis: Issues and Recommendations. CMIS Technical Report 08/36, CSIRO, Perth, Australia.

Turner, G.M., Baynes, T., McInnis, B. (2009), A Water Accounting System for Strategic Water Management. *Water Resources Research* (submitted).

Turner, G.M., Baynes, T.M., McInnis, B.C., West, J. and Hoffman, M. (2007), Water Accounting System for Strategic Resource Management. In: D. Kulasiri (Editor), MODSIM 2007 - International Congress on Modelling and Simulation; Land, Water & Environmental Management: Integrated Systems from Sustainability., Christchurch, NZ.

UN (2006), System of Environmental-Economic Accounting for Water. United Nations, New York.

Van Dijk, A., Evans, R., Hairsine, P., Khan, S., Nathan, R., Paydar, Z., Viney, N. and Zhang, L. (2006), Risks to the Shared Water Resources of the Murray-Darling Basin. Prepared for the Murray-Darling Basin Commission, MDBC Publication 22/06.

Van Dijk A, Kirby M, Paydar Z, Podger G, Mainuddin M, Marvanek S, Arancibia JP (2008) Uncertainty in River Modelling across the Murray-Darling Basin. A report to the Australian Government from the CCSIRO Murray-Darling Basin Sustainable Yields project, CSIRO Australia, Canberra. 93pp.