

Using scientific workflows to calibrate an Australian land surface model (AWRA-L)

J. Vleeshouwer^a, J.M. Perraud^a, D. Collins^b, G. Warren^a, S. Gallant^a and R.J. Bridgart^a

^a CSIRO Land and Water

^b CSIRO Advanced Scientific Computing

Email: jamie.vleeshouwer@csiro.au

Abstract: The AWRA-L landscape hydrology model is one of three model components which form the Australian Water Resources Assessment (AWRA) system that aims to produce interpretable water balance estimates for Australia which (as much as possible) should agree with observations such as point gauging data and satellite observations. It is a 5 km grid-scale model, driven by interpolated climate inputs, that produces continental surfaces representing water stores and fluxes across the landscape, as well as energy and vegetation dynamics at daily timestep. The system is jointly developed by CSIRO and the Bureau of Meteorology (BoM) with model improvements proposed regularly, requiring re-calibration. This regular execution makes scientific workflows an attractive solution for engaging in our large computation and data transformation tasks whilst also providing HPC scheduling, repeatability, traceability and monitoring.

Initially, various software tooling was explored to assist with the calibration of AWRA-L. These included the Catchment Water Yield Water Estimation Tool (CWYET), and various third party products such as PEST, UCODE and PGO. These tooling products were not quite suitable for use with calibrating AWRA-L largely due to the complex calibration requirements (such as the formulation of the AWRA-L objective function) and the lacking flexibility to evolve these products for future AWRA-L calibration requirements. Instead, two prototype systems were developed in Matlab: one with a targeted execution environment on a desktop; and the other with a targeted execution environment on a cluster. From these two prototypes a toolset was then developed based upon the Metaheuristics (Perraud et al, 2012) Application Programming Interface (API), which is a loosely coupled framework for assisting with model optimisation. Scientific workflow software was then used to orchestrate the flow of information between the model calibration processes.

This paper describes the process undertaken to calibrate proposed changes to AWRA-L, design objectives, and current state of the suite of workflow activities developed known as the ‘AWRA Calibration Tools’. Trident workbench was used to develop automated workflows for processing including: parameter optimisation; model simulation; and benchmarking of results. These tools allow domain experts to execute and share workflows, enabling AWRA-L to be steered in a direction that either improves the model’s predictability or is considered to have an improved physical representation without significant degradation in predictions. The paper concludes by identifying key challenges that have emerged, and suggests some improvements for the future.

Keywords: *Trident Workbench, Optimisation, Workflows, AWRA, High Performance Computing*

1. INTRODUCTION

1.1. Project Background

During the mid 1990's the effect of a severe drought throughout extensive parts of Australia caused economic and environmental hardship. A lack of managed water information was recognised as a key problem. Through the Water Act 2007, the Bureau of Meteorology (BoM) was given a statutory responsibility to provide regular reports on the status and usage of Australia's water resources. To meet this mandate the BoM produces two nationally significant water information products: the National Water Account (NWA); and the Australian Water Resources Assessments (AWRA). Underpinning these products is a complex continental scale hydrologic workflow that brings together observational information and models to answer questions on the use trends and status of Australia's water resources (Stenson *et al.*, 2012). The workflow is orchestrated using Delft-FEWS (Gijssbers *et al.*, 2008) and is comprised of: river models (AWRA-R), a groundwater model (AWRA-G) and the landscape water balance model (AWRA-L). AWRA is a world class initiative being developed under the Water Information Research and Development Alliance (WIRADA): a collaboration between the BoM and the CSIRO Water for Healthy Country Flagship, and will for the first time allow consistent reporting at a national level.

1.2. Translating Research to Operations

Throughout the AWRA development life cycle, improvements to model components are made regularly to improve the accuracy of predictions behind the modelled outputs used in preparing the above mentioned reports. These improvements have two aims which are to: improve the models predictability; or improve the physical representation of a model without significant degradation in predictions. For example, maximising streamflow predictability metrics such as Nash Sutcliffe Efficiency (NSE) is of major importance for water accounting, however if an improvement to Leaf Area Index (LAI) correlation can be achieved with only minor degradation of streamflow performance, the change may be considered for acceptance (Figure 1). These improvements are guided by the AWRA Modelling Group (AMG), with results presented regularly to the AWRA Technical Steering Group (ATSG) for approval. The AMG is a CSIRO-based group and guides ongoing system development and considers issues relating to model development, testing and benchmarking. The ATSG is formed of CSIRO and BoM representatives and recommends decisions on the versions release of model components and data streams.

The process of change management for the AWRA-L model is as follows:

1. a 'change' is proposed through the AMG, and is implemented.
2. a calibration experiment takes place which includes the proposed modifications.
3. results of an experiment are benchmarked whereby the modelled outputs are compared to on-ground or satellite observations.

Calibration experiments typically have their changes classified into the following three categories:

- The model (i.e. alternative algorithms for modelling processes)
- The datasets (i.e. updated observed data, catchment boundaries refined, alternative sources)
- The optimisation configuration (i.e. alternative objective function, adjusting parameter constraints)

A new experiment is always configured based upon a previous experiment for easy comparison. To ensure optimal objective function scores are being achieved throughout calibration five replicates are typically run in parallel with a different random number generator seed for the optimiser. Typically three generations of parameter set seeding for each replicate are run (reseeding). In this reseed process the resulting parameters of the initial calibration experiment are used as a direct point in the initial population of a second experiment, and so on. This helps to further improve upon the score of the objective function from the initial experiment.

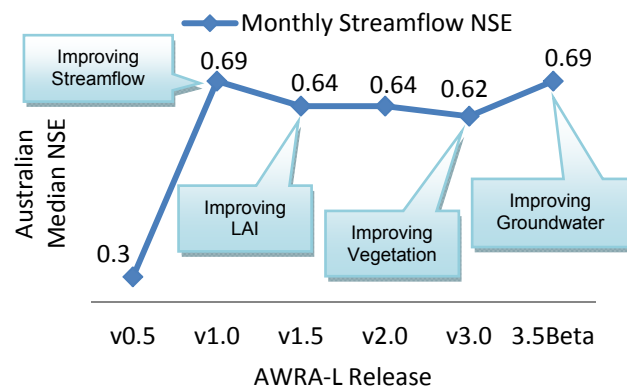


Figure 1: Predictability of streamflow may be traded off to improve other conceptual components of the model

1.3. Choice of using Scientific Workflows

This calibration process takes place in an ad hoc and experimental manner, where a user makes required adjustments to prepare an experiment, results are analysed, and decisions are made by domain experts whether to continue with a proposed change. These qualities are different from an operational workflow which could be scheduled to automatically process information at regular intervals such as Pan-STARRS or NEPTUNE (Barga et al, 2008), however there are characteristics which make using Scientific Workflow Software an attractive option for capturing this process. These include: the ability to automate repetitive tasks; auditability of all data sources used; repeatability of experiments when errors may be discovered; and the provenance and sharing capabilities enabling the sharing of workflows amongst project teams. Figure 2 shows the conceptual components and flow of information required for end to end processing of a calibration experiment. This entire process is what we aim to capture using scientific workflows.

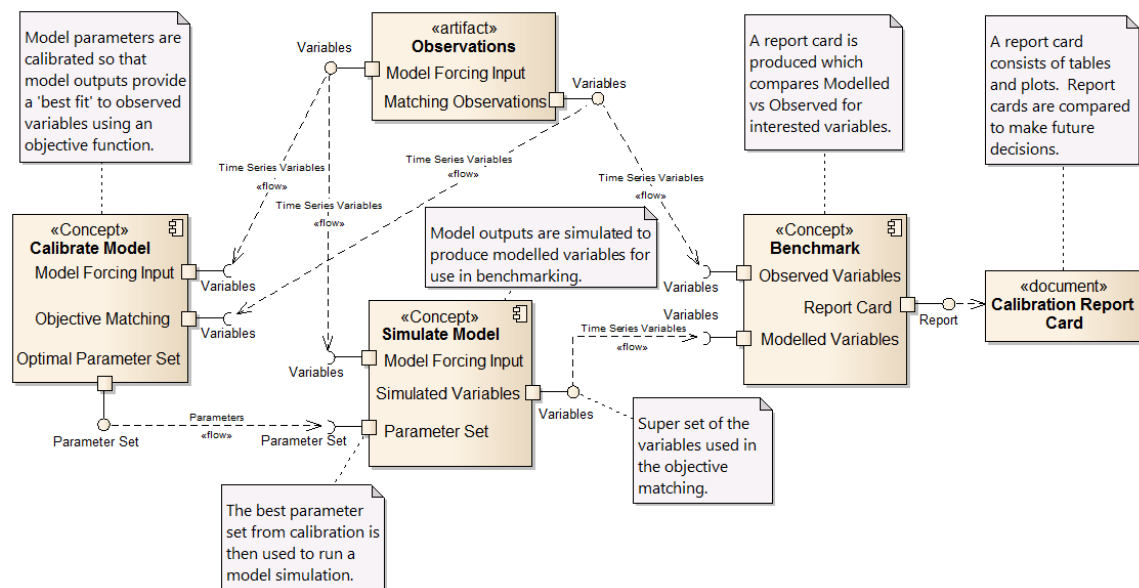


Figure 2: UML Component diagram describing the processes required to undertake a calibration experiment

1.4. AWRA-L calibration requirements

Whilst various techniques have been explored and used for calibration of the AWRA model components, throughout this paper the focus on the calibration of AWRA-L. Initially two developmental systems were prototyped in Matlab using the built in GA optimiser and also a custom implementation of Shuffled Complex Evolution (SCE) (Duan et al, 1993). These prototype systems facilitated the identification of the functional requirements needed to meet the AWRA use cases as follows:

- Calibrate parameters of different model components (including experimental deviations of models)
- Force alternate input and observation data (different sources, catchment sets, simulation period)
- Objective function capable of handling multiple observed data types (streamflow, LAI, etc)
- Ability to modify the continental objective function
- Specify the parameters to be optimised (including adjustable search boundaries)
- Adjust optimisation strategies (such as optimiser configuration, parameterisation)
- Ability to access high performance computing resources

The costs associated with licensing the required Matlab toolboxes prevented us from scaling these prototypes up to meet future requirements. In particular, there was an anticipated desire to eventually research more complex optimisation strategies such as: optimising the AWRA-L parameters at pixel scale rather than catchment scale; and also having the ability to explore the model performance of coupled landscape, river, and groundwater model components. Hence there was a need to migrate to an alternative toolset.

2. SYSTEM OVERVIEW

2.1. Model calibration and validation schemes

A selection of 610 unregulated catchments across Australia is being used in the development of AWRA-L (Zhang et al, 2011). From this dataset half of the catchments are used in calibration while the remaining half is used for cross-validation. The objective function used for calibration, globally optimises the model performance across all 305 catchments resulting in a single parameter set which is then applied to all of Australia (Viney et al, 2009). Observed datasets used within the objective function include both gauged streamflow and MODIS 8-day LAI. These are weighted using a combination of statistics such as daily NSE, monthly NSE, and correlation coefficient. Experiments are then benchmarked against the remaining 305 catchments, and experiments which are deemed to be successful are transferred to the AWRA-L operational system (Figure 3) whereby each grid cell (pixel) is parameterised with the best parameter set.

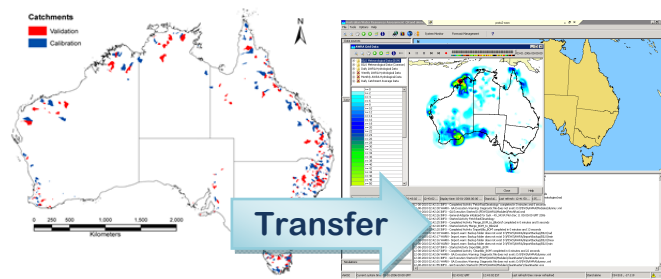


Figure 3: Successful calibration experiments are transferred into the AWRA operation system at six-monthly intervals

2.2. Architecture

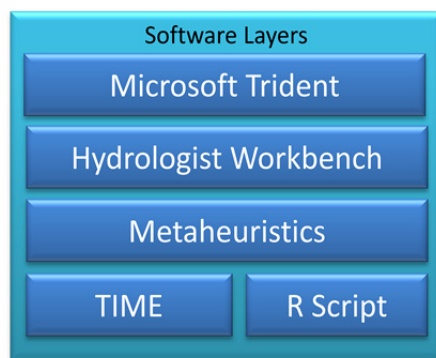


Figure 4: High level overview of system components used within the AWRA calibration tools

The high-level architecture of the AWRA-L Calibration system is shown in Figure 4. Designed upon Microsoft Trident (Barga et al, 2008) – which is based on the Windows Workflow Foundation – the system provides the orchestration framework to execute a calibration experiment. It includes graphical tools for creating, running, managing, and sharing of workflows. The system also makes use of the Hydrologists Workbench (HWB) which is a domain-specific workflow activity library to extend the functionality of Trident Workbench (Cuddy and Fitch, 2010). The particular components of HWB which are utilised are the: Metaheuristics component for configuring parameter optimisation problems for models; and also the task execution components to aid with scheduling calibration experiments on Windows HPC Server clusters. A custom activity library has also been developed for processes specific to AWRA-L such as configuring the model states and spatial parameters for each catchment.

The calibration process is captured in an external executable that is invoked by the workflow. Having this process external to the workflows allows experimental model builds to be managed on a separate file server, thus avoiding conflicts with multiple versions of binaries in the Trident Registry. The external application is built upon the Metaheuristics optimisation framework (Perraud et al, 2012), The Invisible Modelling Environment (TIME) (Rahman et al, 2003), and R. The Metaheuristics framework contains an implementation of Shuffled Complex Evolution (SCE) which is the chosen search algorithm for optimising the model parameters. TIME provides the model simulation functionality such as data I/O handling, and the temporal stepping of the model. R's statistical capabilities are also harnessed in the definition of the continental objective function. R is used as it provides a method for the user to easily modify the objective function via a script rather than compiling one in the C#.NET source code.

2.3. The computational problem

Within the AWRA Calibration Tools two different modelling techniques are available: gridded calibration; and lumped calibration. A gridded calibration experiment replicates the execution of the operational model, while a lumped calibration experiment can be used to reduce computational overheads and still make effective decisions.

Gridded Calibration: Model input data is at cell scale, simulations take place at cell scale, outputs are aggregated to catchment scale, and then a continental objective function is calculated.

Lumped Calibration: Model input data is aggregated to catchment scale as a pre-processing step, so simulations only occur at catchment scale, and then a continental objective function is calculated.

These different calibration techniques are provided to the user through two separate executables. Whilst both executables share the same frameworks, and hence have the same dependencies, the main difference between them is the task parallelisation techniques needed to efficiently compute each problem. The lumped calibration executable parallelisation occurs by running a model simulation for each catchment in its own Message Passing Interface (MPI) process, which has the ability to share data to processes running on other physical computers. Statistics are then calculated on the outputs and passed back to a master MPI process to compute the objective function whereby SCE can perform its directed search continentally (Perraud et al, 2012). This simple approach works well when using lumped catchment data, since the computational requirements are symmetrical (one model instance per catchment). When performing a gridded forcing calibration experiment the workloads of the optimisation problem become extremely asymmetrical. Catchment sizes vary from 1 cell to 264 cells, meaning the work required to calculate statistics (such as NSE) varies for each catchment. For these reasons the gridded calibration executable decouples catchments from worker processes. Cells, corresponding to a model instance, are distributed evenly across the available worker processes, with some workers having the additional role of calculating the final results for a complete catchment. The catchment results are then passed to the master process where optimisation is performed in the same manner as lumped calibration.

The AWRA-L conceptual model does not exhibit lateral fluxes between neighbouring cells, hence parallelising the problem across the spatial domain presents the best opportunity to parallelise units of work for both lumped and gridded modelling techniques. Table 1 provides an indication of the computational differences between performing a lumped and gridded calibration experiment.

	<i>Gridded</i>	<i>Lumped</i>
Model Instances	9600	305
CPU's	464	64
Memory Footprint	120 GB	25 GB
Input Data Volume	18 GB	1 GB
Runtime	12 Hours	4 Hours

Table 1: Comparison of computational requirements for lumped and gridded forcing data

2.4. Scheduling

With regards to the data pipelines used in the workflow activities, a model is used which: defines the model (the system); then defines the optimisation problem; and finally performs the optimisation process. This broad granularity allows the first two steps (configuration) to be easily decoupled from the final step (execution), enabling the tasks to be scheduled across distributed computing environments, such as on a HPC Server scheduler. This style of orchestration also allows for a clear visual representation of the workflow operations to the user as opposed to other possible workflow styles (Perraud et al, 2009a). A conceptualisation of this model is indicated in Figure 5.

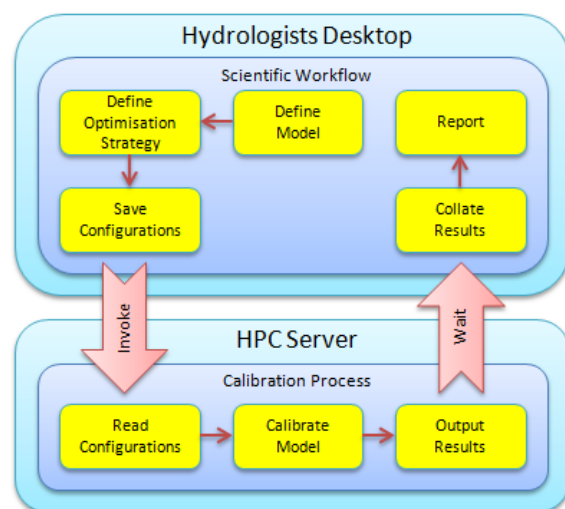


Figure 5: Overview of job scheduling pattern used by the AWRA Calibration Tools

3. WORKFLOW EXAMPLES

Ideally, a scientific workflow which performs the complete end-to-end processing of an AWRA-L calibration experiment is desired. Whilst complete end-to-end processing ('one-click') has not been achieved thus far with the AWRA calibration tools, various components of the process have been automated ('several-clicks'). There are two key limitations of a one-click approach here: (i) domain constraints of job scheduling on shared computing facilities, and (ii) the timing of dataset preparation usually preventing executing the entire process at a single point in time.

Two of the AWRA-L components that have been automated as Trident workflows are model calibration (Section 3.1) and model simulation (Section 3.2). The toolset used for performing the benchmarking of experiments against observed datasets is discussed in Section 3.3.

3.1. Model Calibration

The model calibration workflow globally optimises AWRA-L parameters to yield a single set of parameters that can be applied to all catchments in Australia. It has high level control flow to iterate an experiment configuration multiple times, whereby resulting parameters are used in the initial population of a subsequent iteration. The Seeded Experiment is a workflow which executes the optimisation process for a single iteration, whilst the For Loop and Reseed Controller are used to adjust the input and output paths on this workflow to provide the iterative capability for the reseeding.

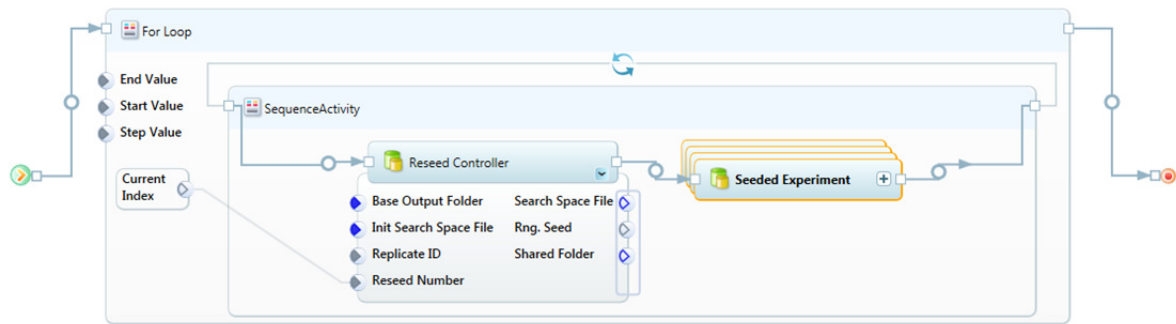


Figure 6: Iterative calibration workflow

Within the Seeded Experiment workflow, the AWRA-L model is configured, and initial catchment conditions averaged from continental grids. A shuffled complex evolution optimiser is configured, along with the catchment objectives, and also an Australian wide objective function. The design of the activities supports a high level of reuse, with only three activities requiring replacement to swap between gridded and lumped calibration techniques.

3.2. Model Simulation

The model simulation workflow is used to produce time series outputs of AWRA-L variables so they can be benchmarked against on-ground or satellite observations. It is used to execute a simulation using the optimal parameter set found during calibration. It contains common activities to the calibration workflow, however can be configured to record a different set of outputs, simulation period, and spatial location to what was used in calibration. This provides the ability to assess the performance of an AWRA-L parameter set using alternative observed phenomenon for benchmarking purposes.

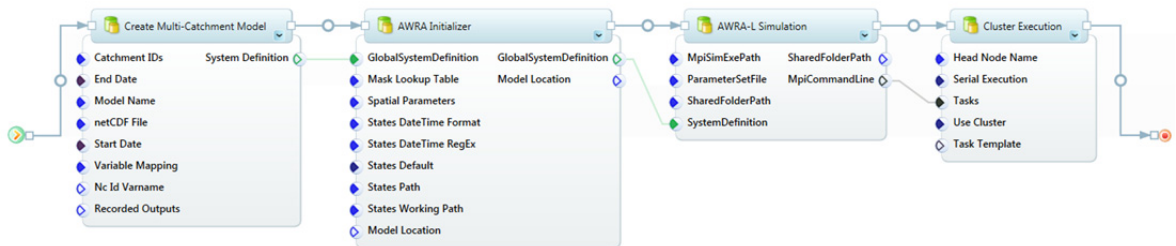


Figure 7: AWRA-L simulation workflow used to provide modelled outputs for use in benchmarking

3.3. Benchmarking modelled results against observations

The benchmarking component enables the evaluation of model components against on-ground or satellite observations, and in doing so, guides future developments and define model caveats. It has been developed as an R based software package (Model-Benchmark) that provides data manipulation routines, statistical measures and report creation features, facilitating rapid, standardised, and repeatable model data benchmark analysis. Within the AWRA development process, Model-Benchmark has been used to generate report cards which evaluate modelled variables versus on ground or satellite observations. These report cards are collated into a PDF file which summarises one variable per page displaying plots, figures and tables reporting on a variety of statistical metrics.

4. DISCUSSION AND CONCLUSION

While the fundamental aspect in this paper has been on developing a core set of workflows to aid with calibrating AWRA-L, there are also pre-processing tasks, as well as post-processing tasks which could provide benefits if they were captured in a scientific workflow.

Pre-processing tasks could include processing the climate input datasets for the catchments used for calibration and validation of AWRA-L. This data set regularly gets re-processed when catchment boundaries may have been modified, or the simulation time period gets extended as more data becomes available. Whilst this may seem rather trivial, it requires large volumes of data to be manipulated and is often protracted, tying up resources both within CSIRO and the BoM. An automated approach using scientific workflows could allow this process to be consistent, more efficient and also minimise the risk of errors.

Post-processing tasks which could also be incorporated into the workflows include tooling to provide reports or diagnostics on how well a particular calibration experiment performed. This could be achieved with sensitivity analysis tooling to identify the most and least influential parameters, potentially allowing the dimensionality of the problem to be reduced. It could also lead to better decisions being made around the optimisation strategies to use so that modelled outputs have improved correlation with observed datasets.

The AWRA calibration tools have been developed to support regular calibration experiments of AWRA-L to further improve the performance of model predictions used in the BoM's reports. Workflows have been developed using the Trident Workbench scientific workflow system to automate calibration, simulation and benchmarking tasks. These workflows integrate a variety of frameworks including HWB, Metaheuristics, TIME, Model-Benchmark, and R statistics. The workflows have been used by hydrologists within CSIRO to further develop the AWRA-L model component for each of the major AWRA releases. The AWRA calibration tools have also been deployed within the BoM so experiments can be undertaken outside this project. These tools will assist the BoM updating the AWRA operational system which is used to support the publication of reports on Australia's water resources highlighting changes in water availability, use, movement and storage at local to national scales and over time scales of months to decades.

ACKNOWLEDGEMENTS

This work was funded through the WIRADA, which is a joint collaboration between the CSIRO Water for a Healthy Country Flagship and of the Bureau of Meteorology Water Division Reporting and Assessment branch. The authors are indebted to senior management in CSIRO and the BoM for the strong and ongoing support of AWRA development. We also thank Heron Brink for his efforts in the development of the AWRA Calibration Tools.

REFERENCES

- Barga, R., Jackson, J., Araujo, N., Guo, D., Gautam, N., and Simmhan Y. (2009) The Trident Scientific Workflow Workbench. *Fourth IEEE International Conference on eScience*
- Cuddy, S.M., and Fitch, P. (2010). Hydrologists Workbench – a hydrological domain workflow toolkit. *2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada*
- Duan, Q.Y., Gupta, V.K., Sorooshian, S. (1993) Shuffled Complex evolution approach for effective and efficient global minimisation. *Journal of Optimisation Theory and Applications. Volume 76, Issue 3, pp501-521.*
- Gijsbers, P.J.A., Werner, M.G.F., and Schellekens, J. (2008). Delft FEWS: A proven infrastructure to bring data, sensors and models together: In Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software (iEMSs 2008). International Environmental Modelling and Software Society, Barcelona, Catalonia, July 2008, pp. 28-36.
- Perraud, J.M., Bai, Q., Hehir, D. (2009a) On the appropriate granularity of activities in a scientific workflow applied to an optimization problem. *International Environmental Modelling and Software Society (iEMSs) 2010 International Congress on Environmental Modelling and Software Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)*
- Perraud, J.M., Vleeshouwer, J., and Stenson, M. (2012). Scaling up the calibration of hydrologic models to a continental scale. *10th International Conference on Hydroinformatics HIC 2012, Hamburg, Germany*
- Perraud, J.M., Vleeshouwer, J., Stenson, M., and Bridgart, R.J. (2009). Multi-threading and performance tuning a hydrologic model: a case study. *18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009* <http://mssanz.org.au/modsim09>
- Perraud, J., Wang, B., Vase, J. (2012) A data model for a sustainable management of parameter sets and optimization results. *International Congress on Environmental Modelling and Software, Germany.*
- Rahman, J.M., Seaton, S.P., Perraud, J.M., Hotham, H., Verrelli, D.I. and Coleman, J.R. (2003) It's TIME for a New Environmental Modelling Framework. *Proceedings of MODSIM 2003, 4, 1727-1732.*
- Stenson, M.P., Fitch, P., Vleeshouwer, J., Frost, A.J., Bai, Q., Lerat, J., Leighton, B., Knapp, S., Warren, G., Van Dijk, A.I.J.M., Bacon, D., Pena Arancibia, J.L., Manser, P., and Shoesmith, J. (2011). Operationalising the Australian Water Resources Assessment system. *WIRADA Science Symposium Proceedings, Melbourne, Australia, 1–5 August 2011.*
- Van Dijk, A.I.J.M., Renzullo, L.J. The Role of Satellite Observation in Australian Water Resources Monitoring. *34th International Symposium on Remote Sensing of Environment, Sydney, Australia 10-15 April 2011.*
- Vieny, N.R., Perraud, J., Vase, J., Chiew, F.H.S., Post, D.A., Yang, A., (2009) The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments. *18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009.*
- Zhang Y.Q., Viney N.R., Chiew F.H.S, van Dijk A.I.J.M., and Y.Y. Liu, “Improving hydrological and vegetation modelling using regional model calibration schemes together with remote sensing data”, *19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011.*