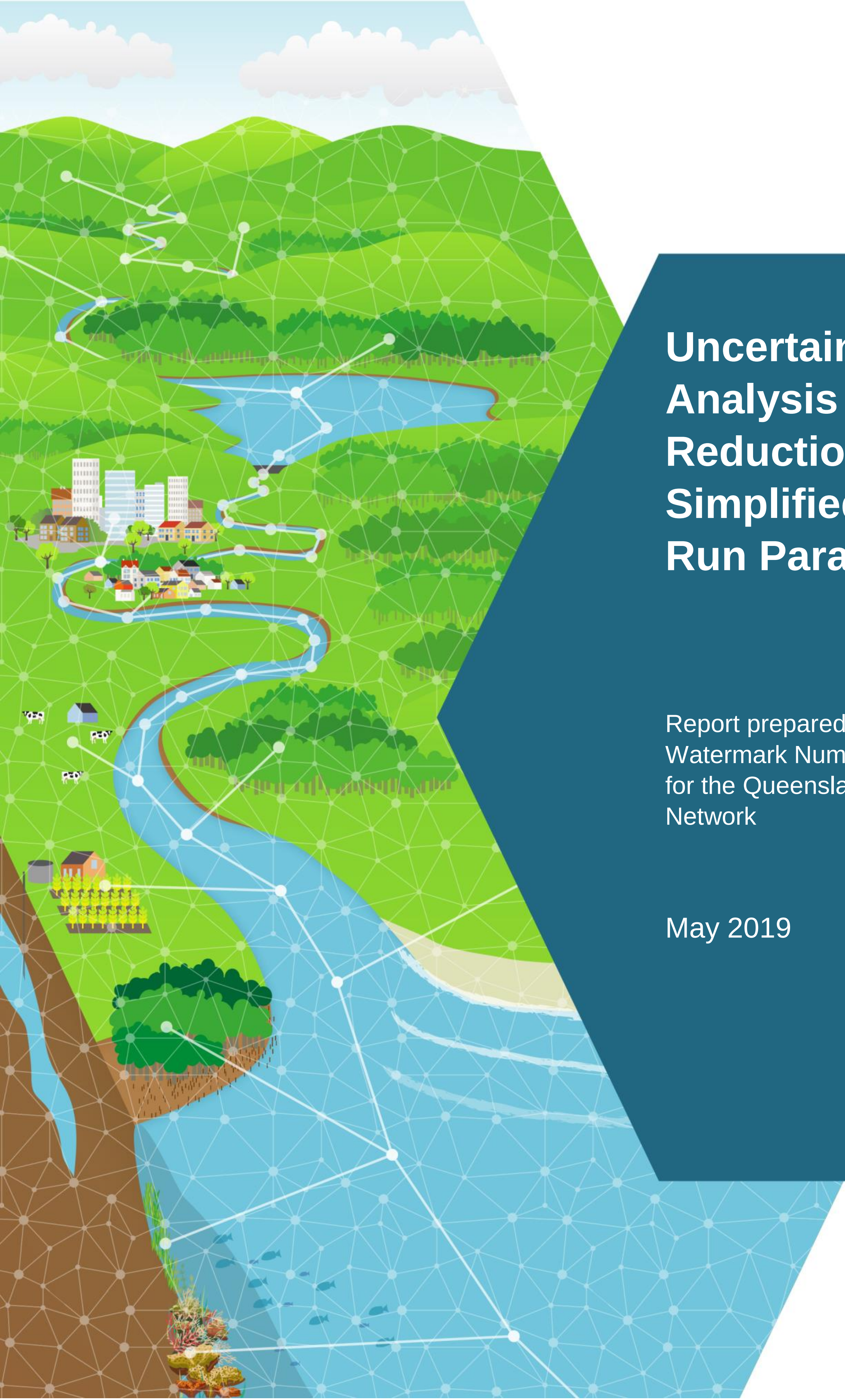
QUEENSLAND WATER MODELLING NETWORK



Uncertainty Analysis and Reduction through Simplified Model Run Parallelisation

Report prepared by John Doherty, Watermark Numerical Computing for the Queensland Water Modelling Network

May 2019

Queensland Government

Contents

Disclaimer

This report is the work of the author and does not represent the views or policies of the Queensland Government.

Queensland Water Modelling Network

The Queensland Water Modelling Network (QWMN) is an initiative of the Queensland Government that aims to improve the state’s capacity to model its surface water and groundwater resources and their quality. The QWMN is led by the Department of Environment and Science in partnership with the Department of Natural Resources, Mines and Energy and the Queensland Reconstruction Authority, with key links across industry, research and government.

Introduction

This document comprises the final report for the QWMN Model Parallelisation Project. Work on this project was carried out over the 2018 calendar year by the following personnel

- John Doherty (Watermark Numerical Computing, Brisbane, Australia)
- Fred Bennett (Resource Assessment, Department of Environment and Science)
- Kiran Bajracharya (Water Planning and Coastal Sciences, Department of Environment and Science)
- Chas Egan (Queensland Hydrology, Department of Environment and Science)
- Dave Welter (South Florida Water Management District).

Project-related technical discussions, including those pertaining to planning and specifications for software development and writing of training documents, were held with the following personnel:

- Jeremy White (GNS, New Zealand, and United States Geological Survey)
- Adam Siade (University of Western Australia and CSIRO)
- Mike Fienen (United States Geological Survey)
- Randy Hunt (United States Geological Survey)
- Craig Simmons (National Centre for Groundwater Research and Training, Flinders University)
- Jim Rumbaugh (Environmental Simulations, USA).

Background

Data Assimilation and Uncertainty

An important, and often under-appreciated, role played by environmental simulation undertaken for the purpose of decision support is that of data assimilation. Many predictions that are made by environmental models (especially models which simulate movement of subsurface water and contaminants) are highly uncertain. This is an outcome of the extreme heterogeneity of natural systems, and of the paucity of data that are available to characterise them.

If undertaken in conjunction with appropriate model-value-adding software, environmental simulation can reduce these uncertainties through assimilation of information-rich data, particularly data comprised of measurements of the historical behaviour of the system that is undergoing simulation. Unfortunately, the uncertainties associated with critical model predictions can remain high, even after a model has been subjected to history-matching. Therefore, when a model is employed to simulate future system behaviour under an existing or altered management regime, these uncertainties must be quantified so that risks associated with environmental management can be assessed and incorporated into that management.

Reduction and quantification of predictive uncertainty requires that many model runs be carried out under the control of numerical algorithms that are designed for these purposes. The linkage between the computer code that encapsulates these algorithms and the simulator must be flexible. Ideally, linkage specifications should include the following details.

1. The interface between the data assimilation algorithm and the simulator should be non-intrusive. That is, the algorithm should be able to communicate with the simulator through the latter’s own input and output files. This promulgates flexibility of data transfer, eliminates the requirement for the simulator’s source code, and fosters innovative model-based processing of environmental data and corresponding simulator outputs.
2. The capability should exist for model runs to be conducted in parallel either on a single computer, across an office network, on a high performance computing cluster, or on the computing cloud. The option chosen for model run parallelisation depends on a modeller’s computing context.

PEST and PEST++

Non-intrusive model-based data assimilation was pioneered by PEST. “PEST” is an acronym for “Parameter ESTimation”. Since its initial release in 1995, it has become the industry standard for calibration and uncertainty analysis of groundwater flow and transport models. It is now widely used for calibration and predictive uncertainty analysis of geothermal, groundwater, surface water, land process and other models. It was written by John Doherty, the lead of the QWMN model parallelisation project, and author of this document.

From 2002 onwards, programs of the PEST suite have supported model run parallelisation. The latest versions of PEST (named BEOPEST and PEST_HP) employ a manager/agent protocol for non-intrusive model run parallelisation. The manager implements the parameter estimation and/or uncertainty analysis algorithm for which repeated running of the model is required. It commissions model runs on an as-needed basis. However it is the agents that supervise these model runs; agents can reside on the same computer, or on different computers, from that on which the manager resides. As many agents should be initiated as there are parallelised model instances required.

Prior to running the model, an agent writes versions of model input files which contain the values of parameters that are specific to that model run. After the model run is complete, the agent reads user-specified numbers from output files that are generated by the model. The signal to run the model, and the values of parameters that the model must employ on any particular model run, are sent from the manager to the respective agent. On completion of a model run, an agent sends numbers read from model output files back to the manager. Information is exchanged between the manager and agents using the TCP/IP protocol.

PEST++ was developed as an enhancement of (and an eventual replacement for) PEST. Initial development of PEST++ (in 2009) was funded by the United States Geological Survey (USGS). Continuation of its development has been funded by both USGS and GNS Science (New Zealand). The original PEST++ development team was comprised of Chris Muffels (S.S. Papadopulous and Associates, Bethesda, MD), Dave Welter (South Florida Water Management district) and Jeremy White (USGS). PEST++ employs the same non-intrusive model interface and parallelisation protocols as does PEST. At the time of writing, the PEST++ suite includes a number of different computer programs; all of these perform a specific data assimilation, decision optimisation, and/or uncertainty quantification task. Of particular interest to the present project is PESTPP-IES. As is discussed below, this program encapsulates an iterative ensemble smoother.

As the name suggests, programming of the PEST++ suite is object-oriented and modular; its programming language is C++ (in contrast to PEST, which is written in FORTRAN). All programs of the PEST++ suite employ the same model run manager. Up until the inception of the QWMN parallelisation project, this was named YAMR. “YAMR” is a loose acronym for “Yet Another Run Manager”. It is programmed to supervise serial or parallel non-intrusive running of a simulator, including a batch/script file which includes more than one simulator.

While design specifications for PEST++ were intended to promulgate its interchangeable use with PEST, there are factors which make complete interchangeability of PEST and PEST++ a difficult undertaking. Of necessity, PEST++ input files require the specification of control variables that are different from those employed by PEST. While PEST itself can accommodate the presence of PEST++ control variables on its input file, the same was not the case for many members of the PEST utility support suite (which is presently comprised of over 200 programs).

The QWMN Parallelisation Project

The purpose of the QWMN parallelisation project was to encourage and support continued and expanded deployment of model-based data assimilation, uncertainty analysis and uncertainty reduction software. Its intention was to achieve this through the following means

- Promoting an increased understanding of the role that model-value-adding software plays in model-based decision support
- Supporting easier deployment of PEST and PEST++ software
- Enabling an easy transition from use of PEST to use of PEST++ programs

- Enabling modellers to readily write their own value-adding software by providing access to an easy-to-use modular, non-intrusive parallel run manager.

It achieved these aims through the following activities

- PEST suite programs were altered to support seamless interoperability with PEST++ suite programs. Repercussions of this are as follows
  - Any utility program which automates construction or modification of a PEST or PEST++ input dataset achieves the same thing for members of the other suite
  - Once a PEST or PEST++ input dataset has been constructed, it can be used by members of both suites.
- The writing of a document that describes the roles played by all members of the PEST and PEST++ suites, and explains how members of the two suites can be used cooperatively and interchangeably.
- A user manual was written for the PEST++ suite. This replaced an outdated USGS open file report and a GitHub-published list of PEST++ control variables which constituted the entirety of PEST++ documentation prior to the writing of this manual.
- Capabilities of the YAMR run manager used by PEST++ were greatly expanded. Run management functionality was made available to a number of programming languages. YAMR was renamed to PANTHER.
- The PESTPP-IES iterative ensemble smoother was tested by DES personnel who conducted history matching and uncertainty analysis on a groundwater and surface water model. Both of these models had previously undergone history matching using more traditional methods.

Details of these activities are now provided.

The PANTHER Run Manager

PANTHER is a loose acronym for “parallel, non-intrusive handling of model runs in environmental management”. It was modified from the PEST++ YAMR run manager to be a stand-alone library to which a programmer can readily link his/her own model value-adding software. Like YAMR, PANTHER manages non-intrusive communication between a model and any program that adds value to that model by running the model many times in support of the specific value-adding task that a user’s program is designed to implement. These model runs can be undertaken in serial or in parallel.

Functionality included in PANTHER that was previously unavailable in YAMR includes the following

- PANTHER functions can be called from the FORTRAN, C, C++ and Python programming languages.
- Model runs can be managed dynamically. Runs can be added to the model run queue and/or subtracted from that queue at any time. The outcomes of individual model runs can be processed immediately following their completion instead of waiting for a batch of model runs to be completed.
- Encrypted files can be passed between a manager and its agents.
- PANTHER is accompanied by a “universal worker”. This worker can act as an agent to supervise local model runs, independently of the program that uses PANTHER for run management.
- A comprehensive programmer’s manual was written for PANTHER.

Through use of the PANTHER run manager, a programmer is relieved of responsibilities for parallel model run management and non-intrusive communication with a model. Use of PANTHER therefore frees a programmer to concentrate on the algorithmic details of his/her model value-adding software, while writing that software in the language of his/her choice.

It is worthy of note that PANTHER provides model run management services for innovative particle swarm optimization software developed by Adam Siade (University of Western Australia and CSIRO). Research work which use of this optimizer has enabled is featured in a number of his publications; see, for example, Rathi et al. (2017) and Siade et al. (2019).

Software and Manuals

PEST

The PEST manual is comprised of two volumes. It is downloadable with PEST itself from the PEST web site.

The PEST manual has been updated to reflect interoperability of PEST suite programs with members of the PEST++ suite. Many of the changes made to PEST suite programs to implement PEST++ interoperability are invisible to the user; hence they require little or no documentation. However a number of utility programs were added to the PEST++ suite specifically to enhance opportunities for collaborative use of PEST and PEST++ programs. Tasks implemented by these new utilities include the following.

- A new utility removes all PEST++ features from a PEST control file. This enables use of a PEST++ control file by a few older members of the PEST family that were not programmed for PEST++ inter-operability.
- A number of new PEST utilities support reading, writing and translation of so-called “JCB” files. These are binary files used for matrix storage by members of the PEST++ suite.
- New PEST utilities generate random realizations of parameter fields, as well as of measurement noise. These provide PEST support for the PESTPP-IES iterative ensemble smoother.

PEST++

A comprehensive manual was written for the PEST++ suite. This manual can be downloaded from the PEST++ GitHub repository.

Education and Training

A document that describes the history and functionality of PEST, PEST++, their respective supporting utility suites, and ways in which these suites can be used collaboratively, can be downloaded from here.

This page (one of the PEST web pages) is dedicated to describing how PEST and PEST++ can be used together. It acknowledges the support provided by the QWMN model parallelisation project in achieving this level of interoperability.

Also downloadable from this page are files and documentation for a tutorial which guides a modeller through conjunctive use of PEST and PEST++ utilities in PESTPP-IES pre/post-processing. This tutorial takes as its starting point a tutorial supplied with the Groundwater Vistas MODFLOW graphical user interface which describes the use of PEST in calibration of a groundwater model in which pilot points are employed as the parameterization device. It was written in order to facilitate adoption of PESTPP-IES by users of Groundwater Vistas.

Worked Examples

General

As was discussed above, PESTPP-IES (a member of the PEST++ suite) is an iterative ensemble smoother. The algorithm that it employs is based on that developed by Chen and Oliver (2013); it is also described by White (2018). The use of ensembles as a mechanism for history-matching, and for exploration of post-history-matching uncertainty, is growing at a rapid rate, particularly in the petroleum industry. Ensemble-based history matching has some advantages over traditional calibration and uncertainty analysis methods such as those employed by PEST. The latter are based on regularised inversion followed by exploration of post-calibration parameter and predictive uncertainty using semi-linear methods such as null space Monte Carlo (Tonkin et al., 2007). The advantages of using PESTPP-IES include the following.

- Because the use of ensembles dispenses with the need to calculate derivatives of model outputs with respect to parameters, parameters can number in the hundreds of thousands, or even millions.
- Beyond a certain number of parameters (a few hundred in most groundwater modelling circumstances), the computational cost of history matching does not increase with the number of parameters requiring estimation. The cost is a function only of the number of realisations which collectively comprise an ensemble. These need to outnumber the dimensionality of the solution space of the inverse, history-matching problem.
- Exploration of post-history-matching uncertainty is not undertaken as an adjunct to the history-matching process. Instead, history-matching and uncertainty analysis are implemented through the same iterative, Bayesian procedure.

Not unexpectedly, the use of ensembles as a basis for history-matching is also accompanied by some disadvantages. (The small numerical burden associated with ensemble-based history-matching does not come for free.) Disadvantages include the following.

- Bias may be introduced to the posterior probability distributions of some predictions.
- Fits attained between model outputs and members of the calibration dataset may not be as good as those that can be achieved using regularized inversion.
- Due to its use of random fields as a parameterization device, numerically delicate models may suffer from numerical instability when used in an Iterative Ensemble Smoother (IES) framework.

As part of the QWMN parallelisation project, DES personnel deployed PESTPP-IES for the history-matching of two very different models. One of these is a highly parameterized groundwater model; the other is a regional surface water model whose adjustable parameter set is large in comparison with that traditionally ascribed to surface water models. The outcomes of these exercises are now briefly described.

Case 1. Central Lockyer Groundwater Model

The Central Lockyer groundwater model was developed in 2000; see Durick and Bleakley (2000) for details. A slightly modified version of the USGS MODFLOW model (McDonald & Harbaugh, 1988) simulates movement of groundwater under a large part of the Central Lockyer Valley, in particular under the Clarendon Sub Artesian Area. Water within alluvial sediments is recharged periodically by Lockyer Creek. Some of this water is extracted from local alluvial aquifers and used for irrigation. Sustainable management of the system relies on predictions made by the Central Lockyer groundwater model.

In late 2017 the Central Lockyer groundwater model was re-calibrated by DES modellers. Aquifer hydraulic properties were not adjusted. However recharge rates in each of 38 creek bed zones were estimated for 70 stress periods of approximately three months duration spanning the period July 1997 to December 2014. 2660 parameters featured in the model calibration process; however, 466 recharge rates were set to very low values at times and locations at which it was known that recharge was virtually absent.

The calibration dataset was comprised of a total of 10,182 heads measured in 108 wells.

DES personnel originally calibrated the model using PEST. Results were good; the standard observation error being was 1.69 m. As part of the QWMN model parallelisation study, the history-matching process was repeated using PESTPP-IES using an ensemble size of 3000 realizations. This is greater than the number of realizations that are commonly employed for groundwater model calibration. However history-matching of the Central Lockyer groundwater model benefitted from this large ensemble size because of the large number of observations featured in the calibration dataset, and because of the high information content of those observations.

With about the same model run cost as that required by PEST, a similar level of fit with the calibration dataset was obtained for most parameter realizations employed by PESTPP-IES. Post history-matching parameter uncertainties were calculated from adjusted realizations comprising the PESTPP-IES posterior ensemble. These were compared with those estimated using linear analysis based on a PEST-calculated Jacobian matrix. While the results were generally similar, there were some differences. The source of these differences is unknown. They may be an outcome of model nonlinearity. Alternatively, they may be an outcome of a small amount of ensemble "collapse" – a phenomenon that sometimes accompanies use of ensemble methods.

Overall, the results of this study suggest that the successes that ensemble methods have enjoyed in history matching and uncertainty analysis of petroleum reservoir models are likely to be repeated for groundwater model history matching and uncertainty analysis. This study also suggests that the Central Lockyer groundwater model could benefit from further application of PESTPP-IES. In particular, the introduction of aquifer hydraulic conductivity and storage parameters to the history matching process would not increase the numerical burden of this process if undertaken by PESTPP-IES. At the same time, good fits are likely to be achieved between model outputs and members of the calibration dataset. Meanwhile addition of these extra, spatially variable, parameters to the history matching process has the potential to reveal greater parametric (and perhaps predictive) uncertainty than that which was revealed through history-matching of this model using creek recharge parameters alone.

A report on this work is available on request.

Case 2. Uncertainty Analysis of Pioneer Catchment Rainfall Runoff Model

The Iterative Ensemble Smoother (IES), described above, represents a powerful approach for dealing with very high dimensional history matching problems that are typically encountered in groundwater and petroleum reservoir modelling. Although rainfall-runoff models are not usually subject to the same "curse of dimensionality" as the aforementioned model categories, they do present their own set of challenges when it comes to calibration and uncertainty analysis.

The influence of calibration data uncertainty on model performance is regarded as important, however is rarely accounted for in rainfall-runoff modelling. One interesting feature of the IES is that it efficiently implements an ensemble based variant of the Randomised Maximum Likelihood (RML) method. The RML directly connects model parameter optimisation to the noise in the observed data that the model is being calibrated against. This means that data uncertainty can be included in the calibration process in an implicit fashion.

A rainfall-runoff model for the Pioneer catchment in the Mackay-Whitsunday region was calibrated using the IES to demonstrate its application to this type of model. Using this approach, it was possible to investigate how model parameter uncertainty evolves with increasing data uncertainty. This work is the subject of a paper that is currently in preparation.

The IES is a highly effective tool for evaluating model uncertainty, especially when calibration data uncertainty is expected to significantly contribute to the overall model performance.

Conclusion

The QWMN model parallelisation project has provided decision-support modellers, both in Queensland and elsewhere, with greater access to model value-adding software than existed prior to the project. This has been a direct outcome of a number of project activities. In particular

- The PANTHER run manager has relieved numerical programmers and environmental modellers of specialist technical programming tasks required for non-intrusive parallel model run management.
- Workshops, manuals and educational material has facilitated adoption by PEST users of powerful history-matching and uncertainty analysis software that has recently become available through the PEST++ suite.
- DES personnel have gained experience in the use of ensemble methods to enable history-matching and uncertainty analysis of highly parameterized ground and surface water models.

References

Chen, Y and Oliver, DS 2013, ‘Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification,’ *Computational Geosciences*, 17(4), 689-703.

Durick, A and Bleakley, A 2000, Central Lockyer Groundwater Model. Water Assessment Branch, Department of Natural Resources and Mines.

McDonald, MG and Harbaugh, AW 1988, ‘MODFLOW: A modular three-dimensional finite difference groundwater model,’ Open File Report 83-875, US Geological Survey, Washington.

Rathi, B, Neihardt, H, Berg, M, Siade, A, and Prommer, H 2017, ‘Processes governing arsenic retardation on Pleistocene sediments: adsorption experiments and model-based analysis,’ *Water Resources Research,* 53, 4344-4360.

Siade, AJ, Cui, T, Karelse, RN and Hampton, C 2019, ‘Reduced order Gaussian process machine learning for groundwater allocation planning using swarm theory,’ submitted to *Water Resources Research*.

Tonkin, M, Doherty, J and Moore, C 2007, ‘Efficient nonlinear predictive error variance for highly parameterized models,’ *Water Resources Research*, 43, W07429, doi:10.1029/2006WR005348.

White, JT 2018, ‘A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions,’ *Environmental Modelling and Software*, 109, 191-201.