A river system modelling platform for Murray-Darling Basin, Australia

Ang Yang, Geoff Podger, Shane Seaton and Robert Power

ABSTRACT

Global climate change and local development make water supply one of the most vulnerable sectors in Australia. The Australian government has therefore commissioned a series of projects to evaluate water availability and the sustainable use of water resources in Australia. This paper discusses a river system modelling platform that has been used in some of these nationally significant projects. The platform consists of three components: provenance, modelling engine and reporting database. The core component is the modelling engine, an agent-based hydrological simulation system called the Integrated River System Modelling Framework (IRSMF). All configuration information and inputs to IRSMF are recorded in the provenance component so that modelling processes can be reproduced and results audited. The reporting database is used to store key statistics and raw output time series data for selected key parameters. This river system modelling platform has for the first time modelled a river system at the basin level in Australia. It provides practitioners with a unique understanding of the characteristics and emergent behaviours of river systems at the basin level. Although the platform is purpose-built for the Murray-Darling Basin, it would be easy to apply it to other basins by using different river models to model agent behaviours.

Key words | agent-based modelling, Murray-Darling Basin, river system model

Ang Yang (corresponding author)
Geoff Podger
Shane Seaton
CSIRO Land and Water,
PO Box 1666,
Canberra,
ACT 2601,
Australia
E-mail: *ang.yang@csiro.au*

Robert Power
CSIRO ICT Centre,
GPO Box 664,
Canberra,
ACT 2601,
Australia

INTRODUCTION

Global climate change and local development make water supply one of the most vulnerable sectors in Australia (Bates *et al.* 2008). A series of projects have been conducted to evaluate water availability and the sustainable use of water resources in Australia, especially in the Murray-Darling Basin (MDB). MDB is Australia's largest river basin and home to two million people accounting for approximately 60% of the water used in Australia (CSIRO 2008). A number of legacy river system models developed by various MDB water management authorities pre-exist for many regions on the MDB. The interconnections between them have either been ignored, or are managed by manually transferring data files between models. In reality, however, the river systems physically interact with each other. Therefore, being able to model the interaction between the different river systems is critical to model the system as a whole (e.g. at the basin level). In recognition of this, some attempts have been made to model interactions by connecting models, e.g. OpenMI (http://www.openmi.org). However, all model connections have to follow the same connection protocol; it is not flexible enough to be adopted in our case.

In this paper, a river system modelling platform is proposed to model the whole MDB as a single drainage basin. The core component of this proposed platform is the modelling engine, an agent-based hydrological simulation system called the Integrated River System Modelling Framework (IRSMF) (Yang & Podger 2010). In IRSMF, each subsection of the MDB is represented as an agent, the behaviour of each agent is modelled by a legacy model (either a surface water model or a groundwater flow model), and the interactions between agents are modelled as the hydrological connectivity between the real-world systems. The proposed river system modelling platform has successfully been used

doi: 10.2166/hydro.2012.153

in the CSIRO MDB Sustainable Yields project (MDBSY) and the Murray-Darling Basin Authority (MDBA) MDB Plan project (MDBP).

MDBSY for the first time estimated current and future water availability at the basin level in Australia (CSIRO 2008). The results from MDBSY have provided the most accurate assessment, to date, of the available water resources in the basin. They have shown how the effects of climate change, forestry and farm dam development, and groundwater development impact on local regions and throughout the MDB. In the MDBP, a series of 'what-if' analyses has been conducted, informing the Australian federal government's first strategic plan for the integrated and sustainable management of water resources in the MDB.

This paper provides an overview of the proposed river system modelling platform, followed by a detailed description of each component of the platform.

OVERVIEW OF THE RIVER SYSTEM MODELLING PLATFORM

System architecture

Figure 1 presents the high level system architecture of the main components of the proposed river system modelling platform and shows the interactions between components. The diagram elements adopt the following conventions: the rounded rectangles are software applications; the ovals are information stores (the file system or a database); solid lines depict the flow of information from source to destination; and the dotted arrow indicates data linkages, an association from one dataset to another. There are three categories of users: *modellers* explore the impact of altering agent behaviours to achieve a particular goal under different climate and development conditions, and the *Reporting Group* and *Policy Planners* examine reports produced from the information generated from the IRSMF output targeted at specific water resource planning objectives.

The *Reporting Database* records the key summary statistics and raw time series outputs of agents and includes a description of the modelling processes used to configure the IRSMF (the settings) to achieve these outputs. Standard reports can be generated with reporting tools. Some of the plots generated require access to the 'raw' time series outputs as well as the summary statistics.

The *Provenance* system ensures the exact same simulation can be reproduced at a later date to generate the exact same results as before. This ensures that it is possible to trace the results back to the source data, model version, and system configurations.

The IRSMF modelling engine is the core of the proposed platform, where all modelling processes are run. It is written in C#.net and has a graphical interface that allows executables and configuration files to be added and assigned to

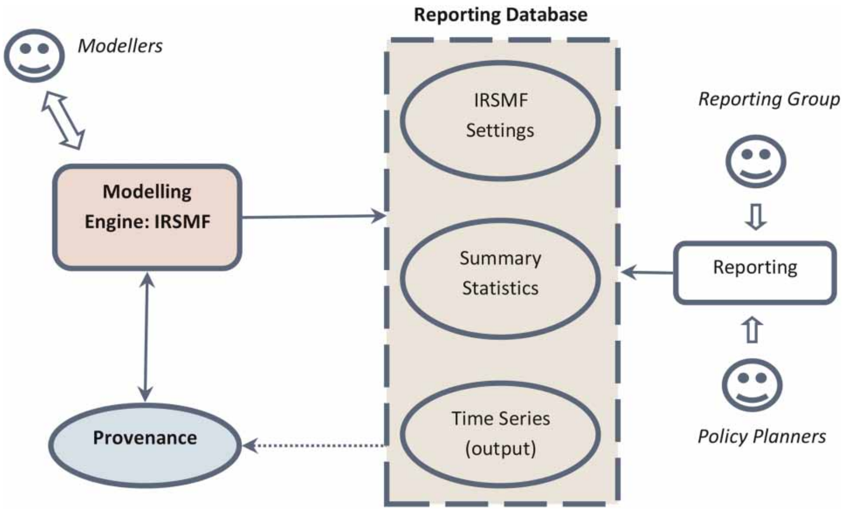


Figure 1 | System architecture.

individual agents, agents connected together and the scenarios run. It performs three basic tasks:

1. Gathers climate and flow scenarios, modifies inputs for various agents, then simulates the whole river system in a specified sequential manner (specified in an XML configuration file).
2. Extracts specific time series information from agents, allows for time step differences and inputs the time series information to connected agents. The information includes flow, height, storage spills, storage volume, demands and resource availability (allocation) information.
3. Post processes results from the agents, for example, converting the outputs into a consistent format for uploading to the reporting database. These results are subsequently used in preparing reports.

When a modeller decides that the results of a study on a valley (one agent or several linked agents), a region (linked valleys), or the entire basin (linked regions) will be of interest to others or should be recorded for comparative purposes or further analysis then the results are archived. It is at this point that the interactions between the IRSMF and the various information stores within the reporting database are activated.

Workflow

Before starting a study, a modeller will either manually prepare all legacy river system models, which are used to model behaviours of agents, and required input scenario data using a specific file structure, or resurrect a snapshot of the whole file structure from the provenance system. This includes the whole file structure of the agent template directory and the scenario data directory, a list of scenarios and a list of parameters for each agent type that can be adjusted for the study. The modeller can then conduct a study with the following six steps (Figure 2).

1. Configure the run time environment including the path of the agent template directory, scenario data directory, and working and output directory, scenario list and tweak tags.

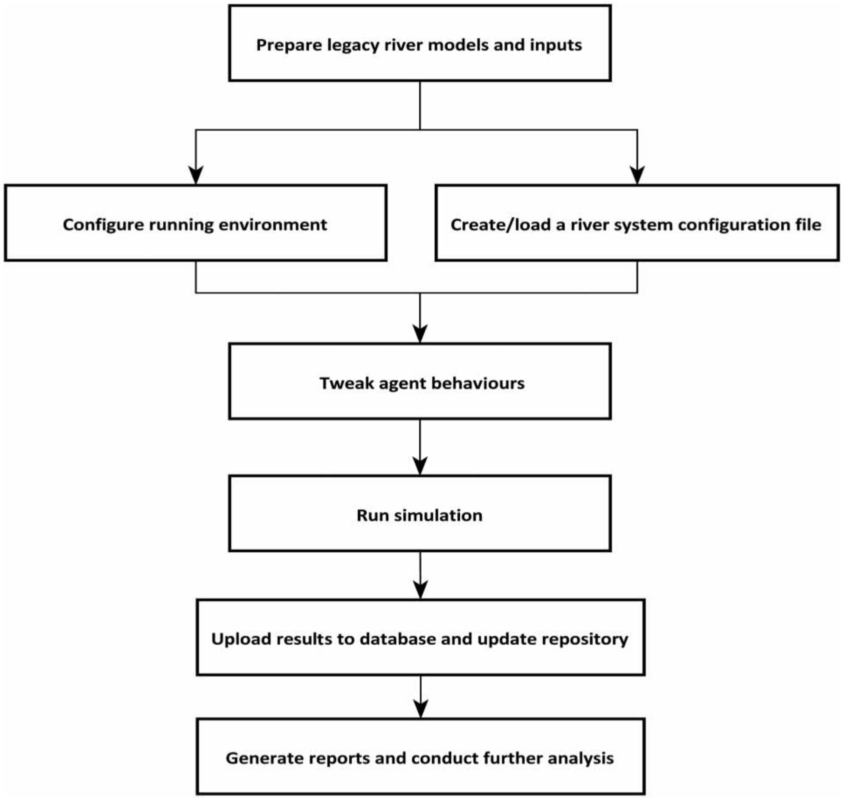


Figure 2 | Typical workflow.

2. Define a river system including creating and configuring all agents and their interactions. After defining a river system, e.g. individual valleys, regions or the entire basin, it can be saved as an XML definition file. At a future time this river system does not need to re-define; IRSMF can load this river system by opening the XML definition file.
3. Tweak individual agent behaviours as required by changing the values of the corresponding parameters, discussed in the next section.
4. Run the simulated river system.
5. Upload key summary statistics, key output time series and IRSMF settings to the reporting database for further analyses. If the inputs or configurations of any agent were changed, then the provenance also needs to be updated.
6. A series of standardised reports is generated and further analyses can be conducted.

MODELLING ENGINE – IRSMF

Understanding the dynamics of coupled social-environmental systems is more complex than non-coupled systems because of nonlinearities and emergent behaviours. The overall system behaviour cannot simply be predicted by aggregating knowledge of each component of such a system. Therefore, multi-agent systems are increasingly adopted as a simulation tool to explore nonlinear interactions among social and natural components within a system, such as environmental changes, human actions, policy interventions, etc. (Becu *et al.* 2003; Bousquet & Page 2004; Yang *et al.* 2006; Bithell & Brasington 2009). In multi-agent systems, the constituent parts are usually modelled as agents with a set of pre-defined characteristics. In this way, a real world system is simulated by an artificial world populated by interacting processes to reproduce nonlinear patterns and emergent behaviour. It is particularly effective to use multi-agent systems to represent the real world systems which are composed of a number of nonlinear interacting parts that have a large space of complex decisions and/or behaviours to choose from.

Dunstan *et al.* (2005) and Ryan & McAlpine (2005) adopted agent-based modelling techniques to study the impacts of human activities (land use) on the catchment's hydrology. Tillman *et al.* (2001), Becu *et al.* (2003), Bithell & Brasington (2009), and Farolfi *et al.* (2010) investigated the role of stakeholders/farmers (agents) in the water supply and demand systems and water management at the catchment level. van Oel *et al.* (2010) proposed a multi-agent approach to represent the interaction between spatial-temporal variability of water availability and water use. Reaney (2008) developed 'hydroAgents' to trace the flow of water through the catchment. The movement of 'hydroAgents' was determined by the local hydrological characteristics. In such a way, the temporal and spatial dynamics of flow generation and transmission can be captured during a storm event. It is interesting that in all the above studies, except Reaney (2008), the agent-based models were all adopted to model the social or economical component of the studied system.

The proposed IRSMF differs from the aforementioned studies; each agent in the IRSMF represents a subsection of the MDB. Its behaviours are modelled by an existing legacy river system model. Eight types of agent are supported in the IRSMF: IQQM (Simons *et al.* 1996), REALM (Diment 1991; Perera *et al.* 2005), PRIDE (Erlanger *et al.* 1992), MsmBigmod (MDBC 2001), SNOWY, St-George, MODFLOW (Harbaugh *et al.* 2000) and URBAN. The whole MDB was divided into 18 reporting regions. These regions represent the extent of the river system models that were used to develop the various water sharing plans in each of the regions in the MDB. There are 63 agents that describe the MDB: 19 IQQM agents, four MODFLOW agents, five REALM agents, 16 PRIDE agents, 16 URBAN agents, one MsmBigmod agent, one St George agent and one SNOWY agent. Most of the agents were developed and calibrated by jurisdictional agencies. Figure 3 shows an example of agents modelled in the 18 regions in the MDB. The boxes represent agents while the arrows represent the interaction between two agents.

Because the behaviours of each agent are modelled by a legacy model, modifying the value of certain parameters in the model will in turn change the behaviours of the corresponding agent. In such a way, it is easy to map between agent behaviours and system outcomes, and thus to study the dynamics of the whole system. The relevant parameters for a model are tagged with special characters, e.g. $G, $H, $E, $A, $F, etc. The modellers can modify the values of tagged parameters proportionally or by percentages, scale factors or absolute values. Figure 4 shows the interface

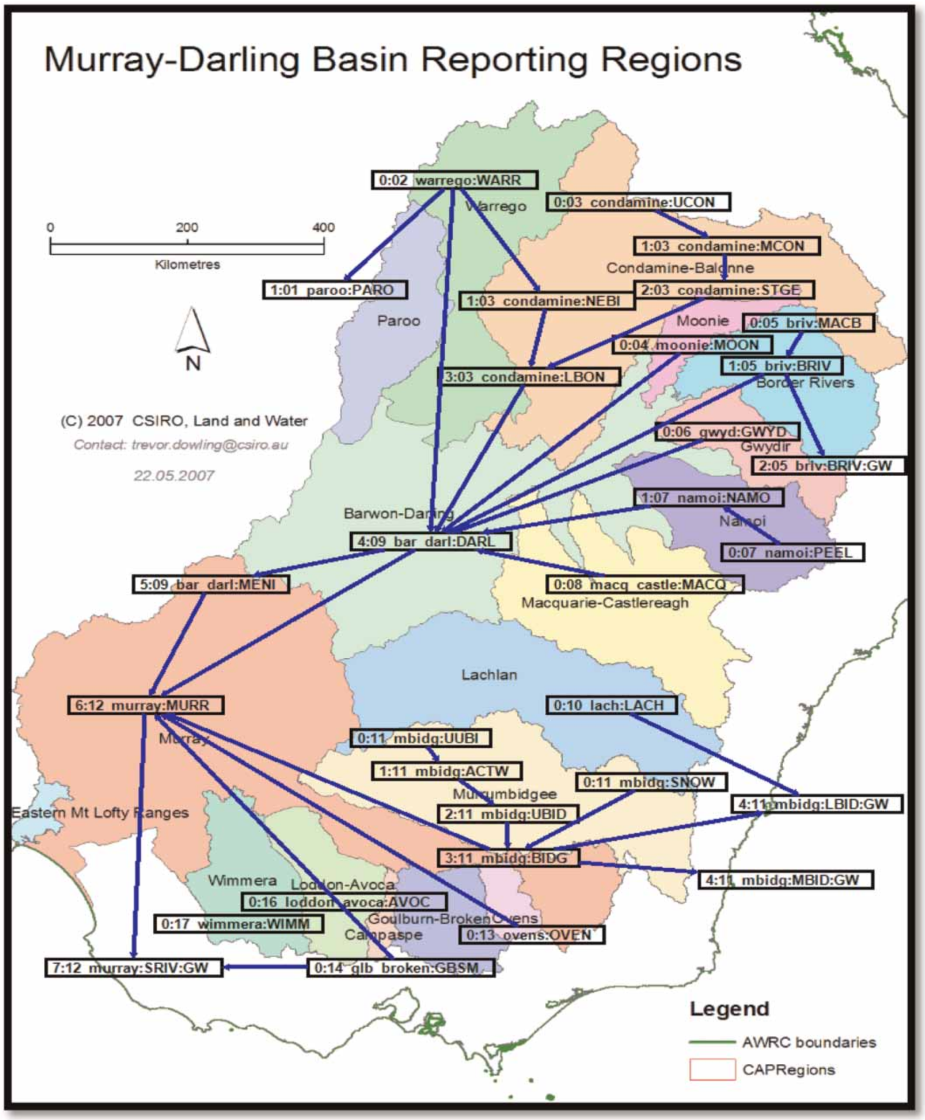


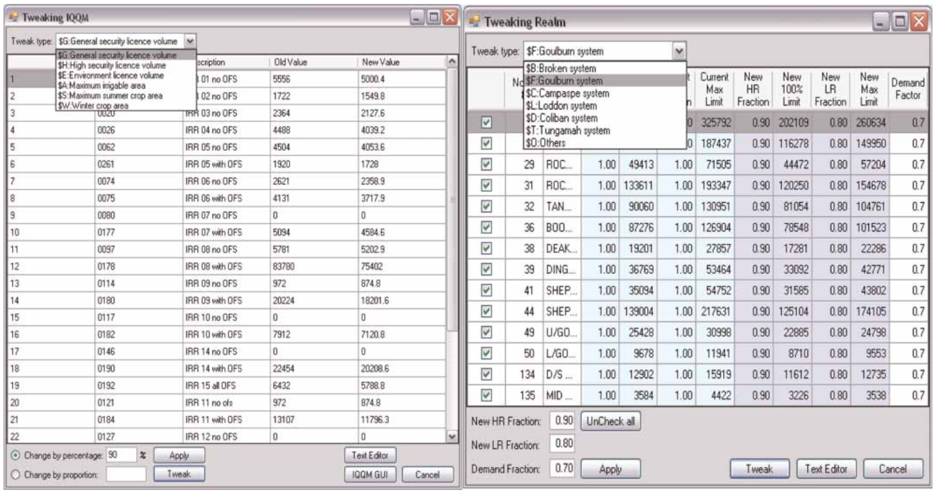
Figure 3 | An example of agents modelled in the 18 reporting regions in the MDB.

used to modify the behaviours of IQQM agents (left) and REALM agents (right). In such a way, modellers can easily investigate the impact, for example of upstream river systems on downstream river systems, and ultimately the behaviour of the entire basin.

Two types of agent interactions are modelled in the IRSMF. If the outputs of Agent A affect the inputs of Agent B, we call this influence from A to B as a 'feed forward' interaction. If A impacts B by a 'feed forward' interaction and the outputs of B affect the inputs of A in turn, we call the influence from B to A as a 'feedback' interaction. To simplify the implementation, there is no explicit implementation of the feedback interaction in the IRSMF. If B has feedback influence on A, we duplicate A as A1 and A2. Then two feed forward interactions are created: A1 → B and B → A2. The feed forward interaction of B → A2 is used to simulate the feedback interaction from B to A.

There are three steps to implement the interaction from agent A to B:

1. Extract the time series data at the connection point from the agent A outputs.



Tweak type: $G:General security licence volume

		scription	Old Value	New Value
1		01 no OFS	5556	5000.4
2		02 no OFS	1722	1549.8
3	0020	IRR 03 no OFS	2364	2127.6
4	0026	IRR 04 no OFS	4488	4039.2
5	0062	IRR 05 no OFS	4504	4053.6
6	0261	IRR 05 with OFS	1920	1728
7	0074	IRR 06 no OFS	2621	2358.9
8	0075	IRR 06 with OFS	4131	3717.9
9	0080	IRR 07 no OFS	0	0
10	0177	IRR 07 with OFS	5094	4584.6
11	0097	IRR 08 no OFS	5781	5202.9
12	0178	IRR 08 with OFS	83780	75402
13	0114	IRR 09 no OFS	972	874.8
14	0180	IRR 09 with OFS	20224	18201.6
15	0117	IRR 10 no OFS	0	0
16	0182	IRR 10 with OFS	7912	7120.8
17	0146	IRR 14 no OFS	0	0
18	0190	IRR 14 with OFS	22454	20208.6
19	0192	IRR 15 all OFS	6432	5788.8
20	0121	IRR 11 no ofs	972	874.8
21	0184	IRR 11 with OFS	13107	11796.3
22	0127	IRR 12 no OFS	0	0

● Change by percentage: 90 %
○ Change by proportion:

Tweak type: $F:Goulburn system

	No				n	Current Max Limit	New HR Fraction	New 100% Limit	New LR Fraction	New Max Limit	Demand Factor
☑						325792	0.90	202109	0.80	260634	0.7
☑						187437	0.90	116278	0.80	149950	0.7
☑	29	ROC...	1.00	49413	1.00	71505	0.90	44472	0.80	57204	0.7
☑	31	ROC...	1.00	133611	1.00	193347	0.90	120250	0.80	154678	0.7
☑	32	TAN...	1.00	90060	1.00	130951	0.90	81054	0.80	104761	0.7
☑	36	BOO...	1.00	87276	1.00	126904	0.90	78548	0.80	101523	0.7
☑	38	DEAK...	1.00	19201	1.00	27857	0.90	17281	0.80	22286	0.7
☑	39	DING...	1.00	36769	1.00	53464	0.90	33092	0.80	42771	0.7
☑	41	SHEP...	1.00	35094	1.00	54752	0.90	31585	0.80	43802	0.7
☑	44	SHEP...	1.00	139004	1.00	217631	0.90	125104	0.80	174105	0.7
☑	49	U/GO...	1.00	25428	1.00	30998	0.90	22885	0.80	24798	0.7
☑	50	L/GO...	1.00	9678	1.00	11941	0.90	8710	0.80	9553	0.7
☑	134	D/S ...	1.00	12902	1.00	15919	0.90	11612	0.80	12735	0.7
☑	135	MID ...	1.00	3584	1.00	4422	0.90	3226	0.80	3538	0.7

New HR Fraction: 0.90
New LR Fraction: 0.80
Demand Fraction: 0.70

Figure 4 | Tune up behaviours of IQQM agents (left) and Realm agents (right).

2. If the time step is different between agent A and agent B, the extracted time series is transferred to the time step required by the agent B using the algorithms described in the following paragraph.
3. Replace the data at the connection point in agent B with the time series data extract from agent A.

The connection points are determined from an understanding of the connectivity of the real hydrological systems. In the MDBSY, 12 types of feed forward interactions between agents are captured: IQQM → IQQM, IQQM → St George, St George → IQQM, IQQM → MsmBigmod, Snowy → IQQM, Snowy → MsmBigmod, IQQM → Realm, Realm → IQQM, Realm → MsmBigmod, IQQM → Modflow, Realm → Modflow and MsmBigmod → Modflow with 130 interactions in total. There are five types of defacto feedback interactions modelled in MDBSY. They are: Modflow → IQQM, Modflow → Realm, Modflow → MsmBigmod, IQQM → Snowy and MsmBigmod → IQQM. There are 106 interactions in total.

The time steps of agents are daily, weekly and monthly. When agents interact with each other, the outputs of an agent are not necessarily at the same time step as the inputs required by the interacted agents. The IRSMF provides several ways of handling both increases and decreases in time steps. The major challenge is to ensure that mass balance is preserved. Transforming time series data from shorter to longer time steps is done by simply adding data together. Most of the interactions between agents are from a larger time step to a shorter time step and five disaggregation methods are provided: (1) daily observed historical flows are added together across the larger time step (weekly or monthly) to get a total flow for that larger time step, then the daily observed historical flows are multiplied by the ratio of the modelled weekly or monthly flow divided by the observed total at the larger time step to get a daily series that preserves the original daily or weekly total; (2) fixed pattern; (3) mean value; (4) constant value; and (5) special method for release forecasts from Snowy (Yang 2010).

The IRSMF works strictly on a predefined file structure containing three main directories (Figure 5).

1. Scenario data directory: contains all required scenario data files (e.g. climate data, flow data). These files are different for different scenarios. They are inputs to agents.
2. Agent template directory: contains all files required to run a simulation of an agent including executable, configuration files and all other input files. Each agent must have an agent template directory and a matched scenario data directory.
3. Working and output directory: the directory where the simulation is running and outputs are generated.

Each of these main directories contains a series of subdirectories with a predefined file structure. When a simulation starts, all relevant files including data files and configuration files are copied to a particular directory under the working and output directory where the

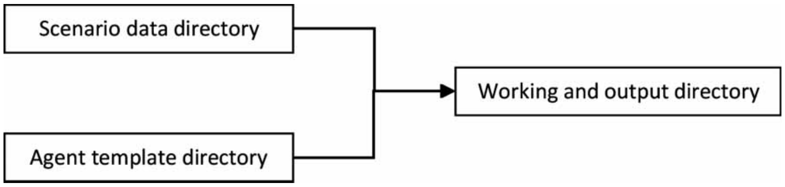


Figure 5 | Main directories in the IRSMF.

simulation is running. All the files and file structures in both the scenario data directory and agent template directory are stored and tracked in the provenance component so that the results from the simulation can be audited.

PROVENANCE

With advances in computation techniques, the provenance is becoming critical to effectively managing exponentially increasing volumes of data. The provenance is a comprehensive data documentation archive containing a description of 'how', 'when', 'where' and 'why' the data were produced and 'who' produced them. A number of applications of provenance systems have been developed in different domains, such as image processing (Braga & Banon 2008), security and privacy (McDaniel 2011), cosmology (Anderson *et al.* 2008), healthcare (Kifor *et al.* 2006), ecology (Osterweil *et al.* 2010) and eScience (Simmhan *et al.* 2005; Sahoo *et al.* 2008; Zhao *et al.* 2011). The provenance system in our case aims to ensure the exact same simulation run can be repeated at a later date to produce the exact same results as the original run. This ensures that it is possible to trace the results back to the source data, model version, and configuration.

All aspects of the agents and the IRSMF software are managed and stored in a Subversion version control system (http://subversion.apache.org/) as the provenance system. Subversion is used to track changes to: agent configurations, input time series datasets, model versions and the IRSMF version. When results are saved to the database for reporting, a unique snapshot is taken of the IRSMF and all its components. This snapshot is stored in the Subversion repository and a unique ID is associated with this snapshot. This way the Subversion repository (the central file store) becomes the provenance information store. When the summary statistics are loaded into the reporting database, the Subversion revision number and a few IRSMF specific details are all that is required to allow the same IRSMF configuration settings to be replicated, allowing a simulation to be repeated at a later date.

To repeat a simulation run, the entire IRSMF software and all corresponding inputs and configuration details are resurrected from the Subversion repository using the corresponding revision number, and re-run.

REPORTING DATABASE

The reporting database stores three types of data: IRSMF settings, summary statistics and raw time series data produced from each agent. A set of selected key statistics are automatically calculated after a simulation and stored in the reporting database. A description of the simulation is also stored in the reporting database. A link to the IRSMF state is made by recording the unique ID of the corresponding snapshot in the provenance system.

The database schema to capture this information is shown in the entity-relationship diagram (Figure 6). A description of these entities is as follows.

RepositoryDetails contains information about the Subversion repository. This includes the path to identify the repository and the revision number corresponding to the simulation used to produce the results. The revision number (unique ID) is the Subversion managed reference that can be used to resurrect the same file structure used for IRSMF. The path is the label for the directory used for the Subversion check out, the Subversion commit comments for that revision number, and the Subversion user name of who performed the commit.

FrameworkDetails records the configuration of the simulation. These details are provided by the user on the

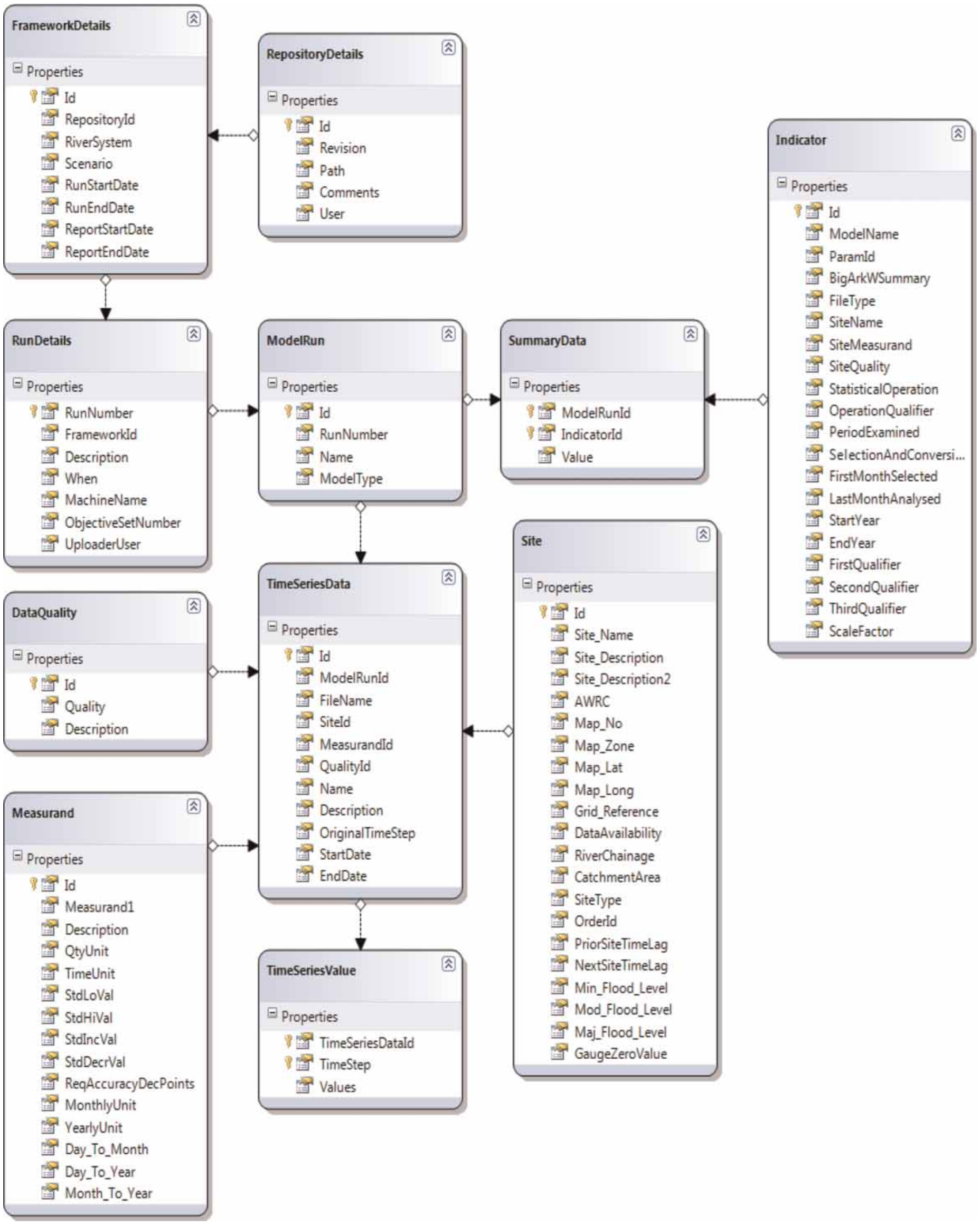


Figure 6 | Entity-relationship diagram.

IRSMF interface: the river system being run (the xml definition file loaded), the scenario being run, the period the models are run for (start and end dates), and the reporting start and end dates.

RunDetails contains the description of the simulation when performing a run. This includes a user-provided description of the purpose of the run, the actual start and end times when the run was performed, and the machine name on which it was run. This is referred to as a *Framework run* and will consist of many simulation runs.

ModelRun contains the information about an individual agent within a simulation. There are one or more ModelRun entries for a single simulation, corresponding to the number of agents in the IRSMF River System being run. The

ModelRun contains the agent name (for example NAMO or PEEL) and the type of agent (e.g. IQQM, REALM and MsmBigMOD). The RunNumber is the linkage to the RunDetails table.

SummaryData the summary statistics produced from a simulation are stored in this entity.

Indicator is a complete description of the parameters in the statistics configuration file used to generate the summary statistics.

Site is a description of a Site. A site has an identifier, the Id, which is unique across the basin, regardless of the region or state it is located in.

Measurand is a description of the data that may be collected at a site.

DataQuality is a further description of the data that may be collected at a site.

TimeSeriesData contains a description of a single time series. This includes a description of the identifiers that collectively uniquely identify a location: SiteId, MeasurandId, QualityId. These identifiers link to the respective tables: Sites, Measurands, and DataQuality. The ModelRunId column links to the Id from the ModelRun table. The time series data can be found in the TimeSeriesValues table.

TimeSeriesValue is the time series data stored in binary format in the Values column. The values are identified using the TimeSeriesDataId, which links to the Id column of the TimeSeriesData table. The same TimeSeriesDataId may record values in different time steps, hence the column TimeStep to distinguish them. The original values will always be stored, but it may be possible to also derive, for example, monthly or annual time series data and record these values also.

DISCUSSION

In the MDBSY project, the whole river system is simulated from 1 June 1895 to 30 July 2006 for four ‘without development’ scenarios and eight developed scenarios. The results are presented and analysed for the period 1 July 1895 to 30 June 2006 as the first month is required as a warm up period.

Figure 7 shows the current average surface water availability across the MDB (CSIRO 2008). It provides a clear picture of how much water is available across the MDB. This will help policy makers to efficiently distribute and use water resources.

Figure 8 shows the impact of the median 2030 climate on average surface water availability across the MDB (CSIRO 2008). The overall impact would be an 11% reduction, or 2,481 Gl year^{-1} less surface water on average. The impact on water availability due to climate change varies from region to region (e.g. 3% reduction in Paroo and a 21% reduction in the Wimmera). Sixty seven percent of the water availability reduction across the MDB is from the Goulburn-Broken, Ovens, Murray and Murrumbidgee regions. The reason that much of the reduction in surface water availability would occur in the south-east of the MDB is because most of the runoff in the MDB is generated in that area. Thus, the impact of climate change is likely to be greatest there.

While the MDBSY focuses on assessment of water availability under different climate and development conditions, the MDBP concentrates on how to efficiently and sustainably use water resources by making an effective water use policy across the basin. Therefore, the MDBA largely takes advantage of the flexibility of the river modelling platform to explore the solution space by adjusting agent behaviours (particular parameters). It uses a similar configuration to that used in the MDBSY, although some of the models have been updated to reflect more recent water management policy. The development of the first Basin Plan in Australia will be informed by the MDBP. The Basin Plan, due for release by MDBA in November 2011, provides for integrated and sustainable management of water resources in the MDB.

xA number of challenges were encountered in building this platform. There are 63 agents modelled by eight legacy surface water or groundwater models, and 218 agent interactions that describe the whole MDB. It is quite difficult to connect them all and get them to run for a common period on a common platform. Some of the models needed to be extended to cover the common modelling period. Connection points needed to be clearly identified at the up and downstream ends to ensure mass balance is preserved. Some models would only work with a specific directory structure and would not handle long

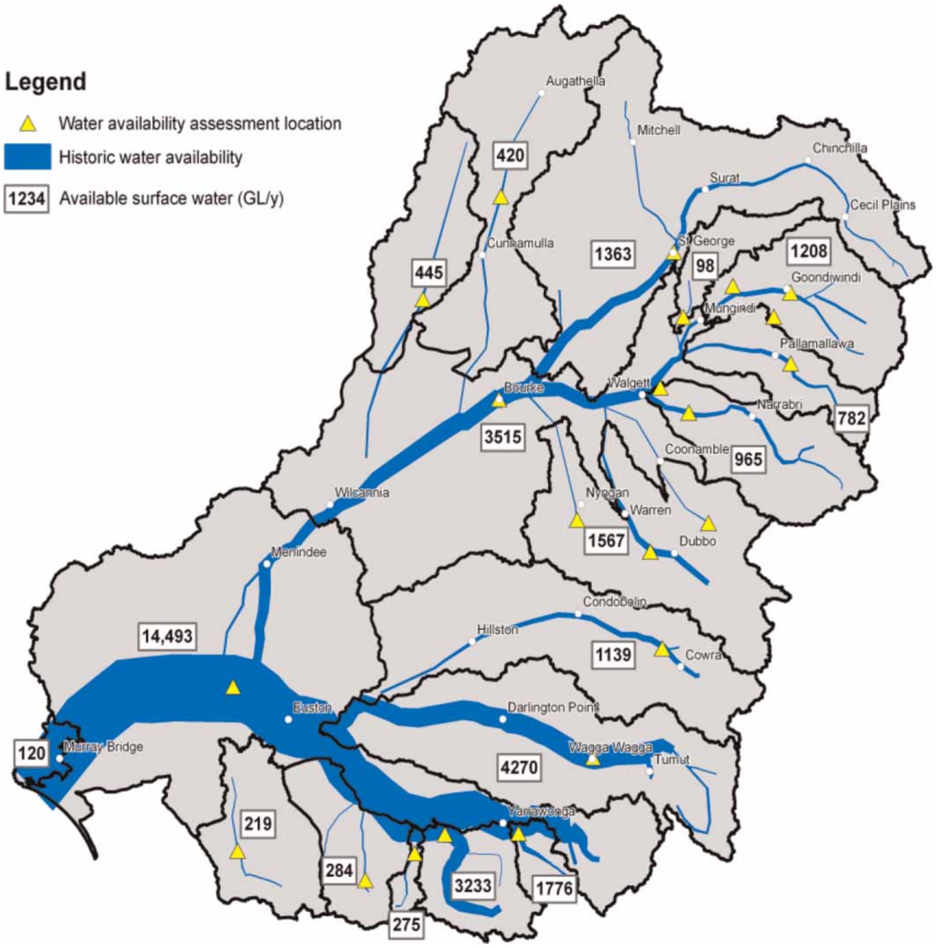


Figure 7 | Average surface water availability across the MDB (CSIRO 2008).

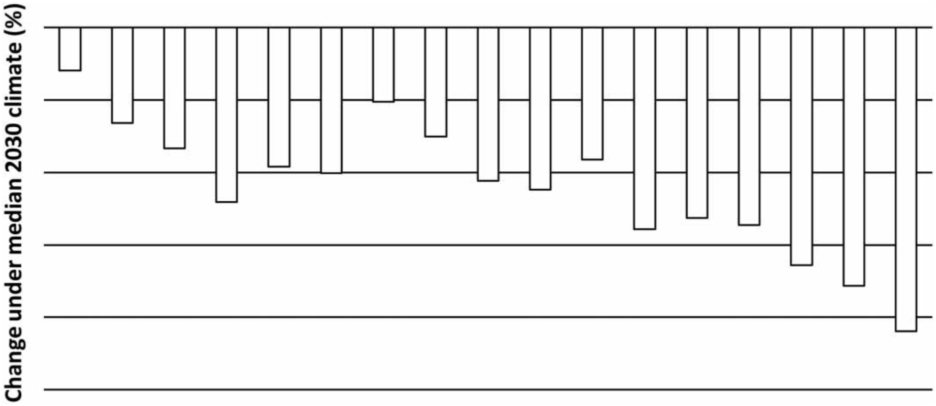


Figure 8 | Percentage changes in average surface water availability by region under the median 2030 climate.

file names and were subsequently modified. Some models would not run with multiple copies at the same time. A unique directory structure on multiple drives has to be developed to support all of the idiosyncrasies of the various models. Some models are so old that they have been compiled for 16-bit processers and needed to be recompiled so that they could be run on a 64-bit operating system. One of the models crashed for some of the extreme scenarios and had to be modified several times. The runtime is another issue. A single simulation of 111 years for the whole MDB takes about 3–4 days even with parallel computing. All the simulations have then to be submitted to a cluster so that the simulations for different scenarios can be run at the same time.

The methods used in this study to disaggregate results to a shorter time step preserve the new monthly values but do not reflect the likely changes in daily flow characteristics. The methods used for calibrating each of the models vary considerably. In some cases, the ungauged inflows are artificially large which is compensated by large unattributed loss relationships. In other cases, losses are included in the ungauged inflows. There is scope to develop a consistent and robust approach to calibration for all models. Many models use regression relationships for inflows and demands, which in some cases are not robust for extreme climate and flow changes. These need to be replaced by physically based models that reflect the change in demand as a function of climate and available resources. This is not an easy task and in the meantime the platform currently developed will still play a very important and useful role.

CONCLUSIONS

The proposed river system modelling platform has been developed to model the MDB at the whole of the basin scale. This has been the first time that catchment managers and policy makers have had access to such a tool to support water planning and policy development. It has provided practitioners a unique understanding of the characteristics and emergent behaviours at the basin level. Although it is purpose-built for the MDB, it would be easy to apply it to other basins by using different river models to model agent behaviours.

The proposed modelling platform allows the assessment of water resource contributions from multiple regions to downstream sites. This provides a useful way of understanding how changes within a region will impact on downstream users. Understanding this connectivity is essential to managing the whole of basin impacts for policy development and how they might impact on the environmental aspects of the entire basin.

The platform has successfully been applied into two high profile national projects: the MDBSY and MDBP, this work has resulted in some further improvements in the future use of the platform, such as a flexible plug-in system to allow the modeller to easily plug a new type of agent (model) into the platform; the integration of data analysis functionalities into the platform. The visualisation tool could be enhanced so that modellers may easily capture the pattern during data analysis; and feedback could be explicitly modelled rather than having multiple replications of models feeding forwards.

ACKNOWLEDGEMENTS

This paper is based on the MDBSY project funded by the National Water Commission under the Raising National Water Standards Program and the MDBP project funded by the MDBA. The authors would like to acknowledge the MDB state governments and private industry partners that contributed surface and groundwater models to the projects: NSW Department of Water and Energy (DWE), Queensland Department of Natural Resources and Water (DNRW), Victoria Department of Sustainable Environment, South Australian Department of Land Water and Biodiversity Conservation (DLWBC), Murray-Darling Basin Authority (MDBA), ACTEW and Snowy Hydro. The authors would also like to thank everyone who contributed to various aspects of the study. The authors would thank to the CSIRO internal reviewers, Dr Wendy Welsh and Dr Julien Lerat, for their constructive comments.

REFERENCES

Anderson, E. W., Silva, C. T., Ahrens, J. P., Heitmann, K. & Habib, S. 2008 Provenance in comparative analysis: a study in cosmology. *Computing in Science and Engineering* **10**, 30–37.

Bates, B. C., Kundzewicz, Z. W., Wu, S. & Palutikof, J. P. (eds) 2008 *Climate Change and Water. Technical Paper of the*

Intergovernmental Panel on Climate Change. IPCC Secretariat, Geneva, 210 pp.

Becu, N., Perez, P., Walker, A., Barreteau, O. & Page, C. 2003 Agent based simulation of a small catchment water management in northern Thailand: description of the CATCHSCAPE model. *Ecological Modelling* **170**, 319–331.

Bithell, M. & Brasington, J. 2009 Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution. *Environmental Modelling and Software* **24**, 173–190.

Bousquet, F. & Page, C. 2004 Multi-agent simulations and ecosystem management: a review. *Ecological Modelling* **176**, 313–332.

Braga, J. C. & Banon, G. J. F. 2008 Data provenance: theory and application to image processing. *IEEE Latin America Transactions (Revista IEEE America Latina)* **6**, 207–214.

CSIRO 2008 *Water Availability in the Murray-Darling Basin*. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. CSIRO, Australia.

Diment, G. A. 1991 Wide use of a generalised headworks and resources model REALM. *International Hydrology and Water Resources Symposium*, I.E., Australia, pp. 579–583.

Dunstan, N., Qiang, X., Cacho, O. & Hean, R. 2005 Multi-agent design for catchment analysis. In *International Conference on Information and Communication Technology*, Cairo, Egypt, IEEE, New York, pp. 101–107.

Erlanger, P. D., Poulton, D. C. & Weinmann, P. E. 1992 Development and application of an irrigation demand model based on crop factors. In *Conference on Engineering in Agriculture*, Australia. IEAust. Nat. Conf. Publication No. 92/11, 923–298.

Farolfi, S., Muller, J.-P. & Bonte, B. 2010 An iterative construction of multi-agent models to represent water supply and demand dynamics at the catchment level. *Environmental Modelling and Software* **25**, 1130–1148.

Harbaugh, A. W., Banta, E. R., Hill, M. C. & McDonald, M. G. 2000 *MODFLOW-2000, the U.S. Geological Survey modular ground-water model*. User guide to modularization concepts and the Ground-Water Flow Process. U.S. Geological Survey, Reston, Virginia, Open-File Report 00–92.

Kifor, T., Varga, L. Z., Vazquez-Salceda, J., Alvarez, S., Willmott, S., Miles, S. & Moreau, L. 2006 Provenance in agent-mediated healthcare systems. *IEEE Intelligent Systems* **21**, 38–46.

McDaniel, P. 2011 Data provenance and security. *IEEE Security and Privacy* **9**, 83–85.

MDBC 2001 *Model River: River Murray System Daily Model*. MDBC Report RMSDM 009, Murray Darling Basin Commission, Canberra.

van Oel, P. R., Krol, M. S., Hoekstra, A. Y. & Taddei, R. R. 2010 Feedback mechanisms between water availability and water use in a semi-arid river basin: a spatially explicit multi-agent simulation approach. *Environmental Modelling and Software* **25**, 433–443.

Osterweil, L. J., Clarke, L. A., Ellison, A. M., Boose, E., Podorozhny, R. & Wise, A. 2010 Clear and precise specification of ecological data management processes and dataset provenance. *IEEE Transactions on Automation Science and Engineering* **7**, 189–195.

Perera, B. J. C., James, B. & Kularathna, M. D. U. 2005 Computer software tool REALM for sustainable water allocation and management. *Journal of Environmental Management* **77**, 291–300.

Reaney, S. M. 2008 The use of agent based modelling techniques in hydrology: determining the spatial and temporal origin of channel flow in semi-arid catchments. *Earth Surface Processes and Landforms* **33**, 317–327.

Ryan, J. G. & McAlpine, C. A. 2005 Simulating the feedback between land cover configuration and ecohydrological functioning in complex adaptive landscapes. MODSIM 2005: International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Nedlands, Univ Western Australia, (pp. 758–764).

Sahoo, S. S., Sheth, A. & Henson, C. 2008 Semantic provenance for eScience: managing the deluge of scientific data. *IEEE Internet Computing* **12**, 46–54.

Simmhan, Y. L., Plale, B. & Gannon, D. 2005 A survey of data provenance in e-science. *SIGMOD Record* **34**, 31–36.

Simons, M., Podger, G. & Cooke, R. 1996 IQQM – a hydrologic modelling tool for water resource and salinity management. *Environmental Software* **11**, 185–192.

Tillman, T., Larsen, T. A., Pahl-Wostl, C. & Gujer, W. 2001 Interaction analysis of stakeholders in water supply systems. *Water Science and Technology* **43** (5), 319–326.

Yang, A. 2010 *The Integrated River system Modelling Framework. A report to the Murray-Darling Basin Authority*. CSIRO: Water for a Healthy Country National Research Flagship, Canberra, June 2010.

Yang, A., Abbass, H. A. & Sarker, R. 2006 Characterizing warfare in red teaming. *IEEE Transactions on Systems, Man, Cybernetics, Part B: Cybernetics* **36**, 268–285.

Yang, A. & Podger, G. 2010 An agent based hydrological simulation system for Murray-Darling Basin, Australia. In *9th International Conference on Hydroinformatics HIC 2010*, 7–11 September 2010, Tianjin, China, pp. 1478–1485.

Zhao, J., Sahoo, S. S., Missier, P., Sheth, A. & Goble, C. 2011 Extending semantic provenance into the web of data. *IEEE Internet Computing* **15**, 40–48.

First received 24 October 2011; accepted in revised form 12 December 2011. Available online 29 March 2012