An integrated catchment yield modelling environment

Yang, A

CSIRO, Water for a Healthy Country Flagship, GPO Box 1666, Canberra, ACT, 2601
Email: ang.yang@csiro.au

Abstract: Catchment yield modelling involves a number of complex processes such as input data scaling and data quality checking, model calibration, parameter set generation for gauged and ungauged catchments, simulation runs, post data analysis, etc. In the CSIRO Murray Darling Basin Sustainable Yields project (MDBSY), a suite of tools was developed and driven by a series of Python scripts to perform these modelling processes repeatable and consistently. Although it was flexible, it was very hard for modellers with non IT background to run them. Keeping that mind, an Integrated Catchment Yield Modelling Environment (ICYME) is developed and applied in the CSIRO Tasmania Sustainable Yields Project (TasSY) and the South-west Western Australia Sustainable Yields Project (SWWASY). It is built on the toolset developed in MDBSY. Given its user friendly GUI, modellers can easily (i) conduct climate and observed runoff data quality checking with a set of selectable criteria, (ii) calibrate rainfall runoff models with multiple calibration optimisation algorithms, (iii) determine the optimal parameter set for ungauged catchments based on gauged catchments using cross verification, and (iv) run both catchment based and cell based simulations for a set of selected catchments. The ICYME also includes a suite of visualisation and statistical tools to facilitate these key modelling activities. The current version of ICYME supports six rainfall runoff models (Sacramento, SimHyd, Ihacres, SMARG, AWBM and 2-tap AWBM) and three calibration optimisation algorithms (Genetic Algorithm, Shuffled Complex Evolution and Uniform Random Sampling). In the future more rainfall runoff models and calibration algorithms can be integrated into ICYME to increase its modelling capacity. ICYME runs on catchments which are spatially discretised into grids, typically 5km x 5km. These grids are also called the grid-cells. All calculations are done on these grids and lumped as the total catchment output. Data for each grid cell is stored in a particular directory and file structures. The usability and flexibility of ICYMF has attracted hydrological modellers' attention. Besides TasSY and SWWASY, it has also been used in the Northern Australia Sustainable Yields Project (NASY) and will be adopted in the Water Forecast and Prediction project in the Water Information Research and Development Alliance (WIRADA) between CSIRO and the Bureau of Meteorology (BOM) and the Condamine-Balonne project funded by CSIRO Water for Health Country Flagship.

Keywords: Catchment yield modelling, calibration, rainfall runoff model

1. BACKGROUND

Water resources are quite scarce for rural and urban use in Australia (Ladson, 2008). Climate change and social and economic development are likely to worsen this situation. Thus improvement of water resources planning and management is crucial for Australian communities, industries and governments. In November 2006, the Australian government funded the Murray-Darling Basin Sustainable Yields project (MDBSY). It was led by CSIRO and used the expertise from state government agencies and leading industry consultants. The MDBSY project is the first study to provide a detailed basin-scale assessment of the impact of climate change, catchment development and groundwater extraction on the availability and use of water resources in Murray-Darling Basin (CSIRO, 2008). The success of MDBSY has directly led to three new sustainable yields projects: the Tasmania Sustainable Yields Project (TasSY), the South-west Western Australia Sustainable Yields Project (SWWASY), and the Northern Australia Sustainable Yields Project (NASY), which will provide water managers and users with good estimates of current and likely future water yields for these three regions.

In order to be consistent with MDBSY, all these three sustainable yields projects follow the same methodology as used in MDBSY. The framework for MDBSY is shown in Figure 1 (CSIRO, 2007). The rainfall-runoff modelling, also known as catchment yield modelling, involves a number of complex processes including input data scaling and data quality checking, model calibration, parameter set generation for both gauged and ungauged catchment, simulation runs and post data analysis. In MDBSY, a suite of tools was developed and driven by a series of Python scripts to perform these modelling processes. The tools were flexible, but it was quite hard for modellers with non IT background to run them. This paper
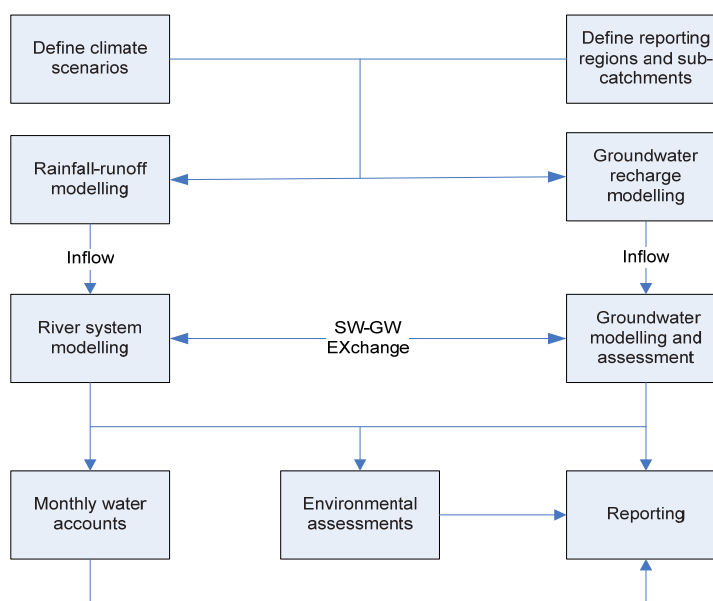

Figure 1. MDBSY Framework

describes the development of an integrated catchment yield modelling environment (ICYME), which was applied to both TasSY and SWWASY. It is built on the toolset developed in MDBSY. With its user friendly GUI, modellers can easily (i) conduct climate and observed runoff data quality checking with a set of predefined selectable criteria, (ii) calibrate rainfall runoff models with multiple calibration optimisation algorithms, (iii) determine the optimal parameter set for ungauged catchments based on gauged catchments using cross verification and, (iv) run both catchment based and cell based simulations for a set of selected catchments. ICYME also integrates a suite of visualisation and statistical tools to facilitate these key modelling activities and the model results analysis. Current version of ICYME supports six rainfall runoff models including Sacramento (Burnash, 1995), SimHyd (Chiew *et al.*, 2002), Ihacres (Jakeman *et al.*, 1990), SMAR (Kachroo, 1992), AWBM (Boughton, 1993 and Boughton, 2004) and 2-tap AWBM (a special modified version of AWBM used by Hydro Tasmania) and three calibration optimisation algorithms which are: Genetic Algorithm, Shuffled Complex Evolution (Duan *et al.*, 1992, Duan, *et al.*, 1993) and Uniform Random Sampling. In the future, more rainfall runoff models and calibration algorithms can be integrated into ICYME. ICYME runs on catchments which are spatially discretised into grids, typically 5km x 5km. These grids are also called the grid-cells. All calculations are done on these grids and lumped as the total catchment output. Data for each grid cell is stored in directory and file structures described later in the paper. The usability and flexibility of ICYMF has already attracted hydrological modellers' attention. Besides TasSY and SWWASY, it has also been used in NASY and will be adopted in the Water Forecast and Prediction project in the Water Information Research and Development Alliance (WIRADA) between CSIRO and BOM and Condamine-Balonne project funded by CSIRO Water for Health Country Flagship.

The rest of the paper is organised as follows. The next section is an overview of catchment yield modelling processes supported in ICYME followed by the file structure and data flow adopted in ICYME. Then how to set up ICYME is described. Finally conclusions are drawn and future work is discussed.

Yang, An integrated catchment yield modelling environment

2. OVERVIEW OF CATCHMENT YIELD MODELLING PROCESSES IN ICYME

2.1. Input data quality checking

The process of input data quality checking is to check if the input data including both climate and rainfall runoff data is good enough to do further modelling tasks, e.g. calibration and simulation. The following criteria are used:

a. Annual: if runoff is less than rainfall for all years (i.e., runoff coefficient < 1)
b. Annual: if evaporation is within a given error tolerance of the long-term mean for all years
c. Annual: if runoff is less than (rainfall – evaporation)
d. Monthly: if runoff is less than rainfall for all months (i.e., runoff coefficient < 1)
e. Monthly: if evaporation is within a given error tolerance of the long-term monthly mean for all months

The modeller may set both tolerances via a friendly GUI. The criteria a and d are compulsory and others are optional. After checking, a summary will be shown to modellers if data breaks the criteria or everything is good for further analysis.

2.2. Calibration

Calibration is the process to search for the optimal parameter set of rainfall-runoff models based on matching of observed and simulated data. In ICYME it is done in two steps: calibration configuration generation and calibration. This process requires three types of data:

- Observed runoff data – The data should be presented on a daily time-step as in a comma separated file (CSV) without headers named by the gauge ID.
- Historical climate data – The data should also be presented on a daily time-step. Each cell is listed in a comma separated file (CSV) with headers named by the grid cell ID (e.g. 1399.csv).
- Cell-catchment mapping – This data specifies which grid cell lies within which catchment and is given in a comma separated file (CSV) without headers.

The calibration is done using the optimisation of an objective function. The objective function of calibration can be chosen from the following three functions:

$$Obj_1 = E_{NS} \qquad Obj_2 = E_{NS} + 5 \times \left| \ln\left(\frac{Max\left(\sum P, 0.01\right)}{\sum O} \right) \right|^{2.5}$$

$$Obj_3 = \begin{cases} E_{NS}, & \left| \frac{\sum P}{\sum O} - 1 \right| < 0.05 \\ E_{NS} - 2000 \times \left(\left| \frac{\sum P}{\sum O} - 1 \right| - 0.05 \right), & otherwise \end{cases}$$

Where E_{NS} is the daily Nash-Sutcliffe efficiency, $\sum P$ is the sum of predicted runoff and $\sum O$ is the sum of observed runoff. . The first objective function is based on the daily Nash-Sutcliffe efficiency alone. The second is a combination of a smooth, less severe bias constraint (the log-bias constraint) with the daily Nash-Sutcliffe efficiency while the third combines the daily Nash-Sutcliffe efficiency with a severe bias constraint which attempts to ensure that the total predicted stream flow is within 5 % of the total observed stream flow (the bucket constraint). Please refer to Viney et al. (2009) for further details.

2.3. Cross verification

Cross verification is the process to determine which parameter sets from the gauged and calibrated catchments is the best suited to ungauged catchments for runoff prediction. It requires a regionisation mapping file, which specifies the parameter sets of a catchment to be transferred to another catchment. This information is supplied in a comma separated file (CSV) without headers. There are three columns: receiver catchment ID, sender catchment ID, distance order. The closest pair (receiver and sender) for each receiver is denoted as 1. The second closest is 2 and so forth.

2.4. Simulation

ICYME supports two types of runoff prediction: cell based simulation and catchment based simulation. The cells based simulation gives runoff generated for each grid cells, while the catchment based simulation produces the runoff data that is aggregated for whole catchment. The simulation process requires three types of data:

- Scenario climate data – The data is in the same format as historical climate data used in calibration.
- Cell parameterisation file – It specifies the parameter set from which calibrated catchment is to be used for grid cells in simulation. It is in a comma separated file (CSV) with two columns: grid cell Id and catchment ID.
- Catchment-report region mapping – It specifies which catchment is within which report region in a comma separated file (CSV) without headers. It contains two columns: catchment ID, report region name.

3. FILE STRUCTURE AND DATA FLOW

ICYME is strictly based on a particular file structure which consists of four main folders: Inputs, Calibrations, CrossVerifications and Simulations (Figure 2). The folder of "Inputs" and its structure is prerequisite for conducting any modelling activity in ICYME. Other folders are generated by the ICYME during modelling process to store outputs generated from each modelling process. The structure of each folder is described below. The inside of "<" and ">" need to be replaced with real name. For example the CalibrationAlgorithmName can be "ShuffledComplexRosenbrock" and RainfallRunoffModelName can be "Sacramento".
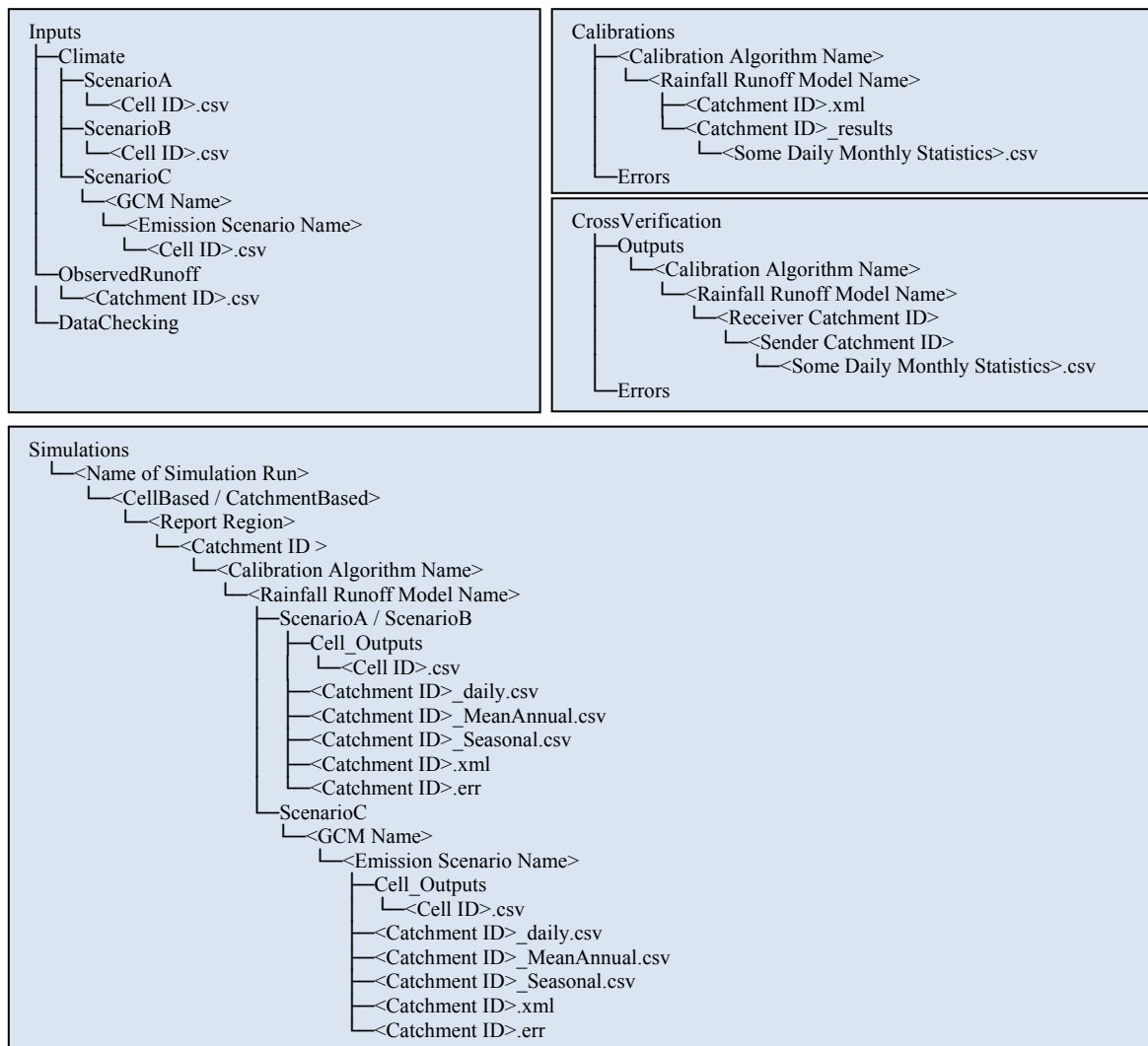

Figure 2. File structure in ICYME

The "Inputs" folder stores all data from hydrologists, including climate data and observed runoff data. There is a subfolder "DataChecking" under "Inputs", used to store the input data quality checking summary.

All data are stored as files with formats of CSV, TXT or XML in ICYME. Figure 3 shows how data is generated in each modelling process. Calibration and simulation is defined by an xml configuration file. The output of calibration, cross verification and simulation is a set of time series stored as a set of CSV files.
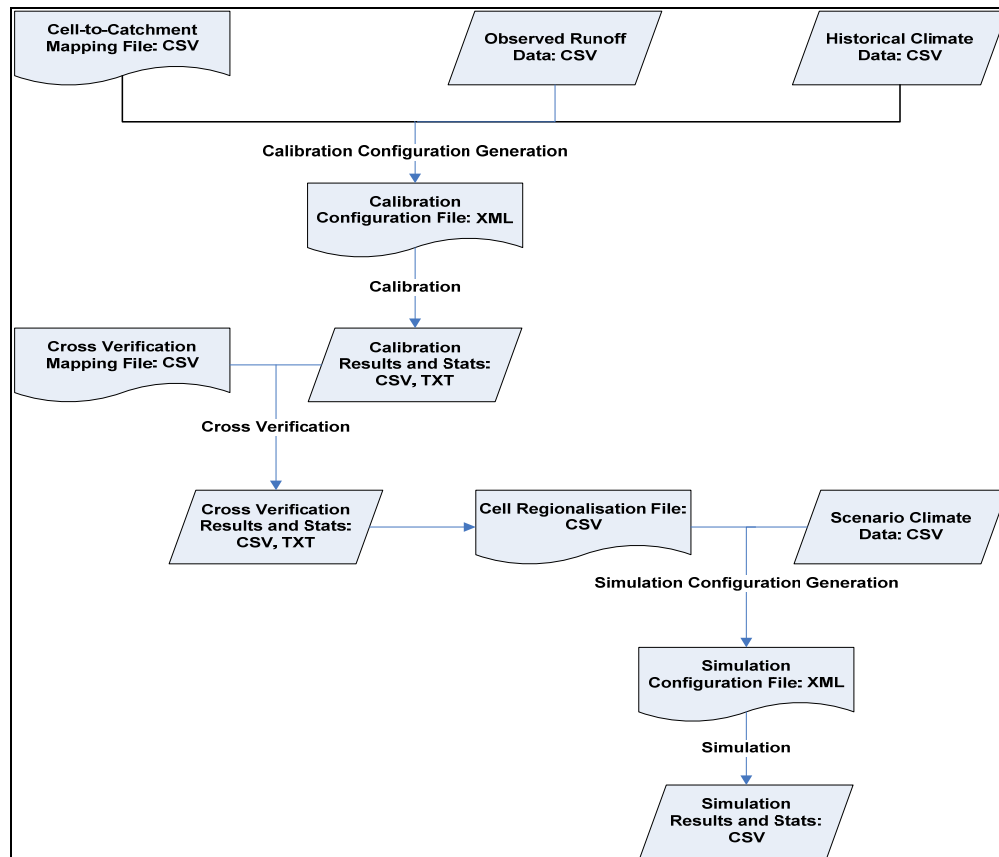
Figure 3. Data flow diagram in ICYME

4.   SETUP OF ICYME

Figure 4 is the main interface of ICYME. The left side is the list of catchments in the current modelling area. It provides information of how many grid cells in that catchment and what the current status is for that catchment. The upper right side is a map of that modelling area. A predefined colour schema is used to graphically show modelling status of each catchment. The bottom right side is a log panel which shows what is happening during modelling process.

Each project in ICYME may consist of multiple modelling areas, each of which may contain multiple catchments. Each catchment consists of a set of grid cells. When a new modelling area is added into a project via Project menu, it has to be set up via a modelling area configuration window which includes four configuration panels (Figure 5). The General configuration panel allows modellers to provide general information for this modelling area, such as its name, shape file, where to store data and where the cell-catchment mapping file is. Modellers also use this panel to set up criteria for input data checking. The Calibration panel allows modellers to set up calibration period, choose calibration objective function, select calibration algorithm and rainfall runoff model. Modellers only need to provide where the regionalisation file is for cross verification in the Cross Verification panel. The Simulation panel allows modellers to set up simulation period, where the cell parameterisation file and catchment-report region mapping file are, to select simulation scenario, rainfall runoff model and the parameter set from which calibration algorithm. All the configuration information is saved in an xml file.

After completing configuration, modellers may select catchments and run modelling process in a desktop PC or on a centralized cluster via the Run menu. Under the Tools menu, there are a suite of visualisation and statistics tools, which may facilitate modelling activities and post data analysis.
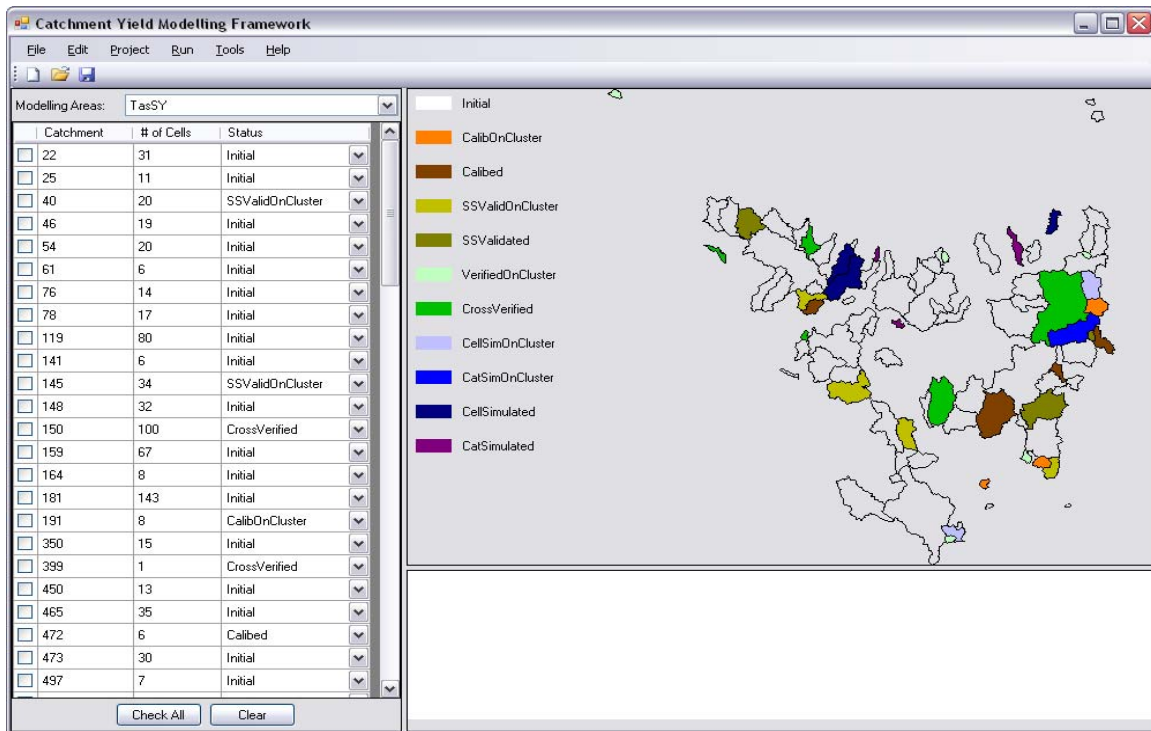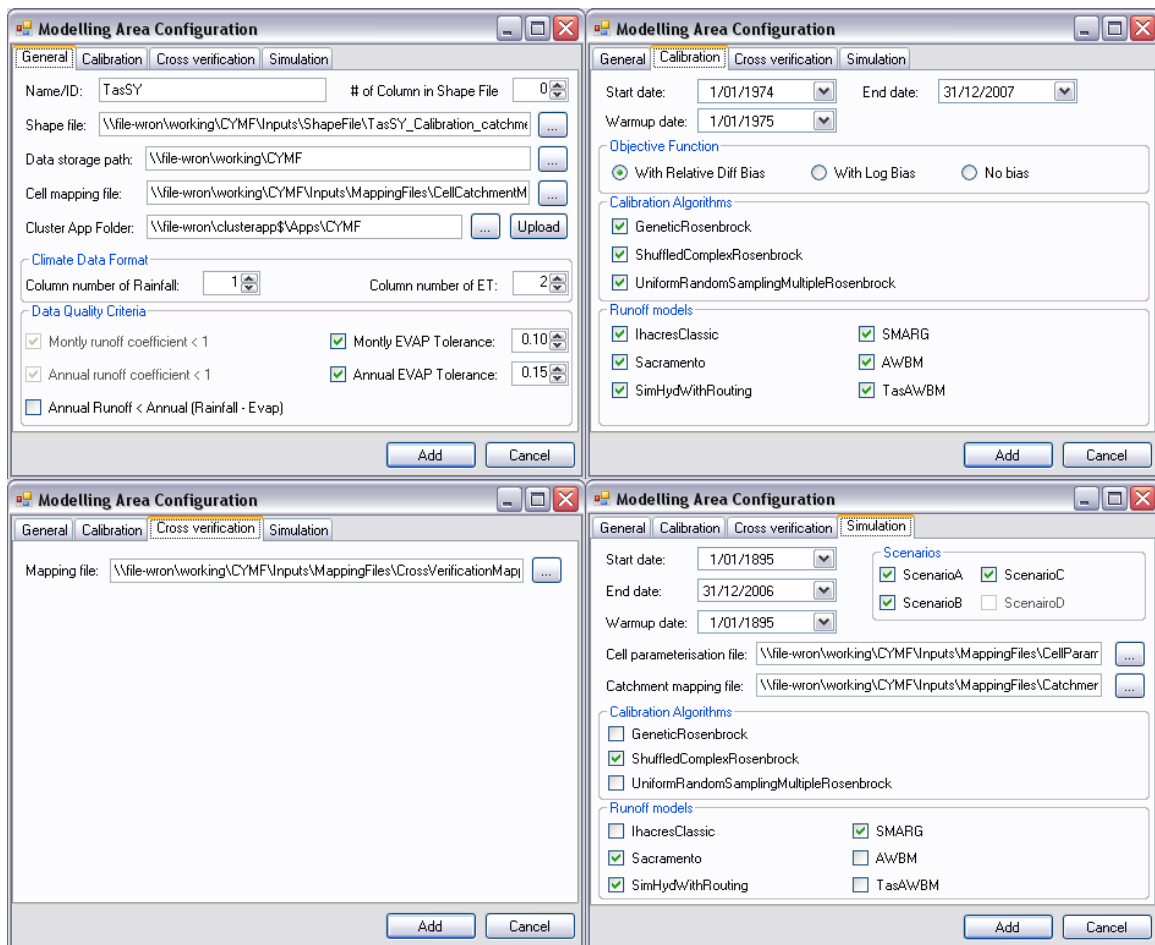
Figure 4. Main interface of ICYME

Figure 5. Configuration panel

5. DISCUSSION AND FUTURE WORK

ICYME creates a friendly window based modelling environment where modellers without any programming background can easily conduct most rainfall runoff modelling activities, e.g. data checking, calibration, cross verification and simulation. All modelling activities can be run either on a desktop PC or on a cluster. With a predefined colour schema and logging system, modellers can quickly capture the modelling progress of each catchment. The integrated visualisation and statistical tools can largely facilitate the post-analysis for modellers. These unique features combined with its usability and flexibility have attracted hydrological modellers' attention. The successful application of ICYME in three regional sustainable yield projects (TasSY, SWWASY and NASY) has shown that ICYME is a valuable tool in catchment yield modelling.

The application in sustainable yields projects has also suggested ways to improve it. ICYME can be enhanced in the following ways:

- More rainfall runoff models can be integrated into it in the future. Alternatively a very flexible plug-in system may be developed to allow the modeller to easily plug a new model into ICYME;
- More calibration algorithms may be added. At present the range of each calibrated parameter of calibration algorithms and the value of parameters of each search algorithm itself are hard coded. In the future, they should be able to be tuned by modellers;
- More data analysis functionalities can be integrated into ICYME. The visualisation tool can also be enhanced so that modellers may easily capture the pattern during data analysis.
- All data are stored as files in ICYME. The file system is simple and straight forward. However, it produces a huge amount of files. It is very hard for maintenance, management and manipulation. It also introduces a huge amount of IO activities which dramatically increase the run time. Using a database may be a good approach to solve these issues and improve its efficiency.

ACKNOWLEDGMENTS

REFERENCES

Boughton, W.C. (1993). A hydrograph-based model for estimating the water yield of ungauged catchments. In: *Proceedings of the 1993 Hydrology and Water Resources Conference*, Institution of Engineers, Australia, National Conference Publication no. 93/14, 317-324.

Boughton, W. C. (2004), The Australian water balance model. *Environmental Modelling and Software.* 19, 943-56.

Burnash, R.J.C. (1995). The NWS river forecast system – catchment modeling. In: Singh, V.P., (Ed). *Computer Models of Watershed Hydrology*, Water Resources Publications, Colorado, USA. 311-366.

CSIRO (2007). Overview of project methods. *A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project*. CSIRO, Australia

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.

Chiew, F. H. S., Peel, M. C. and Western, A. W. (2002), Application and testing of the simple rainfall-runoff model SYMHYD. In: Singh, V.P. and Frevert, D.K. (eds). *Mathematical models of small watershed Hydrology and applications*. Water Resources Publications, Littleton CO, USA. 335-67.

Duan, Q.Y., Sorooshianan, S. and Gupta, V.K. (1992), Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*. 28(4), 1015-1031.

Duan, Q.Y., Gupta, V.K. and Sorooshian, S. (1993). Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *Journal of Optimization Theory and Applications*. 76(3), 501-521.

Jakeman, A.J., Littlewood, I.G. and Whitehead, P.G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology*. 117, 275-300.

Kachroo, R.K. (1992). River flow forecasting. Part 5. Applications of a conceptual model. *Journal of Hydrology*. 133, 141-178

Ladson, A. (2008). Hydrology: an Australian introduction. Oxford University Press. New York.

Viney, N.R., Perraud, J-M, Vaze, J., Chiew, F.H.s., and Yang. A. (2009). The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments. The 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation. Cairns, Australia.