Integration and application of the Rainfall Runoff Library


*Associate Professor, Dept. of Environ. Eng., Pusan National University, San 30, Jangjundong, Gumjunggu, Pusan, Republic of Korea, 609-735 (E-mail: kimsangh@pnu.edu)

**Chief, CSIRO Land and Water, F. C. Pye Laboratory, Clunies Ross Street, Action, ACT 2602 GPO Box 1666, Canberra ACT 2601 Australia (E-mail: rob.vertessy@csiro.au)

***Project Researcher, CSIRO Land and Water (jean-michel.perraud@csiro.au)

****Dept. of Environ. Eng., Pusan National University, San 30, Jangjundong, Gumjunggu, Pusan, Republic of Korea, 609-735 (E-mail: tjd0102@hanmail.net)

Abstract The Rainfall Runoff Library (RRL) provides a convenient platform for implementing environment modelling components such as rainfall runoff models, calibration tools, and objective functions. A rainfall–runoff model widely known and used in South Korea, TANK, is added to the RRL, and used along with the models AWBM and SIMHYD to reproduce the historical time series of daily and monthly runoff at the Soyanggang Dam and Youngcheon Dam catchments located in South Korea. The features of the RRL allow for an easy comparison of different models in a standardised and common framework. Three optimisation methods (Genetic algorithm, Rosenbrock method and Shuffled Complex Evolution algorithm) were applied to calibrate the model parameters using three different objective functions. The applicability of each model to these catchments is discussed based on the resulting statistics.

Keywords Rainfall Runoff library; modelling and decision support; calibration; TANK

Introduction

The modelling of the rainfall–runoff relationship allows not only for the simulation of the soil moisture levels and water yield as such but it is also often a pre-requisite for the modelling of other processes at the catchment scale. Many rainfall–runoff models, most of them lumped and conceptual, have been created in the past few decades. The applicability of a given model depends not only on the robustness of the model structure but also on the variability in the behaviour of the catchments (Perrin et al., 2001). It is consequently highly desirable to apply several models to obtain a comprehensive insight into the hydrological behaviour of the catchment that is studied.

These rainfall–runoff models were often implemented as monolithic software packages with different software requirements, data formats, and user interfaces. Practitioners willing to trial several models on a given study area will thus not only have to gain a substantial understanding of each model structure but also to maintain and use several software tools and specific data sets, the latter making this comparative modelling exercise labour intensive. To address this issue, the Cooperative Research Centre for Catchment Hydrology (CRCCH) has built an extendible library of lumped conceptual rainfall–runoff models, known as the Rainfall Runoff Library (RRL), as part of its Catchment Modelling Toolkit, and made it freely available as a downloadable, stand-alone product (Podger, 2003). The RRL is built upon the CRCCH environmental modelling framework, The Invisible Modelling Environment TIME (Rahman et al., 2003). While the RRL is not meant to be a tool for model development per se, it is designed such that the effort that is required to add a new model or a calibration tool is small. This is achieved by using one of the distinctive features that TIME offers compared to other
existing modelling frameworks, the use of metadata to describe, manage and run its models.

This paper illustrates how the RRL is used in a case study of two catchments in South Korea. A rainfall–runoff model widely known in South Korea and Japan, TANK (Sugawara, 1995), was added to the library. TANK and two other models in the library were then used to reproduce the historical daily and monthly flow records of these catchments. These models were calibrated against observed runoff using several calibration tools and objective functions, as a basis for comparing the model performances.

Materials and methods

Features of the RRL

The RRL is chosen for the comparative study described in the second part of this paper primarily for the following features:

1. It is designed from the ground up to facilitate the addition of a new model, as was needed for this case study
2. It can read and write time series in several existing data file formats
3. For each modelling exercise on a given catchment, the characteristics and options used in the calibration, and the resulting model parameters can be tracked, saved as a project for later retrieval to ensure the reproducibility of the results, irrespective of the specific model at hand
4. It does not rely on any third party software of a commercial nature
5. It offers a unified user interface, tailored to rainfall–runoff modelling, but at the same time as model-neutral as possible. It contains generic and advanced visualisation features deemed useful for rainfall–runoff modelling (log–log scatter plot, log-normal scale for the duration curve)
6. Being developed upon the TIME framework, the models and other software components developed for the RRL are directly reusable in other existing and future products of the CRCCH Catchment Modelling Toolkit

The implementation of these features relies heavily on the use of so-called attributes in addition to traditional object-oriented software engineering (Perraud et al., 2003). Attributes, also referred to as metadata, are a feature of some recent programming languages. The use of metadata is illustrated in more detail in the section describing the addition of TANK to the RRL.

Models in the RRL

The RRL is a library of lumped conceptual rainfall–runoff models. It currently includes five models, of which three are often used in Australia: the Australian Water Balance Model AWBM (Boughton and Mein, 1996), SACRAMENTO (Burnash et al., 1973), and SIMHYD (Chiew et al., 2002).

The AWBM model is a water budget model which assumes a functional relationship between rainfall and runoff using three parallel storages for evaporation and surface storage, with their respective partial catchment areas and capacities, and two additional storages for surface and subsurface runoff depths (Figure 1(a)). The effective rainfall from the three surface storages is partitioned between the two linear storages proportionally to a base flow index parameter. AWBM has up to eight parameters, although those controlling the partial areas are fixed by default in accordance with the latest version devised by the author, W. Boughton (personal communication).

SACRAMENTO is the rainfall–runoff model of choice used in the Integrated Quantity and Quality Model (IQQM) of the New South Wales Department of Infrastructure, Planning and Natural Resources (DIPNR). It is worth noting that the code of the model,
courtesy of DIPNR, could be kept in its original Fortran language since RRL runs on top of the Microsoft.NET framework which facilitates multi-language integration. Many studies used this 17 parameter model for the design and evaluation of calibration methods.

SYMHYD consists of three storages: the interception storage, the soil moisture storage and the groundwater storage. Infiltration excess and saturation excess are modelled to represent surface and interflow runoff (Figure 1(b)). The version of SIMHYD included in the RRL is actually a nine parameter variant of the model used in the Environmental Management Decision Support System (Chiew et al., 2002). Two additional parameters allow for the explicit modelling of an impervious area.

Integration of TANK

TANK (Figure 1(c)) is a conceptual rainfall runoff model introducing four reservoirs to describe the hydrologic processes of a catchment (Sugawara, 1995). Each of the tanks can have an outflow and evapotranspiration as outputs. The runoff from tank 1 is the surface discharge, and the intermediate, sub-base and base discharges are generated respectively from tank 2, 3, and 4. Seventeen parameters were used to describe the model structure, though four of them only deal with antecedent moisture conditions and are more initial conditions than true parameters in effect.

Models implemented in TIME and the RRL are designed to be as small as possible, and thus easy to code. Usually, only the correct declaration of inputs, outputs, parameters and the core algorithm over one time step are required in order to have a fully functioning model. For the RRL, the only additional software requirement is to “inherit” from a generic Rainfall Runoff Model, to allow for the correct automatic detection of additional
models by the program. The handling of data files, time series, time step and other temporal aspects is already taken care of generically for all models. As previously mentioned, metadata is key to the extendibility of the library. Metadata may for instance document a variable as a parameter, with a specified minimum and maximum, a summary explaining its role, and even a constraint dependent on another parameter (e.g. H11 of the Tank model). Besides the obvious usefulness as documentation embedded in the code, other software components in TIME (e.g. calibration tools) will inspect that metadata and extract information to correctly handle any present or future rainfall–runoff model.

**Calibration tools.** An abundant literature is testimony to the difficulty of the calibration of models. The RRL thus includes several features designed to ease the task of calibrating the rainfall–runoff models.

It is possible to classify calibration tools into two categories, besides manual calibration. Generic calibration tools will work on the vast majority of models because they will handle each model parameter indifferently and do not need to know the model structure. Custom calibration tools will work with a specific model only because their algorithm does discriminate between model parameters, e.g. by treating storage depths differently from the recession constants. The RRL currently has four generic calibration tools: the Pattern Search from [Hooke and Jeeves (1961)](https://www.jstor.org/stable/1588250), the Genetic Algorithm (GA) ([Wang, 1991](https://www.worldcat.org/title/optimization-by-genetic-algorithms/oclc/15974741)), the Shuffled Complex Evolution (SCE-UA), based on [Duan et al., (1994)](https://onlinelibrary.wiley.com/doi/abs/10.1002/eqge.560180703), and the Rosenbrock search method ([Rosenbrock, 1960](https://www.jstor.org/stable/1975261)).

The Rosenbrock method is a local search method, similar in principle to a downhill descent search. Its algorithm is superior to the pattern search in that it adapts its step size and search direction to the local conditions around a point in the parameter space. The pattern search or Rosenbrock method are especially handy as a supplement to global optimisation methods which do not always converge enough towards the optimum.

The GA is a powerful search tool that utilises evolutionary-based principles to find optimal solutions. The GA randomly samples parameter sets in the first generation. Three processes, the reproduction, crossover and mutation are used to generate a new population of solutions. The GA optimiser relies mostly on two main principles, as described in [Wang, (1991)](https://www.worldcat.org/title/optimization-by-genetic-algorithms/oclc/15974741). Parameter values are coded into a binary scheme (the genes), and the parameter sets are selected to produce offspring proportionally to a trapezoidal discrete probability density where the best parameter sets have the highest probability so that the search is less likely to become trapped in a non-global optimum.

The Shuffled Complex Evolution algorithm has been the subject of several papers ([Sorooshian et al., 1993](https://link.springer.com/article/10.1007/BF02579744); [Thyer et al., 1999](https://link.springer.com/article/10.1007%2FS10113-003-5405-0)), and is probably the most popular global search method in rainfall–runoff modelling, since several studies suggest that this method performs quite consistently well on non-linear models. To some extent, principles similar to those in the GA can be found in the SCE-UA: it uses a trapezoidal probability density function for selection, and the shuffling of the populations of points bears some similarities with the generation of an offspring.

The RRL offers simple but useful tools for the detection of very wet or dry years in the recorded annual runoff, the definition of a calibration and verification period, and appropriate warm-up periods, in order to improve the robustness of the calibration.

**Objective functions for calibration.** The generic optimisers can calibrate models against existing data using a variety of objective functions and options. There is an option to calibrate with a dual objective, although it should be noted that the internal optimisation engine can handle any number of objectives for possible use in future versions. In this
study, three types of objective functions were used for calibration and validation of model performance. The sum of squares of difference between simulated and observed runoff will be noted in the rest of this paper as

\[ OBJ_1 = \sum_{i=1}^{n} (SIM_i - REC_i)^2 \]  

(1)

where \( SIM \) is the simulated and \( REC \) is the recorded runoff.

The absolute summation of the difference between simulated and observed runoff will be noted as

\[ OBJ_2 = \sum_{i=1}^{n} |SIM_i - REC_i| \]  

(2)

The Nash–Sutcliffe coefficient (Nash and Sutcliffe, 1970), efficiency, is defined as

\[ OBJ_3 = \frac{\sum_{i=1}^{n} (REC_i - \overline{REC})^2 - \sum_{i=1}^{n} (SIM_i - REC_i)^2}{\sum_{i=1}^{n} (REC_i - \overline{REC})^2} \]  

(3)

where \( \overline{REC} \) is the average of recorded runoff.

Furthermore, the models can be calibrated on monthly flows as is usual in studies where the determination of the overall water yield is the purpose or on the duration curves in order to avoid the effects of non-synchronised input time series for calculated and observed runoff.

Description of the case study. Three models, SIMHYD, AWBM and TANK, from the RRL have been applied to simulate the monthly and daily historical records of two catchments in Korea. The Soyanggang Dam catchment is located upstream of the Han River on the north-western parts of South Korea. The catchment area is 2703 km\(^2\), the mean annual rainfall is 1168 mm/year and the mean annual potential evaporation is 658 mm/year. Elevation ranges from 192 m to 1660 m above sea level with slope varying from 0 to 73%. More than 90% of the catchment is forested area and 7% of land is used for agricultural purposes. The temperature varies from \(-11.4\) to 31.5 degrees Celsius over the year. 6 years recorded hydrologic time series from 1986 to 1992 were used to calibrate model parameters and validations of models were carried out using 4 years data between 1993 and 1996. The Youngchun Dam catchment is one of the primary headwaters of Nackdong river located in the middle west part of South Korea. The catchment area is 3901 km\(^2\) and the mean rainfall and evaporation are 1101 mm/year and 524 mm/year, respectively. Digital elevation analysis provides elevations varying between 138 m and 1116 m above sea level. The mean slope is 20.5% and the maximum slope is 55.37%. Approximately 85% of the land use is forest and 10% of the catchment is rural area. Rainfall and runoff data between 1983 and 1986 were used in calibrations of models and the time series from 1986 to 1988 were used in the validation process. All periods have a warm-up of one year before data is included in the calculation of the objective functions.

The performances of models or optimisation methods can be compared by a unified evaluation standard, \( OBJ_3 \), often referred to as efficiency. Table 1 shows the daily and monthly simulation results of the Soyanggang Dam catchment by various models, objective functions and optimisation methods. The daily calibration efficiencies vary from 0.672 to 0.867 and monthly calibration ranges between 0.938 and 0.955. The verification results improve daily and monthly efficiencies as 0.734 to 0.889 and 0.98 to 0.99, respectively. Figure 2 shows rainfall time series, observed and verified runoff time series at the Soyanggang Dam catchment. The simulation results of Youngchun Dam catchment...
are shown in Table 2. Efficiencies of daily calibrations between 0.503 and 0.614 are also improved in the verification process as 0.714 to 0.822. Monthly calibration efficiencies are distributed from 0.728 to 0.766 and verification results are distributed between 0.639 and 0.892. The consistent increase of efficiencies of verification may be explained by improved accuracy of measurements in later verification periods.

Statistical analysis of the simulations provides a comprehensive insight into different methodologies in the catchment modelling exercise. Table 3 presents the averages and standard deviations of simulation efficiencies out of different combinations of models, optimisation methods and objective functions both in the calibration and verification processes. Statistics of daily flow simulation indicate that the TANK model produces the highest average and lowest standard deviation of efficiencies. No significant difference can be found on daily runoff calibration and verification between SIMHYD and AWBM. Monthly flow calibration and verification statistics indicate that the capabilities of runoff calculation among the three models are almost similar. The more complicated structure
of TANK (15 parameters) shows better simulation efficiencies in the daily flow simulation than SIMHYD (9 parameters) or AWBM (8 parameters).

Considering the parsimony of model structures and model performance, TANK is recommended for daily flow simulation and SIMHYD and AWBM are more appropriate for monthly flow generation. GA and SCE-UA produce similar calibration and verification results. A directional searching method, the multi-start Rosenbrock method, shows a performance similar to other global search optimisation methods. The statistical differences of efficiencies among OBJ1, OBJ2 and OBJ3 are small enough to ignore any preference in choosing a specific objective function on these two catchments. However,
the dimensionless formulation of equation (3) is desirable to minimise the impact of scale on the calibration results.

Conclusions
The RRL contains models used primarily on Australian catchments. It is however designed as a flexible framework and allowed for the easy addition of TANK, a model widely used in Korea and Japan. TANK, AWBM and SIMHYD were then applied to two Korean catchments with a combination of three optimisation methods and three objective functions. A comprehensive evaluation of different modelling methodologies could thus be achieved, using a consistent framework. The coefficient of efficiency was increased between the calibration and verification periods in most cases, especially for the well-monitored Soyanggang Dam catchment. For the two catchments, TANK is the most appropriate model for the simulation of the daily runoff while SIMHYD and AWBM are recommended for monthly runoff evaluation since their simpler structure still achieves roughly the same efficiency as TANK.

Acknowledgements
This research was supported by a grant from Brainpool Program from Korean Federation of Science and Technology Societies and Brain Busan 21 Program.

References
Cooperative Research Centre for Catchment Hydrology, Canberra, AU.