



Research papers

Application of random number generators in genetic algorithms to improve rainfall-runoff modelling



Martin Chlumecký^{a,*}, Josef Buchtele^b, Karel Richta^a

^a CTU FEE, Department of Computer Science, Prague 166 27, Czech Republic

^b CAS CZ, Institute of Hydrodynamics, Prague 166 12, Czech Republic

ARTICLE INFO

Article history:

Received 24 March 2017

Received in revised form 20 June 2017

Accepted 15 August 2017

Available online 16 August 2017

This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Corrado Corradini, Associate Editor

Keywords:

Genetic algorithm

Optimisation

Rainfall-runoff modeling

Random generator

ABSTRACT

The efficient calibration of rainfall-runoff models is a difficult issue, even for experienced hydrologists. Therefore, fast and high-quality model calibration is a valuable improvement. This paper describes a novel methodology and software for the optimisation of a rainfall-runoff modelling using a genetic algorithm (GA) with a newly prepared concept of a random number generator (HRNG), which is the core of the optimisation. The GA estimates model parameters using evolutionary principles, which requires a quality number generator. The new HRNG generates random numbers based on hydrological information and it provides better numbers compared to pure software generators. The GA enhances the model calibration very well and the goal is to optimise the calibration of the model with a minimum of user interaction. This article focuses on improving the internal structure of the GA, which is shielded from the user. The results that we obtained indicate that the HRNG provides a stable trend in the output quality of the model, despite various configurations of the GA. In contrast to previous research, the HRNG speeds up the calibration of the model and offers an improvement of rainfall-runoff modelling.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Hydrological data contain large amounts of useful information, which can be fully utilised by rainfall-runoff models that describe the hydrological processes in nature in mathematical terms (Singh and Woolhiser, 2002). Rainfall-runoff models are defined as an interaction of the physical processes that describe the structure and behaviour of observed basins (Beven, 2012). The outputs of well-calibrated models are applied, for instance, in the prediction of flash floods or in assessments of water consumption in vegetation cover. The rainfall-runoff models are primarily designed to estimate basin parameters where measurement is scarcely feasible or even impossible (Robertson and Kirkegaard, 2006).

This article focuses on the Sacramento soil moisture accounting model (SAC-SMA), which is a conceptual water balance physical model that is based on the principles of water movement in a watershed (Burnash, 1995). Calibration configures the behaviour of the modelled catchment based on historical data (Sorooshian et al., 1993). This suggests that the calibration of the model is an important part of modelling and that the credibility of the output is closely associated with the user's experience of the model (Smith et al., 2003).

Optimisation algorithms make the model calibration easier for users with limited technical knowledge. These algorithms enhance the quality of manual calibration automatically, giving a significant improvement (Tolson and Shoemaker, 2007). Genetic algorithms (GAs) are the most common group of optimisation algorithms for rainfall-runoff models, and are based on the principles of evolution (Holland, 1992). The majority of GA research and rainfall-runoff models try to achieve better results by modifying the GA parameters or through various multi-step procedures which calibrate model parameters sequentially in a precisely defined sequence. Gupta et al. (2003) use a vector of estimated model output generated using model parameters. The goal is to find the vector with the best estimate of the parameters so that the observed and modelled outputs are as close as possible. Vector items represent various statistical functions, which determine the quality of a calibrated model. Another technique is to select a function that specifies the quality of a calibration. This function helps GA to find the best solution. The function is selected according to statistical indicators to obtain the best output (Kisi et al., 2013).

Current optimisation techniques are not user-friendly, especially for hydrologists who do not have the necessary background in the software architecture of the GA framework. Users of these optimisation techniques/frameworks must learn their application or architecture, and this is usually tedious, complex, and discouraging. Moreover, this is exacerbated whenever the application

* Corresponding author.

E-mail address: chlumma1@fel.cvut.cz (M. Chlumecký).

has to be reviewed; that is, for each different period or basin. Incorrect parameter settings can cause needless deterioration in the quality of the model. Therefore, the optimisation technique should require minimal user configuration in order to be useful (Fonseca and Fleming, 1995).

The primary motivation of this article is to improve the calibration of the model by the genetic algorithm with minimal user configuration of the genetic algorithm. This should enable a faster and more efficient utilisation of the information stored in the environmental data, which can be extracted by the rainfall-runoff model with the best calibration. Although the internal architecture of the genetic algorithm must be modified to achieve this, any intervention in the SAC-SMA model must be minimal to ensure that the user does not need to understand this additional configuration of the model. This approach provides an important improvement to the output of the SAC-SMA model.

The core of the genetic algorithm is a random number generator that estimates the values of the model's parameters. The generator usually produces numbers with a uniform probability distribution (Gallagher and Sambridge, 1994; Harrison, 2010). This paper shows that the distribution function of the generator influences the quality of the model output. We present a new concept of the genetic algorithm for optimising the SAC-SMA model with a new random number generator.

2. Methodological background

2.1. Sacramento soil moisture accounting model

Sacramento soil moisture accounting model is one of the most widely used rainfall-runoff models. This model is used to convert precipitation input into discharge outputs. The well-designed architecture of SAC-SMA provides motivation for further research and development. This model imitates two-dimensional heterogeneous runoff processes for river catchments.

This model operates using a system of water reservoirs (zones). An ability of retention and draining water storage in the upper zone affects the infiltration of rainfall in the lower zone layers. Infiltrated water is then: (1) retained; (2) pumped by vegetation during evapotranspiration; (3) infiltrated into low-lying system zones; or (4) channelled into a river network. The total streamflow is the sum of all of the partial zone runoffs.

The SAC-SMA model can be configured by 29 model parameters. The first five model parameters and its five coefficients determine a rate of outflow from the model reservoirs. The next seven parameters define static attributes of modelled basin (e.g. afforestation coefficient). The 12 remaining parameters determine monthly adjustment factors for potential evapotranspiration. Most of the second kind of model parameters must be estimated using historical hydrological data, (Burnash, 1995). It should be noted that a maximum of 10 parameters effectively calibrate the model.

2.2. Validation of the model calibration

Validation of the model confirms the quality of the model calibration. Various statistical indicators are used for model validation (Zhang et al., 2012). This study primarily focuses on the statistical method and uses the six most fundamental indicators, as summarised below (Chai and Draxler, 2014):

1. Average absolute monthly volume error

$$AMVE = \frac{\sum_{i=1}^m |MS_i - MQ_i|}{\sum_{i=1}^m MQ_i} 100 \quad (1)$$

where M is number of months, MS is monthly simulated volume, MQ is monthly observed volume.

2. Daily average absolute error

$$DAE = \frac{1}{n} \sum_{i=1}^n |Q_i - S_i| \quad (2)$$

where S is simulated discharge, Q is observed discharge, n is number of events.

3. Root mean square error

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_i - S_i)^2} \quad (3)$$

4. Monthly bias

$$MB = \frac{8.64}{AREA} \frac{1}{n} \sum_{i=1}^n (S_i - Q_i) \quad (4)$$

5. Correlation coefficient

$$R = \frac{n \sum_{i=1}^n S_i Q_i - \sum_{i=1}^n S_i \sum_{i=1}^n Q_i}{\sqrt{(n \sum_{i=1}^n S_i^2 - (\sum_{i=1}^n S_i)^2)(n \sum_{i=1}^n Q_i^2 - (\sum_{i=1}^n Q_i)^2)}} \quad (5)$$

6. Nash-Sutcliffe model efficiency coefficient

$$E = \frac{\sum_{i=1}^n (Q_i - S_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \quad (6)$$

where, \bar{Q} is mean of observed discharge.

The units of all of the used indicators as described by Eqs. (1)–(6) are illustrated in Table 1.

2.3. Genetic algorithm

The genetic algorithm (GA) is an evolutionary technique that is inspired by processes in nature. The structure of the GA is based on Darwin's theory of natural selection. Each individual is a candidate for the solution of a given problem. A fitness function quantitatively expresses the quality of the solution. The task of the GA is to create a "breed" of the individual with the best fitness function. The GA evaluates a population which is composed of chromosomes, which contain data in a similar way to biological chromosomes (Gallagher and Sambridge, 1994). Each chromosome in the population is a set of genes which are coded using binary or real values. The GA attempts to discover the combination of genes which maximises or minimises the fitness function (Holland, 1992). To achieve this, the GA uses equivalents of natural genetic operators; for instance selection, crossover and mutation (Bäck et al., 2000).

New sets of chromosomes are appraised over a number of generations. A fitness value is calculated and assigned to each new estimated chromosome (Mitchell, 1998). The fitness value provides information about the quality of a chromosome and this is used for the probabilistic selection of chromosomes for the next generation. This selection is analogous to natural selection. The GA is terminated if predetermined criteria occur. The new chromosome is a result of genetic operations, the most significant of which are crossover and mutation. Both of these operations are random processes with a defined probability. The crossover takes two chromosomes as parents and produces a new chromosome by exchanging their chromosomes with each other. Parts of chromosomes for exchange are selected at random and can be determined at a single point or at multiple points. The GA can become trapped at a local minimum despite the crossover operation. The mutation operation can avoid this disadvantage because it changes the values of genes

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات