Advances in Water Resources 38 (2012) 81–91

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Calibration of hydrological models on hydrologically unusual events

Shailesh Kumar Singh [a,*], András Bárdossy [b]

[a] National Institute of Water and Atmospheric Research Ltd., Christchurch, New Zealand
[b] Institute of Hydraulic Engineering, University of Stuttgart, Stuttgart D-70569, Germany

ARTICLE INFO

Article history:
Received 11 January 2011
Received in revised form 12 December 2011
Accepted 16 December 2011
Available online 4 January 2012

Keywords:
Critical events
Conceptual hydrological model
Data depth function
HBV
HYMOD
ICE algorithm

ABSTRACT

The length of the observation period used for model calibration has a great influence on the identification of the model parameters. In this contribution it is shown that a relatively small number of so called unusual time periods are sufficient to specify the model parameters with the same certainty as using the whole observation period. The unusual events are identified from discharge or precipitation observations series using the statistical concept of data depth. The idea is to distinguish between model states which are covered by previously observed states (interpolation case), and those for which no similar events occurred (extrapolation case). Depth functions are used to identify unusual events from four days lagged discharge or API (antecedent precipitation index) series. Data with low depth are near the boundary of the multivariate set and are thus considered as unusual. The depth is calculated using the observations, their natural logarithms, their rank and their first differences. Model calibration using the selected critical periods is only slightly worse than using all data. The transferability of the parameters for different time periods is equally good as using all the data and significantly better than random selection. Two different models (HBV and HYMOD) are used to demonstrate the methodology for the Neckar catchment in South-West Germany. The methodology developed in this study can be potentially useful for developing monitoring strategies.

1. Introduction

Hydrological models are used for forecasting and water management, to provide information for decision making. Due to the simplification of the complex natural processes, the high spatial and temporal variability, and the limited availability of observations, the identification of the model parameters is a difficult task. Even physically-based models require parameter adjustments due to differences between the observation and modelling scales, and to a limited observability of certain variables and processes. Conceptual models require model calibration to estimate parameter values [1–3]. The non-uniqueness of the parameters makes this a challenging task.

Past observations of discharge and weather (temperature, precipitation, etc.) are used for calibration of hydrological models. The observation period might include floods, droughts and normal flow periods. It is assumed that the calibration of the model will only be considered successful if the observation period is representative of the hydrological behaviour of the catchment. The information contained in the observations with respect to the parameters is not uniformly distributed along the series. Certain time periods might be useful for the identification of specific parameters while others might be less important. For example, summer observations are of no use to identify parameters related to snow accumulation. Wagener et al. [4] shows that information contained in a data series are non-homogeneous.

The length and information content of a time series play a vital role in the parameter identification of a model. Several authors have investigated data requirements for the identification of stable model parameters [5–12,3]. Even so, it is very difficult to precisely define what length of data is sufficient to identify model parameters so that they can also be used for other time periods. Other investigators have reported one year to eight years of data collection as being sufficient to obtain robust parameters. However, it cannot be generalised because different models have different levels of complexity and different catchments have different information content in each year of hydrological record. Moreover, the information content of hydrological data is generally not known. Hence, we always use the whole data series so that the model uses as much information as possible to identify its parameters. There is a need to establish a method which can be used to identify the critical time periods in a given time series, which contain most of the information need to identify the model parameters.

Time periods with high hydrological variability may useful for calibration as they contain a lot of information for parameter identification [5]. In this study, a similar hypothesis was tested, assuming unusual events in a given series represent most of hydrological

* Corresponding author.
E-mail addresses: shaileshiitm@gmail.com, sk.singh@niwa.co.nz (S.K. Singh).

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.12.006

variability. These unusual events are also called *critical events* since they contain most of the relevant (critical) information to help identify model parameters.

The purpose of this study is to develop a simple and reliable technique to identify those time periods (events) that can best be used for model calibration. The hypothesis of the paper is that if a model works well in critical time periods it also work well for any other time periods. Critical events correspond to unusual circumstances such as the appearance of unusual sequences of rainfall amounts, or increases/decreases in discharge series or long dry periods. These events can be identified from the observations with the help of the statistical depth function.

The paper is organised as follows: After the introduction in Section 2 the methodology and algorithm are described. The case study area and the hydrological models used are presented in Section 3. The results are summarised in Section 4. A possible use of the methodology for ungauged catchments requires precipitation forecasts which then are used to predict the occurrence of future unusual events. This aspect is discussed in Section 5. The paper closes with a discussion of the results and conclusions.

2. Methodology

It is obvious that in many cases only a part of the data collected is important to identify model parameters. This is particularly true for linear or monotonic models. In order to illustrate this idea consider the following example. A linear regression is used to fit a straight line $y = ax + b$ to a data set $Z_n = \{(x_i, y_i);\ i = 1, \ldots, n\}$. This fit can be described with the parameter vector $\theta = (a, b)$ where a and b are slope and intercept, respectively. We use three different datasets to identify the parameters:

1. Case 1: All the n points
2. Case 2: Using a selected n^* *critical* points
3. Case 3: Using a selected n^*(same number as in Case 2) random points

The above experiment was carried out $M = 1000$ times. Each time a set of $n = 100$ points in two dimensional space was generated. For Case 2 ($n^* = 10$) the critical points were selected as those points which have the five largest and the five smallest x value, from the set of $n = 100$ points. In Case 3 $n^* = 10$ points were selected randomly from the set with $n = 100$ points. For each set a regression line was fitted using the least squares principle. This process was repeated 1000 times, where the random set of points was always generated from the same bi-variate normal distribution (with $r = 0.8$). The slopes of the fitted regressions are shown for the three cases in Fig. 1. It is clear that Case 1 (the whole dataset) and Case 2 (critical data) have very similar distributions, where as in Case 3 the slopes show a much higher scatter. This shows that careful selection of a subset of the data may result in parameters which are similar to those obtained from the whole dataset. The general concept is that we can fit better regression lines if we use more data. Hence, at first glance, the above example may appear to challenge this general concept. From this example, we can see that only about 10% of the data can give a regression almost as good as would be the case if we selected the data carefully. This simulation example can be further extended to real hydrological modelling.

2.1. Identification of critical time periods using data depth function

A time series of discharge or precipitation may have a large number of sequences which look very similar and can be regarded as a kind of repetition of previous events. Some, however, (among them the maxima and minima) differ from previous observations.

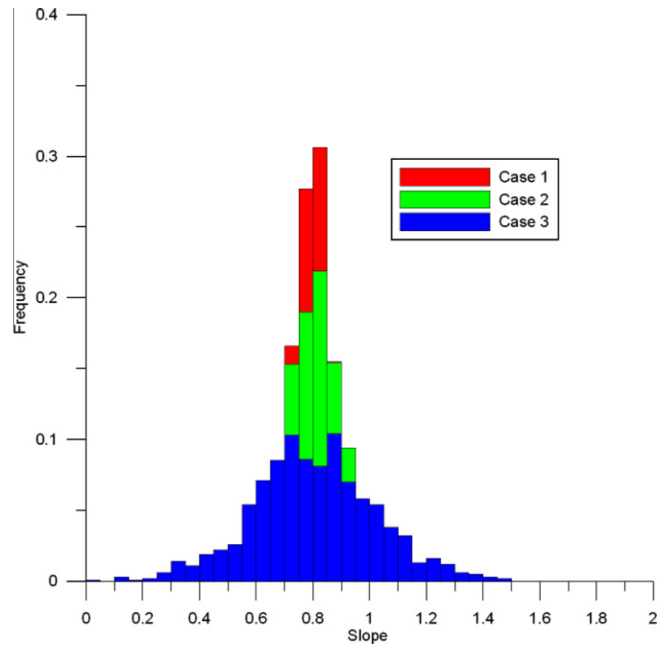

Fig. 1. Histogram for regression slopes in the three cases.

As rainfall/runoff modelling is more complex than a simple function fitting, the critical events cannot be reduced to the biggest and smallest events as in the previous example. Hydrological systems react to precipitation with a certain delay, and further the actual response of the catchment is related to past rainfall. Therefore, instead of selecting critical events based on single daily values we decided to use precipitation or discharge-related variables over a period of a few days.

If, using critical events, we can select data that are hydrologically relevant then we may improve our calibration process [5]. To identify the critical time periods that may contain enough information to identify model parameters, unusual sequences in the series of discharges and precipitation have to be identified. The intermittent nature of precipitation only allows the identification of unusual events for high precipitation amounts. However, for hydrological modelling, droughts might be of equal importance. Thus, instead of precipitation $P(t)$, the series of antecedent precipitation indices is used. The antecedent precipitation index API is defined as [13]:

$$API(t+1) = \alpha API(t) + P(t+1), \tag{1}$$

where $0 < \alpha < 1$ is a constant. This index describes the wetness state of a catchment and can be well related to discharge. The higher α, the more influence is assigned to past precipitation. However the value of α does not have significant influence on the identification of critical events. α may vary from catchment to catchment based on size, form or other characteristics of a catchment. In this study $\alpha = 0.9$ was used.

Another possible way to identify unusual hydrological events is to use the observed discharge series $Q(t)$. For simplicity, denote the selected series (API or discharge) as $X(t)$. Unusual events are defined as a sequence of $X(t)$ values that meet certain criteria. Considering d consecutive time steps (in our case days) leads to a d-dimensional set:

$$\mathbf{X}_d = \{(X(t-d+1), X(t-d+2), \ldots, X(t))\ t = d, \ldots, T\}, \tag{2}$$

where T is the total number of observation time steps available. Again, for simplicity, denote $X_d(t) = (X(t-d+1), X(t-d+2), \ldots, X(t))$.