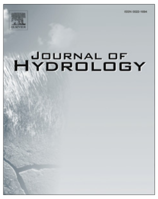
Journal of Hydrology 510 (2014) 447–458

Contents lists available at ScienceDirect

Journal of Hydrology

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



Smart low flow signature metrics for an improved overall performance evaluation of hydrological models

Matthias Pfannerstill *, Björn Guse, Nicola Fohrer

Christian-Albrechts-University of Kiel, Institute of Natural Resource Conservation, Department of Hydrology and Water Resources Management, Kiel, Germany

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 21 December 2013
Accepted 26 December 2013
Available online 8 January 2014
This manuscript was handled by Peter K. Kitanidis, Editor-in-Chief, with the assistance of Christophe Darnault, Associate Editor

Keywords:
Catchment hydrology
Model evaluation
Multi-metric framework
Signature measure
Low flow
Flow duration curve

SUMMARY

Hydrological models have to be calibrated accurately to provide reasonable model results. For a concise model evaluation, the different phases of the hydrograph have to be considered in multi-metric frameworks with appropriate performance metrics. Low and high flows need to be reproduced simultaneously without neglecting the other phases of the hydrograph.

In this paper, we highlight the relevance of very low and low flows with separate performance metrics. We present a multi-metric evaluation framework to identify calibration runs, which represent the different phases of the hydrograph precisely. A stepwise evaluation was done with commonly used statistical performance metrics (Nash–Sutcliffe, percent bias) and signature metrics, which are based on the flow duration curve (FDC). In order to consider a fairly balanced evaluation between high and low flow phases, we divided the flow duration curve into segments of high, medium and low flow phases, and additionally into very high and very low flow phases. The model performance in these segments was evaluated separately with the root mean square error (RMSE).

Our results show that this evaluation method leads to an improved selection of good calibration runs to enhance the overall model performance by the refined segmentation of FDC. By combining performance metrics for high flow conditions with low flow conditions, this study demonstrates the challenge of calibrating a model with a satisfactory performance in high and low phases simultaneously. Consequently, we conclude that an additional performance metric for very low flows should be included in model analyzes to improve the overall performance in all phases of the hydrograph.

1. Introduction

Hydrological models are used in practice and science to assess a wide range of hydrological problems such as climate and land use change or to predict extreme events in terms of flood and low flow events for river management (Tallaksen et al., 1997; Hunter et al., 2007; Laaha and Blöschl, 2007; Thielen et al., 2009). Hydrological complexity is reflected in different phases within the discharge time series. A challenge of hydrological models is to adequately represent all phases with the same model parameter set (Madsen, 2000). To achieve a satisfying reproduction of the hydrological processes, hydrologic models have to be calibrated to the conditions of the study catchments. Generally, model parameters are calibrated for specific catchment characteristics to the measured discharge time series. The most suitable parameters are selected with a sensitivity analysis (see van Griensven et al., 2006) or on the user's experience respectively. The next step is the calibration of selected parameters with following evaluation of model results by visual inspection of the hydrograph fitting and the application of performance measures (e.g. Moriasi et al., 2007).

The Nash–Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970), which is often used to evaluate simulation results in hydrology, is sensitive to differences in the observed and simulated means and variances (Legates and McCabe, 1999). However, this performance measure is more sensitive to extreme values (Legates and McCabe, 1999) and tends to neglect possible deviations in low flow periods as it is not very sensitive to systematic over- and underestimations of the model (Krause et al., 2005). The root mean square error (RMSE) overemphasizes flood peaks and leads to a bad calibration of low flow periods (Boyle et al., 2000; Madsen, 2000; Bekele and Nicklow, 2007). As a consequence, a better performance for high flows than for low flows may result in an underestimation in long dry periods (De Vos et al., 2010). Furthermore, a good performance in some periods with high flows is able to dominate the global performance and masks the poor performance in other periods like low flow periods (Zhang et al., 2011).

Referring to high and low flow calibration, the application of one single criterion tends to measure the difference between the simulated and observed hydrographs by matching one aspect of the hydrograph at the expense of another (Boyle et al., 2000;

---

* Corresponding author. Tel.: +49 0431 880 1798.
E-mail address: mpfannerstill@hydrology.uni-kiel.de (M. Pfannerstill).

0022-1694/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.12.044

Wagener et al., 2001). Furthermore, the application of one single performance measure is insufficient to take into account the representation of all relevant processes (Gupta et al., 1998; Wagener and Gupta, 2005; Gupta et al., 2008). This was also stated by Madsen (2000), who found no overall best performance measures during the calibration process. The reason for this shortcoming is the loss of valuable information by projecting from the high dimension of the data set down to the single dimension of the residual-based summary statistic (Gupta et al., 2008; Herbst and Casper, 2008). Since matching of all parts of the hydrograph is favorable, a trade-off between different phases of the hydrograph has to be accepted. This trade-off effect can be minimized in multi-objective approaches with multiple performance measures, whose importance for discharge calibration was revealed in Boyle et al. (2000), Bekele and Nicklow (2007), De Vos et al. (2010), Zhang et al. (2011), and Guse et al. (2013).

To assess different phases of the hydrograph, van Werkhoven et al. (2009) and Zhang et al. (2012) included statistical and hydrological metrics into the calibration process. They defined statistical metrics for the base and peak flow, hydrological metrics for the midrange flow and long-term water balance. The different parts of the hydrograph reflect different catchment functions (e.g. baseflow recession during dormant season of the vegetation) that can be captured in individual model components through parameter selection informed by careful hydrograph analysis (Carrillo et al., 2011). The importance of different performance metrics was also mentioned in van Werkhoven et al. (2009), Martinez and Gupta (2011) and Herman et al. (2013), who proposed a multiple criteria application for diagnostic model analysis. In a diagnostic analysis on differing watershed behavior during rainfall and dry periods, the hydrograph can be separated into driven, non-driven quick, non-driven slow discharge as defined by Boyle et al. (2000). Also Bekele and Nicklow (2007) applied specific objective functions to fit different portions of time series. Madsen et al. (2002) separated performance measures for high and low flows, which were only considered in periods above or below a threshold for high or low flows, respectively.

Yilmaz et al. (2008) used the overall water balance, vertical redistribution, temporal and spatial redistribution as signature measures for major behavioral functions. Signature measures are defined as hydrologic response characteristics that provide insights into the hydrologic function of catchments (Sawicz et al., 2011). Pokhrel et al. (2012) stated that several signature measures give a better overall representation of the hydrologic characteristics of the catchment. Both studies used the flow duration curve (FDC) to diagnose model performance for different flow characteristics of the catchment. There are several suggestions for splitting up the FDC into segments, which describe characteristic hydrological processes within the catchment (Yilmaz et al., 2008; Yokoo and Sivapalan, 2011; Cheng et al., 2012; Pokhrel et al., 2012).

Dividing the flow duration curve into segments leads to a process-based calibration for the dominant processes within the catchment, which are reflected by the different parts of the hydrograph. However, van Werkhoven et al. (2009) see the limitations of the FDC to fully reflect the quality of simulations, since it includes no information on accurate flow timing. In contrast to time series, FDC indicates only that the right distribution of flow levels occurred throughout the record (van Werkhoven et al., 2009). Thus, van Werkhoven et al. (2009) proposed a combination of statistical and signature metrics to capture the different parts of the hydrograph as well as their timing.

Dunn (1999) found high uncertainty for low flow predictions without specific consideration of a low flow criterion. Especially in lowlands, distinct low flow periods occur frequently but with high variability in the minimum discharge. In this case, it is uncertain, if the traditional segmentation (low flow: 70% time flow equalled or exceeded) of the FDC is sufficient to calibrate low flow periods. For an adequate representation of the very low flow periods with respect to very high flows, additional segmentations could be an approach to calibrate a fairly balanced representation of extreme periods. In our investigations we took up these questions and focused on following topics:

- How can all phases of the hydrograph be combined in a multi-metric framework evaluation?
- Does a multi-metric framework detect calibration runs with a reasonable reproduction of all phases of the hydrograph?
- Does the additional segmentation of the FDC into low flow segments lead to an improved reproduction of low flows?

2. Materials and methods

2.1. Study area

Our investigations were carried out in the Kielstau catchment ($50\ km^2$), which is located within a lowland area of the federal state Schleswig–Holstein in Northern Germany. The topography ranges between 27 m and 78 m above mean sea level with a flat landscape, described by rolling hills and depressions. In the higher regions of the Kielstau catchment, Haplic and Stagnic Luvisols are the dominant soils. Along the stream and its tributaries primarily Sapric Histosols are found (BGR, 1999). As a consequence of this flat topography, the groundwater is a specific characteristic of this lowland catchment. Schmalz et al. (2008) describe the dynamics of the near-surface groundwater at a riparian wetland as a dynamic interaction between groundwater and surface water. The near-surface groundwater is generally controlled by precipitation and, close to the river, also by river water level (Schmalz et al., 2008). Due to high water levels, a high fraction of approximately 38% of the agricultural area is drained (Fohrer et al., 2007). Further information about the catchments and results of investigations can be found in Schmalz and Fohrer (2009) and Fohrer and Schmalz (2012).

The hydrological characteristics are typical for a northern German lowland. The mean annual precipitation and temperature are 918.9 mm and 8.2% (DWD, 2012). The annual discharge is characterized by a mean outlet discharge at the gauging station Soltfeld of $0.42\ m^3\ s^{-1}$, a mean low flow discharge of $0.05\ m^3\ s^{-1}$ and a mean high flow discharge of $2.75\ m^3\ s^{-1}$ (LKN, 2013). Referring to the seasonality of the discharge, high flow events take place from November to January (LKN, 2013). The lowest discharge is observed from June to the late August (LKN, 2013). For our study, we used the mean daily discharge of the gauging station Soltfeld from 1999 to 2010.

2.2. The SWAT model

For our multi-metric framework development, we used the Soil and Water Assessment Tool (SWAT2012; Arnold et al., 1998). With this semi-distributed, eco-hydrological model, the discharge and the water cycle was simulated on a daily time-step for the Kielstau catchment. SWAT is a process-based conceptual model with abstracted, empirical components. These different components result in a very complex model with a high number of parameters (Cibin et al., 2010). The model concept of SWAT divides the processes into a land and a water phase (Neitsch et al., 2011). The water balance at the land phase is calculated by changes in soil water storages for each day, based on the calculation of the relevant processes. The main water input is the precipitation. To solve the water balance equation, the most important processes such as evaporation, runoff, soil water percolation and groundwater flow are considered. After calculating the water balance, the subbasins are connected in the water phase and the water is routed through the subbasins.