Environmental Modelling & Software 103 (2018) 74–89

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Groundwater recharge predictions in contrasted climate: The effect of model complexity and calibration period on recharge rates

Christian Moeck [a, *], Jana von Freyberg [b, c], Mario Schirmer [a, d]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Water Resources and Drinking Water, Überlandstr. 133, 8600 Dübendorf, Switzerland
[b] ETH Zurich, Department of Environmental Systems Science, Universitätstr. 16, 8092 Zurich, Switzerland
[c] Swiss Federal Research Institute WSL, Mountain Hydrology and Mass Movements, Zürcherstr. 111, 8903 Birmensdorf, Switzerland
[d] University of Neuchâtel, Centre of Hydrogeology and Geothermics (CHYN), Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland

A R T I C L E I N F O

Article history:
Received 14 November 2016
Received in revised form
31 August 2017
Accepted 11 February 2018

Keywords:
Groundwater recharge
Model complexity
Calibration period
Split sample test
Contrasted climate
Droughts

A B S T R A C T

We systematically evaluated the effect of model complexity and calibration strategy on estimated recharge using four varyingly complex models and a unique long-term recharge data set. A differential split sample test was carried out by using six calibration periods with climatically contrasting conditions in a constrained Monte Carlo approach. All models performed better during calibration than during validation due to differences in model structures and climatic conditions. The two more complex, physically-based models predicted the observed recharge with relatively small uncertainties, even when calibration and prediction periods had different climatic conditions. In contrast, the more simplistic soil-water balance model significantly underestimated the recharge rates. The fourth, semi-mechanistic model captured the observed recharge rates, but with a larger uncertainty range than the physically-based models. Our results may have relevant implications for a broad range of applications when recharge models are used as decision-making tools.

1. Introduction

Groundwater recharge is one of the main drivers in hydrogeological and hydrological systems (Bakker et al., 2013). Understanding the relationship between groundwater recharge rates and climatic conditions is essential for sustainable water resource management. Typically, hydrogeological models are applied to gain insight into this relationship and to generate recharge predictions to inform decision makers (Meixner et al., 2016).

The selection of a recharge model that is suitable for robust recharge predictions under site-specific climatic conditions is, however, often subjective and might be biased by the common practice of the modeller (Kurylyk and MacQuarrie, 2013; Vansteenkiste et al., 2014). This selection can be critical as the model structure and parametrization can strongly affect the quality of the simulations (Breuer et al., 2009; Li et al., 2015; Velazquez et al., 2013). For instance, Moeck et al. (2016) evaluated a variety of different recharge models in a climate change impact study and concluded that the selected model complexity can lead to significant model bias in the predictions. For hydrological models, Butts et al. (2004) indicated that model bias due to variation in model complexity is an issue for both, lumped and physically-based models. Consequently, utilizing only one model for hydrograph predictions ignores the possible uncertainty associated with the model structure (Doll and Fiedler, 2008). Although sophisticated calibration tools can result in an optimal fit between model simulations and observations for the calibration period (Doherty, 2003; Moeck et al., 2015; Zambrano-Bigiarini and Rojas, 2013), it is indispensable to have solid understanding of the reliability of model predictions beyond the calibration period.

Apart from the model structure and parametrization, implicit assumptions in the calibration approach are additional sources of uncertainty. For instance, many hydrological predictions are based on the assumption that the model calibration based on historical time periods is similarly valid for the prediction period (Seibert, 2003). The assumption of stationarity, however, is not always true, especially under changing climatic conditions. Non-stationarity of model parameters can occur, suggesting that certain historic time periods might be more useful for the identification of the parameter space while others might be less informative (Li et al., 2012; Vaze et al., 2010; Wagener et al., 2003). As it

---

* Corresponding author.
E-mail address: Christian.moeck@eawag.ch (C. Moeck).

https://doi.org/10.1016/j.envsoft.2018.02.005

was shown for rainfall-runoff modelling studies, extreme climatic periods such as heat waves or heavy precipitation events can have a strong impact on the predictions and increase the uncertainties of the predictions (Coron et al., 2012; Merz et al., 2011; Seiller et al., 2012). Cuthbert and Tindimugaya (2010) demonstrated that different groundwater recharge models simulated comparable historic recharge rates but the models responded very differently to changes in precipitation intensity. In a climate change impact study, Crosbie et al. (2011) concluded that the recharge model uncertainty is the smallest source of uncertainty compared to different global climate models or the downscaling methods. However, the effects of the calibration strategy and observation period as potential sources of uncertainty were not considered in the study of Crosbie et al. (2011). It can be speculated that model recharge predictions were not always consistent due to the chosen calibration strategy and period, even though the models performed similarly for the calibration period.

With regard to hydrological models for impact studies, Brigode et al. (2013) emphasized that the model reliability should be enhanced through developing calibration strategies that increase model performance and robustness under dissimilar climatic conditions. In that regard, Kirchner (2006) proposed that prediction models should be tested with a more comprehensive and incisive validation method, such as the differential split-sample test. With such a test, the model is calibrated and validated on time periods with very contrasting climatic and/or hydrological conditions (Klemes, 1986). For instance, Vaze et al. (2010) performed differential split-sample tests on four different rainfall-runoff models and found that models calibrated over wet periods generally tended to predict runoff incorrectly over a dry period. Brigode et al. (2013) also compared two rainfall-runoff models in a split-sample test to evaluate the model robustness and parameter uncertainty. The authors found that the major source of uncertainty and lacking robustness of the models occurred for climatic conditions that were very different from the calibration period. Coron et al. (2012) developed a Generalized Split Sample Test methodology to verify all possible combinations during the calibration-validation period. The authors indicated that the transferability of model parameters can introduce significant errors in the model predictions, which has strong implications in all model applications.

The differential split-sample test is still rarely applied for model testing (Andréassian et al., 2009) and if so, it is typically used for rainfall-runoff models. For the systematic evaluation of groundwater recharge models this method has, to the best of our knowledge, not yet been applied. We speculate that such an analysis might be precluded by the lack of long-term measurements of recharge rates required for model calibration. At the catchment scale, groundwater recharge cannot be measured experimentally (Scanlon et al., 2002; Scanlon et al., 2006; von Freyberg et al., 2015) so that the only direct measurements of vertical groundwater recharge at the plot scale can be obtained from large lysimeters (Groh et al., 2016). However, long time series of vertical groundwater recharge that cover a wide range of climatic conditions are generally rare.

In this study, we aim to address this research gap by utilizing a unique, long-term data set from the large lysimeter in the Rietholzbach research catchment in Switzerland (Gurtz et al., 2003; Seneviratne et al. 2012). Lysimeter seepage measurements from 1976 until today allow us to systematically evaluate how groundwater recharge predictions are affected by i) the type of model structure, ii) the parameterization and iii) the used calibration periods. In the first part of our study we systematically compare different model structures through five different model performance criteria and methods, such as Taylor plots and post-calibrated uncertainty analysis for the sum of annual recharge for the entire simulation period. We investigate how different parameterizations of models (due to parameter non-identifiability) influence the predictions by applying a Monte Carlo approach. Subsequently, we perform differential split-sample tests to investigate the relationship between the model performance (and robustness) and the choice of the calibration period.

2. Study area and groundwater recharge models

2.1. Rietholzbach lysimeter

The large, free-draining weighting lysimeter (2.5 m deep, 2 m diameter) is located in the Rietholzbach research catchment, a pre-alpine head watershed of the Thur river basin in north-eastern Switzerland (Fig. 1a). The lysimeter was constructed in 1975 and is mainly filled with gley-brown soil from the same location (Gurtz et al., 2003). The lysimeter surface is covered with grass to imitate the surrounding conditions. At the bottom of the lysimeter column, outflow is measured with a tipping bucket (Fig. 1b). As groundwater table depths are generally shallow at the site (typically less than 5 m beneath surface, but below plant root depth), lysimeter seepage is assumed to be a reliable indicator of actual vertical groundwater recharge (Ghasemizade et al., 2015; von Freyberg et al., 2015). Observed lysimeter seepage also correlates well with the streamflow signal of the Rietholzbach river, indicating that recharge processes at the plot scale are representative for the 3.14 km^2-large catchment (Seneviratne et al., 2012). In the present study, we use daily lysimeter seepage as a surrogate for vertical

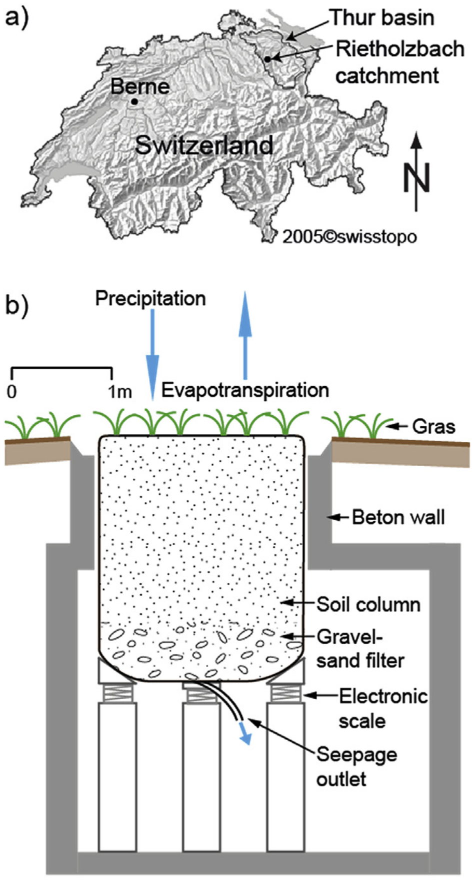

Fig. 1. a) Location of the Rietholzbach research catchment in the Thur river basin in north-east Switzerland; b) Schematic setup of the lysimeter system in the Rietholzbach research catchment (modified after Seneviratne et al. (2012)).