ELSEVIER

Computers & Geosciences 34 (2008) 699–713

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

Effective modelling of percolation at the landscape scale using data-based approaches

Benny Selle[a,*], Gunnar Lischeid[b], Bernd Huwe[a]

[a]Soil Physics Group, University of Bayreuth, D-95440 Bayreuth, Germany
[b]Ecological Modelling, University of Bayreuth, D-95440 Bayreuth, Germany

Received 6 September 2006; received in revised form 22 June 2007; accepted 27 June 2007

Abstract

Process-based models have been extensively applied to assess the impact of landuse change on water quantity and quality at landscape scales. However, the routine application of those models suffers from large computational efforts, lack of transparency and the requirement of many input parameters. Data-based models such as Feed-Forward Multilayer Perceptrons (MLP) and Classification and Regression Trees (CART) may be used as effective models, i.e. simple approximations of complex process-based models. These data-based approaches can subsequently be applied for scenario analysis and as a transparent management tool provided climatic boundary conditions and the basic model assumptions of the process-based models do not change dramatically. In this study, we apply MLP, CART and Multiple Linear Regression (LR) to model the spatially distributed and spatially aggregated percolation in soils using weather, groundwater and soil data. The percolation data is obtained via numerical experiments with Hydrus1D. Thus, the complex process-based model is approximated using simpler data-based approaches. The MLP model explains most of the percolation variance in time and space without using any soil information. This reflects the effective dimensionality of the process-based model and suggests that percolation in the study area may be modelled much simpler than using Hydrus1D. The CART model shows that soil properties play a negligible role for percolation under wet climatic conditions. However, they become more important if the conditions turn drier. The LR method does not yield satisfactory predictions for the spatially distributed percolation however the spatially aggregated percolation is well approximated. This may indicate that the soils behave simpler (i.e. more linear) when percolation dynamics are upscaled.

Keywords: Metamodel; Soil hydrology; Groundwater recharge; Lumped models; Pattern recognition techniques; Neural networks; Upscaling

1. Introduction

The modelling of soil-hydrological processes at landscape scales is seeing an increased need in water-related management purposes, such as for assessing the environmental impact of agricultural management on water quality in drinking water

*Corresponding author. Current address: Future Farming Systems Research Division, Department of Primary Industries, Tatura, Victoria, Australia.
E-mail address: benny.selle@dpi.vic.gov.au (B. Selle).

0098-3004/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2007.06.007

reservoirs and river basins (Kralisch et al., 2003; Tuteja et al., 2003).

Process-based models are usually based on the numerical solutions of partial differential equations (e.g. Richards Equation and the Convection-Dispersion Equation). They have been extensively applied to assess the impact of landuse change on water quantity and quality at the landscape scale (e.g. Huwe and Totsche, 1995; Tuteja et al., 2003; Singh et al., 2006). However, the routine application of these models suffers from large computational efforts, lack of transparency and requirement of many input parameters. The direct measurement of model parameters is difficult under field conditions. Field measurements often do not produce information content sufficient to inversely calibrate all parameters used in process-based models (Vrugt et al., 2004; Schulz and Jarvis, 2004).

Data-based models such as Feed-Forward Multilayer Perceptrons (MLP) and Classification and Regression Trees (CART) (Breimann et al., 1984) are extremely flexible regression techniques. These techniques, as well as other data-based approaches, may be used as effective models, i.e. simple approximations of complex process-based models. These data-based approaches can subsequently be applied for scenario analysis and as a transparent management tool provided climatic boundary conditions and the basic model assumptions of the process-based models do not change dramatically (e.g. Haberlandt et al., 2002; Kralisch et al., 2003). They may also help to identify key variables and thus, to effectively parameterise process-based models (i.e. to decide where to put the most effort into parameterisation through direct measurement).

In this study, we model the percolation in soils at landscape scales using MLP, CART and Multiple Linear Regression (LR) with weather, groundwater and soil information. Percolation is used as a typical example of a process that is complex and practically relevant to many water quantity and quality issues. The data-based approaches are used to predict both, spatially distributed and area-average (“spatially aggregated”) percolation previously obtained via process-based modelling with Hydrus1D. Thus, the complex process-based model is approximated using simpler data-based approaches. An important motivation behind this research is therefore to assess the required level of complexity for different approaches to model the percolation at landscape scales. This may show the effective dimensionality of this process which in turn may have implications for the identifiability of soil parameters.

2. Material and methods

2.1. Site characteristics

The study site “Weißenstädter Becken” is an approximately 10 km^2 plateau with a subdued relief at between 600 and 700 m ASL in the “Fichtelgebirge” (Northern Bavaria, Germany) (Fig. 1). The primary landuse is pasture. The water balance is positive as a result of the high annual precipitation (ranging from 900 to 1000 mm) and also due to the low mean annual temperatures which range from 5 to 6 °C. Brown Podzolic soils occur most frequently in the area, having developed from periglacial muds and the acidic bed rock parent material (granite, mica slate). Histosols and Gley soils occur in convergent parts of the landscape. A very detailed database exists for the study site due to the large number of previously completed studies (e.g. Mertens and Huwe, 2002; Selle et al., 2006).

2.2. Determination of the data set

The data set used to train, validate and test the data-based approaches was obtained by the following procedure.

First, 242 one metre deep soil profiles were obtained by soil core augering (Fig. 1). Samples were taken to represent lower, middle and upper parts of the slopes, the various parent materials according to the geologic map and the distinct soil textures from the soil appraisal[1] within the study area. The soil texture for each horizon was determined by hand and the textural classes were identified according to the German Handbook for soil mapping (Boden, 1994). The observed ranges of soil texture were: percentage silt (0–80%), percentage clay (0–65%) and percentage sand (0–100%). The bulk density for each horizon was estimated using field methods. Typical values according to their organic carbon contents were assumed for histic soils. The bulk density varied from 0.5 g/cm^3 in histic soils to 1.8 g/cm^3 in compacted subsoils (periglacial muds).

Second, the soil-hydraulic parameters (water retention and hydraulic conductivity function, respectively) according to Van Genuchten (1980) were obtained using the pedo-transfer functions

[1]Soil appraisal (German: *Bodenschätzung*) is defined as the assessment of a piece of property for financial purposes, which primarily considers the soil fertility properties.