Science of the Total Environment 505 (2015) 390–401

Contents lists available at ScienceDirect

Science of the Total Environment

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

Modeling water outflow from tile-drained agricultural fields

Vladimir Kuzmanovski [a,b,*], Aneta Trajanov [b], Florence Leprince [c], Sašo Džeroski [a,b,d], Marko Debeljak [a,b]

[a] International Postgraduate School, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[b] Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[c] ARVALIS - Institut du Végétal, 21, Chemin de Pau, 64121 Montardon, France
[d] Center of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, 1000 Ljubljana, Slovenia

HIGHLIGHTS

- We model surface runoff and drainage discharge water outflow from agricultural fields.
- We use machine learning techniques in order to accurately predict the outflow.
- The physical-based models are with high complexity of input and calibration.
- Our models overcome the issue with input complexity with comparable performances as physical-based models.

ARTICLE INFO

Article history:
Received 31 July 2014
Received in revised form 3 October 2014
Accepted 4 October 2014
Available online 21 October 2014

Editor: D. Barcelo

Keywords:
Agriculture
Drainage discharge
Surface runoff
Data mining
Machine learning
Modeling

ABSTRACT

The estimation of the pollution risk of surface and ground water with plant protection products applied on fields depends highly on the reliable prediction of the water outflows over (surface runoff) and through (discharge through sub-surface drainage systems) the soil. In previous studies, water movement through the soil has been simulated mainly using physically-based models. The most frequently used models for predicting soil water movement are MACRO, HYDRUS-1D/2D and Root Zone Water Quality Model. However, these models are difficult to apply to a small portion of land due to the information required about the soil and climate, which are difficult to obtain for each plot separately.

In this paper, we focus on improving the performance and applicability of water outflow modeling by using a modeling approach based on machine learning techniques. It allows us to overcome the major drawbacks of physically-based models e.g., the complexity and difficulty of obtaining the information necessary for the calibration and the validation, by learning models from data collected from experimental fields that are representative for a wider area (region).

We evaluate the proposed approach on data obtained from the La Jaillière experimental site, located in Western France. This experimental site represents one of the ten scenarios contained in the MACRO system. Our study focuses on two types of water outflows: discharge through sub-surface drainage systems and surface runoff.

The results show that the proposed modeling approach successfully extracts knowledge from the collected data, avoiding the need to provide the information for calibration and validation of physically-based models. In addition, we compare the overall performance of the learned models with the performance of existing models MACRO and RZWQM. The comparison shows overall improvement in the prediction of discharge through sub-surface drainage systems, and partial improvement in the prediction of the surface runoff, in years with intensive rainfall.

1. Introduction

The quantity and quality of surface and ground water affect the ability of aquatic environments to sustain healthy ecosystems, while an abundant supply of clean water is a basic requirement for many of its fundamental uses on which humans depend. Between 80% and 85% of human population uses groundwater as a drinking water supply (Bedient et al., 1999). Furthermore, the consumption of water for human needs is approximately doubled every 20 years, while, on the other hand, new sources of water are becoming scarcer and polluted water is becoming more expensive to remediate. Thus, the protection of surface and ground water is crucial for human and ecosystem health.

Industry and agriculture, as the main water polluters, affect water quality through both point and diffuse-source pollution. The former

* Corresponding author at: Department of Knowledge Technologies, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia.
E-mail addresses: vladimir.kuzmanovski@ijs.si (V. Kuzmanovski), aneta.trajanov@ijs.si (A. Trajanov), f.leprince@arvalisinstitutduvegetal.fr (F. Leprince), saso.dzeroski@ijs.si (S. Džeroski), marko.debeljak@ijs.si (M. Debeljak).

http://dx.doi.org/10.1016/j.scitotenv.2014.10.009
0048-9697/

mainly refers to industrial or sewage treatment plants, while the latter comes from many diffuse-sources (e.g., surface runoff and discharge through sub-surface drainage systems for surface water, and infiltration for ground water). Furthermore, several studies have shown that agriculture acts as the main diffuse-source polluter because of the use of plant protection products (referred to as pollutants in the remainder of the paper). These products are applied on a field scale in accordance with conventional agricultural management and practices (Capel et al., 2001; Holvoet et al., 2007).

Diffuse-source pollution is a process characterized by random occurrence and uncertain discharge of pollutants, variable temporal pollution loads, and is difficult to simulate and forecast (Zhenyao et al., 2012). It is a complex process, the modeling of which involves many phases: identification of pollutant transfer pathways, prediction of the possible pollutant transfer and assessment of water pollution risk. Eventually, the modeling helps in selection and application of specific mitigation measures by providing the decision-making process with additional information about the water status and facilitating the pollution risk reduction.

The main transfer pathways of diffuse-source pollution at the field scale are surface runoff, discharge through sub-surface drainage systems (referred to as drainage discharge in the remainder of the paper) and lateral seepage (lateral hypodermic flow on the non-permeable soil substratum). Furthermore, infiltration is identified as a direct pollutant transfer path in ground water (Brown and van Beinum, 2009; Holvoet et al., 2007). The results of the study by Brown and van Beinum (2009) have shown that surface runoff and drainage discharge make a significant contribution to the pollution of surface waters. In addition, drainage discharge appeared as a relevant route for pollutant transport in six out of ten environmental scenarios representative of agricultural conditions across Europe (FOCUS, 2001).

The estimation of pollutant transfers (further referred to as water outflows in this paper) at the field level is mainly physically-based modeling approaches, combining a theoretical description of different water outflows in the soil and data from field experiments (used for calibration and validation of the models). As a disadvantage of such models, Bredehoeft (2005) notes that while the foundations of such modeling are conceptual models, new empirical data typically render invalid predictions. Moreover, he suggests that the solution of this problem is twofold: to collect as much data as feasible and to keep the possibility to change the structure of the conceptual models open.

In the last 20 years, several physically-based models of water outflow processes have been developed (Gerke and van Genuchten, 1993). The most frequently used models for predicting soil water outflows are the physically-based models MACRO (Larsbo and Jarvis, 2005), HYDRUS-1D/2D (Šimůnek and van Genuchten, 2008) and the Root Zone Water Quality Model (Ahuja et al., 2000). A common methodological problem of physically-based models is their parameterization, which can be very time-consuming and expensive due to the requirements for calibration of specific data. Such data are typically obtained from laboratory analysis of the examined soil types, such as designation, thickness, texture, soil structure, pH, organic carbon content or bulk density in order to estimate hydraulic conductivity and water retention. Furthermore, today's monitoring technology allows collecting large amounts of data that can be used by the models in tuning their predictions or to build new and more accurate models. However, the use of such collected data would require complex structure revisions of the existing physically-based models [e.g., Larsbo and Jarvis (2005), Šimůnek and van Genuchten (2008), Ahuja et al. (2000)], which is impractical.

We aim to avoid the methodological limitations of physically-based models and to exploit the large amount of existing empirical data collected in experimental fields. We employ recent advances in information technologies for modeling drainage discharge and surface runoff as the two main pollutant pathways to surface water from hydromorphic agricultural fields. In particular, we propose to use a modeling methodology based on machine learning and data mining techniques to describe the water outflow through the soil and to make accurate predictions of the amount of drained and surface runoff water from agricultural fields.

Machine learning is a methodology that holds a lot of promise for the field of environmental sciences (Debeljak and Džeroski, 2011). It studies methods that build predictive models (in the form of decision trees, decision rules, linear equations, etc.) from examples (instances) described with the values of their features (independent variables) (Fig. 1). As an additional evaluation of our new modeling approach, we compare the performance and predictions of our models with those of the physically-based models MACRO and RZWQM, applied to the same case-study (La Jaillière, France) for the same time period.

The remainder of the manuscript is organized as follows. In Section 2, we introduce the machine learning methods we used in our research. Section 3 gives a description of the case study, while

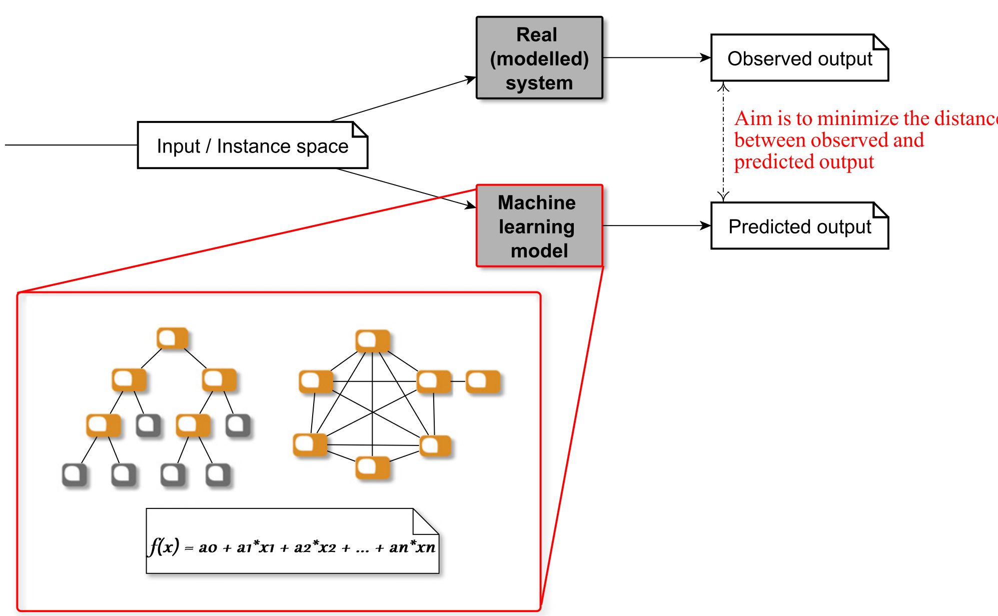

Fig. 1. Scheme of modeling approach using machine learning methodology.