Introduction

Protecting the surface and the ground water from pollution by plant protection products (PPP) used in agriculture, would reduce the negative environmental consequences of PPP usage. The main path of diffuse water pollution with PPP is the drainage water flow from tile-drained agricultural lands. To reduce the water pollution, the amount of outflow of polluted water from the fields has to be reduced. Losses of phytochemicals to drains are closely controlled by the time between their application in the field and the initiation of drainage flow. Therefore both identification of the drainage periods and prediction of the amount of drained water from tile-drained agricultural fields are important tasks in determining a proper set of mitigation strategies for protecting surface water against pollution (Kreuger & Nilsson, 2001).

To this end, we need to have accurate models for predicting the time periods and quantity of outflow. In this study, we address the problem of predicting the drainage period and the quantity of drainage outflow by using machine learning and data mining techniques. Our focus is mainly on describing the conditions which trigger beginning and the end of drainage flows and accurately predicting the amount of drained water.

Data

We used data from the field experiment station La Jaillière, France, which has 11 fields, where data have been collected for each field for 25 agricultural seasons (1987-2011). The data describe the agricultural practices (tillage, sowing, fertilizing and phytochemicals application dates), the amount of drainage and runoff water flows separately for each field with sampling proportional to the water flows. The meteorological data were collected from two meteorological stations in La Jaillière.

Methodology

Predictive models were built with the data mining modelling approach that attempts to find patterns and new knowledge in large datasets. It applies data mining methods, such as machine learning as one of the most active research areas in the field of artificial intelligence. This approach was successfully applied in numerous applications in the field of environmental and agricultural sciences (Debeljak and Džeroski, 2011). For estimating the predictive performance of data mining models, we used a standard technique of cross-validation.

Results

The best model (Fig. 1a) for estimating the start of the drainage period uses past meteorological data and cumulative drainage while the models that that do not use the cumulative drainage from the beginning of a agricultural season can still successfully estimate the beginning of a drainage period (accuracy from 75.06% to 91.24%). For estimating the end of a drainage period we used two different combinations of attributes; the first one uses only past meteorological data and the second one, besides past, uses also future meteorological data (Fig. 1b). Since we applied several different types of data mining methods for predicting drained outflow and because the performances of these models are similar, the model with low complexity was selected as the final one (Fig.2).
Figure 1. The predictive models for the start (a) and the end (b) of a drainage period with accuracy of 94.65% and 85.94%, respectively.

Figure 2. The drainage model with highest performance (average RMSE over 25 agricultural seasons is 0.9) and lowest complexity, where Rainfall A1-A5 mean cumulative rainfall for today and last 1, up to 5 days; RainfallN5 means daily cumulative rainfall 5 days ago; TempN3 means daily average temperature 3 days ago; LM1-16 are linear models (not shown due to the space limitation) constructed for each leaf in the tree; Rgi is Italian rye grass.

Conclusions

The models that estimate a drainage period and the quantity of outflow are of high quality and can be used to give a reliable estimation about the drainage regime in the La Jaillière region. They use data that can be obtained from weather forecasts, so that they can estimate whether the drainage will begin soon (sometime in the next week) or whether it will stop in the following week and they can make reliable prediction about the amount of drained water. These models can be further used as an initial point for exposure assessment in environmental risk analysis, together with additional model for predicting the concentrations of active substances.

References


