Studying the drainage periods for agricultural fields with data mining: La Jaillière case study

Aneta Trajanov[a], Vladimir Kuzmanovski[a], Florence Leprince[b], Benoit Real[c], Alain Dutertre[d], Julie Maillet-Mezeray[e], Sašo Džeroski[a,f,g], Marko Debeljak[a,f]

[a] Department of Knowledge Technologies, Jozef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia

Emails: aneta.trajanov@ijs.si, vladimir.kuzmanovski@ijs.si, saso.dzeroski@ijs.si, marko.debeljak@ijs.si

[b] ARVALIS – Institut du végétal, 21, chemin de Pau, 64121 Montardon, France

Email: f.leprince@arvalisinstitutduvegetal.fr

[c] ARVALIS - Institut du végétal, 2, chaussée Brunehaut, CS 30200, 80208 Peronne Cedex, France

Email: b.real@arvalisinstitutduvegetal.fr

[d] ARVALIS - Institut du végétal, Station expérimentale de La Jaillière, 44370 La Chapelle Saint Sauveur, France

Email: a.dutertre@arvalisinstitutduvegetal.fr

[e] ARVALIS - Institut du végétal, Station expérimentale, 91720 Boigneville, France

Email: julie.mailletmezeray@bayer.com

[f] Jozef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia

[g] Center of Excellence for Integrated Approaches in Chemistry and Biology of Proteins, Jamova cesta 39, 1000 Ljubljana, Slovenia

The identification of intensive drainage periods is important for determining mitigation strategies for protecting water against pollution with plant protection products. Most attempts to estimate the start, duration and the end of a drainage period are based either on mechanistic modeling approach or on empirical knowledge about tile drainage. The mechanistic modeling requires many parameters, while the empirical approach does not allow for making simulations and predictions needed for proposing reliable mitigation measures. In order to complement these two approaches, we have used a data mining approach on data from 25 (1987-2011) agricultural seasons (campaigns) from the experimental station La Jaillière, France. The models for estimating the start and the end of the intensive drainage periods for a particular campaign have the form of decision trees and tell us which factors influence these dates the most. The start of a drainage period depends mostly on the cumulative drainage and the cumulative rainfall since the beginning of the campaign and the average air temperature of the last seven days. For estimating the end of a drainage period, the most important variables are the cumulative rainfall of the last seven days and the average air temperature of the following seven days.

The obtained models for estimating the start and the end of a drainage period could be used not just to estimate the daily status of the drainage regime on a particular field (e.g., presence or absence of drainage), but they can also be used to predict the drainage status of the field for a time period covered with reliable weather forecasts. Using information from weather forecasts to run simulations on models for the beginning of a drainage period (or the end, depending on the decision at hand) would make it easier for farmers and policy makers to take into account the drainage period when deciding to apply plant protection products in the field. Thus, our data mining models, built from measured data, bring decision making flexibility to their users, because they can be used either for ex-ante or ex-post analysis. The combination of both types of analysis presents a very simple decision support system, which significantly increases the certainty and flexibility of management decisions taken by advisors and farmers in the La Jaillière area (ARVALIS) or in other places with the same field and crop management properties.