Computers and Electronics in Agriculture 110 (2015) 259–266

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

SVMLEACH – NK POTATO: A simple software tool to simulate nitrate and potassium co-leaching under potato crop

J.G. Fortin [a,*], A. Morais [a], F. Anctil [a], L.E. Parent [b]

[a] Department of Civil and Water Engineering, Université Laval, 1065 avenue de la Médecine, G1V 0A6 Québec, Canada
[b] Department of Soils and Agrifood Engineering, Université Laval, 2425 rue de l'Agriculture, G1V 0A6 Québec, Canada

ARTICLE INFO

Article history:
Received 22 August 2014
Received in revised form 24 November 2014
Accepted 28 November 2014

Keywords:
Nitrate leaching
Potassium leaching
Least squares support vector machines
Precision farming

ABSTRACT

Nutrient leaching is a major issue in intensive agriculture. Modelling helps evaluating nutrient loss and developing fertilization strategies. This paper introduces the SVMLEACH – NK POTATO software tool, which embeds a series of least squares support vector machines (LS-SVMs) models that simulate nitrate (NO_3) and potassium (K) losses using minimal input data. Results are compared to data collected from 36 suction lysimeters in 2012 and 2013 in two experimental fields under potato crop in the Province of Québec, Canada. Leachate concentrations in NO_3 and K are converted by the software in potential leaching mass flux using water loss simulated with the drainage function of a meteorological land surface scheme, namely the Canadian Land Surface Scheme (CLASS) model. Performance during validation is very satisfying with Root Mean Square Errors (RMSEs) representing 10% of the mean of the observed concentrations.

1. Introduction

Groundwater contamination by fertilizers is a worldwide problem (World Health Organization, 1985; Power and Schepers, 1989; Bouchard et al., 1992; Rupert, 2008). Sandy soils intensively cropped to potato or other high N-demanding crops are particularly at risk (Levallois and Phaneuf, 1994). Studies conducted in the North eastern part of the U.S. (Maine, New York) and Canada (Ontario) have shown that, for potato crops, 45–60% of the N applied at planting is recovered by the tubers (Bouldin and Selleck, 1977). Sharma and Sharma (2013) demonstrated that accompanying anions such as Cl^- and NO_3^- may also lead to substantial K losses, especially in light textured soils. Although K is not usually considered an environmental pollutant, its leaching may also affect plant growth and quality (Broschat, 1995). The K nutrition of crops under varied regimes of nitrogen supply was reviewed by Zhang et al. (2010). For instance, it was shown that the soil K status exerts an important influence on crop uptake and response to added N, as a consequence of the physiological role of K in plants.

Many process-based models have been developed to simulate nutrient leaching as a function of soil properties, climatic conditions, and management practices (e.g. SOILN, LEACHM, and NLEAP). These models require site-specific soil, climatic, and management data as inputs, which are generally not collected unless model calibration is the main intent (Jiang et al., 2012). They also use empirical coefficients that may require long-time local optimization (Varcoe, 1990; Fortin et al., 2011a). Alternative approaches might therefore be needed for operational purposes. Previous studies showed that machine learning techniques, such as least squares support vector machines (LS-SVMs), are particularly powerful for simulating soil N status from minimal input data (Fortin et al., 2014).

The objective of this work is to present a simple software tool simulating the daily seasonal dynamics of N and K leaching under potato crop with minimal input data. The N or K total losses below the root zone, referred herein as leaching mass flux in kg ha^{-1}, are generally computed as the integral of the curve relating concentration to cumulative water losses. In this work, the water loss is first modelled using the Canadian Land Surface Scheme (CLASS) (Verseghy, 1991) and approximated with a specific LS-SVM requiring only rainfall and day of year (DOY) as inputs. Input candidates for the leaching models are selected from prior knowledge on their close relationship with N and K leaching: N fertilization rate (*Nfert*), cumulative temperature ($\Sigma temp$), cumulative rainfall ($\Sigma rain$), percentage of clay content (% Clay) in the arable layer, and DOY. Data collected from 36 suction lysimeters in 2012 and 2013 in two experimental fields under potato crop in the Province of Québec, Canada, were used to calibrate and validate the models. Results of the LS-SVMs were compared to a multiple linear regression (MLR) to evaluate the increase of performance of the

* Corresponding author at: Pavillon Adrien-Pouliot 1065, avenue de la Médecine, Bureau 1916, Université Laval, Québec, G1V 0A6 Québec, Canada. Tel.: +1 418 656 2131(2032).
E-mail address: jerome.goulet-fortin.1@ulaval.ca (J.G. Fortin).

http://dx.doi.org/10.1016/j.compag.2014.11.025
0168-1699/

non-linear approach. A simple user graphical interface was developed and the program is compiled as a stand-alone Windows application. Because the software tool relies on a series of LS-SVMs, it is named SVMLEACH. More specifically, the current tool was developed for potato and NO_3/K leaching and was therefore named SVMLEACH – NK POTATO. The MATLAB libraries needed to run the models are freely distributed by the MATLAB Component Runtime (MCR) in a non-profit perspective. The MCR is automatically installed in the user's computer during the deployment of the SVMLEACH – NK POTATO package.

2. Materials and methods

2.1. The data

Meteorological data were obtained from planting to harvest at a daily time-step from a nearby micro-meteorological station. In N and C models, the minimum climatic input data requirement for simulating soil biological activity generally includes precipitation, temperature, and information on potential evapotranspiration. Because soil organic matter mineralization into inorganic N forms shows little sensitivity to evapotranspiration (Fortin et al., 2011a,b), only cumulative temperature and cumulative precipitation were considered. CLASS however requires a more detailed dataset: short-wave incident radiation ($W\,m^{-2}$), long-wave radiation ($W\,m^{-2}$), specific humidity ($kg\,kg^{-1}$), wind speed ($m\,s^{-2}$), and atmospheric pressure (Pa). Short-wave and long-wave radiation values were retrieved from the NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/) (Kalnay et al., 1996).

Soil concentration data was collected from four rainfed potato (*Solanum tuberosum* L.) fields located at Groupe Gosselin FG Inc. near Pont Rouge, Province of Québec, Canada (Lat. 46°45′N; Long. 71°42′W). Each site was divided into three large blocks submitted to the following N treatments at planting: 0, 90, 130, and 170 $kg\,N\,ha^{-1}$. P (190 $kg\,ha^{-1}$), K (260 $kg\,ha^{-1}$), and Mg (30 $kg\,ha^{-1}$) were applied according to local recommendations (CRAAQ, 2003). Soils were sampled at spring at each site to determine particle-size distribution using the hydrometer method (Sheldrick and Wang, 1993). Suction lysimeters (model 1900L24-B02M2, Soil Moisture Equipment Corp., Ca) were installed into the soil (40 cm) at planting. There were three replications per treatment, for a total of 36 lysimeters. Extraction of the soil water from the suction lysimeters was performed about 7–8 times each year (a total of 64 samples) and daily time series were obtained by linear interpolation (for a total of 1795 daily values). The NO_3 and K concentrations were determined by liquid chromatography using a Dionex ICS2000 equipped with a UV-detector (Dionex Corporation, Sunnyvale, Ca). For each treatment, the results of the replications were averaged prior to modelling.

2.2. N and K leachate concentrations models

The LS-SVM method has been chosen for implementation in LS-SVM after an evaluation of several regression techniques for the estimation of NO_3 concentration on leachates (Fortin et al., 2014). In that comparative study, LS-SVM outperformed neural network and multivariate adaptive regression spline, as well as regular multivariate linear regression. The underlying principles of SVMs are presented in Vapnik (1998).

The LS-SVM algorithm can be described as follows. For a given training set $\{x_k, y_k\}_{K=1}^{N}$ with the input $x_k \in R^N$ and the output $y_k \in R$, a regression model is built using nonlinear mapping function $\varphi(x)$, which maps the input data to higher dimensional feature space:

$$y = w^T\varphi(x) + \beta \tag{1}$$

where T is the transpose, $W \in R^N$ is the weight vector and β is the bias. For function approximation, the following optimization problem is formulated:

$$\min J(w,e) = \frac{1}{2}w^Tw + \frac{\gamma}{2}\sum_{k=1}^{N} e_k^2 \tag{2}$$

subject to the constraint:

$$y_k = w^T\varphi(x) + \beta + e_k, \quad k = 1,\ldots,N \tag{3}$$

where γ is the regularization parameter adjusting the trade-off between model's complexity and the training errors and e_k the random errors. The following Lagrange function is used to solve this optimization problem:

$$L(w,\beta,e,\alpha) = J(w,e) - \sum_{k=1}^{N}\alpha_k\{w^T\varphi(x_k) + \beta + e_k - y_k\} \tag{4}$$

where α_k are Lagrange multipliers, also called support values in that context. The solution of the above equation can be obtained by partially differentiating with respect to each variable:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \to w = \sum_{k=1}^{N}\alpha_k\varphi(x_k) \\ \frac{\partial L}{\partial \beta} = 0 \to \sum_{k=1}^{N}\alpha_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \to \alpha_k = \gamma e_k, \quad k = 1,\ldots,N \\ \frac{\partial L}{\partial \alpha_k} = 0 \to w^T\varphi(x_k) + \beta + e_k - y_k = 0, \quad k = 1,\ldots,N \end{cases} \tag{5}$$

Removing w and e, the equation become the following linear function group:

$$\begin{bmatrix} 0 & 1^{\to T} \\ 1^{\to} & \Omega + \gamma^{-1}I \end{bmatrix}\begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{6}$$

with

$$\begin{cases} y = [y_1,\ldots,y_N] \\ 1^{\to} = [1,\ldots,1] \\ \alpha = [\alpha_1,\ldots,\alpha_N] \\ \Omega = \{\Omega_{kl}|k,l = 1,\ldots,N\} \end{cases} \tag{7}$$

and

$$\Omega_{kl} = \varphi(x_k)^T\varphi(x_1) = K(x_k,x_l), \quad k,l - 1,\ldots,N, \tag{8}$$

where $K(x_k,x_l)$ is the kernel function. Common examples of kernel function contain linear, polynomial and radial basis function (RBF) kernel. In SVMLEACH, the RBF kernel was selected as the kernel function:

$$K(x,x_k) = \exp\left(-||x - x_k||^2/\sigma^2\right) \tag{9}$$

where σ is a width parameter that must set by the user The LS-SVM regression model is therefore expressed as:

$$y = \sum_{k=1}^{N}\alpha_k K(x,x_k) + \beta \tag{10}$$

Determination of the proper kernel function is a crucial issue when building SVM or LS-SVM models. However, no systematic methodology is available for a prior selection of kernel function (Wu et al., 2008). As it is the case in most LS-SVM function approximation problems, the RBF kernel was used. The popularity of the RBF kernel resides in its known capability for solving nonlinear problem, while its compact shape reduces computational