Computers and Electronics in Agriculture 101 (2014) 68–75

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Deriving data mining and regression based water-salinity production functions for spring wheat (*Triticum aestivum*)

Amir Haghverdi [a,b,*], Bijan Ghahraman [b], Brian G. Leib [a], Inmaculada Pulido-Calvo [c], Mohammad Kafi [d], Kamran Davary [b], Behrang Ashorun [b]

[a] Dept. of Biosystems Engineering and Soil Science, University of Tennessee, 2506 E.J. Chapman Dr., Knoxville, TN 37996-4531, USA
[b] Dept. of Water Engineering, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran
[c] Dept. of Agroforestry Sciences, Higher Technical School of Engineering, Campus La Rábida, University of Huelva, 21819 Palos de la Frontera, Huelva, Spain
[d] Dept. of Agronomy, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran

ARTICLE INFO

Article history:
Received 8 May 2013
Received in revised form 25 September 2013
Accepted 15 December 2013

Keywords:
Classification and regression trees
Deficit irrigation
Multilayer perceptron
Response surface
Salinity

ABSTRACT

Production functions (PFs) are practical tools for not only irrigation scheduling but also in economic analysis as a mathematical relationship between relative grain yield and factors like evapotranspiration, irrigation water and salinity. This study was carried out in the Mashhad region of Iran during cropping years 2010 and 2011 to evaluate the performances of two data mining methods, decision tree and neural network, for deriving PFs of spring wheat under simultaneous drought and salinity stress compared with four well known regression-based PFs. The four well known PFs were: Jensen-PF (Jensen, 1968), Minhas-PF (Minhas et al., 1974), modified Stewart-PF (Stewart et al., 1977; Stegman et al., 1980), and Nairizi-PF (Nairizi and Rydzewski, 1977). Heading and flowering were the most sensitive growth stages followed by the stem elongation and booting. Salinity stress also affected grain yield and therefore was an important parameter for deriving PFs. In general, all the PFs were in agreement concerning the sensitivity of spring wheat to water stress. The neural network-based PF performed the best with a root mean square error equal to 44.27 g m^{-2} while the decision tree-based PF ranked fourth out of six in terms of accuracy. The most important advantage of the neural network-based PF was the flexible number of input parameters.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The agriculture sector is the major consumer of water in arid and semi-arid regions of Iran. Consequently, agricultural water management is the most important and sensitive part of any integrated water consumption reduction plan. Practicing deficit irrigation (DI) and using saline water for irrigation are among the most frequently used methods for overcoming water shortages. However, since both salinity and drought reduce the availability of soil water for crops, yield reduction needs to be predicted accurately (Domínguez et al., 2011). Moreover, salinity is a natural consequence of long term DI practices in arid and semi-arid regions.

A considerable number of studies have been published on salinity and DI of winter wheat in Iran, e.g. Dehghanisanji et al. (2009), Gowing et al. (2009), Kiani and Abbasi (2009), yet there are few studies focusing on DI for spring wheat. However, on a world-wide level, spring wheat is a critical crop because water use is reduced since it is harvested before the high evaporative demand of summer and water use productivity is increased since it can take advantage of spring rainfall (López-Urrea et al., 2009).

The mathematical estimation of yield response is called a production function (PF). PFs are very practical in irrigation scheduling and are useful in economic analysis. Traditional PFs like those derived by Jensen (1968) and Nairizi and Rydzewski (1977) use detailed relative evapotranspiration (ET) or irrigation water requirement (IW) in relationship to specified crop-growth stages as the independent variables. Recently researchers have incorporated more independent variables to derive PFs. For instance, Kuang et al. (2012) reformulated some well-known PFs for considering waterlogging stress on corn. Also Ai-hua et al. (2012) investigated fertilization as an independent variable for deriving rice PF. While utilizing more input predictors introduces more complexity and non-linearity in the process, this complexity may be required in to convey an accurate assessment of many situations. However, the traditional regression-based empirical equations have a lower modeling capability to precisely model the non-linear relationships for complex ecological systems (Dai et al., 2011).

Nearly all of the previously published well-known PFs are regression-based. Recently, data mining (DM) procedures, like

* Corresponding author at: Dept. of Biosystems Engineering and Soil Science, University of Tennessee, 2506 E.J. Chapman Dr., Knoxville, TN 37996-4531, USA. Tel.: +1 8652359694.
E-mail address: ahaghver@utk.edu (A. Haghverdi).

0168-1699/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compag.2013.12.009

neural network (NN) and decision tree (DT) are being employed as alternative options for modeling complex non-linear systems (Dai et al., 2011; Haghverdi et al., 2012). Mucherino et al. (2009) noted an increased number of DM studies in agriculture and expects more growth in the future. Moreover, Huang et al. (2010) mentioned the successful modeling result of DM-based procedures in the variety of investigation domains such as soil and water, crop management, soil physics, and precision agriculture. Recently, some studies have utilized NN to model yield levels for different crops (e.g. Fortin et al., 2010; Dai et al., 2011; Ehret et al., 2011). Dai et al. (2011) used NN to simulate the response of sunflower crop yield to soil moisture and salinity. They found that NN is more accurate than regression method for estimating crop yield using soil moisture and salinity at different crop growth stages as input predictors.

Above studies have confirmed the potential of NN for precisely predicting the yield of different crops by means of soil data (Dai et al., 2011) and weather and crop data (Fortin et al., 2010), yet no attempt has been done to predict yield variation in respect to quality and quantity of irrigation water. This study aims to address the need of a robust site specific empirical model for irrigation scheduling which would transfer the data that we have (i.e. easy collected irrigation related information) to what we need (i.e. the amount of yield). The specific objectives of the current study are: (1) to develop DM-based PFs for spring wheat using irrigation water salinity (EC) and IW at different growth stages as input predictors; (2) to identify the domain of influence of each input predictor; (3) to estimate the parameters of some previously well-known regression-based PFs; and (4) to evaluate the performance of DM-based PFs in comparison with regression-based PFs in northeast of Iran.

2. Materials and methods

2.1. Site description and general information

The location of the study was Mashhad region in northeastern Iran. A two year cropping study, 2010–2011, was conducted at the research farm of Ferdowsi University of Mashhad at 36° 16′ N latitude, 59° 38′ E longitude, 985 m above sea level. The spring wheat cultivar (*Triticum-aestivum*) was planted on the 13th of March 2010 and on the 17th of March 2011. Weather data and soil properties for the site are shown in Table 1. From 1985–2010, the mean annual precipitation, the mean minimum and mean maximum annual temperature and the mean annual relative humidity were 265 ± 72.7 mm, 8 ± 1.2 °C, 22 ± 0.9 °C and 55 ± 5.7%, respectively.

The surface (0–40 cm) and the subsurface (40–100 cm) layers are silt loam and clay loam, respectively. Soil sampling was done in the root zone, up to 100 cm at 20 cm intervals, in five different locations, corners and the center of the field. The soil texture and bulk density were measured using the hydrometer method (Gee and Bauder, 1986) and the soil clods method (Blake and Hartge, 2002), respectively. The water content at field capacity (FC, −33 kPa), and permanent wilting point (PWP, −1500 kPa), were estimated using k-Nearest software (www.ars.usda.gov). The k-Nearest software is a soil hydraulic pedotransfer function tool for the estimation of soil water contents at FC and PWP using the k-nearest neighbor approach (Nemes et al., 2008). The estimated water contents at different sampling points were averaged and were used for IW scheduling. The normal irrigation interval in the Mashhad region is 12 days. However for avoiding undesired water stress, a 10 day irrigation interval was applied for the whole irrigation season. The amount of IW for each plot was accurately applied using a volumetric water flow meter sensitive to 0.1 L. There were two sources of saline water and fresh water available with EC equal to 0.5 and 10 dS m^{-1}, respectively. The rest of the salinity levels were applied by combining these sources in storage tanks. Studying the effect of initial soil salinity on wheat yield was out of the scope of this study. The field of study had not been irrigated with saline water before the experiment hence had no initial salinity problem in the first year of the experiment. In the second year the location of the plots were slightly adjusted the way that those plots which were irrigated in the first year with saline water were not inside the experimental region anymore.

Table 1
Weather parameters and soil information during the cropping seasons (i.e. from March to July), 2010 and 2011, in the experimental location, Mashhad region of Iran.

Weather parameters	2010	2011
Average of minimum daily temperature (°C)	13.57	12.99
Average of maximum daily temperature (°C)	26.96	26.51
Average of monthly precipitation (mm)	24.93	56.94
Average of minimum daily relative humidity (%)	27.81	35.87
Average of maximum daily relative humidity (%)	64.75	59.90
Soil characteristics (cm)[b]	0–40	40–100
Texture	Silt loam	Clay loam
Bulk density (g cm^{-3})	1.37	1.48
Water content at FC (cm^3 cm^{-3})[a]	0.31	0.32
Water content at PWP (cm^3 cm^{-3})[a]	0.10	0.16

[a] FC: field capacity; PWP: permanent wilting point.
[b] Field had no initial salinity problem.

For both cropping seasons, the salinity and DI were applied after appearance of the third leaf of the crop. Before that, all of the plots were fully irrigated with the same amount of non-saline water. Crop disease and pest management, fertilizer supplements and tillage practices were identical following research-farm recommendations. Harvest was done on the 22 of June 2010 for the first cropping season and from 30 June to 4 July 2011 for the second cropping season. The plots were hand harvested utilizing the center of each plot (i.e. 1 m^2) to eliminate the possible edge effect of neighboring plots.

2.2. Experimental designs in the first year

A four-factor, two level unreplicated factorial design was employed in the first year. The variables were IW at different wheat growth stages (i.e. seedling growth-tillering (stage 1), stem elongation-booting (stage 2), heading-flowering (stage 3), and dough stage-ripening (stage 4)) and the variable levels were 20% and 100% of IW. In addition, the central point (60% of IW at all growth stages) with two replications was applied. The experiment was replicated for two different EC, 0.5 and 10 dS m^{-1}. Two level factorial designs usually are utilized for screening many factors to discover the vital factors, and how they interact (Myers et al., 2009). The purpose of the first year design was to identify the region of interest of the variables in order to design the second year experiment. Monetary and labor resources for doing field experiments are usually limited, thus the number of replications are typically low. Consequently, available resources only allowed an unreplicated design for these experiments, otherwise some of the original factors would have been omitted. For estimating error, the mean squares of high-order interactions were combined, based on the sparsity of effects principle (Myers et al., 2009). The growth stages were identified weekly based on Zadoks et al. (1974) growth stages code (Table 2). There were 36 (2 m × 2.1 m) plots within 4 rows with 2 m intervals between rows and 1 m interval between plots in each row, for excluding the side effects.

During the first year, IW was calculated based on time-domain reflectometry (TDR) readings from TRASE Model 6050X1 probes (Soil Moisture Equipment, Santa Barbara CA, USA). Prior to applying treatments, four moisture probes were placed at the 20, 40, 70, and 100 cm soil depths in each plot. The soil moisture from the 100% IW treatment using non-saline water was used to

Table 2
Growth stages of spring wheat in Mashhad region and corresponding irrigation water.

Growth stages	Symbol	Zadoks growth stages	1st year irrigation (mm)[b]	2nd year irrigation (mm)
Beginning	0	Emergence-seedling growth	40, 40	40, 40
Beginning[a]	1	Seedling growth-tillering	36, 52	36, 54
Middle	2	Stem elongation-booting	44, 84	63, 87
Middle	3	Heading-flowering	62, 112	96, 122
End	4	Dough stage-ripening	148	150

[a] Beginning of the deficit and saline treatments.
[b] The irrigation values belong to the full irrigation treatment which was applied by means of both non saline (0.5 dS m^{-1}) and saline (10 dS m^{-1}) water resources.

determine the irrigation amounts of all treatments. Soil moisture was measured the day before irrigation and IW requirement was calculated as the difference between actual water content and FC in the root zone. Based on the previous local observations, the maximum root zone depth for spring wheat was assumed to be 1 m at the last irrigation, approximately two weeks before harvesting, with a linear growth rate during the cropping season. Crop evapotranspiration (ET) was calculated by the water balance equation using TDR measured data. Since the groundwater table is very deep (>70 m) and measured irrigation prevented over application, water movement between the root zone layer and deeper layers was ignored. Also plots were sheltered for excluding the effect of rainfall and were surrounded by earth dykes, 30 cm in height, in order to prevent the lateral spread of irrigation water.

2.3. Experimental designs in the second year

A five-factors, five level central composite design (CCD) was employed during the second year (Myers et al., 2009). The variables were IW at 4 different growth stages, same as the first year, and EC. Table 3 shows the detailed information about the second year experiment including applied levels of each variable. The irrigation levels were 30%, 40%, 65%, 90% and 100% of water requirement and salinity levels were 0.5, 1.8, 5.25, 8.6 and 10 dS m^{-1}. The established CCD consists of 3 components: an unreplicated factorial design, axial runs and central point. In addition to CCD, a rainfed and a full irrigation treatment each with 2 replications were applied. Comparing to the first cropping year, the number of plots were increased to 52 while the size and structure of the plots and distance between them and between rows were similar to the first cropping year. However in the second year, plots were not sheltered and TDR probes were not used because the number of plots was high while available financial and labor resources were limited. In fact, this year volumetric sampling was done the day before each irrigation event from the main plot, full irrigation with non-saline water, for calculating the amount of IW. During the second year, root zone salt distribution was monitored by measuring the salinity of saturated paste from samples which were randomly gathered from different plots at different growth stages at two different depths, 0–30 cm and 30–60 cm. Initial random sampling before applying treatments showed the soil profile salinity level of the whole experimental area was uniform and negligible.

2.4. Well known regression-based PFs

Four well known PFs were used in this study: The Jensen-PF ((Eq. (1)) –Jensen, 1968); the Minhas-PF (Eq. (2) – Minhas et al., 1974); the modified Stewart-PF ((Eq. (3)) – Stewart et al., 1977 and Stegman et al., 1980), and the Nairizi-PF (Eq. (4) – Nairizi and Rydzewski, 1977).

Table 3
Detailed information of experimental plots, levels of variables, at the second cropping season.

Plot[a]	Variable					Plot	Variable				
	1	2	3	4	EC		1	2	3	4	EC
Unreplicated factorial design											
1	0.4F	0.4F	0.4F	0.4F	1.89	17	0.4F	0.4F	0.4F	0.4F	8.61
2	0.4F	0.9F	0.4F	0.4F	1.89	18	0.4F	0.9F	0.4F	0.4F	8.61
3	0.4F	0.4F	0.9F	0.4F	1.89	19	0.4F	0.4F	0.9F	0.4F	8.61
4	0.4F	0.9F	0.9F	0.4F	1.89	20	0.4F	0.9F	0.9F	0.4F	8.61
5	0.4F	0.4F	0.4F	0.9F	1.89	21	0.4F	0.4F	0.4F	0.9F	8.61
6	0.4F	0.9F	0.4F	0.9F	1.89	22	0.4F	0.9F	0.4F	0.9F	8.61
7	0.4F	0.4F	0.9F	0.9F	1.89	23	0.4F	0.4F	0.9F	0.9F	8.61
8	0.4F	0.9F	0.9F	0.9F	1.89	24	0.4F	0.9F	0.9F	0.9F	8.61
9	0.9F	0.4F	0.4F	0.4F	1.89	25	0.9F	0.4F	0.4F	0.4F	8.61
10	0.9F	0.9F	0.4F	0.4F	1.89	26	0.9F	0.9F	0.4F	0.4F	8.61
11	0.9F	0.4F	0.9F	0.4F	1.89	27	0.9F	0.4F	0.9F	0.4F	8.61
12	0.9F	0.9F	0.9F	0.4F	1.89	28	0.9F	0.9F	0.9F	0.4F	8.61
13	0.9F	0.4F	0.4F	0.9F	1.89	29	0.9F	0.4F	0.4F	0.9F	8.61
14	0.9F	0.9F	0.4F	0.9F	1.89	30	0.9F	0.9F	0.4F	0.9F	8.61
15	0.9F	0.4F	0.9F	0.9F	1.89	31	0.9F	0.4F	0.9F	0.9F	8.61
16	0.9F	0.9F	0.9F	0.9F	1.89	32	0.9F	0.9F	0.9F	0.9F	8.61
Axial unreplicated runs											
33	0.65F	0.65F	0.65F	0.65F	0.5	38	F	0.65F	0.65F	0.65F	5.25
34	0.65F	0.65F	0.65F	0.3F	5.25	39	0.65F	F	0.65F	0.65F	5.25
35	0.65F	0.65F	0.3F	0.65F	5.25	40	0.65F	0.65F	F	0.65F	5.25
36	0.65F	0.3F	0.65F	0.65F	5.25	41	0.65F	0.65F	0.65F	F	5.25
37	0.3F	0.65F	0.65F	0.65F	5.25	42	0.65F	0.65F	0.65F	0.65F	10
Central replicated runs											
43	0.65F	0.65F	0.65F	0.65F	5.25	45	0.65F	0.65F	0.65F	0.65F	5.25
44	0.65F	0.65F	0.65F	0.65F	5.25	46	0.65F	0.65F	0.65F	0.65F	5.25
Additional runs											
47	0	0	0	0		50	F	F	F	F	0.5
48	0	0	0	0		51	F	F	F	F	10
49	F	F	F	F	0.5	52	F	F	F	F	10

[a] F: Full irrigation; 1, 2, 3, 4: irrigation water (% of full irrigation treatment) at different growth stages (identical to Table 2), EC: irrigation water salinity (dS m^{-1}).

$$\frac{Ya}{Ym} = \prod_{i=1}^{n} \left(\frac{ETa_i}{ETm_i}\right)_i^{\lambda_i} \quad (1)$$

$$\frac{Ya}{Ym} = \prod_{n=1}^{n} \left[1 - \left(1 - \left(\frac{ETa}{ETm}\right)_i\right)^2\right]^{b_i} \quad (2)$$

$$\left(1 - \frac{Ya}{Ym}\right) = \sum_{i=1}^{n} Ky_i \left(1 - \frac{ETa}{ETm}\right)_i \quad (3)$$

$$\frac{Ya}{Ym} = \prod_{i=1}^{n} \left(\frac{IWa_i}{IWm_i}\right)_i^{\gamma_i} \quad (4)$$

where *Ya*, *ETa* and *IWa* are the grain yield (GY), evapotranspiration, and IW from stressed treatments; *Ym*, *ETm* and *IWm* are grain yield, evapotranspiration, and IW from the non-stressed treatment; *i* shows the growth stage; λ, δ and γ are the Jensen's, Minhas' and Nairizis' moisture stress sensitivity indices, respectively; *Ky* is the modified Stewart's moisture stress yield reduction coefficient; *n* is the number of growth stages; and $\sum$ is the additive sign and Π is a multiplicative sign. Since the first three PFs needed ET information, they were derived only using the data from the first cropping season while the fourth Equation, i.e. Nairizi-PF, was established using the combined data from both the first and the second cropping seasons. Regardless of salinity levels, all of the plots were combined and randomized prior to deriving the PFs. From this data set, 75% was used for deriving PFs and 25% for testing the accuracy of the derived PFs. For identifying moisture-stress sensitivity indices, the equations were transformed into a multiply linear function, in which the indices were the coefficients and the relative IW and ET were the input predictors and relative GY was the output predictor. No attempt was made to modify the four regression-based PFs with a salinity predictor. However, salinity effects are incorporated into the ET-based PFs because ET will be reduced by high salt concentrations. The amount of precipitation in the second year at the period of applying treatments was very small (less than 4% of the applied IW) thus contribution of it in fulfilling the IW requirement was assumed to be negligible.

2.5. Decision trees to derive PF

Decision tree (DT) is a well-known data mining procedure. DT has a top-down branched structure containing some if-then rules for modeling the desired attribute in regard to the relative importance of input predictors in the system under study (Huang et al., 2010). The tree is fitted to a dataset by separating the data into homogeneous subsets in regard to the input predictors, and the output is predicted by the tree leaves for all samples. Finally, the top-down pruning process is used to improve the generalization ability of the tree for classifying the new samples and helping to avoid over-fitting (Witten et al., 2011).

In the current study, the Classification and Regression Tree (CART) (Breiman et al., 1984) was employed to derive the PF. CART models repeatedly partition the data to find increasingly homogeneous subsets based on input predictors splitting criteria using variance minimizing algorithms. The number and combination of input predictors is user-defined when one uses DM techniques. That is why in addition to the relative IW at different growth stages, EC was utilized as an input predictor for deriving DT-PF. The first and the second cropping season's data were combined and randomized and then divided into two groups for development (75% of data) and for testing (25%). Maximum tree depth was set to five. Minimum records in the parent and child branches were set to 2% and 1% of data, respectively as the stopping criteria for splitting specific branches of the tree. Furthermore the minimum change in impurity was set to 0.0001; thus if the best split for a branch reduced the impurity by less than the specified amount, the split was not be made.

2.6. Neural networks to derive PF

A Neural Network (NN) is a non-linear statistical data mining method capable of modeling complex relationships between input and output predictors. In the other words, NN is a parallel structured modeling tool inspired by the function of the human brain. Its effectiveness is derived from the fact that it does not require a priori functional form to relate input predictors to the outputs (Mucherino et al., 2009).

A three-layer perceptron, the most widely adopted network used to map input–output relationships (Maren et al., 1990), was used in this study. The NN architecture consisted of an input layer, a hidden layer, and an output layer. The tangent hyperbolic was chosen as the activation function for the hidden layer nodes, which helped in non-linearly by transforming the inputs into an alternative space where the training samples were linearly separable (Brown and Harris, 1994), while the linear activation function was used for the output layer. The Levenberg-Marquardt algorithm (Shepherd, 1997; Demuth and Beale, 2000; Pulido-Calvo and Portela, 2007) was selected for the network training process. All NN modeling steps were performed using Neurosolution 5.07 (www.nd.com) software evaluation version. After randomly combining the first and second cropping season's data, 75% was used for developing PFs and 25% for testing the accuracy of the derived PFs. To avoid overtraining, the development data set was divided into two individual parts: for training (60% of whole data) and for cross-validation (15% of whole data). The cross-validation part stopped the training process using supervised learning control. The default number of iterations for training was set at 1000. The training process stopped when the mean square error of the cross-validation set began to increase. The number of neurons in the hidden layer changed from 1 to 15 and the training was repeated 3 times for each number of hidden neurons (Iyer and Rhinehart, 1999). The input and output predictors were identical with DT-PF.

Sensitivity analysis was done after deriving NN-PF to identify the level of importance of each input predictor on GY modeling. In fact, sensitivity analysis was a method for extracting the cause and effect relationship between the inputs and outputs of the network. Standard deviations were added and subtracted from the mean of each of input predictor. The trained network then was used to calculate the variation in GY corresponding with the variation of each individual input predictor in 100 equal steps when the rest of the predictors were set to be constant and have the average of their magnitudes.

2.7. Evaluation criteria

The procedure for deriving PFs, and therefore testing them, was repeated 4 times to improve the reliability of the results by involving all of the data in testing phase. The whole data set was randomly divided to 4 equal parts and each time 3 parts were selected for derivation while the remaining part was used for testing. The program IRENE (www.isci.it/tools) was adopted for calculating the selected statistics, i.e. root mean square of error (RMSE), correlation coefficient (*r*) and mean bias error (MBE), for evaluating the performances of the PFs as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (E_i - M_i)^2}{n}} \quad (5)$$

$$r = \frac{\sum_{i=1}^{n}(M_i - \bar{M})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2 \sum_{i=1}^{n}(E_i - \bar{E})^2}} \quad (6)$$

$$\mathrm{MBE} = \sum_{i=1}^{n} \frac{E_i - M_i}{n} \quad (7)$$

where M_i is the measured GY; E_i is the estimated GY for ith treatment; $\bar{M}$ is the average of measured GY; $\bar{M}$ is the average of estimated GY and n is the number of treatments.

3. Results

3.1. PF indices and predictor variables importance

The indices and coefficients of the PFs are illustrated in Table 4. The presented indices for each individual PF in fact are the mean and standard deviation of the four times deriving procedure. The highest indices and coefficients occurred at the middle stages (Table 2) for all PFs corresponding to heading and flowering, followed by the stem elongation and booting growth stages. In general, the indices and coefficients at the beginning and at the end of the cropping season were lower than the middle stages of the season. Indeed the lowest indices belonged to the first growth stage for all of the PFs except for the Nairizi-PF.

The structure of DT-PF is presented in Fig. 1. The nods were labeled by corresponding predictor names while the leaves of the tree were labeled by the predicted GY value. The splitting values were also written for the top four layers to make the topology of the tree more clear. The tree in Fig. 1 was derived using the whole data set in order to be more generalized while the actual trees were established by using training data, i.e. 75% of the whole data set.

Table 4
The indices and coefficients of the PFs for the different growth stages of spring wheat in Mashhad region.

CWPF	Growth stages[a]			
	1	2	3	4
Jensen (λ)	0.11 (0.02)[b]	0.34 (0.05)	0.43 (0.04)	0.17 (0.06)
Minhas (δ)	0.25 (0.04)	0.62 (0.10)	0.73 (0.07)	0.31 (0.10)
M-Stewart (Ky)	0.11 (0.03)	0.40 (0.06)	0.43 (0.05)	0.18 (0.06)
Nairizi (γ)	0.19 (0.02)	0.26 (0.03)	0.40 (0.02)	0.16 (0.01)

[a] The growth stages are identical with the Table 2. M-Stewart: modified Stewart.
[b] Numbers in parenthesis are the standard deviation among the derived coefficients for four times deriving procedure.

Therefore, there are some minor differences between the topology of this tree and the 4 individual trees that were derived. The first and the second levels were divided in regard to the relative IW of the third and the second growth stages, respectively. The branching value of the nods in the first level is 0.45 of relative IW at heading and flowering (third growth stage) while 0.52 and 0.87 of IW at stem elongation and booting (second growth stage) were the splitting values at the second level of the left and right branches, respectively. The rest of the predictors (i.e. EC, IW at first and forth growth stages) appeared in the next layers. The EC in the fourth and fifth levels of the right main branch was available at several nodes, but the IW at the first and at the end of the growing seasons occupied all of the nodes in the same position of the left main branch. The magnitude of branching of EC decreased moving from the root toward the leaves of the tree.

Results of the NN-PF sensitivity analysis are presented in Fig. 2. The result was calculated upon training data set of NN-PFs. Columns show the mean of the sensitivity for the four repeated derivations including the standard deviation. The IW at the third level had the highest impact on GY while the salinity had the lowest effect on the modeling.

3.2. Model evaluation

The evaluation statistics and the scatter plots, measured versus estimated GY, of the PFs are presented in Table 5 and Fig. 3, respectively. The first three PFs were derived using the ET information of the plots from the first cropping year experiment while the last three were developed using IW values from combination of the plots of both cropping years. Overall, the performance of the derived PFs using both years of data, i.e. IW-based PFs, are better than those which were derived only using data of first year, i.e. ET-based PFs. Among the PFs, which were established using IW information, the best result was due to the NN-PF, with RMSE equal to 44.27 g m^{-2}, while the performance of the Jensen-PF, with RMSE equal to 52.64 g m^{-2}, was better than the rest of well-known PFs that were derived using ET data from the individual plots. The high r value of all PFs, ranged from 0.85 to 0.90, was in agreement with the linear trends shown in the scattered plots of Fig. 3. The MBE values were low indicating there was no systematic over/under estimation on any of PFs (See Fig. 3). This was considered as a fairly accurate performance of the PFs in general. However, the same pattern of error was recognizable in ET-based PFs; overestimation for some plots started around 200 g m^{-2} and continued for higher GY values (Fig. 3).

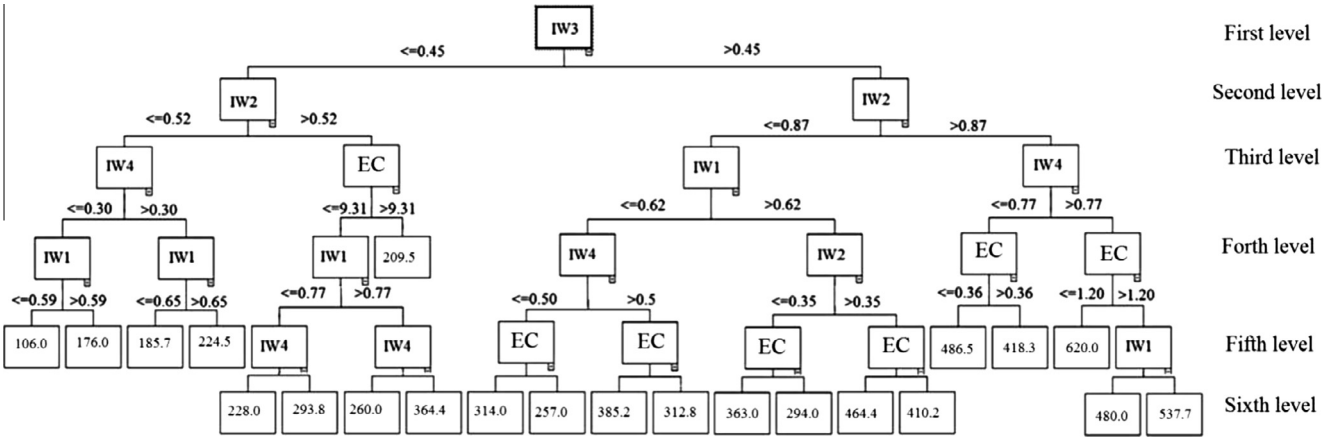

Fig. 1. Schematic structure of the derived DT- PF applied to spring wheat. EC: irrigation water salinity (dS m^{-1}); IW1, IW2, IW3, IW4: relative irrigation water requirement at different growth stages (Table 2). The values inside the leaves are the predicted GY (g m^{-2}).

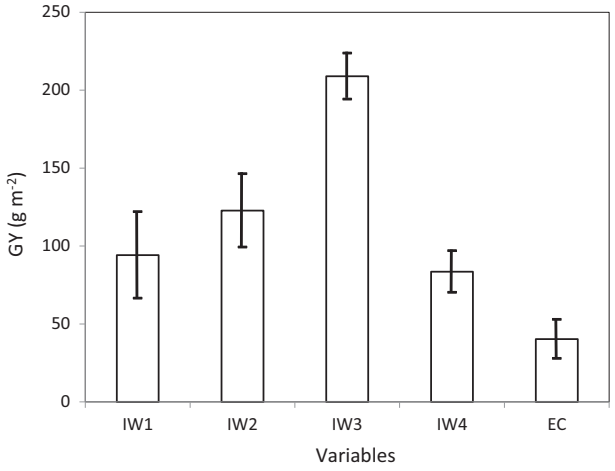

Fig. 2. Sensitivity of the grain yield (GY) to each input predictor calculated using the training data set of NN-PF. IW1, IW2, IW3 and IW4 are the applied irrigation water at different growth stages (Table 2) and EC is the irrigation water salinity.

Table 5
Evaluation statistics for all of the PFs in test set.

PF	Data	Predictors[b]	RMSE (g m^{-2})	MBE (g m^{-2})	r
Jensen-PF	S1	ET_{Rel}	52.64	5.22	0.87
Minhas-PF	S1	ET_{Rel}	63.16	22.60	0.88
M-Stewart-PF[a]	S1	ET_{Rel}	60.82	11.58	0.85
Nairizi-PF	S1 + S2	IW_{Rel}	50.85	9.09	0.90
NN-PF	S1 + S2	IW_{Rel} + EC	44.27	−2.99	0.92
DT-PF	S1 + S2	IW_{Rel} + EC	60.02	9.56	0.86

[a] M-Stewart: modified Stewart. S1: first cropping season data; S2: second cropping season data.
[b] ET_{Rel}: relative ET; IW_{Rel}: relative IW.

4. Discussion

In general, the indices and coefficients at the beginning and the end of the cropping season were lower than those at the middle stages of the season. This trend was observed in PFs which used either ET or IW as the input predictor. The priorities in the branching of the DT-PF (Fig. 1) and the result of sensitivity analysis for the NN-PF (Fig. 2) were in agreement with the indices and coefficients of regression-based PFs (Table 4). These findings reiterated the findings of previous studies: the impact of water stress was most keenly realized in winter wheat during the flowering/grain filling, followed by stem elongation/booting (García del Moral et al., 2003; Karam et al., 2009). The differences between the studies may be related to the differences in weather and soil from location to location. Moreover, the variation of the indices and coefficients during the season indirectly followed the routine shape of the crop coefficient curve derived for spring wheat by López-Urrea et al. (2009). They found that maximum ET was reached around May 25 and decreased during the ripening period as the growing season advanced. According to the Zadok's growth stages in Table 2, the middle stages began with stem elongation and were ended with flowering. Reduction in the GY around flowering by water stress may be due to the declining spike and spikelet number and the fecundity of remaining spikelets (Karam et al., 2009).

An earlier assessment of well-known PFs by Igbadun et al. (2007) found better performance of the Jensen-PF over the Minhas and modified-Stewart PFs in an irrigated maize crop. The present study also showed better performance of the Jensen PF for spring wheat. Among the regression-based PFs, however, the Nairizi-PF worked slightly better than the others. The better result of Nairizi-PF could be related to the number and distribution of the data that was used to derive it. In other words, Nairizi-PF was derived based on data from both cropping years; whereas the other functions just used the information from the first year. The higher number of plots means more information about the experimental

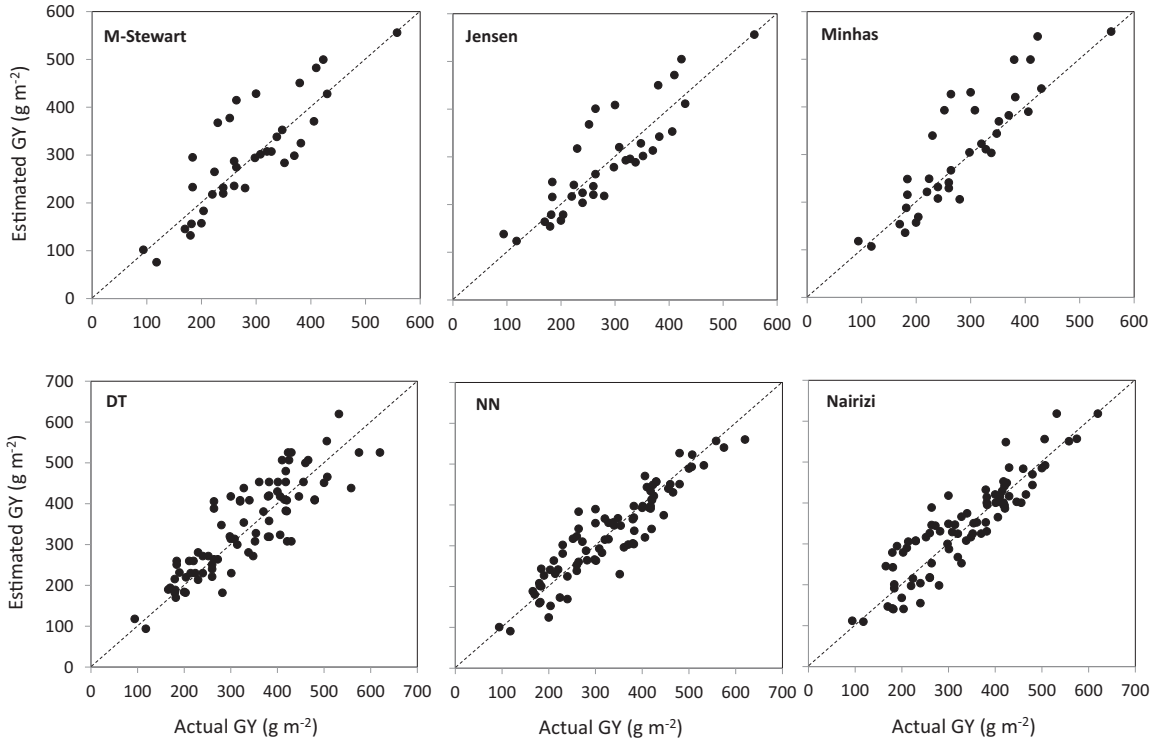

Fig. 3. Scatter plots of measured versus estimated grain yield (GY) (g m^{-2}) for all of the PFs in test sets for spring wheat in Mashhad region.

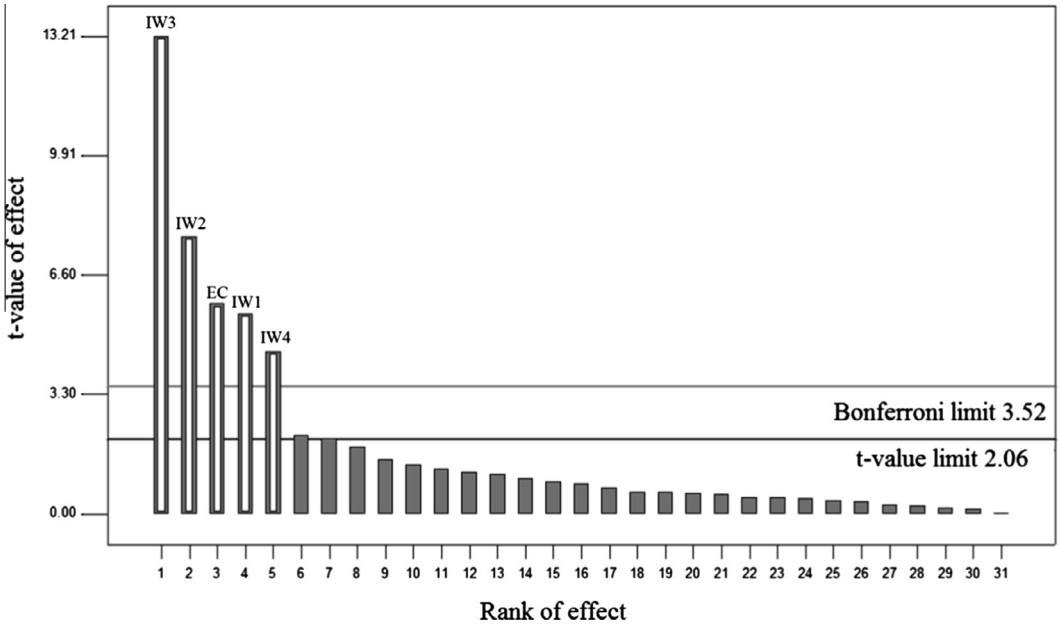

Fig. 4. Pareto chart indicating the level of importance of the input predictors and the possible interactions for the experiment of the first cropping year. IW1, IW2, IW3 and IW4 are the irrigation amount at different growth stages (Table 2) and EC is the irrigation water salinity (This figure was drawn using the trial version of Design-Expert 7 software (Stat-Ease, Minneapolis, MN)).

domain and thus a more accurately derived PF. Performance of Nairizi-PF also showed IW can be used as a valid input predictor. This finding may be valid only if the over irrigation was eliminated by accurate estimation of IW and therefore IW was close to ET.

NN-PF showed the highest accuracy, but DT-PF ranked fourth among all six PFs. RMSE for relative grain yield value was 0.07 (=44.27 g m^{-2}) for NN-PF, lower than the reported error by Igbadun et al. (2007), i.e. RMSE for relative grain yield was equal to 0.11 for Jensen, 0.20 for Minhas and 0.15 for M-Stewart. Dai et al. (2011) recently introduced the NN as more accurate model than multiple linear regression for modeling sunflower responses to soil moisture and salinity. The better performance of NN-PF in comparison with regression-PFs in this study, together with the result of Dai et al. (2011), implies that the NN could be used as a reliable alternative option to regression for deriving the next generation of PFs. The fixed number of predictors is the most remarkable handicap of the well-known regression-PFs that restrict their application. In contrast, there is less limitation for choosing input predictors when using NN-PF. This gives investigators the unique opportunity to design future studies that apply different types of input predictors. In fact, the better performance of NN-PF may be related to the usefulness of EC as an input predictor.

Soil salinity was already used as an input predictor in PF studies to represent the magnitude of salinity tension (e.g. Yang-Ren et al., 2007). Identifying the accumulation of salt in root zones, however, is not easy and requires time and costly equipment. In this study, random monitoring of soil salinity at two different depths (0–30 cm and 30–60 cm), during the second season in plots under various salinity stress showed there was a high positive correlation, $r = 0.95$, between the soil salinity and EC. Although less pronounced, the correlation between the quantities of salt (i.e. EC multiply by the total irrigation throughout the season after applying saline water) and accumulated salt in root zone was positive as well, $r = 0.69$. As a result, the hypothesis of considering EC instead of soil salinity as an input predictor is logically acceptable. Indeed, the EC directly affects the salt accumulation in the root zone which inhibits water uptake by roots. For soils with long term salinity problems, considering EC as an input predictor may be problematic because it does not reflect the initial salinity distribution within the root zone hence it is not able to represent the accumulated salts in the root zone. It should be mentioned that the study field had no initial salinity problems; therefore, this subject was not an issue in this research. Due to the flexible structure of NN-PF, however, it seems that including initial soil salinity as an additional input predictor on top of IW and EC can help one to derive the PF for soils which are struggling with long term salinity problem. Further investigation is needed to evaluate the effectiveness of initial soil salinity as an input predictor.

Although the black box nature of NN-PF did not reveal any information, the white box nature of DT-PF was useful in interpreting the role of each predictor. In the left branch of DT-PF (Fig. 1), water stress was the prevailing tension reducing GY; hence, the presence of EC as an input attribute was reduced to a single node even with a high salinity value; i.e. 9.31 dS m^{-1}. As expected, in the right branch of the tree with low water stress, EC as a predictor was more important. It could be stated that the role of salinity as an input predictor in the structure of PF in the treatments with high water stress was not as important as the treatments with moderate and low water stress. In fact the high water stress seemed to be more effective in reducing yield than salinity. The branching value of salinity in the left part of the tree was higher than the threshold salinity of wheat, 6 dS m^{-1} (Maas and Grattan, 1999), which may be due to the progressive effect of water stress as a predominant factor. However, based on Gowing et al. (2009), the idea of existing threshold salinity for crops by which the yield remains constant may be abandoned. The presence of some nodes that used very low amounts of salinity, although in less critical lower branches of the right part of the tree, may support this idea.

Reasonable accuracy of the derived PFs in this study may be related to the high number of observations as a result of using response surface methodology instead of traditional experimental designs. Discarding the replications provided this opportunity to examine more pair points in the experimental space of salinity and DI. Although discussion about statistical aspects of the adopted response surface design is out of the scope of current research, Fig. 4 supports the *sparsity of the effect* principle and, therefore,

the statistical validity of the un-replicated designs that were established. Indeed, Fig. 4 shows the pareto chart of the *t*-value of the effects, i.e. effects of predictors and their possible interactions, based on the result of the first cropping year. It also contains two t-limits (i.e. the t-limit and Bonferroni adjusted t-limit) as reference points to judge the significances of effects. In this figure, all of the higher order effects are below the Bonferroni limit and t-Value limit, meaning they were not likely to be significant at the 5% risk level.

5. Conclusion

Decision Tree (DT) and Neural Network (NN) methods were employed to estimate spring wheat grain yield (GY) under simultaneous salinity and water stress. These Data Mining (DM) methods were compared with four well known production functions (PFs) (i.e. Nairizi, Jensen, Minhas and M-Stewart). The performance of the NN-PF was better than the other PFs. The ability to introduce new input predictors is an important advantage of DM-based PFs over the existing well-known regression-based PFs. Also, utilizing response surface methodology instead of traditional experimental design, provides an opportunity to employ un-replicated but statistically valid designs in a manner that allows for a greater range of input variables with limited plot space. This is especially important because multifactor NNs require extensive data sets. In addition, evapotranspiration (ET) and soil salinity, which are difficult to measure, could be replaced with irrigation water (IW) and irrigation water salinity (EC) as more easily collected PF predictors with reasonable accuracy. However, more investigation is needed to confirm the usefulness of this replacement under different initial and boundary conditions.

References

Ai-hua, S., Shi-jiang, Z., Ya-fen, G., Zhong-xue, Z., 2012. Jensen model and modified Morgan model for rice water-fertilizer Production Function. Proc. Eng. 28, 264–269.

Blake, G.R., Hartge, K.H., 2002. Bulk density. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis. Part 4. Physical Methods. SSSA Book Ser., 5. American Society of Agronomy, Madison, USA, pp. 363–375.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall/CRC, New York, USA.

Brown, M., Harris, C., 1994. Neurofuzzy Adaptive Modeling and Control. Prentice Hall, New York, USA.

Dai, X., Huo, Z., Wang, H., 2011. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. Field Crops Res. 121, 441–449.

Dehghanisanji, H., Nakhjavani, M.M., Zeggaf Tahiri, A., Anyoji, H., 2009. Assessment of wheat and maize water productivities and production function for cropping system decisions in arid and semiarid reigns. Irrig. Drain. 58, 105–115.

Demuth, H., Beale, M., 2000. Neural Network Toolbox. Mathworks, Inc..

Domínguez, A., Tarjuelo, J.M., de Juan, J.A., López-Mata, E., Breidy, J., Karam, F., 2011. Deficit irrigation under water stress and salinity conditions: the MOPECO-salt model. Agric. Water Manage. 98, 1451–1461.

Ehret, D.L., Hill, B.D., Helmer, T., Edwards, D.R., 2011. Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data. Comput. Electron. Agric. 79, 82–89.

Fortin, J.G., Anctil, F., Parent, L., Bolinder, M.A., 2010. A neural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. Comput. Electron. Agric. 73, 126–132.

García del Moral, L.F., Rhrarrabti, Y., Villegas, D., Royo, C., 2003. Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: an ontogenic approach. Agron. J. 95, 266–274.

Gee, G.W., Bauder, J.W., 1986. Particle size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1. Agron. Monogr. 9. American Society of Agronomy, Madison, WI, USA, pp. 383–411.

Gowing, J.W., Rose, D.A., Ghamarni, H., 2009. The effect of salinity on water productivity of wheat under deficit irrigation above shallow groundwater. Agric. Water Manage. 96, 517–524.

Haghverdi, A., Cornelis, W.M., Ghahraman, B., 2012. A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. J. Hydrol. 442–443, 46–54.

Huang, Y., Lan, Y., Thomson, S.J., Fang, A., Hoffmann, W.C., Lacey, R.E., 2010. Development of soft computing and applications in agricultural and biological engineering. Comput. Electron. Agric. 71, 107–127.

Igbadun, H.E., Tarimo, A.K.P.R., Salim, B.A., Mahoo, H.F., 2007. Evaluation of selected crop water production functions for an irrigated maize crop. Agric. Water Manage. 94, 1–10.

Iyer, M.S., Rhinehart, R.R., 1999. A method to determine the required number of neural-network training repetitions. IEEE Trans. Neural Networks 10 (2), 427–432.

Jensen, M.E., 1968. Water consumption by agricultural plants. In: Kozlowski, T.T. (Ed.), Water Deficits in Plant Growth, 1. Academic Press, New York, USA, pp. 1–22.

Karam, F., Kabalan, R., Breidi, J., Rouphael, Y., Oweis, T., 2009. Yield and water-production functions of two durum wheat cultivars grown under different irrigation and nitrogen regimes. Agric. Water Manage. 96, 603–615.

Kiani, R., Abbasi, F., 2009. Assessment of the water-salinity crop production function of wheat using experimental data of the Golestan Province. Iran. Irrig. Drain. 58, 445–455.

Kuang, W., Xianjiang, Y., Xiuqing, C., Yafeng, X., 2012. Experimental study on water production function for waterlogging stress on corn. Proce. Eng. 28, 598–603.

López-Urrea, R., Montoro, A., González-Piqueras, J., López-Fuster, P., Fereres, E., 2009. Water use of spring wheat to raise water productivity. Agric. Water Manage. 96, 1305–1310.

Maas, E.V., Grattan, S.R., 1999. Crop yields as affected by salinity. In: Skaggs, R.W., van Schilfgaarde, J. (Eds.), Agricultural Drainage Agronomy No. 38. American Society of Agronomy, pp. 55–108.

Maren, A.J., Harston, C.T., Pap, R.M., 1990. Handbook of Neural Computing Applications. Academic Press, San Diego, USA.

Minhas, B.S., Parkhand, K.S., Srinivasan, T.N., 1974. Towards the structure of a production function for wheat yields with dated input of irrigation water. Water Resour. Res. 10, 383–386.

Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. Data Mining in Agriculture. Springer.

Myers, R.H., Montgomery, D.C., Anderson-Cook, C.M., 2009. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. third ed. John Wiley & Sons, NewYork, USA.

Nemes, A., Roberts, R.T., Rawls, W.J., Pachepsky, Ya.A., Van Genuchten, M.Th., 2008. Software to estimate -33 and -1500 kPa soil water retention using the non-parametric k-nearest neighbor technique. Environ. Model. Software 23, 254–255.

Nairizi, S., Rydzewski, J.R., 1977. Effects of dated soil moisture stress on crop yields. Exp. Agric. 13, 51–59.

Pulido-Calvo, I., Portela, M., 2007. Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. J. Hydrol. 332, 1–15.

Shepherd, A.J., 1997. Second-order Methods for Neural Networks. Springer, New York.

Stegman, E.C., Hanks, R.J., Musick, J.T., Watts, D.G. 1980. Irrigation water management-adequate or limited water. In: Challenges of the 80's. Proceedings of the ASAE, 2nd National Irrigation Symposium, October.

Stewart, J.L., Danielson, R.E., Hanks, R.J., Jackson, E.B., Hagon, R.M., Pruit, W.O., Franklin, W.T., Riley, J.P. 1977. Optimizing crop production through control of water and salinity levels in the soil. Utah Water Research Lab., PR. Logan, UT, pp. 151–1.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data Mining Practical Machine Learning Tools and Techniques. third ed. Elsevier.

Yang-Ren, W., Shao-Zhong, K., Fu-Sheng, L., Lu, Z., Jian-Hua, Z., 2007. Saline water irrigation scheduling through a crop-water-salinity production function and a soil-water-salinity dynamic model. Pedosphere 17 (3), 303–317.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421.