Journal of Hydrology: Regional Studies 35 (2021) 100832

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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

Application of artificial neural networks to the design of subsurface drainage systems in Libyan agricultural projects

Murad A. Ellafi [1], Lynda K. Deeks *, Robert W. Simmons

Cranfield Soil and Agrifood Institute, Building 52a, Cranfield University, Bedfordshire, MK 43 0AL, UK

ARTICLE INFO

Keywords:
Saturated hydraulic conductivity
Artificial neural networks
Agricultural drainage design
Pedotransfer functions
Sub-surface drainage
Arid areas

ABSTRACT

Study region: The study data draws on the drainage design for Hammam agricultural project (HAP) and Eshkeda agricultural project (EAP), located in the south of Libya, north of the Sahara Desert. The results of this study are applicable to other arid areas.

Study focus: This study aims to improve the prediction of saturated hydraulic conductivity (K_{sat}) to enhance the efficacy of drainage system design in data-poor areas. Artificial Neural Networks (ANNs) were developed to estimate K_{sat} and compared with empirical regression-type Pedotransfer Function (PTF) equations. Subsequently, the ANNs and PTFs estimated K_{sat} values were used in EnDrain software to design subsurface drainage systems which were evaluated against designs using measured K_{sat} values.

New hydrological insights: Results showed that ANNs more accurately predicted K_{sat} than PTFs. Drainage design based on PTFs predictions (1) result in a deeper water-level and (2) higher drainage density, increasing costs. Drainage designs based on ANNs predictions gave drain spacing and water table depth equivalent to those predicted using measured data. The results of this study indicate that ANNs can be developed using existing and under-utilised data sets and applied successfully to data-poor areas. As K_{sat} is time-consuming to measure, basing drainage designs on ANN predictions generated from alternative datasets will reduce the overall cost of drainage designs making them more accessible to farmers, planners, and decision-makers in least developed countries.

1. Introduction

In the Earth's terrestrial biosphere, soil plays an essential role in managing the mass and energy transferred between the earth and the atmosphere (Amundson et al., 2015; Bittelli et al., 2015). Upper soil horizons have hydraulic properties (governed by pore size distribution, water retention and hydraulic conductivity) that regulate the local water balance through, infiltration, evapotranspiration, surface runoff, and groundwater recharge (Zhang and Schaap, 2019). These soil properties have a fundamental effect on local, regional and global land surface water and energy balances (Montzka et al., 2017; Vereecken et al., 2016; Verhoef and Egea, 2014). For example, approximately 60 % of rainfall is returned to the atmosphere through the soil-plant-atmosphere continuum (Katul et al., 2012; Oki and Kanae, 2006), and > 50 % of the global biomass production and related carbon cycle depend on soil processes (Cleveland et al., 2013).

* Corresponding author.
E-mail address: l.k.deeks@cranfield.ac.uk (L.K. Deeks).
[1] Permanent Address: Department of Soil and Water Sciences, University of Tripoli, Tripoli 13538, Libya.

https://doi.org/10.1016/j.ejrh.2021.100832
Received 4 December 2020; Received in revised form 5 May 2021; Accepted 7 May 2021
Available online 18 May 2021
2214-5818/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Water for agriculture is often unregulated and overexploited. In spite of this, the general belief is that by 2025 irrigated lands have to be extended by 20–30 % in order to meet global food demands (FAO, 2002). However, the world's arable lands are degrading due to soil salinization, desertification, erosion, and urbanisation (FAO, 2002). In arid and semiarid areas water use efficiency is a major issue due to increasing cross-sector demand for limited water resources (Elshemy, 2018). Therefore, understanding soil hydrological behaviour is essential in the global drive to expand crop production (Abdelbaki et al., 2009). For example, drainage systems in arid and semiarid areas have been designed and installed to manage and control soil waterlogging and salinity problems in irrigation schemes. However, less than 30 % of all irrigated land needing drainage have been effectively drained (Schultz et al., 2007), and the situation is worse in least developed countries (OECD, 2020). According to Smedema et al. (2000), the proportion of agricultural land drained in developing countries is 5–10 % as compared to 25–30 % in developed countries. This is primarily due to a fundamental lack of data on key soil attributes needed as inputs to design drainage systems (Ayars and Evans, 2015). This paucity of data is often driven by the cost of sample collection and laboratory analysis and/or a lack of adequate laboratory resources. Historically, in many countries, soil data have been routinely gathered through soil surveys, yet such soil data rarely includes saturated hydraulic conductivity (K_{sat}), an essential parameter in drainage system design (Patil and Singh, 2016).

Accurate quantification of K_{sat}, is essential for multiple aspects of agrohydrology (Zhang and Schaap, 2019). Saturated hydraulic conductivity, is used to quantify water movement through a unit cross-section of soil per unit time (Aimrun et al., 2004) and is used to design subsurface drainage systems, irrigation practices and for modelling agricultural and hydrological processes such as water movement and solute transport in the soil (Abdelbaki et al., 2009). Saturated hydraulic conductivity is also used in subsurface drainage models to determine drain spacing at specific drain depth (Bicknell et al., 2005; Chung et al., 1992; Ma et al., 2012; Skaggs, 1978; Van Dam et al., 1997). Thus, the accurate determination of K_{sat} is important as it influences the technical and economic feasibility of large-scale agricultural drainage projects.

Saturated hydraulic conductivity can be estimated using either in-situ methods such as the Guelph permeameter (above water-table method) (Reynolds and Elrick, 1985) and the auger hole method (below water-table method) (Van Beers, 1958), or laboratory-based methods such as the constant head and falling head permeameters (Klute and Dirksen, 1986). However, K_{sat} is highly spatially variable over a range of scales e.g., plot, field and catchment (Hassler et al., 2014). Therefore, a large number of soil samples and measurements are required to accurately describe the hydraulic properties of an area of interest (Sobieraj et al., 2004). These measurements often are difficult, time-consuming and expensive (Fooladmand, 2011; Patil and Singh, 2016). In addition, due to time and financial constraints, researchers, planners and decision-makers in many countries especially in least developed countries are faced with inadequate data to work with (K_{sat} values are simply unavailable) (Patil et al., 2011, 2010). Consequently, multiple researchers have focused on developing indirect methods for estimating K_{sat} such as utilising Pedotransfer functions (PTFs) to estimate K_{sat} from more readily available soil measurement information (e.g., soil texture and bulk density) (Arrington et al., 2013; Dashtaki et al., 2010; Khodaverdiloo et al., 2011).

PTFs were first represented by empirical regression equations relating water and solute transport parameters to the fundamental soil properties that are accessible in soil survey databases (Cosby et al., 1984; Dane and Puckett, 1994; Julia et al., 2004; Puckett et al., 1985). For example, Ahuja et al. (1986) developed an empirical equation to estimate K_{sat} from the effective porosity only, while Li et al. (2007) developed a PTF to estimate K_{sat} that required percentage (%) sand, silt and clay, bulk density (BD) and soil organic matter. However, Wösten et al. (2001) stated that future developments in PTFs would be derived from more accurate data mining tools such as Artificial Neural Networks (ANNs) and group methods of data mining techniques such as K-Nearest Neighbour, and support vector machine (SVM).

K-Nearest Neighbour have been applied to predict precipitation (Huang et al., 2017), groundwater depth (Kombo et al., 2020) and crop yield (Jothi et al., 2020). Whereas SVM has also been applied to predict groundwater depth (Mallikarjuna et al., 2020) and the discharge coefficient of rectangular side weirs located on trapezoidal channels (Azimi et al., 2019). ANNs have been successfully applied to predict water-table depth and groundwater salinity at different drain depths and spacing (Nozari and Azadi, 2017), estimate soil salinity (Bouksila et al., 2010), and to predict evapotranspiration from limited meteorological data (Zanetti et al., 2007). ANNs have also been applied to predict sediment transport without sedimentation by combining ANNs with the Particle Swarm Optimization (PSO), Imperialist Competitive Algorithm (ICA), Genetic Algorithm (GA) and Decision Tree (DT) methods (Ebtehaj et al., 2020, 2018).

Artificial Neural Networks have been shown to be more flexible and able to deal with multiple sources of input data as compared to PTF-based regression equations (Wösten et al., 2001). In the last two decades, ANNs have been utilised as a special class of PTFs using feed-forward or radial basis functions to approximate any continuous (nonlinear) function (Sobieraj et al., 2004). Detailed statistical and functional reports have been carried out to compare the performance between ANN-based PTFs and regression-type PTFs to predict K_{sat} on different datasets such as soil texture and BD (Agyare et al., 2007; Arshad et al., 2013; Parasuraman et al., 2006). Several authors have demonstrated that ANNs outperform regression techniques, especially when uncertainties in the quality of the data are small (Baker and Ellison, 2008; Merdun et al., 2006; Minasny and Mcbratney, 2002). Therefore, ANNs have the potential to be a more accurate approach to estimating K_{sat} as compared to more conventional PTF regression equations.

This study aims to improve the prediction of saturated hydraulic conductivity (K_{sat}) in order to enhance the efficacy of drainage system design in data-poor areas utilising existing and currently under-utilised datasets. It is envisaged that this will reduce the overall cost of drainage system design making them more accessible to farmers, planners, and decision-makers in least developed countries. The objectives of this study were to: (1) develop ANNs to estimate $K_{sat,}$ (2) compare the performance of applying ANNs as compare to regression-type PTF equations for the estimation of K_{sat}, and (3) evaluate the performance of drainage systems based on these two contrasting methods of estimating K_{sat}.

2. Material and methods

2.1. Case study

Data sets were generated for two established agricultural projects in Libya, Hammam Agricultural Project (HAP) and Eshkeda Agricultural Project (EAP), using seven years (1974–1981) of archived and under-utilised data from the General Authority of Water Resources in Libya. Both study areas are located in the southern part of Libya, north of the Sahara Desert. Both study areas are considered to be in the hyper-arid zone (FAO, 1989) with annual rainfall rarely exceeding 100 mm. There are no perennial rivers in the area of interest. Consequently, the only water source is paleo-groundwater from two aquifers. The Mezda and Murzuq aquifers are the source of irrigation for HAP and EAP respectively (Tantawi, 2000).

The HAP, is located in the north-east part of Sokna Oasis (29°04′01″N 15°47′05″E) at an elevation of *circa* 270 m above sea level (asl). The HAP covers 1200 ha and is divided into 182 farms (6−8 ha for each farm). The work on the project began in 1971, and the reclamation was completed in 1985. The HAP project has a drainage system designed in the 1970s. The soil texture in the HAP is highly variable with eight soil texture classes (Table 1). According to the USDA textural classification the HAP dataset can be classified as, sand (22.3 %), loamy sand (18.3 %), sandy loam (31.4 %), loam (0.6 %), sandy clay loam (21.0 %), clay loam (4.2 %), sandy clay (1.8 %) and clay (0.4 %) (Soil Survey Division Staff, 1993) (see Table 1 for more details). The other project, EAP, is located north of Sabha city (27°32′45″N 14°16′7″E) at an elevation of 320 m asl. The project was initiated in 1975 and it became operational in the second half of 1976, with a drainage system being designed between 1979−1981. The EAP covers 3000 ha and is divided into three subareas: Western Area (2000 ha), Central Area (600 ha) and Eastern Area (370 ha). In accordance with the installed irrigation wells, the area comprises 25 districts each with 12 farms of *circa* 10 ha each. The farms are sub-divided into two farmed plots, each 180 m by 234 m in size. Windbreaks have been planted between the farm plots and around the individual farms. The EAP dataset is associated with four soil texture classes dominated by sand (88.4 %) with 8.0 %, 3.4 % and 0.2 % of the set represented by loamy sand, sandy loam and loam, respectively (Table 1).

2.2. Dataset description

The data collected from the two project sites included particle size distribution (%sand (S), %silt (Si) and %clay (C)), bulk density (BD) (g cm^{-3}), %wilting point (WP), %field capacity (FC), and K_{sat} (m d^{-1}), summarised in Table 1. Soil texture was measured using the hydrometer method following Bouyoucos (1927). The soil BD was measured following the methods described by Buckman and Brady (1960). Field capacity and wilting point were determined by tensiometers, using the method reviewed by Richards (1949). In the EAP K_{sat} was measured in the field using the average of four replications of the auger-hole method (Van Beers, 1958). In the HAP K_{sat} was measured in the laboratory using undisturbed samples following the constant head method (Klute and Dirksen, 1986). In total, for the HAP, there were 770 K_{sat} measurements (c. 1 measurement for every 2 ha) while for the EAP there were 442 K_{sat} measurements (c. 1 measurement for every 5 ha). To facilitate ANN development and validation, in both EAP and HAP an area was randomly selected and designated as a 'data-poor' area for which K_{sat} measurements were assumed to be absent. In the EAP a 600 ha area in the central section of the scheme (5.0 km x 1.2 km) was selected as the 'data-poor' area whilst in the HAP a 500 ha area in the western section of the scheme (1.43 km x 3.5 km) was selected. The total number of K_{sat} measurement points in these designated 'data-poor' areas were 82 and 158, in EAP and HAP respectively, which represent 20 % of the dataset for each site. The remaining areas of EAP and HAP were assumed to be 'data-rich' areas. The data from 'data-rich' areas was used to develop the ANNs to predict the K_{sat} in the designated data-poor areas. Data from the data-rich areas of each scheme were assigned into two groups based on the input requirements for the development of ANNs. The first group had limited parameters that included, S, Si, C, and BD as independent variables and K_{sat} as the dependent variable. The second group included the parameters in the first group as well as WP and FC. A statistical summary of the dataset used in this study is given in Table 1.

Table 1
Statistical summary of the dataset used in this research.

	Sand (%)		Silt (%)		Clay (%)		BD (g cm^{-3})		FC (%)		WP (%)		K_{sat} (m d^{-1})	
Project	EAP	HAP	EAP	HAP	EAP	HAP	EAP	HAP	EAP	HAP	EAP	HAP	EAP	HAP
Min	66.0	25.0	1.0	1.0	1.0	1.0	1.1	1.6	14.0	2.2	1.5.0	0.7	0.1	0.1
Max	98.0	98.0	36.0	46.0	10.0	41.0	1.9	2.0	23.0	44.0	14.0	14.0	24.0	38.0
Average	92.6	73.0	4.7	13.5	2.8	14.0	1.5	1.9	18.5	18.0	8.3	5.6	5.3	7.2
STDEV	4.0	15.5	3.0	7.8	1.8	10.6	0.2	0.1	1.6	7.0	2.3	2.2	4.5	8.2
Median	94.0	72.0	4.0	14.0	2.0	11.0	1.5	1.9	18.5	17.5	8.5	5.5	3.7	4.2
	Percentage of samples associated with textural class													
Project	S		LS	SL	L		SCL		CL		SC		C	
EAP	88.4 %		8.0 %	3.4 %	0.2 %		0.0 %		0.0 %		0.0 %		0.0 %	
HAP	22.3 %		18.3 %	31.4 %	0.6 %		21.0 %		4.2 %		1.8 %		0.4 %	

Where: S is sand, LS is loamy sand, SL is sandy loam, L is loam, SCL is sandy clay loam, CL is clay loam, SC is sandy clay, and C is clay (Soil Survey Division Staff, 1993).

2.3. ANNs model development

Artificial neural networks are a mathematical approach used to characterise synthetic neurons to solve complex problems in a similar way to that of the human brain (Agatonovic-Kustrin and Beresford, 2000). Since the 1940s, there has been growing interest in the study of neurological mechanisms and structures. That interest has led to the development of new computational models, linking systems or ANNs, based on biological processes that have been used to solve complex problems such as pattern recognition and rapid information processing and adaptation (Huang, 2009). ANNs use machine learning based on the principle of self-adjusting internal control parameters and can be used to recognise complex non-linear relationships between input and output datasets (Ayoubi et al., 2011).

ANNs are typically formed of three processing layers, namely an 'input', 'hidden' and 'output' layer (Fig. 1) (Chollet and Allaire, 2017; Ghatak, 2019; More, 2018). The input layer consists of the selected independent variables. Each variable in the input layer is connected to the 'hidden' layer through a 'weighting' algorithm (Bell, 2015; Chollet and Allaire, 2017; Ghatak, 2019). In some cases, the weights given to each independent variable are generated randomly, in other cases, they are fine-tuned and trained through a process called backpropagation (Bell, 2015; Chollet and Allaire, 2017; Ghatak, 2019). Backpropagation quantifies the error at the output by comparing the currently derived outputs to the expected ones and modify the weights to reduce the error (Bell, 2015; Chollet and Allaire, 2017; Ghatak, 2019). The layer located between the input and the output layers is called the 'hidden' layer. The hidden layer may consist of more than one layer with each layer containing a number of units (a grouping of independent variables). These units, known also as 'neurons', are used to guide the processing of the data contained in the input layer and to generate computational connections between the input and the output layer. There is no specific formula to decide the number of units to be included in the hidden layer. It is generally problem-dependent, with the error tested by varying the number of units and layers in order to determine the appropriate number of units/layers to be used (Bell, 2015). The hidden layer is represented by a group of activation functions including linear functions, logistic functions and hyperbolic functions. Logistic and hyperbolic functions are mainly used in classification problems, because they range between 0 and 1 or between -1 and 1, while the linear functions (infinity to infinity) are mainly used in regression problems (Ghatak, 2019). The output layer consists of a series of inverse functions, which are used to refine the training and calculate the difference between the output predictions and the measured $K_{sat,}$ and to minimise error. One of the most widely used algorithms for this purpose is the backpropagation algorithm (Bell, 2015). Backpropagation was applied to develop the ANNs in this study.

For the development of ANNs to predict K_{sat} in the designated data-poor areas within EAP and HAP, several processing steps were undertaken. The first step was to define a training dataset in order to develop the ANNs. The training dataset was based on 80 % of the data from the data-rich areas of the EAP and HAP (leaving 20 % of the data as the pseudo 'data-poor area' for testing purposes against the actual measured K_{sat} data from both sites). The second step was to define the appropriate network (number of layers and number of neurons in each layer). A sequential orthogonal approach (Sun, 2012) was adopted to determine the optimum number of neurons. This approach adds a neuron to the 'hidden layer' until the ANN with the smallest error is identified (Sun, 2012). The ANN with the smallest error in predicting K_{sat} within the data-rich area was subsequently applied to the data-poor area (using data from EAP and HAP that had not been used in the development of the model) to validate and test the performance of ANN by comparing the ANN predicted K_{sat} values with measured K_{sat} values. Table 2 shows the input parameters used to develop six ANN models to predict K_{sat}. All ANNs were developed for both HAP and EAP using two groups of input parameters. The first group of input parameters included %sand, %silt, % clay, bulk density, wilting point and field capacity were applied in ANN1, ANN3, and ANN5. ANN1 and ANN3 used data only from EAP and HAP respectively, and ANN5 used the combined data set from both EAP and HAP. The second group of input parameters included %sand, %silt, %clay and bulk density were applied in ANN2, ANN4, and ANN6. ANN2 and ANN4 used data only from EAP and HAP

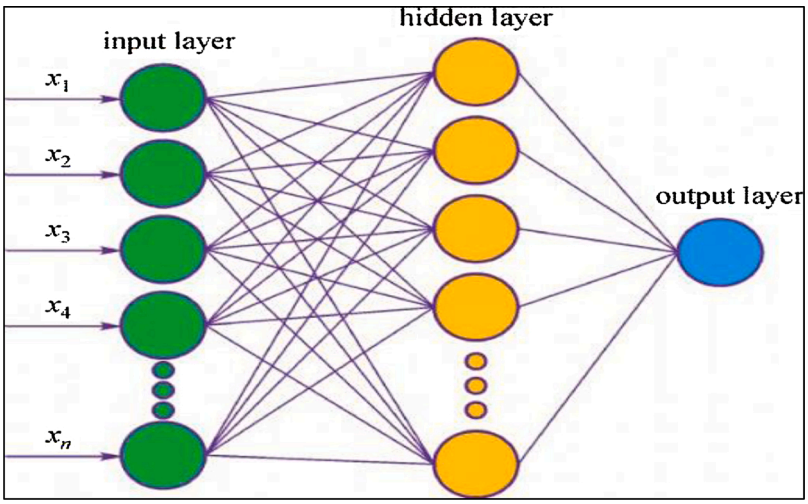

Fig. 1. Structure of an artificial neural network (based on Khademi and Jamal, 2016).

Table 2
The ANN models developed for the EAP and HAP study sites.

ANN Model	Inputs required	Project name
ANN1	%Sand, %Silt, %Clay, BD (g cm^{-3}), %FC, %WP, and K_{sat} (m d^{-1})	EAP
ANN2	%Sand, %Silt, %Clay, BD (g cm^{-3}), and K_{sat} (m d^{-1})	EAP
ANN3	%Sand, %Silt, %Clay, BD (g cm^{-3}), %FC, %WP, and K_{sat} (m d^{-1})	HAP
ANN4	%Sand, %Silt, %Clay, BD (g cm^{-3}), and K_{sat} (m d^{-1})	HAP
ANN5	%Sand, %Silt, %Clay, BD (g cm^{-3}), %FC, %WP, and K_{sat} (m d^{-1})	EAP + HAP*
ANN6	%Sand, %Silt, %Clay, BD (g cm^{-3}), and K_{sat} (m d^{-1})	EAP + HAP*

(ANN) Artificial Neural Network, (EAP) Eshkeda Agricultural Project, and (HAP) Hammam Agricultural Project. (WP) wilting point, (FC) field capacity, (BD) bulk density, (K_{sat}) saturated hydraulic conductivity, and *EAP + HAP = combined data set.

respectively, and ANN6 used the combined data set from both EAP and HAP. The reason for having two groups of input parameters is to assess the minimum data set required to predict K_{sat} accurately.

The critical evaluation of the ability of the six ANN models developed in this study to accurately predict K_{sat} was undertaken in five steps. Step 1 used the training data (data-rich area) for site-specific ANN model training and the data from the designated data-poor areas as testing datasets. The best performing ANN model in terms of predicting K_{sat} in the designated data-poor areas was selected by applying three statistical parameters. These were: (1) coefficient of determination (R^2) between the observed and predicted K_{sat} values for the data poor areas (e.g., best performance is equal to 1), (2) mean square error (MSE) and (3) root mean square error (RMSE) (Patil and Singh, 2016; Zhang and Schaap, 2019), which can respectively be defined mathematically as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}\left(M_i - \overline{M}_i\right)} \tag{1}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_i - M_i)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - M_i)^2}{n}} \tag{3}$$

Where n represents the number of soil samples, M_i is the measured K_{sat}, P_i is the predicted K_{sat}, and $\overline{M}_i$ is the mean measured K_{sat}.

MSE and RMSE were used to give a quantitative indication of ANN model error in units of K_{sat} (the best performance value is zero). Step 2 tested the four site specific ANN models (Table 2) on data-rich areas only using a K-fold cross-validation technique (Bell, 2015; Ghatak, 2019). There are other cross-validation techniques such as, (1) validation set approach (where the data is divided randomly into two sets, one set to train the model and the other set to test the model), (2) Leave one out cross-validation (where one sub-set of the data point is left out, the model is built on the remaining data, and the model is tested using the 'left out' data sub-set) (Bell, 2015; Chollet and Allaire, 2017; Ghatak, 2019). In this study, the K-fold cross-validation method was applied. In the K-fold cross-validation technique, the training data (data from the pre-defined data-rich areas in both sites) were divided into ten randomly generated sub-sets. For EAP and HAP this resulted in ten randomly generated sub-sets of 36 and 61 data points respectively. During each K-fold cross-validation, nine sub-sets were combined and used as the 'training sub-set' and one part used as the 'testing sub-set'. This process was repeated ten times "rotating out" a testing sub-set each time so that each sub-set used in each ANN was applied in the testing. For both EAP and HAP, this K-fold cross-validation process was also repeated, removing 10 data points from the dataset each time until only 10 data points remained.

This approach was used to quantify the performance (ability to accurately predict K_{sat}) of each of the four site specific ANN models (Table 2) against the size of the training set used (Chollet and Allaire, 2017). This technique was used to quantify the ability of the ANNs to predict K_{sat} as the size of training set was reduced randomly and to also determine the 'minimum' training set required to develop the ANN. Step 3 of performance testing compared the outputs of the ANNs against outputs generated from traditional, widely adopted, empirical equations used to predict K_{sat} (Abdelbaki et al., 2009; Minasny and McBratney, 2000a) (Table 3). Step 4 evaluated

Table 3
Pedotransfer functions (PTF) widely adopted to predict saturated hydraulic conductivity (K_{sat}).

Formula (cm hr^{-1})	PTF Code	References
$K_{sat} = 15.696\ EXP[-0.1975C]$	PTF-1	(Puckett et al., 1985)
$K_{sat} = 30.384\ EXP[-0.144C]$	PTF-2	(Dane and Puckett, 1994)
$K_{sat} = 0.0920e^{0.0491*S}$	PTF-3	(Julia et al., 2004)
$K_{sat} = 2.54 * 10^{(-0.6+0.012*S\ -\ 0.0064*C)}$	PTF-4	(Cosby et al., 1984)

Where: C is the % clay, S is the % sand.

the performance of each developed ANNs to predict K_{sat} in the other site (non site-specific). For example, ANN1 and ANN2 were developed and tested in EAP and were applied to predict K_{sat} in HAP and ANN3 and ANN4 were developed and tested in HAP and were applied to predict K_{sat} in EAP. This step was added to evaluate the applicability of applying ANNs, developed in specific locations, to predict K_{sat} in new locations in Libya. In this example, the separation distance between ANN development and application was 225 km (i.e., the separation distance between EAP and HAP). Step 5 developed ANNs by combining the training datasets from the pre-defined data-rich areas of both EAP and HAP to predict K_{sat} in the pre-defined data-poor areas in EAP and HAP. This step was applied to evaluate the accuracy of ANNs based on larger combined datasets from multiple sites, compared to smaller but site-specific data sets, to predict K_{sat}.

2.4. Application of ANNs to drainage design

Drainage systems can be designed as either surface, subsurface or a combination of both surface and subsurface, depending on the prevailing conditions in the area and crops grown (Luthin, 1978). This study focused on subsurface drainage which is used in both EAP and HAP. The determination of appropriate spacing and depth of drains are the main parameters in the design of subsurface drainage systems that need to be considered (Luthin, 1978). A number of drainage theories and associated equations for designing subsurface drainage systems have been developed (Dagan, 1965; Dumm, 1960, 1954; Hooghoudt, 1940, 1937; Kirkham, 1958; Luthin and Worstell, 1959; van Schilfgaarde, 1963). These equations were developed to relate the spacing and depth of drains to water-table depth and drawdown rates. Methods have been developed to design drainage systems under steady or non-steady state (see reviews by van der Ploeg et al. (1999) and Youngs (1999)). EnDrain (Drainage computer program released by Prof. Oosterbaan in the Netherlands) was applied to design drainage systems in the designated data-poor areas of both EAP and HAP using K_{sat} derived from the ANNs developed in this study (Table 2) as compared to widely adopted PTFs (Table 3) and the actual measured K_{sat} for these data-poor areas. EnDrain uses Eq. (4), described by Hooghoudt (1937; cited in Ritzema (1994)), to calculate the distance between drains and the depth of the water-table at the midpoint between two drains (Oosterbaan, 1993):

$$L^2 = \frac{8 \cdot Kb \cdot d \cdot h \; + \; 4 \cdot Ka \cdot h^2}{q} \tag{4}$$

Where L is drain spacing (m), q is the steady-state drainage discharge rate (m d^{-1}), Ka is the hydraulic conductivity of the soil above drain level (m d^{-1}), Kb is the hydraulic conductivity of the soil below drain level (m d^{-1}), h is the height of water-table above the water level in the drain (m), d is the equivalent depth which is a function of the spacing (L), the depth (in m) of impervious layer below the drain (Di-Dd in Fig. 2), and the radius (r) of the drain (m). Values for d can be found in tables presented by Hooghoudt (1940).Van der Molen and Wesseling (1991) developed this further to calculate series solutions for multiple equivalent depths.

EnDrain requires two values of K_{sat} one from above the drain level and one from below the drain level. Therefore, a representative

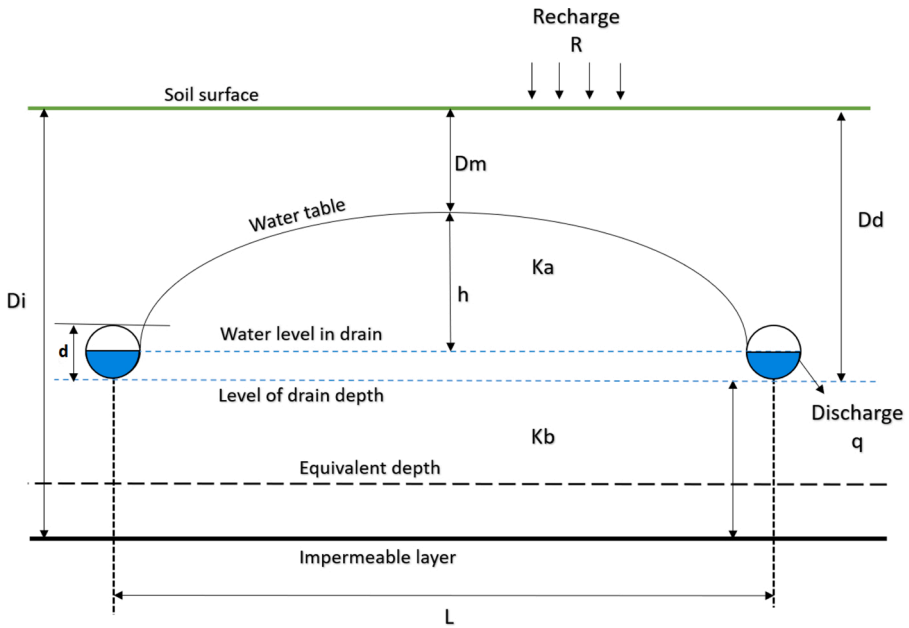

Fig. 2. Schematic of subsurface drainage system indicating input parameters for EnDrain and Eq. 4. (Di) depth from the surface to impermeable layer, (Dd) depth of drain from surface, (Dm) water-table depth from surface, (h) height of water-table above water level in drain, (Ka) saturated hydraulic conductivity above the drain, (Kb) saturated hydraulic conductivity below the drain, (d) drain diameter and (L) is the spacing between drains.

K_{sat} was required as an input to EnDrain. According to Luthin (1978), the geometric mean (K_g) gives the best representative K_{sat} for a study area. Therefore, representative K_g values for EAP and HAP were generated for input to EnDrain. Eq. (5) illustrates how the geometric mean was calculated and Table 4 shows the different calculations of K_g based on the project area and the source of K_{sat} (e.g., K_{sat} derived from ANNs):

$$K_g = \sqrt[N]{K_{sat1} \cdot K_{sat2} \cdot K_{sat3} \cdot K_{satN}} \quad (5)$$

Where the geometric mean (K_g) is found by multiplying the values of K_{sa} (K_{sat1} is the K_{sat} value for sample 1, K_{sat2} is K_{sat} value of sample 2 and so on to N number of values) and then the N^{th} root of the product of those values is then found (N^{th} root is equal to the number of values).

Water-table depth based on ANN and PTF generated K_g values (Table 4) was compared with water-table depth based on K_g values generated from measured K_{sat} (K_g1 and K_g 8 in Table 4). The value of R used in EnDrain was the average water requirement for date palm (*Phoenix dactylifer*) and alfalfa (*Medicago sativa*). Date palm was chosen because it is considered to be the most profitable crop for the south of Libya (Elmeer et al., 2016). Alfalfa was adopted as it is grown as a fodder for livestock in both EAP and HAP. The average water requirements for date palm is 0.02 m d^{-1} in EAP and 0.007 m d^{-1} in HAP, and for alfalfa are 0.01 m d^{-1} for EAP and 0.006 m d^{-1} for HAP. These values were taken from an unpublished report based on work undertaken by the Libyan Government in 1999 (General Water Authority, 1999). Site investigations of soil properties were undertaken at both sites between 1974–1980 in HAP by Danenco (1980) and Holzmann-Wakuti (1974) and between 1976–1981 in EAP by ITALCONSULT (1976) and Cornelius-Brochier (1981), to design agricultural projects including drainage systems in these areas. The following data were taken from these reports: The average depth of the impermeable layer for both EAP and HAP was assumed to be 1.5 m, two soil layers were assumed with two different K_{sat} values with depth, and the depth of water-table was set at 0.5 m. The vertical K_{sat} and the horizontal K_{sat} were assumed to be equal (Hess et al., 1992). Table 5 shows the input parameters applied in EnDrain. Fig. 2 is a schematic representation of a subsurface drainage system and provides an explanation of the input parameters used in Table 5 and Eq. 4.

Finally, an economic analysis was conducted to calculate the initial cost of installing each drainage system in the designated data-poor areas using the measured K_g as well as ANNs and PTFs generated K_g values. The cost of drains installed was assumed to be $2.62/m according to Skaggs (2007).

2.5. Software used

The data was analysed using several software packages. The development of the ANNs utilised a neuralnet package in R (Stefan and Guenther, 2019). Pedotransfer function equations (Table 3) used to predict K_{sat} and the representative K_{sat} (K_g) were calculated in Microsoft EXCEL (2016). Finally, subsurface drain spacing, and depth of water-table were calculated using EnDrain (Oosterbaan, 1993).

3. Results

3.1. Performance of site-specific ANNs and cross-validation

The structure of each of the site-specific ANN models developed in this study, and their Step 1 performance evaluation are presented in Table 6. All ANNs were able to predict, based on R^2, between 0.89−0.96 of the variability in K_{sat} in the designated data-poor areas of EAP and HAP. As expected, the ANN models performed better when using the training data than when the testing data was used. In HAP, the R^2 was slightly better in the testing section ($R^2 = 0.96$ and 0.95) than in the training section ($R^2 = 0.94$ and 0.93;

Table 4
K_{sat} geometric means (K_g) derived from the ANN and PTF models evaluated in this study and used as input parameters to design a sub-surface drainage system using EnDrain.

Project name	K_g Code	Source of K_{sat} data	K_g Above drain level (A_{dl})	K_g Below drain level (B_{dl})
	K_g1	**Measured K_{sat}**	K_g-A_{dl}-1	K_g-B_{dl}-1
	K_g2	K_{sat} from ANN1	K_g-A_{dl}-2	K_g-B_{dl}-2
	K_g3	K_{sat} from ANN2	K_g-A_{dl}-3	K_g-B_{dl}-3
EAP	K_g4	K_{sat} from PTF-1	K_g-A_{dl}-4	K_g-B_{dl}-4
	K_g5	K_{sat} from PTF-2	K_g-A_{dl}-5	K_g-B_{dl}-5
	K_g6	K_{sat} from PTF-3	K_g-A_{dl}-6	K_g-B_{dl}-6
	K_g7	K_{sat} from PTF-4	K_g-A_{dl}-7	K_g-B_{dl}-7
	K_g8	**Measured K_{sat}**	K_g-A_{dl}-8	K_g-B_{dl}-8
	K_g9	K_{sat} from ANN3	K_g-A_{dl}-9	K_g-B_{dl}-9
	K_g10	K_{sat} from ANN4	K_g-A_{dl}-10	K_g-B_{dl}-10
HAP	K_g11	K_{sat} from PTF-1	K_g-A_{dl}-11	K_g-B_{dl}-11
	K_g12	K_{sat} from PTF-2	K_g-A_{dl}-12	K_g-B_{dl}-12
	K_g13	K_{sat} from PTF-3	K_g-A_{dl}-13	K_g-B_{dl}-13
	K_g14	K_{sat} from PTF-4	K_g-A_{dl}-14	K_g-B_{dl}-14

(K_{sat}) saturated hydraulic conductivity, (K_g) geometric mean of Ksat, (PTF) pedotransfer function, (ANN) Artificial Neural Network, (EAP) Eshkeda Agricultural Project, and (HAP) Hammam Agricultural Project.

Table 5
Input parameters for EnDrain required to design a subsurface drainage system for EAP and HAP.

Input parameter	Units	EAP		HAP	
		Date palm	Alfalfa	Date palm	Alfalfa
Time average recharge (R)	m d^{-1}	0.02	0.01	0.007	0.006
Depth to the impermeable layer from surface (Di)	m	1.5	1.5	1.5	1.5
Depth water level in drain from surface (Dm + h)	m	0.9	0.9	0.9	0.9
Depth of drain bottom from surface (Dd)	m	1.0	1.0	1.0	1.0
Diameter of drain (d)	m	0.2	0.2	0.2	0.2
Depth water-table midway between drains (Dm)	m	0.5	0.5	0.5	0.5

Source of data: (Cornelius-Brochier J.V, 1981; Danenco, 1980; General Water Authority, 1999; Holzmann-Wakuti, 1974; Italconsult, 1976). (EAP) Eshkeda Agricultural Project, (HAP) Hammam Agricultural Project.

Table 6
Structure performance evaluation of the ANNs implemented in the designated data-poor areas of EAP and HAP.

Model name	Project name	Model Structure*	Training				Testing			
			Number of samples	RMSE m d^{-1}	MSE m d^{-1}	R^2	Number of samples	RMSE m d^{-1}	MSE m d^{-1}	R^2
ANN1	EAP	6:1:1	360	1.40	1.98	0.95	82	1.92	3.69	0.90
ANN2		4:2:1	360	1.53	2.35	0.94	82	2.00	4.03	0.89
ANN3	HAP	6:4:4:1**	610	2.29	5.26	0.94	158	2.78	7.76	0.96
ANN4		4:5:1	610	2.52	6.36	0.93	158	2.82	7.97	0.95

* Model structure is the structure of the ANNs, for example 6:4:1, 6 is the number neurons representing the input parameters, 4 is the hidden layer neurons, and 1 is the output of the model.

** 6:4:4:1 means that model has two hidden layers with 4 neurons in each layer. (PTF) Pedotransfer function, (ANN) Artificial Neural Network, (EAP) Eshkeda Agricultural Project, (HAP) Hammam Agricultural Project, (R2) coefficient of determination, (RMSE) Root Mean Square Error, and (MSE) Mean Square Error.

Table 6).

Generally, the ANNs developed in this study showed that there was little difference in performance between the ANNs as R^2 ranged from 0.89−0.90 in EAP and 0.95–96 in HAP. According to the performance results in Table 6, models ANN1 and ANN3 performed better than both ANN2 and ANN4. Even though, the number of data points in the ANNs developed for EAP (for training and testing) were almost half the number used in the ANNs developed for HAP (360 in EAP and 610 in HAP), the RMSE and MSE values for the EAP ANNs indicate that they are more accurate.

The results of the Step 2 K-fold cross-validation are shown in Fig. 3. Fig. 3a shows the median RMSE across the ten randomly generated sub-sets (dark line inside the boxplot) when the size of training set is fixed to the maximum size (360 in EAP and 610 in HAP). In Fig. 3a it can be noted that the median RMSE for ANN1, ANN2, ANN3, and ANN4, were 1.65, 0.40, 1.80, and 1.40 m d^{-1}, respectively. Also, the minimum and maximum RMSE ranged between 1.37–1.84, 0.10–1.40, 1.30–8.10 and 0.30–2.70 m d^{-1} in ANN1, ANN2 ANN3 and ANN4, respectively. These results showed that the developed ANNs for both EAP and HAP were in the range of the 10-fold cross-validation. Therefore, the training dataset (from the data-rich areas) was effectively used to develop ANNs able to predict K_{sat}. Fig. 3b indicates how RMSE varies with the size of training dataset. As indicated by the low median RMSE values, the best performing ANNs were developed when the full dataset was used in the training phase (Fig. 3b). However, for both EAP and HAP, even if the training set was reduced to 50, the ANNs were still able to accurately predict K_{sat} (Fig. 3b).

3.2. Performance of PTFs and non-site-specific ANNs

Four PTFs that are widely adopted to predict K_{sat} were applied to predict K_{sat} in the designated data-poor areas of EAP and HAP (Table 3). In relation to RMSE and MSE, PTF-2 (developed by Dane and Puckett, 1994) performed better than the other PTFs in both EAP and HAP (Table 7). However, it is critical to note that all the PTFs were between 2–5 times less accurate at predicting K_{sat} than the site-specific ANNs. For example, the RMSE values of PTF-2 were 4.19 and 12.72 m d^{-1} for EAP and HAP, respectively. In contrast, the RMSE for all the site-specific ANNs applied to EAP and HAP ranged from 1.92 to 2.82 m d^{-1}. Further, the MSE for the PTFs applied in EAP and HAP ranged from 17.60–225.00 m d^{-1}. In contrast, the MSE for the site-specific ANNs applied in EAP and HAP ranged from 3.69 to 7.97 m d^{-1}. For the combined ANNs (ANN5 and ANN6), the accuracy was better than the PTFs but slightly less (3 times less accurate) than the site-specific ANNs. For example, ANN5 in RMSE was 2.11 m d^{-1} while in ANN1 was 1.92 m d^{-1}, while ANN6 was 6.09 m d^{-1} in predicting K_{sat} in EAP. Finally, for the non-site-specific ANNs, the accuracy was less than all other ANNs and PTFs with the highest RMSE and MSE values of 15.88 and 252.23 m d^{-1}, respectively, in ANN4 for predicted K_{sat} in EAP.

Figs. 4 and 5 show the accuracy of the PTFs specific and non-site-specific ANNs in predicting K_{sat} within the data-poor areas of EAP and HAP. The 1:1 line represents the line of perfect agreement between the predicted and the measured values of K_{sat}. It is of note that the predicted K_{sat} of the site-specific ANNs all closely comply with the 1:1 line. The results indicate that the site-specific ANNs were significantly better at predicting the observed K_{sat} values compared to those predicted by the PTFs and non-site-specific ANNs applied

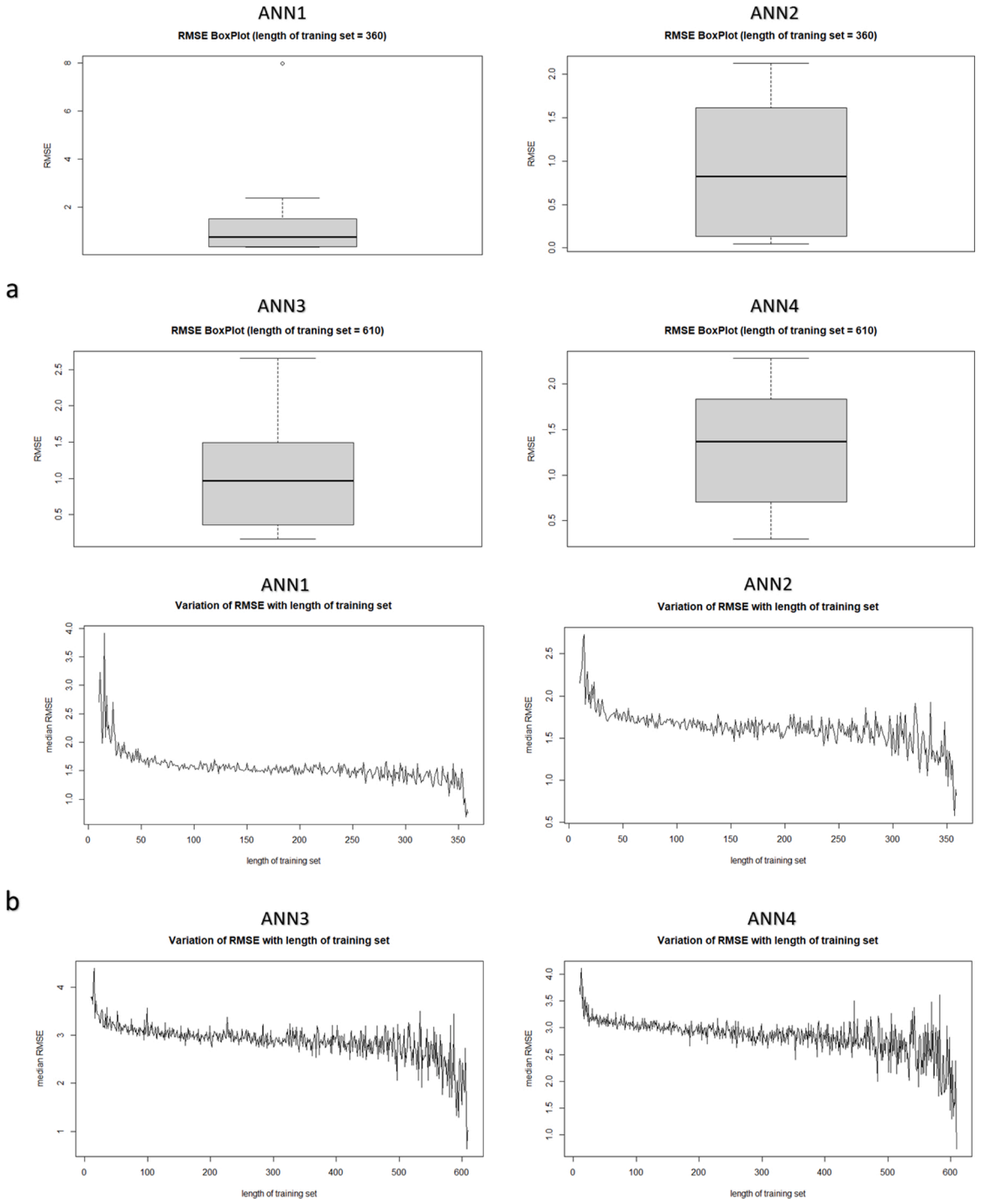

Fig. 3. (a) Boxplots of the median Root Mean Square Error (RMSE) for the Artificial Neural Networks (ANN1-ANN4) models developed for Eshkeda Agricultural Project (EAP) and Hammam Agricultural Project (HAP). (b) The median RMSE for the size of training used in the development of the ANNs for EAP and HAP.

to EAP and HAP. For example, for the highest EAP observed K_{sat} value in EAP of 21.90 m d^{-1}, PTF-1, PTF-2, PTF-3 and PTF-4 predicted K_{sat} values of 2.10, 4.70, 2.20 and 2.00 m d^{-1}, respectively. This is an underprediction of one order of magnitude. In contrast, for the same sample (21.90 m d^{-1}) ANN1 and ANN2 predicted values of 22.40 and 24.50 m d^{-1} respectively. While ANN3 and ANN4 (developed to predict K_{sat} in HAP) overpredict K_{sat} in EAP with the highest value of prediction 42.20 and 29.80 m d^{-1} respectively. Furthermore, in EAP the measured K_{sat} ranged between 0.60 to 21.90 m d^{-1}, whereas the prediction values of K_{sat} for all PTFs ranged

Table 7
The comparative performance of PTFs and ANNs in predicting K_{sat} in the designated data-poor areas of EAP and HAP.

Model name	EAP (82 data points)			HAP (158 data points)		
	RMSE	MSE	R^2	RMSE	MSE	R^2
PTF-1	5.38	29.00	0.00	14.80	218.00	0.73
PTF-2	4.19	17.60	0.00	12.70	162.00	0.74
PTF-3	5.56	31.00	0.01	14.80	220.00	0.90
PTF-4	5.72	32.80	0.00	15.00	225.00	0.88
ANN1	**1.92**	**3.69**	**0.90**	**15.14**	**229.00**	**−0.26**
ANN2	**2.00**	**4.03**	**0.89**	**15.47**	**239.00**	**0.79**
ANN3	**7.26**	**201.87**	**0.07**	**2.78**	**7.76**	**0.96**
ANN4	**15.88**	**252.23**	**0.01**	**2.82**	**7.97**	**0.95**
ANN5	2.11	4.47	0.77	3.39	11.51	0.88
ANN6	6.09	37.11	0.70	4.42	19.61	0.83

(PTF) Pedotransfer function, (ANN) Artificial Neural Network, (EAP) Eshkeda Agricultural Project, (HAP) Hammam Agricultural Project, (R^2) coefficient of determination, (RMSE) Root Mean Square Error, and (MSE) Mean Square Error.

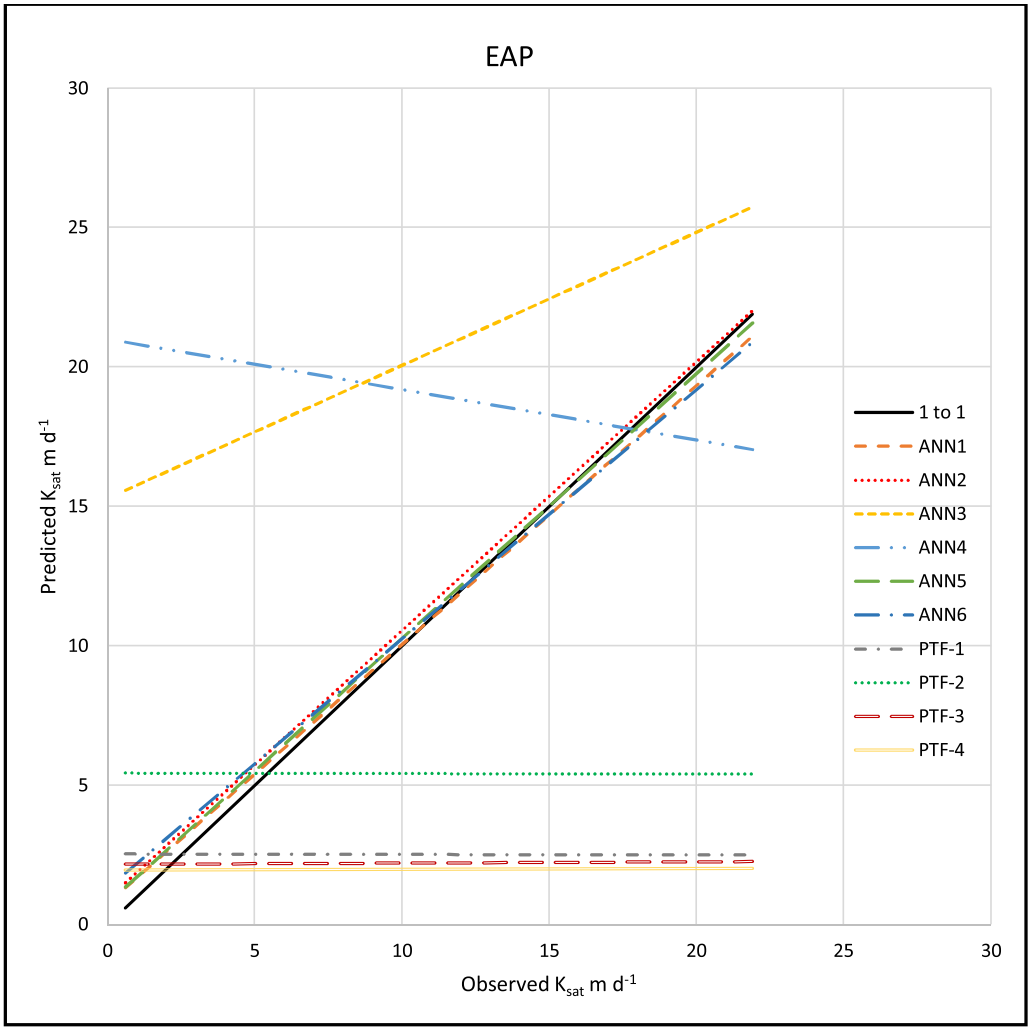

Fig. 4. Predicted Saturated hydraulic conductivity (K_{sat} in m d^{-1}) versus observed K_{sat} (m d^{-1}) of Pedotransfer Functions (PTF-1 to PTF-4) and Artificial Neural Networks (ANN1 to ANN6) in Eshkeda Agricultural Project (EAP).

from 1.20 to 6.30 m d^{-1}. In contrast, the non-site-specific, K_{sat} ranged between 7.00 to 42.20 m d^{-1}. While, in EAP the predicted K_{sat} values for the site-specific ANNs ranged from 0.60 to 25.30 m d^{-1}. A similar pattern was observed in the HAP (Fig. 5), where the measured K_{sat} ranged between 0.40 to 38.00 m d^{-1}, whereas the prediction values for all PTFs ranged between 0.03 to 6.30 m d^{-1}, and

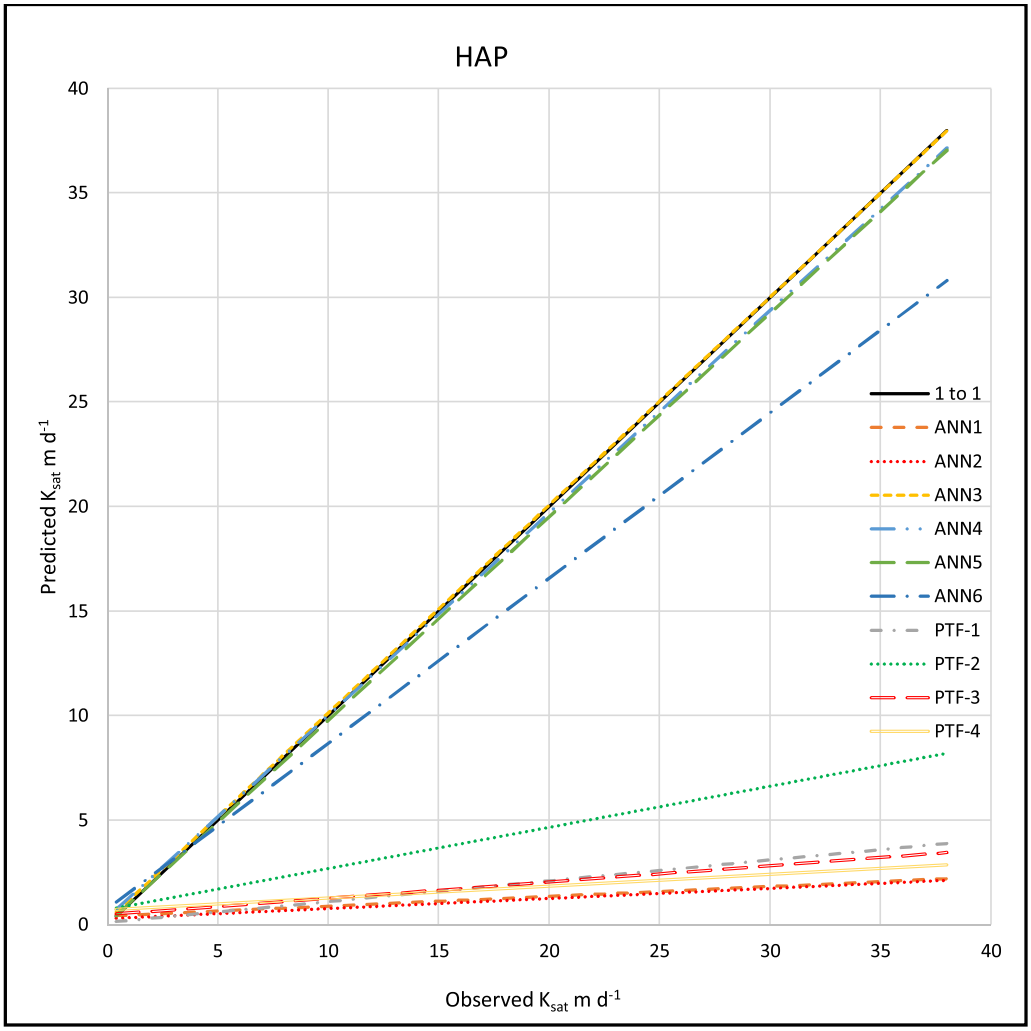

Fig. 5. Predicted Saturated hydraulic conductivity (K_{sat} in m d^{-1}) versus observed K_{sat} (m d^{-1}) of Pedotransfer Functions (PTF-1 to PTF-4) and Artificial Neural Networks (ANN1 to ANN6) in Hammam Agricultural Project (HAP).

for non-site-specific ANNs ranged between 0.01 to 3.50 m d^{-1}. In contrast, the site-specific ANNs Predicted K_{sat} in HAP ranged in value between 0.20 to 27.70 m d^{-1}. However, the site-specific ANNs used to predict K_{sat} in HAP were unable to accurately predict K_{sat} when the measured K_{sat} was $\geq$ 30.00 m d^{-1}. These ANNs underestimate the K_{sat} values $\geq$30.00 m d^{-1} by 10–30 %. However, only 6 samples (out of 158) had a K_{sat} value of $\geq$30.00 m d^{-1}. It can be noted that the PTFs tend to under-predict the K_{sat} values when the observed measurements exceed 6.00 m d^{-1} and 2.00 m d^{-1}, in EAP and HAP respectively. Additionally, it is of note that ANN1 (Fig. 4) and ANN3 (Fig. 5) show slightly better compliance with the 1:1 line in comparison to ANN2 (Fig. 4) and ANN4 (Fig. 5). Also, ANN5 and ANN6 showed a comparable prediction to the ones obtained by the site-specific ANNs where K_{sat} values ranged between 0.70 to 23.20 m d^{-1} in EAP and 0.50 to 30.80 m d^{-1} in HAP. However, ANN6 was slightly less accurate in predicting K_{sat} especially the observed values of $>$15 m d^{-1}.

3.3. Drain spacing and water-table depths

Eq. 5 was used to estimate the K_g for the different scenarios presented in Table 4, the results of which are given in Table 8. Estimation of K_g was made above (A_{dl}) and below (B_{dl}) the drain level. These estimated values of K_g were applied in EnDrain to predict the drain spacing and water-table depths for the designated data-poor areas of EAP and HAP. Measured K_{sat} values used to calculate K_g (EAP K_g1 and HAP K_g8), resulted in a drain spacing between two drains of 24.7 m (date palm) and 34.9 m (alfalfa), and 56.0 m (date palm) and 60.5 m (alfalfa), for EAP and HAP respectively. In general, K_g derived from the four ANNs K_{sat} predictions had the lowest variation in distance between drains compared to the target value based on measured data (K_g1 and K_g8). For EAP, the % variation in drain spacing from the target for ANN1 and ANN2 was for date palm, 1.1 and 4.6 % respectively. In contrast, for the PTFs (K_g 4–7) % variation in drain spacing from the measured data ranged from 28.6 to 36.8% (Table 8). For HAP, the % variation in drain spacing from

Table 8
Performance of PTF and ANN models in terms of variance (%) from target drain spacing (m) and water-table depth (m) when applied to the designated data poor areas of EAP and HAP.

Drainage Project	K_g Code	Data source	Spacing (m)		% variation from the target design		Initial installation costs (USD$ 1000′s)		Variation in water-table depth (cm) with respect of spacing and K_g measured	K_g Above drain level (A_{dl}) (m d^{-1})	K_g Below drain level (B_{dl}) (m d^{-1})
			Date palm	Alfalfa	Date palm	Alfalfa	Date palm	Alfalfa			
	K_g1	**Measured K_{sat}**	**24.7**	**34.9**	**0%**	**0%**	**638.0**	**450.0**	**50.0***	**2.6**	**5.6**
	K_g2	K_{sat} from ANN1	24.4	34.5	1.3 %	1.1 %	645.0	456.0	51.0	3.0	5.3
	K_g3	K_{sat} from ANN2	25.7	36.5	4.0%	4.6 %	610.0	431.0	47.0	3.1	6.0
EAP (600 ha)	K_g4	K_{sat} from PTF-1	16.2	22.9	34.4 %	34.4 %	975.0	685.0	71.3	2.1	2.1
	K_g5	K_{sat} from PTF-2	17.6	24.9	28.7%	28.6 %	896.0	632.0	68.2	2.1	2.6
	K_g6	K_{sat} from PTF-3	16.4	23.2	33.6%	33.5 %	962.0	676.0	70.9	2.0	2.2
	K_g7	K_{sat} from PTF-4	15.6	22.1	36.8 %	36.7 %	1010.0	711.0	72.6	1.8	2.0
	K_g8	**Measured K_{sat}**	**56.0**	**60.5**	**0%**	**0%**	**238.0**	**220.0**	**50.0***	**11.2**	**4.5**
	K_g9	K_{sat} from ANN3	47.0	50.8	16 %	16 %	275.0	257.0	50.3	11.5	4.3
	K_g10	K_{sat} from ANN4	46.4	50.1	17.1%	17.2 %	284.0	266.0	51.0	11.2	4.2
HAP (500 ha)	K_g11	K_{sat} from PTF-1	12.8	13.8	77.1%	77.2 %	1027.0	954.0	85.0	1.2	0.2
	K_g12	K_{sat} from PTF-2	23.4	25.3	58.2 %	58.2 %	559.0	514.0	75.9	3.1	1.0
	K_g13	K_{sat} from PTF-3	19.2	20.9	65.7%	65.5 %	679.0	624.0	79.7	1.5	0.9
	K_g14	K_{sat} from PTF-4	20.1	21.6	64.1%	64.3 %	651.0	605.0	79.1	1.5	1.0

* target depth of water-table (Cornelius-Brochier J.V, 1981; Danenco, 1980; Holzmann-Wakuti, 1974; Italconsult, 1976). (K_{sat}) saturated hydraulic conductivity, (K_g) geometric mean of K_{sat}, (PTF) pedotransfer function, (ANN) Artificial Neural Network), (EAP) Eshkeda Agricultural Project, and (HAP) Hammam Agricultural Project.

the measured data for ANN3 and ANN4 was 16.0 and 17.2 % respectively. In contrast, for the PTFs (K_g 11–14) % variation in drain spacing from the measured data ranged from 58.2 to 77.2%. The target depth of water-table for EAP and HAP was 50 cm from the soil surface (as obtained by K_g1 and K_g8). Generally, the variation of water depth based on K_g derived from the four ANNs were closer to the target water-table depth with % variation ranging between 2.0–6.0% and 0.6–2.0%, in EAP and HAP respectively. In contrast, the % variations in water depth for the PTFs compared to the target design ranged from 27.0 to 31.0% in K_g (4–7) and 34.0–41.0% in K_g (11–14).

The estimation of the initial cost of installing a drainage system in the designated data-poor areas of EAP (600 ha) and HAP (500 ha) using measured K_g1 and K_g8 (Table 8) was US$ 638,000 and US$ 450,000 in EAF and US$ 238,000 and US$ 220,000 in HAP, for date palm and alfalfa, respectively. The results show that for both EAP and HAP, installing drainage schemes based on K_{sat} derived from ANNs is more economical than designs using K_{sat} derived from PTFs. In EAP K_g2 derived from ANN2 resulted in only a 1.1 and 1.3 % increase, compared to measured data, in initial installation costs for date palm and alfalfa, respectively (Table 8). In contrast, K_g3 derived from ANN3 resulted in a 4.4 % and 4.3 % decrease in initial installation costs, compared to measured data, for date palm and alfalfa respectively. In HAP, K_g9 and K_g10 derived from ANN3 and ANN4 result in a 15.5–19.3% and 16.8–20.9% increase in initial installation costs, compared to measured data, for date palm and alfalfa respectively. In contrast, K_g11-14 derived from PTFs 1–4 result in substantial 234.0–431.0% and 233.0–433.0% increases in initial installation cost, compared to measured data, for date palm and alfalfa respectively.

In comparison to the initial drainage costs calculated based on measured data, the extra initial cost of designing drainage systems in EAP and HAP using K_g2, K_g3, K_g9, and K_g10 ranges between $6,000 to $46,000. In contrast, the extra initial cost of designing drainage systems in EAP and HAP using K_g values derived from PTFs is an order of magnitude higher ranging from $182,000 to $790,000.

4. Discussion

4.1. Development of ANNs

All site-specific ANNs developed in this study were able to accurately predict K_{sat} (see Figs. 4 and 5). In this study, to predict K_{sat} values of soils in designated data-poor areas of EAP and HAP, ANNs were developed using selected key input variables. The input parameters for models ANN2 and ANN4 (% S, Si and C and BD) do not describe the structure or connectivity of void spaces in soil, therefore, they do not explain the connections between soil structures and functions very well (Nemes et al., 2003; Pachepsky et al., 2006). Anderson and Bouma (1973) found that an excellent prediction (as close as possible to the observed values) of K_{sat} can be obtained when pore space is measured directly. Therefore, the ANNs that also include FC and WP (ANN1 and ANN3) increased the accuracy of prediction of K_{sat} in comparison to ANN2 and ANN4 (see Figs. 4 and 5).

The K-fold cross-validation was undertaken to estimate the test error rate within the training dataset in order to confirm the confidence of applying the developed ANN to the designated data-poor areas. The results presented in Fig. 3b show that for areas with similar soil characteristics to EAP and HAP the number of K_{sat} measurements required to generate an accurate ANN can be reduced to *circa* 50 (taking into account the total area of each scheme ranging between 1000–3000 ha). These results support the findings of Yang (1995) where the quality of an ANNs training dataset is more important than the quantity of values used. This can be seen in mis-prediction of K_{sat} values of 30.0 m d^{-1} or above in ANN3 and ANN4, when the highest measured K_{sat} value used to develop these ANNs was 28.0 m d^{-1}. In addition, Yang (1995) added that, if the model was trained with a full range of expected situations, then It would require fewer input values to develop the training dataset, which means having all the expected values of the dependent variable within the training dataset. For example, in this study, K_{sat} was the dependent variable needed to be predicted for a specific location (data-poor area). By having all ranges of K_{sat} within the training dataset, the developed ANN will be more accurate.

Generally, the non-site-specific ANNs failed to accurately predict K_{sat} in EAP and HAP. In EAP, ANN3 and ANN4 over-predict K_{sat} especially the measured values between 0.6 to 10.0 m d^{-1}. Whereas in HAP, ANN1 and ANN2 under-predict K_{sat}. These results show that an ANN developed in a specific location may not be transferrable to another location even within the same pedoclimatic region. However, the combined ANNs (ANN5 and ANN6) illustrate that soil datasets based on more than one location are able to accurately predict K_{sat}.

4.2. PTFs performance

The PTFs applied in this study failed to accurately predict K_{sat} values especially for values $\geq$10.0 m d^{-1}, which represent 15.0 % and 54.0 % of the designated data-poor areas in EAP and HAP respectively. This is due to the PTFs evaluated being developed (trained and tested) for very different soils. For example, the dataset sources used to develop PTF-1, PTF-2 and PTF-4 were US soils, while the dataset source used to develop PTF-3 was from Spain (Cosby et al., 1984; Dane and Puckett, 1994; Julia et al., 2004; Puckett et al., 1985). Critically, the soil texture classes used to develop the PTFs were different to the dominant soil textures in EAP and HAP. For example, the percentage of the dataset associated with sand and loamy sand in EAP was 96.4 % (Table 1), while the percentage of these soil textures in the dataset used to develop PTF-1, PTF-2, PTF-3 and PTF-4 were 14.0 %, 53.0 %, 1.1 % and 3.0 %, respectively. Similarly, the combined percentage of sand, loamy sand and sandy loam soils in HAP was 72.0 %. In contrast, the combined percentage of these soil texture classes in the dataset used to develop PTF-1, PTF-2, PTF-3 and PTF-4 were, respectively, 38.0 %, 60.0 %, 7.0 % and 11.6 %. Consequently, the K_{sat} values predicted using the PTFs were poorly correlated to the measured values of K_{sat}. In fact, the PTFs predicted K_{sat} with values ranging between 0.002–6.3 m d^{-1} while the measured values ranged between 0.4–38.0 m d^{-1}. The PTFs underestimated K_{sat} by 83.0–99.0%. Consequently, at both EAP and HAP, the K_{sat} values predicted by the PTFs was significantly less

accurate than those predicted using ANNs. These results support the findings of Pringle et al. (2007), who found that all regression-based PTFs naturally tend to smooth the predicted values. The underestimation of K_{sat} by the selected PTFs is predominantly due to differences in soil texture. The soil texture classes in the dataset used to develop PTF-1, PTF-2, PTF-3 and PTF-4 were dominated by fine textured classes such as loam, sandy clay loam, clay loam, sandy clay and clay (Cosby et al., 1984; Dane and Puckett, 1994; Julia et al., 2004; Puckett et al., 1985).

PTF-2, developed by Dane and Puckett (1994) as compared to the other PTFs had low RMSE (22.0–27.0% and 14.0–15.0% in EAP and HAP respectively) and low MSE (39.0–46.0% and 26.0–28.0% in EAP and HAP respectively) in predicted K_{sat}. This observation agrees with the findings of Minasny and McBratney (2000a), where PTF-2 gave better prediction of K_{sat} for sandy soils compared to other PTFs. In addition, this is due to the soil texture classes in the dataset used to develop PTF-2 that comprised 53.0 % sandy and loamy sand and 60.0 % sandy, loamy sand and sandy loam. However, it is known that PTFs are not as reliable in their predictions when applied to different soils beyond the geomorphological and pedoclimatic region or soil texture from which it was originally derived (Minasny and McBratney, 2000b; Tietje and Hennings, 1996). Therefore, it was advised by Wösten (1997); Nemes (2015) and Van Looy et al. (2017) that PTFs should only be applied to the same pedoclimatic conditions from where they were developed. Nemes et al. (2003) suggested that it is better to use a small set of relevant data to develop PTFs, than to use a large and irrelevant data set for a specific area. Moreover, the application of PTFs in estimating K_{sat} at a specific location that contains different dominant soil textures might yield results with very limited accuracy and relevance if these structural features are disregarded (Lin et al., 1999; Vereecken et al., 2010). These PTFs will therefore be of limited use for estimating water transport in a specific field and/or a specific zone within a field (Parasuraman et al., 2006). The results of the current study corroborate these findings. Therefore, for field-scale application purposes, developing PTFs from a small set of relevant site-specific data may turn out to be more successful than using PTFs derived from a large but more general dataset where locally specific conditions such as structural features might be missing (Nemes et al., 2003). For the study areas (EAP and HAP) the developed ANNs are a more appropriate and accurate way of estimating K_{sat} than adoption of existing PTFs.

4.3. Drainage design implications

In EnDrain K_{sat} was the only parameter that varied in each design, all other parameters remained constant. An underestimation of K_{sat} by the PTFs resulted in an underestimation of the geometric mean (K_g) by 31.0%–96.0% from the measured $K_{sat,}$ in EAP and HAP, respectively. The consequence of this was a 29.0–37.0% and 58.0–77.0% shorter predicted distance between drains in EAP and HAP, respectively. A drainage design based on these predictions will have: (1) a lower than expected water level, and (2) have a higher drainage density. Such a design will incur significant extra initial cost as well as increased maintenance costs, compared to a design based on measured data. However, the drainage designs based on ANNs predictions gave drain spacing's and water-table depth equivalent to the target water-table depth, meaning the initial cost and subsequent maintenance costs would be almost equivalent to that based on measured data. For EAP this variation ranged between -4.4 – 1.3 % and in HAP between 14.4–17.3% respectively. This makes the use of ANNs to design drainage systems significantly more economical than the PTFs. Consequently, the ANNs developed in this study have considerable potential to reduce the cost of the design and maintenance of drainage systems as compared to the adoption of PTFs and are a less time-consuming alternative to traditional soil sampling and analyses for a specific area. Therefore, so long as the increase in installation cost in EAP and HAP is less than the saving made in the sampling and analytical programmes, adoption of ANN1–4 is cost-effective. In addition, ANN development indicates that if site-specific ANNs are to be developed, the number of soil samples required to create an accurate ANN can be substantially reduced. Finally, the results in Table 8 show that the K_{sat} derived from ANNs can be applied to the design of a drainage system that would maintain the water-table depths at equivalent depths to directly measured data (differences from the target designs ranged between 1.0–6.0% in EAP and HAP) providing those models are developed using datasets that meet at least the minimum quantity and quality requirements. In contrast, designs based on PTFs predictions would result in a water-table depth that differed from the target design. In EAP the difference from the target value ranged between 27.0–30.0% and in HAP ranged between 34.0–41.0%.

5. Conclusions

ANNs as an indirect method of predicting K_{sat} have been applied by several researchers (Agyare et al., 2007; More and Deka, 2018; Schaap et al., 2001; Sedaghat et al., 2016). As a rapid and reliable method, it shows good potential, but its accuracy is dependent on the quality and quantity of the dataset used to develop the ANN. In this research, ANNs were applied to predict K_{sat} in designated data-poor areas of two case study drainage schemes located in the south of Libya. It was possible to demonstrate that ANNs developed using easily measured existing and under-utilised data such as the % of sand, silt and clay, bulk density, field capacity and wilting point, were able to provide highly accurate predictions of K_{sat}. It was also shown that K_{sat} predictions based on ANNs were significantly more comparable to measured K_{sat} results than those predicted using widely adopted PTFs. Only K_{sat} predicted from ANNs could be used to accurately calculate both the target drain spacing as well as water-table depth in a drainage scheme designed using EnDrain. Therefore, assuming sampling cost is not considered, initial drainage design costs (based on length and depth of inserted pipe) would be comparable between designs based on ANN predictions and those based on measured K_{sat}. These results also suggest that ANNs can be successfully developed using existing and under-utilised datasets. As K_{sat} is time-consuming to measure, basing drainage designs on ANN predictions generated from a minimum dataset will be more cost-effective and therefore more accessible to farmers, planners, and decision-makers in least developed countries. However, further research is needed to evaluate the impact of these designs on crop production, salinity control and waterlogging, using a simulation model such as DRAINMOD, as well as, calculating the predicted

income based on these designs for different cropping systems.

CRediT authorship contribution statement

Murad Ellafi: writing-original draft, conceptualization, methodology, validation, formal analysis, investigation. **Lynda Deeks:** Supervision, writing-review and editing, review of analysis. **Rob Simmons:** Supervision, writing-review and editing, review of analysis.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgement

This work was supported by Libyan Ministry of Higher Education & Scientific Research, via the Government of National Unity, Libyan Academic Attaché – London (FA042-185-5447).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrh.2021.100832.

References

Abdelbaki, A., Youssef, M.A., Naguib, E.M., Kiwan, M., El-giddawy, E.I., 2009. Evaluation of pedotransfer functions for predicting saturated hydraulic. In: ASABE Annual International Meeting. ASABE, Reno, Nevada. https://doi.org/10.13031/2013.27309.

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22, 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1.

Agyare, W.A., Park, S.J., Vlek, P.L.G., 2007. Artificial neural network estimation of saturated hydraulic conductivity. Vadose Zone J. 6, 423–431. https://doi.org/10.2136/vzj2006.0131.

Ahuja, L., Naney, J., Green, R., Nielsen, D., 1986. Macroporosity to characterize spatial variability of hydraulic conductivity and effects of land management. Soil Sci. Soc. Am. J. 48, 699–702.

Aimrun, W., Amin, M.S.M., Eltaib, S.M., 2004. Effective porosity of paddy soils as an estimation of its saturated hydraulic conductivity. Geoderma 121, 197–203. https://doi.org/10.1016/j.geoderma.2003.11.010.

Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. Science (80-.) 348 (1–6), 1261071. https://doi.org/10.1126/science.1261071.

Anderson, J.L., Bouma, J., 1973. Relationships between saturated hydraulic conductivity and morphometric data of an argillic horizon. Soil Sci. Soc. Am. J. 37, 408–413.

Arrington, K.E., Ventura, S.J., Norman, J.M., 2013. Predicting saturated hydraulic conductivity for estimating maximum soil infiltration rates. Soil Sci. Soc. Am. J. 77, 748–758. https://doi.org/10.2136/sssaj2012.0288.

Arshad, R.R., Sayyad, G., Mosaddeghi, M., Gharabaghi, B., 2013. Predicting saturated hydraulic conductivity by artificial intelligence and regression models. ISRN Soil Sci 2013, 308159, 8 pages.

Ayars, J.E., Evans, R.G., 2015. Subsurface drainage - what's next? Irrig. Drain. 64, 378–392. https://doi.org/10.1002/ird.1893.

Ayoubi, S., Pilehvar, A., Mokhtari, P., L., K, 2011. Application of Artificial Neural Network (ANN) to predict soil organic matter using remote sensing data in two ecosystems. In: Atazadeh, I. (Ed.), Biomass and Remote Sensing of Biomass. InTech Open Access, London, UK, pp. 181–196. https://doi.org/10.5772/18956.

Azimi, H., Bonakdari, H., Ebtehaj, I., 2019. Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel. Appl. Water Sci. 9, 78. https://doi.org/10.1007/s13201-019-0961-5.

Baker, L., Ellison, D., 2008. Optimisation of pedotransfer functions using an artificial neural network ensemble method. Geoderma 144, 212–224. https://doi.org/10.1016/j.geoderma.2007.11.016.

Bell, J., 2015. Machine learning with R. Machine Learning, 2nd ed. PACKT, Birmingham. https://doi.org/10.1002/9781119642183. ch14.

Bicknell, B., Imhoff, J., Kittle, J., Jobes, T., Donigian, A., 2005. Hydrological Simulation Program—Fortran (HSPF).

Bittelli, M., Campbell, G., Tomei, F., 2015. Soil Physics With Python: Transport in the Soil-plant-Atmosphere System. Oxford University Press, New York, USA.

Bouksila, F., Persson, M., Berndtsson, R., Bahri, A., 2010. Estimating soil salinity over a shallow saline water table in semiarid Tunisia. Open Hydrol. J. 4, 91–101.

Bouyoucos, G., 1927. The hydrometer as new method for mechanical analysis of soils. Soil Sci. 23, 343–353.

Buckman, H., Brady, N., 1960. The Nature and Properties of Soil, 6th ed. Macmillan company, New York, USA.

Chollet, F., Allaire, J., 2017. Deep Learning With R. MEAP ed. Manning.

Chung, S., Ward, A., Schalk, C., 1992. Evaluation of the hydrologic component of the ADAPT water table management model. Trans. ASAE 35, 571–579.

Cleveland, C.C., Houlton, B.Z., Smith, W.K., Marklein, A.R., Reed, S.C., Parton, W., Del, S.J., Running, S.W., 2013. Patterns of new versus recycled primary production in the terrestrial biosphere. PNAS 110, 12733–12737. https://doi.org/10.1073/pnas.1302768110.

Cornelius-Brochier, J.V., 1981. Wadi Al Shatti: Drainage Design. Tripoli, Libya.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20, 682–690.

Dagan, G., 1965. Steady drainage of a two-layered soil. J. Irrigat. Drainage Div. Proc. ASCE 91, 51–64.

Dane, J., Puckett, W., et al., 1994. Field soil hydraulic properties based on physical and mineralogical information. In: van Genuchten, MTh (Ed.), The International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. University of California, Riverside, CA, pp. 389–403.

Danenco, 1980. Technical Study for Jufra el Hamam Project. Copenhagen, Denmark.

Dashtaki, S.G., Homaee, M., Khodaverdiloo, H., 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. Soil Use Manage. 26, 68–74. https://doi.org/10.1111/j.1475-2743.2009.00254.x.

Dumm, L., 1954. Drain spacing formula: new formula for determining depth and spacing of subsurface drins in irrigated lands. Am. Soc. Agric. Eng. 10, 726–730.

Dumm, L., 1960. Validity and use of the transient-flow concept in subsurface drainage. Am. Soc. Agric. Eng.
Ebtehaj, I., Bonakdari, H., Zaji, A.H., 2018. A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes. Alexandria Eng. J. 57, 1783–1795. https://doi.org/10.1016/j.aej.2017.05.021.
Ebtehaj, I., Bonakdari, H., Zaji, A.H., Gharabaghi, B., Hossein, A., Bahram, Z., 2020. Evolutionary optimization of neural network to predict sediment transport without sedimentation. Complex Intell. Syst. https://doi.org/10.1007/s40747-020-00213-9.
Elmeer, K., Ahmed, A., Serqiwa, S., 2016. Morphological diversity of three elite date palm (Phoenix dactylifera L.) cultivars grown in South of Libya. In: The Fourth Scientific Conference of Environment and Sustainable Development in the Arid and Semi-Arid Regions. University of Ejdabya, Ejdabya, Libya.
Elshemy, M., 2018. Review of technologies and practices for improving agricultural drainage Water quality in Egypt. Unconventional Water Resources and Agriculture in Egypt. Springer Verlag, pp. 163–188. https://doi.org/10.1007/698.
FAO, 1989. Arid Zone Forestry: A Guide for Field Technicians. Food & Agriculture Org, Rome, Italy.
FAO, 2002. Agricultural Drainage Water Management in Arid and Semi-arid Areas. FAO Irrigation and Drainage. https://doi.org/10.1177/1362361399003003007.
Fooladmand, H.R., 2011. Pedotransfer functions for point estimation of soil moisture characteristic curve in some Iranian soils. African J. Agric. Res. 6, 1586–1591. https://doi.org/10.5897/AJAR10.1152.
General Water Authority, 1999. Water and Irrigation Requirements for the Most Important Crops in Libya [unpublished Report]. Tripoli.
Ghatak, A., 2019. Deep Learning with R. https://doi.org/10.1007/978-981-13-5850-0.
Hassler, S., Lark, R., Zimmermann, B., Elsenbeer, H., 2014. Which sampling design to monitor saturated hydraulic conductivity? Eur. J. Soil Sci. 792–802. https://doi.org/10.1111/ejss.12174.
Hess, K.M., Wolf, S.H., Celia, M.A., 1992. Large-scale natural gradient tracer test in sand and gravel, hydraulic conductivity variability and calculated macrodispersivities. Water Resour. Res. 28, 2011–2027.
Holzmann-Wakuti, 1974. Hydrogeology of The Joufrah Project Area. Tripoli, Libya.
Hooghoudt, S., 1937. Contribution of the Knowledge of Some Physical Soil Properties. NO 6. Detremination of the Conductivity of Soils of the Second Kind (in Dutch). Bijdragen Tot De Kennis Vaneenige Natuurkundige Van Den Grand,6. Algemeene Beschonwing Van Het Problem Van De D. The Hauge, Netherlands.
Hooghoudt, S., 1940. Contribution of the Knowledge of Some Physical Soil Properties. NO 7. Detremination of the Conductivity of Soils of the Second Kind (in Dutch). Bijdragen Tot De Kennis Vaneenige Natuurkundige Van Den Grand, 7. Algemeene Beschonwing Van Het Problem Van De. The Hauge, Netherlands.
Huang, Y., 2009. Advances in artificial neural networks - methodological development and application. Algorithms 2, 973–1007. https://doi.org/10.3390/algor2030973.
Huang, M., Lin, R., Huang, S., Xing, T., 2017. A novel approach for precipitation forecast via improved K -nearest neighbor algorithm. Adv. Eng. Informatics 33, 89–95. https://doi.org/10.1016/j.aei.2017.05.003.
Italconsult, 1976. Fezzan Land Reclamation Project: Wadi-Shati Eshkeda Area. Rome, Italy.
Jothi, V.L., Neelambigai, A., Nithish, S., Santhosh, K., 2020. Crop yield prediction using KNN model. Int. J. Eng. Res. Technol. 8, 4–7.
Julia, M., Monreal, T., Sanchez del Corral Jimeneza, A., Garc Melendez, E., 2004. Constructing a saturated hydraulic conductivity map of Spain using Pedotransfer functions and spatial prediction. Geoderma 123, 257–277.
Katul, G.G., Oren, R., Manzoni, S., Higgins, C., Parlange, M.B., 2012. Evapotranspiration : a process driving mass transort and energy exchange in the soil-plant-atmosphere-climate system. Am. Geophys. Union 50, 1–25. https://doi.org/10.1029/2011RG000366.1.INTRODUCTION.
Khademi, F., Jamal, S.M., 2016. Research papers predicting the 28 days compressive strength of concrete using artificial neural network. i-manager's. J. Civ. Eng 6, 1–6. https://doi.org/10.26634/jce.6.2.5936.
Khodaverdiloo, H., Homaee, M., Th, M., Genuchten, V., Ghorbani, S., 2011. Deriving and validating pedotransfer functions for some calcareous soils. J. Hydrol. 399, 93–99. https://doi.org/10.1016/j.jhydrol.2010.12.040.
Kirkham, D., 1958. Seepage of steady rainfall through soil into drains. Trans. Am. Geophys. Union 39, 892–908.
Klute, A., Dirksen, C., 1986. Hydraulic conductivity and diffusivity: laboratory methods. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods. ASA, Madison, WI, pp. 687–734.
Kombo, O.H., Kumaran, S., Sheikh, Y.H., Bovim, A., Jayavel, K., 2020. Long-term groundwater level prediction model based on hybrid KNN-RF technique. Hydrology 7, 1–24.
Li, Y., Chen, D., White, R., Zhu, A., Zhang, J., 2007. Estimating soil hydraulic properties of Fengqiu County soils in the North China Plain using pedo-transfer functions. Geoderma 138, 261–271.
Lin, H., Mclnnes, K., Wilding, L., Hallmark, C., 1999. Effects of soil morphology on hydraulic properties: II. Hydraulic Pedotransfer Functions. Soil Sci. Soc. Am. J. 63, 955–961.
Luthin, J., 1978. Drainage Engineering. Krieger Publishing, Huntington, N.Y.
Luthin, J., Worstell, R., 1959. The falling water table in tile drainage: III factors affecting the rate of fall. Trans. ASAE 2 (45–47), 51.
Ma, L., Ahuja, L.R., Nolan, B.T., Malone, R.W., Trout, T.J., Qi, Z., 2012. Root zone water quality model (RZWQM2): model use, calibration, and validation. Trans. ASAE 55, 1425–1446.
Mallikarjuna, B., Sathish, K., Venkata Krishna, P., Viswanathan, R., 2020. The effective SVM-based binary prediction of ground water table. Evol. Intell. https://doi.org/10.1007/s12065-020-00447-z.
Merdun, H., Çınar, Ö., Meral, R., Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil Tillage Res. 90, 108–116.
Minasny, B., Mcbratney, A.B., 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66, 352–361. https://doi.org/10.2136/sssaj2002.0352.
Minasny, B., McBratney, A.B., 2000a. Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. Aust. J. Soil Res. 38, 905–926. https://doi.org/10.1071/SR99110.
Minasny, B., McBratney, A.B., 2000b. Estimation of sorptivity from disc-permeameter measurements. Geoderma 95, 305–324. https://doi.org/10.1016/S0016-7061(99)00096-8.
Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., Vereecken, H., 2017. A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. Earth Syst. Sci. Data Discuss. 9, 529–543.
More, P., 2018. Using Machine Learning to Predict Water Table Levels in a Wet Prarie in Northwest Ohio (MSc Thesis). Bowling Green State University.
More, S.B., Deka, P.C., 2018. Estimation of saturated hydraulic conductivity using fuzzy neural network in a semi-arid basin scale for murum soils of India. ISH J. Hydraul. Eng. 24, 140–146. https://doi.org/10.1080/09715010.2017.1400408.
Nemes, A., 2015. Why do they keep rejecting my manuscript - do's and don'ts and new horizons in pedotransfer studies. Agrokem. es Talajt. 64, 361–371. https://doi.org/10.1556/0088.2015.64.2.4.
Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67, 1093–1102. https://doi.org/10.2136/sssaj2003.1093.
Nozari, H., Azadi, S., 2017. Experimental evaluation of artificial neural network for predicting drainage water and groundwater salinity at various drain depths and spacing. Neural Comput. Appl. https://doi.org/10.1007/s00521-017-3155-9.
OECD, 2020. DAC List of ODA Recipients Effective for Reporting on 2020 Flows [WWW Document]. URL http://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/DAC-List-of-ODA-Recipients-for-reporting-2020-flows.pdf (Accessed 5.20.20).
Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. Science (80-.) 313, 1068–1073.
Oosterbaan, R., 1993. EnDrain: Software Program for Subsurface Drainage Equations Using the Energy Balance of Groundwater Flow, Permitting Anisotropic and Stratified Soils, As Well As Entrance Resistance to Drains.
Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. Geoderma 131, 308–316. https://doi.org/10.1016/j.geoderma.2005.03.012.

Parasuraman, K., Elshorbagy, A., Si, B.C., 2006. Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Sci. Soc. Am. J. 70, 1851–1859. https://doi.org/10.2136/sssaj2006.0045.
Patil, N.G., Singh, S.K., 2016. Pedotransfer functions for estimating soil hydraulic properties : a review. Pedosph. An Int. J. 26, 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6.
Patil, N.G., Rajput, G., Nema, R., Singh, R., 2010. Predicting hydraulic properties of seasonally impounded soils. J. Agric. Sci. Cambridge 148, 159–170.
Patil, N.G., Pal, D., Mandal, C., Mandal, D., 2011. Soil water retention characteristics of Vertisols and pedotransfer functions based on nearest neighbor and neural networks approach to estimate AWC. J. Irrig. Drain. Eng. 138, 177–184.
Pringle, M.J., Romano, N., Minasny, B., Chirico, G.B., Lark, R.M., 2007. Spatial evaluation of pedotransfer functions using wavelet analysis. J. Hydrol. 333, 182–198. https://doi.org/10.1016/j.jhydrol.2006.08.007.
Puckett, W.E., Dane, J.H., Hajek, B.F., 1985. Physical and mineralogical data to determine soil hydraulic properties 1. Soil Sci. Soc. Am. J. 49, 831–836.
Reynolds, W., Elrick, D., 1985. In situ measurement of field-saturated hydraulic conductivity, sorptivity, and the α parameter using the Guelph Permeameter. Soil Sci. 140, 172–180.
Richards, L., 1949. Methods of measuring soil moisture tension. Soil Sci. 68, 95–112.
Ritzema, H., 1994. Subsurface flow to drains. In: Ritzema, H. (Ed.), Drainage Principles and Applications. International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands, pp. 236–304.
Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 2001. Rosetta : a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251, 163–176.
Schultz, B., Zimmer, D., Voltman, W.F., 2007. Drainage under increasing and changing requirements. Irrig. Drain. 56, S3–S22. https://doi.org/10.1002/ird.
Sedaghat, A., Bayat, H., Sinegani, A.A.S., 2016. Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soils. Eurasian Soil Sci. 1 (49), 377–387. https://doi.org/10.1134/S106422931603008X.
Skaggs, R.W., 1978. A Water Management Model for Shallow Water Table Soils.
Skaggs, R.W., 2007. Criteria for calculating drain spacing and depth. Am. Soc. Agric. Biol. Eng. 50, 1657–1662.
Smedema, L.K., Abdel-Dayem, S., Ochs, W.J., 2000. Drainage and agricultural development. Irrig. Drain. Syst. Eng. 14, 223–235. https://doi.org/10.1023/A:1026570823692.
Sobieraj, J.A., Elsenbeer, H., Cameron, G., 2004. Scale dependency in spatial patterns of saturated hydraulic conductivity. Catena 55, 49–77. https://doi.org/10.1016/S0341-8162(03)00090-0.
Soil Survey Division Staff, 1993. Soil Survey Manual. U.S. Department of Agriculture Handbook 18. Natural Resources Conservation Service.
Stefan, F., Guenther, F., 2019. Package 'neuralnet' [WWW Document]. URL https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf (Accessed 5.13.20).
Sun, J., 2012. Learning algorithm and hidden node selection scheme for local coupled feedforward neural network classifier. Neurocomputing 79, 158–163. https://doi.org/10.1016/j.neucom.2011.09.019.
Tantawi, A., 2000. Water Resources in Libya. Egyptian Office for the Distribution of Publications, Cairo.
Tietje, O., Hennings, V., 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69, 71–84. https://doi.org/10.1016/0016-7061(95)00050-X.
Van Beers, W.F.J., 1958. The Auger-hole Method.
Van Dam, J., Huygen, J., Wesseling, J., Feddes, R., Kabat, P., Van Walsum, P.E., Groenendijk, P., van Diepen, C., 1997. Theory of SWAP Version 2.0. Simulation of Water Flow, Solute Transport and Plant Growth in the Soil-water-Atmosphere-Plant Environment. Wageningen.
van der Molen, W.H., Wesseling, J., 1991. A solution in closed form and a series solution to replace the tables for the thickness of the equivalent layer in Hooghoudt's drain spacing formula. Agric. Water Manage. 19, 1–16. https://doi.org/10.1016/0378-3774(91)90058-Q.
van der Ploeg, R., Horton, R.M., Kirkham, D., 1999. Steady flow to drains and wells. In: Skaggs, R., van Schilfgaarde, J. (Eds.), Agricultural Drainage. Wisc.: SSSA, Madison, pp. 213–264.
Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer functions in earth system science: challenges and perspectives. Rev. Geophys. 55, 1199–1256. https://doi.org/10.1002/2017RG000581.
van Schilfgaarde, J., 1963. Design of tile drainage for falling water tables. J. Irrig. Drain. Div. 89, 1–12.
Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten, M.T., 2010. Using pedotransfer functions to estimate the van genuchten-mualem soil hydraulic properties: a review. Vadose Zone J. 9, 795–820. https://doi.org/10.2136/vzj2010.0045.
Vereecken, A., Schnepf, J., Hopmans, M., Javaux, D., Or, D., Roose, T., Vanderborght, J., Young, M.H., Amelung, W., Aitkenhead, M., Allison, S.D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Franssen, H.J.H., Heppell, J., Horn, R., Huisman, J.A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., Mcbratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E.C., Schwen, A., Šimůnek, J., Tiktak, A., Van Dam, J., Van Der Zee, S.E.A.T.M., Vogel, H.J., Vrugt, J.A., 2016. Modeling soil processes : review, key challenges, and new perspectives brief history of soil modeling. Vadose Zone J. 15, 1–57. https://doi.org/10.2136/vzj2015.09.0131.
Verhoef, A., Egea, G., 2014. Modeling plant transpiration under limited soil water : comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models. Agric. For. Meteorol. 191, 22–32. https://doi.org/10.1016/j.agrformet.2014.02.009.
Wösten, J.H.M., 1997. Pedotransfer functions to evaluate soil quality. In: Gregorich, E., Carter, M. (Eds.), Soil Quality for Crop Production and Ecosystem Health. ELSEVIER SCIENCE PUBLISHERS B.V, Amesterdam, The Netherlands, pp. 221–245.
Wösten, J.H., Pachepsky, Y., Rawls, W., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251, 123–150.
Yang, C., 1995. Applications of Artificial Neural Network Technology in the Design of Water-Table Management Systems. McGill University, Montreal, Quebec, Canada.
Youngs, E., 1999. Non-steady flow to drains. In: Skaggs, R., van Schilfgaarde, J. (Eds.), Agricultural Drainage. Wisc.: SSSA, Madison, pp. 265–296.
Zanetti, S.S., Sousa, E.F., Oliveira, V.P.S., Almeida, F.T., Bernardo, S., 2007. Estimating evapotranspiration using artificial neural network and minimum climatological data. J. Irrig. Drain. Eng. 133, 83–89.
Zhang, Y., Schaap, M.G., 2019. Estimation of saturated hydraulic conductivity with pedotransfer functions : a review. J. Hydrol. 575, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058.