International Journal of

Agricultural Research

ISSN 1816-4897

Academic
Journals Inc.

www.academicjournals.com

International Journal of Agricultural Research 7 (3): 166-168, 2012
ISSN 1816-4897 / DOI: 10.3923/ijar.2012.166.168

Estimating Water Content at Field Capacity and Permanent Wilting Point Using Non-parametric K-NN Algorithm

[1]Ali Keshavarzi, [1]Fereydoon Sarmadian, [1]Ali Asghar Zolfaghari and [2]Paria Pezeshki
[1]Department of Soil Science Engineering, University of Tehran, P.O. Box 4111, Karaj 31587-77871, Iran
[2]Department of Geology and Soil Science, Physical Land Resources Group, Ghent University, Krijgslaan 281 (S8), 9000, Ghent, Belgium

Corresponding Author: Ali Keshavarzi, Department of Soil Science Engineering, University of Tehran, P.O. Box 4111, Karaj 31587-77871, Iran

ABSTRACT

Through a comparison study in Ziaran region, Iran, we used a hierarchical set of inputs by using Bulk Density (Bd), sand, silt, clay and Organic Matter (OM) contents and developed the Artificial Neural Network (ANN) models and run the k-NN estimation algorithms to estimate water content at field capacity and permanent wilting point.

Key words: Field capacity, permanent wilting point, pedotransfer functions, k-NN algorithm

INTRODUCTION

Accurate measurement of soil water contents - to determine water availability which is a crucial factor in assessing the suitability of a land area for producing a given crop- is important in various applications in hydrology and soil science. Several attempts have been made to estimate indirectly these properties from more easily measurable and more readily available soil properties. Such relationships are referred to as Pedotransfer Functions (PTFs).

However, it has been observed that today's PTFs are all based on parametric approaches (equations) and identifying the right equation which ensure similar distribution of errors across the data space is not always an easy task. So, non-parametric techniques have been considered an alternative technique which recognizes patterns and similarities rather than fitting equations to data (Nemes *et al*., 2006). The use of non-parametric algorithm has been found beneficial when the form of relationship between input and output data is unknown a-priori (Yakowitz and Karlsson, 1987; Lall and Sharma, 1996). Such is the case with soil hydraulic properties, where the form of their dependence on other soil attributes is not known in advance. The k-Nearest Neighbors (k-NN) algorithm is an analogue approach (Lall and Sharma, 1996). This approach has its origin as a non-parametric statistical pattern recognition procedure to distinguish between different patterns according to a selection criteria.

In pattern recognition, the k-NN algorithm is a method for classifying objects based on closest training examples in the feature space and is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k-nearest neighbors (k is a positive integer, typically small). The k-NN algorithm can also be adapted for use in estimating continuous variables. One such implementation uses an inverse distance weighted average of the k-nearest multivariate neighbors. This algorithm functions as follows:

- **Step 1:** Compute Euclidean or Mahalanobis distance from target plot to those that were sampled
- **Step 2:** Order samples taking for account calculated distances
- **Step 3:** Choose heuristically optimal k-nearest neighbor based on RMSE done by cross validation technique
- **Step 4:** Calculate an inverse distance weighted average with the k-nearest multivariate neighbors

Through, a comparison study in Ziaran region, Iran, we used a hierarchical set of inputs by using Bulk Density (Bd), sand, silt, clay and Organic Matter (OM) contents and developed the Artificial Neural Network (ANN) models and run the k-NN estimation algorithms based on the method that proposed by Nemes *et al*. (2006) to estimate water content at field capacity and permanent wilting point. The k-NN method is based on recognizing a similar pattern of specified target soil parameters within a database of measured soil data which could be used as prediction of other parameters of the target soil. Target soil information as the initial seed of data with the soil database are required as input files for running the algorithm. The k-NN algorithm typically selects a specified number of soil parameters similar to the pattern of same variables in the database. The k-NN algorithm searches its memory (the soils database) to find the most similar soil from its database. According to Jagtap *et al*. (2004), it does not have to be an exact match but can be the closest match, called "best soil".

The k-NN technique and many of its derivatives belong to the group of 'lazy learning algorithms'. It is 'lazy', as it passively stores the development data set until the time of application; all calculations are performed only when estimations need to be generated. Application of this technique means identifying and retrieving the nearest (most similar) stored objects to the target object (Nemes *et al*., 2006).

In summary while ANN-based PTFs have been relatively successful, there are a number of reported weaknesses that need to be considered. These are number of coefficients (weights) that do not permit easy physical interpretation (Schaap *et al*., 2001), the ANNs's structure which has to be selected a priori and therefore, may not be optimal since there are many types of neurons and possible connections (Wosten *et al*., 2001) and there is no assurance that the learning algorithm will find optimum weights that minimize prediction errors. ANNs are very data demanding and such techniques have only been become possible when are used together with a large database. Considering the problems associated with ANN-based PTFs, it seems that k-NN approach is able to improve the PTFs accuracy and reliability.

ACKNOWLEDGMENTS

The authors are grateful to anonymous reviewers for their valuable comments and suggestions, Dr. Attila Nemes (Department of Environmental Sciences, Univ. of California, Riverside, CA), Dr. Walter J. Rawls (USDA-ARS Hydrology and Remote Sensing Lab., Beltsville, MD) and Dr. Yakov A. Pachepsky (USDA-ARS Environmental Microbial Safety Lab., Beltsville, MD) for providing the software used in the present study.

REFERENCES

Jagtap, S.S., U. Lall, J.W. Jones, A.J. Gismn and J.T. Ritchie, 2004. Dynamic nearest neighbor method for estimating soil water parameters. J. Am. Soc. Agric. Bio. Eng., 47: 1437-1444.

Lall, U. and A. Sharma, 1996. A nearest-neighbor bootstrap for resampling hydrologic time series. Water Resour. Res., 32: 679-693.

Nemes, A., W.J. Rawls and Y.A. Pachepsky, 2006. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. Soil Sci. Soc. Am. J., 70: 327-336.

Schaap, M.G., F.J. Leij and M.T. Van Genuchten, 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Hyrology, 251: 163-176.

Wosten, J.H.M., Y.A. Pachepsky and W.J. Rawls, 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol., 251: 123-150.

Yakowitz, S. and M. Karlsson, 1987. Nearest-Neighbor Methods with Application to Rainfall/Runoff Prediction. In: Stochastic Hydrology, Macneil, J.B. and G.J. Humphries (Eds). Springer, New York, pp: 149-160.