Development of a Spatial Hydrologic Soil Map Using Spectral Reflectance Band Recognition and a Multiple-Output Artificial Neural Network Model

Khamis Naba Sayl[1,3], Haitham Abdulmohsin Afan[1], Nur Shazwani Muhammad[1], Ahmed ElShafie[2]

[1]Department of Civil and Structural Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[2]Department of Civil Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
[3]Department of Dams and Water Resources, Engineering College, University of Anbar, Ramadi, Iraq

Correspondence to: Nur Shazwani Muhammad (shazwani.muhammad@ukm.edu.my)

The authors express sincere appreciation and thanked the Editorial Board and Reviewer #1 for the constructive comments. These comments will definitely improve the quality of the manuscript. The authors have addressed all of the concerns raised by Reviewer#1 in an itemized fashion below. The authors will include all suggestions in the revised version of the manuscript.

Reviewer #1:

- My main problem with the article is the extremely small dataset that the authors are using for their analysis. They used in total 25 soil samples, a small dataset that was split in 19 samples for validation and 6 for validation. Estimating an ANN on the basis of 19 samples that predicts sand, clay and silt contents on the basis of 9 bands of a LANDSAT 8 image just seems practically impossible or the authors should come with very good arguments. In addition, a validation on the basis of just 6 samples cannot provide accurate results.

Reply:

The authors completely agreed with the reviewer that the used data for developing ANN model is somewhat small in terms of number of collected data, which may not be recommended for the development of an ANN model. The main objective of this research is to propose a methodology for the recognition of soil textures and proof it using available data. The authors believe that the changes in the soil texture at certain pilot area are expected to be relatively minor. However, more data would be more helpful at different sites of the study area. Therefore, the authors suggest to change the title of this manuscript to reflect the overall study, i.e. “Towards the Development of a

Spatial Hydrologic Soil Map Using Spectral Reflectance Band Recognition and a Multiple-Output Artificial Neural Network Model".

- The authors talk about "soils", "soil types", and "soil survey", but they only talk about topsoil texture in terms of sand, silt, clay.

Reply:

The authors agreed with the reviewer on this issue. This study does focus on the topsoil texture in terms of sand, silt and clay because they represent the main variables of hydrological soil group. The authors will make sure that this term is consistent throughout the whole manuscript.

- Line 54: proximal sensing can also be done under field conditions.

Reply:

The following paragraphs will be added to give more details on this:

"Most reported studies revealed the high potential of proximal sensing to estimate soil properties based on clear absorption features at the laboratory and local scale. However, for large scale mapping, this exercise need to be extended beyond the plot scale. Important qualitative and quantitative soil information can be obtained from remote sensing data."

- Line 57: more important for efficient use of RS to characterize soil conditions is the soil cover (e.g., weeds, crop residues) and soil structure/roughness.

Reply:

Unsupervised classification was adopted in the first step of the methodology to determine vegetation cover. The Normalized Difference Vegetation Index (NDVI) is used primarily for vegetation identification and to determine the lushness of vegetated land surfaces. Also, as mentioned in lines 122-126, "The methodology has been used the satellite image of the study area,

represent arid region, (Landsat 8, August 2014) has been taken in dry season when the barren soil covers large area.

- Line 68: I agree that 50% is a poor correlation, but the 100% certainly is not a poor correlation. One should be more specific.

Reply:

The authors thank the reviewer for this valuable comment. This line is typo error. The authors have decided to delete this line to avoid any confusion regarding the correlation index.

- Line 102-106: The authors indicated in the introduction that soil texture is an important property for the assessment of runoff. However, here the authors talk about a flat arid area, where the high infiltration rates are considered to be a problem. This does not seem to be very consistent.

Reply:

The west desert of Iraq is classified as an arid region, with an uneven distribution of precipitation in time and space. The authors believe that the quantity of effective runoff is much more important as compared to the amount of annual rainfall. Most of the precipitation events in the study area are short duration and high intensity that occurred for short periods in a year. As a result, there is a call for more efficient water conservation alternatives.

- Line 110: if a large part of the plateau is rocky, this should be considered in the subsequent analysis. There is now reference how the authors dealt with the rocky area.

Reply:

The authors thank the reviewer for his comment. Actually, the rocky area is very small as compared to the overall study area. The authors will improve this statement in order to indicate that the rocky

part is ignored in the analysis due to the small percentage and hence it has insignificant influence on the results.

- The authors should be more specific on the procedures. For example, they mention that the unsupervised classification was used to identify sampling locations but not how (line 138). They were certainly not selected by a GPS device. Most likely they were first identified on the classified map and subsequently located with the GPS. Another example, is that the soil classifications were "manipulated" with ArcGIS (line 156). However, no details on what the manipulation meant.

Reply:

As mentioned in lines 124-127, the main purpose of unsupervised classification is to select the soil sampling locations on a primitive map based on a good depiction of some spectral classes and they are classified by colour. Then, we used GPS device to identify these locations on site.

The following paragraphs will be added to give more information about manipulation:

" Arc GIS spatial analyst extension can convert the themes, depending on vector features to grids. Additionally, grids can be derived from various spatial analysis operations, and it is added to be viewed as grid themes. These grid cells have been classified in various ways and different colors are chosen for each class where the colors represent the progression of values for a data attribute specified. It is achieved after the raster themes are converted into a shape file, which includes the environmental characteristics that represents the hydrological soil group"

- Line 140. Distinguishing properly between silt and clay is practically impossible with sieving. Details on the sample treatment are required.

Reply:

The authors agreed with the reviewer that this part is unclear. The authors have included the following paragraph in order to give more information on the difference between silt and clay in this study.

"The distribution of particle sizes larger than 75µm (retained on the No.200 sieve) is determined by sieving, while the distribution of particle sizes smaller than 75 µm is determined by sedimentation process, using a hydrometer to get the necessary data".

- Line 152: Details on the Radial Basis Neural Network Model are required.

Reply:

Authors agree with the reviewer that additional information on radial basis neural network is required. The following details will be added to the revised manuscript:

“Radial Basis Neural Network (RBNN) is an artificial method that based on the interpolation of a multivariate function. RBNN consists of three layers, i.e. input layer for feeding feature vector to the network, hidden layer where the calculation of outcome of basis function is processed, and finally the output layer for linear combining the basic functions. The following figure shows the structure of RBNN.

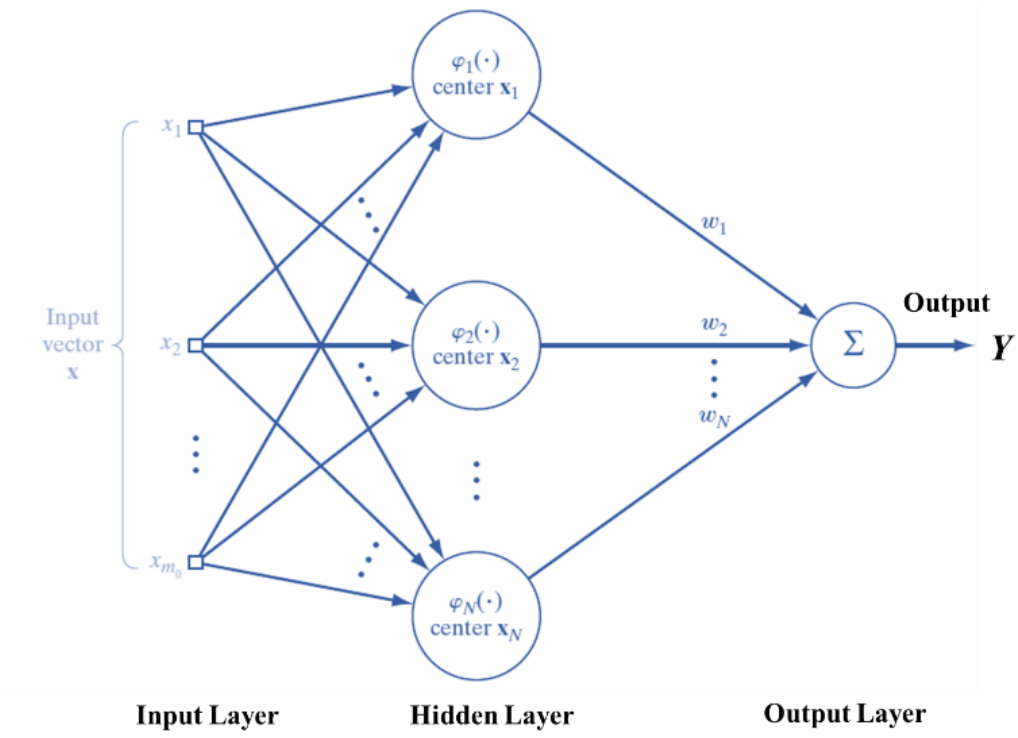

Figure (). Structure of Radial Basis Function Neural Network

The hidden layer applies a non-linear transformation from the input space to the hidden space. The output layer applies a linear transformation from the hidden space to the output space. The radial basis functions φ 1, φ 2, ... φ *N* are known as hidden functions while $\{\varphi i(x)\}_{i=1}^{N}$ is called the hidden space. The number of basic functions (*N)* is typically less than the number of data points available for the input data set. Among several radial basis functions, the most commonly used is the Gaussian, which in its one-dimensional representation takes the following form:

$$\varphi(x,\mu) = e^{-\frac{\|x-\mu\|^2}{2d^2}}$$

where μ is the center of the Gaussian function (mean value of *x)* and *d* is the distance (radius) from the center of $\varphi(x,\mu)$, which gives a measure of the spread of the Gaussian curve.

The hidden units use the radial basis function. If a Gaussian function is used, the output of each hidden unit depends on the distance of the input *x* from the center μ. During the training procedure,

the center μ and the spread *d* are the parameters to be determined. It can be deduced from the Gaussian radial function that a hidden unit is more sensitive to data points near the center. This sensitivity can be adjusted by controlling the spread *d*. It can be observed that the larger the spread, the less sensitivite radial basis function to the input data. The number of radial basis functions inside the hidden layer depends on the complexity of the mapping to be modeled and not on the size of the data set, which is the case when utilizing multi-layer perceptron ANN. Moreover, RBNN has the ability to recognize a complex relation between the input and output of the model. This research identify the relationship between the bands and soil types. RBNN model requires some important parameters to be established before perform the training process, such as the performance goal of 0.0005 and the spread constant of 1."

- In general, you could say that the description of the methodology is insufficient to be able to even apply the methodology in another area.

Reply:

Owing to the reviewer feedback, the authors have significantly improved the methodology section in order to include more detailed information on the proposed methodology for developing a hydrological soil map. The methodology section in the revised manuscript has been split into two sub-sections, first, the Unsupervised Classification Process and the second part for the Radial Basis Function Neural Network (RBNN) method as pattern recognition technique. As reported in the previous comment, the authors added more detailed information on RBNN model. In addition, the authors will the following paragraphs in order to provide more details on the unsupervised classification process performed in this study:

"Classification is the process of sorting pixels into a finite number of classes or categories of data based on their data file values. If a pixel satisfies a certain set of criteria, then it is assigned to the class that corresponds to that set of criteria. A pixel may be characterized by its spectral signature, which is determined by the relative reflectance in the different wavelength bands. Multispectral classification is an information-extraction process that analyzes these spectral signatures and then assigns pixels to categories based on similar signatures. There are two methods to classify pixels, i.e. unsupervised and supervised classification.

Unsupervised classification identifies clusters or groupings in a feature space. A cluster is a set of points in the feature space where their local density is large (relative maximum) compared to the density of feature points in the surrounding region. Techniques are useful for image segmentation and for classification of raw data to establish classes and prototypes. Clustering is also a useful vector quantization technique for compression of images. The clustering algorithm has no regard for contiguity of pixels that define each cluster. One of the most known unsupervised classifier method is the ISODATA. ISODATA is an iterative process where it repeatedly performs an entire classification (outputting a thematic raster layer) and recalculates the statistics. Self-Organizing refers to the way in which it locates clusters with minimum user input.

Each iteration calculates means and reclassifies pixels with respect to the new means. Iterative class splitting, merging, and deleting are done based on the input threshold parameters. All pixels are classified to the nearest class unless standard deviation or distance threshold is specified. One of the primary advantages of unsupervised classification is many of the classes are created automatically."

The authors believed that with this in-depth information in the methodology, including the added details on the unsupervised classification process and the radial basis function neural network (RBNN) , this section has improved significantly and could be applied in similar study area.

- Line 212: I wonder whether the authors can conclude on the basis of this study that the methodology is fast, reliable, and cost-effective.

Reply:

The authors conclude such benefits of the proposed methodology because the study area is located in a remote area and developing country such as Iraq. As mentioned in line 96, the total catchment area is 13370 km^2. For such a large and remote area, there are limited infrastructures and financial resources. Additionally, there is a need to collect soil samples from this area which in other word means that it is time-consuming, unreliable and expensive. Therefore, the authors consider that the proposed methodology with its level of accuracy could be faster, reliable and cost-effective as compared with the time required to collect the soil samples for this remote and relatively large area. In addition, the proposed method could be more reliable due to the untruthfulness of the

collected soil sampling in such remote area and disfigurement for the samples due to transferring them from remote area to the laboratory for analysis. Finally, the authors consider that the proposed methodology for identifying the hydrological soil map is cost-effectiveness as compared with the standard procedure because collecting samples from such remote areas is very expensive, in addition to the cost of analyzing the samples. The authors improved the conclusion to include more details as reported here in order to be more accurate on the benefits of the proposed methodology and its advantages over the standard procedure at a similar study area.