Soil salinity mapping utilizing sentinel-2 and neural networks

R.S. Morgan*1, M. Abd El-Hady2 and I.S. Rahim1

Soils and Water Use Department, Agricultural and Biological Research Division, National Research Centre, El Behouth St., Dokki, Cairo, Egypt.
Received: 09-12-2017 Accepted: 20-06-2018 DOI: 10.18805/IIARe.A-316

ABSTRACT
Soil salinity is the most important soil property that affects the agriculture productivity. Periodical monitoring of its status is considered a crucial factor in the selection of appropriate agricultural practices to attain a sustainable production. The availability of remote sensing data processed by a somewhat novel method such as artificial neural networks (ANN) offer a potential solution that could easily and affordably replace the in-site monitoring methods. The aim of this work is to use high spectral resolution Sentinel-2 (S2) data for soil salinity prediction utilizing neural networks. The study evaluated three approaches in processing the S2 data for inclusion in the artificial neural network for soil salinity prediction. These approaches included S2 spectral reflectance data, spectral indices and principal component analysis (PCA) of the S2 data. The results revealed that a combination of these approaches including the reflectance data of band 11(shortwave infrared band) of S2, the normalized differential vegetation index (NDVI) and the second PCA (dominated by the near infrared band) gave the best performance when used as input when designing the artificial neural networks to predict the soil salinity. The overall accuracy of this approach has a coefficient of determination (R2) of 0.94 between the actual and predicted soil salinity.

Key words: Artificial neural networks, Sentinel-2, Soil salinity.

INTRODUCTION
Soil salinity drastically influences the agricultural productivity especially in arid and semi-arid regions. Furthermore, salinity also affects other major soil degradation problems such as soil dispersion, increased soil erosion, and engineering problems (Metternicht and Zinck, 2003). On average, 20% of the world’s irrigated lands are affected by salts, but this figure increases to more than 30% in countries such as Egypt (Ghassemi et al., 1995). To reduce the negative environmental impacts of soil salinity and ensure the long-term sustainability of irrigated agriculture it is important of establish a rapid monitoring systems that could track the changes in soil salinity (Metternicht, 1998).

Over the last decades remote sensing has been favoured over conventional methods to map soil salinity. Most frequently used approaches for detecting soil salinity using remote sensing data employ spectral indices. Soil salinity indices are the most prominently used (Al-Khaier, 2003 and Abbas and Khan, 2007), while other studies integrated these indices with vegetation indices (VIs) (Iqbal, 2011 and Azabdaftari and Sunar, 2016).

Recently novel methods such as neural networks have been introduced for processing remote sensing data (Morgan et al., 2016 and Lakhankar et al., 2009). An Artificial Neural Network (ANN) is defined as a mathematical model that tries to simulate the structure and functionalities of biological neural networks (Krenker et al., 2011). These techniques are data driven where a functional relationship is determined based on some training datasets (Lakhankar et al., 2009).

This paper presents a novel approach utilizing neural networks in processing Sentinel-2 data for detecting soil salinity in a selected area in Egypt.

MATERIALS AND METHODS
Study area and field work: The study area is located about 84 km to the north-west of Cairo and covers an area of about 1.5 Km². The area is irrigated via drip irrigation system using groundwater. The study area is considered a relatively newly reclaimed area cultivated with olive trees (Morgan et al., 2017). Thirty-two soil samples were collected at approximately 200x200 m grid Fig 1.

Laboratory analysis: The field work included collecting soil augering samples at depth 0 – 25 cm. The collected soil samples were air-dried, ground gently and sieved through 2mm sieve. They were analyzed for salinity measured as total dissolved solids (TDS) in extract 1:1 (soil : water), using Hanna Instruments HI 2550 Benchtop Meter according to Soil Survey Division Staff (1993). The soil TDS was expressed in ppm.

Sentinel-2 data: Sentinel-2 (S2) is a wide-swath, high-resolution, multi-spectral imaging earth observation mission developed by European Space Agency (ESA) based on the
technology and experience acquired from the SPOT and Landsat missions over the past decades. It consists of two identical satellites flying in the same sun-synchronous orbit but phased at 180° to each other at a mean altitude of 786 km. It is designed to give a high revisit frequency of 5 days at the Equator. The satellites are equipped with the state of the art multispectral imager instrument that offers high-resolution optical imagery. It provides a set of 13 spectral bands: four bands at 10m, six bands at 20m and three bands at 60m spatial resolution Table 1 (ESA, 2015 and Vajsová and Aastrand, 2015).

Data pre-processing: A geographic database was designed containing the location and soil salinity data of the observation points. The data were projected into Universal Transverse Mercator (UTM) and World Geodetic System 1984 (WGS 84) datum-zone 36 N.

Sentinel-2 data acquired on 16/8/2015 was downloaded in Level-1C top of atmosphere (TOA) reflectance data. These data are radiometrically and geometrically corrected and in Geographic Markup Language JPEG2000 (GMLJP2) format. They were imported into QGIS 2.18.0 software which supports the GMLJP2 format and exported as Geotiff into ILWIS software for processing. The visible bands (B2-B4), near infrared band (B8) and shortwave infrared band (B11-B12) were used in this study. The 20 m bands (B11 and B12) were resized into 10m pixel size as the visual bands. Thereafter, these bands were stacked into one image then a subset of the study area was produced.

Data Processing: In our study we evaluated three different approaches in processing Sentinel-2 data. These approaches included the use of the reflectance data of selected six S2 bands, the principal component analyses of these selected bands and spectral indices driven from S2 bands.

The PCA is a powerful statistical technique which aims at reducing variance and dimensionality of the data set (Gupta et al., 2013). Because some of the original bands may be highly correlated such bands could be combined into new, less correlated eigen images using PCA (Munyati, 2004). The first two or three components will carry most of the real information of the original data set (Santiesteban, 2003).

Three spectral indices including two vegetation indices (VI) and a salinity index (SI) were evaluated for predicting of soil salinity. VIs are dimensionless, radiometric measures of vegetation (Huete et al., 1999). The spectral ranges or bands used in VIs are selected depending on the spectral properties of plants (Mróz and Sobieraj, 2004). Studies using these indices in soil salinity detection depends on that the soil salinity affects vegetation density and crop growth. The VI used in this study included the normalized differential vegetation index (NDVI) and the moisture stress index (MSI). The NDVI, the most well-known and often used vegetation index, was introduced by Rouse et al. (1974) [Eq. 1]. NDVI enables to eliminate topographic effects and variations in the sun illumination angle, as well as other atmospheric elements such as haze (Mróz and Sobieraj, 2004).

\[
\text{NDVI} = \frac{P_{\text{NIR}} - P_{R}}{P_{\text{NIR}} + P_{R}} \quad \text{Eq. 1}
\]

Where \( P_{\text{NIR}} \) and \( P_{R} \) are the reflectance in the near infrared and red band respectively.

Salinity inhibits plant access to soil water by increasing the osmotic strength of the soil solution. As the soil dries, the soil solution becomes increasingly concentrated, further limiting plant access to soil water.
(Sheldon et al., 2004) and consequently affecting its water content. The MSI, suggested by Rock et al. (1986), is one of the vegetation indices formulated to estimate vegetation water content [Eq. 2]. The index is a simple ratio between reflectance in the shortwave infrared which is sensitive to water content (decreases with increase water content of leaves) to that in the near infrared which is sensitive to changes in leaf internal structure and nearly unaffected by changing water content (Danson and Bowyer, 2004). Higher values of this index indicate greater water stress and less water content.

\[ \text{MSI} = \frac{P_{\text{SWIR}}}{P_{\text{NIR}}} \]

Where \( P_{\text{SWIR}} \) and \( P_{\text{NIR}} \) are the reflectance in the shortwave infrared and near infrared respectively.

This study also used a salinity index developed by Abbas and Khan (2007)[Eq. 3]. The index is a simple blue to red band ratio, depending on that the blue band has significant higher reflectance with salt increase.

\[ \text{SI} = \frac{P_{\text{B}}}{P_{\text{R}}} \]

Where \( P_{\text{B}} \) and \( P_{\text{R}} \) are the reflectance in the blue and red band respectively.

**Artificial Neural Networks (ANN):** An Artificial Neural Network (ANN) consists of a network of artificial neurons which imitate the biological neurons of human brain. These neurons are logically arranged in layers: an input layer, an output layer and one or more hidden layers. The input layer is the mean by which data are presented to the network. The output layer holds the response of the network to the input. The hidden layers enable these networks to represent and compute complicated associations between the inputs and outputs. The neurons interact with each other via weighted connections and each neuron is connected to all the neurons in the next layer (Erzin et al., 2010).

In this study, we used a feed-forward neural network where the information flow is unidirectional, i.e. information flow from input to output in one direction with no back-loops (Krenker et al., 2011). When designing the neural network 70% of the samples were used for training and 15% for testing and 15% for validation. The appropriate neural network was assessed based on the coefficient of determination (R\(^2\)) between the actual and predicted soil salinity. The structure of the neural network was manually alternated using a trial-and-error process until the highest R\(^2\) was achieved.

The ILWIS software was used to extract the image values (from the S2 bands, PCAs and VIs) corresponding to the field observation points. The data was initially exported into Excel to apply the correlation analysis, the Simple Linear Regression (SLR) and the Multiple Linear Regression (MLR) analysis as a preliminary step before applying the neural networks. Thereafter, the data was exported into MATLAB as spreadsheet file for designing the neural networks using the Neural Network Toolbox. The neural network with the best performance expressed as R\(^2\) was used for processing the images that are imported in tiff format into MATLAB. The results were exported in geotiff format into ILWIS software for analysis and visualization.

**RESULTS AND DISCUSSION**

**S2 Reflectance data:** The SLR analysis was applied to each band reflection to predict soil salinity. This analysis revealed that R\(^2\) between the predicted and actual soil salinity did not exceed 0.16 for all visible and near infrared bands and slightly increased to 0.31 and 0.35 for the shortwave bands (band 11 and 12 respectively). Furthermore, applying the MLR analysis for the six bands together resulted R\(^2\) of 0.47 and decreased reaching R\(^2\) of only 0.37 when applying the MLR for the two shortwave infrared bands.

MATLAB was used for designing and an ANN and the best designed network had a six nodes input layer (representing the S2 bands) and a hidden layer with 15 nodes and an output layer with one node representing soil salinity had its performance reached R\(^2\) of 0.77 Fig 2a. Contrary to the MLR results, the performance of the ANN could be enhanced using the SWIR bands only as inputs and reaching R\(^2\) of 0.87 between the predicted and actual soil salinity Fig 2b. The designed network included a two node input layer representing the shortwave infrared bands and a hidden layer with 10 nodes and one output layer with one node representing soil salinity.

**Principal component analyses (PCA):** Utilizing ILWIS the PCA was applied to the six studied S2 bands. Table 2 shows

![Fig 2: Performance of the designed ANN for a) the six S2 spectral bands b) the SWIR bands.](image)
the correlation matrices between the studied S2 six bands. The data indicated that the studied bands were highly correlated with R² more than 0.84 between each pair of bands which can cause a lot of redundancy and consequently validating the importance of applying the PCA.

The PCA results indicated that the first three principal components contained about 99.6 of the data variability with the first PCA containing most of the variability (93.3 %), followed by the second PCA (3.3 %) and lastly the third PCA (3.0%) and thus only these three PCAs were used in this study. The data also revealed that the first PCA was somewhat dominated by the SWIR bands (B11-B12), while the second and third PCAs were dominated by the near infrared Table 3. Therefore, the first PCA could represent the water content status, while the second and third PCAs represent the health condition as healthy vegetation reflects highly in near infrared (Munyati, 2004).

The correlation analysis between the three PCAs and soil salinity revealed that there is a negative correlation (r = -0.60) between the second PCA which is dominated by the NIR band and the soil salinity. This could be rendered to the vegetation health condition which is negatively correlated to the increased soil salinity. The first and third PCAs had coefficient of correlation of 0.45 and 0.30, respectively. The MLR analysis was applied to the first three PCA bands and resulted R² of 0.51. These results indicated improvement in the soil salinity predictions compared to the SLR and MLR analyses of the S2 six band reflectance data.

When processing the first three PCAs using ANN, a neural network was designed with one input layer consisting of three nodes representing the three PCAs and a hidden layer with 10 nodes and one output layer with one node representing soil salinity. The network performance reached only R² of 0.49 Fig 3a.

A second network was designed using the first and second PCAs only. The first PCA represent most of the variability in the six bands while the second had the highest correlation with the soil salinity. This network has the first two PCAs as nodes for the input layer and a hidden layer with ten nodes and one node output layer and its performance reached R² of 0.77 between the actual and predicted soil salinity Fig 3b. Accordingly, neither the first three nor two PCAs neural networks could achieve the high performance attained by the shortwave infrared bands in predicting soil salinity.

Spectral indices: The soil salinity was negatively correlated with NDVI and positively correlated with MSI (r = -0.57 and 0.77 for NDVI and MSI, respectively). The vegetation density and consequently NDVI decrees with soil salinity increment. On the other hand, the increase in MSI indicates a greater water stress which is associated with increasing soil salinity. The SI had the lowest coefficient of correlation (r = 0.34).

The SLR analysis was applied using each single index for predicting soil salinity. Accordingly, R² between the predicted and actual soil salinity was 0.30, 0.41 and 0.24 for the NDVI, MSI, SI, respectively. Moreover, applying the MLR analysis for the three indices resulted R² of 0.41. The same results were achieved when using only the NDVI and MSI.

**Table 2: The correlation matrix between the S2 bands.**

| Band No. | B2   | B3   | B4   | B8   | B11  | B12  |
|----------|------|------|------|------|------|------|
| B2       | 1    |      |      |      |      |      |
| B3       | 0.99 | 1    |      |      |      |      |
| B4       | 0.97 | 0.99 | 1    |      |      |      |
| B8       | 0.84 | 0.87 | 0.86 | 1    |      |      |
| B11      | 0.88 | 0.91 | 0.93 | 0.88 | 1    |      |
| B12      | 0.88 | 0.91 | 0.93 | 0.86 | 0.99 | 1    |

**Table 3: The eigenvectors of the calculated covariance matrix of the six S2 bands.**

| PCA Band | B2   | B3   | B4   | B8   | B11  | B12  |
|----------|------|------|------|------|------|------|
| PC 1     | 0.207| 0.298| 0.432| 0.388| 0.535| 0.494|
| PC 2     | 0.213| 0.267| 0.19 | 0.661|-0.417|-0.486|
| PC 3     | -0.323|-0.366|-0.553| 0.628| 0.231| 0.097|

**Table 4: Area and percentage of the different soil salinity classes.**

| Salinity class | Area (m²) | Area (%) |
|----------------|-----------|----------|
| <500           | 573600    | 40.06    |
| 500-1000       | 502400    | 35.09    |
| 1000-1500      | 311600    | 21.76    |
| >1500          | 44300     | 3.09     |
|                | 1431900   | 100.00   |

**Fig 3:** Performance of the designed ANN for a) the first three PCAs b) the first two PCAs.
Thereafter, an ANN was designed for the NDVI and had one input layer with one node representing the NDVI and a hidden layer with 5 nodes and one output layer with one node representing soil salinity. A similar neural network was designed for the MSI. The networks performances had $R^2$ of 0.69 and 0.64 for the NDVI Fig 4a and MSI Fig 4b, respectively. The third designed neural network included an input layer with two nodes representing the NDVI and MSI and a hidden layer with 10 nodes and one output layer with one node representing soil salinity and its performance reached $R^2 = 0.79$ Fig 4c. As concluded from the PCA networks, the performance could not surpass those obtained by the shortwave infrared data of S2.

Taking into account only the data which had the highest coefficient of correlation with the soil salinity including the first shortwave reflectance data (B11), the NDVI data and the second PCA data, a neural network was designed. This network had a three nodes input layer representing the S2 band 11, NDVI and the second PCA, a hidden layer with 15 nodes and one node output for soil salinity Fig 5. Utilizing this enhancement the performance of the network could reach $R^2 = 0.94$ Fig 6. and the resulting soil salinity map is shown in Fig 7.

Furthermore, a soil salinity classification map was produced using ILWIS slicing operation to accurately locate the problem areas with the highest soil salinity Fig 8. The results revealed that most of the studied area (about 75%) has soil salinity less than 1000 ppm (Table 4). The slightly
problematic salinity class with salinity more than 1500 ppm covered only an area of about 3.1% of the studies area and all of this class was located at the south western part of the study area comparable to the findings of Morgan et al. (2017).

CONCLUSION

Upon these results it was concluded that the artificial neural network offers much higher potentiality in predicting the soil salinity using remote sensing data compared to traditional methods such as SLR and MLR. Furthermore, it was observed that the vegetation density associated with the NDVI and vegetation health associated with the second PCA represented an important part when designing the neural network for predicting the soil salinity. Furthermore, the short wave infrared data also had an integral part in soil salinity prediction and the reflectance data was superior to the spectral indices such as MSI. It is also recommended that other salinity indices should be tested for inclusion in the designing of the network used for salinity prediction.

REFERENCES

Abbas, A. and Khan, S. (2007). Using remote sensing techniques for appraisal of irrigated soil salinity. Proceedings of the International Congress on Modelling and Simulation (MODSIM ’07), Oxley L. and Kalusiri D., (Eds.). Modelling and Simulation Society of Australia and New Zealand, Brighton, pp. 2632–2638.

Al-Khaier, F. (2003). Soil salinity detection using satellite remotesensing. Master of Science in Geo-information Science and Earth Observation, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Azabdaftari, A. and Sunar, F. (2016). Soil Salinity Mapping using Multitemporal LANDSAT Data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B7, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic.

Danson, F.M. and Bowyer, P. (2004). Estimating live fuel moisture content from remotely sensed reflectance. Remote Sensing of Environment, 92: 309-321.

Erzin, Y., Rao, B.H., Patel, A., Gumaste, S.D. and Singh, D.N. (2010) . Artificial Neural Network Models for Predicting Electrical Resistivity of Soils from Their Thermal Resistivity. International Journal of Thermal Sciences, 49:118-130.

ESA(2015). Sentinel-2 User Handbook. ESA Standard Document, Issue 1 Rev 2.

Ghassemi, F., Jakeman,A.J., and Nix, H.A. (1995). Salinisation of land and water resources: human causes, extent, management and case studies. Sydney, N.S.W.: NSW University Press.

Gupta, R.P., Tiwari, R.K., Saini, V., Srivastava, N. (2013). A Simplified Approach for Interpreting Principal Component. Advances in Remote Sensing Images, 2:111-119. http://dx.doi.org/10.4236/ars.2013.22015.

Huete, A., Justice, C., van Leeuwen, W. (1999). MODIS vegetation index (MOD13) algorithm theoretical basis document, Ver. 3.

Iqbal, F. (2011). Detection of salt affected soil in rice-wheat area using satellite image African. Journal of Agricultural Research, 6(21): 4973-4982. DOI: 10.5897/AJAR11.634.

Krenker, A., Bester, J., Kos, A. (2011). Introduction to the Artificial Neural Networks, Artificial Neural Networks - Methodological Advances and Biomedical Applications, (Kenji Suzuki (Ed.)) Available from: http://www.intechopen.com/books/artificial-neural-networks-methodological-advances-and-biomedical-applications/introduction-to-the-artificial-neural-networks.

Lakhankar, T., Ghedira, H., Temimi, M., Sengupta, M., Khanbilvardi, R., Blake, R.(2009). Non-parametric Methods for Soil Moisture Retrieval from Satellite Remote Sensing Data, Remote Sensing, 1:3-21. doi:10.3390/rs1010003

Metternicht, G.I. (1998). Fuzzy classification of JERS-1 SAR data: an evaluation of its performance for soil salinity mapping. Ecological Modelling, 111:61–74

Metternicht, G.I., Zinck, J.A. (2003). Remote sensing of soil salinity: potentials and constraints, Review article. Remote Sensing of Environment, 85:1-20.

Morgan, R.S., Abd El-Hady, M., Rahim, I.S., Silva J, Ribeiro S. (2017). Evaluation of various interpolation techniques for estimation of selected soil properties. International Journal of GEOMATE, 13(38): 23-30.

Morgan, R.S., Abd El-Hady, M., Rahim, I.S., Silva, J., Berg, A. (2016). Integrating microwave and optical data for monitoring soil moisture. in Proc. ‘Living Planet Symposium 2016’, Prague, Czech Republic, 9–13 May 2016 (ESA SP-740, August 2016).

Mróz, M. and Sobieraj, A. (2004). Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series, and their suitability for land cover and agricultural crop identification. Technical Sciences, 740-66.

Munyati, C. (2004). Use of Principal Component Analysis (PCA) of Remote Sensing Images in Wetland Change Detection on the Kafue Flats, Zambia. Geocarto International, 19(3):11-22.

Rock, B.N., Vogelmann J.E., Williams, D.L., Vogelmann, A.F., Hoshizaki, T. (1986). Remote detection of forest damage. Bioscience, 36: 439-445.

Rouse, J.W., Haas, R.W., Schell, J.A., Deering, D.W., Harlan, J.C. (1974). Monitoring the vernal advancement and retrogradation (Greenwave effect) of natural vegetation. NASA/GSFC Type III Final Report, Greenbelt, MD, USA.

Santiesteban, I.A.(2003). Multi-sensor image fusion applied to morphodynamics of coastal environments in Belize. Master of Science Thesis in Geo-information Science and Earth Observation, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Sheldon, A., Menzies, N.W., Bing, S. H., Dalal R. (2004). The effect of salinity on plant available water. SuperSoil 2004: 3rd Australian New Zealand Soils Conference, 5 – 9 December 2004, University of Sydney, Australia. http://www.regional.org.au/au/asssi/SuperSoil2004Conference/Program/Day2/p38.pdf

Vajsová, B., Aastrand, P.J. (2015). New sensors benchmark report on Sentinel -2A. EUR 27674 EN. doi: 10.2788/544302.