Soil salinity mapping and hydrological drought indices assessment in arid environments based on remote sensing techniques

Mohamed Elhag and Jarbou A. Bahrawi
Department of Hydrology and Water Resources Management, Faculty of Meteorology, Environment & Arid Land Agriculture, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Correspondence to: Mohamed Elhag (melhag@kau.edu.sa)

Received: 19 November 2016 – Discussion started: 8 December 2016
Revised: 14 February 2017 – Accepted: 1 March 2017 – Published: 15 March 2017

Abstract. Vegetation indices are mostly described as crop water derivatives. The normalized difference vegetation index (NDVI) is one of the oldest remote sensing applications that is widely used to evaluate crop vigor directly and crop water relationships indirectly. Recently, several NDVI derivatives were exclusively used to assess crop water relationships. Four hydrological drought indices are examined in the current research study. The water supply vegetation index (WSVI), the soil-adjusted vegetation index (SAVI), the moisture stress index (MSI) and the normalized difference infrared index (NDII) are implemented in the current study as an indirect tool to map the effect of different soil salinity levels on crop water stress in arid environments. In arid environments, such as Saudi Arabia, water resources are under pressure, especially groundwater levels. Groundwater wells are rapidly depleted due to the heavy abstraction of the reserved water. Heavy abstractions of groundwater, which exceed crop water requirements in most of the cases, are powered by high evaporation rates in the designated study area because of the long days of extremely hot summer. Landsat 8 OLI data were extensively used in the current research to obtain several vegetation indices in response to soil salinity in Wadi ad-Dawasir. Principal component analyses (PCA) and artificial neural network (ANN) analyses are complementary tools used to understand the regression pattern of the hydrological drought indices in the designated study area.

1 Introduction

Remote sensing data are considered to be a convenient source to perform several vegetation indices in either simple or complicated band ratio combinations. Satellite images offer a large amount of data that could be analyzed, processed and stored to better understand several vegetation indices based on the type of the satellite sensor used (Govaerts et al., 1999; Pinty et al., 2009). Hypothetical backgrounds have been implemented to improve and enhance the optimization of particular satellite sensors to support certain vegetation indices (Verstraete et al., 1996; Gobron et al., 2000; Psilovikos and Elhag, 2013).

Spectral vegetation indices are mathematical combinations of different spectral bands mostly in the visible and near-infrared regions of the electromagnetic spectrum. Vegetation activities can be measured comprehensively through semi-analytical methods of spectral band ratios that have been extensively used to detect not only seasonal variability of the vegetation cover but also local scale spatial variability (Broge and Mortensen, 2002; Xiao et al., 2002).

The generic principle of utilizing vegetation indices is to improve the interpretation of the spectral data reflected from a vegetation cover. Spectral reflectance variabilities tend to differentiate between different vegetation characteristics, based on crop water relationships and other surrounding features of soil components, and atmosphere, based on the maximization of vegetation characteristics over the surrounding environments (Moulin and Guerif, 1999; Boegh et al., 2002). Color, roughness and water content are the main soil components that affect soil spectral reflectance (Curran, 1983a, b; Bouman and Tuong, 2001).
Soil parameter variation tends to draw a line on a plenary scattergram. Nevertheless, this line, used as a reference point and known as a “soil line” in vegetation studies, involved both red and infrared spectral bands (Colombo et al., 2003; Elhag, 2014a, b). The utilization of vegetation indices has always been challenged by a major difficulty, which is the minimization of soil component interferences and sensitivity maximization of atmospheric variations (Qi et al., 1994; Leprieur et al., 2000). The atmospherically resistant vegetation index (ARVI), developed by Kaufman and Tanné (1992), and the global environmental monitoring index (GEMI), developed by Pinty and Verstraete (1992), are the less sensitive vegetation indices to atmospheric variation. Additionally, Qi et al. (1994) reported that the GEMI is soil noise sensitive. The higher noise sensitivity of GEMI has completely disabled the index and classified it as inadequate for arid regions.

Implementations of vegetation indices varied, from a local leaf scale to a continental vegetation scale. Moreover, certain indices tend to be site and/or species specific (Clevers, 1989; Elhag, 2014a), and they cannot be applied to different species or different leaf structures and canopy geometry (Xiao et al., 2002). The scholarly work of Kerr and Ostrovsky (2003), Pettorelli et al. (2005), Huete et al. (2008) and Elhag (2014b) reported that several vegetation indices were used to estimate different vegetation parameters extensively, including the leaf area index (LAI), the fractional vegetation cover (FC), the crop water stress index (CWSI), the drought severity index (DSI) and the water supply vegetation index (WSVI).

Soil salinization is a dynamic process that basically arises when an excess of irrigational water is frequently used in the drainage capacity of the fields (Warlow and Egbert, 2010). Implementations of remote sensing techniques in soil salinity mapping achieved comprehensive results on the regional scale (Montandon and Small, 2008). The brightness index (BI), the normalized difference salinity index (NDSI) and the salinity index (SI) are widely distinguished in soil salinity mapping in an arid environment (Douaoui et al., 2006; Jia paer et al., 2011). The current research aims to evaluate the suitability of different vegetation indices for a different level of remotely sensed soil salinity contrasting to crop water relationship in Wadi ad-Dawasir.

2 Materials and methods

2.1 Study area

The study area, the Wadi ad-Dawasir town, is located in the plateau of Najd at 44°43′ lat and 20°29′ long, about 300 km south of the capital city, Riyadh. The study area illustrated in Fig. 1 is comprised of gravelly tableland disconnected by insignificant sandy oases and isolated mountain bundles. Across the Arabian Peninsula, as a whole, the tableland slopes toward the east from an elevation of 1360 m in the west to 750 m at its easternmost limit. Wadi ad-Dawasir and Wadi al-Rummah, which are the most important patterns of the ancient riverbeds, remain in the study area. The Wadi ad-Dawasir and Najran regions are the major irrigation water abstractors from the Al-Wajid aquifer. Agriculture in the Wadi ad-Dawasir area consists of technically highly developed farm enterprises that operate with modern pivot irrigation systems. The size of center pivot ranges from 30 to 60 ha, with farms managing hundreds of them with the corresponding number of wells. The main crop grown in winter is wheat and occasionally potatoes, tomatoes or melons. All-year fodder consists of alfalfa, which is cut up to 10 times a year for food. Typical summer crops for fodder are sorghum and Rhodes grass, which is perennial but dormant in winter. The shallow alluvial aquifers could not sustain the high groundwater abstraction rates for a long time and groundwater level declined dramatically in most areas. Meteorological features of the area are speckled. Five elements of meteorology are constantly recorded through a fixed weather station located within the study area. Temperature varies from a minimum of 6°C to a maximum of 43°C. Relative humidity is mostly stable at 24%. Solar radiation of average sunrise duration is generally 11 h day⁻¹. Average wind speed is closer to 13 km h⁻¹ and may reach up to 46 km h⁻¹ in thunderstorm incidents. Finally, mean annual rainfall is about 37.6 mm (Al-Zahrani and Baig, 2011).

2.2 Methodological framework

The current research work is based on assessing a regression correlation between different vegetation indices and their spatial corresponding soil salinity values conducted from satellite images. The principal component analysis (PCA) was used to envisage the impacts of soil salinity on the current vegetation.

Figure 1. Location of the study area (Elhag, 2016).
2.3 Estimation of vegetation indices

2.3.1 Water supply vegetation index (WSVI)

The water supply vegetation index is calculated by

\[ WSVI = \frac{\text{NDVI}}{T_s}, \]

where \( T_s \) is the estimated brightness temperature channel or related remote sensing imagery, and NDVI is the normalized difference vegetation index. The smaller this index is, the more severe the drought is.

2.3.2 Soil-adjusted vegetation index (SAVI)

The soil-adjusted vegetation index is calculated by

\[ \text{SAVI} = \frac{(\text{NIR} - \text{R})}{(\text{NIR} + \text{R}) \cdot (1 + L)}, \]

where NIR is the near-infrared band, R is the red band and \( L \) is the soil brightness correction factor, commonly \( L = 0.5 \) (Huete, 1988).

2.3.3 Moisture stress index (MSI)

The moisture stress index is calculated by

\[ \text{MSI} = \frac{\text{SWIR}_1}{\text{NIR}}, \]

where SWIR\(_1\) is the short-wave infrared band 1.

2.3.4 Normalized difference infrared index (NDII)

The normalized difference infrared index is calculated by

\[ \text{NDII} = \frac{(\text{NIR} - \text{SWIR}_1)}{(\text{NIR} + \text{SWIR}_1)}. \]

2.4 Estimation of soil salinity index

Soil salinity indices are principally adjusted to detect salt mineral in soils based on the different responses of salty soils to various spectral bands. The following equation to map soil salinity was used following Elhag (2016).

\[ \text{SI} = (\text{G} \times \text{R})/\text{B}, \]

where B is the blue band, G is the green band and R is the red band.

2.5 Regression analyses

The purpose of the regression analyses is to envisage the regression potentials between the soil salinity index from one side and the rest of the hydrological drought indices from the other side. The principal component analyses and artificial neural network (ANN) analyses were the implemented approaches. The PCA is used to transform a set of likely correlated with unlikely correlated variables. The principal components number is less than or equal to the variables’ original number. Following Lorenz (1956), the PCA fundamental equations are described as follows.

First, vector \( W_{(1)} \) has to be calculated as follows:

\[ w_{(1)} = \arg \max_{\|w\|=1} \left\{ \sum_{i} (t_i)^2 \right\} = \arg \max_{\|w\|=1} \left\{ \sum_{i} (x_i \cdot w)^2 \right\}. \]

(6)

The matrix form of the above equation gives the following:

\[ w_{(1)} = \arg \max_{\|w\|=1} \left\{ \|Xw\|^2 \right\} = \arg \max_{\|w\|=1} \left\{ w^T X^T X w \right\}. \]

(7)

\( W_{(1)} \) has to be calculated as follows:

\[ w_{(1)} = \arg \max \left\{ w^T X^T X w \right\}. \]

(8)

The resulting \( w_{(1)} \) suggests that the first component of a data vector, \( x_{(i)} \), can then be expressed as a score of \( t_{(i)} = x_{(i)} \cdot w_{(1)} \) in the transformed co-ordinates or as the corresponding vector in the original variables, \( \{x_{(i)} \cdot w_{(1)}\}w_{(1)} \).

The neural network regression model is written as

\[ Y = \alpha + \sum h \phi_h \left( \alpha_h + \sum i w_{ih}X_i \right), \]

where \( Y = E(Y|X) \). This neural network model has one hidden layer, but it is possible to have additional hidden layers.

The \( \phi(z) \) function used is hyperbolic tangent activation function. It is used for logistic activation for the hidden layers.

\[ \phi(z) = \tanh h(z) = \frac{1 - e^{-2z}}{1 + e^{-2z}}. \]

(10)

Significantly, the final output should be stochastically linear, with no prediction limitations being between 0 and 1. A simple diagram of a skip-layer neural network is illustrated in Fig. 2. The equation for the skip-layer neural network for regression is shown below.

\[ Y = \alpha + \sum_{i=1}^{p} \beta_i X_i + \sum h w_{ih} \phi_h \left( \alpha_h + \sum i w_{ih}X_i \right). \]

(11)

It should be clear that these models are highly parameterized and, thus, will tend to overfit the training data. Cross-validation is, therefore, critical to making sure that the predictive performance of the neural network model is adequate.

The determination of the adequate performance of the ANN model is a must. Five different criteria are used: the Pearson correlation coefficient (\( R \)), the root mean square error (RMSE), the mean absolute deviation (MAD), the negative log likelihood and the error sum of squares (SSE). Basically, RMSE is the examined parameter for comparability.
Figure 2. Artificial neural network scheme with one hidden layer and three nodes.

reasons. RMSE can be computed as follows:

\[
\text{RMSE} = \sqrt{\frac{1}{T_0} \sum_{t=1}^{T_0} (\hat{y}_t - y_t)^2},
\]

where \( t \) is the time index and \( \hat{y}_t \) and \( y_t \) are the simulated and measured values. Principally, the higher value of \( R \) and smaller values of RMSE ensure the better performance of the model.

3 Results and discussion

The realization of the hydrological drought indices was exercised after a comprehensive remote sensing data correction. Basically, atmospheric correction and spatial enhancement were practiced utilizing Landsat 8 OLI data acquired over the designated study area. The four hydrological drought indices were shown in Figs. 3 to 6. Stochastic algorithms of WSVI and SAVI mapping (Figs. 3 and 4) showed spatial coherence with higher drought indices’ values within the agricultural area rather than the surrounding area (Ceccato et al., 2001; Daughtry et al., 2004).

On the contrary, MSI functioned as a deterministic drought index, it was nearly unaffected by changing water content. Conducted results showed two classes of stresses: stressed and no stress. The no stress class was located within the agricultural area, and the stressed area was represented along the agricultural peripheral areas (Fig. 5), where higher values indicate greater water stress and less water content. This could be explained rationally by the presence of irrigational sprinkles (Hunt Jr. and Rock, 1989; Ceccato et al., 2001). NDII is also a stochastic algorithm and was used in the current research due to the higher sensitivity of the infrared band to detect changes in water content of plant canopies (Hardisky et al., 1983). The spatial distribution of NDII (Fig. 6) was mapped accordingly with WSVI and SAVI indices, in which higher NDII values meant higher water content (Jackson et al., 2004). There are several algorithms to map soil salinity based on utilization of different remote sensing data and different ecological systems. An adequate NDSI algorithm was carried out according to Elhag (2016) findings in arid ecosystems. In Fig. 7, NDSI showed spatial variation, especially within the new agricultural expansion at the southwestern part of the study area. The sprinkle movement drove the salt accumulation to be located at the peripherally of the agricultural areas (Lunetta et al., 2006; Konukcu et al., 2006).

Further statistical analyses were carried out to construe the correspondences between salted soils and different horological drought indices. The regression analysis demonstrated in Fig. 8 showed that salinity increases with lower WSVI and SAVI (Fig. 8a, b), which is explained by the salt accumulation in soils in parts per million (ppm). Under salinity stress conditions, there is not enough available water in soils for proper vegetation growth (Lunetta et al., 2006; Yang et al., 2011).

Generally, MSI values (Fig. 8c) are high in the study area because of the excess irrigation regime adopted to overcome the high evaporation rates (Elhag and Bahrawi, 2014; Elhag, 2016). Excess irrigation regimes in poor-drain soils lead to waterlogging problems and salts accumulation (Elhag, 2016).

Due to NDII’s higher sensitivity to water, NDII values increase with higher NDSI values. Salts accumulation caused by excessive irrigation is the driving force behind the proportional increment of NDII values in conjunction with NDSI values as demonstrated in Fig. 8d (Jackson et al., 2004; Shi et al., 2015).

Figure 9 demonstrated the principal component analysis along with the Factor Analysis. Moreover, eigenvalue de-
composition was also demonstrated. WSVI and SAVI were grouped together. Additionally, NDII and MSI were individually plotted against the former indices.

Similar results conducted from the scatterplot matrix and the accompanying correlation matrix are shown in Fig. 10 and Table 1. A high correlation is distinguished between WSVI and SAVI, while a negative correlation is noted between WSVI and SAVI from one side and MSI and NDII from the other side.

In Table 2, NDSI regression analysis shows that NDII is the proper fit based on different regression parameters (Rodgers and Nicewander, 1988). The spearman’s correlation demonstrated in Table 3 supports the PCA results. Hydrological drought indices were classified into two categories: MSI and NDII in one category and WSVI and SAVI in the other one. The elements of each category are positively correlated. MSI and NDII were significantly correlated, and WSVI and SAVI were highly correlated. Moreover, any other
Regression analyses of NDSI (ppm) against hydrological drought indices.

Figure 8.

The ANN analysis was carried out under one hidden layer, three nodes, and hyperbolic tangent activation function conditions. These conditions were carefully exercised to prevent the algorithm from overfitting; the ANN analysis is demonstrated in Table 4. NDII expressed the highest RMSE, which indicates that NDSI and NDII are statistically the best fit (Jiang, 2013). SAVI is at the second best fit, followed by combinations of the four hydrological drought indices were not correlated.

The ANN analysis was carried out under one hidden layer, three nodes, and hyperbolic tangent activation function conditions. These conditions were carefully exercised to prevent the algorithm from overfitting; the ANN analysis is demonstrated in Table 4. NDII expressed the highest RMSE, which indicates that NDSI and NDII are statistically the best fit (Jiang, 2013). SAVI is at the second best fit, followed by

Table 1. Correlation matrix.

<table>
<thead>
<tr>
<th></th>
<th>NDII</th>
<th>MSI</th>
<th>SAVI</th>
<th>WSVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDII</td>
<td>1</td>
<td>0.7182080406</td>
<td>−0.708975719</td>
<td>−0.703572559</td>
</tr>
<tr>
<td>MSI</td>
<td>1</td>
<td>−0.888156103</td>
<td>0.9977255509</td>
<td>0.0022</td>
</tr>
<tr>
<td>SAVI</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSVI</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Figure 9. Principal component analysis.
Table 2. Regression analysis.

<table>
<thead>
<tr>
<th></th>
<th>NDII</th>
<th>MSI</th>
<th>SAVI</th>
<th>WSVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root square error</td>
<td>0.798566127</td>
<td>0.254999657</td>
<td>0.246131379</td>
<td>0.243463225</td>
</tr>
<tr>
<td>Root square error adjusted</td>
<td>0.797205088</td>
<td>0.249965871</td>
<td>0.241037672</td>
<td>0.23835149</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>31.88199207</td>
<td>0.384262574</td>
<td>0.202130562</td>
<td>0.000447112</td>
</tr>
<tr>
<td>Mean of response</td>
<td>124.5466667</td>
<td>0.733333333</td>
<td>0.28631262</td>
<td>0.000611978</td>
</tr>
<tr>
<td>Observations (Sum wgts)</td>
<td>150</td>
<td>150</td>
<td>150</td>
<td>150</td>
</tr>
</tbody>
</table>

WSVI. MSI failed to fit NDSI values comprehensively, like the former hydrological drought indices (Jones and Marshall, 1992; Jiapaer et al., 2011).

4 Conclusions

The findings of the current research emphasize the importance of the hydrological drought indices to envisage the adverse effects of salts accumulation in poorly drained soils similar to the study area under investigation. The soils of Wadi ad-Dawasir are poorly drained and still under heavy pressure of heavy irrigation schemes to overcome the high evaporation rates. Therefore, the implemented irrigation schemes should be adjusted for better natural resources management. Remote Sensing techniques were satisfactorily implemented and interpreted in terms of soil salinity mapping in consort with hydrological drought indices. The normalized difference infrared index was statistically proven to be the profound normalized difference salinity index, followed by soil-adjusted vegetation index and water supply vegetation index, respectively. The principal component analyses and artificial neural network analyses are complementary tools used to understand the regression patterns of the hydrological drought indices in the designated study area. Further work needs to be considered towards the restrictiveness of the drastic effect of salts accumulation within the study area.

Data availability. The data used in this paper are free. The open access online data are available at the earth explorer website.
Table 3. Spearman’s correlation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>By variable</th>
<th>Correlation</th>
<th>Count</th>
<th>Lower 95 %</th>
<th>Upper 95 %</th>
<th>Significance probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSI</td>
<td>NDII</td>
<td>0.7182</td>
<td>150</td>
<td>0.6305</td>
<td>0.7878</td>
<td>*</td>
</tr>
<tr>
<td>SAVI</td>
<td>NDII</td>
<td>−0.7090</td>
<td>150</td>
<td>−0.7805</td>
<td>−0.6191</td>
<td>NS</td>
</tr>
<tr>
<td>SAVI</td>
<td>MSI</td>
<td>−0.8882</td>
<td>150</td>
<td>−0.9178</td>
<td>−0.8487</td>
<td>NS</td>
</tr>
<tr>
<td>WSVI</td>
<td>NDII</td>
<td>−0.7036</td>
<td>150</td>
<td>−0.7763</td>
<td>−0.6124</td>
<td>NS</td>
</tr>
<tr>
<td>WSVI</td>
<td>MSI</td>
<td>−0.8825</td>
<td>150</td>
<td>−0.9136</td>
<td>−0.8412</td>
<td>NS</td>
</tr>
<tr>
<td>WSVI</td>
<td>SAVI</td>
<td>0.9977</td>
<td>150</td>
<td>0.9969</td>
<td>0.9984</td>
<td>**</td>
</tr>
</tbody>
</table>

Note that * is significant, ** is highly significant and NS is non-significant.

Table 4. Neural network analysis.

<table>
<thead>
<tr>
<th>Training measures</th>
<th>Validation measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDII</td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.7574526</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0999530</td>
</tr>
<tr>
<td>Mean Abs Dev</td>
<td>0.0571881</td>
</tr>
<tr>
<td>-LogLikelihood</td>
<td>−88.411680</td>
</tr>
<tr>
<td>SSE</td>
<td>0.9990600</td>
</tr>
<tr>
<td>Sum Freq</td>
<td>100</td>
</tr>
<tr>
<td>MSI</td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.3032101</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.2388872</td>
</tr>
<tr>
<td>Mean Abs Dev</td>
<td>0.1203075</td>
</tr>
<tr>
<td>-LogLikelihood</td>
<td>−1.2825260</td>
</tr>
<tr>
<td>SSE</td>
<td>5.7067096</td>
</tr>
<tr>
<td>Sum Freq</td>
<td>100</td>
</tr>
<tr>
<td>SAVI</td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.7565419</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1499295</td>
</tr>
<tr>
<td>Mean Abs Dev</td>
<td>0.0857822</td>
</tr>
<tr>
<td>-LogLikelihood</td>
<td>−47.865170</td>
</tr>
<tr>
<td>SSE</td>
<td>2.2478847</td>
</tr>
<tr>
<td>Sum Freq</td>
<td>100</td>
</tr>
<tr>
<td>WSVI</td>
<td></td>
</tr>
<tr>
<td>RSquare</td>
<td>0.7533827</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0003280</td>
</tr>
<tr>
<td>Mean Abs Dev</td>
<td>0.0001876</td>
</tr>
<tr>
<td>-LogLikelihood</td>
<td>−660.35100</td>
</tr>
<tr>
<td>SSE</td>
<td>1.08E-05</td>
</tr>
<tr>
<td>Sum Freq</td>
<td>100</td>
</tr>
</tbody>
</table>

RSquare = root square error and RMSE = root mean square error.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. 155-36-1437-D. The authors, therefore, gratefully acknowledge the DSR’s technical and financial support.

Edited by: L. Eppelbaum
Reviewed by: S. Boteva and N. Yilmaz

References


