Prediction of Soil Total Nitrogen Content Using Spectroradiometer and GIS in southern Iraq
Qassim A. Talib Alshujairy¹, Nooreldeen Showqi Ali²

¹Soil science and water resources dept. Al-Muthana University, Iraq
²Soil science and water resources dept. Baghdad University, Iraq

Abstract — In this study, soil samples were collected from two locations: Samawa and Rumetha in southern Iraq. The samples from each location were split into two datasets: calibration set and validation set. VNIR reflectance (350-2500 nm) and GIS-Kriging were used in combination with Partial Least Square (PLS) to predict total N. only two regions reported higher determination coefficient $R^2$ and lower Root Mean Square Error (RMSE) than the other wavelength regions. PLS calibration models yielded an $R^2$ of 0.96 and 0.97 for Rumetha and 0.87 and 0.94 for Samawa location in bands at 500-600 and 800-1000 nm, respectively. The potential of VNIR-based and GIS-Kriging models to predict new unknown soil samples were assessed by using validation datasets from both studied locations. The cross-validation of GIS-Kriging models were unsatisfactory predicted with an $Q^2$ of 0.28 between laboratory-measured and predicted total N values for Rumetha and 0.43 for Samawa location. While VNIR-based validation models achieved highly predictive power with an $R^2$ of 0.84 between laboratory-measured and predicted total N values for Rumetha and 0.85 for Samawa location. These results reveal extremely decreasing in model predictive ability when shifting from VNIR Spectroscopy method to GIS-Kriging.

Keywords — GIS, Kriging, Spectroradiometer, Total Nitrogen.

I. INTRODUCTION

Visible Near Infrared Radiation (VNIR) is a tool for molecular structure determination (Pasquini, 2003). It's concept based on the interactions between the investigated sample and the VNIR at the region of electromagnetic spectrum in range of 400-2500 nm (Canasveras et. al., 2012).

There are several factors contribute in the bandwidth of infrared absorptions (narrow or broadening). Collision between molecules and the limited of timelife of an excited state are the main sources of line broadening, where the less well describe energy associated with the shorter lifetime of transition to the excited state (Stuart, 2004). In spite of that the absorption intensity is highly related to the change in molecular dipoles, hence, the large change lead to a very strong absorption and inversely, a very weak absorption associated with a very small change in dipole (Günzler and Gremlich, 2002; Hollas, 1996).

Over the last 30 years there has been a growing interest in VNIR-Spectroscopy as a potential analytical technique for studying materials in many different fields such as Agriculture, food, textiles, polymers, wool, biomedical and pharmaceutical (Guerrero, et al., 2010; Niemoller and Behmer, 2008). VNIR-Spectroscopy has attracted much attention from soil scientist as a promising technique to determine a lot of soil properties from a single scan, reduce sample preparation, diminish the hazardous of using chemicals, and measurement can be taken both in laboratory and field with a few seconds (Ben-Dor and Banin, 1995; Chang et al., 2001; Viscarra Rossel et al., 2006). Soil spectrum is a characteristic shape that is caused by absorbing light to various degrees depending on the constituents of soil. Absorption of energy occurs mainly due to electrons transition between molecular orbits in visible region and vibrations in molecular bonds in near infrared region (Wetterlind et al., 2013). The frequencies at which light are absorbed match the difference between two energy levels and are displayed in %R (or transformed to absorbance) for analytical purposes (Miller, 2001). Much research has been widely using the diffuse reflectance ( VNIR-Spectroscopy) to determine soil texture, structure, soil organic matter (SOM), fertility, clay mineralogy, and microbial activity (Bowers and Hanks, 1965; Brown et al., 2006; Stenberg et al., 1995).
The soil spectra is characterized by few, broad, and overlapping absorption features due to several reasons such as sample chemical constituent, neighboring functional groups, hydrogen bonds (Miller, 2001). Diffuse reflectance of the soil is mostly influenced by Soil particle size, surface structure, and water films on soil surface (Twomey et al., 1986).

The assessment of soil fertility is usually routine work associated with soil nutrients level and organic matter content as main soil properties for either to Precision agriculture or to maintain soil from degradation (Gobeille et al., 2006). Such estimation needs to acquire data, since the soil sampling and laboratory analysis are costly and time consuming. At this point, spectroscopic techniques become a promising method to make rapid laboratory soil analysis as well as on-line field soil analysis with aid of a field portable instrument (Chang et al. 2001; Pirie et al. 2005; Brown et al. 2006; Nanni and Dematte, 2006). As described by Desbiez et al. (2004), soil fertility is function of soil properties therefore; it is comprehensive concept more than to measure directly.

Soil nitrogen and organic carbon have generally yielded most optimistic correlation coefficients ($R^2$ higher than 0.90) between the actual and estimated concentrations. Dalal and Henry (1986) studied the prediction of total soil nitrogen based upon selection of the combination of three wavelengths 1702, 1870 and 2052 nm. The correlation coefficient of prediction was $R^2=0.92$ and the standard errors for the prediction were much larger for soil samples with low contents of organic matter and total nitrogen. Wavelengths range1100- 2300 nm reported to be the best wavelengths to predict soil total N with correlation coefficient $R^2=0.94$ (Reeves and McCarty, 2001).

The sensitivity of N and C to infrared radiation is the main driving force that pushes researchers toward soil N and organic C analyses, which proved in very successful calibration coefficient $R^2$ in the region of 0.80 - 0.98 (Reeves et al., 2001; Cozzolino and Moron, 2006; Stevens et al., 2008). Several parameters have been evaluated in addition to total N and C, microbial C and N with reported $R^2$ in the range 0.60 to over 0.90 (Change et al., 2001; Ludwig et al., 2002). In regarding of C and N mineralization: the accumulated mineral N under aerobic and anaerobic incubation conditions have been studied with promising $R^2$ values between 0.70 and 0.80 (Palmberg and Nordgren, 1993). Others studies have reported less correlation values ($R^2<0.5$). The disappointing results have been attributed to the small subset samples, diverse locations and/or predicting one experiment with a calibration on the other (Reeves et al., 1999; Change et al., 2005). Terhoeveb-Urselmans et al., (2006) underlined the importance of the sufficient sample sets to predict biological properties with NIR spectroscopy. The less reliable prediction with very low correlations ($R^2<0.50$) is usually related to the low concentration or ephemeral and temporal changes in soil solution chemistry (Janik, et al., 1998).

Several factors influencing soil N content, soil particle-size fractions is one of these factors which has been studied by an indirect measurement of soil available N by using NIR spectroscopy (Barthès et al., 2008). The infrared absorption bands are often overlapped, some of interference is associated with soil carbonate contents which is considered as another factor influencing the accuracy of soil N content (Linker et al., 2005; Jahn et al., 2006). Borenstein et al. (2006) considered carbonate overlapping as the largest errors for nitrate determination in calcareous soils, where nitrate band is disturbed by absorbance band of carbonate.

### II. MATERIAL AND METHODS

The study area is Al-Muthana province, located in Southern Iraq. The second location is located in Al-Rumetha (45.252034-45.255029°E, 31.497139-31.498901°N) the northern border of the first study location Al-Samawa (45.255029-45.2525034°E, 31.497139-31.498901°N), about 25 Km from the center of Al-Samawa.

In the laboratory, the soil samples were air dried, hand cleaned to remove foreign particles and ground to pass through a 2 mm sieve and analyzed on organic matter content, pH, Ec, (Page et al.1982), CEC,texture(Black, et al.(1965) The soil total N content of each soil sample was determined by using conventional chemical analysis by Kjeldahl method (Bremner and Mulvaney1982).Table 1
TABLE 1
SOME SOIL PROPERTIES OF THE STUDY LOCATIONS.

<table>
<thead>
<tr>
<th>Location</th>
<th>Soil Classification</th>
<th>Parent Material</th>
<th>Electrical conductivity (Ece)</th>
<th>potential of Hydrogen (pH)</th>
<th>Carbonate</th>
<th>Organic Matter (O.M)</th>
<th>Cation Exchange Capacity (CEC)</th>
<th>Soluble Cations and Anions</th>
<th>Soil Particles</th>
<th>Texture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rumetha</td>
<td>8.7</td>
<td>7.5</td>
<td>215</td>
<td>12</td>
<td>24.4</td>
<td>Ca</td>
<td>Clay</td>
<td>Clay loam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Samawah</td>
<td>23</td>
<td>7.7</td>
<td>190</td>
<td>0.9</td>
<td>21.3</td>
<td>Mg</td>
<td>Silt</td>
<td>Sand Clay Loam</td>
</tr>
</tbody>
</table>

The reflectance spectra of air-dry soil samples were obtained in the laboratory by ASD spectrometer- FieldSpec® 3, with a spectral range of 350-2500 nm.

The near infrared reflectance spectra were transformed into absorbance spectra using the Log (1/Reflectance) function. According to Beer Lambert Law, absorbance is directly proportional to the concentration of studied properties. Prior to any model development the soil samples were randomly divided into a calibration set with 32 soil samples and a validation set with 8 soil samples. Spectral absorbance was correlated to total N concentration using statistical analysis software XLSTAT 2014 program to create predictor models. Models with the highest coefficient of determination (R²) and lowest root mean square error (RMSE) were plotted to identify the significant portion of wavelength for total N prediction.

III. RESULT AND DISCUSSION

Soil spectra result from overlapping and combination absorption features of soil components made VNIR Spectra complex to interpret. Partial least square regression (PLS) method was used to translate the spectral curve into values of predicted concentrations of total Nitrogen for both locations under study. The importance of different wavelength regions can be assessed using correlation plots. The performance statistical characteristics of Total Nitrogen for both Rumetha and Samawa locations are summarized in Table 2. In this study, among the entire spectral region of 350-2500 nm for soils of Rumetha and Samawa locations, only two regions reported higher determination coefficient R² and lower Root Mean Square Error (RMSE) than the other wavelength regions. Absorbance spectra were well correlated with total Nitrogen (TN) contents in bands at 500-600 nm with R² = 0.95 and 0.869 for Rumetha and Samawa locations, respectively and 800-1000 nm with R² = 0.97 for Rumetha and 0.938 for Samawa location. The best correlation values for TN were corresponded to the findings of other researchers who reported a wavelength region of 900-2400 nm as sensitive wavelengths to detect soil TN (Dalal and Henry, 1986; Martin et al., 2002; Shao and He, 2011; Dick et al., 2013, and Vohland et al., 2014). The wavelengths regions around 1400-2200 nm are dominated by NIR absorption of soil water content. For that reason, the models built from wavelength region 800-1000 nm are adopted to predict soil Total Nitrogen (TN) in this study according to Bullock et al., (2004) and Tiruneh, (2014).
### TABLE 2
STATISTICAL CHARACTERISTICS (CALIBRATION) FOR TOTAL NITROGEN IN SAMPLES OF RUMETHA AND SAMAWA LOCATIONS.

<table>
<thead>
<tr>
<th>Location</th>
<th>Wavelength nm</th>
<th>Observations</th>
<th>R²</th>
<th>RMSE</th>
<th>Std.deviation (Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rumetha</td>
<td>500-600</td>
<td>32</td>
<td>0.955</td>
<td>0.023</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>800-1000</td>
<td>32</td>
<td>0.979</td>
<td>0.016</td>
<td>0.019</td>
</tr>
<tr>
<td>Samawa</td>
<td>500-600</td>
<td>32</td>
<td>0.869</td>
<td>0.045</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>800-1000</td>
<td>32</td>
<td>0.938</td>
<td>0.039</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Statistically predictions of TN for both locations are shown in Fig 1. Coefficient of determination (R²) for Partial least square regression (PLS) analysis was used to relate the NIR absorbance data to the measured concentration of Total Nitrogen, which determined by the chemical analysis. Coefficient of determination (R²c) for calibration set of 32 soil samples for Rumetha and Samawa location obtained by PLS analysis were 0.979 for Rumetha and 0.938 for Samawa. However, In this study, the success of PLS models were proved by determination Coefficients (R²v) for the external validation set (the remaining 8 soil samples) of both locations, which were 0.84 and 0.85 for Rumetha and Samawa locations, respectively.

![Figure 1](image)

**FIGURE 1:** PLOTS OF MEASURED VERSUS PREDICTED TOTAL NITROGEN CONTENT IN THE CALIBRATION AND VALIDATION SETS OBTAINED VIA PLS MODELS AT 800-1000 nm FOR: (A) SOILS IN RUMETHA LOCATION AND (B) SOILS IN SAMAWA LOCATION.

The predicted and measured total Nitrogen contents for both locations shown the ability to obtain calibration models with reliable quantitative prediction within the range of samples from both locations. The high predictive power of the model was confirmed by the lower RMSE values 0.016 and 0.039 for Rumetha and Samawa locations, respectively. Results obtained from PLS analyses are quite promising to make approximate quantitative predictions for new unknown samples.

In terms of coefficients of determination (R²) and Root mean square error (RMSE), the calibration of total Nitrogen was fairly accurate. Based upon literature, the prediction is considered reliable with R² > 0.50 and accurate with R² > 0.80 (Alrajehy, 2002; Ludwig et al., 2002). The results are in trend with other studies (Bilgili et al., 2010; Kleinebecker et al., 2013; Bansod and Thakare, 2014) for achieving highly predictive power of NIR spectroscopy model for Total Nitrogen in soil.

#### 3.1 Nutrients Prediction by GIS-Kriging

Laboratory-measured soil data set of 32 point samples from samawa and rumetha were used for statistical modeling and prediction process. The spatial concentration of total N was generated with the use GIS-Kriging Ordinary Spherical technique for the (x, y) locations point samples from Samawa and Rumetha. The spatial characteristics of predicted total N concentrations showed increasing trend from center to southwest part of Samawa as shown in Fig. 2. In contrast, the spatial distribution of total N showed increasing trend from southwest to northeast part of Rumetha location (Fig. 3). The spatial
distribution trend of total N, status as shown in Fig. 2 and 3 could be significantly affected by slope gradient, land use, and soil type (Wang et al., 2008).

**Figure 2:** Spatial distribution map of total N in Samawa location

**Figure 3:** Spatial distribution map of total N in Rumetha location.

The quality of the predicted total N values by GIS-Kriging models was generally assessed by using validation dataset with 8 soil points for total N. The same validation dataset as was used in the validation of NIR model prediction was used to evaluate the Kriging models prediction accuracy. Partial Least Square Regression analysis was done between total N Kriging-predicted values and lab-measured values. The evaluation criteria for assessing the final prediction GIS-Kriging model are summarized in Table 3. The results as shown in Table 3, indicated that the coefficient of determination $R^2$ for both studied locations showed convergent $R^2$ values for total N with values of 0.60 and 0.61 for Samawa and Rumetha, respectively. Compared with $R^2$ results, other statistical measurements values were unsatisfactory predicted with low values of $Q^2$ and high values of standard deviation and Root Mean Square Error. Based upon these evaluation criteria, the smaller the RMSE and Standard deviation values, the better the prediction ability of the model (Hengl, 2007).

<table>
<thead>
<tr>
<th>Location</th>
<th>Target Variables</th>
<th>Observations</th>
<th>$R^2$</th>
<th>$Q^2$</th>
<th>RMSE</th>
<th>Std.deviation (Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rumetha</td>
<td>Total N</td>
<td>8</td>
<td>0.60</td>
<td>0.28</td>
<td>0.023</td>
<td>0.027</td>
</tr>
<tr>
<td>Samawa</td>
<td>Total N</td>
<td>8</td>
<td>0.61</td>
<td>0.43</td>
<td>0.059</td>
<td>0.068</td>
</tr>
</tbody>
</table>

**IV. CONCLUSION**

In aid of qualitative interpretation (PLS), the spectral bands that showed good correlations with total N have been identified. For total N, only two regions reported high $R^2$ and low RMSE: for bands at 500-600 nm with $R^2=0.95$ for Rumetha and 0.87 for Samawa and 800-1000 nm with $R^2 = 0.97$ for Rumetha and 0.94 for Samawa location.

The reported wavelength regions was well-known as N. At all reported bands,

The performance of both prediction model: Kriging- based models and VNIR-based models have been assessed by using validation set. In terms of all statistically evaluation parameters, prediction capability of the GIS- Kriging models for total N was as much lower and poor in comparison with VNIR-based model.

**REFERENCES**


