COMBINING PLSR AND KRIGING METHOD TO RETRIEVE SOIL SALINITY WITH ALI DATA IN YELLOW RIVER DELTA

Xingwang Fan[1], Jinmei Tao[1], and Yongling Wengp*[2]

[1]Graduate Student, Surveying and Mapping Department, Transportation School, Southeast University, Nanjing, 210096, PR China; Tel: +86-25-83795236
Email: hantanghe@sina.com

[2]Associate Professor, Surveying and Mapping Department, Transportation School, Southeast University, Nanjing, 210096, PR China; Tel: +86-25-83795236
Email: mgwyl@yahoo.com.cn

KEY WORDS: Yellow River Delta, Remote Sensing, ALI, PLSR, Universal Kriging, Soil Salt Content

ABSTRACT: Soil salinization is one of the most common land degradation progresses in Yellow River Delta (YRD) of China. In this study, deterministic and geostatistical methods were jointly applied to estimate soil salt content (SSC) in this area on the basis of field spectra and Advanced Land Imaging (ALI) data. SSC and field spectra of 50 soil samples were collected to investigate the relationship between soil salinity and ALI-convolved field spectra using partial least square regression (PLSR) method. Significant correlation was observed with determination coefficient (R^2) of 0.837. Subsequently, SSC map was predicted using ALI reflectance data and the PLSR model. The predictions show the similarity to the field observations. To detect soil salinity in the vegetated area, 314 samples were systematically collected on PLSR-derived SSC map and were then interpolated using universal kriging (UK) method. Cross validation results show that 95.9% of the 314 samples are included in the confidence intervals under confidence level of 95% through a Bland-Altman plot. Therefore, it is convinced that the combination of methods provides an inexpensive and labor-saving approach to the estimation of soil salinity in the entire study area with desirable accuracy.

1. INTRODUCTION

As reported in 1990's, 7% of the earth's terrestrial surface, especially irrigated lands were contaminated with excessive salt (Metternicht and Zinck 2003). Remote sensing technique has shown great advantages in identifying and mapping soil salinity in landscape scale for recent decades. Since 1980's, computer-aided image classification has been well documented as another method extensively used to perform soil salinity research. Several studies have demonstrated its efficiency in differentiating saline soil and non-saline soil (Panah *et al*. 1999, Melendez-Pastor *et al*. 2010). However, none of these methods were claimed competent to identify slightly salinized soil, moderately salinized soil and intensively salinized soil. In recent years, the emergence of hyperspectral sensor and quantitative remote sensing technique seem to provide the ability to solve the problem. The capabilities of hyperspectral imagery for salinity mapping have been recently investigated by Ben-Dor *et al*. (2002), Weng *et al*. (2008) and Dutkiewicz *et al*. (2009). The techniques such as continuum removal analysis used in these studies enable the detection of spectral absorption features and facilitate the modelling processes. By using continuum removal analysis, Weng (2010) detected salinity-sensitive bands and constructed salinity spectral index (SSI) based on the continuum removed reflectance.

Quantitative remote sensing technique has also been examined using multi-spectral data (Shrestha 2006, Neild *et al.* 2007). These studies predicted soil salinity by establishing linear and nonlinear models between soil salinity and sensitive indices. To the best of our knowledge, few studies have incorporated all bands of muliti-spectral data into model establishing process to estimate soil salinity.ALI data has recently been introduced into the studies of

soil salinity (Bannari *et al.* 2008). Their results showed that the short wave infrared region (SWIR) of simulated ALI data was a good indicator sensitive to slight and moderate saline soils. To the best of our knowledge, models for estimating salt salinity were only based in general on laboratory spectrometric measurements and simulated data. Applications of field spectrometric measurements remain to be examined. And relatively few studies have investigated the possibility of using ALI data for quantitative mapping of soil salinization. In addition, the soil surface is not always clear of vegetation coverage, which makes no soil information can be extracted from remote sensing images. To estimate soil salinity under vegetated areas, the type and the growing status of vegetation are taken into account as indirect indicators (Dehaan and Taylor 2002, Shrestha 2006, Khan *et al*. 2008). For the area covered with a small amount of vegetation, these vegetation-related studies are reluctant to be effective.

In this study, we focus on estimating soil salinity in a typical area of Yellow River Delta. For this purpose, a PLSR model was created between laboratory measured soil salinity and ALI-convolved field spectra. The model was then applied to ALI reflectance image to produce soil salinity map of bare soils. For the vegetated area, universal kriging was performed on the basis of 314 samples collected on PLSR-derived soil salinity map. We also examined the predictive capability of this combined approach.

2. MATERIALS AND METHODS

2.1 Study Area

The Yellow River Delta is located in Dongying City, Shandong Province of eastern China. A typical study area was selected in 118°41′-119°18′ E and 37°11′-38°07′ E, which covers a rectangle region of about 37 km in width and 106 km in length, corresponding to a scene of ALI image mosaiced with four-image strips (figure 1). Field observations show that the sub-district takes on different levels of saline soils, ranging from non-saline to intensive saline soils, which makes our study more reasonable. The slight and moderate saline soils are generally distributed near the center of the study area, where the Yellow River provides favorable irrigation conditions. The intensive saline soils and salinized soils are concentrated in the low-lying and flat terrain near the beach (Tian *et al*. 2003).

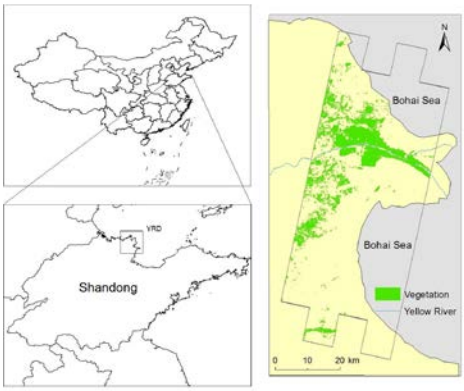

Figure 1. Map of study area with the boundary of ALI image. The area is covered with vegetations such as cotton, reed, *Suaeda salsa* and *Tamarix chinensis*. The vegetated area was masked based on NDVI.

2.2 Soil Sampling and laboratory analysis

Soil sampling was conducted in clear days in April, 2005. A total of 55 topsoil samples were collected using a stratified sampling strategy. Spectrometric measurements were conducted using analytic spectral device (ASD) FieldSpec FR spectroradiometer. Three to five measurements were taken on each target. For each sampling site, variables such as sampling time, status of vegetation and GPS readings were recorded. The

samples were air-dried to eliminate adverse influences of moisture and passed through 2-mm sieve to remove debris, stones and stubbles. Then they were taken for physico-chemical analysis. The concentrations of eight major ions, including K^+, Na^+, Ca^{2+}, Mg^{2+}, Cl^-, HCO_3^-, CO_3^{2-} and SO_4^{2-} were measured and the final SSC were defined as the total of the above ions. The 55 samples were divided into five classes according to national saline soil classification standard of China: nineteen non-saline soil samples with SSC < 1 $g \cdot kg^{-1}$, eight slight saline soil samples with SSC of 1-2 $g \cdot kg^{-1}$, twelve moderate saline soil samples with SSC of 2-4 $g \cdot kg^{-1}$, ten intensive saline soil samples with SSC of 4-6 $g \cdot kg^{-1}$ and six salt soil samples of SSC > 6 $g \cdot kg^{-1}$.

2.3 Partial least square regression and Universal kriging

PLSR is a method that specifies a linear relationship between dependent variables and a set of independent variables. The general idea of PLSR is to extract the orthogonal variables using principal component analysis (PCA) and determine the optimal number of orthogonal variables, accounting for as much variance of the dependent variable with the least residuals (Garthwaite 1994). The only dependent variable in this work is soil salinity, and thus a univariate PLSR algorithm was adopted. The PLSR modeling was performed using the software Unscrimbler (Edition 9.5) developed by CAMO Inc. Spectrometric measurements should be examined and processed before they could be used for model establishment. To reduce noises, reflectance within water-vapor absorption region was removed and the curves were smoothed with a five-point smoothing algorithm. The smoothed spectra were resampled and convolved following FWHM (full-width at half maximum) specifications of ALI imagery (http://www.eoc.csiro.au/). Two-thirds of samples were used to establish a PLSR model and the remaining samples were reserved for the purpose of model validation.

Kriging is a technique for estimating values of continuous random spatial variables from data without bias and with minimum variance (Webster and Mcbratney 1987). Being different from other interpolation methods, kriging takes into account the variables' spatial correlation to interpolate values in an unknown location. Kriging assumes that an estimated value is more likely to resemble the observations nearby than those in a remote location. The assumption is based on the fact that the difference of values in two points increases as their distance increases. To describe the relationship between difference and distance, semi-variogram should be fitted as a plot of semi-variance *vs.* distance. A set of predefined semi-variance functions can be used to determine a semi-variogram such as Gaussian model, spherical model and exponential model. Typically, a semi-variogram displays an increase in semi-variance as distance increases (Burgess and Webster 1980). When the distance reaches a certain value, the semi-variance remains unchanged. The unchanged semi-variance is referred to as sill and the specific distance is referred to as range. It is noted that the semi-variance has a positive value when the sample interval approaches zero. This value is usually described as nugget (Oliver and Webster 1986). These parameters are determined by semivariance analysis and used for interpolation in the further step.

2.4 Image processing

ALI image was collected on April 14th, almost synchronous with the time of field investigation. The data was georeferenced to WGS-84 coordinate in terms of GPS readings with a root mean square error (RMSE) of 0.429 pixels. To match image data to field measured spectra, Empirical Line (EL) algorithm in Envi 4.7 package was used to perform atmospheric correction with spectrometric measurements of eight ground objects. These objects comprise of airport pavements (bright object), bare soils (moderate) and clear reservoirs (dark object). The algorithm forces the corresponding DN value to match the calibration samples

and a linear regression is used to equal DN and reflectance for each band. The calibration was satisfying with relative errors of 0.29% - 2.19%. Nonsoil such as vegetation pixels were removed from the reflectance image using masks created on the basis of NDVI.

3. RESULTS AND DISCUSSION

3.1 Soil salinity of bare soils

The model was established with R^2 of 0.837 and RMSE of 0.747 g·kg^{-1}(figure 2), indicating a good fit between the two datasets. To derive a quantitative soil salinity map of the study area, the model was applied to ALI reflectance image on a pixel-to-pixel basis. In order to reduce the noises, a low pass filter of 3×3 kernel was applied. Statistics of the map show an average SSC value of 4.32 g·kg^{-1} with the maximum value of 15.62 and the minimum value of -5.23. According to the image histogram, appropriately 1.15% soils have negative SSC values, which is unreasonable. To quantitatively assess the accuracy of the SSC map, estimated values of 50 samples were collected on SSC map and compared with the corresponding measured SSC values (figure 2). The result shows that the extracted values are strongly correlated with the true soil salinity (R^2 = 0.645, RMSE = 0.948 g·kg^{-1}).

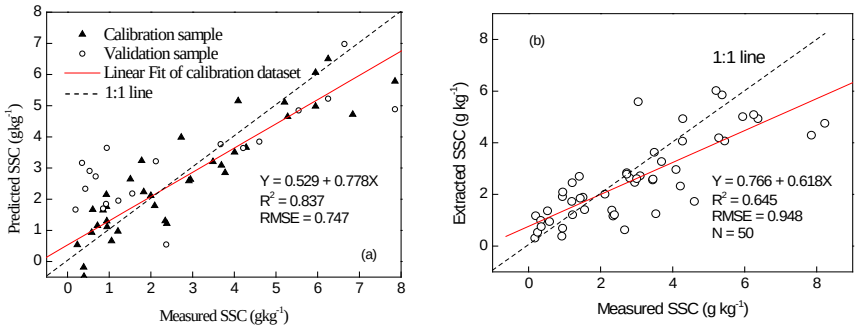

Figure 2. (a) Scatter plot of predicted SSC *vs.* measured SSC. (b)Scatter plot of extracted SSC *vs*. measured SSC.

3.2 Soil salinity under vegetation area

To detect soil salinity in the vegetated area, 314 samples were systematically collected from the PLSR-derived SSC map. These samples include soil enclosed by vegetation and soil close to vegetation area, ranging from 0.795 g·kg^{-1} to 10.236 g·kg^{-1} with an average of 4.394 g·kg^{-1}. Thus, a first-order trend was fitted to represent the general distribution of soil salinity in the study area. The detrended dataset is more likely to obey normal distribution. The semi-variogram shows that it fitted with spherical model. Parameter estimation and interpolation processes were accomplished in ArcGIS geostatistical analysis (GA) module. The magnitude of spatial correlation can be expressed by a ratio of nugget to sill. For the semi-variogram established in this study, the ratio of nugget and sill of SSC is 0.296, suggesting a strong spatial correlation. Based on analysis above, a soil salinity map was interpolated using kriging method. Then Soil salinity of vegetated area was cropped from the kriging-derived soil salinity map with mask created in advance. The estimated SSC values in the vegetation area range from 1.84 to 7.08 g·kg^{-1} with an average value of 3.09 g·kg^{-1}. Descriptive statistics show that 15.11%, 67.67% and 16.50% of the vegetation area is affected with slight, moderate and intensive saline soils, respectively. Only 0.72% of the vegetation area is covered with salt soils. The agricultural lands have a typical SSC value of appropriate 2-3 g·kg^{-1} where salt-tolerant crops such as cotton and wheat are cultivated.

3.3 SSC Mapping

Combining PLSR-derived and the kriging-derived soil salinity map, the entire soil salinity map including bare soils and vegetated area was produced (figure 3).

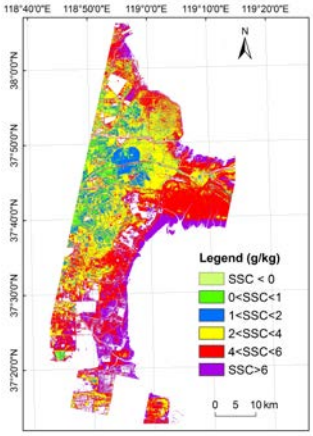

Figure 3. Soil salinity map produced with PLSR and UK combined method.

Descriptive statistics show that 1.03% of the soils have negative SSC values, which are generally classified as non-saline. About 5.55% of the study area is not affected with salt, and 10.25% of the area is slightly salinized, and 31.37% is moderately salinized. The three types of soils comprise the majority of agricultural lands. The soils with SSC value between 4 and 6 $g \cdot kg^{-1}$ reaching as much as 35.00% of the entire area, are dominantly located near the coastal area where the groundwater table is high and seawater provides plenty of salt resources. The rest of the area is salt soils with SSC values higher than 6 $g \cdot kg^{-1}$, accounting for 17.83% of the study area. Most of these soils occur in the area regularly saturated with seawater, for instance, saltpans, dried seawater aquacultural ponds and mudflat.

4. CONCLUSIONS

The PLSR model was established between soil salinity and ALI-convolved field spectra with R^2 of 0.837 and RMSE of 0.747 $g \cdot kg^{-1}$. Transferring the model to ALI reflectance image shows a satisfying result between predicted SSC and measured SSC (R^2=0.645, RMSE=0.948 $g \cdot kg^{-1}$). To estimate soil salinity under vegetated area, a universal kriging method was performed on the basis of 314 soil samples collected on the PLSR-derived soil salinity map. Statistics of cross validation indicate that the result of universal kriging interpolation was promising. The spatial distribution under vegetated area shows coherent spatial patterns with higher soil salinity in eastern coastal area and lower soil salinity in inland area. This paper provides a new approach to estimate soil salinity in the area with sparse vegetation and young plants. By using this approach, it is convenient and inexpensive to obtain soil salinity information in an entire area. Based on soil salinization under vegetation area, the authorities will have sufficient evidence to guide agricultural development.

ACKNOLEDGEMENTS

This work was supported by National Natural Science Foundation of China (No. 41071264). We would like to thank the students at Nanjing University for their assistance in field work and those who make suggestions for revision of this paper.

References

Bannarl, A., Guedon, A. M., El-Harti, A., Cherkaoui, F. Z. and El-Ghmari, A., 2008, Characterization of Slightly and Moderately Saline and Sodic Soils in Irrigated Agricultural Land using Simulated Data of Advanced Land Imaging (EO-1) Sensor. Communications in Soil Science and Plant Analysis, **39**, pp, 2795-2811.

Ben-Dor, E., Patkin, K., Banin, A. and Karnieli, A. 2002, Mapping of several soil properties using DAIS-7915 hyperspectral scanner data: A case study over soils in Israel. International Journal of Remote Sensing, 23, **pp**, 1043-1062.

Burgess, T. M. and Webster, R., 1980, Optimal interpolation and isarithmic mapping of soil properties. I. the semi-variogram and punctual kriging. Journal of Soil Science, 31, pp, 315-331.

Crowley, J. K., 1991, Visible and near-infrared (0.4-2.5μm) reflectance spectra of playa evaporite minerals. Journal of Geochemical Research, 96, pp, 231-242.

Dehaan, R. L. and Taylor, G. R., 2002, Field-derived spectra of salinized soils and vegetation indicators of irrigation-induced soil salinization. Remote Sensing of Environment, 80, pp, 406-417.

Dutkiewicz, A., Lewis, M. and Ostendorf, B., 2009, Evaluation and comparison of hyperspectral imagery for mapping surface symptoms of dryland salinity. International Journal of Remote Sensing, 30, pp, 693-719.

Garthwaite, P. H., 1994, An interpretation of partial least squares. Journal of American Statistical Association, 89, pp, 122-127.

Khan, N. M., Rastoskuev, V. V., Sato, Y. and Shiozawa, S., 2008, Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. Agriclutural Water Management, 77, pp, 96-109.

Melendez-Pastor, I., Nabarro-Pedreno, J. and Koch, M., Gomez, I., 2010, Applying imaging spectroscopy techniques to map saline soils with ASTER images. Geoderma, 158, pp, 55-65.

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

Neild, S. J., Boettinger, J. L. and Ramsey, R. D., 2007, Digitally mapping gypsic and natric soil areas using Landsat ETM data. Soil Science Society of America Journal, 71, pp, 245-252.

Oliver, M. A. and Webster, R., 1986, Semi-variograms for modelling the spatial pattern of landform and soil properties. Earth Surface Processes and Landforms, 11, pp, 491-504.

Panah, S. K., Goossens, R. and Dapper, M., 1999, Study of soil salinity in the Ardakan area, Iran, based upon field observations and remote sensing. 18th EARSeL symposium on Operational Remote Sensing for Sustainable Development, 11-14 May, 1999, Enschede, The Netherlands, pp, 419-426.

Shrestha, R. P., 2006, Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand. Land Degradation & Development, 17, pp, 677-689.

Webster, R. and Mcbratney, A. B., 1987, Mapping soil fertility at Broom's Barn by simple kriging. Journal of the Science of Food and Agriculture, 38, pp, 97-115.

Weng, Y. L., Gong, P. and Zhu, Z. L., 2008, Soil salt content estimation in the Yellow River delta with satellite hyperspectral data. Canadian Journal of Remote Sensing, 34, pp, 259-270.

Weng, Y. L., Gong, P. and Zhu, Z. L., 2010, A Spectral Index for Estimating Soil Salinity in the Yellow River Delta Region of China Using EO-1 Hyperion Data. Pedosphere, 20, pp, 378-388.