RETRACTED ARTICLE: Monitoring soil salinization in Manas River Basin, Northwestern China based on multi-spectral index group

Xiaoyan Shi, Jianghui Song, Haijiang Wang and Xin Lv
College of Agriculture, Shihezi University, Shihezi, People’s Republic of China; The Key Laboratory of Oasis Ecological Agriculture of Xinjiang Production and Construction Group, Shihezi University, Shihezi, Xinjiang, People’s Republic of China

ABSTRACT
Large-scale and accurate monitoring soil salinization is essential for controlling soil degradation and sustainable agricultural development. The agricultural irrigation area of the Manas River Basin in the arid area of Northwest China was selected as the test area. The soil salinization monitoring model based on spectral index group was constructed by comparing the accuracy of PCR, PLSR and MLR models using the transformation of multi-spectral index group and index screening. The results showed that there was a certain correlation between each index and soil salinity. The correlation coefficients for original data for the logarithm Ln(R), exponential eR and square root R1/2 respectively, the correlation between each index and soil salinity was significantly improved, with the maximum correlation coefficient was up to 0.7564 of R1/2. The salt content estimation models were constructed by different data transformation using PLSR, PCR and MLR methods, respectively. This study provides a fast and accurate method for monitoring regional soil salinity content and the results can provide a reference for soil salinity grading management in arid and semi-arid areas.

Introduction
Soil salinization is a typical soil degradation phenomenon occurring in arid and semi-arid regions (Fernandez-Bucetas et al., 2006; Metternicht & Zinck, 2003; Wang & Jia, 2012). At present, global salinized soil accounts for about 3% of global land resources, which increases by 2.0 × 10^6 hm² per year (Peng et al., 2019). Secondary salinization area is about 77.00 hm², 58% of which occurs in irrigated agricultural areas (Metternicht & Zinck, 2003). Due to the high soluble salt content of soil parent material, low rainfall, high evaporation intensity, shallow groundwater burial depth and unreasonable water resources management, soil salinization is increasingly severe in arid and semi-arid areas (Shrivastava & Kumar, 2015). The total area of saline soil in China has currently reached 36 million hectares, of which, saline soil in cultivated land accounts for 6.62% (The National Soil Survey Office, 1998). Salinized soil not only changes the physicochemical properties of the soil but also inhibits crop growth in severe cases (F. Wang et al., 2015). Accordingly, rapid and accurate acquisition of regional soil salinity dynamics is a prerequisite for better utilization of salinized farmland.

The current laboratory determination of soil salinity by traditional soil sampling has high precision. However, the sparse sampling points and poor sample representativeness make it difficult to meet the needs of large-area and simultaneous continuous monitoring of soil salinity. Compared with traditional technology, remote sensing technology has wide coverage and fast cycle, which has provided an effective and rapid non-destructive new approach for dynamic monitoring of regional soil salinization. Studies have shown that reflectance value of Landsat data has great potential for resource environment investigation and mapping. By Landsat remote sensing image, it is possible to estimate regional land use and cover (Gao et al., 2019; Q. Huang et al., 2019a; Q. Liu et al., 2019), water resource distribution (Du et al., 2019; Xiao & Ouyang, 2019; Yan & Guo, 2019), soil salinity pattern evolution (Essahlaoui et al., 2019), avalanche hazards (Singh et al., 2019), surface temperature (Dhar et al., 2019), crop yield (Filippi et al., 2019), soil nutrients (Darwish & Fadel, 2017), etc. proving that the technology plays a huge role in quantitative analysis of resource environment.

Determination of soil salinity by remote sensing is mainly to perform quantitative inversion based on the law of soil reflectance change with soil salinity content (Hick & Russell, 1990). For higher soil salt content, spectral response of visible and near-infrared band is stronger (Rao et al., 1995). These studies have laid the theoretical foundation for remote sensing monitoring of soil salinity. However, spectral reflectance of soil provides a comprehensive reflection of soil physical and chemical properties, which is influenced by soil nutrients (Hengl et al., 2017), atmospheric water vapor (Song & Grejner-}

CONTACT Haijiang Wang  wanghaijiang@shzu.edu.cn  College of Agriculture, Shihezi University, Shihezi 832000, People’s Republic of China

This article has been retracted. Please see Retraction (http://dx.doi.org/10.1080/22797254.2020.1762247)  © 2020 The Author(s). Published by Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Brzezinska, 2009), ground-surface temperature (Anderson & Johnson, 2016), and so on. Therefore, it is difficult to accurately estimate the soil salt content based on the original spectral reflection characteristics (Jiang & Shu, 2019; Tilley et al., 2007). After some mathematical transformation of the soil spectral curve, the reflection and absorption characteristics of the spectrum can be better highlighted, so as to distinguish it from other spectral characteristics, and screen-sensitive bands. Studies have shown that spectral transformation such as logarithmic transformation (Fu et al., 2019), Square root transformation (ZY Liu et al., 2018), reciprocal transformation (Fu et al., 2019; ZY Liu et al., 2018). The prediction accuracy can be effectively improved by adding auxiliary variables (such as topographic factors, spectral indexes, environmental factors, etc.) (Chuangye et al., 2016; Wang et al., 2018). In soil salinity monitoring using Landsat remote sensing image, Jiang and Shu (2019) added vegetation index NDVI and soil salinity index DSI significantly related to soil salinity based on the original band index, thus effectively improving the accuracy of salt monitoring. In the study on soil salinity of three oases in Xinjiang, China, Fei Wang et al. (2013) found that accuracy of the salt prediction model would gradually increase as near-infrared band index NIR, soil sand index, vegetation index EVI, slope and surface temperature LST were gradually introduced. Y. Zhang et al. (2019) found that models constructed using original band index (O), normal spectral index (S), red edge index (R) and environmental variables (E) under bare soil conditions exhibited the best predictive ability. Soil salinity not only affects reflectance of the soil itself, but also affects degree of surface looseness, water condition and growth of aboveground vegetation. Hence, in soil salinity study, quantitative inversion of salt content can be achieved by combining vegetation index, water index and soil index.

In summary, current inversion of soil salinity by remote sensing technology is mainly to build salt content estimation model based on salt-sensitive spectral band or in combination with simple spectral index. Studies on optimization, evaluation principles of the model construction index as well as comprehensive estimation accuracy of multi-type spectral indexes are still insufficient (Hao et al., 2018; Jibo et al., 2018). Therefore, taking saline-alkali soil in Manas River Basin in Xinjiang, China as the research object, based on extensive situ measurement of soil salt samples, this study integrates Landsat8 remote sensing data and vegetation, soil and water spectral index groups, comparatively analyzes the impact of spectral data transformation, spectral index screening group on accuracy of soil salinity inversion model using principal component regression (PCR), partial least squares regression (PLSR) and multiple linear regression (MLR). A quantitative estimation model of soil salinity is constructed, accuracy of the salt content inversion model is assessed under different model input variables and different model methods to provide a reference for quantitative inversion of salinity by remote sensing in arid and semi-arid regions.

Materials and methods

Overview of the study area

Manas River Basin, located in Xinjiang, China, is adjacent to Tianshan Mountains in the north and Junggar Basin in the south. The terrain is high in the south and low in the north. It has mountain in the south, with plain in the middle and Gurbantunggut desert in the north. The total area is 3.35 × 10^4 km^2. The whole basin consists of five large rivers, namely the Taxi River, the Manas River, the Ningjia River, the Jingou River and the Bayingou River, which are mainly supplemented by melting of glaciers and precipitation. The basin is located in the hinterland of Eurasia, with an average annual temperature of 4–7°C. It is hot in summer and cold in winter. The precipitation varies greatly in time and space, with average annual precipitation at 110–200 mm and annual average evaporation at 1500–2100 mm. With high underground water level, it has typical arid continental climate. With 148–187 days of frostless period, it is a typical desert oasis and irrigated agricultural area. The groundwater level is high, and the groundwater is mainly supplemented by river channels, canal systems, leakage of plain reservoirs and infiltration of field irrigation. Affected by climatic conditions, hydrological conditions, soil parent material, vegetation, etc., it has complex soil types and soluble salt can easily form large-area saline soil on the surface under natural conditions.

The geomorphology of the Manas Basin has a typical mountain basin system structure, which is mountainous area-piedmont plain-desert in turn (Shao & Cui, 2003). After the river developed in the southern mountainous region enters the basin, the soil carried by it gradually deposits, forming alluvial flood fans, impact plains, and dry deltas in this order. Affected by factors such as regional soil parent material, groundwater level, and climatic conditions, a large number of different types of saline soils are distributed in the area, especially the alluvial fan edges and flood plain with high groundwater level, abundant groundwater volume, and poor flow, which cause severe salinization in the top part of the alluvial plain (Z. H. Wang et al., 2010). Under natural conditions, the soil leaching and desalination processes in the irrigation area of the Manas River Basin are very weak, and the unreasonable irrigation and drainage methods at the beginning of reclamation led to the increase of groundwater level and the high degree of mineralization of irrigation water, which finally form secondary salt in the basin. Waterlogging is frequent and seriously affects local agricultural development.
Field data acquisition and processing

In this study, a field sampling survey was carried out on Manas River Basin from July to August 2017. The layout of the samples gave comprehensive considerations to different land use types, planting structures, landform types, irrigation methods and soil types. A total of 337 observation sample regions were set in sampling points with wide coverage, representativeness and diverse terrain types (Figure 1). GPS was used to accurately locate and record coordinates of each sampling point. Each plot area is about 30 m × 30 m (corresponding to single pixel area in Landsat image). In each sample area, five soil surface samples (0–20 cm) were collected by five-point sampling method, uniformly mixed to form a mixed soil sample and brought back to the laboratory. After drying, sieving and grinding, the samples were screened through 2 mm sieve. Total salt content of the soil was determined by gravimetric method with a 1:5 soil-to-water solution (Shi-dan, 2010). A total of 337 samples were randomly divided into two groups, of which 253 were used for modeling (75% of the samples) and 84 samples were used for verification. The salt content of the samples is shown in Table 1.

Image data processing and index selection

The selection of remote sensing image data was basically consistent with field sampling of soil in time. The Landsat8 remote sensing image of the Manas River Basin in Xinjiang, China on 13 July 2017 was collected (www.gscloud.cn/). The projected coordinate system is UTM-WGS84, with a total of 11 bands (B). The image preprocessing work of geometric correction, radiation calibration and FLAASH atmospheric correction was carried out using ENVI 5.3 software, and the surface reflectance value of each band corresponding to the field sampling points was extracted. Spectral index can magnify weak connection between the amplification bands, reduce model complexity and remove redundant variables (Shi et al., 2014). Many researchers have tried to introduce it into remote sensing inversion model to improve estimation accuracy of the model (Allbed et al., 2014; Kertész & Tóth, 1994). After consulting domestic and foreign literatures, 17 spectral indexes related to soil salinity inversion were selected as analysis indexes, including 5 vegetation indexes (V), 11 soil indexes (S) and 1 water body index (W) (Table 2).

Figure 1. Distribution map soil salinity at sampling points in the test area.
Remote sensing inversion model of soil salt

In order to find the relationship between the selected indexes and soil salinity, all remote sensing indexes were subject to mathematical transformation. There will be errors in remote sensing images due to lighting conditions and terrain factors, and mathematical transformation can lower the impact of noise on spectrum to a certain extent, thereby strengthening the relationship between remote sensing indexes and soil salinity. To further analyze the relationship between remote soil sensing index and soil salinity, the soil salt content data of each sampling point and the standardized soil remote sensing index (indicated by R) were mathematically transformed, and the transformation forms include reciprocal (1/R), exponent (e^R), logarithm (ln(R)), reciprocal of logarithm (1/ln(R)), reciprocal of exponent (1/e^R), root mean square (R^(1/2)). All the indexes were standardized using the formula as follows:

\[ x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

whereas \( x \) is an original value, \( x' \) is the normalized value.

For the selection of sensitive bands for remote sensing modeling of soil salinity, soil salinity content and all indexes under each transformation are usually selected for correlation analysis. A greater absolute value of the correlation coefficient indicates higher correlation with salt. Screening of modeling methods means great significance for remote sensing interpretation of soil salinity. At present, remote sensing technology mainly adopts regression model for inversion of soil salinity. In this study, PCR, PLSR and MLR modeling methods were selected for remote sensing inversion of soil salinity. In PCR, principal components are extracted based on the principle that the selected principal components retain the information of the original variables as much as possible and the two are uncorrelated with each other. Then, the new variable is modeled by MLR and linearly transformed to obtain the linear regression model of dependent variable and independent variable (Sun, 1995). PLSR modeling is to select step by step a number of new comprehensive variables irrelevant to each other and capable of whole system interpretation from independent variables to replace the original variables. The dependent variable is used for regression modeling of m principal components, and the regression equation between the dependent variable and the independent variable is then obtained by transformation, thus fulfilling the modeling. To some extent, the modeling can effectively eliminate the effect of the correlation between the independent variables on the model coefficients (Gusnanto et al., 2003). MLR is to predict the target variables through fitting of linear equations using multiple explanatory variables (Y. Huang et al., 2019b).

The accuracy of the salt interpretation model was tested by cross-validation, and the optimal model was selected by comparing accuracy and reliability of the model based on the determination coefficient of the test model, the root mean square error value. For higher R^2 (V), the stability and fitting degree of the model is higher; for lower RMSE(V), the model has better estimation capacity.

### Table 1. Descriptive statistical analysis table of soil sample salt content.

| Sample (number)     | Mean (g kg^-1) | Max. (g kg^-1) | Min. (g kg^-1) | SD (g kg^-1) | CV (%) |
|---------------------|----------------|----------------|----------------|--------------|-------|
| Calibration data-set (253) | 15.67          | 35.29          | 0.38           | 7.26         | 46.33 |
| Validation data-set (84)      | 17.56          | 34.28          | 0.30           | 7.39         | 42.08 |
| Total sampling points (337)   | 16.14          | 35.29          | 0.30           | 7.32         | 45.35 |

### Table 2. Spectral index and formula.

| Salinity indices | Full name | Spectral functions | Reference |
|------------------|-----------|--------------------|-----------|
| NDVI             | Normalized difference vegetation index | \( \text{NDVI} = \left( \frac{\rho_{\text{ NIR}} - \rho_{\text{ R}}} {\rho_{\text{ NIR}} + \rho_{\text{ R}}} \right) \) | Rouse et al. (1974) |
| GDVI             | Generalized difference vegetation index | \( \text{GDVI} = \left( \frac{\rho_{\text{ SWIR}} - \rho_{\text{ R}}} {\rho_{\text{ SWIR}} + \rho_{\text{ R}}} \right) \) | Sripada et al. (2006) |
| WDIVI            | Weighted difference vegetation index | \( \text{WDVI} = \left( \frac{\rho_{\text{ NIR}} - a \times \rho_{\text{ R}}} {\rho_{\text{ NIR}} + \rho_{\text{ R}}} \right) \) | Basso et al. (2000) |
| SAVI             | Soil adjusted vegetation index | \( \text{SAVI} = \left( \frac{\rho_{\text{ NIR}} - \rho_{\text{ R}}} {\rho_{\text{ NIR}} + \rho_{\text{ R}}} \right) \) | Huete (1988) |
| RVI              | Ratio vegetation index | \( \text{RVI} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Pearson and Miller (1972) |
| NDSI             | Normalized difference salinity index | \( \text{NDSI} = \left( \frac{\rho_{\text{ NIR}} - \rho_{\text{ R}}}{\rho_{\text{ NIR}} + \rho_{\text{ R}}} \right) \) | Tripathi et al. (1997) |
| SI               | Salinity index | \( \text{SI} = \sqrt{\rho_{\text{ R}} + \rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S1               | Salinity index 1 | \( \text{S1} = \sqrt{\rho_{\text{ R}} + \rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S2               | Salinity index 2 | \( \text{S2} = \sqrt{\rho_{\text{ R}} + \rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S3               | Salinity index 3 | \( \text{S3} = \sqrt{\rho_{\text{ R}} + \rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S4               | Salinity index 4 | \( \text{S4} = \sqrt{\rho_{\text{ R}} + \rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S5               | Salinity index 5 | \( \text{S5} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S6               | Salinity index 6 | \( \text{S6} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S7               | Salinity index 7 | \( \text{S7} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S8               | Salinity index 8 | \( \text{S8} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S9               | Salinity index 9 | \( \text{S9} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| S10              | Salinity index 10 | \( \text{S10} = \frac{\rho_{\text{ R}}}{\rho_{\text{ NIR}}} \) | Khan et al. (2005) |
| LSWI             | Land surface water index | \( \text{LSWI} = \left( \frac{\rho_{\text{ NIR}} - \rho_{\text{ SWIR}}}{\rho_{\text{ NIR}} + \rho_{\text{ SWIR}}} \right) \) | Nield et al. (2007) |
\[
R^2(C) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

\[
R^2(V) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_p)^2}
\]

RMSE(C) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}

RMSE(V) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_p)^2}

where \(y_i\) is the observed value, \(\hat{y}_i\) is the interpreted value, \(\bar{y}\) is the mean value, and \(n\) is the number of data points. C represents for calibration and P represents for validation.

**Result analysis**

*Correlation between soil salinity and band, spectral index*

In order to understand the sensitivity of different remote sensing bands to soil salinity, the correlation between soil salinity in the study area and the original band reflectance of Landsat was analyzed. It can be seen from Figure 2 that great correlation exists between the remote sensing bands, and the correlation between visible light bands is high, up to 0.969. Good correlation is shown between short-wave infrared 1 and short-wave infrared 2, with a correlation coefficient of 0.939, and good correlation is also exhibited between near-infrared band and visible light band.

Correlation analysis between 11 Landsat image bands and salt content reveals that salt has a positive correlation with visible light band and a negative correlation with near-infrared band. Except B6 and B8 bands (whose correlation coefficients are 0.094, -0.058, respectively), indexes of other bands have a significant correlation with salt, and the largest relevant B10 band has a correlation coefficient of 0.3689, which is consistent with the results of Peñuelas et al. (1997). As the soil salinity increases, visible light bands have increased reflectance while near-infrared bands have decreased reflectance. It is mainly because spectral reflectance of visible light is subject to influence of chlorophyll (Osborne et al., 2002); when the plant faces salt stress, chlorophyll level decreases with the decrease of photosynthesis, so absorption of blue light and red light is weakened, leading to increased blue and red light reflectance of vegetations. Reflectance of the near-infrared region decreases with increasing salt content, which is probably because the near-infrared spectrum reflection is absorbed by water and OH ionic group in the crystal (Wiegand et al., 1992). Soil salt is significantly negatively correlated with B10 and B11 bands, with correlation coefficients r of -0.369 and -0.359, respectively. It is possibly because sulphated soil is widely distributed in surface soil of the Manas River Basin (Abuduwaili et al., 2012), while sulfuric acid absorption band exists in TIRS2 band (Metternicht & Zinck, 1997). A significant correlation

![Figure 2. Correlation heat map between original index and soil salinity.](image-url)
is shown between soil salt content and salinity index, vegetation index. Where, the correlation coefficient between SI2 and salt content is −0.3129, and the correlation coefficients of RVI, NDVI, WDVI and SAVI are −0.3027, −0.3078, −0.3143 and −0.3112, respectively. Each vegetation index shows a significant negative correlation, which is probably because the sampling time is July when the surface has different degrees of vegetation coverage, if the soil salt content exceeds a certain value, plant growth will be inhibited, leading to increased reflectance of red light band and decreased reflectance of near-infrared band, thereby affecting the associated vegetation index.

Different forms of transform processing of spectral reflectance can help eliminate background interference, improve spectral sensitivity and correlation (Gong & Yu, 2001). Figure 3 shows the correlation coefficient between each index and salt content after $1/Ln(R)$, $1/e^R$, $1/R$, $Ln(R)$, $e^R$ and $R^{1/2}$ transformation of the original data. It can be seen that $1/Ln(R)$, $1/e^R$, $1/R$ data transformations fail to significantly increase the correlation between each index and salt, but after logarithmic transformation $Ln(R)$, exponential transformation $e^R$ and square root transformation $R^{1/2}$, the correlation between each index and soil salinity is significantly improved. Where, $Ln(R)$ and $e^R$ transformations can significantly improve the positive correlation with B5, SI2, S1, S2, RVI, NDVI, GDVI, WDVI, SAVI and LSWI, and improve negative correlation with B2, SI,S11,S13,S3,S5,SI-T and NDSI; $R^{1/2}$ transformation can improve positive correlation with B1, B2, B3, B4, B6, B7, SI, S11, SI3, S3, S5, SI-T and NDSI, and improve negative correlation with B5, B10, B11, SI2, S1, S2, S6, RVI, NDVI, GDVI, WDVI, SAVI and LSWI. The vegetation index in each transformation is negatively correlated to the salt index, which also indicates that the vegetation index decreases as soil salinity increases.

**Construction of soil salinity interpretation model**

**Estimation accuracy of different data transformation models**

The soil salt content estimation model was constructed by PLSR, PCR and MLR methods respectively, and the effect of different data transformations on the accuracy of the estimation model was compared and analyzed (Table 3). Seen from determination coefficient and root mean square error of the model, after $e^R$, $1/e^R$, $Ln(R)$, $1/Ln(R)$ and $R^{1/2}$ transformations of the original spectrum, determination coefficient $R^2(C)$, root mean square error RMSE(C) of the modeling set, plus determination coefficient $R^2(V)$, root mean square error RMSE(V) of the prediction set are significantly improved. After $e^R$, $1/e^R$, $Ln(R)$, $1/Ln(R)$, $R^{1/2}$ transformations, PLSR method-based modeling set $R^2(C)$ are 0.6347, 0.5899, 0.6083, 0.5752 and 0.6489, respectively, and RMSE(C) are 4.4924, 5.1904, 4.6521, 5.2503 and 4.0345 g kg$^{-1}$, respectively. PLSR method has superior estimation accuracy than PCR and MLR methods, indicating that data transformation through spectral indexes has certain effect on improving accuracy and stability in soil salt content prediction. The inversion accuracy of five mathematical transformation forms is comprehensively compared. $R^{1/2}$ transformation model has more significant modeling and prediction effects than models $e^R$, $1/e^R$, $Ln(R)$ and $1/Ln(R)$, and $R^2$ of the modeling set and the prediction set are 0.6489 and 0.6033, respectively, while RMSE are 4.0345 and 4.5456, respectively, suggesting that $R^{1/2}$ transformation can better eliminate the effect of natural factors such as soil texture and soil parent material as well as human

![Figure 3. Correlation coefficient maps of different index and soil salt content after six transformations.](image-url)
Table 3. Estimation accuracy of different data transformation.

| Data transformation | PLSR       | PCR        | MLR        |
|---------------------|------------|------------|------------|
|                     | \(\text{RMSE(C)}\) (g kg\(^{-1}\)) | \(\text{RMSE(V)}\) (g kg\(^{-1}\)) | \(\text{RMSE(C)}\) (g kg\(^{-1}\)) | \(\text{RMSE(V)}\) (g kg\(^{-1}\)) | \(\text{RMSE(C)}\) (g kg\(^{-1}\)) | \(\text{RMSE(V)}\) (g kg\(^{-1}\)) |
| \(R\)               | 0.4749     | 0.2487     | 0.3511     | 0.2204     | 0.3048     | 0.2631     |
| \(1/R\)             | 0.3283     | 0.2318     | 0.3057     | 0.2611     | 0.2961     | 0.2611     |
| \(e^R\)             | 0.6347     | 0.4924     | 0.5339     | 0.4466     | 0.5021     | 0.4466     |
| \(1/e^R\)           | 0.5899     | 0.459      | 0.5065     | 0.3978     | 0.4964     | 0.3978     |
| \(\text{Ln}(R)\)    | 0.6083     | 0.4671     | 0.4647     | 0.3785     | 0.4632     | 0.3785     |
| \(1/\text{Ln}(R)\)  | 0.5752     | 0.4154     | 0.5106     | 0.4287     | 0.4806     | 0.4287     |
| \(R^{1/2}\)         | 0.6498     | 0.6033     | 0.5634     | 0.5237     | 0.5281     | 0.5237     |
factors on spectral index, thus enhancing accuracy of spectral index in estimating soil salt content.

**Calibration index and accuracy of model**

In the study, 28 spectral group indexes were selected and added to the model from big to small according to absolute values of correlation, so that the relationship between the number of screening index factors and equation determination coefficient as well as equation prediction accuracy was obtained. Figure 4 shows accuracy verification of the PLSR model when different variables are selected under different data transformation. When the number of factors participating in the modeling increases gradually, except 1/R transformation form, the other six transformation forms show consistent variation trend in model verification accuracy. That is, as the number of model factors increases, model stability increases first and then decreases, while fitting performance of 1/R transform processing model increases with the increase of index factor, reaching optimal state when all factors participate in the modeling \((R^2(V) \text{ and RMSE}(V) \approx 0.2318 \text{ and } 8.0962 \text{ g kg}^{-1})\), respectively. Seen from determination coefficient of soil salinity and prediction accuracy, except 1/R and Ln(R), the soil salt monitoring models of other transformations are superior to the original spectral modeling results, indicating that certain data transformation of the original spectral data helps improve monitoring accuracy of soil salinity. Considering soil salinity prediction accuracy and model complexity, \(R^{1/2}\) transformation not only has higher verification accuracy (the model verification accuracy \(R^2(V) \text{ is increased from } 0.1751 \text{ to } 0.2786 \text{ in the original spectral index model to } 0.4351 \text{ and } 0.6472\)), but also achieves the highest prediction accuracy when fewer factors \((n = 10)\) are introduced to the model.

**Optimization and verification of the estimation model**

Based on the principle of maximum \(R^2(V)\) and minimum \(\text{RMSE}(V)\), the PLSR model with 10 factors of \(R^{1/2}\) transform participating in modeling \((R^2(V) \approx 0.6472 \text{ and RMSE}(V) \approx 4.3106)\) is selected as the optimal model. Figure 5a shows the regression coefficient of the optimal model. Using 84 measured values, the selected optimal model is validated to obtain accuracy of modeling set and verification set (Figure 5b). It can be seen that a good linear relationship is shown between the predicted value and the measured value as a whole. Most samples are distributed on both sides of the 1:1 line. However, some samples deviate from the 1:1 line, and measured salt mass fraction between 0~10 g kg\(^{-1}\) inverse values between 7~12 g kg\(^{-1}\), which are higher than the measured value.

**Remote sensing inversion of soil salinity**

The optimal model of this study was applied to remote sensing images, and the soil salt content distribution map of Manas River Basin was obtained by inversion (Figure 6). Most of the farmland soil in the study area belongs to moderate and severe salinized soil, accounting for 58.97% and 25.41% of farmland in the irrigation area of Manas River, respectively. The areas with severe soil salinization are mainly located in the southeast of Manas River Basin, which is mainly because the area is located at alluvial–proluvial fan edge of the basin, with shallow burial depth of groundwater and unsmooth flow, and the soil texture is mainly clay loam with high salt content.

**Discussion**

**Effect of spectral data transformation on correlation**

This study found that after data transformation of the original indexes by mathematical algorithm, some indexes increased the correlation with soil salinity (Ln (R), \(e^{R}\), \(R^{1/2}\)), increased differences between the indexes and increased the correlation between the transformed index and soil salt content. Also, some failed to significantly increase the correlation coefficient between spectral index and salinity \((1/R, 1/	ext{Ln}(R), 1/e^{R})\). All of these data transformation forms

Figure 4. Influence of variable numbers to model reliability of PLSR model under six transformations.
involve the process of solving reciprocal, indicating that reciprocal processing has little effect on enhancing characteristics of the index. It can thus be known that during analysis of remote sensing data, data transformation is a method to improve modeling accuracy, but not every data transformation can improve modeling accuracy. Therefore, in the research, it is a very critical step to select a data transformation form suitable for modeling. The results show that the three data transformation processing (Ln(R), e^R and R^{1/2}) of standardized variable can improve the correlation between soil salt content and variable index.

**Figure 5.** The coefficient of optimal model and verification (10 variables of PLSR model under $R^{1/2}$ transformation).

**Figure 6.** Inversion of soil salinity content in Manas River Basin.
Moreover, in the $R^{1/2}$ data transformation form, PLSR model achieves the maximum verification accuracy ($R^2(C) = 0.6472$) and the minimum RMSE (4.3106 g kg$^{-1}$) at 10 factors, indicating that the model has good stability. Mainly targeting at soil salinity, this paper filters data transformation forms favorable to remote sensing modeling of soil salinity. However, band reflectance is subject to combined influences of soil parent material, water, salt, organic matter and texture, and influence factors differ for regional soils (Dor et al., 2015), so further research is needed to study whether different soil ion composition, parent material, water content factors will affect the optimal transformation form. Therefore, whether the model on square root of soil salinity index established in this study is applicable to other regions needs further analysis.

**Importance of different types of indexes in modeling**

Since soil is closely related to multiple factors (such as vegetation, parent material, soil moisture), these factors should also be used as predictors for soil property modeling in the study on spatial distribution of soil properties (Camera et al., 2017; Minasny & McBratney, 2016; Mulder et al., 2011). Table 4 lists the evaluation parameters of various indexes and their combinations in the modeling of soil salinity under square root data transformation. When only a single variable is introduced in the model, one of B, S, V and W models with the highest accuracy is selected for modeling total vegetation indexes. As other indexes are added to the model, modeling accuracy of the model is further improved. Where, $B + S + V + W$ model with four types of predictors has the highest modeling accuracy ($R^2(C) = 0.5426$), its verification accuracy is lower than that of three predictor variable models ($B + S + V$, $B + S + W$, $B + V + W$, $S + V + W$), and the model stability is poor. This may be because the study area has typical arid continental climate with evaporation greater than rainfall, and the surface soil moisture is low, leading to small correlation between soil salinity and soil water body index. Addition of the index to the model will lower reliability of the equation to a certain extent. Moreover, vegetation index and soil index in this study are derived from band calculation. When there are too many indexes, there will be more redundant information, resulting in over-fitting of the model. Asfaw et al. (2018) pointed out that spectral reflectance of soil is subject to influence of spectral characteristics. Salt crust, soil color, moisture and crops in salinized soil will affect reflection amount. Therefore, combination of spectral bands with image enhancement techniques has broad prospects in quantitative analysis of soil salinization using remote sensing. This paper finds that combination of band index (B), soil salinity index (S) and vegetation index (V) can achieve better results in soil salinity evaluation of arid and semi-arid regions, which is consistent with the results of Fei Wang et al. (2013), T. Zhang et al. (2015), and Yu et al. (2018). That is to say, in the quantitative analysis of the resource environment, it is more reliable and accurate to comprehensively consider multiple indexes than considering a single type of index.

**Conclusion**

In this paper with Xinjiang Manas River Basin as the research area, a quantitative estimation model of soil salinity is constructed by integrating field measured data, Landsat8 remote sensing data and vegetation, soil and water spectral index groups. The conclusions are drawn as follows:

(1) There is low correlation between the reflectance of original band of Landsat satellite image, with the highest in B10 band (0.3689). Logarithmic transformation $\ln(R)$, exponential transformation $R^5$ and square root transformation $R^{1/2}$ of the spectral group data significantly improve the correlation between soil salinity and various indexes. Where, transformation form with the most significant increase is $R^{1/2}$, with a maximum increase of S5, which is increased by 0.4510 compared with the original correlation coefficient, indicating that spectral transformation is an effective way to improve the correlation between spectral index and salt content.

(2) In comparison of the three methods of PLSR, PCR and MLR for constructing soil salt estimation model, based on all spectral index modeling, PLSR model is the best, followed by PCR model and MLR model is the worst.

(3) By adding the spectral group index of the seven data transformation forms to the PLSR model from big to small according to the absolute value of the correlation coefficient, it is found that

| Table 4. Reliability analysis of PLSR model in various index type and combination type under $R^{1/2}$ transformation. |
|---------------------------------------------------------------|
| $R^2(C)$ | RMSE(C) | $R^2(V)$ | RMSE(V) |
|---|---|---|---|
| B | 0.6713 | 3.9024 | 0.6375 | 4.3189 |
| B + S | 0.6859 | 3.8214 | 0.6348 | 4.3333 |
| B + W | 0.6883 | 3.8074 | 0.6404 | 4.3035 |
| B + S + V | 0.6721 | 3.8983 | 0.6387 | 4.3126 |
| B + S + W | 0.6971 | 3.7618 | 0.6458 | 4.2745 |
| B + S + V + W | 0.6913 | 3.7907 | 0.644 | 4.2844 |
| B + V | 0.6942 | 3.7745 | 0.6444 | 4.2858 |
| B + V + W | 0.6691 | 3.9145 | 0.628 | 5.0126 |
| S + V | 0.6866 | 3.817 | 0.6425 | 4.2923 |
| S + W | 0.6721 | 3.9055 | 0.6402 | 4.3047 |
| S + V + W | 0.6924 | 3.7846 | 0.6471 | 4.2675 |
| V | 0.6845 | 3.8288 | 0.6411 | 4.2998 |
| V + W | 0.7652 | 3.8753 | 0.6347 | 4.3339 |
| W | 0.5598 | 4.4851 | 0.4973 | 5.3643 |
| B + S + V + W | 0.6498 | 4.0345 | 0.6033 | 4.5456 |
except 1/R transformation, stability of the other six transformational models first increases and then decreases as the number of factors increases. Finally, according to the principle of maximum $R^2(V)$ and minimum RMSE(V), the PLSR model with 10 factors of $R^{1/2}$ transformation participating in modeling is selected as the optimal interpretation model for soil salinity ($R^2$ (V) is 0.6472 and RMSE (V) is 4.3106 g kg$^{-1}$).

(4) Comparison of model evaluation parameters under $R^{1/2}$ transformation, band index (B), soil salinity index (S), vegetation index (V), water index (W) and their combined forms reveals that it is more reliable and accurate to take B, S and V indexes into account than using a single type of index.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the International Science & Technology Cooperation Program of China [2015DFAI11660]; Major Scientific Research Projects of Xinjiang Production and Construction Corps [2018AA004, 2018AA005].

References

Abuduwaili, J., Yang, T., A. N. G., Abulimiti, M., DongWei, L. I. U., & Long, M. A. (2012). Spatial distribution of soil moisture, salinity and organic matter in Manas River watershed, Xinjiang, China. Journal of Arid Land, 4(4), 441–449. https://doi.org/10.3724/SP.J.1227.2012.00441

Allbed, A., Kumar, L., & Aldakheel, Y. Y. (2014). Assessing soil salinity using soil salinity and vegetation indices derived from ikonos high-spatial resolution imageries: Applications in a date palm dominated region. Geoderma, 230–231, 1–8. https://doi.org/10.1016/j.geoderma.2014.03.025

Anderson, W., & Johnson, T. (2016). Evaluating global land degradation using ground-based measurements and remote sensing. Asfaw, E., Suryabagavan, K. V., & Argaw, M. (2018). Soil salinity modeling and mapping using remote sensing and GIS: The case of Wonji sugar cane irrigation farm, Ethiopia. Journal of the Saudi Society of Agricultural Sciences, 17(3), 250–258. https://doi.org/10.1016/j.jssas.2016.05.003

Bannari, A., Guedon, A. M., El-Harti, A., Cherkaoui, F. Z., & 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(19–20), 2795–2811. https://doi.org/10.1080/00103620802432717

Basso, F., Bove, E., Dumontet, S., Ferrara, A., Pisante, M., Quaranta, G., & Taberner, M. (2000). Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: An example covering the Agri basin (Southern Italy). Catena, 40(1), 19–35. https://doi.org/10.1016/S0341-8162(99)00062-4

Camera, C., Zomeni, Z., Noller, J. S., Zissimos, A. M., Christoforou, I. C., & Bruggeman, A. (2017). A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. Geoderma, 285, 35–49. https://doi.org/10.1016/j.geoderma.2016.09.019

Chuangye, S., Hongxu, R., & Chong, H. (2016). Estimating soil salinity in the yellow river delta, eastern china—an integrated approach using spectral and terrain indices with the generalized additive model. Pedosphere, 26(5), 626–635. https://doi.org/10.1016/S1002-0160(15)60071-6

Darwish, T., & Fadel, A. (2017). Mapping of soil organic carbon stock in the Arab countries to mitigate land degradation. Arabian Journal of Geosciences, 10(21), 474. https://doi.org/10.1007/s12517-017-3267-7

Dhar, R. B., Chakraborty, S., Chattopadhyay, R., & Sikdar, P. K. (2019). Impact of Land-Use/Land-Cover Change on Land Surface Temperature Using Satellite Data: A Case Study of Rajarhat Block, North 24-Parganas District, West Bengal. Journal of the Indian Society of Remote Sensing, 47(2), 331–348. https://doi.org/10.1007/s12524-019-00939-1

Dor, E. B., Ong, C., & Lau, I. C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma, 245, 112–124. https://doi.org/10.1016/j.geoderma.2015.01.002

Douaoui, A. E. K., Nicolas, H., & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. Geoderma, 134(1–2), 217–230. https://doi.org/10.1016/j.geoderma.2005.10.009

Du, J., Cheng, L., Zhang, Q., Yang, Y., & Xu, W. (2019). Different Flooding Behaviors Due to Varied Urbanization Levels within River Basin: A Case Study from the Xiang River Basin, China. International Journal of Disaster Risk Science, 10(1), 89–102. https://doi.org/10.1007/s13753-019-0195-4

Essahlaoui, A., Teodoro, A. C., & Mohajane, M. (2019). Modeling and mapping of soil salinity in Tetifalel plain (Morocco). Arabian Journal of Geosciences, 12(2), 35. https://doi.org/10.1007/s12517-018-4202-2

Fernandez-Buces, N., Siebe, C., Cram, S., & Palacio, J. L. (2006). Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico. Journal of Arid Environments, 65(4), 644–667. https://doi.org/10.1016/j.jaridenv.2005.08.005

Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D., Pozza, L. E., Ugbaje, S. U., . . . Bishop, T. F. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precision Agriculture, 1–15. https://doi.org/10.1007/s11119-018-09628-4

Fu, C. B., Xiong, H. G., & Tian, A. H. (2019). Study on the effect of fractional derivative on the hyperspectral data of soil organic matter content in arid region. Journal of Spectroscopy, (2019, 1–11. https://doi.org/10.1155/2019/7159317

Gao, X., Cheng, W., Wang, N., Liu, Q., Ma, T., Chen, Y., & Zhou, C. (2019). Spatio-temporal distribution and transformation of cropland in geomorphologic regions of China during 1990–2015. Journal of Geographical Sciences, 29(2), 180–196. https://doi.org/10.1007/s11442-019-1591-4

Gong, P., & Yu, B. (2001). Conifer species recognition: Effects of data transformation. International Journal of
Singh, D. K., Mishra, V. D., Gusain, H. S., Gupta, N., & Singh, A. K. (2019). Geo-spatial Modeling for Automated Demarcation of Snow Avalanche Hazard Areas Using Landsat-8 Satellite Images and In Situ Data. *Journal of the Indian Society of Remote Sensing*, 47(3), 513–526. https://doi.org/10.1007/s12524-018-00936-w

Song, D. S., & Grejner-Brzezinska, D. A. (2009). Remote sensing of atmospheric water vapor variation from gprs measurements during a severe weather event. *Earth Planets & Space*, 61(10), 1117–1125. https://doi.org/10.1186/BF03352964

Sripada, R. P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal*, 98(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Sun, J. (1995). A correlation principal component regression analysis of NIR data. *Journal of Chemometrics*, 9(1), 21–29. https://doi.org/10.1002/cem.1180090104

The National Soil Survey Office. (1998). *Soils of China*. China Agriculture Press.

Tilley, D. R., Ahmed, M., Son, J. H., & Badrinarayanan, H. (2007). Hyperspectral reflectance response of freshwater macrophytes to salinity in a brackish subtropical marsh. *Journal of Environmental Quality*, 36(3), 780–789. https://doi.org/10.2134/jeq2005.0327

Tripathi, N. K., Rai, B. K., & Dwivedi, P. (1997, October). Spatial modeling of soil alkalinity in GIS environment using IRS data. In *The 18th Asian Conference on Remote Sensing*, Kualalumpur.

Wang, F., Chen, X., Luo, G., Ding, J., & Chen, X. (2013). Detecting soil salinity with arid fraction integrated index and salinity index in feature space using landsat tm imagery. *Journal of Arid Land*, 5(3), 76–89. https://doi.org/10.14928/CNKI:SUN:GHXX.0.2013-03-010

Wang, F., Chen, X., Luo, G., & Han, Q. (2015). Mapping of regional soil salinities in Xinjiang and strategies for amelioration and management. *Chinese Geographical Science*, 25(3), 321–336. https://doi.org/10.1007/s11769-014-0718-x

Wang, H., & Jia, G. (2012). Satellite-based monitoring of decadal soil salinization and climate effects in a semi-arid region of China. *Advances in Atmospheric Sciences*, 29(5), 1089–1099. https://doi.org/10.1007/s00376-012-1150-8

Wang, X, Zhang, F, Ding, J, Kung, H-T, Latif, A, & Johnson, V. C. (2018). Estimation of soil salt content (ssc) in the ebinur lake wetland national nature reserve (elwnnr), northwest china, based on a bootstrap-hp neural network model and optimal spectral indices. *Science Of The Total Environment*, 615(12), 918. doi:10.1016/j.scitotenv.2017.10.025

Wang, Z. H., He, X. L., Yang, G., & Li, S. (2010). Sustainable strategies and status of unconventional water resources development and utilization in manas river basin. *Advanced Materials Research*, 113-116, 565–570. https://doi.org/10.4028/www.scientific.net/AMR.113-116.565

Wiegand, C. L., Everitt, J. H., & Richardson, A. J. (1992). Comparison of multispectral video and SPOT-1 HRV observations for cotton affected by soil salinity. *International Journal of Remote Sensing*, 13(8), 1511–1525. https://doi.org/10.1080/01431169208904205

Xiao, Y., & Ouyang, Z. (2019). Spatial-temporal patterns and driving forces of water retention service in China. *Chinese Geographical Science*, 29(1), 100–111. https://doi.org/10.1007/s11769-018-0984-0

Yan, J., & Guo, J. (2019). Extraction and Analysis of Natural Vegetation Cover and Soil Moisture Using Landsat TM Data in the Arid Oasis of Xinjiang, China. *Journal of the Indian Society of Remote Sensing*, 47(2), 213–221. https://doi.org/10.1007/s12524-018-0922-8

Yu, H., Mingyue, L., Baojia, D., Zongming, W., Liangjun, H., & Bai, Z. (2018). Mapping soil salinity/sodicity by using landsat oli imagery and pls algorithm over semi-arid west jilin province, china. *Sensors*, 18(4), 1048. https://doi.org/10.3390/s18041048

Zhang, T., Zhao, G., Chang, C., Wang, Z., & Jia, J. (2015). Information extraction method of soil salinity in typical areas of the yellow river delta based on landsat imagery. *Agricultural Sciences*, 6(1), 71–77. https://doi.org/10.4236/as.2015.61006

Zhang, Y., Sui, B., Shen, H., & Ouyang, L. (2019). Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Computers and Electronics in Agriculture*, 160, 23–30. https://doi.org/10.1016/j.compag.2019.03.015