Remote Sens. **2013**, *5*, 4533-4550; doi:10.3390/rs5094533

OPEN ACCESS

Remote Sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

A Study of Soil Line Simulation from Landsat Images in Mixed Grassland

Dandan Xu * and Xulin Guo

Department of Geography and Planning, University of Saskatchewan, 117 Science Place, Saskatoon, SK S7N5C8, Canada; E-Mail: xulin.guo@usask.ca

* Author to whom correspondence should be addressed; E-Mail: xudd0621@gmail.com; Tel.: +86-130-6966-1488.

Received: 26 June 2013; in revised form: 20 August 2013 / Accepted: 6 September 2013 / Published: 12 September 2013

Abstract: The mixed grassland in Canada is characterized by low to medium green vegetation cover, with a large amount of canopy background, such as non-photosynthetic vegetation residuals (litter), bare soil, and ground level biological crust. It is a challenge to extract the canopy information from satellite images because of the influence of canopy background. Therefore, this study aims to extract a soil line, a representation of bare soil with litter and soil crust in the surface, from Landsat images to reduce the background effect. Field work was conducted in the West Block of Grasslands National Park (GNP) in Canada, which represents the northern mixed grassland from late June to early July 2005. Six TM images with either no or only a small amount of cloud content were collected in 2005. In this study, soil lines were extracted directly from images by quantile regression and the (R, NIR_{min}) method. The results show that, (1) both cloud and cloud shadow have obvious influence on simulating soil line automatically from images; (2) green up and late senescence seasons are relatively better for soil line simulation; (3) the (R, NIR_{min}) method is better for soil line simulation than quantile regression to extract green biomass or green cover information.

Keywords: Remote sensing; mixed grassland; soil line; quantile regression; PVI; ground cover

1. Introduction

Grassland ecosystems, characterized by high amounts of carbon stock, are one of the main terrestrial ecosystems besides forest, tundra, and cropland ecosystems, which provides multiple habitats for wildlife [1]. More importantly, grasslands provide a variety of functions without requiring fertilizer input, such as high productivity, site stability, capture and beneficial release of water [2], grazing, and recreation. For grassland monitoring, biomass is an indicator of grassland productivity and grazing capacity while playing an important role in understanding ecosystem response to disturbances such as climate change, grazing intensity, and extreme weather events (flood and drought) [3]. Vegetation indices extracted from remotely sensed data have a high correlation with green biomass and green cover [4–8], which have the advantages of large scale biomass and cover estimation and long term biomass and cover change evaluation. However, soil along with ground level biological crust and covered vegetation residuals (litter) presents a problem to the interpretation of vegetation indices [9].

Richardson and Wiegand [10] first introduced the concept of soil line and perpendicular vegetation index (PVI) to reduce interruption from the soil background when extracting vegetation signals from Landsat MSS images. Soil line is a line which shows the linear relationship between near-infrared (NIR) reflectance and red (R) reflectance of bare soil [11]. Theoretically, the main factor that characterizes soil line is soil type because of variation in organic matter and chemical components. In addition, soil line is also influenced by soil surface condition (e.g., roughness and vegetation residuals) [12,13]. However, soil line does not vary from soil brightness (causing by soil moisture [14] and roughness [15]) without soil type changing. Soil line concept has been widely accepted to interpret remotely sensed imagery [12], especially to normalize soil background effect for vegetation discrimination [16]. Some vegetation indices are designed based on soil line, such as PVI, Adjusted Transformed Soil-Adjusted Vegetation Index (ATSAVI) [17] and Transformed Soil Adjusted Vegetation Index (TSAVI) [18] to minimalize the effect of soil background. Studies based on soil line are also in estimation of fraction cover [19], vegetation residual cover [20], soil organic matter [21,22] and soil degradation [23]. Besides, because soil line is relatively stable for a certain soil type, Jaishanker, Thomaskutty, Senthivel and Sridhar [14] applied soil line transformation method to relative atmospheric correction and Stabile and Searcy [24] indicated that the crop parameter can be compared directly from multi-temporal images after the images are normalized to a reference soil line.

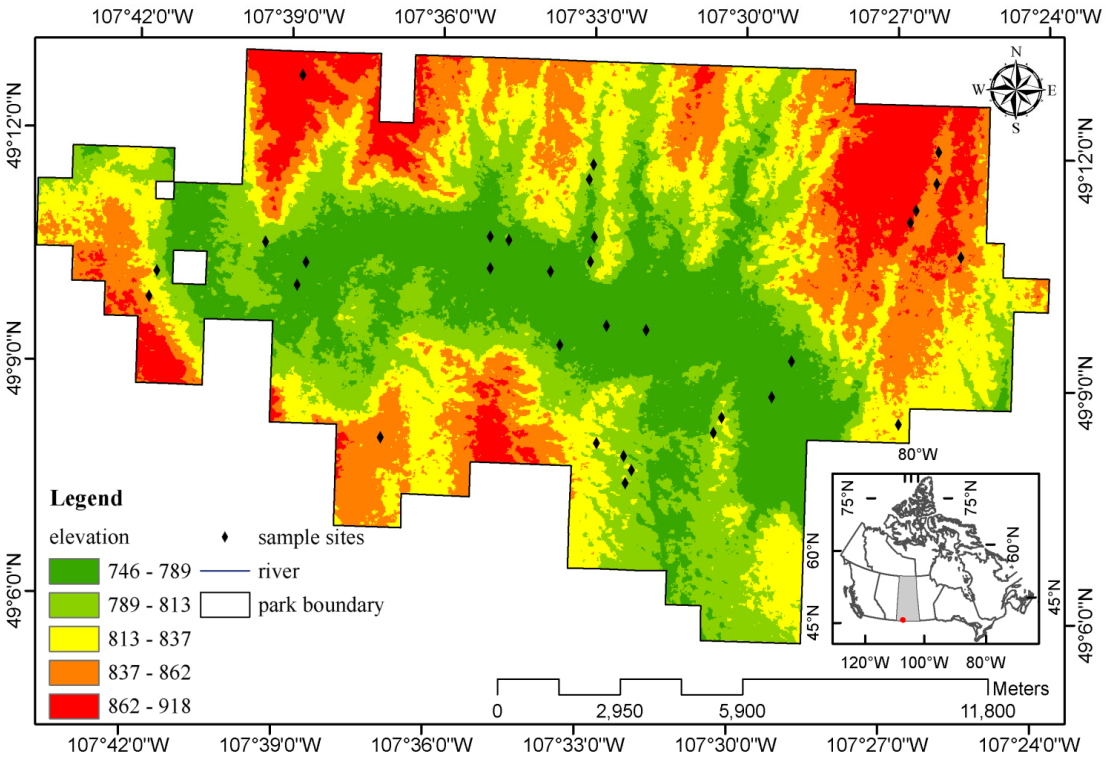
Commonly, researchers conducted indoor experiments [12], simulated models (for example, radioactive simulation model) [25], or tested the reflectance of soil samples from field work to obtain soil lines. Fox, Sabbagh, Searcy and Yang [11] developed an automated method for extracting soil lines from remotely sensed images in cropland that fits a linear regression by deriving a set of minimum near-infrared digital numbers across R and NIR bands while removing contradicting soil line pixels with an iterative process. Unlike cropland, grassland which is long term conserved has more soil crust (moss and lichen) and litter coverage and more heterogeneous vegetation structure. Since conserved grassland is more complex than grassland, this study aims to develop a suitable method to extract the soil line from Landsat images automatically. The objectives of this study are as follows, (1) to examine the ability of quantile regression and (R, NIR_{min}) method for extracting soil lines from a two dimensional scatter plot of NIR and R bands; (2) to test the best time period for extracting soil lines linked with vegetation phenological sections; (3) to evaluate the cloud effect on extracting soil

lines; (4) to compare soil lines extracted directly from imagery and the soil line obtained from reflectance of dug soil samples.

2. Study Area

The study area was the West Block of Grasslands National Park (GNP) (49°N, 107°W), located in the southern part of Saskatchewan, Canada (Figure 1). The west block has an area of approximately 521 km^2. The annual mean temperature and total precipitation of this area are 3.4 °C and 340 mm, respectively [26]. GNP falls within the semi-arid mixed grassland ecosystem [27] which contains upland, sloped, and valley grasslands. The uplands are dominated by needle-and-thread grass (*Stipa comata Trin. & Rupr.*), blue grama grass (*Bouteloua gracilis (HBK*) *Lang. ex Steud.*), and western wheatgrass (*Agropyron smithii Rydb.*). The valley grasslands are dominated by western wheatgrass and northern wheatgrass (*Agropyron dasystachym*) along with higher densities of shrubs and occasional trees [28]. The sloped grasslands consists of vegetation species from both upland and valley grasslands.

Figure 1. Sample sites distribution map in the west block of Grasslands National Park (GNP).



The main soil types in the park region are chernozemic and solonetzic soils [29]. Chernozemic soil is the most common soil type in grasslands, characterized by a dark color and high amount of organic content. Comparatively, solonetzic soils, with high salinity and lighter color, are formed due to drought and high evaporation conditions. The study area has been conserved for over 20 years [30]. As a result, grasslands in this region contain large amounts of non-photosynthetic vegetation residuals including litter and standing dead materials. In addition, under the low to medium cover of vegetation canopy, most of the surface is covered by microphytic communities of small non-vascular plants, mainly including mosses and lichens, which form biological crusts.

3. Materials

3.1. Field Data

Field work was conducted in late June and early July of 2005; the maximum growing season of the northern mixed prairie. Twenty-six sites (Figure 1) were selected based on a stratified random design and accessibility with 8 sites in upland, 4 sites in sloped land, and 14 sites in valley grassland. Two 100 m × 100 m plots were set up in each site. Each plot was composed of two 100 m transects perpendicularly crossed with each other in their centers. Twenty 50 cm × 50 cm quadrats were set up in each plot at 10 m intervals, excluding the center. Ground percentage covers, such as grass, forb, shrub, standing dead material, litter, lichen, moss, bare soil, and rock, were estimated by observation in each quadrat. Due to restrictions in the park, biomass was collected at 20 m intervals with a 20 cm × 50 cm quadrat using the harvesting method. Fresh biomass clipped with scissors was sorted into four groups; grass, forb, shrub, and dead materials. Biomass was then measured by weight after they were dried in the oven for 48 hours at 60 °C. Soil samples were collected by digging the soil in the quadrats after the biomass was removed. The reflectance of soil samples was examined in laboratory conditions. To correlate with data from satellite imagery, all the biomass data, ground percentage, and soil reflectance was averaged for each individual site. Soil reflectance and dead material reflectance of R and NIR bands were averaged by R band location and NIR band location of Landsat images, respectively.

3.2. Satellite Images

Table 1 contains the information of all the Landsat images which were downloaded from the United States Geological Survey (USGS) website, including six TM images with no or small amounts of cloud content. All the images acquired from the USGS website had already been geometrically corrected, so image preprocessing only included atmospheric correction. Atmospheric correction was conducted in PCI Geomatica using the ATCOR2 algorithm. After atmospheric correction, all the images were clipped into the study area. The spatial resolution of all the images was 30m × 30m and the projection was WGS_1984_UTM_Zone_13N.

Table 1. Acquisition dates and cloud cover of Landsat TM images.

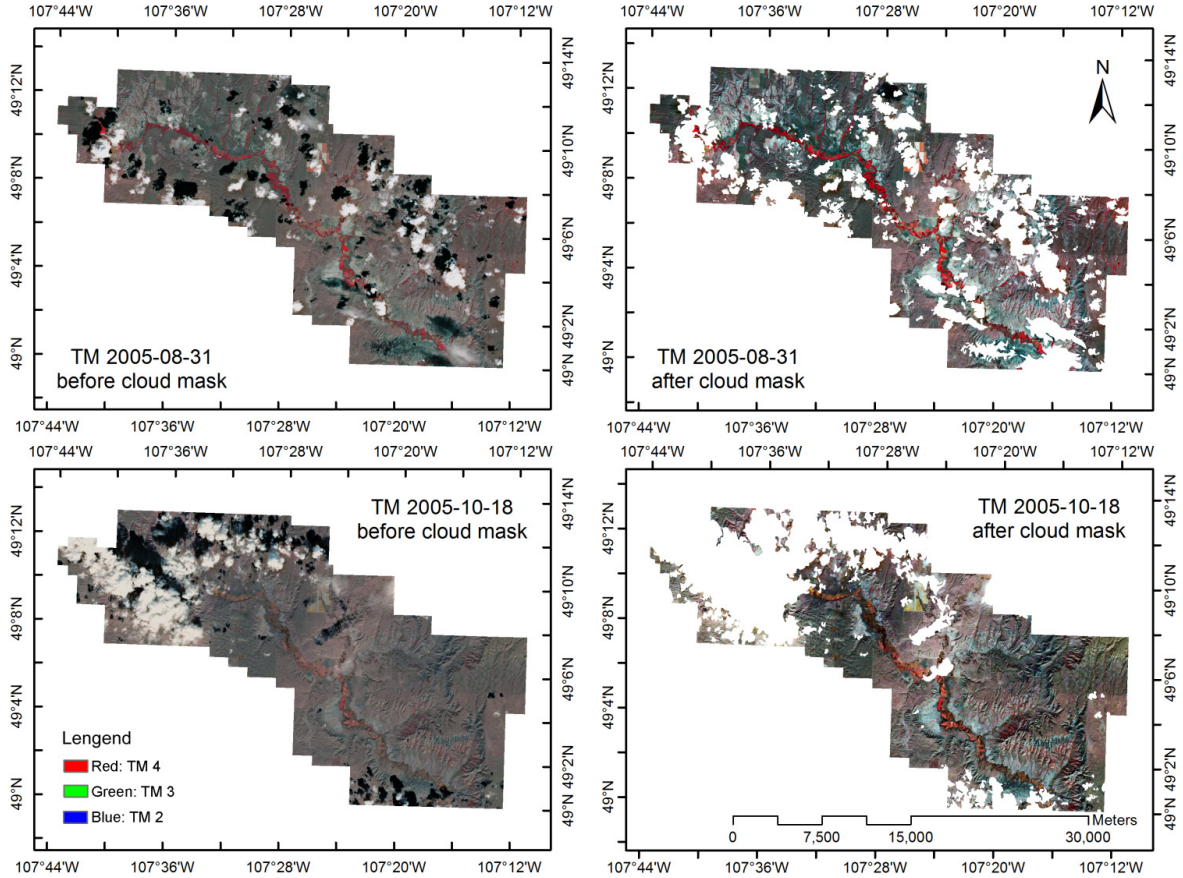
ID	File Name	Date	Sensors	Cloud%
1	LT50370262005115PAC01	25 April 2005	TM	0
2	LT50370262005131PAC01	11 May 2005	TM	0
3	LT50370262005195PAC01	14 July 2005	TM	0
4	LT50370262005211PAC01	30 July 2005	TM	0
5	LT50370262005243PAC01	31 August 2005	TM	24.1
6	LT50370262005291PAC01	18 October 2005	TM	27.5

Clouds in the two cloudy images were masked manually (Figure 2) in ArcGIS and percentage of cloud content in Table 1 was calculated by "(1-area of cloud masked image/area of west block) × 100%".

Twenty three MODIS vegetation images with 16 days composites and 250m × 250m resolution (MOD13Q1) were also acquired from the USGS website which were used for testing the vegetation phenology. Normalized Difference Vegetation Index (NDVI) images, registered to a Universal

Transverse Mercator (UTM) projection, were acquired by processing MODIS images using the MODIS Reprojection Tool.

Figure 2. Images before and after cloud mask.



4. Methods

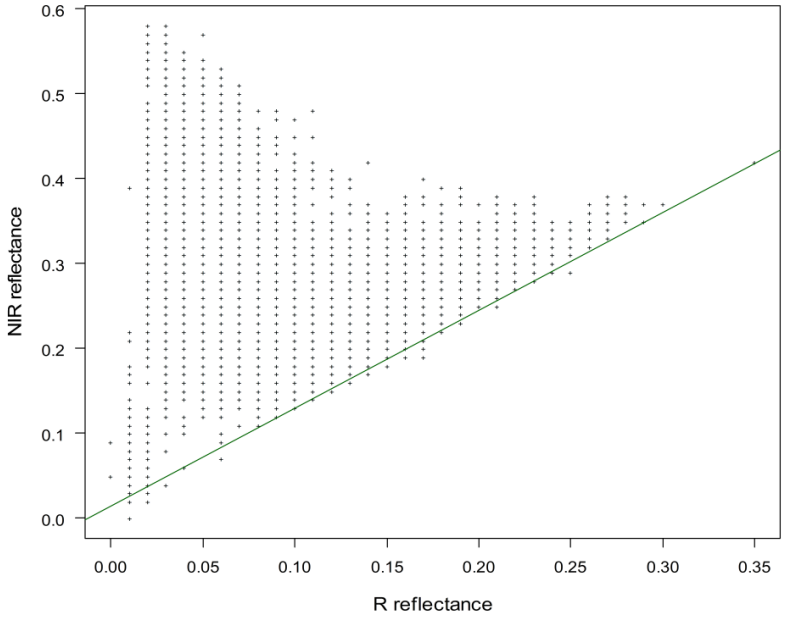
The two dimensional scatterplot of NIR band and R band is a fan shape scatter plot (Figure 3). The bottom line of this fan shape represents the soil line [11]. In this study, two statistical methods, discussed in Sections 4.1 and 4.2, were chosen to quantify this line.

4.1. Linear Regression of a Set of (R, NIR_{min})

The direct way to extract the bottom line of scatter plot is to fit the linear regression of a set of points characterized by minimum NIR value within the certain R values, which is named the (R, NIR_{min}) "method" in this paper. Fox, Sabbagh, Searcy and Yang [11] introduced an automated soil line identification routine based on the basic idea of the (R, NIR_{min}) method. The linear regression of a set of (R, NIR_{min}) was conducted in R software. Firstly, the whole dataset was divided into several datasets with 0.005 intervals of R reflectance ($0 < R \leq 0.005$, $0.005 < R \leq 0.01$, $0.01 < R \leq 0.015 \ldots 0.630 < R \leq 0.635$). Secondly, one point with minimal NIR value (R, NIR_{min}) was selected in each divided dataset. Finally, a general linear model was fitted for those points (R, NIR_{min}) (Figure 3). The dataset for Figure 2 and Figure 3 were the R and NIR reflectance of TM 2005-07-14 within the study area which were rounded

off to two decimal places (This transformed reflectance was not used in any analysis of this study and it was just used here to show the basic concept of the two methods).

Figure 3. Linear regression of a set of (R, NIR_{min}).



4.2. Quantile Regression

After it was first introduced by [31], quantile regression is gradually becoming a comprehensive linear and non-linear regression method [32]. Instead of focusing on the mean, quantile regression fits the linear or non-linear regression on the selected quantile (0–1) of the distribution of the response variable [33–35]. Because quantile regression fits regression curves to part of the distribution, this method has proved its ability to extract boundary lines in two-dimensional scatter plots [36,37]. In this study, quantile regression was conducted in R software for extracting the bottom boundary line in NIR *vs.* R scatter plots (Figure 4). Instead of setting the quantile to zero for extracting the bottom line in this study, the quantile was set by "the number of points below the bottom line" divided by "total number of points". In most situations, the fan shape scatter plot showed a clear, straight bottom boundary line with some points below this bottom boundary. By setting the quantile to zero in those situations, all the points would be above the simulated line. Therefore, extracting the bottom line from the scatter plots with some outliers with lower NIR reflectance, the quantile of quantile regression was set as a number close to zero but not zero, for example, 0.00001.

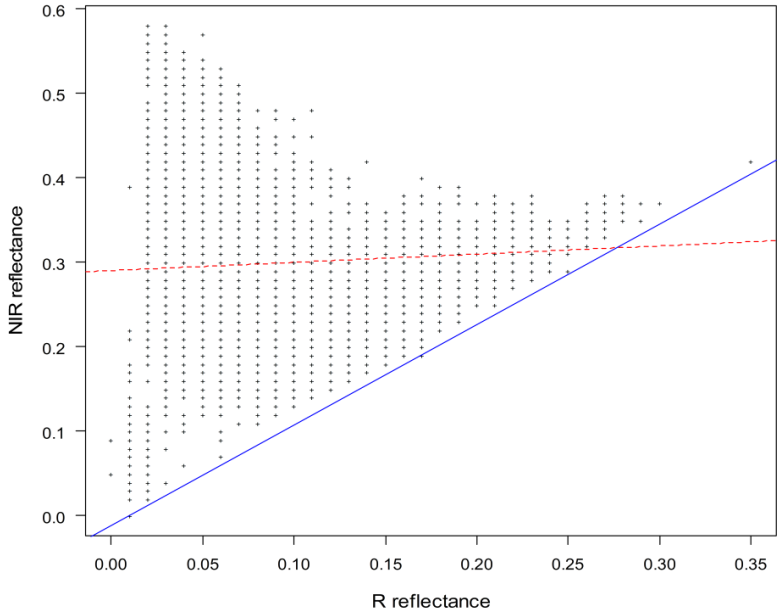
4.3. Vegetation Phenological Stages

Besides investigating the method for extracting soil lines, examining when the best time period to identify soil lines will also be tested. Grassland composition (the percentage of green grass, bare soil and dead material) varies from vegetation phenological stages. As a result, the structure of the scatter plots change dramatically through different vegetation phenological phases.

The three main vegetation phenological stages are green-up, maturity, and senescence. The vegetation phenology was assessed by the curvature-change rate method which was developed by

Zhang *et al.* [38]. This method has the ability of vegetation phenological stage examination at large scales by using MODIS vegetation index products. In this study, the mean NDVI of the whole study area was obtained by zonal statistics in ArcGIS software.

Figure 4. Quantile regression (the blue solid line is 0.1% quantile regression line; the red dash line is linear regression for the whole dataset and it is also the 50% quantile regression).



4.4. Validation of Soil Line

Vegetation indices (Table 2, [10,17,18,39]), based on soil line parameters (ATSAVI, TSAVI and PVI), were used to compare soil lines extracted from this study with the soil lines estimated from bare soil reflectance. These vegetation indices were also used to evaluate the soil lines extracted from this study by comparing their ability to extract green biomass and green cover through NDVI. To correlate with the field data collection date, only TM 2005-07-14 was used to calculate the vegetation indices.

Table 2. Vegetation indices based on soil line parameters and Normalized Difference Vegetation Index (NDVI).

Vegetation Index	Algorithm	Citation
ATSAVI	$ATSAVI = \frac{a^*(NIR - aR - b^*)}{aNIR + R - ab + X(1+a^2)}$	[17]
TSAVI	$TSAVI = \frac{a(NIR - aR - b)}{aNIR + R - ab}$	[18]
PVI	$PVI = \frac{NIR - aR - b}{\sqrt{1+a^2}}$	[10]
NDVI	$NDVI = \frac{NIR - R}{NIR + R}$	[39]

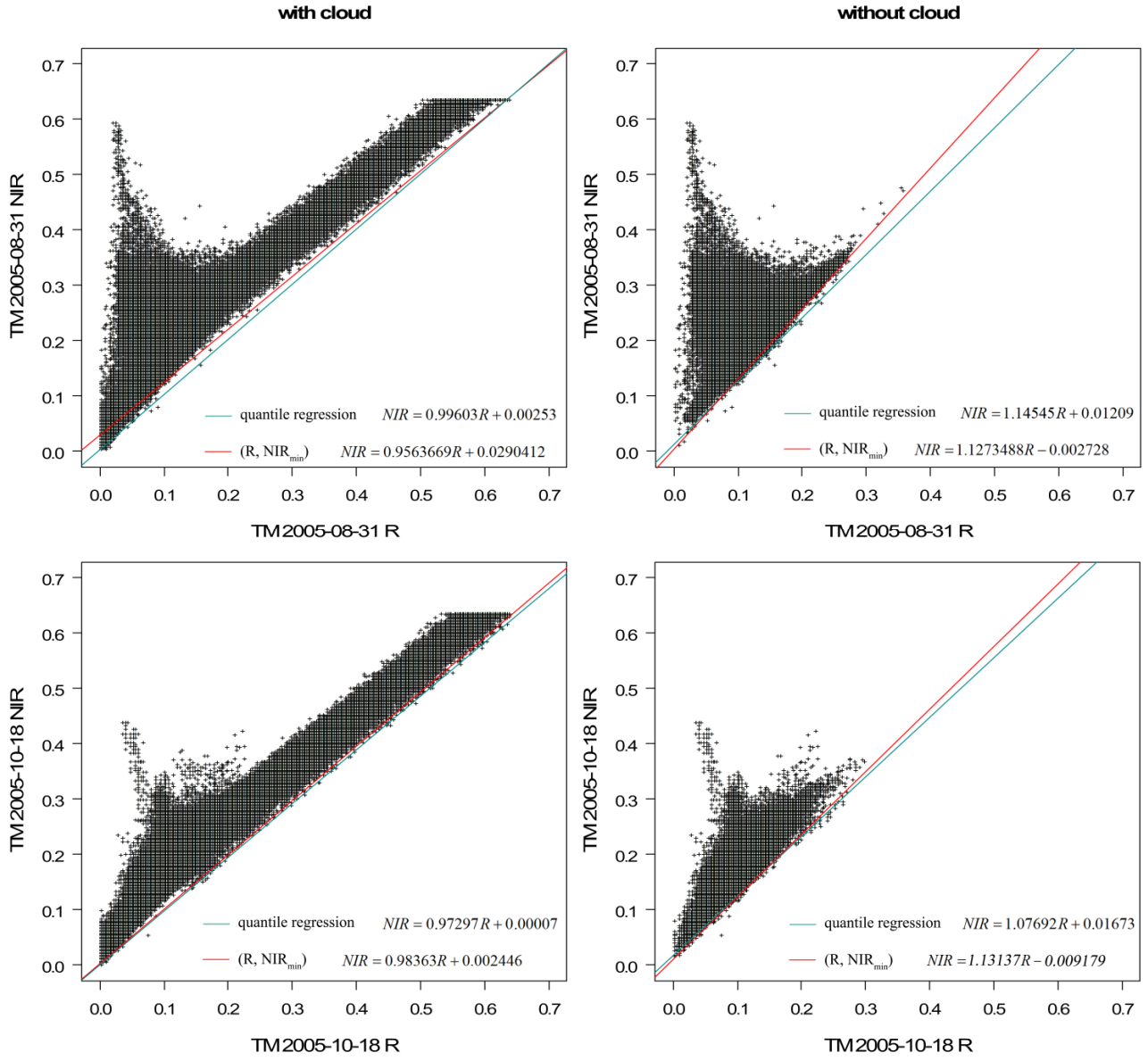
a^* means the slope of soil line; b means the intercept of soil line; X = 0.08.

5. Results and Discussion

5.1 Cloud Effect on Soil Line Extraction by Two Methods

In Figure 5, cloud pixels have high R and NIR value and cloud shadow pixels have low R and NIR value in the scatterplots of cloudy pixels. The cloud and cloud shadow effect of soil line by both methods are really obvious. For both images, cloud and cloud shadow lessen the slope of extracted soil lines. Cloud and cloud shadow usually have similar reflectance in both R and NIR bands, which means the points representing cloud and cloud shadow are in line "NIR = R" in the scatter plot of two dimension with NIR and R bands. In this case, both images actually have soil line slopes over 1 (Figure 5 without cloud), so the consequence of cloud effects is a lower slope for soil lines. Otherwise, if the slope of actual soil line is lower than 1, then the cloud effect would result in a higher slope. In addition, cloud and cloud shadow pixels cause more difference of both slope and intercept of soil lines with the (R, NIR_{min}) method than that with quantile regression.

Figure 5. Extracted soil lines of TM images before and after cloud masked.

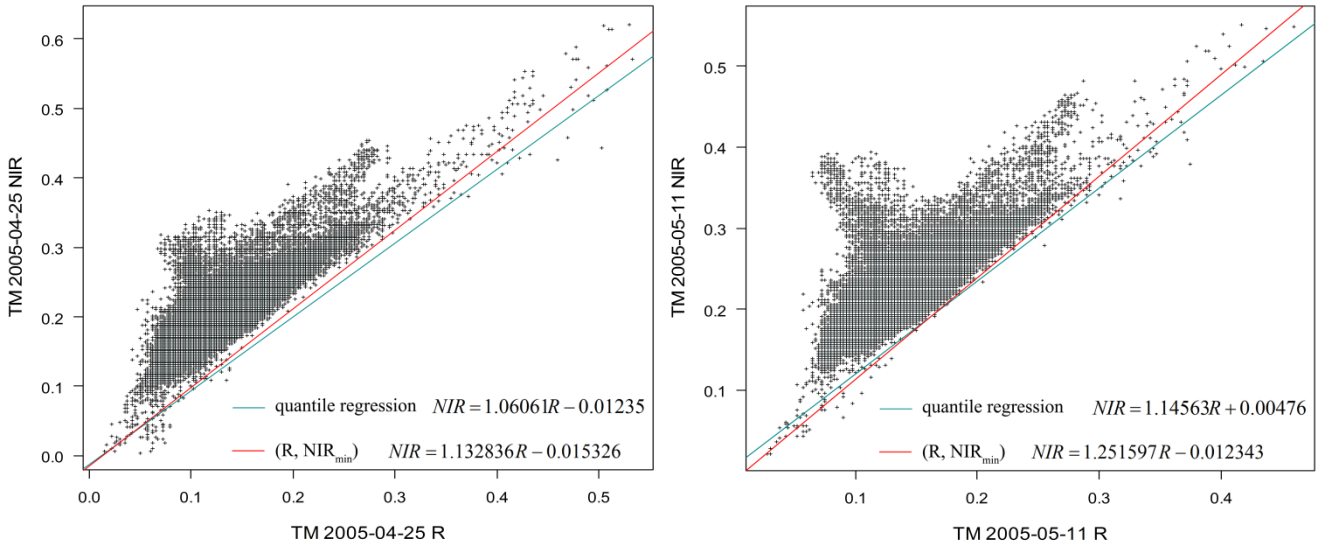


5.2 Soil Line Extracted by the Two Methods

From the result of vegetation phenological stage testing for 2005, vegetation started to green up in the middle of April, became mature in middle and late June, and senesced through early July to middle October.

The scatter plots images, in the green up section, contain some sparse points in both high and low R value parts (Figure 6) which poses challenges for extracting soil lines from those images. In the scatterplots, the bottom boundary is clearly a straight line excluding some sparse points. The soil lines extracted by the (R, NIR_{min}) method are closer to the straight bottom line than that by quantile regression. It has been exhibited previously that quantile regression is influenced by the outliers in the X axis [36]. With the two images about 15 days different (Figure 6), the different slope and intercept extracted from quantile regression is much higher than that extracted from the (R, NIR_{min}) method, which also means the sparse points below the bottom line cause some challenge for quantile regression to extract soil line. Especially, the scatterplot of TM 2005-04-25 (Figure 6) with more sparse points below the bottom line causes more difficulty for quantile regression to capture the bottom line.

Figure 6. Soil lines by the two methods in the grass green up period.



TM 2005-07-14 is not exactly in the grass mature period, but it is near the mature period (Figure 7). Grass mature period has a scatterplot with relatively higher NIR reflectance in low R value region because those points represent mature grass. Furthermore, the bottom line of the grass mature period scatterplot is a banana shaped curve. This boundary shape is caused by the relatively higher NIR value in low R value region during maturity which clearly shows the grass growth when comparing to green-up and senescence. In this case, the (R, NIR_{min}) method fits a soil line with a higher intercept than quantile regression. Quantile regression actually fits a tangent line of the bottom curve in the position of R value around 0.18.

On 30 July 2005, grass has already started to senesced. The points with lower R and NIR value represent the senesced grass (Figure 8). In this early senescence period, the bottom of the scatterplot forms two straight bottom lines. The lower straight line shows the senesced grass. For the (R, NIR_{min}) method, the dataset for the final linear regression of soil lines includes points with much higher NIR

reflectance a in high R value region which causes the regression line upward. In this case, quantile regression shows its ability to simulate the lower bottom line. Even in the middle senescence period (Figure 5: TM 2005-08-31 without cloud), the scatterplot still has this unique signature with two straight bottom lines and quantile regression still shows its ability to capture the lower bottom line. In the late senescence period (Figure 5: TM 2005-10-18 without cloud), the scatterplot has a clear straight line. In this case, both quantile regression and the (R, NIR_{min}) method fit the bottom line.

Figure 7. Soil lines by the two methods near the grass mature period.

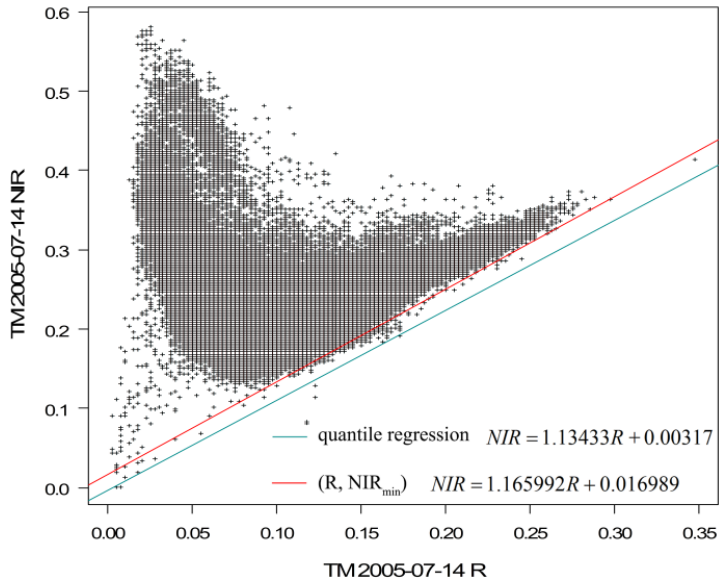
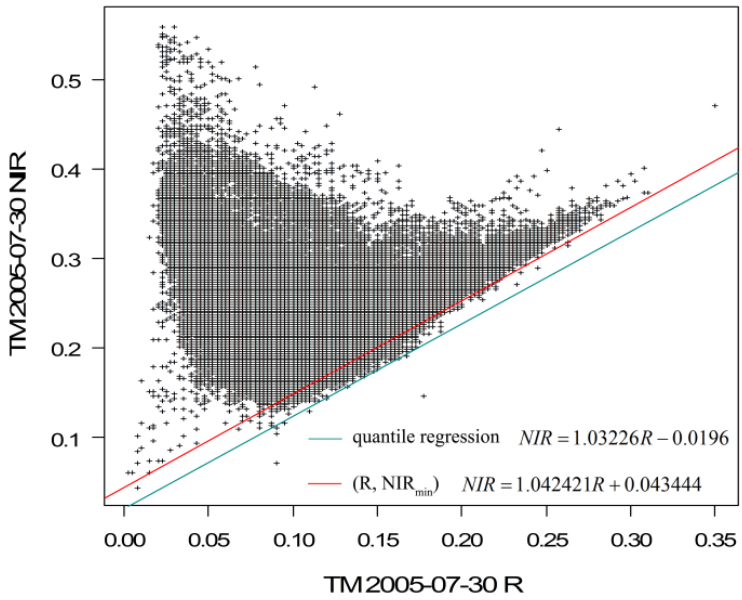


Figure 8. Soil lines by the two methods in the early senescence period.



The ability of both statistic methods for bottom line extraction varies from vegetation phenology because the bottom lines change through vegetation phenological cycle. For the scatterplots in the late senescence stage with clear straight bottom line, both methods simulate the bottom lines quit well. In the early green up stage, the soil lines extracted by quantile regression do not precisely fit the bottom

line because of the influence of the sparse points in the parts with very low R value and very high R value. Quantile regression is based on the quantile distribution of the response variable, NIR reflectance in this case. Therefore, the density or the distribution of the scatter plot will influence the quantile regression results, which is the reason the sparse points with low R and NIR values and those with high R and NIR values affect the quantile regression results in early green-up. In the maturity period, it is hard to capture the bottom straight line for both methods because the bottom line of the mature stage scatterplot is actually a curve. With unique characteristic of two straight bottom lines for the early and middle senescence period, quantile regression captures the lower bottom line, which represents the senesced grass rather than the (R, NIR_{min}) method. In this study, quantile regression is more likely to simulate the boundary distribution than the (R, NIR_{min}) method. However, the two methods have different abilities in different circumstances. It is hard to tell if quantile regression is a better method for simulating soil lines than the (R, NIR_{min}) method, but quantile regression is better to catch the boundary characteristics of fan shape scatter plots in most situations. Actually, both methods have similar results when the bottom boundary of the fan shape scatter plot is clearly a straight line (Figure 5: TM 2005-10-18 with cloud) or a straight line with very few outliers (Figure 5: TM 2005-10-18 without cloud).

Theoretically, the quantile should be set to zero, but in reality, it is set to a number close to zero because not all the points are above the bottom boundary. Quantile setting certainly contribute to the quantile regression the most. For a large dataset in this study, slight change in quantile setting (e.g., from 0.00001 to 0.0000099) does not cause obvious difference of the soil line extraction. The basic idea of (R, NIR_{min}) method is selecting the points located in the bottom. There will be some omission and commission errors. In the certain range of R value, the point with minimum NIR value may locate upward the boundary. On the other hand, the points chosen by this method includes the sparse points under the bottom line, which influences the regression results. Fox, Sabbagh, Searcy and Yang [11] deleted some points to reduce the omission and commission errors based on the long distance between the points and the initial soil line, which statistically slightly improves the R^2 of the regression model.

5.3 Relatively Better Time Period for Extracting Soil Line

The slope and intercept of the extracted soil lines from the six images are in Table 3. In Table 3, "slope_min" and "intercept_min" mean the slope and intercept of soil lines extracted by (R, NIR_{min}) method; "slope_qr" and "intercept_qr" mean the slope and intercept of soil lines simulated by quantile regression; data is named by sensors name, date, and presence or absence of cloud. For example, "tm050425" means a TM image on 25 April 2005 and "tm051018_cf" means a cloud masked imagery of TM image on 18 October 2005.

Table 4 shows the correlation coefficients of green biomass and cover with the indices based on the extracted soil lines from this study. The coefficients of the biomass and cover with indices based on soil line of lab reflectance and NDVI were also presented in Table 4 as reference. All the coefficients are in significant level of P value less than 0.001. The "soil line names" in the table are named by extracted method and image acquired data. For example, "min_051018" means the soil line was

extracted from the TM image acquired in 18 October 2005 by the (R, NIR_{min}) method, while "qr_051018" means the soil line was extracted from the same image by quantile regression.

Table 3. Slope and intercept of extracted soil lines by the two methods.

Date	Data	Slope_Min	Intercep_min	Slope_qr	Intercept_qr
25-Apr-05	tm050425	1.133*	−0.015*	1.061*	−0.012
11-May-05	tm050511	1.252*	−0.012*	1.146*	0.005
14-Jul-05	tm050714	1.166*	0.017*	1.134*	0.003
30-Jul-05	tm050730	1.042*	0.043*	1.032*	0.020
31-Aug-05	tm050831_cf	1.273*	−0.003	1.145	0.012
18-Oct-05	tm051018_cf	1.131*	−0.009*	1.077	0.017

* means P value < 0.05

Table 4. Correlation coefficients of the indices based on the extracted soil lines with green biomass and cover.

Soil Line Names	Indices	r Green Biomass	r Green Cover	Soil Line Names	Indices	r Green Biomass	r Green Cover
min_051018	ATSAVI	0.600	0.741	min_050714	ATSAVI	0.589	0.707
	TSAVI	0.629	0.786		TSAVI	0.626	0.760
	PVI	0.406	0.403		PVI	0.437	0.458
qr_051018	ATSAVI	0.584	0.692	qr_050714	ATSAVI	0.596	0.725
	TSAVI	0.627	0.760		TSAVI	0.629	0.776
	PVI	0.346	0.302		PVI	0.409	0.408
min_050831	ATSAVI	0.601	0.746	min_050511	ATSAVI	0.602	0.753
	TSAVI	0.629	0.783		TSAVI	0.628	0.789
	PVI	0.501	0.583		PVI	0.492	0.563
qr_050831	ATSAVI	0.591	0.712	qr_050511	ATSAVI	0.595	0.724
	TSAVI	0.628	0.767		TSAVI	0.629	0.775
	PVI	0.419	0.426		PVI	0.420	0.427
min_050730	ATSAVI	0.539	0.601	min_050425	ATSAVI	0.602	0.747
	TSAVI	0.603	0.695		TSAVI	0.629	0.790
	PVI	0.301	0.230		PVI	0.408	0.406
qr_050730	ATSAVI	0.577	0.677	qr_050425	ATSAVI	0.600	0.737
	TSAVI	0.626	0.755		TSAVI	0.629	0.788
	PVI	0.287	0.208		PVI	0.325	0.268
lab soil line	ATSAVI	0.599	0.739				
	TSAVI	0.629	0.777				
	PVI	0.503	0.587				
	NDVI	0.630	0.778				

ATSAVI, based on the soil line simulated from 25 April image by both quantile regression and the (R, NIR_{min}) methods, slightly improved the correlation coefficient with green biomass than that based on soil line obtained from lab reflectance of soil samples.TSAVI remained the same correlation coefficient with green biomass (Table 4: min_050425, qr_050425).For the green biomass, ATSAVI,

based on soil lines extracted by the (R, NIR_{min}) method from images acquired in 11 May, 31 August, 18 October, slightly improved the coefficients than the ATSAVI using lab soil line parameters while TSAVI still kept the same coefficients (Table 4: min_050511, min_050831, min_051018). TSAVI calculated from the soil line, extracted from the 25 April image, by both quantile regression and the (R, NIR_{min}) method, improved the correlation coefficient with green cover than that calculated by lab soil line while ATSAVI improved the coefficient with green cover only by the (R, NIR_{min}) method (Table 4: min 050425, qr_050425). TSAVI and ATSAVI using the parameters from the soil line obtained from images acquired in 11 May, 31 August, and 18 October with only the (R, NIR_{min}) method improved the relation with green cover (Table 4: min_050511, min_050831, min_051018). For TM 2005-04-25, both methods were working quite well for improving relation between VIs and green cover/green biomass while only the (R, NIR_{min}) method actually enhanced the relation between VIs and green biomass/green cover for TM 2005-05-11, TM 2005-08-31 and TM 2005-10-18. The possible reason that quantile regression did not make the relation better as the (R, NIR_{min}) method did is that the soil line slopes and intercepts for TM 2005-08-31 and TM 2005-10-18 by quantile regression are not in the significant level which is $P < 0.05$ (Table 3). In the study area, the soil line parameters simulated from images acquired in early green up period (25 April and 11 May) and late senesced season (31 August and 18 October) improved the relationship between ATSAVI or TSAVI with green biomass and green cover, which means early green up and late senescence seasons are both relatively better seasons than late green up, mature and early senesced period for soil line simulation.

In this study, NDVI explains more variation for green biomass than other soil adjusted indices based on soil line, either from lab reflectance or simulations in this study. For green cover, NDVI still shows stronger relation than three soil adjusted indices based on the lab reflectance. However, TSAVI based on the soil lines of TM 2005-04-25 by both methods and TM 2005-05-11, TM 2005-08-31, TM 2005-10-18 by only the (R, NIR_{min}) method improved green cover extraction rather than NDVI and other VIs (Table 4). This means TSAVI based on the soil lines from early green up and late senescence enhanced the relation between NDVI and green cover. In this study, PVI is not suitable for either green biomass or green cover estimation. Based on the research of Yoshioka, *et al.* [40], the influence of soil background on NDVI decreases when Leaf Area Index (LAI) increases. The average LAI from the research conducted is higher than 1.0, which means the effects of soil background on NDVI is quite low. It is the possible reason why NDVI shows stronger relation with green biomass/green cover in most of the cases in the study area. It is probable in other regions with relatively lower vegetation cover, soil adjusted vegetation indices show much stronger relationship with biomass or green cover than NDVI.

5.4. Comparison of Soil Line Extracted from Images with Soil Line Extracted from Field Work

The intercept of soil lines from field work is much higher than that of soil line extracted from images (Figure 9). In Section 5.3, soil lines extracted from early green up period by both methods are better than the soil line calculated from lab reflectance of soil samples. As a result, reflectance of soil samples is much higher than actual surface soil reflectance with litter and biological crust on it in the study area. More importantly, both methods can obtain soil line automatically from imagery with better accuracy and less cost than soil line obtained by lab reflectance of soil samples.

In addition, soil samples that were dug after the biomass was removed have different soil structure from land surface soil. More importantly, the spectral reflectance of litter is much lower than that of bare soil (Figure 10). In the study area, which has been conserved for more than 20 years, the ground coverage of dead material is much larger than that of bare ground. Therefore, the reflectance of soil samples collected after the above ground vegetation, litter and soil crust removed is not a representative of actual soil surface.

Figure 9. Soil line extracted from images with two methods and the soil line from field work (The equation represent the soil line based on the lab reflectance of soil samples).

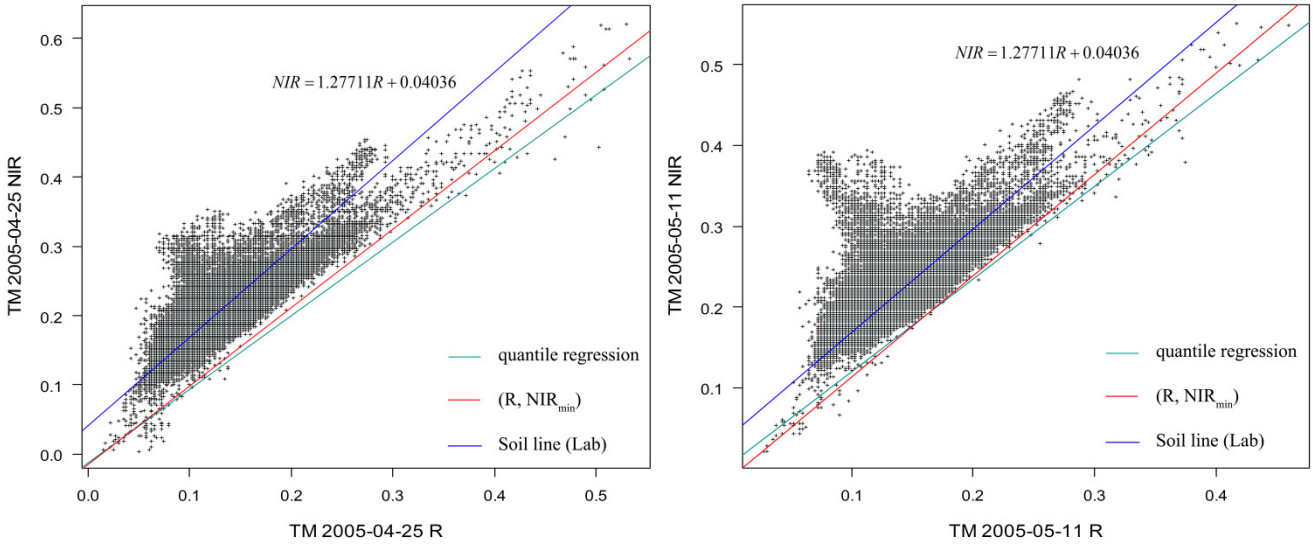
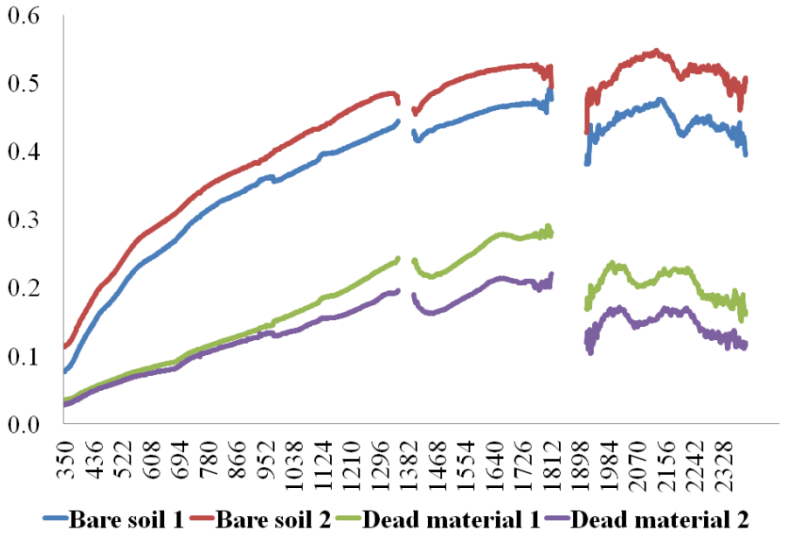


Figure 10. Comparison of bare soil reflectance and dead material reflectance.



The application of soil line is very wide in agriculture aspect, such as reducing soil background effects when extracting vegetation information [10,12,16–18], estimating vegetation fraction cover [19], vegetation residue cover (litter cover) [20], and atmospheric correction [24]. In the grassland, soil line is not widely used probably due to the availability of soil line. In cropland, the bare soil surface is normally covered by some vegetation residues, while in the grassland it is covered by plenty of litter

and soil crust (e.g., moss and lichen). When collecting reflectance of soil samples in cropland, the surface component is pretty much the same as the real surface structure. However, in grassland, collected soil samples are different from the real surface soil because the components and structure are all changed. In this study, it has been proved that soil line extracted automatically from images improved the soil line obtained from the reflectance of soil samples, which means the automatically extracted soil line from this study represent the actual soil line more closely. Hence, with the automatic soil line extraction methods provided by this study, the soil line of grassland can be obtained more accurately with less work. The ability of the two methods for extracting soil lines in cropland have not been tested to date, but definitely worth to be further studied in future research.

Uncertainty of the results of this study probably came from the green cover data because the green cover data was measured by observation (human sense of the researcher) from field work. However, the green cover data was checked by how strong the relationship is between green cover and LAI, which is measured by LAI-2000. In this study, the correlation coefficient between green cover and LAI is 0.724, which is acceptable. Another uncertainty from this study is that there are two soil types (chernozemic soil and solonetzic soil) in the study area. As Yoshioka, *et al.* [40] mentioned in their research, the general soil line of the local area caused slight difference for each soil type with an unique soil line. However, the positive aspect is that there are two soil types in this study area and chernozemic soil is the main soil type. More importantly, solonetzic are found where there are no or low vegetation cover. Thus, chernozemic soil is the only soil type under consideration for extracting green cover/green biomass, which is also the reason for simulating one soil line for each image of the study area. Applying the methods of this study for another study area with more dominated soil types, it has a possibility to improve the accuracy of soil line simulation directly from imagery by separating the different soil types.

6. Conclusions

1. Cloud effect is obvious with soil line simulation from imagery. In addition, cloud lessens the slope of soil line when the slope is larger than 1 while the influence of cloud on soil line actually enlarges the slope when it is less than 1.

2. Both early green up season and late senescence season are relatively better for soil line simulation from images than late green up, mature and early senesced period for soil line simulation. Because soil line does not have seasonal variation since soil type is the main factor that changes soil line. This conclusion is not about if soil line extracted from images in early green up season or late senescence season have better soil lines than soil lines from other seasons. This conclusion is to fill the gap where other studies proved that soil line could be extracted automatically from images but few studies few studies have focused on which season was better for soil line extraction since "near-infrared band" *vs*. "red band" scatter plots have large seasonal variations where as soil line does not.

3. For extracting green biomass or green cover information from Landsat images in mixed grassland, the (R, NIR_{min}) method is better for soil line simulation than quantile regression. Soil adjusted vegetation indices (Adjusted Transformed Soil-Adjusted Vegetation Index and Transformed Soil Adjusted Vegetation Index in this study) based on soil line simulated from images in early green

up and late senescence seasons by the (R, NIR_{min}) method explains more variation for green biomass or green cover than those based on soil line from reflectance of soil samples.

Acknowledgements

The authors would like to acknowledge the Natural Sciences and Engineering Research Council (NSERC), ISTP Canada, and China scholarship Council (CSC) for financial support, as well as Grasslands National Park and the University of Saskatchewan for logistical support. The authors would also like to thank Eric Lamb for his contributions to this project and all the field crew for collecting the field data.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Zhang, C.; Guo, X. Measuring biological heterogeneity in the northern mixed prairie: A remote sensing approach. *Can. Geogr.* **2007**, *51*, 462–474.
2. Milton, S.J.; Dean, W.R.J.; Ellis, R.P. Rangeland health assessment: A practical guide for ranchers in arid Karoo shrublands* 1. *J. Arid Environ.* **1998**, *39*, 253–265.
3. Eisfelder, C.; Kuenzer, C.; Dech, S. A Review on derivation of biomass information in semi-arid regions based on remote sensing data. *Proc. SPIE* **2010**, *7831*, doi:10.1117/12.868505.
4. Frank, A.B.; Karn, J.F. Vegetation indices, CO_2 flux, and biomass for Northern Plains Grasslands. *J. Range Management* **2003**, *56*, 382–387.
5. Feng, X.M.; Liu, Y.; Zhao, Y.S. Remote Sensing Linked Modeling of the Aboveground Biomass of Semiarid Grassland in Inner Mongolia. In Proceedings of 2005 IEEE International Geoscience and Remote Sensing Symposium, IGARSS ’05, Seoul, South Korea, 25–29 July 2005; Volume 5, pp. 3047–3050.
6. Bao, A.M.; Cao, X.M.; Chen, X.; Xia, Y. Study on models for monitoring of aboveground biomass about Bayinbuluke grassland assisted by remote sensing. *Proc. SPIE* **2008**, *7083*, doi:10.1117/12.791724.
7. Shen, M.; Tang, Y.; Klein, J.; Zhang, P.; Gu, S.; Shimono, A.; Chen, J. Estimation of aboveground biomass using *in situ* hyperspectral measurements in five major grassland ecosystems on the Tibetan Plateau. *J. Plant. Ecol.* **2008**, *1*, 247–257.
8. Chen, J.; Gu, S.; Shen, M.G.; Tang, Y.H.; Matsushita, B. Estimating aboveground biomass of grassland having a high canopy cover: An exploratory analysis of *in situ* hyperspectral data. *Int. J. Remote Sens.* **2009**, *30*, 6497–6517.
9. Zhang, C.; Guo, X. Monitoring northern mixed prairie health using broadband satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 2257–2271.
10. Richardson, A.J.; Wiegand, C. Distinguishing vegetation from soil background information (by gray mapping of Landsat MSS data). *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.

11. Fox, G.A.; Sabbagh, G.J.; Searcy, S.W.; Yang, C. An automated soil line identification routine for remotely sensed images. *Soil Sci. Soc. Am. J.* **2004**, *68*, 1326–1331.
12. Baret, F.; Jacquemoud, S.; Hanocq, J.F. About the soil line concept in remote-sensing. *Adv. Space Res.* **1993**, *13*, 281–284.
13. Galvao, L.S.; Vitorello, I. Variability of laboratory measured soil lines of soils from southeastern Brazil. *Remote Sens. Environ.* **1998**, *63*, 166–181.
14. Jaishanker, R.; Thomaskutty, A.V.; Senthivel, T.; Sridhar, V.N. Soil line transformation based relative radiometric normalization. *Int. J. Remote Sens.* **2006**, *27*, 5103–5108.
15. Yoshioka, H.; Miura, T.; Demattê, J.A.; Batchily, K.; Huete, A.R. Derivation of soil line influence on two-band vegetation indices and vegetation isolines. *Remote Sens.***2009**, *1*, 842–857.
16. Huete, A.; Jackson, R.; Post, D. Spectral response of a plant canopy with different soil backgrounds. *Remote Sen. Environ.* **1985**, *17*, 37–53.
17. Baret, F.; Guyot, G. Potentials and limits of vegetation indexes for LAI and APAR assessment. *Remote Sens. Environ.* **1991**, *35*, 161–173.
18. Baret, F.; Guyot, G.; Major, D. TSAVI—A Vegetation Index which Minimizes Soil Brightness Effects on LAI and APAR Estimation. In Proceedings of 12th Canadian Symposium on Remote Sensing and 1989 International Geoscience and Remote Sensing Symposium, IGARSS'89, Vancouver, BC, Canada, 10–14 July 1989; pp. 1355–1358.
19. Gitelson, A.A.; Stark, R.; Grits, U.; Rundquist, D.; Kaufman, Y.; Derry, D. Vegetation and soil lines in visible spectral space: A concept and technique for remote estimation of vegetation fraction. *Int. J. Remote Sens.* **2002**, *23*, 2537–2562.
20. Thoma, D.; Gupta, S.; Bauer, M. Evaluation of optical remote sensing models for crop residue cover assessment. *J. Soil Water Conserv.* **2004**, *59*, 224–233.
21. Fox, G.A.; Sabbagh, G.J. Estimation of soil organic matter from red and near-infrared remotely sensed data using a soil line Euclidean distance technique. *Soil Sci. Soc. Am. J.* **2002**, *66*, 1922–1929.
22. Fox, G.A.; Metla, R. Soil property analysis using principal components analysis, soil line, and regression models. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1782–1788.
23. Wang, J.; Li, Y.H.; Chen, Y.Q.; He, T.; Lv, C.Y. Hyperspectral Degraded Soil Line Index and Soil Degradation Mapping in Agriculture-Pasture Mixed Area in Northern China. In Proceedings of International Workshop on Earth Observation and Remote Sensing Applications, 2008 (EORSA 2008), Beijing, China, 30 June–2 July 2008 ; pp. 1–10.
24. Stabile, M.C.C.; Searcy, S.W. Validation of the soil line transformation technique. *Trans. ASABE* **2009**, *52*, 633–640.
25. Paz-Pellat, F.; Reyes, M.; Medrano, E. Design of spectral vegetation indexes using iso-soil curves. *Agrociencia* **2011**, *45*, 121–134.
26. Guo, X.L.; Wilmshurst, J.F.; Li, Z.Q. Comparison of laboratory and field remote sensing methods to measure forage quality. *Int. J. Environ. Res. Public Health* **2010**, *7*, 3513–3530.
27. Black, S.C.; Guo, X. Estimation of grassland CO_2 exchange rates using hyperspectral remote sensing techniques. *Int. J. Remote Sens.* **2008**, *29*, 145–155.
28. Li, Z.; Guo, X. Detecting climate effects on vegetation in northern mixed prairie using NOAA AVHRR 1-km time-series NDVI data. *Remote Sens.* **2012**, *4*, 120–134.

29. Fargey, K.S.; Larson, S.D.; Grant, S.J.; Fargey, P.; Schmidt, C. *Grasslands National Park Field Guide*; Prairie Wind & Silver Sage, Friends of Grasslands Inc.: Val Marie, SK, Canada, 2000.
30. Guo, X.; Wilmshurst, J.; McCanny, S.; Fargey, P.; Richard, P. Richard, P. Measuring spatial and vertical heterogeneity of grasslands using remote sensing techniques. *J. Environ. Inform.* **2004**, *3*, 24–32.
31. Koenker, R.; Bassett, G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50.
32. Koenker, R.; Machado J.A.F. Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.* **1999**, *94*, 1296–1310.
33. Tsionas, E.G. Bayesian quantile inference. *J. Stat. Comput. Simul.* **2003**, *73*, 659–674.
34. Young, T.M.; Shaffer, L.B.; Guess, F.M.; Bensmail, H.; Leon, R.V. A comparison of multiple linear regression and quantile regression for modeling the internal bond of medium density fiberboard. *For. Prod. J.* **2008**, *58*, 39–48.
35. Cade, B.S.; Terrell, J.W.; Neely, B.C. Estimating geographic variation in allometric growth and body condition of blue suckers with quantile regression. *Trans. Am. Fish. Soc.* **2011**, *140*, 1657–1669.
36. Mills, A.J.; Fey, M.V.; Grongroft, A.; Petersen, A.; Medinski, T.V. Unravelling the effects of soil properties on water infiltration: Segmented quantile regression on a large data set from arid south-west Africa. *Australian J. Soil Res.* **2006**, *44*, 783–797.
37. Mills, A.; Fey, M.; Donaldson, J.; Todd, S.; Theron, L. Soil infiltrability as a driver of plant cover and species richness in the semi-arid Karoo, South Africa. *Plant. Soil* **2009**, *320*, 321–332.
38. Zhang, X.Y.; Friedl, M.A.; Schaaf, C.B.; Strahler, A.H.; Hodges, J.C.F.; Gao, F.; Reed, B.C.; Huete, A. Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* **2003**, *84*, 471–475.
39. Rouse, J.; Haas, R.; Schell, J.; Deering, D.; Harlan, J. Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation Greenbelt, MD; NASA/GSFC: College Station, TX, USA, 1974.
40. Yoshioka, H.; Miura, T.; Demattê, J.A.; Batchily, K.; Huete, A.R. Soil line influences on two-band vegetation indices and vegetation isolines: A numerical study. *Remote Sens.* **2010**, *2*, 545–561.