Remote Sens. **2015**, *7*, 11801-11820; doi:10.3390/rs70911801

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Data-Gap Filling to Understand the Dynamic Feedback Pattern of Soil

Shanxin Guo [1,†], Lingkui Meng [1], A-Xing Zhu [2,3,4,5,6,†,*], James E. Burt [6], Fei Du [6], Jing Liu [6] and Guiming Zhang [6]

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; E-Mails: guoshanxin@whu.edu.cn (S.G.); lkmeng@whu.edu.cn (L.M.)

2 Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing 210023, China

3 State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing, 210023, China

4 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

5 State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

6 Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA; E-Mails: jeburt@wisc.edu (J.E.B.); fdu@wisc.edu (F.D.); jliu93@wisc.edu (J.L.); gzhang45@wisc.edu (G.Z.)

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: axing@njnu.edu.cn or azhu@wisc.edu; Tel.: +1-608-262-0272.

Academic Editors: Nicolas Baghdadi and Prasad S. Thenkabail

Received: 5 May 2015 / Accepted: 1 September 2015 / Published: 14 September 2015

Abstract: Detailed and accurate information on the spatial variation of soil over low-relief areas is a critical component of environmental studies and agricultural management. Early studies show that the pattern of soil dynamics provides comprehensive information about soil and can be used as a new environmental covariate to indicate spatial variation in soil in low relief areas. In practice, however, data gaps caused by cloud cover can lead to incomplete patterns over a large area. Missing data reduce the accuracy of soil information and make it hard to compare two patterns from different locations. In this study, we introduced a new method to fill data gaps based on historical data. A strong correlation between MODIS

band 7 and cumulated reference evapotranspiration (CET_0) has been confirmed by theoretical derivation and by the real data. Based on this correlation, data gaps in MODIS band 7 can be predicted by daily evaporation data. Furthermore, correlations among bands are used to predict soil reflectance in MODIS bands 1–6 from MODIS band 7. A location in northeastern Illinois with a large area of low relief farmland was selected to examine this idea. The results show a good exponential relationship between MODIS band 7 and $CET_0^{0.5}$ in most locations of the study area (with average R^2 = 0.55, p < 0.001, and average NRMSE 10.40%). A five-fold cross validation shows that the approach proposed in this study captures the regular pattern of soil surface reflectance change in bands 6 and 7 during the soil drying process, with a Normalized Root Mean Square Error (NRMSE) of prediction of 13.04% and 10.40%, respectively. Average NRMSE of bands 1–5 is less than 20%. This suggests that the proposed approach is effective for filling the data gaps from cloud cover and that the method reduces the data collection requirement for understanding the dynamic feedback pattern of soil, making it easier to apply to larger areas for soil mapping.

Keywords: dynamic feedback pattern of soil; soil mapping; MODIS; soil surface reflectance; soil drying process; soil evaporation

1. Introduction

Information on the spatial variation of soil is a critical component of environmental research and agricultural management [1]. The rapid development of remote sensing technology provides significant potential to extract soil information at scales from regional to global [2]. In most studies, soil surface characteristics, such as soil texture and soil surface roughness, are generally obtained based on only one or a few images captured by multispectral or hyperspectral sensors at one or a few specific times [3–5]. As a result, the accuracy of the results may be highly related to sensor noise, atmospheric conditions and the soil surface status at that specific time [6,7]. In fact, however, soil reflectance varies during the soil drying process, which is closely related to both soil properties and soil surface water content [8]. Just one or two images cannot provide comprehensive information about overall changes in soil reflectance. In order to capture the changes of soil reflectance from wet to dry for the purpose of differentiating soil conditions, a dynamic feedback pattern for soil was introduced as a new data variable captured by satellite image to study soil characteristics [9]. Previous research has established that the feedback pattern can be used as a new environmental covariate to indicate soil spatial variation over low relief areas [9,10].

The structure of the soil feedback pattern is a three dimensional space arranged with time as the X-axis, wavelength as the Y-axis, and soil surface reflectance as the Z-axis (Figure 1). At a given location after a rainfall, the soil surface reflectance was captured by the MODIS sensor for a short period (5–7 days) and organized in this three-dimensional space. With this design, rainfall was considered as an input to land surface. The changes in soil reflectance that occurred in the process of drying after a rain event were considered as feedback in response to the rain [10]. MODIS was chosen because it has a high temporal resolution compared to other hyperspectral sensors. This dynamic feedback pattern

records soil reflectance changes from wet to dry conditions for each location and can be used to predict soil characteristics at that location [9].

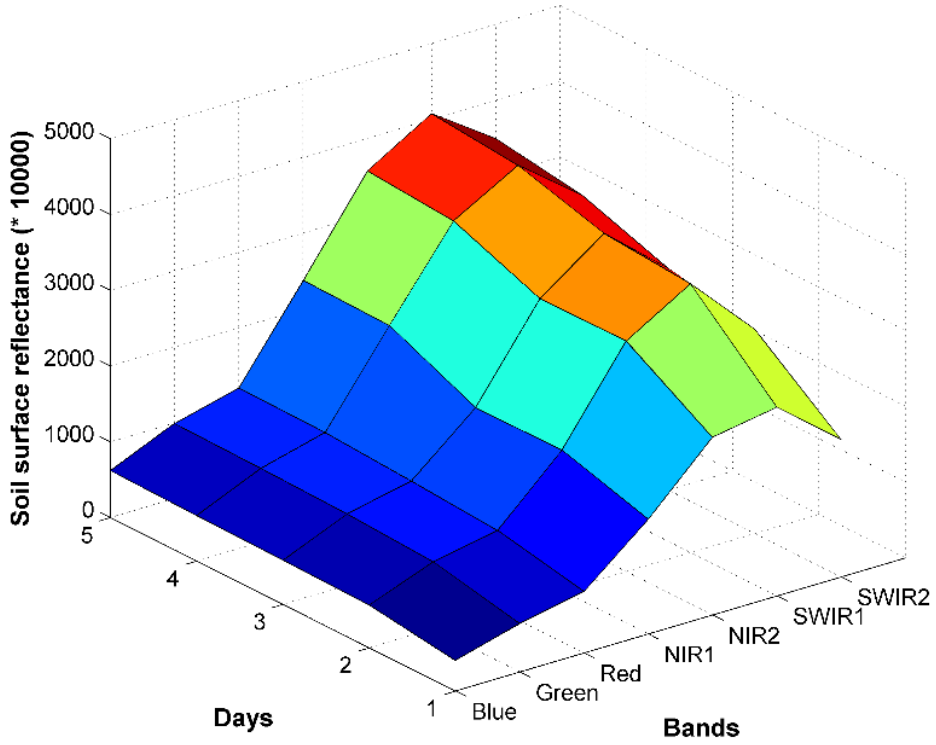

Figure 1. Dynamic soil feedback pattern at a given location in the study area.

In practice, however, there are limitations to using this model to predict soil characteristics over a large area. The data gaps caused by cloud cover may lead to many incomplete soil feedback patterns. In addition, when data gaps occur on different days for two patterns with different locations, it is hard to directly compare the two incomplete patterns. For example, after one rain event, the feedback pattern from Location *A* misses the data for the first and third days and the pattern from Location *B* misses the data for the second and fifth days. These two incomplete patterns cannot be compared directly. In the case of soil mapping, this comparison is essential for differentiating the soils from different locations.

In this study, we focus on how to fill those data gaps to rebuild the complete feedback pattern for each location. Previous research attempting to address the problem of missing data can be grouped into three categories. In the first, methods use multi-source satellite sensors to make up for the missing gaps by assuming low temporal variability between the auxiliary and target image with a well-controlled system error [11]. The transform function which is used to transform an auxiliary pixel to a missing data pixel might come from a spectral BRDF model [12] or a Principal Component Transformation (PCT) or other statistics-based models like Local Linear Histogram Matching (LLHM) [13]. These approaches may have acceptable accuracy when the focus is on temporally stable spectral targets such as land use/cover classes or units in urban mapping. However, they have limited performance with highly dynamic properties like soil reflectance.

The second group uses neighboring pixels around missing points to fill the gap with moving windows. These gap-filling techniques take advantage of spatial autocorrelation that assumes target variables are related to each other over a short distance. The simplest case is to replace the missing pixels with the

mean value of the neighboring pixels. More advanced techniques such as geostatistic-based techniques can also be used [14]. However, problems exist when the gap size is bigger than the semivariogram range, or the target value does not have a strong spatial autocorrelation. These approaches are not suitable for filling data gaps in soil feedback patterns. Because the purpose of the dynamic soil feedback pattern is to describe soil spatial variation, a data process that includes neighboring soil information will reduce the accuracy of soil mapping and increase the uncertainty in the data-gap areas.

The third group uses time-series cloud-free observations at the same location to fill the gaps due to cloud cover. The simplest way is to use the mean value of the images before and after the time of the current image to fill the gap at the same location [15]. These strategies have been applied in MODIS EVI products [11] and TM NDVI [16] gap filling. However, the approaches will fail because soil surface reflectance is highly variable due to soil surface moisture changes that occur over a short time. Thus, data-gap filling in soil feedback patterns remains a challenge.

In this study, we present a new approach to fill data gaps caused by cloud cover, using historical data. The approach was applied to a large area of farmland in northeastern Illinois, USA, as a case study. The proposed method effectively fills data-gaps and allows the soil feedback pattern in every pixel to be used as a data source to predict soil types and properties regardless of cloud cover, using a methodology created in previous studies [9] and [10].

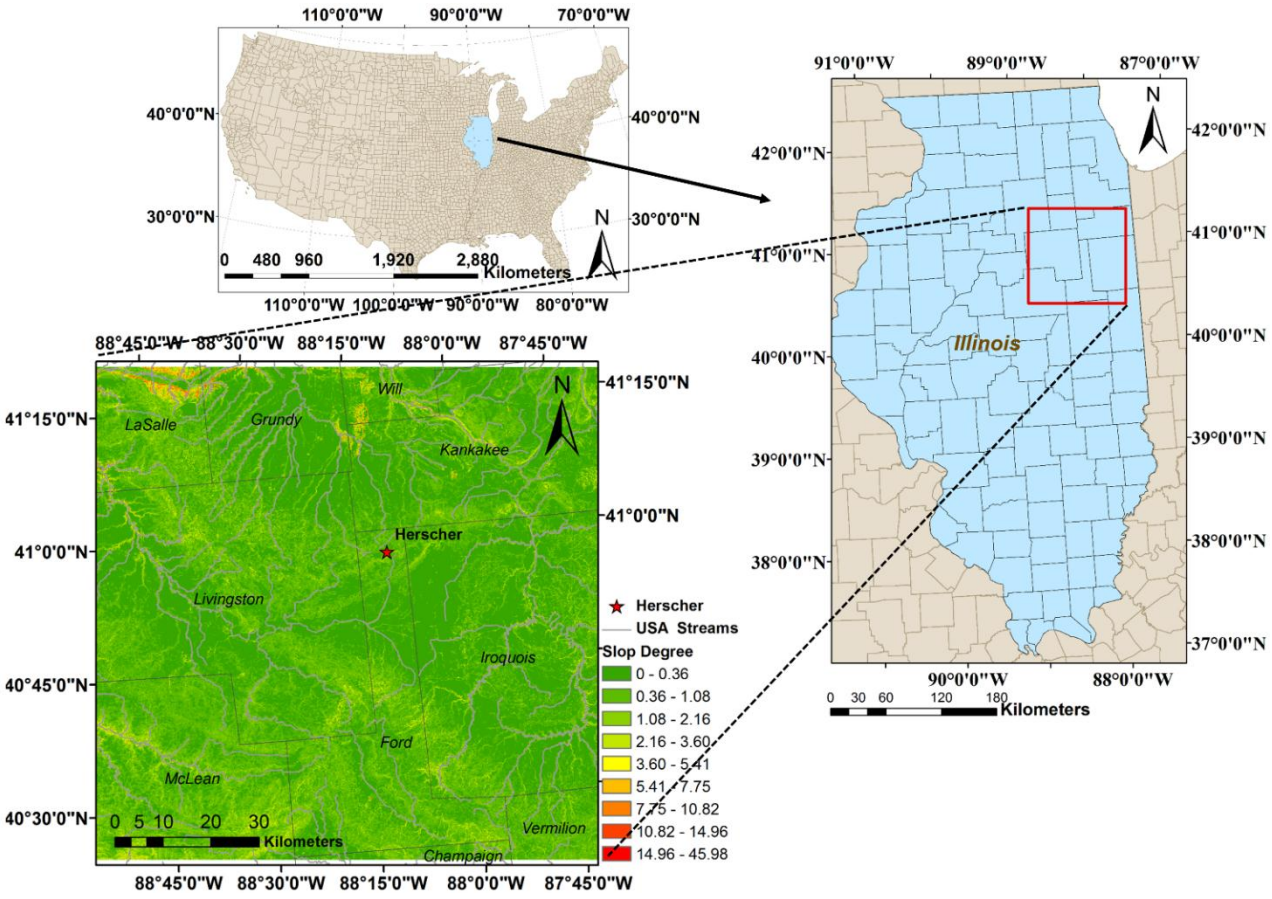

Figure 2. Location of study area with a slope map background.

2. Study Area

The study area is a large, flat area of farmland (about 10,000 km^2) located around Herscher (87°39′36″W–88°52′53″W, 40°24′6″N–41°19′18″N) in northeastern Illinois, USA (Figure 2). The study area is part of a semiarid region in northern Illinois. According to the NOAA National Climatic Data Center weather data (1981–2010), average precipitation in April and May are around 82 and 91 mm, respectively. Average temperatures in April and May are around 17 and 23 °C. Average sunny days and partly sunny days (40%–70% cloud cover) are seven and eight days in April and May [17]. This low number of cloud-free observation days creates many data gaps in the feedback pattern. The major soil order in this area is Mollisol. According to the USDA Soil Survey Geographic Database (SSURGO) map, there are five major soil subgroups in this area, including Vertic Endoaquolls, Vertic Argiaquolls, Typic Endoaquolls, Oxyaquic Argiudolls and Aquic Argiudolls. The study area is low-relief farmland. The average slope degree is 0.66°, with a standard deviation of 0.92. Elevation is in the range of 137–305 m, with a mean of 210 m.

3. Methodology

3.1. Basic Concept

Soils are relatively stable, so at a given location they should reveal similar behavior during the drying process after rain events. Soil surface reflectance is also highly influenced by near-surface environmental conditions. So, there is a correlation between soil surface reflectance and other near-surface environmental parameters. These near-surface parameters are not affected by cloud cover and can be easily obtained. A statistical model can be built to capture this correlation using historical data. With this model, missing soil reflectance data can be predicted by using the near-surface environmental values for the dates with cloud cover.

We use evaporation in this study as the near-surface environmental parameter, because soil moisture in the surface layer (1–2 cm) is highly influenced by near-surface evaporation conditions [18]. Soil surface moisture change is the major cause of soil surface reflectance change during the soil drying process [8]. Compared with soil surface moisture, near-surface evaporation conditions can be easily calculated or estimated from weather observations and are not affected by cloud cover. We can therefore use evaporation as a surrogate for soil surface moisture to predict soil surface reflectance based on a regression model.

This method can effectively predict soil reflectance in water sensitive regions (band 7 in MODIS sensor) because soil surface moisture has more significant influence on soil surface reflectance in 1.7–2.5 μm wavelength regions, the location of MODIS band 7 [8,19–21]. The question then becomes how to predict soil reflectance in water non-sensitive bands (bands 1–6 in MODIS sensor). In fact, early soil science findings show that when soil surface moisture decreases, all regions from the visible (400 nm) to shortwave infrared (2500 nm) regions of the electromagnetic spectrum of soil reflectance increase simultaneously. The difference is that reflectance in the water-sensitive bands (1100 nm–2500 nm) increases more significantly compared to the non-sensitive bands (400–1100 nm) [22]. There is thus also a correlation between water sensitive bands and insensitive bands during the soil drying process. This correlation has already been confirmed and used to predict 1.6 μm band 6 from band 7 to fix the noise

problem in MODIS Aqua band 6 in previous studies [23]. Based on the correlation, soil reflectance in water-insensitive bands can be predicted based on reflectance in water-sensitive bands.

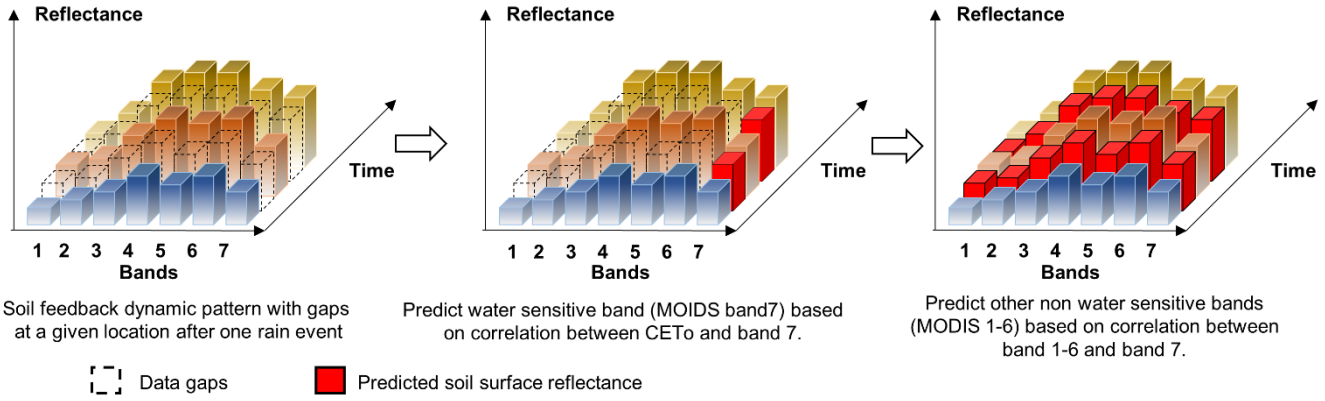

Figure 3. Basic concept for using two correlations to fill missing data gaps in a soil feedback pattern during one rain event.

An overall illustration of our gap-filling method is provided in Figure 3. The method consists of two steps. First, the correlation between water-sensitive bands and the evaporation parameter at each location is obtained and the soil reflectance in water-sensitive bands based on the evaporation parameter value on cloud cover days is predicted. The second step is to find a correlation between soil reflectance in a water-sensitive region (band 7 in MODIS) and in insensitive regions (bands 1–6 in MODIS). Using these correlations, the soil reflectance of bands 1–6 in MODIS is predicted. This two-step process is repeated until all the missing values in the feedback pattern are populated. The same strategy is also applied to other pixels with data gaps. It should be noted that the two correlations are acquired independently from historical data at that location and can only be applied to that location. So, the predictions at locations are independent.

3.2. Choosing a Proper Model to Capture Correlation between Soil Surface Reflectance and Evaporation Parameters

To fill the data gaps, the key step for the study is to predict soil reflectance in a water-sensitive band based on evaporation. In order to build the correlation model between soil surface reflectance and evaporation parameter, a proper regression model must be chosen. This model selection is done based on equation derivation with previous studies' findings on soil evaporation and soil spectroscopy.

Soil surface evaporation can be divided into two stages [18,24–26]; the first stage is the constant-rate stage. In this stage, evaporation (E_s) is limited only by the supply of energy to the soil surface. The evaporation rate is similar to the rate for a free water surface because enough water is present to allow for its free evaporation on the top-soil layer. This stage typically lasts for a few hours depending on soil surface characteristics (structure, texture, *etc.*). The second stage is the falling-rate stage. In this stage, water movement on the soil surface layer is controlled by soil moisture conditions and soil hydraulic properties [27–29]. Stroosnijder (1987) and Gallardo (1996) found a good relationship between cumulative bare soil evaporation (CE_s) and cumulative reference evapotranspiration (CET_0) [24,30]. These variables are connected by maximum possible cumulative soil evaporation (CE_x), which refers to the

maximum possible water that the atmosphere can take from the soil surface when soil evaporation is not limited by soil hydraulic properties. A detailed definition is given as follows.

$$E_x = K_x ET_0 \tag{1}$$

$$CET_{0(k)} = \sum_{i=1}^{k} ET_{0(i)} \tag{2}$$

$$CE_{x(k)} = \sum_{i=1}^{k} E_{x(i)} = K_x \sum_{i=1}^{k} ET_{0(i)} = K_x CET_{0(k)} \tag{3}$$

where K_x is the maximum expected crop coefficient (K_{co}) value; E_x is the daily maximum possible soil evaporation; ET_0 is the daily reference evapotranspiration (potential evapotranspiration); $CET_{0(k)}$ is the k^{th} day's cumulative reference evapotranspiration after a rain event; $CE_{x(k)}$ is the k^{th} day's maximum possible cumulative soil evaporation after a rain event; the cumulative process starts from the first day after a rain event to the k^{th} day. There is a requirement that there is no other rain event during this period. Based on the two-stage soil evaporation method developed by Ritchie (1972) and Stroosnijder (1987) [27,30], the relationship between the k^{th} day's cumulative bare soil evaporation CE_s and the k^{th} day's maximum possible cumulative soil evaporation CE_x is given by Equations (4) and (5).

$$CE_{s(k)} = CE_{x(k)} \; when \; \sqrt{CE_{x(k)}} < \beta \; in \; Stage \; 1 \tag{4}$$

$$CE_{s(k)} = \beta \sqrt{CE_{x(k)}} \; when \; \sqrt{CE_{x(k)}} \geq \beta \; in \; Stage \; 2 \tag{5}$$

where the soil hydraulic factor (β) represents the turning point at which the evaporation rate changes from Stage 1 to Stage 2. In Stage 1, soil evaporation is equal to the maximum possible cumulative soil evaporation since soil hydraulic properties do not limit evaporation. In Stage 2, actual soil evaporation is much smaller than the evaporation in Stage 1 and depends on soil surface characteristics and soil hydraulic properties.

We assume that most soil feedback data are captured when soil evaporation moves into Stage 2, because for most soil types, Stage 1 can typically be maintained for just a short period. This simplification may underestimate the soil evaporation on the first day after a rain event, but will nevertheless fit the actual evaporation of soils for most cases.

To calculate daily reference evapotranspiration, Hargreaves's equation (1985) has been used based on daily maximum and minimum air temperature data [31,32]. The equation was chosen because its requirements can be extended over a large area with fewer weather parameters than other methods. Only daily temperature data is required [33].

$$ET_0 = 0.0023(T_{mean} + 17.8)(T_{max} - T_{min})^{0.5} \, R_a \tag{6}$$

where T_{mean} is equal to$(T_{max} - T_{min})/2$; R_a is extraterrestrial radiation for daily periods that can be calculated by a given latitude and date based on the equation given in the FAO-56 document.

Water balance in the soil surface layer (1–5 mm) can be used to describe how the water content at the soil surface is related to soil evaporation after a rainfall. Data were first collected by MODIS sensor starting from the second day after a rain event. By this point, the infiltration and runoff process has

already finished in the soil surface layer. The water content has reached the soil field capacity (FC) which is the upper limit of moisture content that soil can hold. For a few days after a rain event, soil evaporation is the primary cause of water loss during the soil drying process in the soil surface layer. The water balance after rainfall can be described as a simplified equation as follows.

$$\theta_k = \theta_s - CE_{s(k)} \tag{7}$$

where θ_k represents the soil surface water content on the k^{th} day after rain; θ_s is the surface soil water content after a rain event (field moisture capacity, FC). $CE_{s(k)}$ is the k^{th} day cumulative soil evaporation.

As described above, soil surface reflectance increases as soil surface water content decreases, especially in water sensitive bands. Various models have been developed to build the relationship between soil surface water content and soil surface reflectance [8,19]. The common models show that soil surface reflectance relates to both soil reflectance in dry conditions and soil water content. An exponential relationship most commonly used in these models is based on experimental data in the lab [8] and satellite image data [19]. We use the general model described by Muller (2001) with SPOT image data [19] for the study. The soil moisture-reflectance relationship is shown in Equation (8):

$$\rho_{s(\lambda,k)} = \rho_{so(\lambda)} Exp(a_{s(\lambda)}\theta_k) \tag{8}$$

where $\rho_{s(\lambda,k)}$ is the k^{th} day's soil surface reflectance in the spectral band λ after the rain event, $a_{s(\lambda)}$ is the reflectance attenuation factor for the soil s in the spectral band λ, θ_k represents the soil surface water content on the k^{th} day after rain, and $\rho_{so(\lambda)}$ is the theoretical dry soil reflectance of the soils in the spectral band λ.

To combine the Equations (3), (5), (7) and (8), the derivation steps are shown in Equations (9) and (10), and the relationship between the k^{th} day's soil surface reflectance ($\rho_{s(\lambda,k)}$) and the k^{th} day's cumulative reference evapotranspiration ($CET_{0(k)}$) is shown in Equation (11) as follows.

$$\rho_{s(\lambda,k)} = \rho_{so(\lambda)} Exp\left\{a_{s(\lambda)}\left(\theta_s - \beta\sqrt{K_x CET_{0(k)}}\right)\right\} \tag{9}$$

$$\rho_{s(\lambda,k)} = \rho_{so(\lambda)} Exp\left(a_{s(\lambda)}\theta_s\right) * Exp\left(-a_{s(\lambda)}\beta\sqrt{K_x CET_{0(k)}}\right) \tag{10}$$

$$\begin{gathered} \rho_{s(\lambda,k)} = B * Exp\left(A\sqrt{CET_{0(k)}}\right) \\ where: B = \rho_{so(\lambda)} Exp\left(a_{s(\lambda)}\theta_s\right) \\ A = -a_{s(\lambda)}\beta\sqrt{K_x} \end{gathered} \tag{11}$$

When soil at given location remains stable during multiple rain events, the basic characteristic properties of soil, such as the reflectance attenuation factor $a_{s(\lambda)}$, initial soil water content (field moisture capacity) θ_s, soil reflectance in air dry conditions($\rho_{so(\lambda)}$), soil hydraulic factor (β) and maximum expected crop coefficient (K_x) can be seen as constants during different rain events. In other words, both coefficient A and B can be considered a constant value at a given location. An exponential relationship (Equation (11)) is then shown between the k^{th} day's soil surface reflectance ($\rho_{s(\lambda,k)}$) and the k^{th} day's cumulative reference evapotranspiration ($CET_{0(k)}$) for each location with coefficient A and B. These coefficients can be confirmed by using a regression model based on the historical soil surface reflectance data. For practical convenience, $\rho_{s(\lambda,k)}$ was transformed by a natural logarithm function in order to change the initial exponential relationship into a linear relationship.

3.3. Filling Data Gaps

The details for using daily cumulated ET_0 (CET_0) and historical soil surface reflectance data to predict soil surface reflectance on an overcast day is illustrated in Figure 4. Two correlations are used for the prediction. The first is a correlation between MODIS band 7 and daily cumulated ET_0 (CET_0). The second is a correlation between MODIS bands 1–6 and MODIS band 7. There are five main steps in this gap-filling approach. For each location, daily CET_0 is first calculated based on daily highest and lowest air temperature with Hargreaves's equation (1985); second, a regression model is used to describe the correlation between MODIS band 7 and daily CET_0; third, the correlation learned from Step 2 and the daily CET_0 for days with missing data are used to predict MODIS band 7 with CET_0 on overcast days; fourth, a correlation between band 7 and the other bands is built using a regression model with historical soil surface reflectance data for the given location; fifth, bands 1–6 are predicted based on the correlation between bands and the value of band 7 predicated in Step 3 for a cloud-covered day. The steps are repeated for every location in the study area until all data gaps are filled in.

To describe the correlation between MODIS band 7 and other bands, the linear regression model is a reasonable method. And, as mentioned before, this correlation is learned from historical data, is independent for each location and can only be applied in that location.

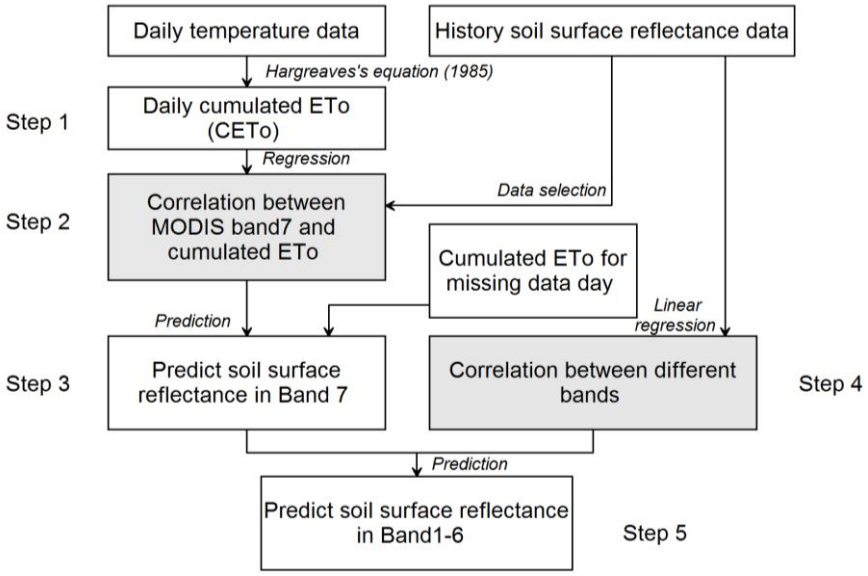

Figure 4. Data-processing workflow.

4. Data Preprocessing and Validation

4.1. Data Collection and Preprocessing

The objective of the dynamic soil feedback pattern is to provide a new environmental covariate for digital soil mapping by recording soil surface reflectance changes during the soil drying process. The most basic requirement of data collection is that all MODIS reflectance data must derive from the reflectance of bare soil. Data for vegetated soil in the growth season are excluded. USGS MODIS daily surface reflectance data in Terra (MOD09GA, MOD09GQ) for every April and May from 2000–2011 (USGS Earth explorer: http://earthexplorer.usgs.gov/) were used. The data collection period was set to

April and May because the entire study area has no snow cover and has not fully started the planting season in these months. After each rain event in the two months, images for the seven days immediately after the precipitation were collected to capture the soil drying processes. Band 1 through band 7 reflectance data were used to build the feedback pattern, which documented the reflection of shortwave radiation from soil surfaces in the visible and near-infrared (VNIR: 400–1100 nm) and short wave infrared region (SWIR: 1100–2500 nm) of the electromagnetic spectrum.

In data preprocessing, reflectance data from all bands were resampled to 250 m pixel size to utilize the spatially detailed spectral data from the red and NIR1 bands. In addition, a NASA cloud mask product (MOD35) was used to remove cloud cover in the image (Data source: http://ladsweb.nascom.nasa.gov/data/search.html). Data selection was based on a series of bounding thresholds to ensure the data used to build the regression relationship were not influenced by the open water, thin clouds and partial vegetation cover. For example, reasonable bare soil data bounds for MODIS band 1 (459–479 nm) values were set to 0.03–0.0625, which can avoid the open water accumulated after rain and any thin cloud influence [8,22,34]. Areas with NDVI of less than 0.3 were considered to be bare land or to contain very sparse vegetation coverage [10].

The NASA Daymet dataset (http://daymet.ornl.gov/) was used as the input meteorological data to calculate daily CET_0, which provides gridded estimates of daily weather parameters, such as daily highest and lowest air temperature and precipitation. The spatial resolution of both air temperature and precipitation is 1km x 1km. According to Thornton (1997), cross-validation mean absolute errors (MAE) for daily predicted *vs.* observed highest and lowest air temperatures are 1.8 and 2.0 ℃, and MAEs for annual averages of daily estimates are 0.7 and 1.2 ℃, respectively. Biases for annual average highest and lowest air temperature are −0.1 and +0.1 ℃. MAE for predicted annual total precipitation was 13.4 cm. and the success rate for predictions of daily precipitation occurrence was 83.3% [35]. Overall, the accuracy of the Daymet dataset is acceptable for this study. Precipitation greater than 0.5 inch (12.7 mm) was considered a valid rain event that ensures that the soil surface layer would be saturated at the beginning (the 0.5 inch minimum was chosen according to precipitation and soil moisture observations in the SCAN weather station in Mason County, Illinois. http://www.wcc.nrcs.usda.gov/scan/).

4.2. Validation

Cloud free observations in MODIS for April and May from 2000–2011 were used as validation data, and a five-fold cross validation was used to evaluate prediction accuracy. The root mean square error (RMSE) was used to assess the average magnitude of the errors between the observed and the predicted values on cloud free days. RMSE was estimated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{12}$$

where P_i is the predicted soil surface reflectance of cloud free day i. O_i is the observed value of soil surface reflectance on day i. n is the number of cloud free days during the data collection period. The range of RMSE is from 0 to ∞. The lower a RMSE value the better the performance of the prediction.

To make RMSE comparable between different bands, the Normalized Root Mean Square Error (NRMSE) was calculated by:

$$NRMSE(\%) = \frac{RMSE}{Max - Min} * 100\% \quad (13)$$

The Max and Min values for each bands are from MODIS historical observations for April and May from 2000–2011. A lower RMSE (lower NRMSE) represents a higher prediction accuracy.

5. Results and Discussion

5.1. Correlation between MODIS Band 7 and Cumulative ET₀ (CET₀)

As described in Equation (11), an exponential relationship exists between soil surface reflectance in the water sensitive band (band 7 in MODIS sensor) and the square root of cumulative ET_0 ($CET_0^{0.5}$). The coefficients A and B can be identified using regression modelling with soil surface reflectance at historical rain events. First, we focus on the result at the pixel level, after which we will discuss the result for the entire study area.

Figure 5a shows the relationship between MODIS band 7 and the square root of cumulated reference evapotranspiration ($CET_0^{0.5}$) in one location (87.909109 °W, 40.514948 °N) at the pixel level. There are 33 records for April and May from 2000–2011 based on 23 rain events. To simplify the analysis, the exponential relationship with band 7 is converted to a linear relationship using a natural logarithm transformation, ln (band 7). The relationship plot is shown in Figure 5b. The result shows a strong linear relationship between $\ln(band\ 7)$ and $CET_0^{0.5}$ at this location (correlation coefficient = 0.8638, $p < 0.001$, $R^2 = 0.7463$).

Over the entire study area, each pixel independently builds the linear relationship between $\ln(band7)$ and $CET_0^{0.5}$. The coefficient of determination (R^2) was used to evaluate how much data variation can be captured by this linear model. The spatial distribution of R^2 is shown in Figure 6a. Rivers, vegetation and urban areas were masked out in the initial data processing. In addition, pixels with limited numbers of observation samples (less than 13) were considered unstable for the regression prediction. Those pixels are also masked with black in Figure 6a. A poor linear relationship (*i.e.*, lower values of R^2) appeared near the rivers and urban areas, while good linear relationships (*i.e.*, higher values of R^2) were mainly located in large farmland areas with bare soil. There are several possible reasons for this: (1) soil surface moisture is not only determined by soil evaporation for the close-to-river pixels, but is also influenced by the horizontal flow from nearby regions because the water table in these areas is close to the soil surface [36]. (2) urban impermeable areas, which have completely different spectrum behavior than soil, can significantly disturb the performance of the linear regression model, and (3) data noise caused by atmospheric conditions, mixed pixels of coarse spatial resolution in MODIS data, and changes in soil surface roughness may also influence model performance. The histogram of R^2 for the entire study area is shown in Figure 6b, with a mean of 0.55 and a standard deviation of 0.17.

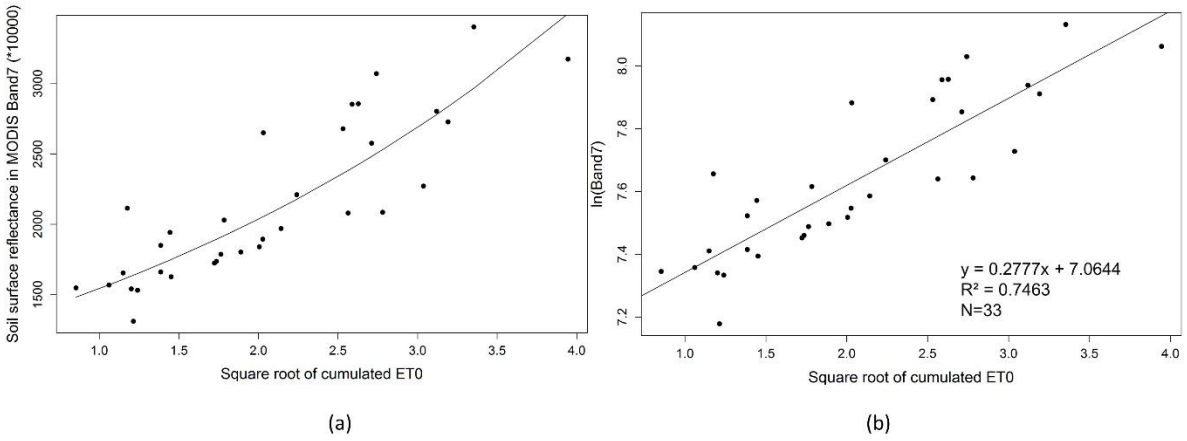

Figure 5. Correlation between soil surface reflectance in water sensitive band (MODIS band 7) and square root of cumulated CET_0 at one pixel (87.909109 °W, 40.514948 °N). (**a**) Exponential relationship; (**b**) Linear relationship after natural logarithm transform.

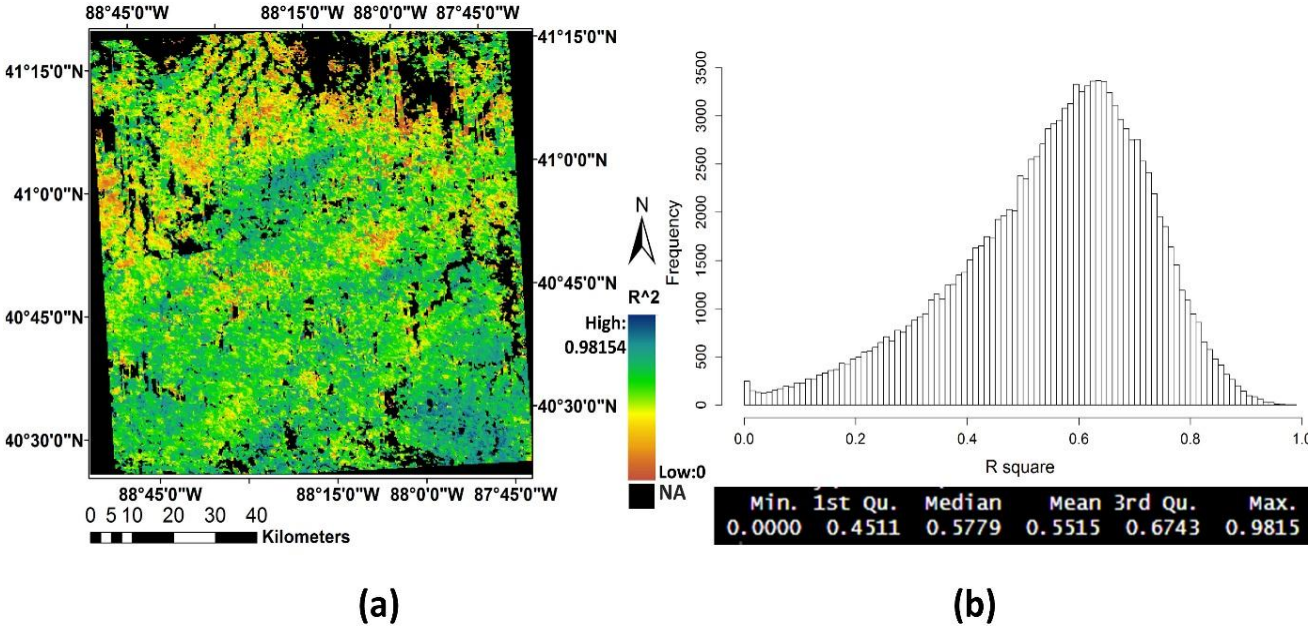

Figure 6. Coefficient of determination (R^2) of linear regression model to describe relationship between $ln(band\ 7)$ and $CET_0{}^{0.5}$ for each location in study area. (**a**) Spatial distribution of R^2; (**b**) Histogram of R^2 and statistical values.

Figure 7a shows the spatial distribution of average residual (%) for each location in the study area. It is no surprise that the pixels with low R^2 in Figure 6 show higher residuals; the low prediction residual appears in large areas of farm land. The value of the residual is the percentage of the real residual in soil reflectance divided by the mean of band 7 in each location. The histogram of average residual (%) in band 7 is shown in Figure 7b with a mean of 13.37% and standard deviation of 3.15.

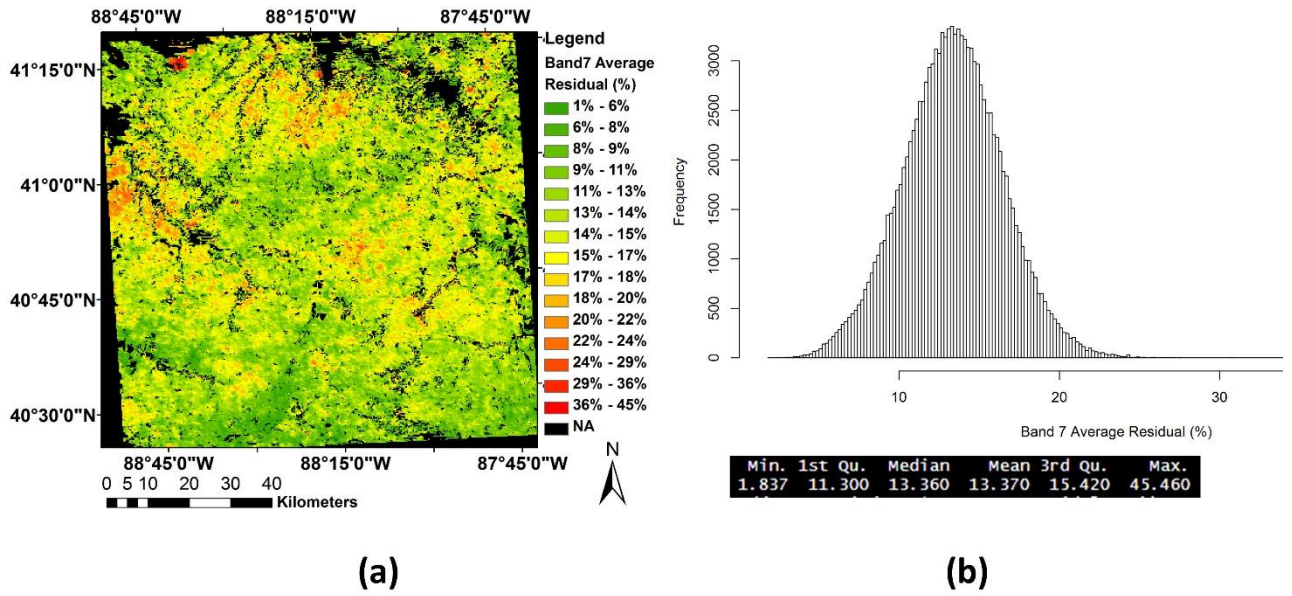

Figure 7. Percentage of average residual of band 7 predicted by linear regression model for each location in study area. (**a**) Spatial distribution of average residual (%) of band 7; (**b**) Histogram of band 7 average residual (%).

Soil refectance data (scaled 10,000 times) from cloud-free images were used to evaluate model performance. The scatter plot between observation and prediction values of all pixels in the entire study area (Figure 8) shows that the linear regression model has captured the soil surface reflectance variation in MODIS band 7 during the soil drying process after rain events. As noted earlier, there is a small group of outliers lying above the 1:1 line (with observation values between 1250 and 1750 and prediction values between 2250 and 3100). This phenomenon might be caused by a disturbance, such as a morning dew, that occured during the soil drying process, because morning dew could significanly increase soil surface moisture, which would lead to a lower-than-normal surface reflectance.

As noted in Figure 8, most of the observation data are gathering reflectance between 1000 and 3250, and there is low data density seen over 3500. This is because for most rain events, soil drying processes can be captured by satellite for only a few days (3–5 days), and then the processes are interrupted by another rain event. In other words, there is a much lower chance to observe soil reflectance on the sixth or seventh day after a rain event. The overall RMSE of band 7 prediction is 316.02 (scaled to 10,000 times real value) and the NRMSE is 10.40%.

In this prediction model, the more accurate daily ET_0 estimation will describe the real soil evaporation more accurately during the soil drying process. The more accurate ET_0 estimation can therefore help the regression model to derive a more accurate relationship between soil surface reflectance and CET_0. The 1985 Hargreaves equation is an experimental equation and is sensitive to local calibration.. According the FAO56 document, the 1985 Hargreaves equation should be verified in each new region by comparing it with estimates from the FAO Penman-Monteith equation at weather stations where solar radiation, air temperature, humidity, and wind speed are measured. The Hargreaves equation error better reveals the linear stability bias. If necessary, the Hargreaves equation can be calibrated on a monthly or annual basis by determining empirical coefficients where ET_0 (Penman-Monteith equation) = a + b ET_0 (1985 Hargreaves equation). The coefficients a and b can be determined by regression analyses. However, this

calibration is not necessary here because the prediction model of band 7 is also a linear regression model. The bias in ET_0 estimated by the 1985 Hargreaves equation will not influence the accuracy of the band 7 prediction since the bias only causes different coefficients rather than the result of band 7 prediction. Besides, the prediction's linear regression model is built into each location and the models in different locations are independent. It is thus not necessary to calibrate ET_0 using the 1985 Hargreaves equation for this study.

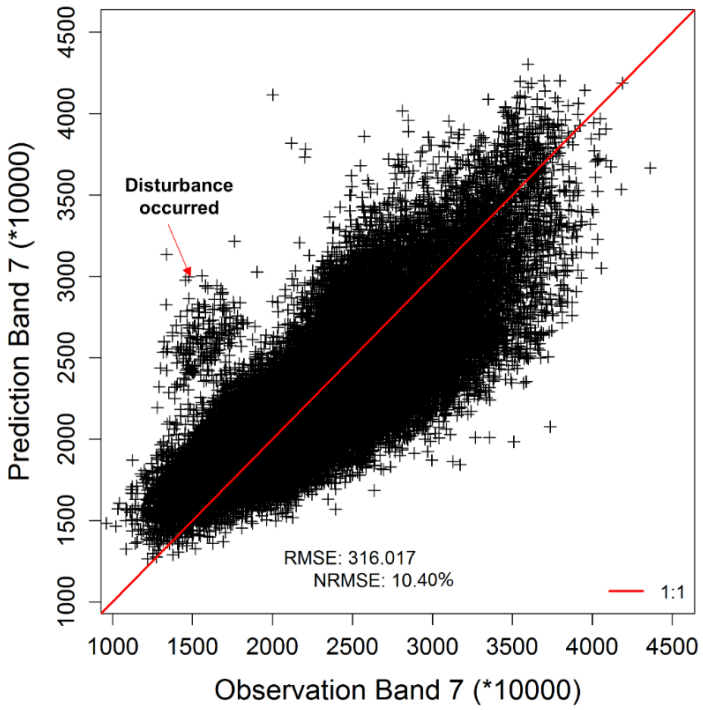

Figure 8. Observation and prediction plot of band 7 based on linear relationship between $\ln(\text{band 7})$ and $CET_0^{0.5}$.

5.2. Correlation among MODIS Bands

A linear regression model was used to describe the correlation between band 7 and other bands at each location. First, we focus our discussion on the results at the pixel level, after which we will discuss the results for the entire study area.

At the pixel level, the adjusted R^2 between bands 1–6 and band 7 from a given location (87.887897 °W, 40.624924 °N) is shown in Table 1. The adjusted R^2 was chosen as an indicator because data size in each location is different, depending on how many rain events can be accepted as valid from 2000–2011 and on how many days the soil surface reflectance can be observed by the MODIS sensor after each rain event. For each location, the data size is thus different. The adjusted R^2 was used as an indicator to avoid the bias of an R^2 estimation that was caused by different data sizes at each location. At this specific location, we can see closer the band is to band 7, the higher R^2 we obtain. When the two bands are closer to each other, the reflectances are more likely influenced by the same factors. Band 6 has the highest R^2 because it is also located in SWIR region [23]. The lowest R^2 is located in band 1.

Table 1. Adjusted coefficient of determination (adjusted-R^2) at a given location between MODIS bands 1–6 and band 7.

	Band 1	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**
Band 7	0.461	0.505	0.496	0.642	0.607	0.915

Pixel location: 87.887897 °W, 40.624924 °N; Number of data samples: 67.

At the regional level, the adjusted-R^2 between bands 1–6 and band 7 are shown in Figure 9. Similar with the result in the pixel level, band 6 has the highest correlation with band 7 over the full study area, with an average adjusted-R^2 of 0.9042, and bands 1 and 3 have a low correlation with band 7, with an average adjusted-R^2 0.4688 and 0.4168, respectively. The low adjusted-R^2 value for each plot indicates that the linear regression model cannot efficiently discribe the data variation at those locations. This has two possible explanations: (1) observation noise in those locations, such as systematic noise, atmospheric conditions and partly vegetated noise, during the 12-year observation period caused the low R^2; (2) changes in soil surface conditions over this long time period caused the low R^2. Figure 9 shows that linear regression models have better performance in NIR and SWIR (Bands 5 and 6), and relatively lower preformance in visible light regions.

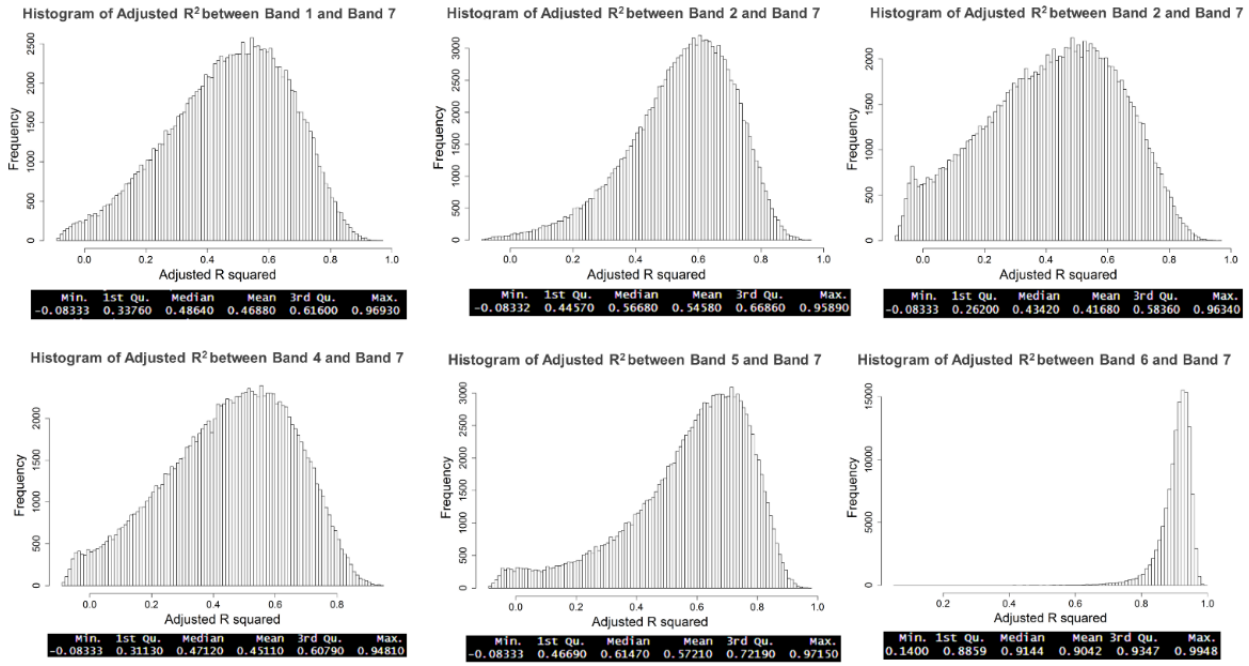

Figure 9. Histogram of adjusted R^2 for each location in entire study area between bands 1–6 and band 7. The highest mean of adjusted R^2 appears in band 6, the lowest in band 1 and band 3.

5.3. Feedback Pattern Gap Filling

Once the two correlations in previous sections are checked, data gaps from overcast days can be filled. Soil reflectance data from clear sky days and a cross validation were used to evalute the prediction preformance. Figure 10 shows the scatterplots of observed values against predicted values for bands 1–6 for the entire study area. bands 3 and 4 are underestimated in some of these validation data because the

two bands are more sensitive to partial vegetation than they are to soil surface moisture change [37]. band 5 has numerous outliers because of the stripe noise in the MODIS terra band 5 dataset. The cut-off edge in the plot of band 1 was due to the elimination of observations over 625 in band 1 to avoid thin cloud data. For band 6 in Figure 10 and band 7 in Figure 8, the direction of most of the prediction is close to a 45-degree line, which indicates that the prediction model captures most of the variety in soil reflectance changes in the two bands. The slight offset of direction in bands 1–5 was caused by the aggregated prediction uncertainty in the two-step prediction. In other words, at one location, if a high level of uncertainty appears when predicting band 7 from CET_0, the high uncertainty will also appear in predicting bands 1–5 at this location.

As described earlier, water content is not the only factor influencing soil surface reflectance in bands 1–5. The model introduced here can explain only part of the data variation from band 1 to band 5, which only concern the influence of water content changes to soil refluence changes during the soil drying process. The points at a distance from the 45-degree line were caused by other factors, which may relate to data noise, soil structure change or soil color change during the soil drying process.

Compared to the visible light region, soil surface reflectance has a higher fluctuation in NIR and SWIR during the soil drying process [8,19,21]. As shown in Figure 10, the range of soil surface change in band 6 during the soil drying process is from 2100–4600, compared to the fluctuation range in band 1 of from 350–650. In order to describe the soil surface reflectance changes during the soil drying process, the prediction accuracy in NIR and SWIR regions is more important than in other regions for describing how soil surface reflectance changes relate to soil surface moisture changes.

The overall RMSE and NRMSE values for the five-fold cross validation are shown in Table 2. The lowest NRMSE is in band 7 because it is directly predicted from evaporation data. band 6 has the second lowest NRMSE because it is highly correlated to band 7. Band 5 has a relatively high NRMSE value due to the strip noise in the original data in MODIS in Terra. The relatively high errors shown in bands 1 and 3 are not surprising because soil surface moisture change is not the only reason for soil reflectance changes in the two bands. Other factors, like soil surface roughness or soil color also influence reflectance change.

Figure 11a illustrates an incomplete dynamic soil feedback pattern with gaps for one location (87.693245 °W, 40.614263 °N) after a rain event starting from 7 April 2003 to 12 April 2003. Only the fourth and fifth day's data were available. Figure 11b shows the dynamic soil feedback pattern after filling in all the data gaps using the proposed approach. This feedback pattern represents the soil surface reflectance changes during this particular rain event at this particular location, and can be further used to understand soil surface properties based on approaches created in previous research [9,10].

Table 2. Overall RMSE and NRMSE from five-fold cross validation for each band prediction.

	Band 1	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**	**Band 7**
RMSE	67.397	101.606	162.355	257.654	391.780	330.659	316.017
NRMSE(%)	18.11%	16.92%	19.08%	16.67%	19.98%	13.04%	10.40%

RMSE has been scaled to 10,000 times real value because initial MODIS data have been scaled for smaller storage.

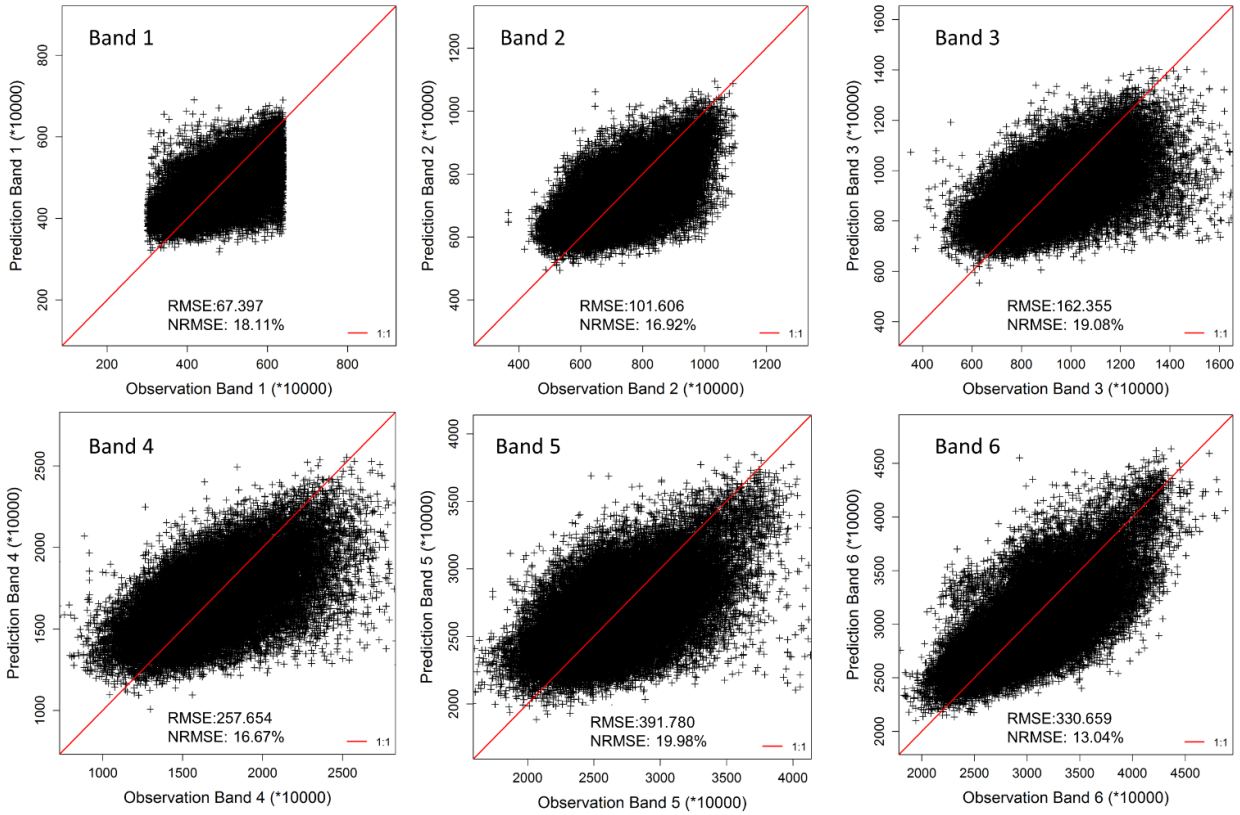

Figure 10. Observation and prediction plot of bands 1–6 based on two correlations of each location in entire study area.

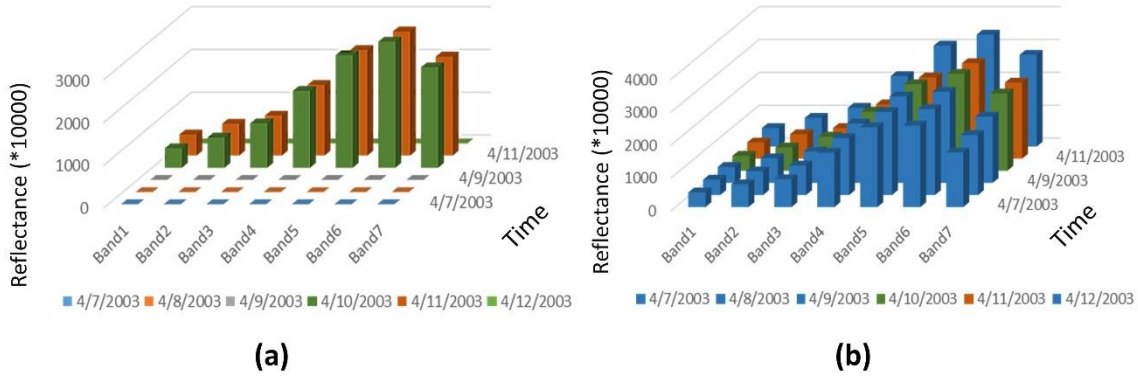

Figure 11. Data-gap filling at a given location (87.693245 °W, 40.614263 °N) in one rain event (from 7 April 2003 to 12 April 2003). (**a**) Incomplete dynamic soil feedback pattern before gap filling; (**b**) Feedback pattern after gap filling. Predicted reflectance is shown with blue bar.

6. Conclusions

This study presents a method for solving the problem of missing data caused by cloud cover in the construction of dynamic soil feedback patterns. The main argument is that the cumulative reference evapotranspiration can be used as auxiliary data to assist gap filling in the feedback pattern. To fill soil reflectance data gaps, a new equation has been derived mathematically to illustrate the relationship

between soil reflectance in water in the sensitive band (MODIS band 7) and the square root of cumulated reference evapotranspiration ($CET_0^{0.5}$). Soil reflectance in MODIS band 7 on an overcast day was predicted using this correlation. The correlation between bands 1–6 and band 7 over multiple rain events for each location was used to predict MODIS band 1 to band 6 on days with missing data.

A case study in flat farmland in northeastern Illinois shows a good relationship between MODIS band 7 and the square root of cumulated reference evapotranspiration ($CET_0^{0.5}$) in most of the bare soil farm land (with average $R^2 = 0.55$, $p < 0.001$; and average NRMSE 10.40%). The five-fold cross validation shows: the approach proposed in this study captured the soil surface reflectance change in bands 6 and 7 during the soil drying process; and the Normalized Root Mean Square Error (NRMSE) is 13.04% and 10.40%, respectively. Acceptable errors are shown in band 1 to band 5; overall NRMSE for those bands is lower than 20%.

The equation derived in the study presents a mathematical explanation of the relationship between soil reflectance and evaporation conditions. This derivation provides a new way to explain soil surface reflectance changes during the soil drying process after rain events. With this new explanation, quantitative and qualitative analyses can be established for further soil reflectance studies. After solving data gaps on overcast days, the soil feedback pattern can be applied to larger flat areas to predict soil properties or soil types using the methodologies presented by Zhu, *et al*. and Liu, *et al*., in previous studies [9,10].

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project No.: 41431177), the Landscape Observation by High Resolution Images project: hydrological monitoring system by high spatial resolution remote sensing image (08-Y30B07-9001-13/15), Natural Science Research Program of Jiangsu (14KJA170001) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. Support to A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, the Manasse Chair Professorship from the University of Wisconsin-Madison, and the "One-Thousand Talents" Program of China are greatly appreciated. We thank all members of the GIS group at the University of Wisconsin-Madison for their encouragement and discussion of the work presented here.

Author Contributions

Shanxin Guo and A-Xing Zhu conceived and designed the experiments; Shanxin Guo and Lingkui Meng performed the experiments; Shanxin Guo analyzed the data; Jim Burt, Fei Du, Jing Liu, Guiming Zhang contributed to the analysis of the result; Shanxin Guo, A-Xing Zhu, Fei Du and Jing Liu wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Moran, M.S.; Inoue, Y.; Barnes, E.M. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ.* **1997**, *61*, 319–346.
2. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52.
3. Croft, H.; Anderson, K.; Kuhn, N.J. Characterizing soil surface roughness using a combined structural and spectral approach. *Eur. J. Soil Sci.* **2009**, *60*, 431–442.
4. Santanello, J.A.; Peters-Lidard, C.D.; Garcia, M.E.; Mocko, D.M.; Tischler, M.A.; Moran, M.S.; Thoma, D.P. Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed. *Remote Sens. Environ.* **2007**, *110*, 79–97.
5. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75.
6. Anderson, K.; Croft, H. Remote sensing of soil surface properties. *Prog. Phys. Geogr.* **2009**, *33*, 457–473.
7. Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **2011**, *162*, 1–19.
8. Lobell, D.B.; Asner, G.P. Moisture effects on soil reflectance. *Soil Sci. Soc. Am. J.* **2002**, *66*, 722–727.
9. Zhu, A.-X.; Liu, F.; Li, B.; Pei, T.; Qin, C.; Liu, G.; Wang, Y.; Chen, Y.; Ma, X.; Qi, F.; *et al*. Differentiation of soil conditions over low relief areas using feedback dynamic patterns. *Soil Sci. Soc. Am. J.* **2010**, *74*, 861–869.
10. Liu, F.; Geng, X.; Zhu, A.-X.; Fraser, W.; Waddell, A. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS. *Geoderma* **2012**, *171–172*, 44–52.
11. Poggio, L.; Gimona, A.; Brown, I. Spatio-temporal MODIS EVI gap filling under cloud cover: An example in Scotland. *ISPRS J. Photogramm. Remote Sens.* **2012**, *72*, 56–72.
12. Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal MODIS-Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130.
13. Boloorani, A.D.; Erasmi, S.; Kappas, M. Multi-Source remotely sensed data combination: Projection transformation gap-fill procedure. *Sensors* **2008**, *8*, 4429–4440.
14. Zhang, C.; Li, W.; Travis, D. Gaps-fill of SLC-off Landsat ETM+ satellite image using a geostatistical approach. *Int. J. Remote Sens.* **2007**, *28*, 5103–5122.
15. Brooks, E.B.; Thomas, V.A.; Wynne, R.H.; Coulston, J.W. Fitting the multitemporal curve: A fourier series approach to the missing data problem in remote sensing analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3340–3353.
16. Pringle, M.J. Robust prediction of time-integrated NDVI. *Int. J. Remote Sens.* **2013**, *34*, 4791–4811.
17. NOAA National Climatic Data Center Weather Data (1981–2010). Available online: http://www.ncdc.noaa.gov/cdo-web/datatools/normals (accessed on 16 May 2014).
18. Ventura, F.; Snyder, R.L.; Bali, K.M. Estimating evaporation from bare soil using soil moisture data. *J. Irrig. Drain. Eng.* **2006**, *132*, 153–158.

19. Muller, E.; Decamps, H. Modeling soil moisture—Reflectance. *Remote Sens. Environ.* **2001**, *76*, 173–180.
20. Liu, W.; Baret, F.; Gu, X.; Tong, Q.; Zheng, L.; Zhang. B. Relating soil surface moisture to reflectance. *Remote Sens. Environ.* **2002**, *81*, 238–246.
21. Somers, B.; Gysels, V.; Verstraeten, W.W.; Delalieux, S.; Coppin, P. Modelling moisture-induced soil reflectance changes in cultivated sandy soils: A case study in citrus orchards. *Eur. J. Soil Sci.* **2010**, *61*, 1091–1105.
22. Fabre, S.; Briottet, X.; Lesaignoux, A. Estimation of soil moisture content from the spectral reflectance of bare soils in the 0.4–2.5 μm domain. *Sensors* **2015**, *15*, 3262–3281.
23. Gladkova, I.; Grossberg, M. Quantitative restoration for MODIS band 6 on Aqua. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1–8.
24. Gallardo, M.; Snyder, R.; Schulbach, K.; Jackson, L. Crop growth and water use model for lettuce. *J. Irrig. Drain. Eng.* **1996**, *122*, 354–359.
25. Wythers, K.R.; Lautnroth, W.K.; Paruelo, J.M. Bare-soil evaporation under semiarid field conditions. *Soil Sci. Soc. Am. J.* **1999**, *63*, 1341–1349.
26. Mellouli, H.; van Wesemael, B.; Poesen, J.; Hartmann, R. Evaporation losses from bare soils as influenced by cultivation techniques in semi-arid regions. *Agric. Water Manag.* **2000**, *42*, 355–369.
27. Ritchie, J. Model for predicting evaporation from a row crop with incomplete cover. *Water Resour. Res.* **1972**, *8*, 1204–1213.
28. Suleiman, A.A.; Ritchie, J.T. Modeling soil water redistribution during second-stage evaporation. *Soil Sci. Soc. Am. J.* **2003**, *67*, 377–386.
29. Lal, R.; Shukla, M.K. *Principles of Soil Physics*; Marcel Dekker: New York, NY, USA, 2004.
30. Stroosnijder, L. Soil evaporation: Test of a practical approach under semi-arid conditions. *Netherlands J. Agric. Sci.* **1987**, *35*, 417–426.
31. Hargreaves, G.H.; Samami, Z.A. Reference crop evapotranspiration from temperature. *Appl. Eng. Agric.* **1985**, *1*, 96–99.
32. Martínez-Cob, A.; Tejero-Juste, M. A wind-based qualitative calibration of the Hargreaves ET_0 estimation equation in semiarid regions. *Agric. Water Manag.* **2004**, *64*, 251–264.
33. Allen, R.G. FAO Irrigation and drainage paper. *Irrig. Drain.* **1998**, *300*, 64–65.
34. Platnick, S.; King, M.D.; Ackerman, S.A.; Menzel, W.P.; Baum, B.A.; Riédi, J.C.; Frey, R.A. The MODIS cloud products: Algorithms and examples from terra. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 459–472.
35. Thornton, P.E.; Running, S.W.; White, M.A. Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.* **1997**, *190*, 214–251.
36. William A, J. *Soil Physics*, 6th ed.; John Wiley: New York, NY, USA, 2004.
37. Yang, W.; Wang, M.; Shi, P. Using MODIS NDVI time series to identify geographic patterns of landslides in vegetated regions. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 707–710.