

# Supplementary Materials: Estimating Soil Moisture with Landsat Data and its Application in Extracting the Spatial Distribution of Winter Flooded Paddies. *Remote Sensing* 2016, 8, 38

Bolun Li, Chaopu Ti, Yongqiang Zhao and Xiaoyuan Yan

## 1. Procrustes analysis

Procrustes analysis is a popular method of statistical shape analysis used to provide least squares matching of two or more sets of landmark data (point matrices represented in given shapes) [1–4]. Ordinary Procrustes analysis matches two sets of landmark data to find out the optimal shape-preserving Euclidian transformations through rotation, translation and uniformly scaling. This transformation minimizes the differences in location between the comparison shape and the target shape [5].

If  $X$  is the target shape matrix of dimension  $n \times p$ ,  $Y$  is a compared shape matrix of dimension  $n \times q$  ( $q \geq p$ ), the transformed matrix  $Z$  is calculated as follows:

$$Z = bYT + a \quad (1)$$

where,  $b$  is a scaling factor for adjusting the size of the matrix,  $T$  is an orthogonal rotation matrix and  $a$  is a constant matrix, used to translate the points. The parameters  $b$ ,  $T$ , and  $a$  are determined by the least squares criterion:

$$\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - Z_{ij})^2 = \min \quad (2)$$

Generalized Procrustes analysis applies Ordinary Procrustes analysis to optimally superimpose  $m$  ( $\geq 3$ ) sets of shapes instead of only two shapes (aligning them to one target shape). With GPA the  $m$  sets can be aligned to each other under least-squares correspondence [6,7].

The consensus shape matrix  $C$  can be initially defined as geometrical centroid of the transformed matrices given by:

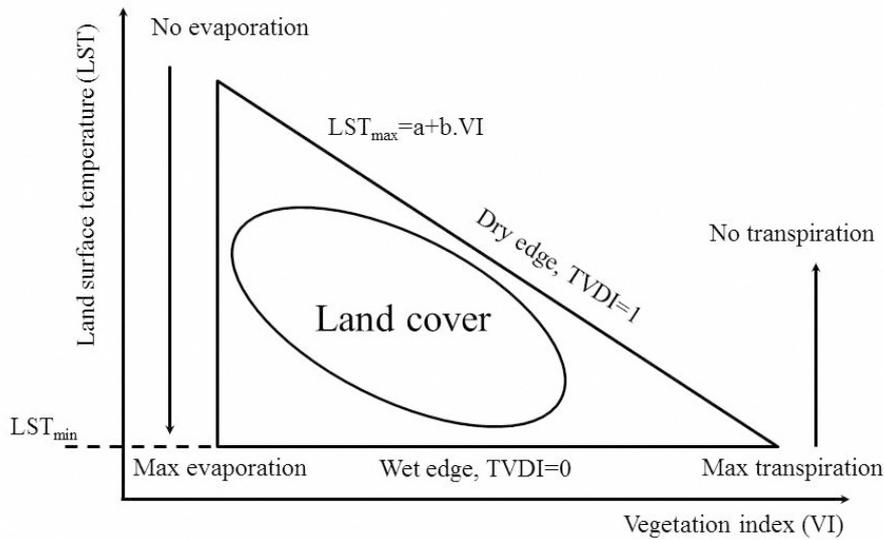
$$C = \frac{1}{m} \sum_{ij=1}^m Z_{ij} \quad (3)$$

The final consensus shape matrix can be obtained using the following minimum condition (least squares objective function) in an iterative computation scheme as follows:

$$\sum_{i=1}^m (Z_{ij} - C)^2 = \min \quad (4)$$

Further details and statement of the algorithm can be referred to the work of Crosilla and Beinat [8] and Beinat and Crosilla [9].

## 2. TVDI



**Figure S1.** Conceptual diagram of LST-NDVI scatterplot for defining Temperature Vegetation Dryness Index (TVDI). Adapted from Sandholt et al. (2002).

On the basis of combining remotely sensed land surface temperature (LST) and normalized difference vegetation index (NDVI) data, the Temperature Vegetation Dryness Index (TVDI) was developed to capture information of surface soil moisture and can be applicable over partially vegetated area [10]. Figure S1 shows a conceptual representation for the TVDI. The scatterplot in LST-NDVI space normally forms a triangular shape [10–12], if a wide range of fractional vegetation cover and soil moisture conditions is covered in the data. In the triangle space, the location of a data point is mainly dominated by soil wetness condition. The wet edge formed by the minimum temperature ( $LST_{\min}$ ) indicates the maximum soil moisture status and evapotranspiration, while the dry edge formed by the maximum temperature ( $LST_{\max}$ ) represents the limiting soil moisture availability and evapotranspiration.  $LST_{\max}$  is defined (originally suggested) as empirical functions of the NDVI. Based on this triangle, the TVDI is originally defined by the following equation:

$$TVDI = \frac{LST - LST_{\min}}{a + b \text{ NDVI} - LST_{\min}} \quad (5)$$

where LST is the observed surface temperature at a given pixel,  $LST_{\max}$  ( $= a + bVI$ ) is the maximum temperature at a given NDVI interval,  $a$  and  $b$  are the regression parameters of the linear regression fit between NDVI and  $LST_{\max}$ ,  $LST_{\min}$  is the minimum temperature in the triangle. The method used for retrieving LST is presented in the Section 3. However, actually, previous studies have revealed that the minimum surface temperature at the wet edge does not remain constant with variations in NDVI values under non-water stress conditions and this will give rise to a overestimation of TVDI at low fractional vegetation cover [13,14]. Consequently, it is unconvincing to use a fixed wet edge to calculate TVDI for all land cover types. Considering the above findings, both wet and dry edges in the NDVI–LST space are linearly regressed in this study and the regressed equations can be given by:

$$LST_{\min} = a_1 + b_1 \times NDVI \quad (6)$$

$$LST_{\max} = a_2 + b_2 \times NDVI \quad (7)$$

where  $LST_{\min}$  is the minimum surface temperature, defining the wet edge,  $LST_{\max}$  is the maximum surface temperature, defining the dry edge,  $a_1$  and  $a_2$  are the intercepts for the wet and dry edges, respectively, and  $b_1$  and  $b_2$  are the slopes for the wet and dry edges, respectively. Accordingly, the TVDI can be calculated according to:

$$TVDI = \frac{LST - (a_1 + b_1 \times NDVI)}{(a_2 + b_2 \times NDVI) - (a_1 + b_1 \times NDVI)} \quad (8)$$

The values of TVDI range from 0 (maximum evapotranspiration) to 1.0 (limited water availability) based on its position in the feature space.

Numerous experiments showed a strong linear relationship between NDVI and plant biomass. Besides, when the fractional vegetation cover is less than 15%, the NDVI is difficult to indicate the regional plant biomass due to the low-vegetation cover. As fractional vegetation cover increased from 15% to 80%, the NDVI values rises rapidly. When the vegetation coverage become greater than 80%, there will be a slighter growth in NDVI with lower sensitivity to vegetation [15,16]. In view of the above, a range of NDVI between 0.15 and 0.75, in which an apparent linear fit of the NDVI–LST space appears (both  $LST_{\max}$  and  $LST_{\min}$  were well linearly regressed), were selected by visual interpretation to achieve the optimal corresponding equations for defining wet edge and dry edge. The pixels with NDVI value of  $< 0.1$  were generally assumed to be water body, cloud or snow, and were thereby ruled out from the regression fitting process [17].

### 3. Land surface temperature retrieval algorithms

In this section, methods for retrieving LST from the Landsat thermal channel of both ETM+ and OLI will be presented.

#### *The Qin et al.'s mono-window algorithm*

Qin *et al.* [18] developed a mono-window (MW) algorithm in order to estimate LST particularly for TM/ETM+:

$$T_s = \left[ a(1 - C - D) + (b(1 - C - D) + C + D)T_{\text{sensor}} - DT_a \right] / C \quad (9)$$

with

$$C = \varepsilon\tau \quad (10)$$

$$D = (1 - \tau)[1 + (1 - \varepsilon)\tau] \quad (11)$$

where  $a = -67.355351$ ,  $b = 0.458606$ ,  $\varepsilon$  represents the land surface emissivity (LSE),  $\tau$  represents the atmospheric transmissivity,  $T_{\text{sensor}}$  is the at-sensor brightness temperature and  $T_a$  is the mean atmospheric temperature (for mid-latitude summer) defined as:

$$T_a = 16.0110 + 0.92621 T_0 \quad (12)$$

$T_0$  is the near-surface air temperature (at about 2 m height). The method used for retrieving LSE

is further presented in the Section 4. The atmospheric transmissivity  $\tau$  can be calculated by the web-based atmospheric correction tool [19,20], in which the control parameters can usually be obtained from local meteorological data.

#### *The split-window algorithm*

The Split-window (SW) techniques basically use the differences in the atmospheric absorption of two thermal infrared (TIR) spectral bands to retrieve the land surface temperature [21]. With the addition of one thermal band compared to Landsat ETM+, the SW algorithm proposed by Jimenez-Munoz *et al.* [22] was applied in this study to obtain LST from Landsat 8 thermal infrared data, in which the LST is given by the following equation:

$$T_s = T_i + c_1(T_i - T_j) + c_2(T_i - T_j)^2 + c_0 + (c_3 + c_4w)(1 - \varepsilon') + (c_5 + c_6w)\Delta\varepsilon \quad (13)$$

where  $T_i$  and  $T_j$  represent the at-sensor brightness temperatures at the two TIR bands  $i$  and  $j$ ,  $c_0$  to  $c_6$  are constant coefficients determined from simulated data,  $\varepsilon'$  ( $= (\varepsilon_i + \varepsilon_j)/2$ ) is the mean emissivity and  $\Delta\varepsilon$  ( $= \varepsilon_i - \varepsilon_j$ ) is the emissivity difference,  $w$  is the total atmospheric water vapor content that can be approximated according to [23]:

$$w = (1.0335 - \tau_{10}) / 0.1134 \quad (14)$$

where  $\tau_{10}$  represents the atmospheric transmissivity at TIRS Band 10. More details for Landsat 8 SW algorithm can be found in the work of Jimenez-Munoz *et al.* [22].

#### **4. Land surface emissivity estimation**

The land surface emissivity (LSE) is an essential parameter to apply the methods mentioned above to a Landsat image, it is important to estimate its value for each pixel prior to performing the MW or SW algorithm. However, as being subjected to the heterogeneity of earth surface materials, the LSE varies remarkably over even short distances [24]. A number of studies have attempted to estimate the emissivity from infrared and visible data for thermal infrared sensors [25,26]. The NDVI Thresholds Method—NDVI<sup>THM</sup>, which is first introduced by Sobrino and Raissouni [25], shows a good working in using the NDVI to indirectly obtain the LSE of the thermal infrared bands in natural vegetation and agricultural areas. However, the NDVI<sup>THM</sup> can merely be applied over surfaces composed by bare soil and vegetation, it fails to work when over other land cover surfaces like water, snow, or manmade materials.

Therefore, an extension of the approach to obtain LSE of these non-vegetation/soil classes of pixels is indispensable in our study area. An operative solution could be to assign an emissivity value for each non-vegetation/soil class according to a classification image as non-vegetation surfaces are normally considered to be homogeneous [27]. To begin this process, we built a decision tree from a given set of thresholds based on different remote sensing indices as indicated by prior studies [28–30] (Figure S2), since decision tree classifier has been proven to be a simple yet widely used classification technique [30]. The study area was then roughly classified into five land cover classes including water body, vegetation, mixture of soil and vegetation, bare soil, and manmade material. It should be noted that a post-classification recoding would be required prior to grouping the land cover types into corresponding emissivity classes. Our approach was to determine the surface emissivity by calculating the average emissivity value of each category of materials (collection of main sample spectra) presented in the MODIS UCSB Emissivity Library

(<http://www.icesb.ucsb.edu/modis/EMIS/html/em.html>) and filtered in terms of ETM+/TIRS band filter function (Table S1). Finally, the improved procedure for retrieving LSE is summarized in Table S2, considering different Land cover types.

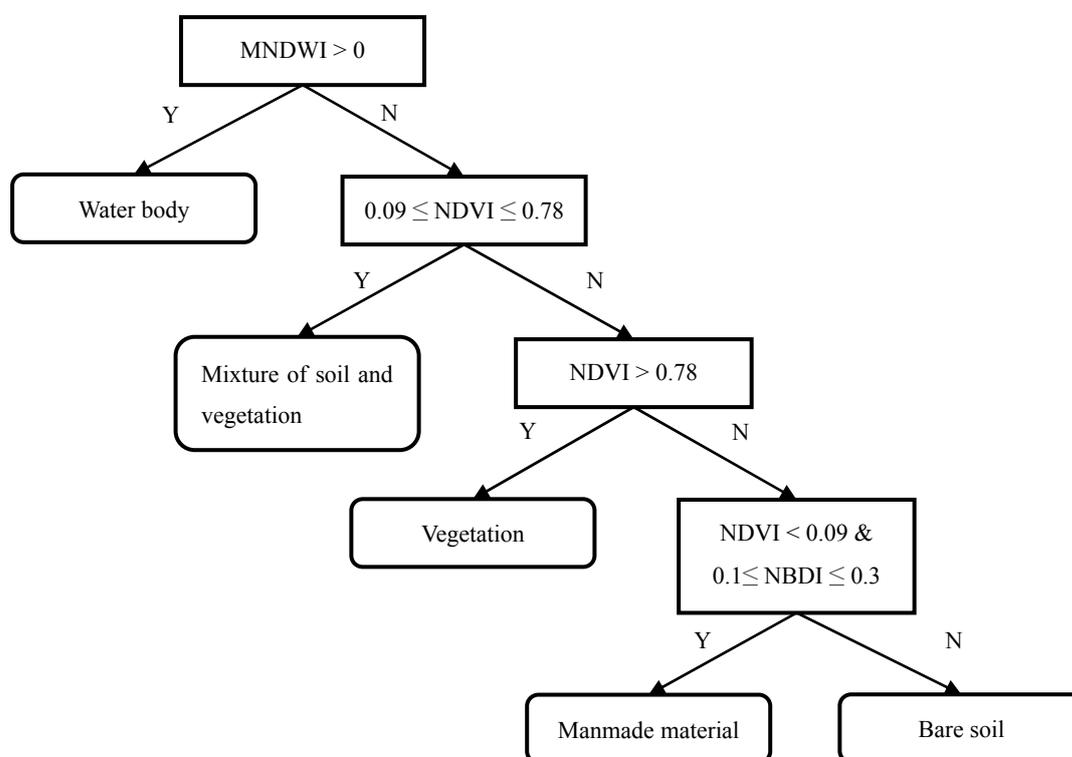


Figure S2. Decision tree classification model.

Table S1. Emissivities of different land cover types for Landsat 8 TIRS band 10 and 11.

Classes	Vegetation	Bare soil	Manmade material	Water body
Band 10	0.986	0.973	0.962	0.993
Band 11	0.988	0.978	0.971	0.987

Table S2. The retrieval of LSE for different cases.

Classes	Expression for LSE
Vegetation	$\epsilon_v + C$
Bare soil	$\epsilon_s$
Mixture of soil and vegetation	$\epsilon_v P_v + \epsilon_s(1 - P_v) + C$
Manmade material	$\epsilon_m$
Water body	$\epsilon_w$

$\epsilon_v$ ,  $\epsilon_s$ ,  $\epsilon_m$  and  $\epsilon_w$  are the vegetation emissivity, the soil emissivity, the manmade material emissivity and the water emissivity, respectively.  $C$  is the cavity term  $= (1 - \epsilon_s) + \epsilon_v F(1 - P_v)$ , where  $F$  is a geometrical factor and a typical value of 0.55 is chosen;  $P_v$  is the vegetation proportion  $= \left( \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$ ,

where  $NDVI_{max} = 0.78$  for fully cover vegetation, and  $NDVI_{min} = 0.09$  for bare soil [31].

## References

1. Crosilla, F. Procrustes analysis and geodetic sciences. In *Geodesy-the Challenge of the 3rd Millennium*; Springer: Heidelberg, Germany, 2003; pp. 287–292.
2. Goodall, C. Procrustes methods in the statistical analysis of shape. *J. R. Stat. Soc. Seri. B (Methodol.)* **1991**, 285–339.
3. Gower, J.C.; Dijksterhuis, G.B. *Procrustes Problems*; Oxford University Press: New York, NY, USA, 2004.
4. Dryden, I.L.; Mardia, K.V. *Statistical Shape Analysis*; Wiley: Chichester, UK, 1998.
5. Schönemann, P.H.; Carroll, R.M. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* **1970**, 35, 245–255.
6. Gower, J.C. Generalized procrustes analysis. *Psychometrika* **1975**, 40, 33–51.
7. Ten Berge, J.F. Orthogonal procrustes rotation for two or more matrices. *Psychometrika* **1977**, 42, 267–276.
8. Crosilla, F.; Beinat, A. Use of generalised procrustes analysis for the photogrammetric block adjustment by independent models. *ISPRS J. Photogramm. Remote Sens.* **2002**, 56, 195–209.
9. Beinat, A.; Crosilla, F. A generalized factored stochastic model for the optimal global registration of lidar range images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2002**, 34, 36–39.
10. Sandholt, I.; Rasmussen, K.; Andersen, J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* **2002**, 79, 213–224.
11. Carlson, T.N.; Gillies, R.R.; Perry, E.M. A method to make use of thermal infrared temperature and ndvi measurements to infer surface soil water content and fractional vegetation cover. *Remote Sens. Rev.* **1994**, 9, 161–173.
12. Gillies, R.R.; Kustas, W.P.; Humes, K.S. A verification of the 'triangle' method for obtaining surface soil water content and energy fluxes from remote measurements of the normalized difference vegetation index (NDVI) and surface e. *Int. J. Remote Sens.* **1997**, 18, 3145–3166.
13. Moran, M.S.; Clarke, T.R.; Inoue, Y.; Vidal, A. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* **1994**, 49, 246–263.
14. Patel, N.R.; Anapashsha, R.; Kumar, S.; Saha, S.K.; Dadhwal, V.K. Assessing potential of modis derived temperature/vegetation condition index (TVDI) to infer soil moisture status. *Int. J. Remote Sens.* **2008**, 30, 23–39.
15. Percent vegetation coverPurevdorj, T.S.; Tateishi, R.; Ishiyama, T.; Honda, Y. Relationships between percent vegetation cover and vegetation indices. *Int. J. Remote Sens.* **1998**, 19, 3519–3535.
16. Yao, C.; Zhang, Z.; Wang, X. Evaluating soil moisture status in xinjiang using the temperature vegetation dryness index (TVDI). *Remote Sens. Technol. Appl.* **2005**, 19, 473–478 (In Chinese).
17. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sens. Environ.* **2012**, 118, 83–94.
18. Qin, Z.; Karnieli, A.; Berliner, P. A mono-window algorithm for retrieving land surface temperature from landsat tm data and its application to the israel-egypt border region. *Int. J. Remote Sens.* **2001**, 22, 3719–3746.
19. Barsi, J.A.; Barker, J.L.; Schott, J.R. An atmospheric correction parameter calculator for a single thermal band earth-sensing instrument. In Proceedings of the IGARSS IEEE International

- Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003; pp. 3014–3016.
20. Barsi, J.A.; Schott, J.R.; Palluconi, F.D.; Hook, S.J. Validation of a web-based atmospheric correction tool for single thermal band instruments. *Proc. SPIE* **2005**, *5882*, doi:10.1117/12.619990.
  21. McMillin, L.M. Estimation of sea surface temperatures from two infrared window measurements with different absorption. *J. Geophys. Res.* **1975**, *80*, 5113–5117.
  22. Jimenez-Munoz, J.C.; Sobrino, J.A.; Skokovic, D.; Mattar, C.; Cristobal, J. Land surface temperature retrieval methods from landsat-8 thermal infrared sensor data. *Geosci. Remote Sens. Lett. IEEE* **2014**, *11*, 1840–1843.
  23. Rozenstein, O.; Qin, Z.; Derimian, Y.; Karnieli, A. Derivation of land surface temperature for landsat-8 tirs using a split window algorithm. *Sensors* **2014**, *14*, 5768–5780.
  24. Salisbury, J.W.; D'Aria, D.M. Emissivity of terrestrial materials in the 8–14  $\mu\text{m}$  atmospheric window. *Remote Sens. Environ.* **1992**, *42*, 83–106.
  25. Sobrino, J.A.; Raissouni, N. Toward remote sensing methods for land cover dynamic monitoring: Application to morocco. *Int. J. Remote Sens.* **2000**, *21*, 353–366.
  26. Van De Griend, A.A.; Owe, M. On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. *Int. J. Remote Sens.* **1993**, *14*, 1119–1131.
  27. Sobrino, J.A.; Jiménez-Muñoz, J.C.; Sòria, G.; Romaguera, M.; Guanter, L.; Moreno, J.; Plaza, A.; Martínez, P. Land surface emissivity retrieval from different VNIR and TIR sensors. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 316–327.
  28. Chen, X.-L.; Zhao, H.-M.; Li, P.-X.; Yin, Z.-Y. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sens. Environ.* **2006**, *104*, 133–146.
  29. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033.
  30. Pang-Ning, T.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*; Library of Congress: Library Washington, DC, USA, 2006; p 74.
  31. Zhang, R.; Tian, J.; Su, H.; Sun, X.; Chen, S.; Xia, J. Two improvements of an operational two-layer model for terrestrial surface heat flux retrieval. *Sensors* **2008**, *8*, 6165–6187.

