Sensitivity Analysis of Hot/Cold Pixel Selection in SEBAL Model for ET Estimation

Leyang Feng

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
In
Geography

Yang Shao, Chair
James B. Campbell
Zachary Easton

April 30th 2015
Blacksburg, VA

Keywords: ET, SEBAL, Hot/cold pixel

Sensitivity Analysis of Hot/Cold Pixel Selection in SEBAL Model for ET Estimation

Leyang Feng

ABSTRACT

The objective of this study was to evaluate the sensitivity of instantaneous latent heat flux (LE) estimation from Surface Energy Balance Algorithm for Land (SEBAL) by changing hot/cold pixel selections. The SEBAL model was programed in a Matlab environment and applied to Lower Fox Watershed in northeast Illinois using two Landsat 5 Thematic Mapper images acquired in summer 2006. Unlike most previous studies where hot/cold pixel were manually selected by image analysts, we emphasized an automated hot/cold pixel selection based on land cover map, normalized difference vegetation index (NDVI) map, and land surface temperature (LST) map. Various combinations of hot/cold pixels were automatically selected along the LST gradient. The LE estimations were then validated against ground-based eddy covariance observation. Results show that the LE estimations from SEBAL were sensitive to both hot and cold pixel selections and tend to be more sensitive to cold pixel selection. The absolute percentage difference (APD) of LE estimation compared with field observation data can range from 0.67% to 67.2% by varying hot and cold pixel combinations. The location of hot/cold pixels appears to have minor impact on SEBAL LE estimation. The LE estimation become acceptable (APD < 10%) when using the hot/cold pixels with a slightly higher/lower LST than LST extremes from the study area. This study provides insights into hot/cold pixel selection and the sensitivity of SEBAL-based LE estimation. Future research on SEBAL ET estimation should focus on enhancing automated hot/cold pixel selection algorithm to improve the model's operational use.

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to Dr. Yang Shao as my master's advisor, for his invaluable support and constant guidance during my two years of MS program. Working under his supervision has been an amazing learning experience. His hard work, determination, commitment to the advancement of science, and kindness has been very inspiring to me. I thank him for his patience and insightful suggestions for my research.

I am grateful to my graduate committee members, Dr. James B. Campbell, and Dr. Zachary Easton, for their constant support, encouragement and detailed guidance that has been an invaluable input for my work.

I would like to thank Geography Department for my two years wonderful time staying here and my fellow graduate students in the department for their help and kindness: Suwen Zhao, Jie Ren, Haitao Wang, and Brandon Wheeler. I also want to thank my friends that have been a constant source of support and fun and for being there when I needed them the most.

My special thanks go to Dan Cheng and my family in China for their love and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Land surface evapotranspiration (ET) is the sum of evaporation from soil and transpiration from plant to the atmosphere. The process of evapotranspiration involves large amount of water and heat transfer between land surface and lower atmosphere (Allen, Tasumi and Trezza 2007c, Long and Singh 2013), therefore ET is a critical component of the hydrologic cycle (Long, Singh and Li 2011). Accurate estimation of ET and understanding of its spatial-temporal dynamics are critical to many applications in developing a water balance, drought monitoring, crop yield forecasting, precision agriculture, and water resource management in general (Bastiaanssen et al. 2005, Bastiaanssen and Ali 2003, Roerink et al. 1997, Ines et al. 2006, Li et al. 2013, Meng et al. 2014, Awan and Ismaeel 2014, Courault, Seguin and Olioso 2005).

ET at the field scale can be directly measured using conventional techniques such as lysimeter system (Cole 1968, Ritchie and Burnett 1968, Wright 1990, Evett et al. 2012), energy balance Bowen ratio (BR) (Prueger et al. 1997, Cragoa and Brutsaert 1996, Devitt et al. 1998, Todd et al. 2000), and eddy covariance (EC) system (Dugas et al. 1991, Baldocchi 2003, Goulden et al. 1996, Williams et al. 2004). These field measurements are often limited by sample size and the extrapolation of ET point measurement over a large area can be challenging. Over the last three decades, advances in remote sensing techniques provide an opportunity to obtain ET in an indirect approach but offer the ability of estimating ET across a variety of spatial and temporal scales (Courault et al. 2005, Gowda et al. 2007). From remotely sensed data, vegetation index, surface albedo, land surface temperature, and other variables critical to ET estimation can

be routinely derived from visible, near-infrared, and thermal spectral bands of various satellite systems (e.g., Landsat and MODIS).

A large number of remote sensing-based ET models rely on energy balance theory. A review of existing energy balance-based ET models was provided by Gowda et al. (2007). A range of ET models were summarized and compared: Surface Energy Balance Index (SEBI) developed by Meneti and Choudhary (1993), Surface Energy Balance System (SEBS) developed by Su (2002), Simplified Surface Energy Balance Index (S-SEBI) developed by Roerink, Su and Menenti (2000), Two-Source Energy Balance model (TSEB) developed by Anderson et al. (1997), Surface Energy Balance Algorithm for Land (SEBAL) developed by Bastiaanssen et al. (1998c), and other SEBAL-like ET estimation models including Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) (Allen et al. 2007c), T-SEBAL (Wang et al. 2014), and M-SEBAL (Long and Singh 2012).

Among various energy balance-based ET models, the Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen et al. 1998c, Bastiaanssen et al. 1998j) is one of the mostly widely used (Gowda et al. 2007). SEBAL is designed to solve the Latent Heat (LE) consumed by ET as a residual from surface energy balance equation:

$$LE = R_n - G - H$$

where R_n is the net solar radiation, G is the sensible heat conducted into the ground, and H is the sensible heat transferred to air. R_n and G can be estimated using surface albedo, vegetation index, and surface emissivity derived from remote sensing data (Dubayah 1992, Wang, White and Robinson 2000, Teixeira et al. 2009, Su 2002, Bastiaanssen et al. 1998c). H is estimated assuming a linear relationship between surface temperature and

near surface temperature so that the heat transfer from ground to air can be quantified. SEBAL has been validated across different regions of the world (Bastiaanssen 2000, Lenk 2013, Immerzeel, Gaur and Zwart 2008, Li and Zhao 2010, Yao, Han and Xu 2010). Based on the past validation results compiled from applications in more than 30 countries around the world, Bastiaanssen et al. (2005) reported that the overall accuracy of SEBAL ET estimation was from 67% to 95% for instantaneous ET and 70% to 98% for daily ET. The accuracy increased to 95% on a seasonal or annual integrated ET estimation (Bastiaanssen et al. 2010).

The SEBAL model requires a clear sky image to perform ET estimation, thus cloud contamination or haze is one of the main concerns for the model's operational use. With a user selected cloud free image, the uncertainty of SEBAL model can be attributed to each component of R_n, G, and H flux estimation and error accumulations. For instance in the net solar radiation calculation, the surface albedo and emissivity are used to calculate the amount of outgoing shortwave radiation and longwave radiation, respectively. The momentum roughness length, an important variable for sensible heat flux (H) calculation, is generally derived from leaf area index and may have large uncertainties (Paul et al. 2014c). Sensitivity analyses of SEBAL parameters have been conducted by a number of researchers (e.g., Wang et al. (2009) and Long et al. (2011)). For example, Long et al. (2011) tested H estimation against different input variables including net solar radiation, roughness length, domain size, pixel resolution, and the temperature of selected extreme pixels. They suggested that one of the most challenging procedures in SEBAL model was to select two extreme pixels, hot pixel and cold pixel,

to determine the boundary hydro-thermal conditions. Similar findings were reported by Wang et al. (2009), Trezza, Allen and Tasumi (2013), Wang et al. (2014).

The hot pixel needs to be identified in a dry agricultural field where ET is assumed to be zero. Cold pixel should be located in a well-irrigated agricultural field representing the maximum ET. In practice, the hot/cold pixels are difficult to identify. The cold pixel can be hard to find in arid and semiarid regions if there is not enough irrigation over agricultural land. The hot pixel can also be difficult to identify when limited dry or bare agricultural land exist during growing season or in a humid region. In most previous studies, image analysts manually performed the selections of extreme pixels. Image visual interpretation by different image analysts may vary substantially and lack of local knowledge often leads to unsatisfactory SEBAL model results. Therefor an automated approach of extreme pixel selection is highly desirable, especially in cases where SEBAL model needs to be implemented operationally for large study areas.

The overall objective of this study was to quantify the sensitivity of SEBAL LE estimation by changing combinations of hot and cold pixel selection and evaluate uncertainties in LE estimation related to changing pixel pairs. We proposed an automated framework of hot/cold selection based on land surface temperature (LST), Normalized Difference Vegetation Index (NDVI), and land cover map. Hot/cold pixels were automatically selected and gradually adjusted following LST gradients. SEBAL LE estimation from various hot/cold pixel pairs were then compared to field measures to identify the best combination of hot/cold pixel selection.

SEBAL Background

Model Description

SEBAL computes LE as residual of surface energy balance equation:

$$LE = R_n - G - H \quad (1)$$

The net solar radiation (R_n) is a sum of all incoming and outgoing shortwave and longwave radiation, computed as:

$$R_n = (1-\alpha)R_{si} + (R_{li} - R_{lo}) - (1-\epsilon)R_{li} \quad (2)$$

where R_{si} is the incoming shortwave radiation (W/m^2), R_{li} is the incoming longwave radiation (W/m^2), R_{lo} is the outgoing longwave radiation (W/m^2), α is the surface albedo, and ε is the surface emissivity, which can be approximated as a function of Normalized Difference Vegetation Index (NDVI) (Bastiaanssen et al. 1998c) and Leaf Area Index (LAI) (Allen et al. 2007c).

The soil heat flux, G, is considered as a fraction of R_n in SEBAL, computed as:

$$G = cR_n \quad (3)$$

the coefficient c, can be estimated using land surface temperature (LST), surface albedo, and NDVI based the empirical equation developed by Bastiaanssen et al. (1998c):

$$c = \frac{Ts}{\alpha}(0.0038\alpha + 0.0074\alpha^2)(1 - 0.98NDVI^4) \quad (4)$$

Sensible heat flux (H) is estimated using:

$$H = \rho C_p \frac{dT}{r_{ah}} \quad (5)$$

where ρ is the air density (kg/m^3), C_p is the air specific heat (1004 J/Kg/K), dT is the difference between the aerodynamic temperature and the air temperature, and r_{ah} is the aerodynamic resistance to heat transport (s/m).

SEBAL assumes a linear relationship between dT and LST for the entire image:

$$dT = aT_s + b \quad (6)$$

The coefficients a and b are obtained through an iterative process using two anchor pixels, the so-called hot pixel and cold pixel, representing the extreme hydro-thermal conditions on the image. A hot pixel is assumed to be on a dry or bare soil surface with no soil water thus LE is assumed to be 0. A cold pixel is selected on water body or well-irrigated cropland with 0 energy in H.

The value of rah is calculated as:

$$r_{ah} = \frac{\ln(z_2/z_1)}{u^* k} \quad (7)$$

where z_1 (m) is the zero plane displacement height, z_2 (m) is the reference height above the plant canopy, u^* is the friction velocity (m/s), and k is the von Karman constant (0.4). The value of u^* can be computed as:

$$u^* = \frac{u(z)k}{\ln(\frac{z}{z_{om}})} \quad (8)$$

where u(z) (m/s) is the wind speed at reference height z (m), z_{om} (m) is the roughness length, which can be estimated using NDVI for all pixels (Bastiaanssen et al. 1998a, Allen et al. 2007c).

A Review on Hot/cold Pixel Selection

Selection of hot/cold pixel is critically important for performing SEBAL model for ET estimation since the two extreme pixels play a key role in estimating H. Theoretically, the hot pixels should be located in somewhere with no ET. Pixels with high surface temperature on a landsat scene normally can be found on land cover types such as impervious surfaces, sandy soils, and dry/bare agricultural fields. Previous SEBAL modeling efforts suggested that dry and bare cropland is favored since the relationship between dT and LST on urban land cover may be nonlinear and more

complicated to define. As for cold pixels, they should be selected in somewhere with maximum ET. Two approaches are often applied in cold pixel selection. One approach is to select cold pixels from water body (Bastiaanssen et al. 2005, Bastiaanssen et al. 1998c). In this case the H is assumed to be 0 for the cold pixel meaning all available energy is used to evaporate water. The second approach is to select cold pixels from a well-irrigated and fully vegetated patch, following the assumption:

$$LE = R_n - G - 1.05ET_r$$

where ET_r is the hourly tall crop reference ET calculated using the Standardized ASCE Penman-Monteith equation and local meteorological observations. This approach is described and applied in Mapping Evapotranspiration with Internalized Calibration Model (METRIC Model) published by Allen et al. (2007c). The first approach is supported by a number of published studies. For example, Long et al. (2011) pointed out that the selection of cold pixels from water body is acceptable since only LST is involved in the procedure of determining coefficients a and b. However, other variables in SEBAL model, such as the surface roughness length and net solar radiation of the cold pixels are also part of the ET computation therefore the location and surface characteristics of cold pixels cannot be ignored. Several investigations also addressed that cold pixels selected from open water cause problem in estimating the heat storage (Timmermans et al. 2007).

Most published studies manually selected hot/cold pixels by image analysts and descriptions on specific pixel selection procedure were often brief. Finding appropriate hot/cold pixels can be very challenging without existing knowledge about area of interest. Consequently, there have been a few recent studies suggesting the development of an automated pixel selection framework. For example, Conrad et al. (2007) and Allen et al.

(2007b) suggested using LST and LAI to locate the potential candidates of hot/cold pixels as low LAI values and high LST characterize hot pixels while high LAI and low LST characterize cold pixels. Similar principles were adapted for automated calculation of ET over European forest using MODIS data, but instead of using LST and LAI, a combination of LST and NDVI was used as an indicator of vegetation cover and water conditions of each pixel.

Methods

Study site

This study was conducted in the Lower Fox watershed of north Illinois, USA. The Lower Fox watershed covers an area of 2856.08 Km^2, located in latitude between 41.35 ℃N and 42.06 ℃N and longitude between 88.19 ℃W and 89.06 ℃W. Elevation increasing from southeast to northwest, ranging from 134m to 323m, with a mean elevation of 219m. The average annual temperature in the Lower Fox watershed ranges from 7.75 ℃ to 11.75 ℃ according to the weather records from 1989 to 2004 at Illinois Climate Network St. Charles Station. The mean annual precipitation is 783mm and the mean annual pan evaporation is 967mm. The dominant land use types in Lower Fox Watershed include urban areas (22%) and agricultural lands (69%). The remaining area is covered by forest (6%), water (1%) and others (2%). More than 48% of the watershed belongs to hydrologic group C, associated with slow infiltration. About 47% of the watershed soils belong to hydrologic group B with a moderate infiltration rate.

Data

Remotely sensed imagery, meteorological data from AmeriFlux site, eddy covariance flux data from AmeriFlux site, digital elevation model (DEM), and National Land Cover Dataset 2006 (NLCD 2006) were used in this study.

Two cloud-free Landsat 5 TM images acquired at 10: 28 am local standard time on July 31st (DOY212) and August 16th (DOY228) were downloaded from the USGS EarthExplorer (http://earthexplorer.usgs.gov/). Landsat 5 TM images were used to derive the surface characteristics such as surface albedo, surface emissivity, net solar radiation,

surface temperature, and NDVI. The atmospheric correction was not conducted because the internal calibration of SEBAL (Allen et al. 2007c).

Meteorological data and eddy covariance flux data on and around Landsat 5 image acquisition dates were obtained from AmeriFlux Site US-IB1(41.86 ℃N, 88.22 ℃W), located in northwest of the Lower Fox watershed. The meteorological monitoring instruments as well as eddy covariance flux tower were installed on a corn/soybean rotation agricultural field at elevation of 226.5m. The tower was equipped with a variety of sensors to measure turbulent fluxes of latent heat (LE), soil heat (G), sensible heat (H), and radiation including outgoing/incoming solar radiation and net solar radiation, which can be used for SEBAL output validation. LE, G, air temperature and wind speed were measured at height of 4.05m. Radiation components were measured at 2.4m. All data were collected at 30 minutes interval. The Meteorological data and eddy covariance flux data at AmeriFlux Site US-IB1 are available from 2005 to 2011 and can be accessed through AmeriFlux Site and Data Exploration System (http://ameriflux.ornl.gov/).

The Digital elevation model (DEM) and National Land Cover Dataset 2006 (NLCD 2006) were obtained from USDA Geospatial Data Gateway (https://gdg.sc.egov.usda.gov/). Both the DEM and NLCD2006 are at 30m resolution.

Model Implementation

Detailed description of SEBAL's theoretical basis, validation, and application can be found in published papers of Bastiaanssen et al. (1998c), Bastiaanssen et al. (1998j). As a modified ET estimation model stemming from SEBAL, the published studies about METRIC (Allen et al. 2007c, Allen et al. 2007a) also provide important information regarding parameter estimation and calculation. However, the actual SEBAL modules or

source code are not available to public. In this study, SEBAL model was implemented in a Matlab environment and arranged in 44 modules to perform data preparation, variable estimation, extreme pixel selection, ET calculation, and sensitivity analysis. The structure and arrangement of all 44 computation modules can be viewed in Figure 4.

Hot/Cold Pixels Selection and Automation

In this study, we followed Verstraeten's approach of using LST and NDVI to characterize hot/cold pixels (Verstraeten et al. 2005). We used NLCD 2006 data to extract agricultural lands (row crops), which provides the image mask for hot/cold pixel selection. In addition, small agricultural patches were removed and edges of every patch were eliminated to make sure that hot/cold pixels could be selected from relatively large homogenous patches. A scatter plot in NDVI - LST space for agricultural area was then used to identify potential pixel candidates. For hot pixel selection, pixels with relatively high LST and low NDVI values should be investigated. A threshold of bottom 10% for NDVI values was used to identify low NDVI values. The first hot pixel was then identified as the pixel with the highest LST. The selection of cold pixel is similar to hot pixel selecting procedure. The cold pixel should be selected from pixel group with low LST and high NDVI. A threshold of top 10% of NDVI value was used to identify the high NDVI values. Then the first cold pixel was identified as the coldest one in terms of LST.

Sensitivity Analysis

In addition to the above two extreme hot/cold pixels, we added four different hot/cold pixels for sensitivity analysis. For the hot pixels, four more pixels were selected from the extreme high LST value to the lower temperature following LST gradient. We

used a -5% step to determine specific LST of hot pixel. The 5% interval was computed as:

$$5\%\ interval = 0.05(T_{s,max} - T_{s,min}) \quad (8)$$

where $T_{s,max}$ is the highest LST among cropland pixels and $T_{s,min}$ is the lowest LST among cropland pixels. Similar approach was applied for additional cold pixel selection following LST gradient originated from the coldest pixel. SEBAL sensitivity analysis was then performed using the different combination of those 10 pixels.

Results

Pixel Selection Results

Selected hot and cold pixels for each image date are plotted in NDVI – LST space shown in Figure 6. Examples of hot and cold pixels selected from automated selection procedure are shown in Figure 7 and Figure 8, respectively. The hot pixel example in Figure 7 was selected as 5% lower in temperature than the hot pixel extreme. The LST map indicates the selected hot pixel representing a high LST of 308.80K, while the NDVI map shows this hot pixel is located in a low NDVI area with a relatively low NDVI value of 0.22. Visual interpretation in false color composite of Landsat 5 TM image confirmed the pixel is located in the center of a dry/bare agricultural field, which meets the selection criterions discussed in section 2.2. The cold pixel example in Figure 8 was selected as 5% higher in temperature than the cold extreme. The cold pixel represents a relatively low LST (298.6K) shown in LST map and a high NDVI value (0.81) shown in the NDVI map. Other pixels in the same patch also have high NDVI values indicating the whole patch is a potentially well-irrigated agricultural field. Visual interpretation in false color composite of Landsat 5 TM image confirmed the pixel is located in the center of an agricultural field.

It should be noted that multiple pixel candidates might exist for any given LST value. Figure 9 shows an example of six different pixel locations for a same LST value defined by the automated selection approach. A sensitivity analysis regarding the spatial location of these candidate pixels was conducted. Table 4 indicates that different locations yield no obvious difference in LE estimation. This provides information to SEBAL users that when dealing with multiple pixel candidates, pixel location is not

important since SEBAL mainly rely on specific LST values of hot/cold pixels. Therefore, the automated pixel selecting framework should focus on LST gradients, rather than pixel locations.

Sensitivity to hot/cold pixels selection

Total of 25 different combinations for 5 hot pixels and 5 cold pixels were provided to SEBAL model for images DOY 212 and DOY228, respectively, to analyze sensitivity of SEBAL to varying hot and cold pixel selections. Figure 10 compares LE estimation from SEBAL model and ground-based LE observation; Table 5 and Table 6 show absolute percentage difference (APD) at the pixel where the AmeriFlux Site US-IB1 is located. As shown in Figure 10, different combinations of hot and cold pixels can lead to large differences in LE estimation. The APD range from 0.68% to 67.17% on DOY 212 and 0.41% to 76.23% on DOY 228, respectively.

Theoretically, the lowest APD should occur when two extreme pixels (the hottest and the coldest pixel) were used since those two pixels anchoring the boundary conditions for H and LE estimation of the rest pixels. Other studies (Wang et al. 2014, Papadavid et al. 2013, Long and Singh 2013) also suggested that selection of the hottest and coldest pixels might yield the most accurate ET estimation. However, according to our investigation, using the extreme pixels did not necessarily result in most accurate estimations. For image on DOY 212, the APD is 48.95% when using the extreme pixels, but the APD dropped to 5.34% when using pixel with 5% lower temperature than hot extreme and pixel with 5% warmer temperature than cold extreme pixels. Similar situation also occurred on image DOY 228, the APD decreased from 15.56% to 7% when changing the extreme pixels to pixel with 5% lower/higher temperature than the

hottest/coldest pixels. For both images, the use of extreme pixels (the hottest and coldest pixels) generated high APD in instantaneous LE estimations. The error percentages were acceptable (under 10%) when we selected pixels with 5% lower/higher in temperature than hottest/coldest pixels.

For a given cold pixel, the APD of LE estimation tends to increase when the temperature of hot pixel decreases. And similarly, for a given hot pixel, the APD of LE estimation increase rapidly when the temperature of cold pixel increases (Table5 and Table 6), indicating the output LE estimation of SEBAL tends to be more sensitive to a changing cold pixel, which could be caused by the procedure of R_n calculation. The temperature of cold pixel is involved in incoming longwave radiation estimation which would cause variation on R_n calculation. Thus, the selection of proper cold pixel not only had effects on adjusting the coefficient a and b in the $dT - T_s$ linear relationship but also affected R_n estimation. Table 7 shows the R_n estimation when using different cold pixels and comparison between those R_n values to in situ observations.

Model performance

The spatial distribution of instantaneous LE over cropland in Lower Fox Watershed is shown in Figure 11. Figure 11 (a) and (b) was generated using 2006 Landsat 5 TM image DOY 212 and DOY 228. The hot and cold pixel pairs used in each image were selected by the automated framework described above. The hot pixel temperature was 5% lower than the highest LST for both images and the cold pixel temperature was 5% higher than the lowest LST for both images. For image date DOY 212, the LE estimation from SEBAL is 434.53W/m^2 at the pixel where the AmeriFlux Site US-IB1 is located. The eddy covariance measured from AmeriFlux Site US-IB1 LE

at 10:30 CST was 412.4 W/m^2. The APD between SEBAL LE estimation and in situ observation is 5.4%. The average LE over cropland was about 360.55 W/m^2. For image date DOY 228, the LE estimation from SEBAL is 259.19W/m^2 at AmeriFlux Site US-IB1. The LE estimation US-IB1 site at 10:30 CST was 278.7W/m^2. The APD between SEBAL LE estimation and in situ observation is 7%. The average LE over cropland was about 174.19W/m^2. The comparison of other variables between SEBAL estimated value and ground-based observation value is presented in Table 3.

The pixels containing negative value of LE estimation are mainly located in the center of several dry/bare agricultural fields and on the edge of an agricultural field patch. For pixels in dry/bare agricultural field, negative values for LE estimation may occur due to systematic errors caused by various assumptions in the energy balance process. For negative pixels on the edge of cropland patches, the negative values may be caused by spectral mixture and image classification error.

Discussion

The sensitivity analysis results indicate that LE estimation of SEBAL is very sensitive to the selection and use of $T_{s,hot}$ and $T_{s,cold}$. Those two extreme pixels determine the accuracy of SEBAL estimation and different extreme pixel pairs lead to large magnitude change of LE estimation. For image on DOY 212, a 2.03K decrease in hot pixel temperature resulted in a 25.42% mean variation in LE estimation. For the image on DOY 212, a 2.78K decrease in hot pixel temperature results in an 11.49% mean variation in LE estimation. The temperature change in cold pixel may cause even larger variation in LE estimation. For the image on DOY 212, a 1.75K increase in cold pixel temperature generated a 43.16% mean variation in LE estimation. For the image on DOY 228, a 3.07K decrease in cold pixel temperature generated a 36.76% mean variation in LE estimation.

The comparison between the SEBAL estimated LE value and the ground-based LE observation value can only be validated at the pixel where the AmeriFlux Site US – IB1 is located. Since the use of pixel pair with 5% lower/higher in temperature than hot/cold extremes generated accurate LE estimation for the validation pixel, we assume the LE estimation based on that pixel pair to be ground truth for the rest pixels in the image. The spatial distribution of error can then be generated using 'ground truth' LE estimation subtracted by LE estimation using other hot/cold pixel pairs. Figure 12 shows the LE estimation difference over cropland between LE 'ground truth' and LE estimation using pixel pair with 10% lower/higher in temperature than hot/cold extremes. Overall, the LE estimation with 10% lower/higher temperature tends to underestimate LE values across the cropland. The large positive values indicate underestimating of LE on dry/bare

agricultural fields and the large negative values indicate overestimating for agricultural fields with relatively high NDVI values. The substantial errors in two extreme hydro-thermal conditions are not surprising because the selected hot and cold pixels (+-10%) defined boundary conditions of LE estimation. Pixels with higher/lower temperature than the user-defined boundary could have large errors in LE estimation.

Selection of hot and cold pixels is critically important to perform the SEBAL model in a proper manner. Improper pixel selection may cause over 50% of error compared to ground based ET measurement in this study. The automated framework proposed in our study mainly used LST and NDVI map and threshold values to identify hot/cold pixel candidates, yielding acceptable LE estimation, especially when non-extreme pixels were used as model input. We note that the threshold values used for NDVI image may need to change when this method is implanted for different images. A histogram of NDVI values for cropland can be used in determining the threshold. If NDVI values are more stretched, a threshold of bottom 10% to 30% may apply for hot pixel selection. If NDVI values are more condensed in high values according to its histogram, a threshold of top 10% to 5% may apply for cold pixel selection. An alternative is to use actual NDVI values as threshold (e.g. <0.2 for hot pixel selection and >0.8 for cold pixel selection).

It is claimed by the SEBAL model developer that the selection of extreme pixels requires skill and experience; therefore the selection can be difficulty and subjectivity (Allen et al. 2007c, Bastiaanssen et al. 2010, Trezza et al. 2013). A number of other studies (Long et al. 2011, Conrad et al. 2007, Verstraeten et al. 2005, Paul et al. 2014a) confirmed this claim. However, manual selection of hot/cold pixel can be subjective and

limit the model's operational use for large study areas. In our study, a pixel pair with 5% lower/higher than hottest/coldest pixel generated the best LE estimation. This approach has potential for automated pixel selection and SEBAL modeling, although further investigations for different growing seasons or locations need to be conducted.

Conclusion

In this study, an automated framework of hot and cold pixels selection for SEBAL LE estimation was proposed by using LST, NDVI and NLCD maps. The pixels selected using this framework were used as input to quantify the sensitivity of SEBAL LE estimation to different combination of hot/cold pixel pairs. These results show: 1. The LE estimation from SEBAL is sensitive to selection of both hot and cold pixels and tends to be more sensitive to cold pixels. 2. As the automated approach may select multiple qualified hot/cold pixels, the use of certain pixel among candidates has little impact on LE estimation. 3. For end user with no prior knowledge of extreme pixel selection, the temperature of pixels with 5% lower/higher than hottest/coldest pixels are suggested to use as model input in order to yield reasonable LE estimation.

References

Allen, R. G., M. Tasumi, A. Morse, R. Trezza, J. L. Wright, W. Bastiaanssen, W. Kramber, I. Lorite & C. W. Robison (2007a) Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - Applications. *Journal of Irrigation and Drainage Engineering-Asce,* 133**,** 395-406.

Allen, R. G., M. Tasumi & R. Trezza (2007c) Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - Model. *Journal of Irrigation and Drainage Engineering-Asce,* 133**,** 380-394.

Anderson, M., J. Norman, G. Diak, W. Kustas & J. Mecikalski (1997) A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. *Remote sensing of environment,* 60**,** 195-216.

Awan, U. K. & A. Ismaeel (2014) A new technique to map groundwater recharge in irrigated areas using a SWAT model under changing climate. *Journal of Hydrology,* 519**,** 1368-1382.

Baldocchi, D. D. (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology,* 9**,** 479-492.

Bastiaanssen, W., H. Pelgrum, J. Wang, Y. Ma, J. Moreno, G. Roerink & T. Van der Wal (1998a) A remote sensing surface energy balance algorithm for land (SEBAL).: Part 2: Validation. *Journal of hydrology,* 212**,** 213-229.

Bastiaanssen, W., B. Thoreson, B. Clark & G. Davids (2010) Discussion of "Application of SEBAL Model for Mapping Evapotranspiration and Estimating Surface Energy Fluxes in South-Central Nebraska" by Ramesh K. Singh, Ayse Irmak, Suat Irmak, and Derrel L. Martin. *Journal of Irrigation and Drainage Engineering-Asce,* 136**,** 282-283.

Bastiaanssen, W. G. M. (2000) SEBAL-based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. *Journal of Hydrology,* 229**,** 87-100.

Bastiaanssen, W. G. M. & S. Ali (2003) A new crop yield forecasting model based on satellite measurements applied across the Indus Basin, Pakistan. *Agriculture Ecosystems & Environment,* 94**,** 321-340.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes & A. A. M. Holtslag (1998c) A remote sensing surface energy balance algorithm for land (SEBAL) - 1. Formulation. *Journal of Hydrology,* 212**,** 198-212.

Bastiaanssen, W. G. M., E. J. M. Noordman, H. Pelgrum, G. Davids, B. P. Thoreson & R. G. Allen (2005) SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering-Asce,* 131**,** 85-93.

Bastiaanssen, W. G. M., H. Pelgrum, J. Wang, Y. Ma, J. F. Moreno, G. J. Roerink & T. van der Wal (1998j) A remote sensing surface energy balance algorithm for land (SEBAL) - 2. Validation. *Journal of Hydrology,* 212**,** 213-229.

Cole, D. W. (1968) A system for measuring conductivity, acidity, and rate of water flow in a forest soil. *Water resources research,* 4**,** 1127-1136.

Conrad, C., S. Dech, M. Hafeez, J. Lamers, C. Martius & G. Strunz (2007) Mapping and assessing water use in a Central Asian irrigation system by utilizing MODIS remote sensing products. *Irrigation and Drainage Systems,* 21**,** 197-218.

Courault, D., B. Seguin & A. Olioso (2005) Review on estimation of evapotranspiration from remote sensing data: From empirical to numerical modeling approaches. *Irrigation and Drainage Systems,* 19**,** 223-249.

Cragoa, R. & W. Brutsaert (1996) Daytime evaporation and the self-preservation of the evaporative fraction and the Bowen ratio. *Journal of Hydrology,* 178**,** 241-255.

Devitt, D. A., A. Sala, S. Smith, J. Cleverly, L. Shaulis & R. Hammett (1998) Bowen ratio estimates of evapotranspiration for Tamarix ramosissima stands on the Virgin River in southern Nevada. *Water Resources Research,* 34**,** 2407-2414.

Dubayah, R. (1992) Estimating net solar radiation using Landsat Thematic Mapper and digital elevation data. *Water resources research,* 28**,** 2469-2484.

Dugas, W., L. Fritschen, L. Gay, A. Held, A. Matthias, D. Reicosky, P. Steduto & J. Steiner (1991) Bowen ratio, eddy correlation, and portable chamber measurements of sensible and latent heat flux over irrigated spring wheat. *Agricultural and forest meteorology,* 56**,** 1-20.

Evett, S. R., R. C. Schwartz, T. A. Howell, R. L. Baumhardt & K. S. Copeland (2012) Can weighing lysimeter ET represent surrounding field ET well enough to test flux station measurements of daily and sub-daily ET? *Advances in Water Resources,* 50**,** 79-90.

Goulden, M. L., J. W. Munger, S. M. FAN, B. C. Daube & S. C. Wofsy (1996) Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy. *Global change biology,* 2**,** 169-182.

Gowda, P., J. Chavez, P. Colaizzi, S. Evett, T. Howell & J. Tolk (2007) Remote sensing based energy balance algorithms for mapping ET: Current status and future challenges. *Transactions of the ASABE,* 50**,** 1639-1644.

Immerzeel, W. W., A. Gaur & S. J. Zwart (2008) Integrating remote sensing and a process-based hydrological model to evaluate water use and productivity in a south Indian catchment. *Agricultural Water Management,* 95**,** 11-24.

Ines, A. V. M., K. Honda, A. Das Gupta, P. Droogers & R. S. Clemente (2006) Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. *Agricultural Water Management,* 83**,** 221-232.

Lenk, O. (2013) Satellite based estimates of terrestrial water storage variations in Turkey. *Journal of Geodynamics,* 67**,** 106-110.

Li, S. B. & W. Z. Zhao (2010) Satellite-based actual evapotranspiration estimation in the middle reach of the Heihe River Basin using the SEBAL method. *Hydrological Processes,* 24**,** 3337-3344.

Li, Z. Y., X. H. Liu, T. X. Ma, D. Kejia, Q. P. Zhou, B. Q. Yao & T. L. Niu (2013) Retrieval of the surface evapotranspiration patterns in the alpine grassland-wetland ecosystem applying SEBAL model in the source region of the Yellow River, China. *Ecological Modelling,* 270**,** 64-75.

Long, D. & V. P. Singh (2012) A modified surface energy balance algorithm for land (M-SEBAL) based on a trapezoidal framework. *Water Resources Research,* 48.

Long, D. & V. P. Singh (2013) Assessing the impact of end- member selection on the accuracy of satellite- based spatial variability models for actual evapotranspiration estimation. *Water Resources Research,* 49**,** 2601-2618.

Long, D., V. P. Singh & Z. L. Li (2011) How sensitive is SEBAL to changes in input variables, domain size and satellite sensor? *Journal of Geophysical Research-Atmospheres,* 116.

Meneti, M. & B. Choudhary (1993) Parameterization of land surface evapotranspiration using a location dependent potential evapotranspiration and surface temperature range. Exchange Processes at the land Surface for a Range of Space and time series, Bolle, HJ, Feddes, RA, and Kalma, JD. *International Association of Hydrological Sciences Publication,* 212**,** 561-568.

Meng, L., D. Long, S. M. Quiring & Y. J. Shen (2014) Statistical analysis of the relationship between spring soil moisture and summer precipitation in East China. *International Journal of Climatology,* 34**,** 1511-1523.

Papadavid, G., D. Hadjimitsis, L. Toulios & S. Michaelides (2013) A Modified SEBAL Modeling Approach for Estimating Crop Evapotranspiration in Semi-arid Conditions. *Water Resources Management,* 27**,** 3493-3506.

Paul, G., P. H. Gowda, P. V. V. Prasad, T. A. Howell, R. M. Aiken & C. M. U. Neale (2014a) Investigating the influence of roughness length for heat transport (z(oh)) on the performance of SEBAL in semi-arid irrigated and dryland agricultural systems. *Journal of Hydrology,* 509**,** 231-244.

Paul, G., P. H. Gowda, P. V. Vara Prasad, T. A. Howell, R. M. Aiken & C. M. U. Neale (2014c) Investigating the influence of roughness length for heat transport (zoh) on the performance of SEBAL in semi-arid irrigated and dryland agricultural systems. *Journal of Hydrology,* 509**,** 231-244.

Prueger, J. H., J. L. Hatfield, J. K. Aase & J. L. Pikul (1997) Bowen-ratio comparisons with lysimeter evapotranspiration. *Agronomy Journal,* 89**,** 730-736.

Ritchie, J. T. & E. Burnett (1968) A precision weighing lysimeter for row crop water use studies. *Agronomy Journal,* 60**,** 545-549.

Roerink, G. J., W. G. M. Bastiaanssen, J. Chambouleyron & M. Menenti (1997) Relating Crop Water Consumption to Irrigation Water Supply by Remote Sensing. *Water Resources Management,* 11**,** 445-465.

Roerink, G. J., Z. Su & M. Menenti (2000) S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere,* 25**,** 147-157.

Su, Z. (2002) The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences Discussions,* 6**,** 85-100.

Teixeira, A., W. G. M. Bastiaanssen, M. D. Ahmad & M. G. Bos (2009) Reviewing SEBAL input parameters for assessing evapotranspiration and water productivity for the Low-Middle Sao Francisco River basin, Brazil Part A: Calibration and validation. *Agricultural and Forest Meteorology,* 149**,** 462-476.

Timmermans, W. J., W. P. Kustas, M. C. Anderson & A. N. French (2007) An intercomparison of the surface energy balance algorithm for land (SEBAL) and the two-source energy balance (TSEB) modeling schemes. *Remote Sensing of Environment,* 108**,** 369-384.

Todd, R. W., S. R. Evett & T. A. Howell (2000) The Bowen ratio-energy balance method for estimating latent heat flux of irrigated alfalfa evaluated in a semi-arid, advective environment. *Agricultural and Forest Meteorology,* 103**,** 335-348.

Trezza, R., R. G. Allen & M. Tasumi (2013) Estimation of actual evapotranspiration along the Middle Rio Grande of New Mexico using MODIS and Landsat imagery with the METRIC Model. *Remote Sensing,* 5**,** 5397-5423.

Verstraeten, W. W., F. Veroustraete & J. Feyen (2005) Estimating evapotranspiration of European forests from NOAA-imagery at satellite overpass time: Towards an operational processing chain for integrated optical and thermal sensor data products. *Remote Sensing of Environment,* 96**,** 256-276.

Wang, J., T. W. Sammis, V. P. Gutschick, M. Gebremichael & D. R. Miller (2009) SENSITIVITY ANALYSIS OF THE SURFACE ENERGY BALANCE ALGORITHM FOR LAND (SEBAL). *Transactions of the Asabe,* 52**,** 801-811.

Wang, J., K. White & G. J. Robinson (2000) Estimating surface net solar radiation by use of Landsat-5 TM and digital elevation models. *International Journal of Remote Sensing,* 21**,** 31-43.

Wang, X. G., W. Wang, D. Huang, B. Yong & X. Chen (2014) Modifying SEBAL Model Based on the Trapezoidal Relationship between Land Surface Temperature and Vegetation Index for Actual Evapotranspiration Estimation. *Remote Sensing,* 6**,** 5909-5937.

Williams, D., W. Cable, K. Hultine, J. Hoedjes, E. Yepez, V. Simonneaux, S. Er-Raki, G. Boulet, H. De Bruin & A. Chehbouni (2004) Evapotranspiration components determined by

stable isotope, sap flow and eddy covariance techniques. *Agricultural and Forest Meteorology,* 125**,** 241-258.

Wright, J. L. (1990) Comparison of ET measured with neutron moisture meters and weighing lysimeters.

Yao, W., M. Han & S. G. Xu (2010) Estimating the regional evapotranspiration in Zhalong wetland with the Two-Source Energy Balance (TSEB) model and Landsat7/ETM+ images. *Ecological Informatics,* 5**,** 348-358.

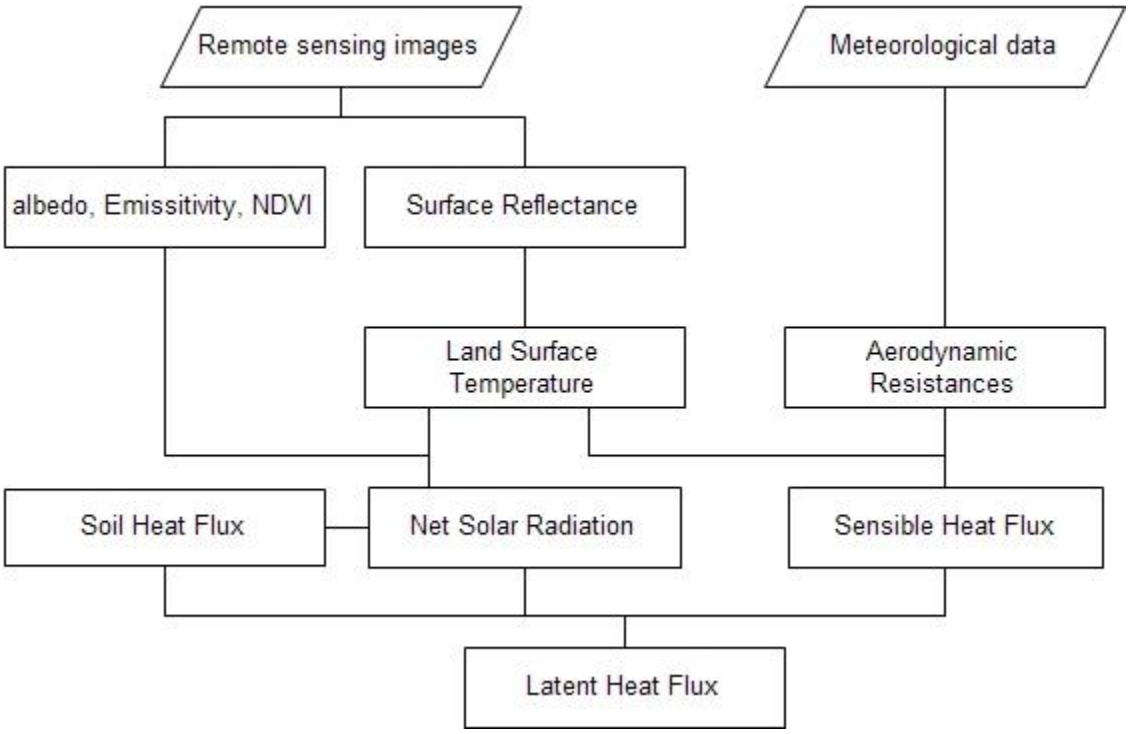

Figure 1 Flowchart showing the logic of SEBAL ET estimation

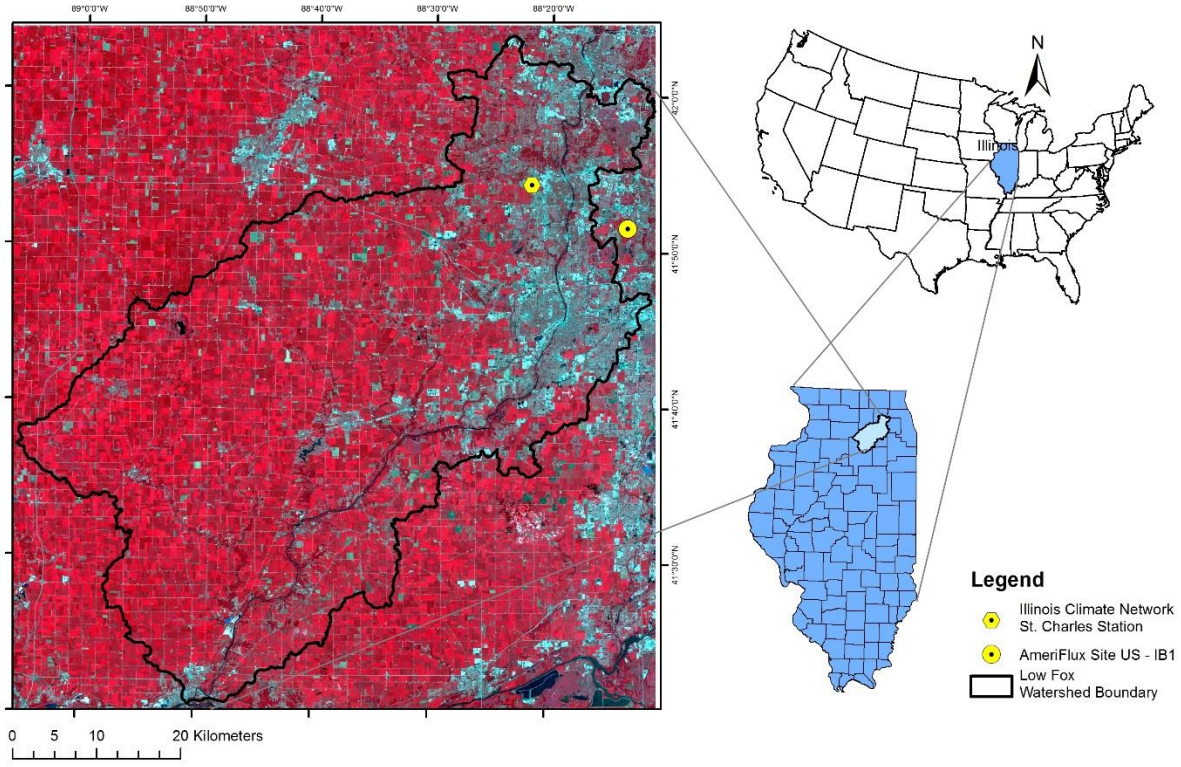

Figure 2 Location and false color composite of Landsat 5 TM images in experimental Site. The Lower Fox Watershed is delineated in black. The AmeriFlux Site US – IB1 and Illinois Climate Network St. Charles Station is shown in yellow marks.

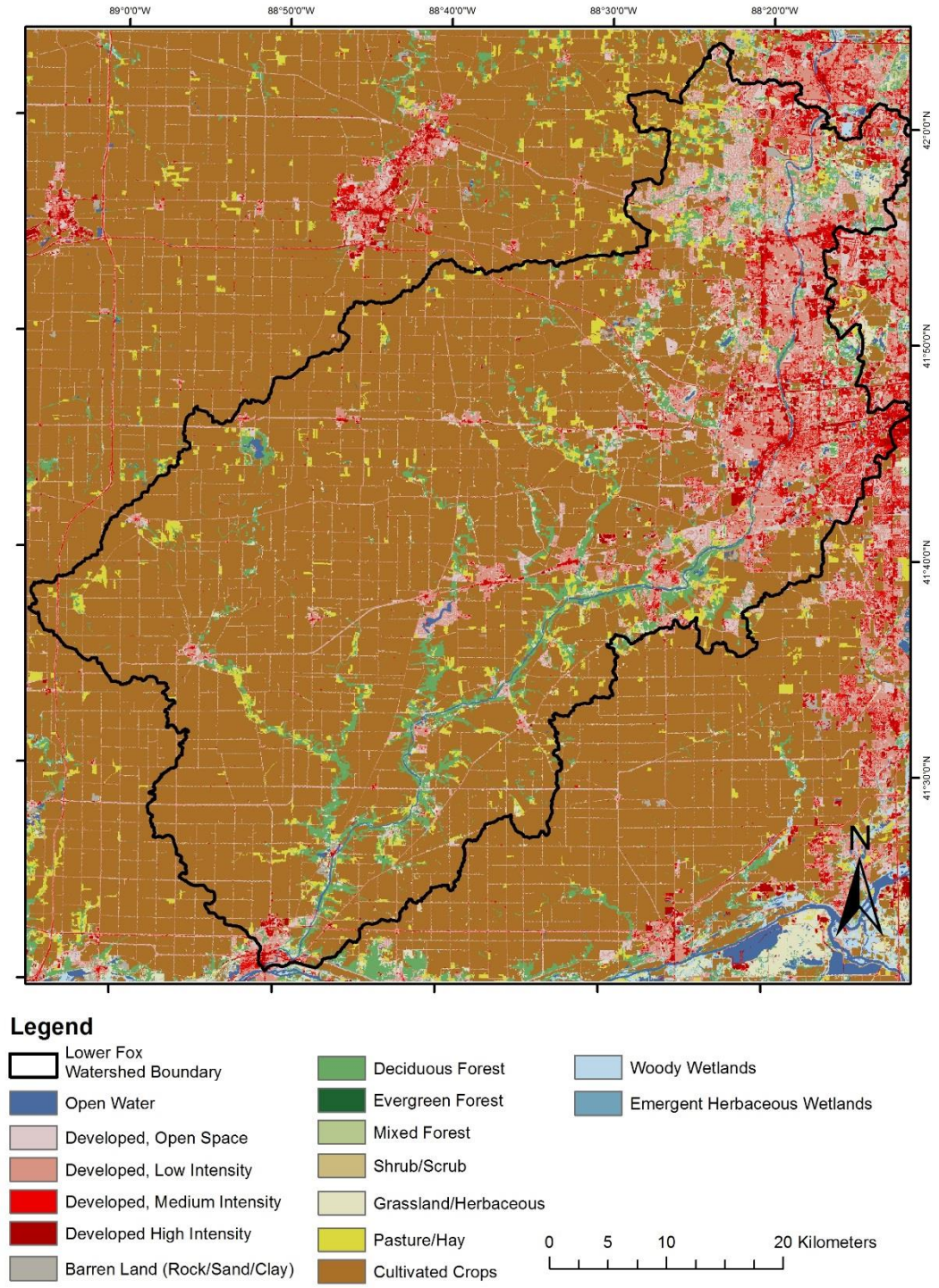

Figure 3 Land use and land cover type of Lower Fox Watershed from NLCD 2006 Dataset

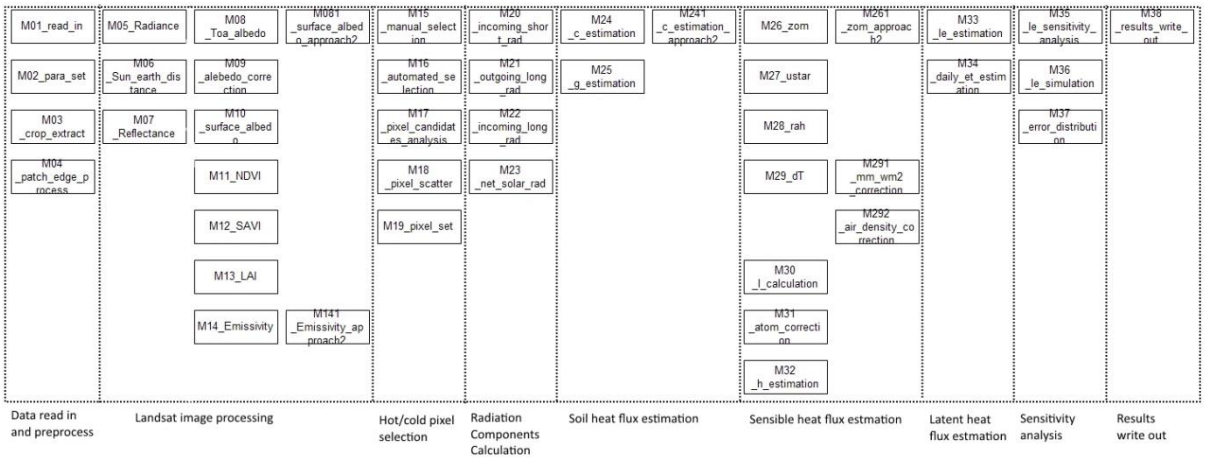

Figure 4 Structure and arrangement of Matlab coding modules

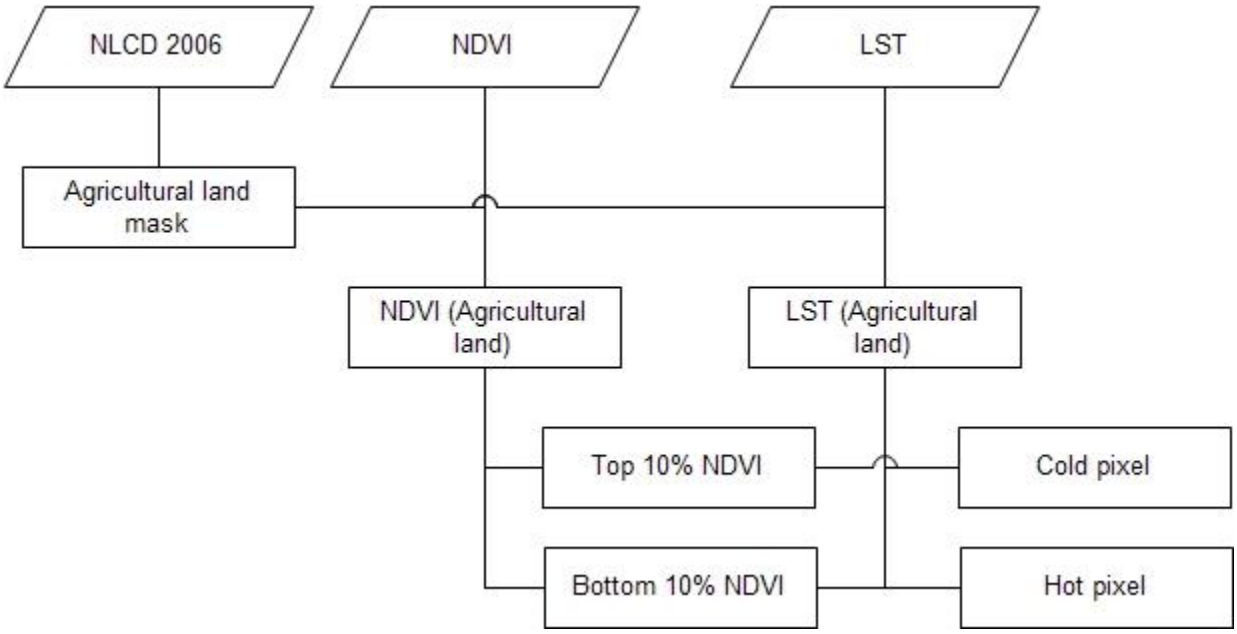

Figure 5 Automated framework flowchart

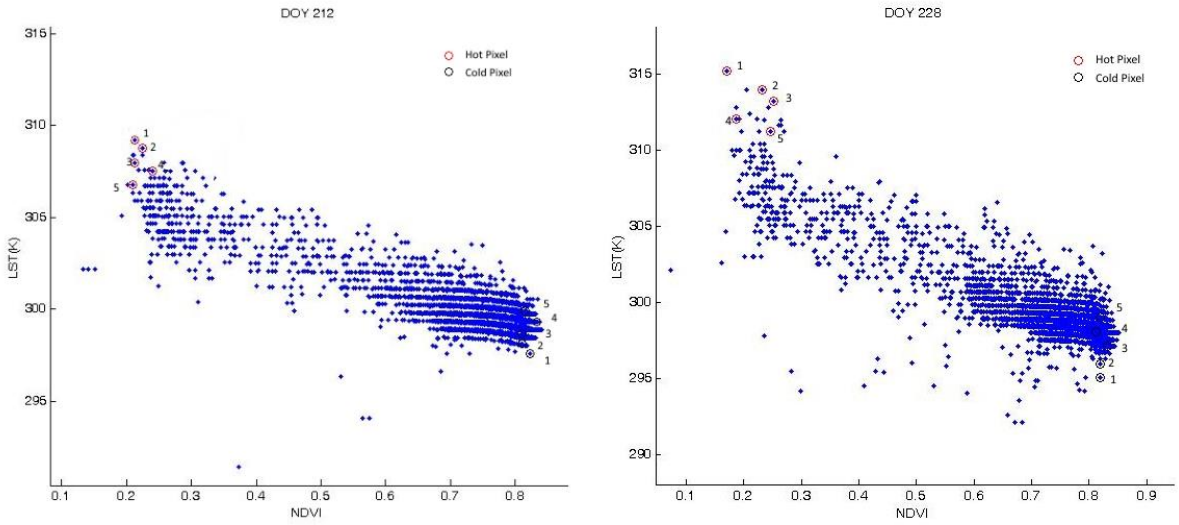

Figure 6 Scatterplots of LST – NDVI space for Lower Fox Watershed on DOY 212 and DOY 228. Numbered red circles represent hot pixels selected by automated approach, and numbered black circles represent cold pixels selected by automated approach.

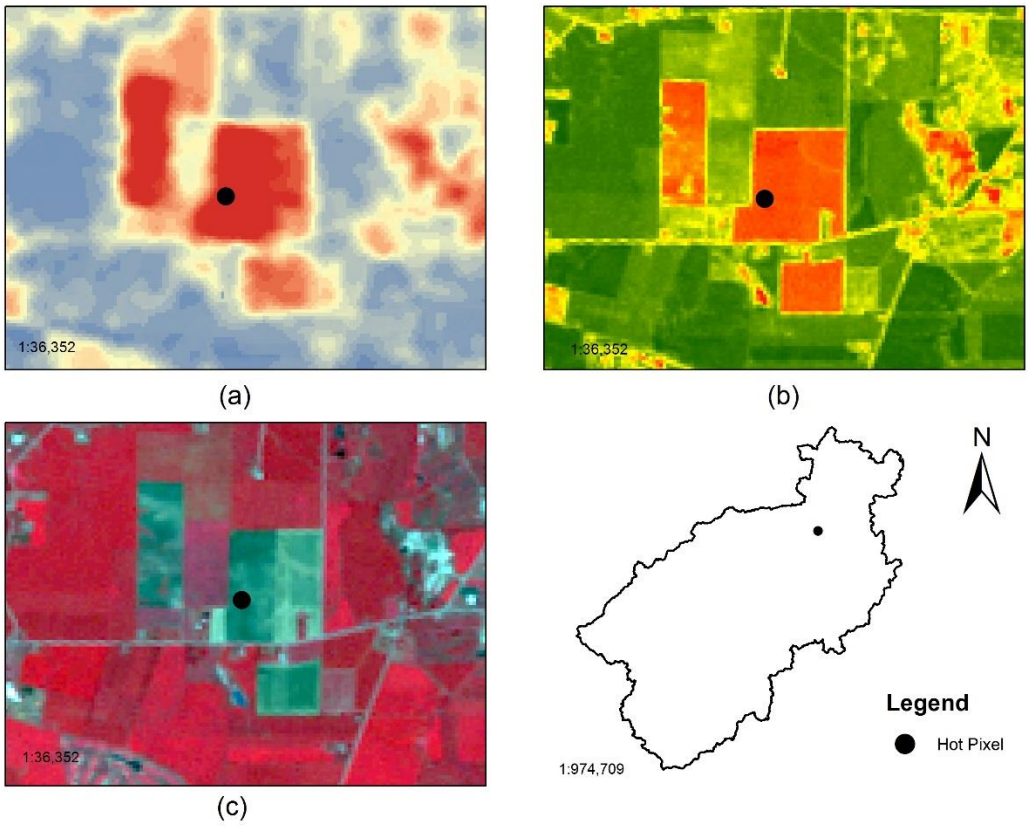

Figure 7 Hot pixel example selected by automated approach and displayed in LST map (a), NDVI map (b), and false color composite of Landsat 5 TM images (c).

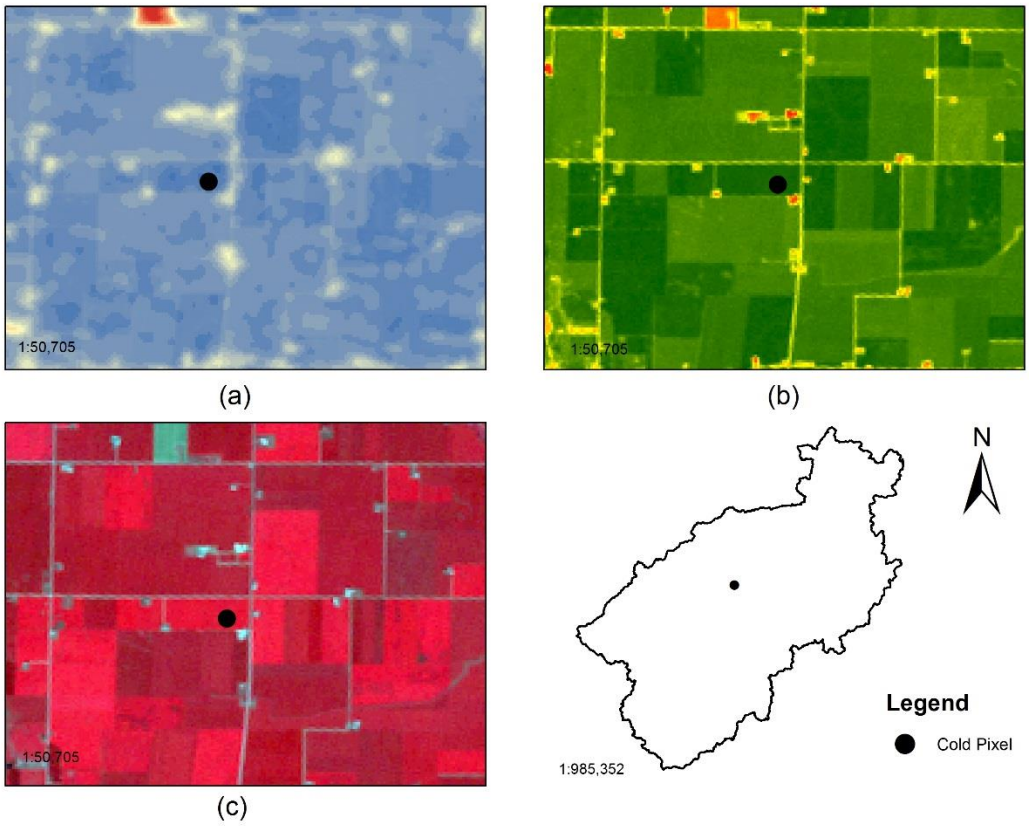

Figure 8 Cold pixel example selected by automated approach and displayed in LST map (a), NDVI map (b), and false color composite of Landsat 5 TM images (c).

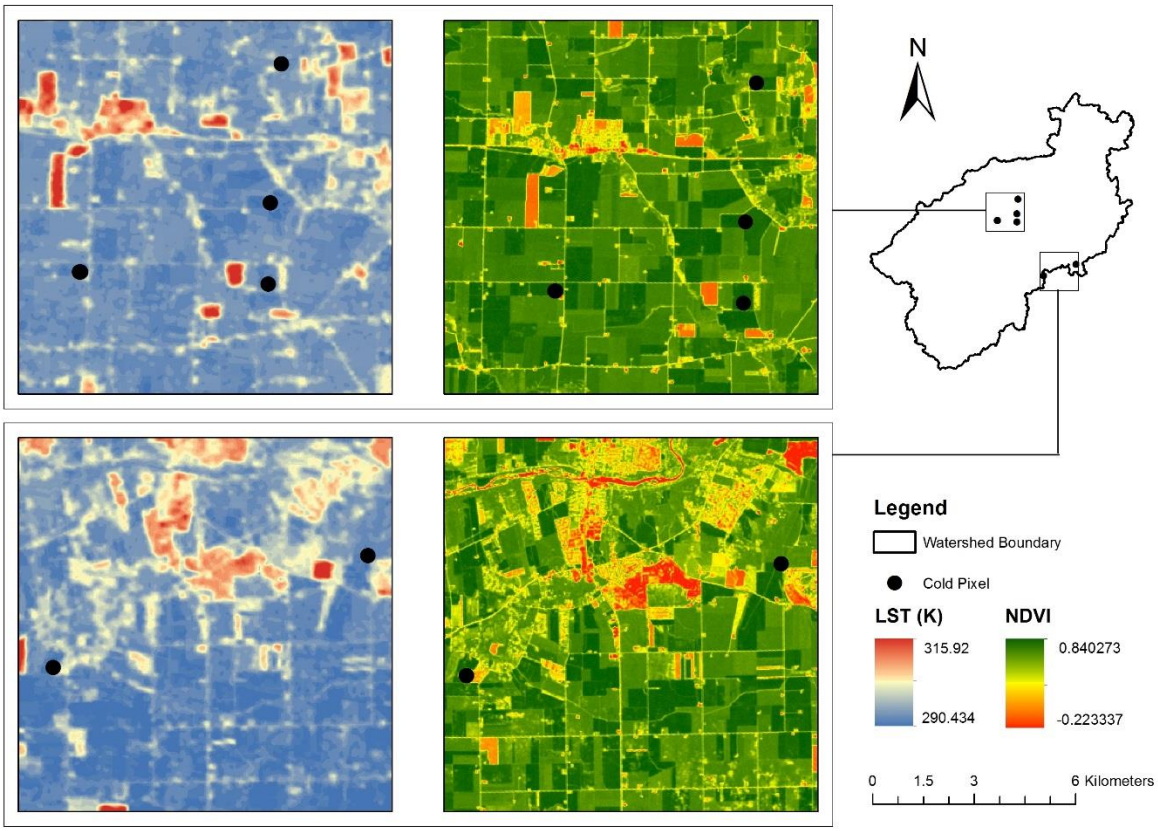

Figure 9 Multiple cold pixel candidates selected by automated selection approach

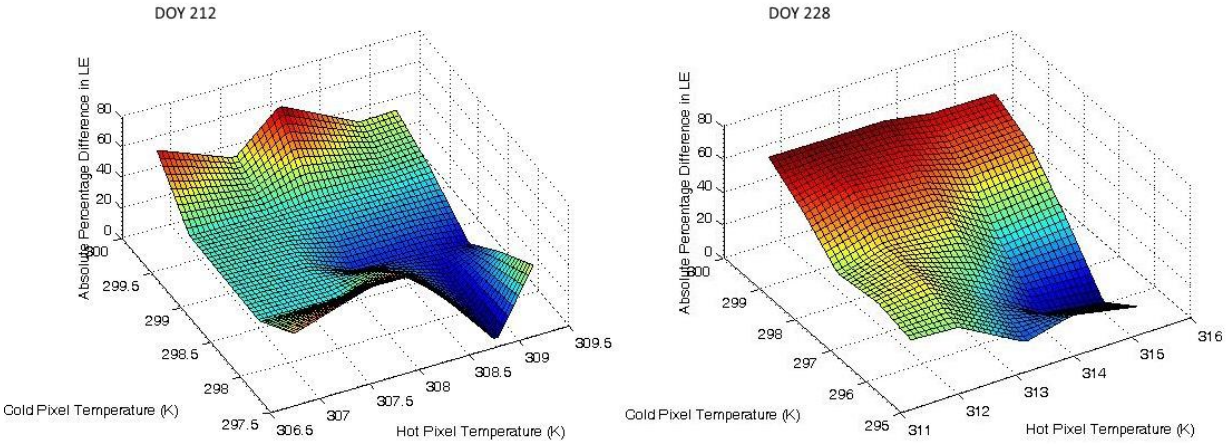

Figure 10 Absolute percentage difference of simulated instantaneous LE estimation from SEBAL compared with ground-based ET measurements using image of DOY 212 and DOY 228.

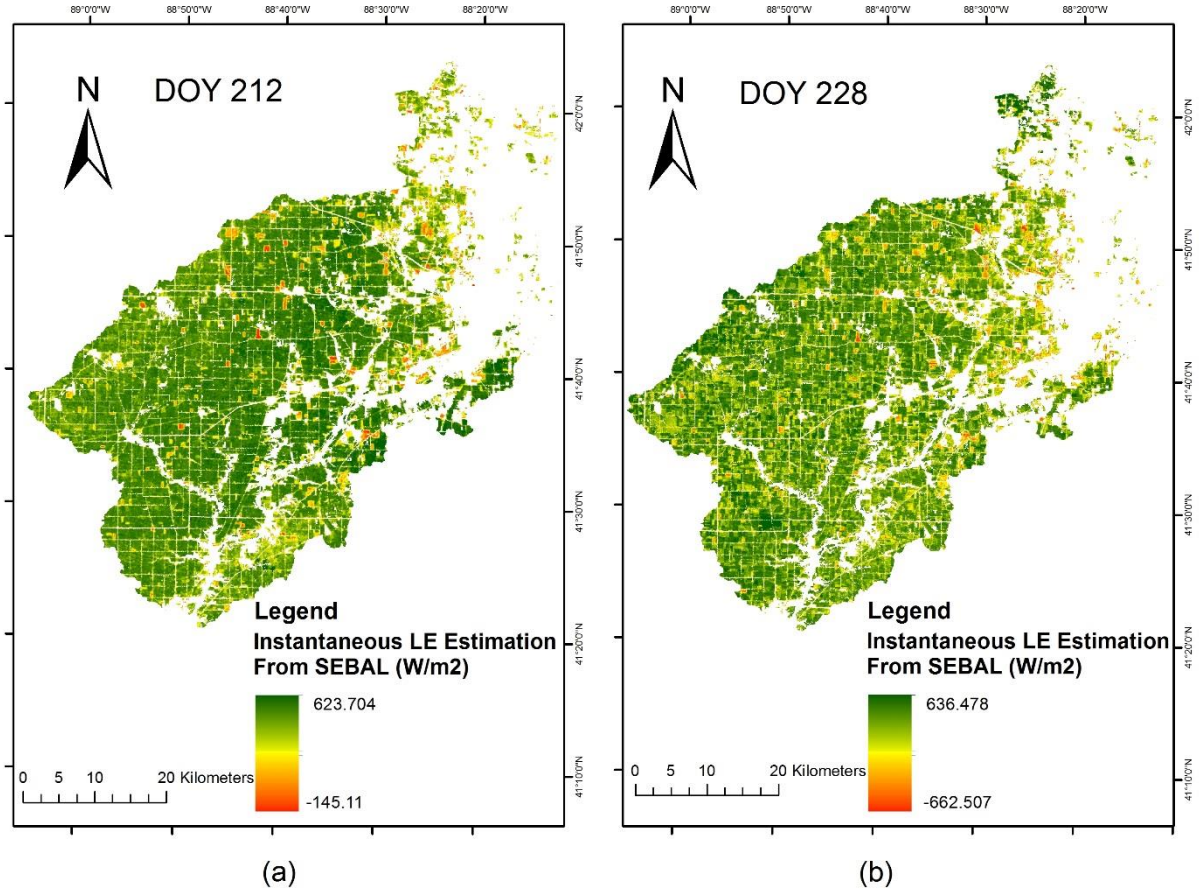

Figure 11 Instantaneous LE estimation from SEBAL in Lower Fox Watershed over agricultural area on image date DOY 212 (a) and DOY 228 (b).

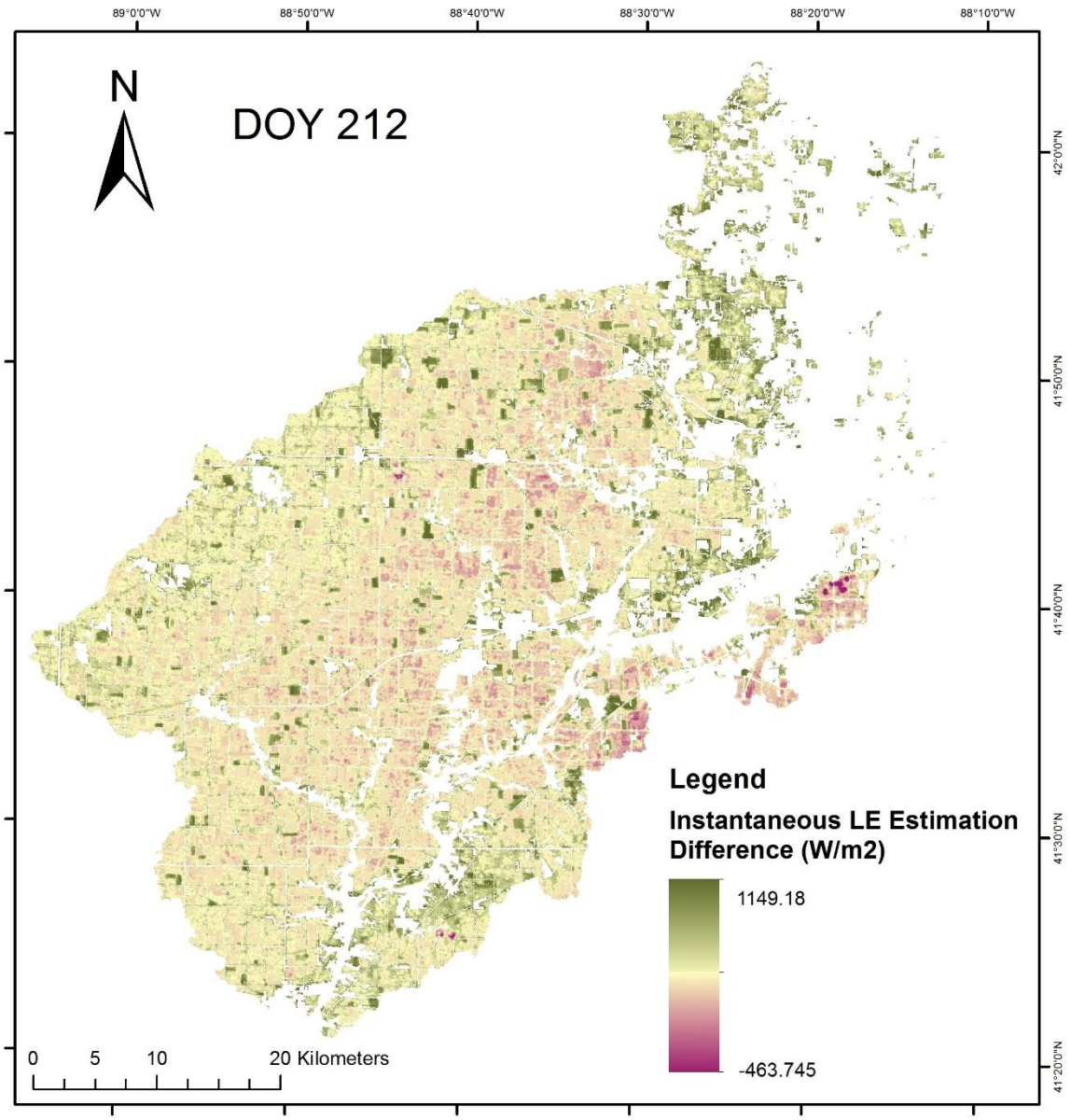

Figure 12 Instantaneous LE estimation Difference between using pixel pair with 5% lower/higher in temperature than hot/cold extremes and using pixel pair with 10% lower/higher in temperature than hot/cold extremes over Cropland for DOY 212

Table 1 Characteristics of Landsat, landcover, and DEM data used

Data Set	Description	Spatial Resolution	Date
Landsat 5 TM	Level-1 data	30m	2006 DOY 212; DOY 228
NLCD	National Land Cover Database	30m	2006
NED	National Elevation Dataset	30m	-

Table 2 Meteorological and fluxes variables for AmeriFlux Site US-IB1

Year	DOY	Time	Air Temperature (℃)	Precipitation (mm)	Wind Speed (m/s)	Rn (W/m^2)	G (W/m^2)	H (W/m^2)	LE (W/m^2)
2006	212	10:30	32.7	0	2.06	574.9	110.9	70.1	412.4
2006	228	10:30	25.8	0	1.35	507.7	80.6	138.7	278.7

Table 3 Comparison of SEBAL-estimated values and filed observations for AmeriFlux Site US-IB1

DOY	Air Temperature (℃)	Rn (W/m^2)	G (W/m^2)	H (W/m^2)	LE (W/m^2)
212 (US-IB1)	32.70	574.9	110.9	70.1	412.4
212 (SEBAL)	31.9	614.36	58.72	121.10	434.53
228 (US-IB1)	25.8	507.7	80.6	138.7	278.7
228 (SEBAL)	25.7	589.18	54.88	275.11	259.19

Table 4 LE Estimation of multiple cold pixel candidates selected by automated selection approach

	Cold Pixel Candidates					
	1	2	3	4	5	6
LST (K)	298.143	298.1215	298.143	298.1324	298.1324	298.1324
LE (W/m^2)	434.5319	434.0551	434.692	434.1808	434.3777	436.646
z_{om} (m)	0.0433	0.0449	0.0433	0.0441	0.0441	0.0441

Table 5 Instantaneous LE Estimation on DOY 212

		LE (W/m^2)				
		Hot Pixel LST (K)				
		306.7666	**307.5469**	**307.9958**	**308.7961**	**309.2139**
Cold Pixel LST (K)	**299.8956**	664.6668	565.022	689.4199	577.0182	666.4531
	299.314	528.3854	523.9663	530.6249	524.9708	530.547
	298.636	384.5874	474.9535	360.0065	463.3785	380.418
	298.143	305.8356	454.3765	263.1552	434.5319	295.3658
	297.5601	229.6376	438.4093	166.8627	409.6073	210.5445

Table 6 Instantaneous LE Estimation on DOY 228

		LE (W/m^2)				
		Hot Pixel LST (K)				
		311.198	**312.0155**	**313.1738**	**313.9802**	**315.1469**
Cold Pixel LST (K)	**299.0756**	491.147	541.8141	488.1522	502.6674	499.9369
	298.0743	448.3109	357.4037	457.1848	433.564	439.5078
	297.1224	395.8373	195.0129	413.4321	359.5299	371.85
	296.0075	322.3077	334.92405	347.5404	259.1899	277.5496
	295.0585	294.4809	309.89315	325.3054	212.803	235.319

Table 7 Net solar radiation (R_n) estimation using different cold pixels

			DOY 212		
LST (K)	299.08	298.07	297.12	296.01	295.06
Rn (W/m^2)	611.81	614.36	616.53	619.52	622.11
			DOY 228		
LST (K)	299.90	299.31	298.64	298.14	297.56
Rn (W/m^2)	586.14	589.18	593.99	598.13	602.55