



Article

Automated Geospatial Models of Varying Complexities for Pine Forest Evapotranspiration Estimation with Advanced Data Mining

Sudhanshu Panda ^{1,*}, Devendra M. Amatya ², Rhett Jackson ³, Ge Sun ⁴ and Asko Noormets ⁵

¹ Institute of Environmental Spatial Analysis, University of North Georgia, 3820 Mundy Mill Road, Oakwood, GA 30566, USA

² Center for Forested Wetlands Research, USDA Forest Service, 3734 Highway 402, Cordesville, SC 29434, USA; damatya@fs.fed.us

³ Warnell School of Forestry, University of Georgia, Athens, GA 30606, USA; rjackson@warnell.uga.edu

⁴ Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, Raleigh, NC 27606, USA; gesun@fs.fed.us

⁵ Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27606, USA; noormets@tamu.edu

* Correspondence: sudhanshu.panda@ung.edu

Received: 25 April 2018; Accepted: 12 September 2018; Published: 19 November 2018



Abstract: The study goal was to develop automated user-friendly remote-sensing based evapotranspiration (ET) estimation tools: (i) artificial neural network (ANN) based models, (ii) ArcGIS-based automated geospatial model, and (iii) executable software to predict pine forest daily ET flux on a pixel- or plot average-scale. Study site has had long-term eddy-flux towers for ET measurements since 2006. Cloud-free Landsat images of 2006–2014 were processed using advanced data mining to obtain Principal Component bands to correlate with ET data. The regression model's r^2 was 0.58. The backpropagation neural network (BPNN) and radial basis function network (RBFN) models provided a testing/validation average absolute error of 0.18 and 0.15 Wm^{-2} and average accuracy of 81% and 85%, respectively. ANN models though robust, require special ANN software and skill to operate; therefore, automated geospatial model (toolbox) was developed on ArcGIS ModelBuilder as user-friendly alternative. ET flux map developed with model tool provided consistent ET patterns for landuses. The software was developed for lay-users for ET estimation.

Keywords: evapotranspiration; modeling; Remote Sensing Landsat; pine forests; artificial neural networks; BPNN; RBFN; Modelbuilder; Visual Basics

1. Introduction

Evapotranspiration (ET) processes at the leaf-to-landscape scales in multiple land uses have important controls and feedback for the local, regional and global climate and water resource systems [1–3]. ET comprises a large part of the hydrologic budget for forests and serves as an indicator of forest growth and ecosystem productivity [2,3]. Leaf area index (LAI), canopy temperature (T_c), canopy (g_c) or leaf stomatal (g_s) conductance, net radiation, wind velocity, and soil moisture are the key drivers of ET [2,4–7]. By altering forest species composition, tree age distributions, and tree densities, land management directly affects ET [8] and consequently watershed hydrology. An interesting observation by Jaramillo et al. (2018)'s research concluded that in climate change scenarios, increased CO_2 concentrations in atmosphere decreases ET due to decreasing stomatal conductance is certain species but the increase in forest biomass is the cause of increased forest ET [9]. Accurate estimation of ET across the landscape is needed for many hydrologic analyses, particularly the assessment of

land cover change's impact on water budgets [2,3]. However, acceptable accuracy of ET estimation over landscape and regional scales at high spatial (plot level) and temporal (daily) scale has yet to be achieved [10,11].

Forest ET processes are inherently complex due to the many ecohydrological interactions within a forest ecosystem that often consist of multiple plant species and multiple vegetation layers with heterogeneous spatial distribution and variable microclimates over space and time [12,13]. Forest ET can be directly estimated by in-situ measurements with the use of either sparsely spaced weighing lysimeters [14]; sap flux sensors [15]; plant chambers [16]; or highly efficient but expensive eddy covariance flux towers [17]. The eddy covariance method that offers simultaneous measurements of ET with high temporal resolutions has gained popularity in recent years due to performance improvements and reduced costs of fast-response monitoring sensors [18–22]. However, the eddy covariance method has limitations in spatial representation of a landscape due to the small footprint of meteorological sensors that cover the tower air-shed. Soil water budget [23], watershed water balance [3,24–27], Bowen ratio or aerodynamics techniques supported by weather station data [20,28], and the soil moisture and potential ET (PET) method [5] are also being used for estimating plant and ecosystem ET. Watershed-constrained water balance calculations of ET followed by change-driver separating techniques are also used by researchers to estimate forest ET [9,29]. ET measurements through all these methods are costly, time consuming, and require elaborate measuring equipment [30] and thus are limited to small spatial scales.

Researchers are also employing hydrologic models with weather, soil, topography, and land cover data, and estimation from satellite remote sensing data [2,10,22] to estimate forest ET. Forest managers, land use planners, and/or hydrologists would potentially benefit from such a simple, efficient, and more cost-effective remote sensing-based technique for forest cover ET estimation [31] over a large area.

Estimating forest ET using remote sensing data is not a new concept. Li and Lyons [32] used 1.1 km resolution NOAA-14 AVHRR remote sensing data to derive surface temperature that combined with limited routine meteorological data like soil moisture to estimate ET of central Australia with limited success. Cristobal et al. (2011) [33] tested the reliability of remote sensing data of TERRA and LANDSAT to estimate forest vegetation ET in the Vallcebre research catchment in Spain from 2003–2005 using 27 AQUA-MODIS images, 11 Landsat-7 and 10 Landsat-5 images in comparison with stand transpiration obtained from sap flow measurements. However, the best estimation of forest ET using the Landsat images were obtained with a 30% uncertainty [33].

Hwang and Choi (2013) [34] studied the seasonal trends of satellite-based evapotranspiration algorithms over a complex ecosystem in East Asia using the Trapezoid Interpolation Model (TIM) [35–37] and the Remote-Sensing Penman-Monteith (RS-PM) [38] method to accurately estimate ET on heterogeneous landscapes including forest cover. The revised RS-PM model employs vapor pressure deficit and LAI to compute canopy conductance and substitutes NDVI with Enhanced Vegetation Index to estimate soil evaporation [38]. The TIM model, which was based on the Priestley-Taylor ET equation, created a relationship between remotely sensed surface temperature and NDVI to estimate ET. Byun et al. (2014) [39] used the dual-model approaches for evapotranspiration analyses over homo- & heterogeneous land surface conditions via the Surface Energy Balance System (SEBS) [40] and the revised Remote Sensing based Penman-Monteith Model (R-RSPM; [38]). Lu et al. (2014) [41] used a hybrid dual-source (H-D) model to estimate ET over different ecosystems using eddy covariance data as reference. The researchers obtained large mean absolute errors (MAE) of 38.6 and 37.6 Wm^{-2} for the woody savannah and forest sites, but low MAE for the grassland site.

Hendrickx et al. (2016) [31] benchmarked optical/thermal satellite imagery for estimating ET for various land uses including forest. The authors concluded that the use of NASA/USGS optical/thermal satellite imagery could considerably improve hydrologic decision support tools compared to their traditional implementations. The benefits of improved decision-making through hydrologic support systems from optical/thermal satellite imagery should substantially exceed the

costs for acquiring such imagery and implementing the remote sensing. Verstraeten et al. (2008) [20] noted that, in general, the finer the spatial resolution of the satellite, the better the agreement between the resulting ET estimates and flux tower measurements. The authors also found that when compared to flux tower measurements, Landsat slightly underestimates ET by 6 Wm^{-2} , ASTER overestimates ET by 18 Wm^{-2} and MODIS underestimates ET by 57 Wm^{-2} . This resulted in errors of 2%, 5% and 15% respectively. Senay et al. (2013) [42] presented a simple, but robust method that uses remotely sensed thermal data and model-assimilated weather fields to produce ET for the contiguous United States (CONUS) at monthly and seasonal time scales. This method, Simplified Surface Energy Balance model, parameterized for operational practices yielded ET results that compared reasonably well with monthly eddy covariance ET data, explaining 64% of the observed variability across diverse ecosystems in the CONUS during 2005. The authors cautioned that more research is required to improve the representation of the predefined boundary conditions in complex terrain at small spatial scales. The approach could use other thermal sensors such as Landsat.

Most existing ET studies use all types of forests or vegetation cover as objects of interest. However, very few studies have used remote sensing to determine ET of individual forest plantations like row crop pine plantation. Remote estimation of pine plantation is essential to observe the change in water availability for its growth when switchgrass or other bioenergy crops are grown as intercrop instead of the natural understory growth. Ha et al. (2015) [43] compared ET estimates of Ponderosa pine forests by the eddy covariance measurements and meteorological and remote sensing-based models and discovered that the Moderate Resolution Imaging Spectroradiometers (MODIS) based remotely sensed method had a limited capability for estimating ET in both undisturbed and disturbed ponderosa pine forests. Therefore, Ha et al. (2015) [43] recommended using finer spatial-resolution-based remote sensing models, such as the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) or Landsat to improve remote-sensing-based estimates of ET.

Ceron et al. (2015) [28] used MODIS digital vegetation data and solar radiation data to derive regional weekly actual ET of pine forests for the South Florida Management District with a simplified surface energy balance method which provided good correlation between the MODIS based ET estimates and the actual in-situ meteorological ET data with average $r = 59$ and $p < 0.0005$. However, refinement in their model was suggested to obtain reliable annual estimation of ET [28]. In addition, suggestions implied that inclusion of more accurate temperature data available from thermal bands of other satellites like ASTER or Landsat would provide better estimates of the forest cover ET, especially the pine forest cover.

Yang et al. (2016) [22] used a multi-sensor ET data fusion system over a mixed forested/agricultural landscape including loblolly pine in North Carolina, USA during 2013 growing season. This study estimated ET from a Two-Source Energy Balance (TSEB) model using thermal infrared remote sensing digital information of land surface temperature from multiple satellite platforms: Geostationary satellite data at 4-km resolution (hourly), 1-km imagery from MODIS (daily), and 30 m Landsat (bi-weekly). By combining these spatial and temporal resolution data to the Spatial-Temporal Adaptive Reflectance Fusion Model (STARFM) to estimate daily ET at 30-m resolution, mean absolute errors of approximately 29% at the daily time step, 12% at the monthly time step, and 3% over the full study period at two flux tower sites in the study area were obtained [22].

Panda et al. (2016) [10] studied the usefulness of remote sensing, especially freely available Landsat ETM+ imageries to estimate homogenous mature pine evapotranspiration (ET) and other ET parameters, such as soil moisture, T_c , LAI, and g_s . They obtained excellent correlation between the actual and predicted upper-canopy temperature ($R^2 = 0.93$, $n = 42$) using the thermal infrared band of Landsat 7 ETM+. However, they obtained low correlation ($R^2 = 0.36$, $n = 6$) for the g_c prediction using band 5 (mid-infrared) and very poor correlation ($R^2 = 0.05$, $n = 42$) for soil moisture prediction using band 7 image info of Landsat 7 ETM+. Panda et al. (2016) [10] used Bands 5, 6, and 7, NDVI, and Vegetation Vigor Index as input parameters to estimate mature pine ET flux using statistical multivariate regression algorithm and obtained satisfactory correlation ($R^2 = 0.61$, $n = 35$) with smaller

spatial and temporal variations in the datasets along with no data mining application in model building. The authors proposed that the inclusion of more spatial variability, sound data mining technique, high-resolution imagery and high-end image-processing approach could provide robust models to help estimate/predict site-specific daily ET flux with better accuracies.

All these models, discussed above, used complicated working procedures and are difficult for novice users to comprehend. For example, the models that were developed for heterogeneous vegetation covers using low spatial resolution imageries were less robust due to lack of adequate data samples covering a large temporal variation in climate. Therefore, based on the recommendation of these studies, there is a need for developing less complex, simpler and user-friendly remote sensing-based ET estimation models with the help of freely available moderate spatial resolution digital information (Landsat, 30 m) that contains bands to estimate canopy temperature, vegetation vigor, biomass, and soil and plant water content.

Most of the vegetation ET studies, including Panda et al. (2016) [10], used statistical modeling approaches to estimate ET values with remote sensing based digital information of the land cover and obtained modest correlation. Use of the artificial neural network (ANN) modeling approach provides a better correlation compared to the conventional statistical modeling in prediction studies related to site-specific crop management, biomass yield prediction, ecological estimation, and hydrologic forecasting with remote sensing data [44–49]. Adeyoye et al. (2012) [50] had used Kohonen Self-Organizing Map (SOM) and feed-forward supervised artificial neural networks to develop models to predict irrigated crop ET with high reliability, i.e., more than 90% ($R^2 > 0.9$). As no known studies have been conducted so far in Forest ET estimation using ANN, this study conducted an analysis to test the efficiency of ANN in modeling homogenous pine forest ET flux.

Thus, the goal of this study was to develop automated remote sensing-based ET flux estimation models of various complexities to estimate/map the daily ET of a homogenous pine forest. The specific objectives of the study were to develop:

1. Artificial neural networks (ANN) models using Landsat based digital information with back-propagation neural network (BPNN) and radial basis function network (RBFN) algorithms to predict homogenous pine forest daily (noon–2 PM average) ET flux on pixel- or plot average- scale;
2. An automated geospatial tool supported by the remotely sensed digital information and ET flux statistical correlation algorithm to help land managers to map pine spatially distributed daily (noon–2 PM average) ET rates in ArcGIS software; and
3. Executable software that estimates daily (noon–2 PM average) ET rates of homogenous pine forest on pixel- or plot average- scale with the PC bands 1 and 2 information.

2. Methodology

2.1. Study Site

The study site is a 4500 ha managed pine forest called the Parker Tract (PT) owned and managed by Weyerhaeuser Company in Washington County, Coastal North Carolina, USA. We selected a mature (~25 years old planted in 1992) loblolly pine (*Pinus taeda*, L.) (LP) stand (35°48' N, 76°40' W) and a young plantation (~5 years old in 2010 clear-cut (CC) (35°48' N, 76°42' W) stands with about 1.5 km apart for this modeling study [51]. The topography of the study area is flat with a ground elevation less than 8 m a.m.s.l. and drained with ditches approximately 90 to 100 m apart to manage the water table and improve tree productivity [19,25]. The area has poorly drained soils with predominant soil types classified as a loamy, mixed, dysic, thermic *Terric Haplosaprists* in the Cape Fear and Belhaven series based on the STATSGO soil analysis. The long-term (1945–2008) average annual precipitation of the site is 1320 ± 211 mm and is evenly distributed with mean annual temperature of 15.5 °C [19]. The annual growing season of pine averages about 195 days [19].

The research site was instrumented for climatologic data acquisition and plant growth monitoring with two AmeriFlux (<http://ameriflux.lbl.gov/>) towers on both the LP and the CC sites (Towers B & A, respectively, Figure 1). The matured loblolly pine (LP) site is a mid-rotation plantation stand with 90 ha area established after clear-cutting a previous rotation of loblolly pine and replanting with 2-year old seedlings at 1.5 m by 4.5 m spacing in 1992 [19]. The clear-cut (CC) site was established in 2003 with 2-year old seedlings in a stand that was clear-cut in 2002 and the flux tower (Tower B) is located in the middle of the site with an approximate distance of 1.5 km from Tower A [22].

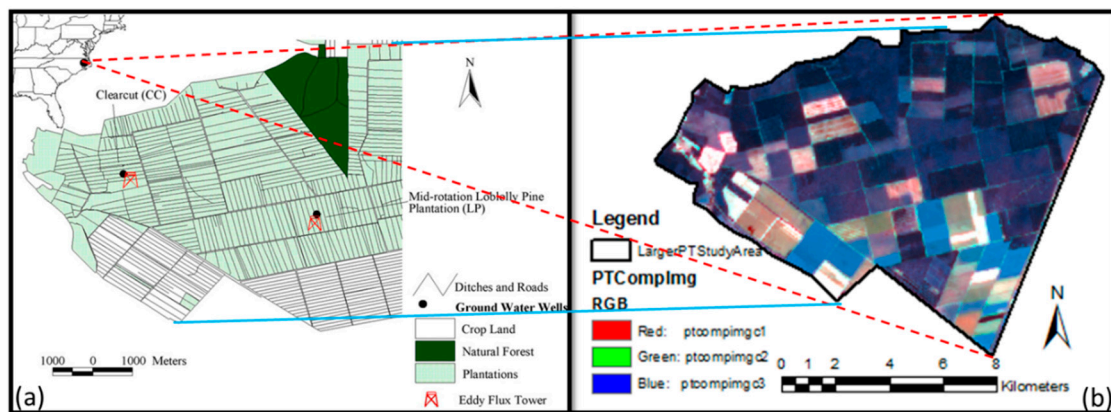


Figure 1. Study area location and instrumentations for climatologic data acquisition and hydrological monitoring—(a) Study area plot design and (b) False Color Composite (FCC) Landsat Image of the study area. Note: PTCCompImg in Figure 1b = PT (larger) site Composite Image.

2.2. Data Acquisition

2.2.1. Micrometeorological Data Acquisition and Preparation

Latent heat fluxes or ET were measured using an open-path eddy covariance system that includes a CSAT3 three-dimensional sonic anemometer (Campbell Scientific Instrument-CSI, Logan, UT, USA), a CR5000 data logger (CSI), an infrared gas analyzer (IRGA, Model LI-7500, LI-COR, Lincoln, NE, USA) and a relative humidity and air temperature sensor (model HMP-45C; Vaisala Oyj, Helsinki, Finland) [19]. Soil heat flux was measured at the matured pine site with three heat flux plates (model HFT3, CSI, Logan, UT, USA) at the depth of 2 cm, which were placed in three contrasting microsites—one in a row of trees, in relative shade, another between rows in a mostly open environment and one about half-way in-between [19]. The flux data were collected at each 30-min time steps and quality checked, as judged by atmospheric stability and flux stationarity [52]; but for few dates, due to malfunctioning of instruments, correct ET flux data were not available. The missing data were gap-filled using the monthly regression between observed and potential ET models created from observed data with high quality [52]. Measured ET (Wm^{-2}) from both eddy flux towers corresponding to the acquired image dates were extracted for the 2006–2014 study period. The ET Flux data were acquired by the tower in the study area on a continuous real-time on every half an hour basis. The data was averaged for 2 h (between noon and 2 PM, covering the image acquisition time of the day) were used as outputs of the estimation model.

2.2.2. Satellite Image Acquisition and Band Selection

Altogether, 70 Landsat images were freely acquired from USGS Earth Explorer (<http://earthexplorer.usgs.gov/>) site, which has completed geometric and radiometric correction on them. They included Path # 14 and Row # 35 cloud-free Landsat 7 ETM+ (2006–2013) and Landsat 8 imageries (2013–2014). However, only 48 sets (for both towers) of suitable images and ET flux data were available for ET model development. For the CC site, only the 2014 data could be used, as 2013 was the year of plantation establishment and pine saplings were almost covered by understories

of red maple, greenbrier and loblolly pine of similar or greater height. All Landsat 7 ETM+ data had scan-line issues but in all ETM+ images used in the study, scan-line issues were avoided when including only 5×5 (=25) pixels around the two eddy flux towers. With a moderate spatial resolution of 30 m, pine vegetation spatial area is covered uniformly. Together Landsat 7 and 8 have temporal resolution of 8 days, so that ET variation in the vegetation can be determined approximately twice a month, capturing seasonal variations. The researchers completed required spectral resolution for ET estimation—having individual bands suitable to estimate the eco-hydrologic parameters related to ET such as LAI, stomatal conductance, soil moisture, and canopy temperature, etc. ([10]; Tables 1 and 2). The images have simple radiometric resolution, i.e., 8-bit for Landsat 7 ETM+ and 16-bit for Landsat 8 (extra processing was needed—raster calculator use to reduce the Landsat 8 pixel values by a factor of 256—to make both imagery data compatible in model development). Landsat satellite wavebands were specifically designed and useful for remote estimation of ecological features, especially ET on earth (Tables 1 and 2).

Table 1. Landsat 7 ETM+ band information and applications (adopted from [10]).

Band # & Name	Band Width	Environmental Monitoring Application Ability
Band 1 (Blue)	0.45–0.52	Bathymetric mapping, distinguishing soil from vegetation and deciduous from coniferous vegetation
Band 2 (Green)	0.52–0.60	Emphasizes peak vegetation, i.e., plant vigor; iron content in rocks and soil differentiation
Band 3 (Red)	0.63–0.69	Vegetation slopes and plant chlorophyll content differentiation
Band 4 (Near Infrared)	0.76–0.90	Biomass content estimation and shorelines delineation/differentiation
Band 5 (Middle Infrared (MIR))	1.55–1.75	Moisture content of soil and vegetation discrimination: cloud and snow differentiation
Band 6 (Thermal Infrared)	10.4–12.5	Thermal mapping and soil moisture estimation
Band 7 (MIR)	2.08–2.35	Soil analysis
Band 8 (Panchromatic)	0.52–0.90	Higher image resolution (15 m)—sharper image definition

Table 2. Landsat 8 band information and its applications (adopted from http://landsat.usgs.gov/best_spectral_bands_to_use.php).

Band # & Name	Band Width	Environmental Monitoring Application Ability
Band 1 (Coastal Aerosol)	0.43–0.45	Coastal and aerosol studies
Band 2 (Blue)	0.45–0.51	Bathymetric mapping, distinguishing soil from vegetation and deciduous from coniferous vegetation
Band 3 (Green)	0.53–0.59	Emphasizes peak vegetation, i.e., plant vigor assessment
Band 4 (Red)	0.64–0.67	Vegetation slopes and plant chlorophyll content differentiation
Band 5 (Near Infrared)	0.85–0.88	Biomass content estimation and shorelines delineation/differentiation
Band 6 (Shortwave Infrared 1)	1.57–1.65	Moisture content of soil and vegetation discrimination: cloud and snow differentiation
Band 7 (Shortwave Infrared 2)	2.11–2.29	Improved moisture content of soil and vegetation discrimination: improved cloud and snow differentiation
Band 8 (Panchromatic)	0.50–0.68	Higher image resolution (15 m)—sharper image definition
Band 9 (Cirrus)	1.36–1.38	Improved detection of cirrus cloud contamination
Band 10 (Thermal Infrared 1)	10.60–11.19	Lower resolution (100 m)—thermal mapping and soil moisture estimation
Band 11 (Thermal Infrared 2)	11.5–12.51	Lower resolution (100 m)—improved thermal mapping and soil moisture estimation

Band 1 to 7 of Landsat 7 ETM+ and Band 2–7 and Band 10 of Landsat 8 were used in the pine ET estimation. All these selected bands are directly or indirectly helpful in the estimation of ET or ET governing parameters (Tables 1 and 2). A 5×5 pixel (150 m \times 150 m) window was chosen surrounding both eddy covariance flux tower locations. Individual bands were separated from each of the images

compatible to the study. Each band (individual dates) was processed together to produce digital information for remote sensing-based ET estimation model development.

2.2.3. Reduction of Spectral Variables through Principal Components Analysis

Principal component analysis (PCA) is a powerful method of analyzing correlated multidimensional data [53,54]. PCA is an orthogonal linear transformation that reorganizes the variance in a multi-band image into a new set of image bands. Wood and McCarthy (1984) [55] effectively explained the PCA algorithm and its advantages in data reduction. Each individual band in the output PCA image receives some contribution from all of the input image's bands. Therefore, PCA was used as a data compression technique for solving the multiple correlation problems. An automated principal component analysis model was developed in ArcGIS 10.3 (ESRI TM) ModelBuilder (Figure 2) using seven bands depending on the satellite (Bands 1–7 for Landsat 7, Bands 2–7 and 10 for Landsat 8). The resulting correlation matrix, covariance matrix, Eigen values and vectors results were analyzed (Table 3, an example of one date Landsat 7 ETM+ mature pine site image PCA) for this study. The first two PCA bands (PC1 and PC2) provided an average of 95% accumulative Eigen values suggesting that 95% of digital information of the seven bands of Landsat images are embedded in the first two PC bands. Therefore, only these two PC bands were used in this study instead of all seven bands as input parameters for ET remote estimation model development. This procedure supported the study objective to develop a model as SIMPLE as possible. A python script was written to use in the automated model in ArcGIS ModelBuilder (Appendix A.1) to batch process all 48 sets (seven selected bands) of Landsat band images to produce corresponding principal component bands from which only the PC1 and PC2 bands were selected to be used in model development. Users would replace the Landsat input bands with their study area that can modify the python script according to their file path and use in the automated geospatial model Appendix A.2) in ArcGIS software to obtain their PC bands easily. The Summary Statistics tool (Appendix A.2) would then provide them with the principal component ASCII information for model development.

Table 3. PCA results to provide decision support for using fewer images in model development.

The Number of Components = 7 Output Raster(s) PERCENT AND ACCUMULATIVE EIGENVALUES			
# PC Layer	Eigen Value	Percent of Eigen Values	Accumulative of Eigen Values
1	395.29	94.73 *	94.73
2	14.64	3.51	98.24
3	3.12	0.75	98.99
4	2.16	0.52	99.51
5	1.00	0.24	99.75
6	0.78	0.19	99.94
7	0.26	0.06	100.00

* The Eigen value percentage shows that only the first principal component band contains majority of seven band information of the studied Landsat 7 ETM+ image and it holds good for other date images including Landsat 8.

2.3. Data Mining Approaches for Model Data Preparation

Data re-preprocessing and data quality control are necessary in continuous real-time data collection process as data gaps were common due to flux tower instrument malfunctioning, weather anomalies, and human error in field measurements. Preprocessing of input data was completed with the following methods. Missing values in a dataset can be filled in by using the “attribute mean” of the dataset [56], most probable value [56], or a global constant [56]. In this study, the mean of the group to fill the missing numbers was used. Data outliers were determined by a separability function that determines how two groups of data clusters would be distinct in a dataset. A higher separability value

suggested a great degree of distinctness in groups. The separability equation used was represented as follows:

$$V = \frac{\mu_i - \mu_j}{\sqrt{\delta_i + \delta_j}} \quad (1)$$

where, μ_i is mean of the group i , μ_j is mean of the group j , δ_i is standard deviation of the data in group i , and δ_j is standard deviation of the data in group j . Group i was the PC bands digital values and group j was the corresponding ET values. Researchers did expect some data volatility due to the drastic weather changes during the year and over nine years in the study area. In case of data volatility (check the standard deviation), intermediate functions (ratio) can reduce volatility [57]. Data integration was also essential; imagery data were collected from two different sources, i.e., Landsat 7 ETM + and Landsat 8 and ET data with instrumentation. Thus, proper integration of these two datasets was needed to work as a single entity which assisted in bringing the PC band digital data into similar range as the Landsat 7 DN (reflectance percentage) values were of 0 to 255 (8-bit data) and the Landsat 8 digital values were 0 to 65535 (16-bit data). Thus, the correlation between attributes A (image data) and B (ET data) was obtained by the following formulae [56]:

$$r_{A,B} = \frac{\sum (A - \bar{A})(B - \bar{B})}{(n - 1)\sigma_A\sigma_B} \quad (2)$$

where, n is the number of data points, A and B are means of two different attribute values, and σ_A and σ_B are standard deviations of the respective attribute values. When the value of $r_{A,B}$ was greater than 0 and A and B are positively correlated, the formula resulted in a high value and thus implied redundancy in the attributes. Data transformation was required for our study due to data volatility as described above. Therefore, in order to enable the dataset to be ANN model friendly, data normalization was completed with our dataset to bring them (PC1, PC2, and ET flux values) into similar ranges. In the NN literature, data normalizing also often refers to rescaling by the minimum and range of the vector to make all elements lie between 0 and 1. Normalization can be done by means of subtracting a measure of location and dividing by a measure of scale; e.g., if the vector contains random values with a Gaussian distribution, subtract the mean and divide by the standard deviation to obtain a 'standard normal' variable with mean 0 and standard deviation 1 (Equation (3))

$$X_N = \frac{X_i - \mu}{\delta}, \quad (3)$$

where, N is number of training cases; X_i is value of the raw input variable, X , for the i th training case; X_N is the normalized value of X ; μ is the mean of data points; and δ is the standard deviation of data points. After the data outlier determination, data volatility reduction with normalization, perfect datasets were created for both statistical and neural network model development.

2.4. Neural Network Model Development

An advanced ANN modeling approach with training, testing and validation model development was also undertaken in this study for accurate estimation of ET flux from Landsat image band based PC1 and 2 bands. We used two different ANN modeling algorithms to develop a predictive ET model using the digital image information. Both Back-propagation Neural Network (BPNN) and Radial Basis Function Network (RBFN) modeling approaches were used in our study.

In a typical BPNN, with every iteration, error generated from the input-output correlation is back-propagated to the input layer from the output layer to make the network learn better so that the global error could be minimized [58]. The increment and decrement of randomly generated weight factor and assigned to individual input parameters, in each iteration can help in reducing the error [58]. We modified the change in weight in the BPNN model with the application of a momentum term to the model. Since the change in weight in a layer is always added to the previous weight to update it

in the gradient descend method, the error in prediction was expected to decrease. The BPNN model architecture pertaining to our research is shown later in the Results section. According to [59]), a typical Radial Basis Function Network (RBFN) consists of three different layers with the successive layer fully connected by feed-forward arcs as shown in the RBFN model architecture pertaining to our research that is shown later in the Results section. There is no provision of weight between the input layer and the hidden layer (prototype). The transfer function used at the hidden layer is the 'radial basis functions', a nonlinear transfer function. This study presents only two input parameters (Landsat PC1 and 2 band digital information) in the input layer with only one hidden layer present in the RBFN model. Generally, in the RBFN model, the output layer is linear [59] but in this study, the RBFN model was nonlinear due to the application of Gaussian transfer function in the network.

A step-by-step model optimization procedure developed by Panda et al. (2010) [44] to obtain the best correlation between input and output parameters in BPNN and RBFN. The same process was employed in this study by changing learning rate, momentum term, and iteration rate to optimize the ANN models for obtaining optimal prediction accuracies. Randomly, with a 70-30% ratio the data were divided as training and testing dataset. The models were optimally trained first and the testing data were used in both models to validate the optimal training model's estimation efficacy as discussed below.

Neural Network Model Performance Evaluation

Both types of the ANN model performances were evaluated based on root mean square error (RMSE), prediction accuracy, and standard error of prediction (SEP). Moreover, the correlation coefficient (r) between the actual and predicted output along with the slope and intercept of the linear regression model was used. The equation for RMSE is given by

$$RMSE = \sqrt{MSE} = \sqrt{\frac{SSE}{n-p}} \quad (4)$$

where, n is the number of observations; p is the number of the parameter to be estimated; and SSE and MSE are sum of squared errors and mean square error, respectively.

Average test prediction accuracy is calculated based on Equation (5), where N is total number of observations and OP_A and OP_P are actual and predicted output, respectively.

$$Average\ Test\ Accuracy = \left(1 - \frac{1}{N} \sum_{i=1}^N \frac{|OP_A - OP_P|}{OP_A} \right) \times 100 \quad (5)$$

An executable file developed in the Biolab of North Dakota State University, Fargo ND, using Visual C++ (Microsoft Corporation, Bellevue, WA) was used to determine the predicted ET accuracy and the subsequent actual and predicted correlation coefficient (r) between the measured and predicted ET and, intercept (a), slope (β), and SEP from the back-propagation neural network result. The predicted and actual output regression analysis was done using the following linear equation:

$$Y = \beta X + a \quad (6)$$

where, X and Y are predicted and actual output, respectively; β is slope; and a is the intercept. The SEP of the predictive model is calculated by using the following equation (Kramer, 1998):

$$\sqrt{\frac{\sum_{i=1}^n [(Y_i - X_i) - d_m]^2}{n - 1}} \quad (7)$$

where, d_m is the mean of the difference between actual and predicted values Y and X (of i^{th} individual), respectively; and n is the total number of observations.

2.5. Automated Geospatial Model and Software Development for ET Estimation

2.5.1. Correlation Algorithm Development and Software Development for ET Prediction

Artificial Neural Network (ANN) models have some limitations in their usage by users. The ANN models would provide ET estimation only with the use of ‘trained’ models as developed in this study and discussed later. However, such users with no access to NN software may not be able to predict the pine ET for their area of interest using these models. Therefore, there is a big need of developing statistical correlation algorithms using the processed PC1 and PC2 band digital information and the corresponding ET Flux data. A multiple Stepwise regression was completed in SPSS (IBM, Armonk, NY, USA). The p-value statistics were used to test the null hypothesis that the coefficient is equal to zero (no effect). With low p-value (<0.05), it was understood that a predictor was meaningful to our model because changes in the predictor’s value are related to changes in the response variable. The model regression coefficients represent the mean change in the response variable for one unit of change in the predictor variable while holding other predictors as constant and it is important in analyses as it isolates the role of one variable from all of the others in the model. Regression coefficients obtained from the statistical model were used by the researchers to develop the ET prediction algorithm and used the algorithm to develop an ET estimation software in Visual Basics Studios platform. The model algorithm was also used in the automated geospatial model based spatial ET map development.

The relationship between input and output data, i.e., the statistical multiple regression correlation algorithm shown (Equation (8)) in the Results and Discussion section was used to develop an executable file (software) that can be used by laypersons to estimate the pine ET from image digital information. The form is very simple as it uses only two inputs (Landsat imagery based PC1 and PC2 band as inputs and just by clicking on the Calculate button, the end-users can get the ET Flux (W/m^2) in the appropriate text box.

2.5.2. Automated Geospatial Model Development for ET Estimation Map Production

Figure 2 is a continuous workflow of developing the ET map of a single date for the study area using the principal component band images as discussed previously. The workflow is presented in the automated geospatial model developed in ArcGIS 10.3 ModelBuilder platform.

5 September 2013 Landsat 8 image was used to develop the ET flux spatial variation map and for the comparison/validation study. As discussed in the previous section and shown in Appendix A.2, seven bands (bands 2–7 and 10) of Landsat 8 were masked to the study area (Figure 1) and principal component analyses were performed to produce seven and two principal component bands, respectively (Figure 2). The cumulative Eigen values for first two PC bands were 98.9% in the seven bands PC analysis. Thus, the two PC bands, whose cumulative Eigen values were 100%, obtained from two bands PC analysis were used in further analysis. Data transformation of 16-bit PC1 and PC2 band rasters was conducted using Raster calculator (Figure 2) with a reduction factor of 256. The transformed PC band rasters were converted as a float (*.flt) file to enable map algebra application in ArcGIS. Finally, through the Raster Calculator tool, the statistical correlation algorithm (Equation (8)) was applied using both transformed PC bands to develop the ET flux raster (map) of the delineated Parker Tract study area (Figure 2). The equations used for raster calculation are provided as labels below the tools in Figure 2.

End-users should be able to develop a single date ET Flux map of their interest study area by just replacing the Landsat image band rasters and study area boundary vector of their choice (left side blue ellipses) in the automated geospatial model (Figure 2). This will be available to users in an ArcGIS Toolbox (*.tbx) form for running the model to produce their intended final output. Output data file path should be changed to suit to their requirement by double clicking on each output data (green ellipses) and modifying the path. It is to be noted that the automated model (Figure 2) developed in ArcGIS has intuitive color scheme, i.e., blue ellipses are INPUT data (vector/raster), yellow rectangle (ArcGIS tools, user developed separate models, or user developed script tools developed in ArcPython as developed

and used as shown in Figure 2, and the green ellipses are output data (vector/raster). From here onwards, the developed automated geospatial model will be referred to as ‘Pine ET-flux Remote Estimation (PETFRE) Model.’ The PETFRE model use is perhaps the simplest way of developing the daily ET Flux maps using Landsat image bands.

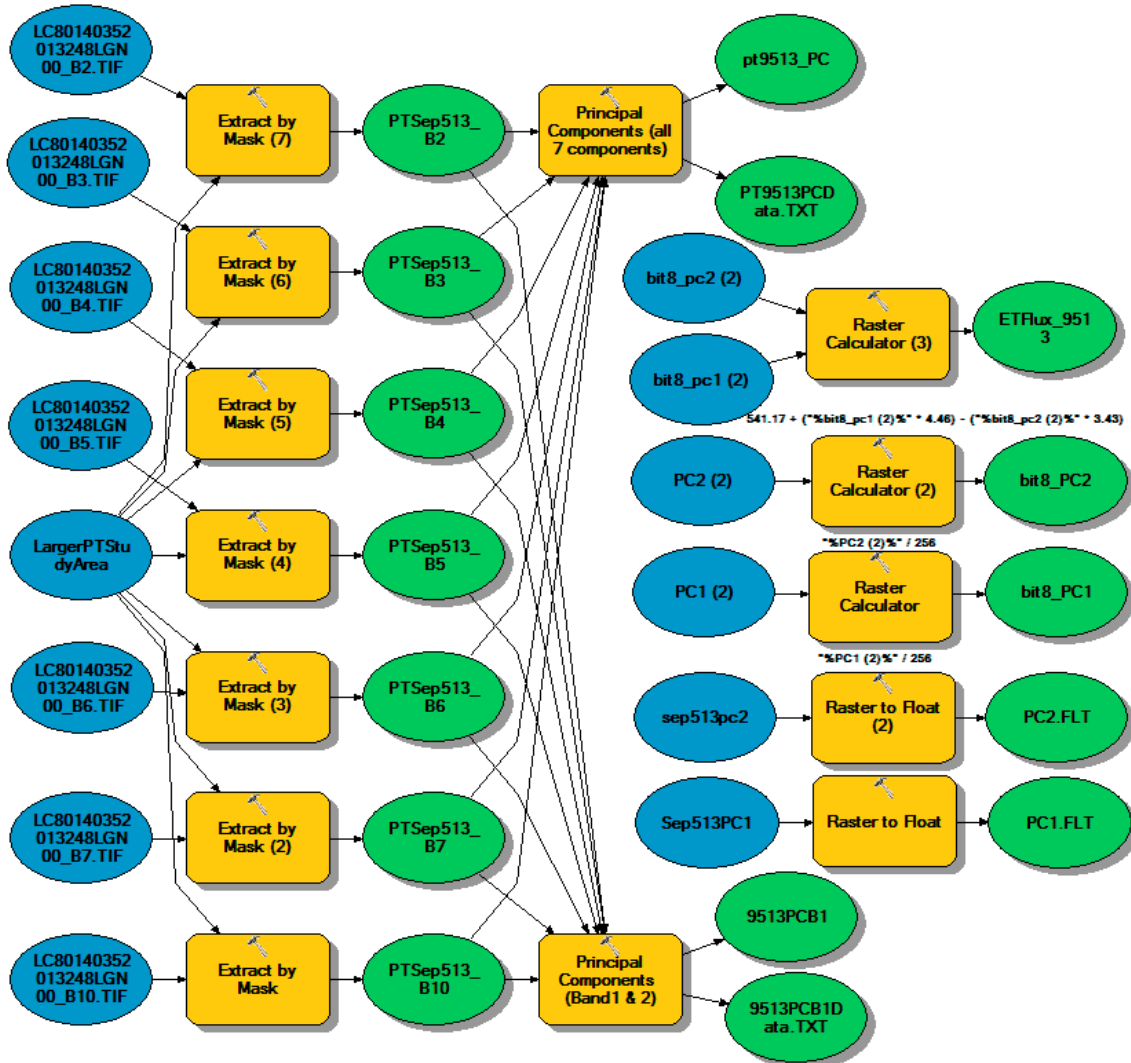


Figure 2. Automated geospatial model developed in ArcGIS ModelBuilder to be used by end-users to develop a single date ET Flux map of a study area.

2.5.3. Study Result Validation

Initially, a visual comparison was conducted between our resultant ET Flux (Wm^{-2}) map developed for 5 September 2013 (scan-line free Landsat 8 image) with the ET (mm/day) map of 20 July 2013 developed by Yang et al. (2016)’s team in the same study area of Parker Tract. The authors (Yang et al., 2016) [22] used cumbersome process-based high-end models like Atmospheric-Land Exchange Inverse (ALEXI) and disaggregation ALEXI (DisALEXI) that combine two-source energy balance (TSEB) and atmospheric boundary layer (ABL) models. The 2011 NLCD landuse classification map were two years apart from our ET map development date; as a result, researchers developed another automated geospatial model (Figure 3) in ArcGIS ModelBuilder platform using the unsupervised ISODATA clustering algorithm, WARD minimum variance statistical function based DENDROGRAM analysis, and reclassification of landuse types with the dendrogram result (Table 4) support to produce the landuse classification map of the same 5 September 2013 Landsat image

(Figure 3). Initially, the ISODATA based 13 landuse classes merged together with dendrogram suggestion into 9 classes and later a five landuse class reclassified map was created with classes like (1) pine forest + mixed forest, (2) bare soil, (3) bare soil intermixed with sparse vegetation cover, (4) bare soil and vegetation intermix, and (5) agriculture.

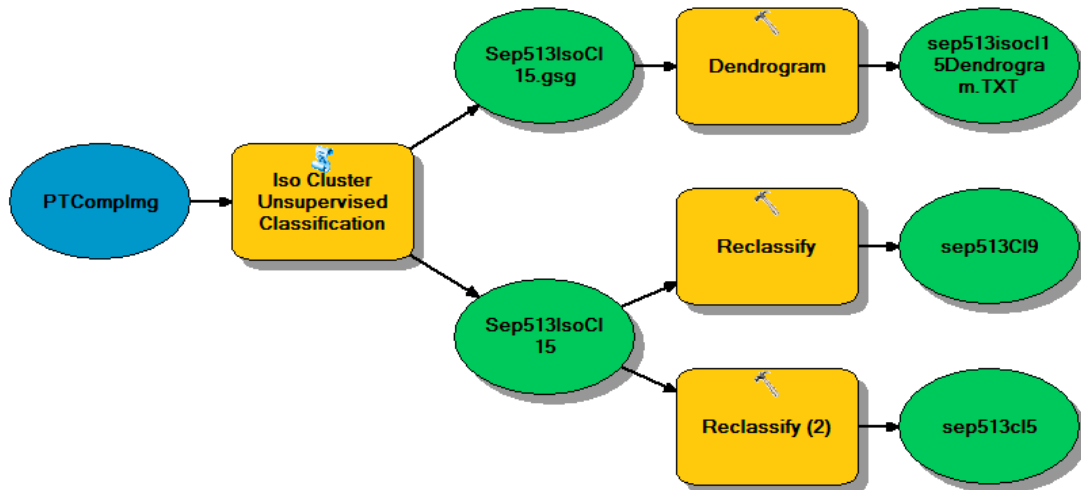


Figure 3. Automated geospatial model developed in ArcGIS ModelBuilder to obtain the landuse classification map (raster) of the study area.

Table 4. WARD minimum variance result obtained with DENDROGRAM analysis of the unsupervised classification of the 5 September 2013 Landsat False Color Composite image.

Distances between Pairs of Combined Classes (In the Sequence of Merging)			Dendrogram of e:envmod~ 1nn_pca~ 1imagec~ 1sep513~ 1.gsg
Remaining Class	Merged Class	Between-Class Distance	
4	5	2.9	
6	7	3.0	
8	10	4.0	
1	3	4.2	
9	11	4.2	
4	6	5.4	
8	9	6.2	
8	12	6.2	
1	2	7.1	
4	8	11.6	
1	4	13.8	
1	3	16.7	

The ET flux (Wm^{-2}) map of 5 September 2013 (Figure 2 output) developed with PETFRE was validated with landuse classification map (Figure 3 output) with five classes. Stratified random sample locations (20 numbers) were selected to obtain the sample landuse types and their corresponding ET flux values. Figure 4, the automated geospatial model, provides the workflow of the ET flux estimation validation process with landuse types of the study area. The ET flux would compare with the landuse types qualitatively, i.e., wetted bare soil characteristic to low-gradient coastal landuse would have higher ET flux than other vegetative land cover. According to Yang et al. (2016) [22] agriculture would have higher ET Flux compared to forest cover depending upon the study area.

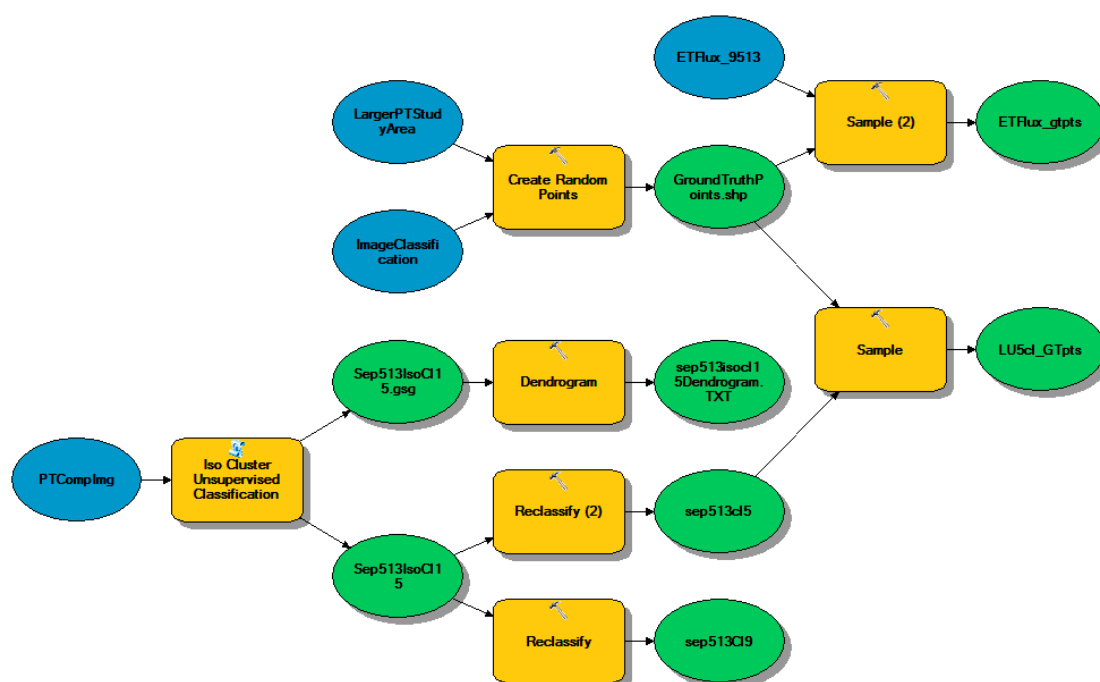


Figure 4. Automated geospatial model developed in ArcGIS ModelBuilder to conduct ET flux estimation validation with study area Landuse with ground-truth samples.

3. Results and Discussion

3.1. ET Prediction Model Results

3.1.1. BPNN Training and Validation Models

Out of the 35 outlier free data for 35 different days of the 2006–2014 period obtained after the data mining on the available dataset, 26 data values for 26 days were randomly chosen as ‘Training’ data and the rest of the nine data values for 9 days were used as the ‘Testing/Validation’ data for the BPNN model development. Both the training and validation data sets contained two input parameters (PC1 and PC2 band digital values) and one corresponding output parameter (ET flux values) for each day. The actual BPNN model architecture is provided in Figure 5. The model structure is presented in the bottom left corner of the figure showing two input parameters, one hidden layer, and one output parameter along with a bias factor associated with the hidden layer, or the processing unit of the model. The top of the figure shows the model graphics that was obtained after the model was run for training. The model parameters were set to the optimum following the neural network optimization step-by-step approach. The learning coefficient of 0.5, momentum term of 0.9, learning rule as Delta-Rule algorithm, transfer function as sigmoid, and a training epoch of 50,000 were set for the optimal training model (Left-Middle of Figure 5). The training model RMSE error of 0.13, data correlation of 0.75, network weight variation dynamics, actual versus desired classification rate, and the confusion matrix were obtained after the model was trained in the Neural Ware Professional II Plus software. Once, the model training was completed, the BPNN trained model was used to validate the model with the randomly prepared testing dataset. The model tested output data was written in *.nnr format. The optimized model obtained average absolute error of 0.18 (18% of the average actual normalized absolute ET flux values used as testing data) and an average accuracy of 81.3% (Appendix C).

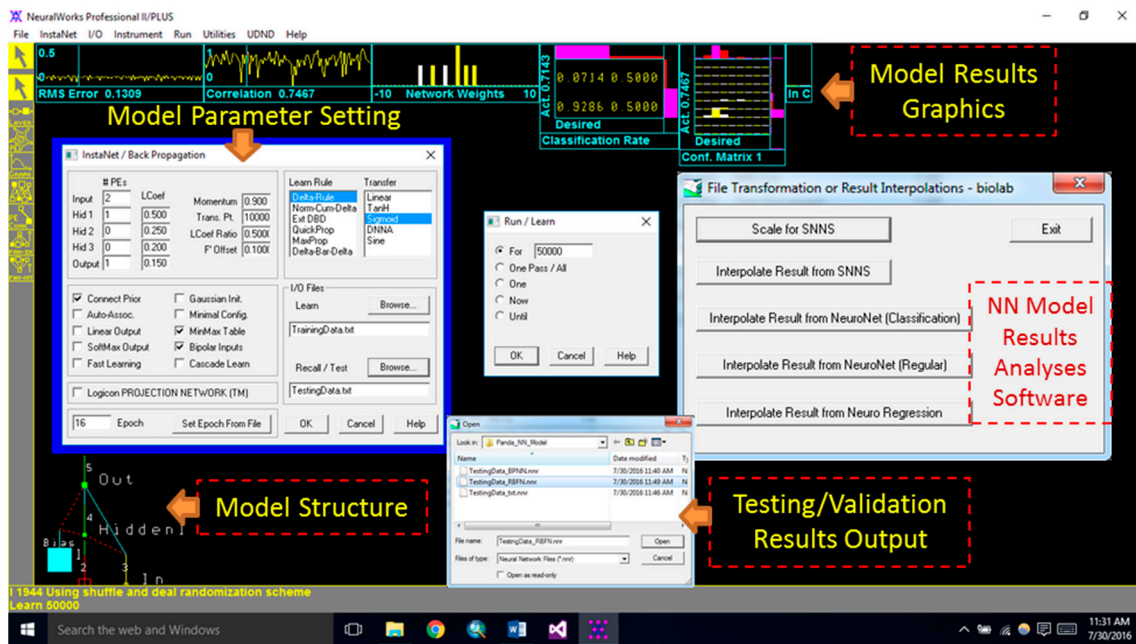


Figure 5. Visual representation of remotely sensed digital information based ET-estimation BPNN model architecture design, training model output graphics, and validation result interpolation (software) process.

3.1.2. RBFN Training and Validation Models

The same ‘Training’ and ‘Testing/Validation’ datasets used for the BPNN model development were also used in the RBFN modeling. The actual RBFN model architecture is provided in Figure 6. The model structure is presented in the bottom left corner of the figure showing two input parameters, one hidden layer, and one output parameter along with a bias factor associated with the hidden layer, or the processing unit of the model. The top of the figure shows the model graphics that was obtained after the model was run for training. The model parameters were set to the optimum following the neural network optimization step-by-step approach. The same (like in the BPNN model) learning coefficient of 0.5, momentum term of 0.9, and 45,000 epochs were found to be optimum for the RBFN model architecture, along with the learning rule of Delta-Rule algorithm and transfer function of sigmoid (Left-Middle of Figure 6). The training model RMSE error of very minimal 0.0034, data correlation of almost 100%, i.e., 0.99, and actual versus desired classification rate of almost 1.00 were obtained. It is shown at the top of Figure 6 visual along with the network weight variation dynamics and the confusion matrix. The RBFN trained model was used to validate the model using the prepared testing dataset. Our optimized model obtained average absolute prediction error of 0.15 (15% of the average actual normalized absolute ET flux values used as testing data) and an average accuracy of 84.8% (Appendix D).

In comparison of both ANN modeling approaches, the RBFN model provided somewhat better ET estimation compared to the BPNN model. More data inclusion may enhance the model ability to predict ET of homogenous pine using the Landsat satellite imagery information. This procedure was developed as an alternative to the less complex, simple, and user-friendly automated geospatial model and the executable file developed for ET flux estimation with remote sensing digital information shown below. However, the ANN model usage is not simple for lay-users but forest managers with access to neural network software possibly would be able to apply them with the availability of ANN software.

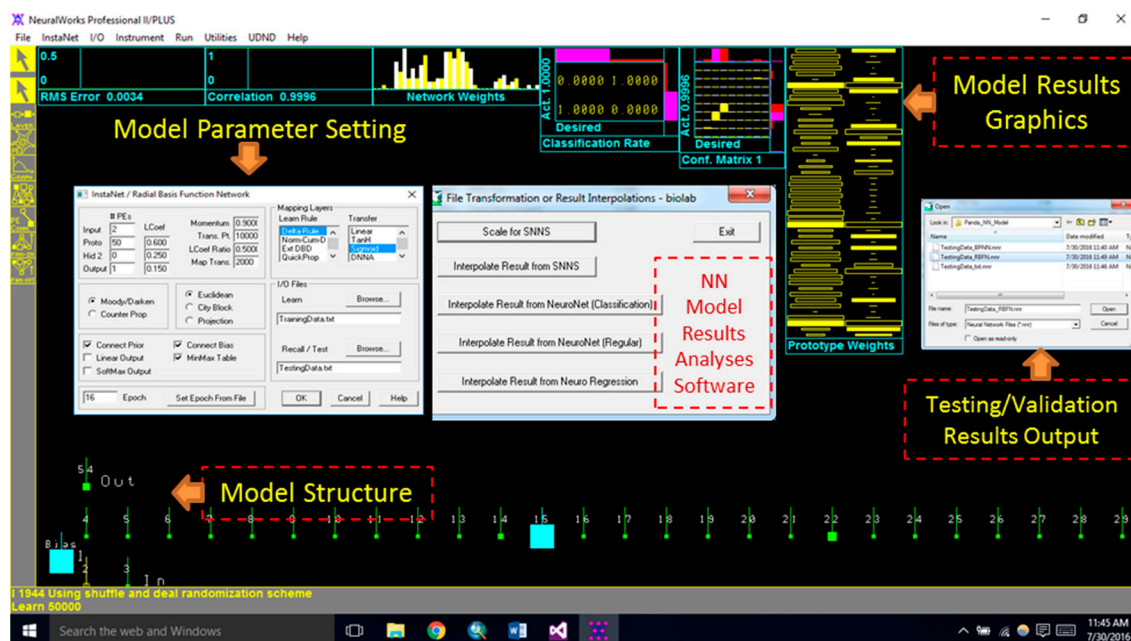


Figure 6. Visual representation of remotely sensed digital information-based ET-estimation RBFN model architecture design, training model output graphics, and validation result interpolation (software) process.

3.1.3. Software Development for ET Prediction

Artificial Neural Network models have their limitations in usage by lay-users. Therefore, researchers developed PETFRE, an automated geospatial model (Figure 4), and a software (executable file) in Visual Basics Studio (Appendix B) using the algorithm developed with the statistical multiple regression correlation study. The regression model developed using the ‘preprocessed’ PC bands digital data as input parameters and the eddy flux tower based ET flux data as output parameter for 35 different days of the study period, provided good coefficient of determination ($R^2 = 0.58$, standard error of prediction (SEP) = 18.55%, p -values of PC1 and PC2 as 0.047 and 0.000000327, respectively). The algorithm (equation) developed from the regression analysis is shown in Equation (8).

$$\text{Pine ET Flux (Wm}^{-2}\text{)} = 541.17 + (\text{PC1} \times 4.46) - (\text{PC2} \times 3.43) \quad (8)$$

Equation (8) was used to develop the executable file (software) to estimate daily ET flux on pixel- or plot average-scale basis. When users input the Landsat image supported principal component bands (PC1 and PC2) based ASCII data for a given day into the appropriate text boxes and click on the Calculate button, the corresponding ET value in Wm^{-2} would be shown on ‘Evapotranspiration (ET) Flux’ text box. The software is provided in Appendix B along with the codes developed in Visual Basic Studio 2015. Figure 7 shows the results of the entire automated model process as a work flow that creates a daily ET flux map of the study area using 5 September 2013 Landsat 8 image bands. Figure 7a is the automated model run with the initial seven band rasters of the Landsat 8 (Figure 7b) and the FCC image is shown at the end of Figure 7b, which was used in image landuse classification (Figure 3). Figure 7c is the PC1 and PC2 band rasters with 16-bit pixel values which were converted to 8-bit rasters (Figure 7d) with raster transformation algorithm. Equation (8) was used to produce the study area ET flux map (Figure 7e for 5 September 2013, which varied from 373.9 Wm^{-2} to 794.9 Wm^{-2} , with an average of 575.45 Wm^{-2} . Predicted ET values for different landuse types are discussed later as a validation of the ET flux estimation for those landuse types.

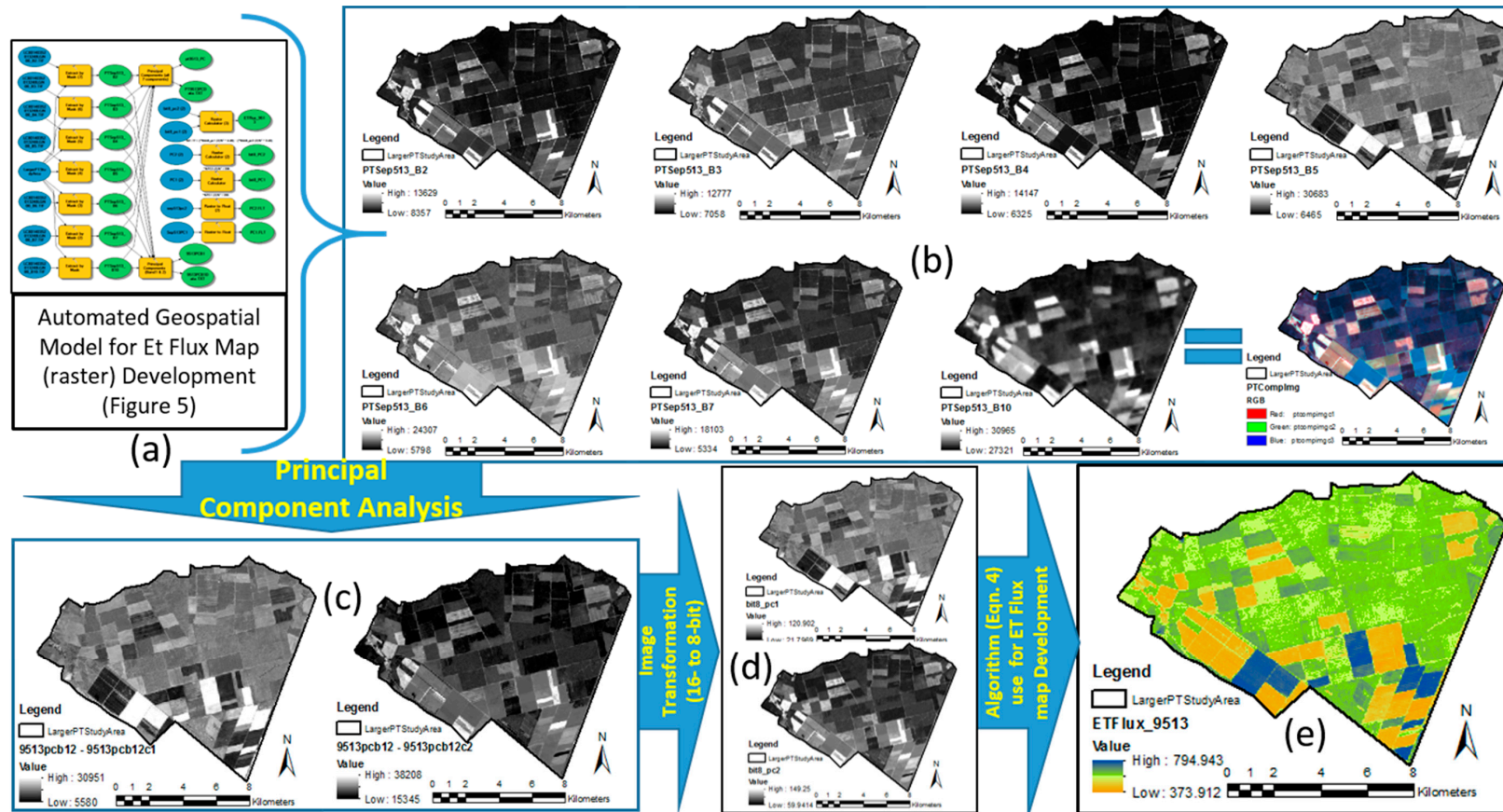


Figure 7. Resultant maps of the entire step-by-step process of the study to develop the automated geospatial model in ArcGIS ModelBuilder to develop the ET flux map of the study area using Landsat image rasters: (a) Automated geospatial model for ET Flux raster development using remote sensing data; (b) individual band raster of a selected composite Landsat 8 image; (c) 16-bit principal component 1 and 2 rasters obtained from the PCA analysis of the composite image; (d) 8-bit converted PC1 and 2 raster from Landsat 8 imagery 16-bit raster to compliment with 8-bit Landsat 7 ETM+ PC rasters; (e) Equation (8) based ET Flux raster of the study area.

3.1.4. ET Flux Map Comparison with Other Studies

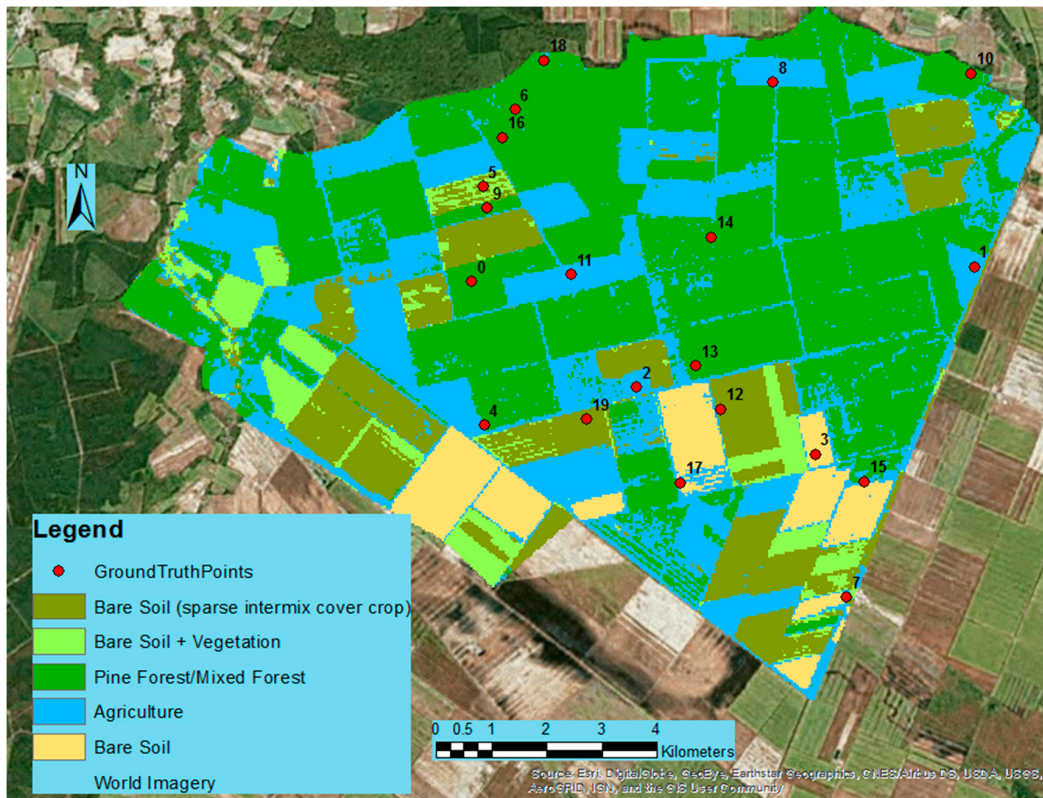
5 September 2013 false color composite (FCC) Landsat image was classified into five dominant landuse classes in the study area. Through randomly sampled 20 ground locations, we analyzed the landuse (Figure 8a) versus ET flux (Figure 8b) correlation trend to ascertain that the developed process of creating an ET flux map with Landsat images works properly (Figure 7e). Table 5 provides the correlational information about the landuse versus ET flux in the study area. Though, our study algorithm for ET flux estimation was developed for pine only still were able to estimate ET fluxes of other vegetation in the larger study area (Figure 7e). We found consistency between the landuse and ET flux values in the study area. The table (Table 5) is arranged with ET flux values from largest to smallest. From the analysis, it is clear that ‘bare soil’ land cover, which is actually the non-forested wetlands (As corroborated in Yang et al., 2016 [22] study), has the highest ET fluxes. Next higher ET flux is from the ‘agriculture’ landuse. Forest (pine forest/mixed forest) has moderate ET flux whereas obviously the upland ‘bare soil intermixed with sparse vegetation’ landuse has lowest ET flux values. In general, the ET patterns for different landuse types found in this study are similar to other studies reported in Yang et al. (2016) [22], Liu et al. (2010) [60]—Table 5 & Figure 4, and Wang et al. (2012) [61]—Figure 5.

Table 5. Landuse versus ET flux values comparison with classified landuse map and the developed ET map of the study area for 5 September 2013.

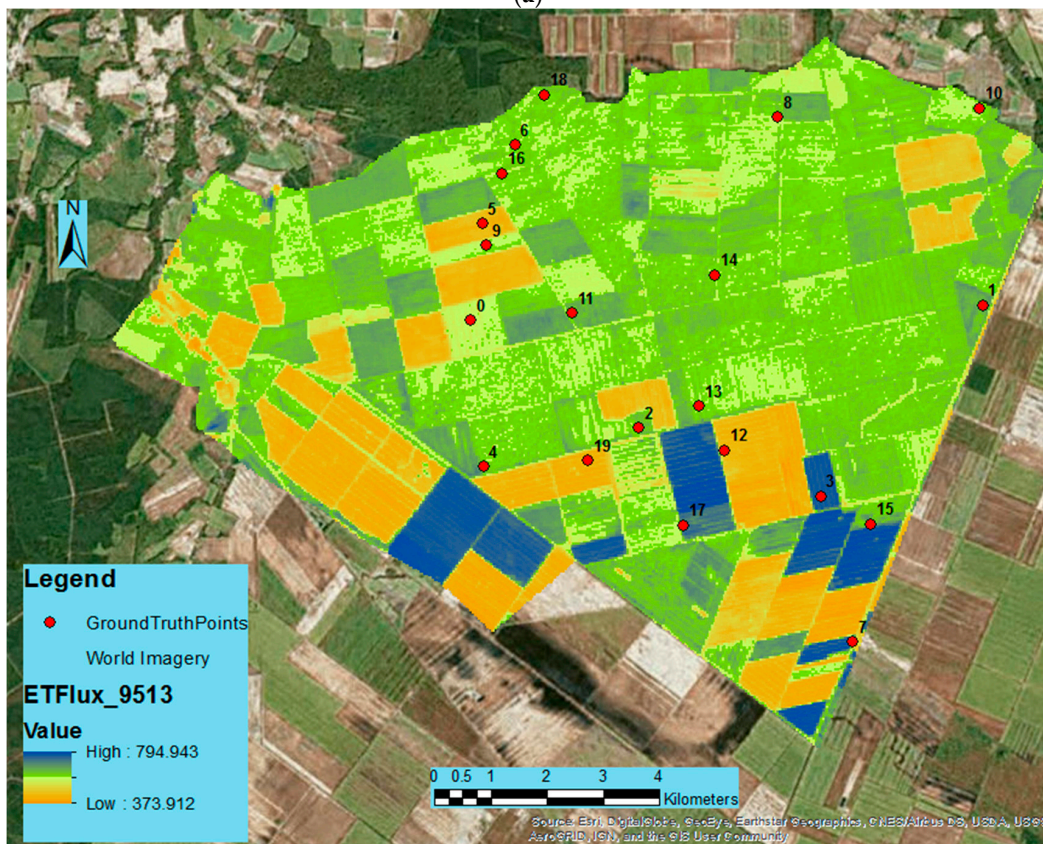
# ID	Ground Truth Points #	Latitude	Longitude	Landuse Code	Class Names	ET Flux Values (Wm^{-2})
17	3	895073.224	3969010.052	5	Bare Soil (non-forested wetland)	753.6
19	17	893109.853	3968512.898	5	Bare Soil (non-forested wetland)	716.8
10	11	891361.022	3971487.833	4	Agriculture	633.7
20	7	895632.953	3966936.526	4	Agriculture	632.9
13	2	892399.327	3969877.053	4	Agriculture	621.5
3	8	894212.281	3974438.179	4	Agriculture	620.9
9	1	897289.221	3971866.671	4	Agriculture	609.5
18	15	895808.303	3968642.686	3	Pine Forest/Mixed Forest	608.1
8	14	893402.825	3972129.405	3	Pine Forest/Mixed Forest	603.7
16	4	890178.624	3969222.054	3	Pine Forest/Mixed Forest	598.9
7	9	890084.101	3972406.477	3	Pine Forest/Mixed Forest	591.2
2	18	890819.792	3974615.016	3	Pine Forest/Mixed Forest	590.6
12	13	893250.508	3970230.524	3	Pine Forest/Mixed Forest	585.9
4	6	890432.435	3973870.562	3	Pine Forest/Mixed Forest	581.7
1	10	897114.404	3974702.01	3	Pine Forest/Mixed Forest	566.7
11	0	889903.614	3971319.138	3	Pine Forest/Mixed Forest	566.5
5	16	890263.664	3973450.483	3	Pine Forest/Mixed Forest	565.4
14	12	893648.877	3969616.451	1	Bare Soil sparse intermix cover crop)	433.9
6	5	890008.174	3972719.761	2	Bare Soil + Vegetation	431.8
15	19	891683.419	3969375.281	1	Bare Soil sparse intermix cover crop)	427.3

3.1.5. Uncertainty and Limitations

As explained in the study, the ET flux data used as output for the model development is an average of 2 h (noon—2 PM) of data collected every 30 min through the Eddy Flux towers. This average data was used to correlate with the image acquisition time. Therefore, tools (ANN training models, automated geospatial models, and executable files) may not be able to predict the daily average ET and obviously true that no image (satellite or aerial) is a representative of the daily (entire day) environmental changes as the images are taken during a particular time of a day. The tools developed in this study may not appropriately estimate that 2 h average ET flux for other landuse types even when tested as the model used for the homogenous pine features in model development. It is likely that some errors might have been introduced in using only two significant image bands (PC1 and PC2) in the analysis and use of image processing software and algorithms used for ANN models. Furthermore, some prediction errors may be attributed to potential errors in ET data measured from eddy covariance-based flux towers [1].



(a)



(b)

Figure 8. (a) Landuse versus (b) ET flux analysis with stratified random ground-truth points of 5 September 2013 in the study area.

4. Summary and Conclusions

In this study, researchers developed three different types of satellite image-based ET models varying from a complex to a more simple and user-friendly level as a software package for users to apply based on their knowledge and the availability of data for pine forests. The software includes (1) artificial neural network based trained models that would predict ET flux values when user provides the PC1 and PC2 ASCII values of their study area as input neurons; (2) an automated geospatial model (PETFRE) to produce an ET flux map of a study area by using Landsat image bands as input rasters; and (3) an executable file to produce the ET flux values of a study pixel or plot- average with the use of PC bands 1 and 2 ASCII values. As forest ET is governed by soil moisture, canopy temperature, leaf area index, stomatal conductance, and vegetation vigor besides other environmental factors, the relevant spectral bands from Landsat were processed using principal component analysis algorithm. Principal component bands PC1 and PC2 (that contain more than 95% inherent digital information of all bands) were used as predictors (input parameters) and actual field ET flux values as the output parameter. A multivariate regression modeling approach was used to develop the pine ET flux estimation algorithm. The algorithm was the basis of the ET flux estimation software development in Visual Basics Studio and ET flux map development with Raster Calculator in ArcGIS. The ET flux map development process was validated through landuse –ET flux correlation analysis. The automated geospatial model-based ET flux map developed for 5 September 2013 for the Parker Tract study area with various other landuse types provided a normal trend in ET Flux values with the type of landuses, which were consistent with other studies. Thus, our less complex, simple, and user-friendly geospatial models and the software should help end-users to develop their ET flux map using minimal data and skill. However, more reliable prediction results could potentially be obtained using the neural network software like BPNN and RBFN model building ability.

With the randomly selected training data for RBFN model, a perfect (0.99) data correlation were obtained along with meager RMSE of 0.0034. Therefore, we believe the trained ANN models (BPNN and RBFN) developed in this study could be used by endusers for predicting the ET flux of pine forest in other locations with similar topographic (coastal flat) and climatic (costal temperate) condition. Although the models are likely robust for such conditions as dataset spanning several years and two different age groups of forest were used in model development, they can still be improved if data from other sites with varying topographic, climatic, and vegetation types could be included to capture spatial variability.

Author Contributions: S.P. is involved in the development of this manuscript in its entirety with the help of all co-authors. The jobs to develop this research publication includes conceptualization of the research study of developing a user-friendly tool for lay-person with limited to no knowledge of remote sensing/image processing, model building, and mapping of ET fluxes of pine and other vegetation. He developed the methodology with other co-authors, conducted data collection, investigation, data pre-processing (image and field data) through data mining and analysis, model development (training, testing, and validation), and results investigation through groundtruthing. He developed the simple executable software for lay person with ArcGIS ModelBuilder platform to run geospatial models with available Toolbox so that they can get the ET map of their intended study area with ArcGIS software and little modification knowledge of the automated geospatial model and filling the input text boxes of the software. S.P. wrote the initial draft and got it edited with the help of entire author-team until it was published. Journal anonymous review team helped a lot in enhancement of the manuscript. D.M.A., PI along with Co-PI S.P. administered this project with a funding acquisition from North Carolina State University through US Depart of Energy. S.P. along with other co-authors conceptualized the research idea and helped in the completion of this manuscript through field data analyses, model verification, accuracy assessment, and model/study uncertainty analyses. He is instrumental in seeing the research is completed and published for real-world application—suggest preparing an ‘easy-to-use’ tool for layperson to estimation ET flux of forest species. R.J., as the forest hydrologist, helped in the calculation of ET flux (from Flux Towers) for specific time periods coinciding with Landsat data to be used in the model developed. G.S. is the forest hydrologist who works on the study area for climate change related hydrological behavior changes analyses of the forest. His in-depth knowledge of the field (study area) was the basis of this entire study including his other publication [22] information from which helped make the resultant tool simpler and compare our study results to his study. G.S. and A.N., an Eddy Flux tower data expert, provided the processed Eddy Flux tower based ET and other associated data to the research team. The entire author team edited the manuscript with research data analyses, result verification, and software testing.

Funding: This research was partially funded by USDA Forest Service agreement # 11-CO-11330140-079 with North Carolina State University through its funding from US Department of Energy’s (US DOE) Funding Opportunity Announcement # DE-FOA-0000314, CFDA # 81.087.

Acknowledgments: Researchers acknowledge George Chescheir, PI of the US DOE Grant recipient at North Carolina State University for support, the U.S. Department of Energy for partial funding to complete this study and Jami Nettles and Cliff Tyson of Weyerhaeuser Company for their logistical support of this study. Many undergraduate students at the Institute for Environmental and Spatial Analysis at the University of North Georgia assisted in the data analysis of the project. The open access publishing fees for this article have been partially covered by the Texas A&M University Open Access to Knowledge Fund (OAKFund), supported by the University Libraries and the Office of the Vice President for Research along with USDA Forest Service and University of Georgia.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A.

Appendix A.1. Python Script for PCA Batch Scripting (Partially Shown in Figure Above)

```

-----
# PCA.py
# Created on: 2016-10-4 10:19:06.00000
# (generated by ArcGIS/ModelBuilder)
# Description:
-----

# Import arcpy module
import arcpy

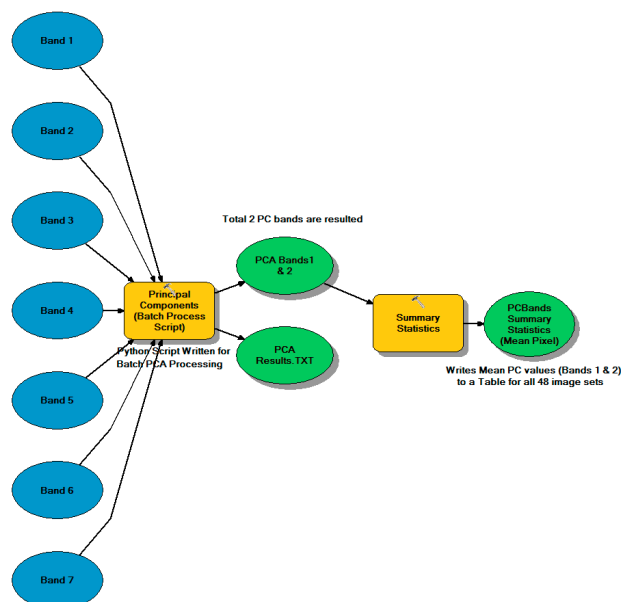
# Set the workspace. List all of the TIFF files
#
arcpy.env.workspace=
'E:\GSC_Data\Projects\DOE_CenterForForestedwetlandResearch\SC_Biomass\Ppresentations\ParkerTractModelDevelo
pment'

# For each raster in the list of rasters
#
for tiff in arcpy.ListRasters("*", "TIF"):
    # Create pyramids
    #
    arcpy.BuildPyramids_management(tiff)

# Select variables from workspace:
Band_7 = "2222_b2.img"
Band_6 = "2222_b3.img"
Band_5 = "2222_b4.img"
Band_4 = "2222_b5.img"
Band_3 = "2222_b6_1.img"
Band_2 = "2222_b7.img"
Band_1 = "2222_b8.img"
Princip_2 = "E:\\GSC_data\\Projects\\DOE_CenterForForestedwetlandResearch\\SC_Biomass\\Presentations\\
ParkerTractModelDevelopment\\Princip_2"
PCA_results_TXT = "E:\\GSC_data\\Projects\\DOE_CenterForForestedwetlandResearch\\SC_Biomass\\Presentations
\\ParkerTractModelDevelopment\\PCA_results.TXT"

# Process: Principal Components
arcpy.gp.PrincipalComponents_sa
("2222_b2.img;2222_b3.img;2222_b4.img;2222_b5.img;2222_b6_1.img;2222_b7.img;2022_B8.img", Princip_2, "2",
PCA_results_TXT)
    
```

Appendix A.2. Principal Component Analysis Model Development to Automate the Process in Model Input Parameters Creation



Appendix B. ET Flux Estimation Software Developed in Visual Basics Studio

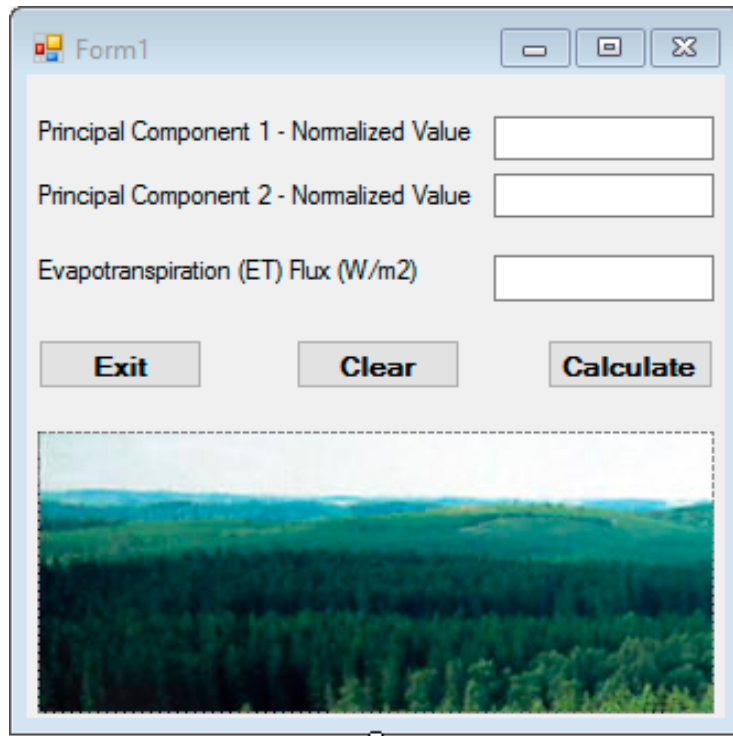


Figure A1. Form developed in VB Studio 2015.

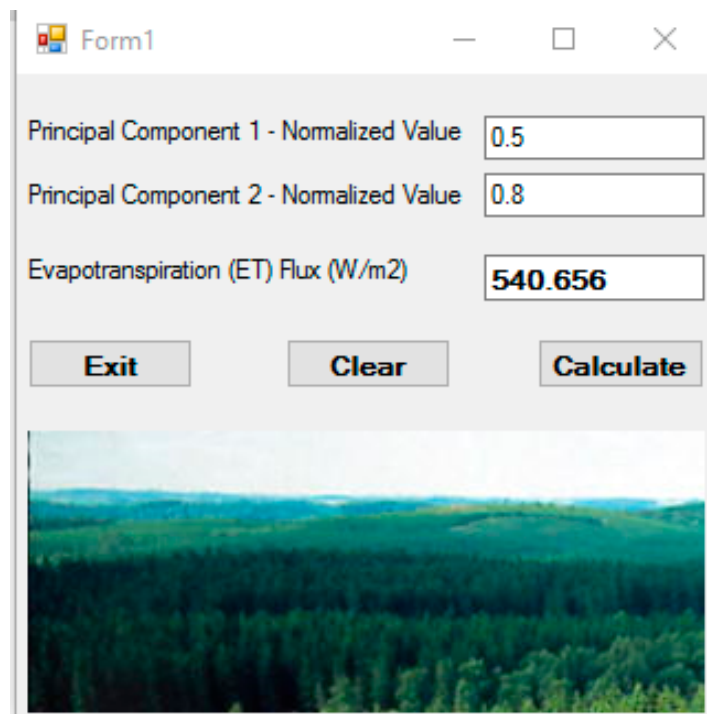


Figure A2. Executable file worked out example.

Codes:

```
Public Class Form1
```

```
Private Sub Button3_Click(sender As Object, e As EventArgs) Handles Button3.Click
Me.Close()
End Sub
```

```
Private Sub Button2_Click(sender As Object, e As EventArgs) Handles Button2.Click
decPC1.Text = String.Empty
decPC2.Text = String.Empty
txtETFlux.Text = String.Empty
End Sub
```

```
Private Sub Button1_Click(sender As Object, e As EventArgs) Handles Button1.Click
Dim decimalPC1 As Decimal 'Landsat imagery Principal Component Band 1 value
Dim decimalPC2 As Decimal 'Landsat imagery Principal Component Band 2 value
Dim decimalETFlux As Decimal 'ET Flux (w/m2) value as result
```

```
decimalPC1 = CDec(decPC1.Text)
decimalPC2 = CDec(decPC2.Text)
```

```
'Calculate the ET Flux (W/m2) using the following algorithm
decimalETFlux = 541.17 + (decimalPC1 * 4.46) - (decimalPC2 * 3.43)
txtETFlux.Text = decimalETFlux.ToString
```

```
End Sub
```

```
End Class
```

Appendix C. BPNN Testing Model Prediction Result

Actual0	Predicted0	Absolute Error0	Accuracy0
0.875407	0.582837	0.29257	66.579
0.520206	0.462012	0.058194	88.8133
0.676164	0.585652	0.090512	86.6139
0.943428	0.712031	0.231397	75.4727
0.668164	0.579851	0.088313	86.6345
1.42101	1.54351	0.12250	91.3323
1.12945	0.945626	0.18382	83.7248
1.41502	1.53342	0.118395	91.633
1.32868	1.64654	0.317865	76.0766

Max Absolute Error is: 0.317865; Min Absolute Error is: 0.058194; Average Absolute Error is: 0.1847; Average Accuracy is: 81.2743%.

Appendix D. RBFN Testing Model Prediction Result

Actual0	Predicted0	Absolute Error0	Accuracy0
0.875407	0.806028	0.069379	92.0747
0.520206	0.605002	0.084796	83.6995
0.771994	0.638304	0.13369	82.6825
0.676164	0.581602	0.094562	86.0149
0.943428	0.807281	0.136147	85.5689
1.12945	0.683315	0.446131	60.5
1.61502	1.75015	0.135132	91.6328
1.4444	1.35203	0.092362	93.6055
1.32868	1.50054	0.171864	87.0651

Max Absolute Error is: 0.446131; Min Absolute Error is: 0.069379; Average Absolute Error is: 0.151563; Average Accuracy is: 84.7604%.

References

1. Amatya, D.M.; Irmak, S.; Gowda, P.; Sun, G.; Nettles, J.E.; Douglas-Mankin, K.R. Ecosystem Evapotranspiration: Challenges in Measurements, Estimates, and Modeling. *Trans. ASABE* **2016**, *59*, 555–560.
2. Sun, G.; Caldwell, P.; Noormets, A.; Cohen, E.; McNulty, S.G.; Treasure, E.; Chen, J. Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model. *Rev. Geophys.* **2011**, *116*. [[CrossRef](#)]
3. Sun, G.; Alstad, K.; Chen, J.; Chen, S.; Ford, C.R.; Lin, G.; Lu, N.; McNulty, S.G.; Noormets, A.; Vose, J.M.; et al. A general predictive model for estimating monthly ecosystem evapotranspiration. *Ecohydrology* **2011**, *4*, 245–255. [[CrossRef](#)]
4. Amatya, D.M.; Skaggs, R.W. Hydrologic modeling of pine plantations on poorly drained soils. *For. Sci.* **2001**, *47*, 103–114.
5. Fisher, J.B.; DeBiase, T.A.; Qi, Y.; Xu, M.; Goldstein, A.H. Evapotranspiration models compared on a Sierra Nevada forest ecosystem. *Environ. Model. Softw.* **2005**, *20*, 783–796. [[CrossRef](#)]
6. Panda, S.S.; Amatya, D.M.; Hoogenboom, G. Stomatal conductance, canopy temperature, and leaf area index estimation using remote sensing and OBIA techniques. *J. Spat. Hydrol.* **2014**, *12*, 1–25.
7. Tian, S.; Youssef, M.A.; Skaggs, R.W.; Amatya, D.M.; Chescheir, G.M. DRAINMOD-Forest: Integrated modeling of hydrology, soil carbon and nitrogen dynamics, and plant growth for drained forests. *J. Environ. Qual.* **2012**, *41*, 764–782. [[CrossRef](#)] [[PubMed](#)]
8. Pereira, L.S.; Oweis, T.; Zairi, A. Irrigation management under water scarcity. *Agric. Water Manag.* **2002**, *57*, 175–206. [[CrossRef](#)]
9. Jaramillo, F.; Cory, N.; Arheimer, B.; Laudon, H.; Van Der Velde, Y.; Hasper, T.B.; Teutschbein, C.; Uddling, J. Dominant effect of increasing forest biomass on evapotranspiration: Interpretations of movement in Budyko space. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 567–580. [[CrossRef](#)]
10. Panda, S.S.; Amatya, D.M.; Sun, G.; Bowman, A. Remote Estimation of a Managed Pine Forest Evapotranspiration with Geospatial Technology. *Trans. ASABE* **2016**, *59*, 1695–1705.
11. Wang, K.; Dickinson, R.E. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.* **2012**, *50*, 2011RG000373. [[CrossRef](#)]
12. Canny, M.J. Transporting water in plants. *Am. Sci.* **1998**, *86*, 152–159. [[CrossRef](#)]
13. Sun, G.; Domec, J.-C.; Amatya, D.M. Forest Evapotranspiration: Measurement and Modelling at Multiple Scales. In *Forest Hydrology—Processes, Management, and Assessments*; Chapter 3; Amatya, D.M., Williams, T.M., Bren, L., de Jong, C., Eds.; CABI Publisher: Wallingford, UK, 2016.
14. Wullschleger, S.D.; Meinzer, F.C.; Vertessy, R.A. A review of whole-plant water use studies in tree. *Tree Physiol.* **1998**, *18*, 499–512. [[CrossRef](#)] [[PubMed](#)]
15. Ford, C.R.; Hubbard, R.M.; Kloeppel, B.D.; Vose, J.M. A comparison of sap flux-based evapotranspiration estimates with catchment-scale water balance. *Agric. For. Meteorol.* **2007**, *145*, 176–185. [[CrossRef](#)]
16. Cienciala, E.; Lindroth, A. Gas-exchange and sap flow measurements of *Salix viminalis* trees in short rotation forest. *Trees* **1995**, *9*, 289–294. [[CrossRef](#)]
17. Baldocchi, D.; Falge, E.; Gu, L.; Olson, R.; Hollinger, D.; Running, S.; Anthoni, P.; Bernhofer, C.; Davis, K.; Evans, R. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* **2001**, *82*, 2415–2434. [[CrossRef](#)]
18. Shuttleworth, W.J. Evapotranspiration measurement methods. *Southwest Hydrol.* **2008**, *7*, 22–23.
19. Sun, G.; Noormets, A.; Gavazzi, M.J.; McNulty, S.G.; Chen, J.; Domec, J.C.; King, J.S.; Amatya, D.M.; Skaggs, R.W. Energy and water balance of two contrasting loblolly pine plantations on the lower coastal plain of North Carolina, USA. *For. Ecol. Manag.* **2010**, *259*, 1299–1310. [[CrossRef](#)]
20. Verstraeten, W.W.; Veroustraete, F.; Feyen, J. Assessment of evapotranspiration and soil moisture content across different scales of observation. *Sensors* **2008**, *8*, 70–117. [[CrossRef](#)] [[PubMed](#)]
21. Wilson, K.B.; Hanson, P.J.; Mulholland, P.J.; Baldocchi, D.D.; Wullschleger, S.D. A comparison of methods for determining forest evapotranspiration and its components: Sap-flow, soil water budget, eddy covariance and catchment water balance. *Agric. For. Meteorol.* **2001**, *106*, 153–168. [[CrossRef](#)]

22. Yang, Y.; Anderson, M.C.; Gao, F.; Hain, C.R.; Semmens, K.A.; Kustas, W.P.; Noormets, A.; Wynne, R.H.; Thomas, V.A.; Sun, G. Daily Landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA using multi-satellite data fusion. *Hydrol. Earth Syst. Sci. Discuss.* **2016**. [[CrossRef](#)]
23. Cuenca, R.H.; Stangel, D.E.; Kelly, S.F. Soil water balance in a boreal forest. *J. Geophys. Res. Atmos.* **1997**, *102*, 29355–29365. [[CrossRef](#)]
24. Amatya, D.M.; Skaggs, R.; Gregory, J. Effects of controlled drainage on the hydrology of drained pine plantations in the North Carolina coastal plain. *J. Hydrol.* **1996**, *181*, 211–232. [[CrossRef](#)]
25. Domec, J.-C.; Sun, G.; Noormets, A.; Gavazzi, M.J.; Treasure, E.A.; Cohen, E.; King, J.S. A comparison of three methods to estimate evapotranspiration in two contrasting loblolly pine plantations: Age-related changes in water use and drought sensitivity of evapotranspiration components. *For. Sci.* **2012**, *58*, 497–512. [[CrossRef](#)]
26. Klein, T.; Rotenberg, E.; Cohen Hilaleh, E.; Raz Yaseef, N.; Tatarinov, F.; Preisler, Y.; Ogée, J.; Cohen, S.; Yakir, D. Quantifying transpirable soil water and its relations to tree water use dynamics in a water limited pine forest. *Ecohydrology* **2014**, *7*, 409–419. [[CrossRef](#)]
27. Smith, D.M.; Allen, S.J. Measurement of sap flow in plant stems. *J. Exp. Bot.* **1996**, *47*, 1833–1844. [[CrossRef](#)]
28. Ceron, C.N.; Melesse, A.M.; Price, R.; Dessu, S.B.; Kandel, H.P. Operational actual wetland evapotranspiration estimation for South Florida using MODIS imagery. *Remote Sens.* **2015**, *7*, 3613–3632. [[CrossRef](#)]
29. Jaramillo, F.; Destouni, G. Comment on “Planetary boundaries: Guiding human development on a changing planet”. *Science* **2015**, *348*, 1217. [[CrossRef](#)] [[PubMed](#)]
30. Monteith, J.; Unsworth, M. *Principles of Environmental Physics*; Academic Press: New York, NY, USA, 2007.
31. Hendrickx, J.M.H.; Allen, R.G.; Brower, A.; Byrd, A.R.; Hong, S.B.; Ogden, F.L.; Pradhan, N.R.; Robison, C.W.; Toll, D.; Trezza, R.; et al. Benchmarking Optical/Thermal Satellite Imagery for Estimating Evapotranspiration and Soil Moisture in Decision Support Tools. *J. Am. Water Resour. Assoc. (JAWRA)* **2016**, *52*, 89–119. [[CrossRef](#)]
32. Li, F.; Lyons, T.J. Remote estimation of regional evapotranspiration. *Environ. Model. Softw.* **2002**, *17*, 61–75. [[CrossRef](#)]
33. Cristobal, J.; Poystos, R. Combining remote sensing and GIS climate modeling to estimate daily forest evapotranspiration in a Mediterranean mountain area. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1563–1575. [[CrossRef](#)]
34. Hwang, K.; Choi, M. Seasonal trends of satellite based evapotranspiration algorithms over a complex ecosystem in East Asia. *Remote Sens. Environ.* **2013**, *137*, 244–263. [[CrossRef](#)]
35. Batra, N.; Islam, S.; Venturini, V.; Bisht, G.; Jiang, L. Estimation and comparison of evapotranspiration from MODIS and AVHRR sensors for clear sky days over the Southern Great Plains. *Remote Sens. Environ.* **2006**, *103*, 1–15. [[CrossRef](#)]
36. Jiang, L.; Islam, S. Estimation of surface evaporation map over southern Great Plains using remote sensing data. *Water Resour. Res.* **2001**, *37*, 329–340. [[CrossRef](#)]
37. Wang, K.; Li, Z.; Cribb, M. Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley–Taylor parameter. *Remote Sens. Environ.* **2006**, *102*, 293–305. [[CrossRef](#)]
38. Mu, Q.; Heinsch, F.A.; Zhao, M.; Running, S.W. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sens. Environ.* **2007**, *111*, 519–536. [[CrossRef](#)]
39. Byun, K.; Liaqt, U.W.; Choi, M. Dual-model approaches for evapotranspiration analyses over homo & heterogeneous land surface conditions. *Agric. For. Meteorol.* **2014**, *197*, 169–187.
40. Su, Z. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 85–99. [[CrossRef](#)]
41. Lu, H.; Liu, T.; Yang, Y.; Yao, D. A hybrid dual-source model of estimating evapotranspiration over different ecosystems and implications for satellite-based approaches. *Remote Sens.* **2014**, *6*, 8359–8386. [[CrossRef](#)]
42. Senay, G.B.; Bohms, S.; Singh, R.K.; Gowda, P.H.; Velpuri, N.M.; Alemu, H.; Verdin, J.P. Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. *JAWRA J. Am. Water Resour. Assoc.* **2013**, *49*, 577–591. [[CrossRef](#)]
43. Ha, W.; Kolb, T.E.; Springer, A.E.; Dore, S.; O'Donnell, F.C.; Martinez Morales, R.; Lopez, S.M.; Koch, G.W. Evapotranspiration comparisons between eddy covariance measurements and meteorological and remote-sensing-based models in disturbed ponderosa pine forests. *Ecohydrology* **2015**, *8*, 1335–1350. [[CrossRef](#)]
44. Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of vegetation indices for agricultural crop yield prediction using neural network. *Remote Sens.* **2010**, *2*, 673–696. [[CrossRef](#)]

45. Panda, S.S.; Steele, D.D.; Panigrahi, S.; Ames, D.P. Precision water management in corn using automated crop yield modeling and remotely sensed data. *Int. J. Remote Sens. Appl.* **2011**, *1*, 11–21.
46. Ranaweera, D.K.; Hubele, N.F.; Papalexopoulos, A.D. Application of radial basis function a neural network model for short-term load forecasting. *IEEE Proc. Gener. Transm. Distrib.* **1995**, *142*, 45–50. [[CrossRef](#)]
47. Sahoo, K.; Hawkins, G.L.; Yao, X.A.; Samples, K.; Mani, S. GIS-based biomass assessment and supply logistics system for a sustainable biorefinery: A case study with cotton stalks in the Southeastern US. *Appl. Energy* **2016**, *182*, 260–273. [[CrossRef](#)]
48. Sahoo, K.; Mani, S.; Das, L.; Bettinger, P. GIS-based assessment of sustainable crop residues for optimal siting of biogas plants. *Biomass Bioenergy* **2018**, *110*, 63–74. [[CrossRef](#)]
49. Zhuang, X.; Engel, B. *Classification of Multi-Spectral Remote Sensing Data Using a Neural Network vs. Statistical Methods*; ASABE Paper No. 90-7552; American Society of Agricultural and Biological Engineers (ASABE): St. Joseph, MI, USA, 1990.
50. Adeloje, A.J.; Rustum, R.; Kariyama, I.D. Neural computing modeling of the reference crop evapotranspiration. *Environ. Model. Softw.* **2012**, *29*, 61–73. [[CrossRef](#)]
51. United States Department of Agriculture (USDA) Forest Service. Database for Landscape-scale Carbon Monitoring Sites. 2014. Available online: <http://www.nrs.fs.fed.us/data/lcms/tpt/> (accessed on 20 November 2014).
52. Noormets, A.; McNulty, S.G.; DeForest, J.L.; Sun, G.; Li, Q.; Chen, J. Drought during canopy development has lasting effect on annual carbon balance in a deciduous temperate forest. *New Phytol.* **2008**, *179*, 818–828. [[CrossRef](#)] [[PubMed](#)]
53. Byrne, G.F.; Crapper, P.F.; Mayo, K.K. Monitoring land-cover change by principal component analysis of multitemporal Landsat data. *Remote Sens. Environ.* **1980**, *10*, 175–184. [[CrossRef](#)]
54. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
55. Wood, C.C.; McCarthy, G. Principal component analysis of event-related potentials: Simulation studies demonstrate misallocation of variance across components. *Electroencephalogr. Clin. Neurophysiol.* **1984**, *59*, 249–260. [[CrossRef](#)]
56. Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 2001.
57. Stein, R. Preprocessing data for neural networks. *AI Expert* **1993**, *7*, 31–37.
58. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall, Inc.: Upper Saddle River, NJ, USA, 1999.
59. Moody, J.; Darken, C.J. Fast learning in networks of locally-tuned processing units. *Neural Comput.* **1989**, *1*, 281–294. [[CrossRef](#)]
60. Liu, W.; Hong, Y.; Khan, S.I.; Huang, M.; Vieux, B.; Caliskan, S.; Grout, T. Actual evapotranspiration estimation for different land use and land cover in urban regions using Landsat 5 data. *J. Appl. Remote Sens.* **2010**, *4*, 041873.
61. Wang, S.; Fu, B.J.; Gao, G.Y.; Yao, X.L.; Zhou, J. Soil moisture and evapotranspiration of different land cover types in the Loess Plateau, China. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 2883–2892. [[CrossRef](#)]

