

Article

## Assessment of Surface Soil Moisture Using High-Resolution Multi-Spectral Imagery and Artificial Neural Networks

Leila Hassan-Esfahani \*, Alfonso Torres-Rua, Austin Jensen and Mac McKee

Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84341, USA; E-Mails: a.torres@aggiemail.usu.edu (A.T.-R.); austin.jensen@usu.edu (A.J.); mac.mckee@usu.edu (M.M.)

\* Author to whom correspondence should be addressed; E-Mail: Leila.esfahani@aggiemail.usu.edu; Tel.: +1-435-764-3151; Fax: +1-435-797-3663.

Academic Editors: Nicolas Baghdadi and Prasad S. Thenkabail

Received: 15 September 2014 / Accepted: 17 February 2015 / Published: 5 March 2015

---

**Abstract:** Many crop production management decisions can be informed using data from high-resolution aerial images that provide information about crop health as influenced by soil fertility and moisture. Surface soil moisture is a key component of soil water balance, which addresses water and energy exchanges at the surface/atmosphere interface; however, high-resolution remotely sensed data is rarely used to acquire soil moisture values. In this study, an artificial neural network (ANN) model was developed to quantify the effectiveness of using spectral images to estimate surface soil moisture. The model produces acceptable estimations of surface soil moisture (root mean square error (RMSE) = 2.0, mean absolute error (MAE) = 1.8, coefficient of correlation ( $r$ ) = 0.88, coefficient of performance ( $e$ ) = 0.75 and coefficient of determination ( $R^2$ ) = 0.77) by combining field measurements with inexpensive and readily available remotely sensed inputs. The spatial data (visual spectrum, near infrared, infrared/thermal) are produced by the AggieAir™ platform, which includes an unmanned aerial vehicle (UAV) that enables users to gather aerial imagery at a low price and high spatial and temporal resolutions. This study reports the development of an ANN model that translates AggieAir™ imagery into estimates of surface soil moisture for a large field irrigated by a center pivot sprinkler system.

**Keywords:** remote sensing; high-resolution; multi-spectral; soil moisture; Artificial Neural Network; UAV

---

## 1. Introduction

Soil moisture content (SMC) is an important factor in managing irrigated farms. SMC includes two main components: surface soil moisture (SSM) (held in the upper 10 cm of soil) and root zone soil moisture (held in the upper 200 cm of soil). Surface soil moisture is a key component for addressing energy and water exchanges at the land surface/atmosphere interface and can be estimated using different techniques, such as *in situ* measurements, physically based models, remote sensing, *etc.* Grayson and Western addressed the estimation of soil moisture by applying: (1) field (or *in situ*) measurements; (2) remote sensing techniques; and (3) soil water balance simulation models [1,2]. Soil moisture constitutes a very small volume in terms of the total global water balance, but it plays a significant role in water resources planning and management [2]. Many current crop production management decisions that are made by growers, production managers, and crop advisors in precision agriculture are already based on observation from remotely sensed data such as satellite imagery. The objective of this research is to generate surface soil moisture (SSM) estimates using high-resolution, remotely sensed data, collected at 15 cm pixel resolution, as inputs to a learning machine algorithm (Artificial neural networks (ANNs)) developed under supervised learning procedures. ANNs are used to build the SSM estimation model. To our knowledge, this is the first study to document estimation of surface soil moisture using remotely sensed data at such a fine spatial resolution and readily available in the sense of temporal resolution. The results will contribute not only to efficient and reliable high-resolution multi-spectral remote sensing validation, but also to better utilization of remotely sensed soil moisture products for enhanced irrigation modeling and scheduling.

Various techniques for retrieving soil moisture content have been the subject of research for almost four decades. Gravimetric measurements of soil moisture are very reliable but are laborious. Measuring SMC with imbedded sensors, such as time and frequency domain reflectometers (TDRs and FDRs), does not require a huge investment of time or facilities; however, most of these methods suffer from some of these same disadvantages. *In situ* measurements can be exhaustive and expensive if large areas are involved, as these measurements are mainly “local,” with a particular footprint representing moisture conditions in only a fraction of a cubic meter of soil [3]. Because of the spatial heterogeneity of soil moisture due to different soil conditions, vegetation, topography, or impact of human activities, local measurements when are carried out on a larger scale such as fields or watersheds, might result in inaccuracies [4]. Remote-sensing techniques might provide a useful tool to address these data acquisition difficulties.

Some of the early works in estimating SMC using remote sensing [5–10] established that thermal remote sensing, in concert with *in situ* measurements, can be used to measure, or at least quantitatively infer, soil moisture content. The possibility of estimating SSM (0–7.6 cm) from visible and near-infrared (NIR) reflectance data has also been demonstrated [11]. Optical and thermal remote-sensing techniques or passive and active microwave sensors offer large-scale monitoring of SSM [11–13]. Some meteorological satellites, such as the Advanced Microwave Scanning Radiometer (AMSR-E), the European Remote Sensing (ERS) satellite scatterometer or the Meteorological Satellite (METEOSAT), offer the possibility of monitoring operational SSM [3]. However, the coarse spatial resolutions (ERS-Scat: 50 km, AMSR-E: 56 km and METEOSAT: visible and infrared (IR) 5 km) of the instruments are often not consistent with the scale of hydrologic processes of interest [14,15]. A number of studies on soil moisture estimations introduced the error sources that have degraded the accuracy of satellite remotely sensed soil

moisture content such that it is critical to calibrate soil moisture estimation algorithms and to validate derived products using ground-truth data. The error sources comprise radio-frequency interference (RFI) [16], vegetation water content [13,17], surface roughness [16], and land surface heterogeneity [18]. It has been stated in the literature that a space-borne sensor designed to interpret SMC on the basis of soil microwave emission, and therefore the relationship between soil dielectric constant and water content, will show considerable systematic uncertainty of around 4% with maximum figures at relatively low water content in SMC retrieval [19].

Remote-sensing measurements in the thermal IR band has given rise to the thermal inertia (TI) approach for SMC retrieval. The TI approach relates SMC to the magnitudes of the differences between daily maximum and minimum soil and crop canopy temperatures [6]. This approach retrieves SMC from models that describe TI as a function of water content [20,21]. The implementation of the TI approach is simple because knowledge of soil physical properties and climate can produce representative SMC profiles up to a depth of 1 m. The limitation of the approach, however, is its sensitivity to the uncertainty of soil physical properties, which are complex to determine spatially and are typically obtained with point measurements [22]. The TI method provides large-scale spatial coverage, but the functions are empirical and have the drawback of being site- and time-specific, such that none of them are general enough to be applied extensively [21]. Monitoring soil moisture by remote sensing includes another set of approaches that permits SSM retrieval from the information contained in satellite-derived surface temperature (Ts) and vegetation index (VI). However, one of the major drawbacks of the Ts-VI method is that, in order to have enough points in a remote-sensing image to use in the determination of the boundaries of extreme conditions, a sufficiently large number of pixels must be sampled. This limitation is a handicap when dealing with smaller scale imagery on the order of the size of a typical farm field [11].

The difficulties associated with the above introduced approaches have led researchers to look for data-driven modeling tools, such as artificial neural networks (ANNs), support vector machines (SVMs), and relevance vector machines (RVMs), to estimate soil moisture [2,3,23–27]. For example, Landsat data has been used for soil moisture estimation using relevance vector and support vector machines [3]. One of the major advantages of the machine learning approach to SMC estimation is that it can provide estimates having resolutions commensurate with remotely sensed data [3].

## 2. Materials and Methods

### 2.1. Artificial Neural Networks (ANNs)

This section presents a brief description of ANNs relevant to this study. A three-layered feed-forward neural network (FFNN) model was developed that includes “I” input neurons, “h” hidden neurons, and “o” output neurons, which can be shown symbolically as ANNs (i,h,o) [28]. Connection weights and bias connect these neurons. Input is multiplied by the connection weights. These products are simply summed, fed through a transfer function to generate a result, and then output. The hidden layer neurons usually use a sigmoidal activation function, while the output layer neurons utilize a linear activation function. The activation functions are used to transform inputs to targeted outputs with a nonlinear regression procedure. Each ANN model requires training and testing operations. In the training operation, by minimizing the cost function (Mean Squared Error (MSE) in this study), the connection

weights and bias values are optimized. Once trained, an independent set of data that was not used for training is applied to test the neural network model [26]. The issue that threatens the application of ANN-based models is the randomness of predicted output, which is fixed in this study [29]. This was carried out by applying seed generation function. Since weights are initialized randomly, seed generation function was reset to overcome the randomness of the results by fixing the weights initialization and make the results reproducible. Also the models were run for a wide range of seed values. The training operation of ANNs was performed by a back-propagation algorithm, which is the most commonly used supervised training algorithm in the multilayer feed-forward networks. The network weights are simultaneously modified by the back-propagation algorithm which seeks to minimize the difference between the targets and the computed outputs. In this kind of algorithm the processing operation is performed in a forward direction, from inputs to hidden layers and eventually to an output layer [30]. A back-propagation method uses a least-mean-square-error method and generalized-delta rule to optimize the network weights. The derivative chain rule and the gradient-descent method are utilized to adjust the network weights [31]. Forward pass and reverse pass are two main phases of the training operation. In the first phase, the input data are multiplied by the initial weights, forming weighted inputs that then are added to yield the net to each neuron. This net generates the output of the neuron after passing through an activation or transfer function.

In back-propagation networks, a derivative of the activation function modifies the network weights. Therefore, continuous-transfer functions are targeted. The Log-sigmoid transfer function and Hyperbolic tangent sigmoid transfer function are the most common continuous-transfer functions in back-propagation networks [32]. The Log-sigmoid transfer function was used in this study. The output of the neuron is transmitted to the next layer as an input, and this procedure is repeated until it reaches the output layer. The error between the network outputs and the target outputs is computed at the end of each forward pass and it is checked with a specific value. If the error passed this value, the procedure continues with a reverse pass; otherwise, training is stopped [33]. In the reverse pass, the weights in the network are modified using the error value. The modification of weights in the output layer is different from the modification of weights in the hidden layers. In the output layer, the target outputs are provided, whereas in the intermediate layers, target values do not exist [31]. Therefore, back propagation uses the derivatives of the objective function regarding the weights in the entire network to distribute the error to neurons in each layer in the entire network [33].

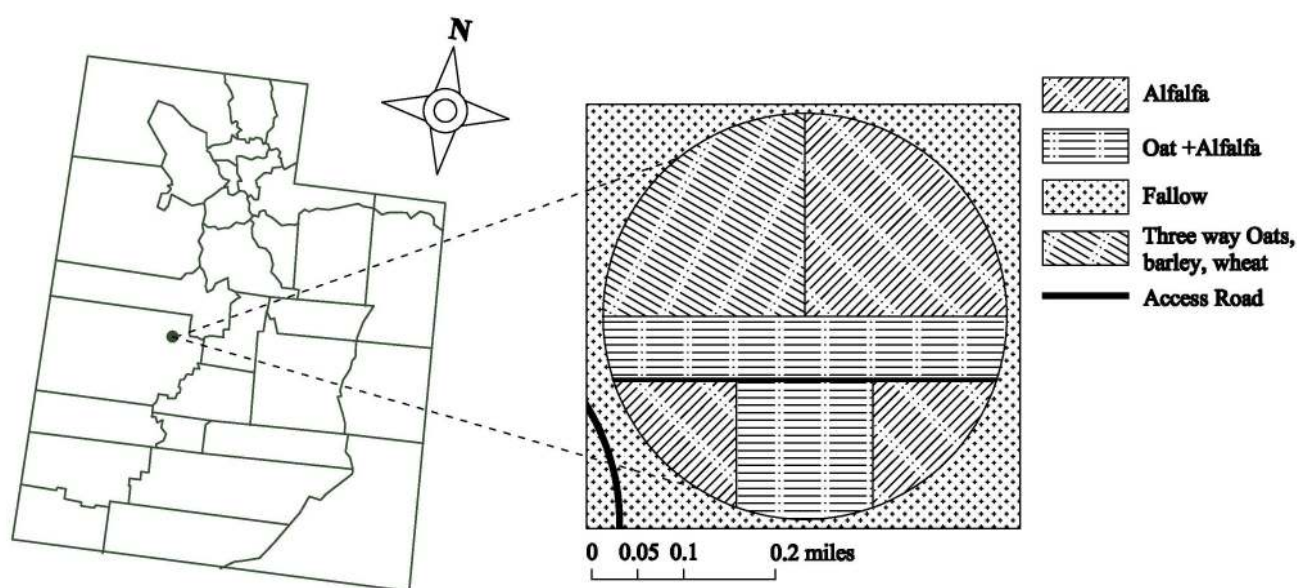
## *2.2. Selection of Possible Input Variables*

One of the critical issues in training learning machine algorithms such as ANNs is to select the appropriate input variables. The idea is to choose the combination of variables that are highly correlated with soil moisture. Previous studies have shown good correlation between soil water content and infrared (IR) skin temperature and normalized difference vegetation index (NDVI), and between IR heating rate and thermal images [34,35]. Optical and microwave remotely sensed data have been used for surface soil energy balance modeling [6,11,12,36]. After collecting these variables from independent datasets, the correlation and dependency among these variables were evaluated in the study reported here. Some Vegetation Indices (VIs) are considered as input variables with some contributions in soil moisture estimations [4,37–39].

### 2.3. Study Area, Instrumentation, Techniques and Data

#### 2.3.1. Study Area

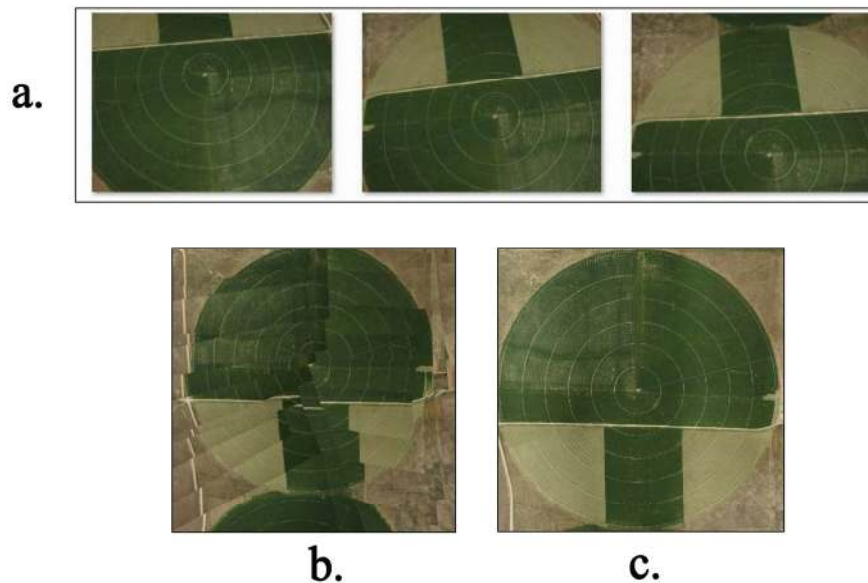
The study area is a farm in Scipio, Utah ( $39^{\circ}14'N$ ,  $112^{\circ}6'W$ ), equipped with a center pivot irrigation system covering an area of approximately 84 acres. The main crops are alfalfa and oats, grown from April to October. Figure 1 shows the location of the farm in Utah, and provides information about the heterogeneity within the farm due to different crop types and the presence of an access road. Generally the center pivot lateral rotates clockwise and supplies irrigation water to the field at a constant rate from an upstream reservoir. In the current study a full rotation of the center pivot takes three days and six hours to irrigate the field fully to field capacity. This study was carried out for the crop growing cycle starting 16 May 2013 and ending 17 June 2013 (4 days).



**Figure 1.** The location of the study area in Utah (schematic Utah counties map (on the left) and cropping pattern for 2013 irrigation season (on the right)), ( $39^{\circ}14'N$ ,  $112^{\circ}6'W$ ).

#### 2.3.2. Instrumentation: AggieAir Minion (Remote Sensing Platform)

AggieAir is the remote sensing platform applied in the current study. This platform is comprised of an autonomous unmanned aerial vehicle (UAV) that carries a multispectral sensor payload. The UAV navigates over the area of interest based on a pre-programmed flight plan and captures images using the on-board sensor payload system. The UAV is a small aircraft (8 feet wing span, 14-pound take-off weight) that can fly for an hour at a speed of 30 miles per hour. In this study, the UAV was equipped with visual, near-infrared, and thermal cameras and flew over the study area on four dates in 2013 (16 May, 1 June, 9 June, and 17 June), acquiring imagery with the optical cameras at 0.15 m resolution and with the thermal camera at about 60 cm. The wavelength range peaks around 420, 500, 600 and 800 nm, respectively, for blue, green, red and NIR sensors. Detailed information about the operation of the AggieAir system has been previously published by Jensen [40].



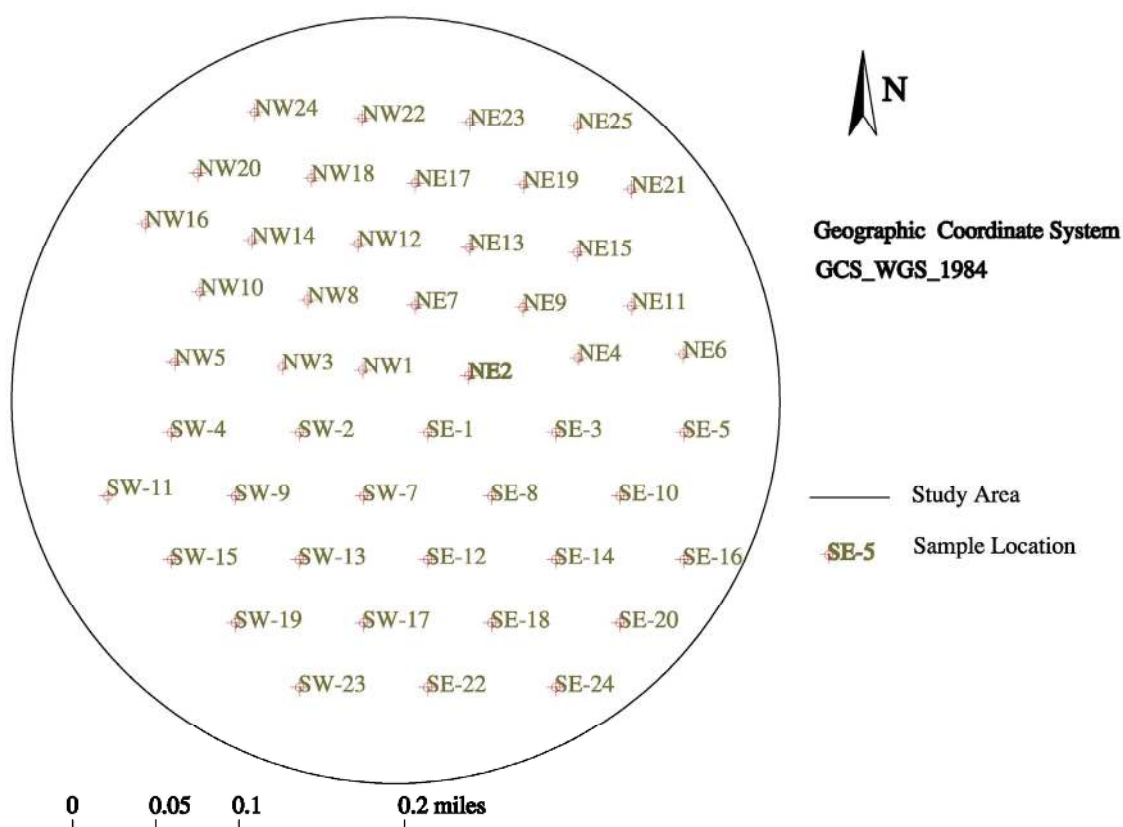
**Figure 2.** (a) Some raw natural color images from the unmanned aerial vehicle (UAV) taken from the study area ( $39^{\circ}14'N$ ,  $112^{\circ}6'W$ ); (b) Orthorectified image using position and orientation of UAV aircraft during image capture; and (c) Accurate orthorectified mosaic image from EnsoMOSAIC.

After the AggieAir UAV completes a flight mission, the aircraft may have acquired 300–400 images from each camera: visual, near-infrared, and thermal (Figure 2a). The images can be georeferenced directly using the position and orientation of the UAV when the image was exposed (Figure 2b) [40]. EnsoMOSAIC is used to orthorectify the AggieAir imagery with high accuracy [40,41]. EnsoMOSAIC generates hundreds of tie-points between overlapping images and uses photogrammetry and block adjustment to refine the position and orientation information for each image, thereby accurately georeferencing each image (Horizontal Accuracy: 1–2 pixels Vertical Accuracy: 1.5–2 pixels (when all error sources are controlled)). EnsoMOSAIC also generates an internal digital elevation model (DEM) to compensate for distortions in the imagery caused by changing elevations. The resulting product is an orthorectified mosaic (Figure 2c) that is in 8-bit digital format. AggieAir uses a modified “reflectance mode” method to convert the digital numbers of the mosaic to reflectance values [40]. This radiometric normalization is the ratio of the digital number from the mosaic to the digital number from a spectralon white reflectance panel with known reflectance coefficients, multiplied by the reflectance factor which accounts for the zenith angle of the sun at the time, date, and location of the photos. The product of this method is an orthorectified mosaic in reflectance values. The reflectance values (for all four flights) range from 0.11 to 0.36, 0.20 to 0.49, 0.15 to 0.51 and 0.51 to 0.61 for blue, green, red and NIR, respectively. Thermal values range from 10.2 to 43.3 degrees Celsius.

### 2.3.3. Ground-Based Data Collection

In order to perform ground truthing, at the same time the AggieAir UAV flew over the study area, intensive ground sampling was conducted at precisely determined locations in the field [42]. Soil samples were collected based on a pre-defined spatial distribution map that was developed in light of the crop types and soil characteristics in the field. The data collection included almost 50 samples per

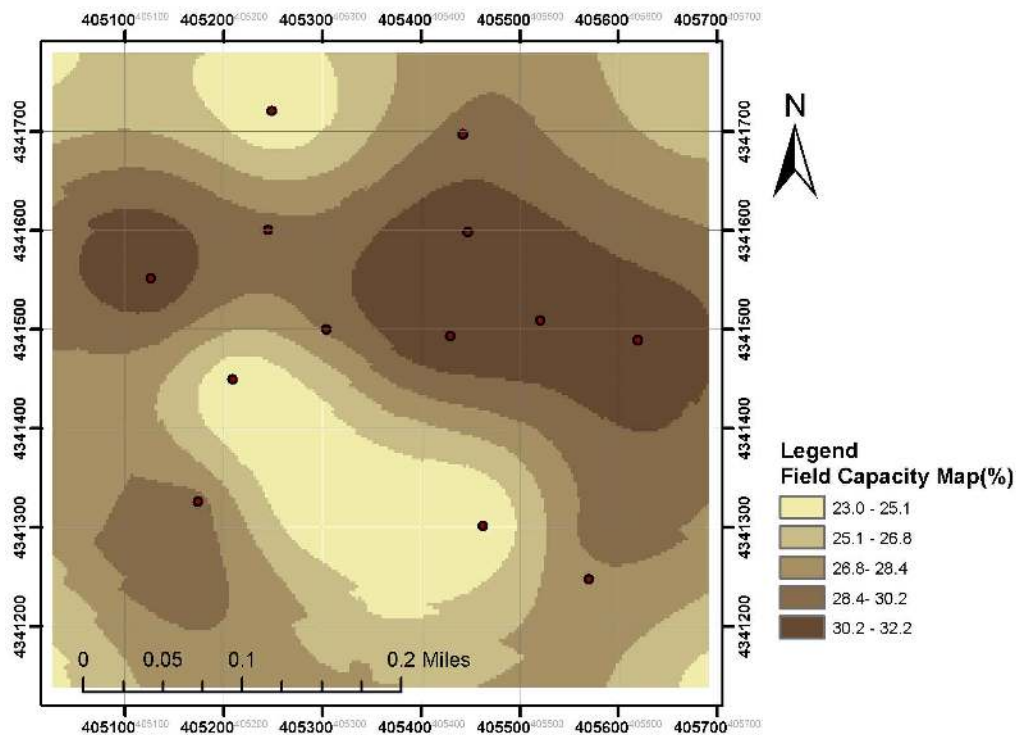
AggieAir flight scattered all over the field (minimum of 12 in each quarter) to cover the soil condition properties. Further, the unusable samples were discarded and the data collected from the four days were pooled (making a data set of 184 points) and utilized in the modeling procedure. The research crew collected soil samples from the surface soil and determined gravimetric soil moisture values after the samples were oven dried and weighted. The crew also used a hand-held measuring device to make in-field measurements and double-check the laboratory soil moisture results. The device, manufactured by Decagon Inc. (Pullman, WA, USA), includes a sensor read-out and storage system for real-time readings. Called “Procheck,” it was connected to a GS3 soil moisture, temperature, and EC sensor from Decagon Inc as well [43]. Figure 3 illustrates the location of soil moisture samples in the study area.



**Figure 3.** Spatial distribution of soil moisture sample locations in the study area.

#### 2.3.4. Soil Texture Analysis

The upper and lower limits of soil moisture storage in the root zone are a function of soil texture. After the soil has been saturated and drained by gravity, the soil is said to be at “field capacity,” and the amount of water that remains in the root zone but which the crop can no longer extract is called the “wilting point” [44]. In order to take these two parameters in to account, 14 different points from around the field were selected for soil texture sampling. After soil type determination, the corresponding field capacity values were acquired from previously published values and considered as model inputs [45]. Figure 4 illustrates the soil field capacity map developed by utilizing a Spherical Kriging interpolation method for the information from the 14 available sampling locations.



**Figure 4.** Map of field capacity based on soil texture type and plot of the location of soil samples.

### 2.3.5. Relevant Vegetation Indices (VIs) from AggieAir Imagery

Visual spectrum (red, green, and blue, or RGB), near-infrared (NIR), and infrared/thermal remotely sensed data and some vegetation indices (VIs) are used as input variables for the soil moisture model. All AggieAir data (RGB, NIR, and thermal imagery), normalized difference vegetation index (NDVI), vegetation condition index (VCI), enhanced vegetation index (EVI), vegetation health index (VHI), and field capacity were chosen as model inputs with surface soil moisture as the target or output. The VHI was proposed by Kogan (1995), which is an additive combination of VCI and Temperature Condition Index (TCI) [36]. Equations ((1)–(5)) represent the vegetation indices included in this study:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + \rho_{\text{RED}}} \quad (1)$$

$$\text{VCI} = 100 \times \frac{\rho_{\text{NDVI}} - \rho_{\text{NDVI}_{\min}}}{\rho_{\text{NDVI}_{\max}} - \rho_{\text{NDVI}_{\min}}} \quad (2)$$

$$\text{EVI} = 2.5 \times \frac{\rho_{\text{NIR}} - \rho_{\text{RED}}}{\rho_{\text{NIR}} + C_1 \times \rho_{\text{RED}} - C_2 \times \rho_{\text{BLUE}} + L} \quad (3)$$

$$\text{TCI} = 100 \times \frac{\text{BT}_{\max} - \text{BT}}{\text{BT}_{\max} - \text{BT}_{\min}} \quad (4)$$

$$\text{VHI} = 0.5 \times \text{VCI} + 0.5 \times \text{TCI} \quad (5)$$

where  $\rho_{\text{NIR}}$ ,  $\rho_{\text{RED}}$  and  $\rho_{\text{BLUE}}$  are NIR, red, and blue reflectance bands;  $C_1$ ,  $C_2$  and  $L$  are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band; and  $\text{BT}$  is the thermal brightness, which is the thermal band reflectance.



### 2.3.6. Model Validation

A K-fold cross validation was used as the model validation technique in order to generalize an independent dataset. In general, in K-fold cross validation the original dataset (including all samples) is partitioned into K sub-data sets. Each time, a single sub-data set is retained for evaluation and the remaining (K-1) sub-data sets are used for training. This process repeats K times, and the errors for each time are estimated. Furthermore, the K model errors are averaged to represent the best model [46,47]. Since the authors were not confident about the optimal percentage of data being considered for training, testing and validation to avoid over-fitting, a 5-fold cross validation technique is applied to the original data set and Mean Squared Error (MSE) is the calculated evaluation criterion. The 5-fold cross-validation was done repeatedly, and during the training phase different values for the training technique's parameters were used in concert with different network architectures. Further, the authors ended up with the best values for number of hidden nodes and training parameters. Then with these in hand, finally the network was trained using all the data, with the best number of hidden nodes and training parameters.

### 2.3.7. Wrapper Selection

For model construction, it is necessary to identify the best combination of input variables from the available data. A wrapper selection method was used to accomplish this. Guyon (2003) introduced the advantages of applying this method with reference to three main aspects: (1) improving the performance of predictors; (2) obtaining faster and more cost-effective predictors; and (3) providing a better understanding of the underlying process that generated the data [48]. This method is recommended over the backward selection method and is applicable to cases with a small number of inputs. Wrapper selection considers all possible combinations of input variables and develops a separate model for each combination. The models are then scored based on their predictive power, and the best model can be selected based on the corresponding score [48,49]. In order to check the goodness of fit, root mean square error (RMSE), mean absolute error (MAE), coefficient of correlation (r), coefficient of performance (e), and coefficient of determination ( $R^2$ ) are the statistical parameters that were calculated to evaluate the performance of the many alternative models and score their predictive power [50].

### 2.3.8. Division Set Up in ANN Model Architecture

The input data division set up can have a significant influence on the performance of an ANN model. Bowden (2002) presented two methodologies for dividing data into representative subsets (training, testing and validation) with similar statistical properties. These methods were proven to develop more robust results compared to conventional approaches in which the dataset was simply divided into arbitrary subsets [51]. The methods were applied by using a 5-fold cross validation method for data generalization. Other water resources related studies have utilized Bowden's approach and concluded that it ensures that the training, testing, and validation sets are representative of the same population [52–57].

It is difficult to assess beforehand how large an artificial neural network model should be for a specific application to avoid over-fitting. Model size strongly relates to sample size, and collecting more data and increasing the size of the training set or reducing the size of a network are recommended as

solutions [28]. In this study, collecting more data was impossible; therefore, the error of the validation data set was checked as alternative method of investigation [28]. As training initiates, the error for all three data sets (training, testing, and validation) decreases, and in the case of over-fitting, the error for validation set increases while the error in the training set maintains a decreasing trend. If the error in the validation set continues in a reducing trend, there is no danger of over-fitting.

### 3. Results and Discussion

#### 3.1. Input Data

##### 3.1.1. Soil Moisture Data Calculation Results

In order to ground truth data and relate soil moisture values to remotely sensed data, gravimetric soil moisture measurements were checked with the corresponding in-field measurements of volumetric soil moisture using soil bulk density values that were extracted from soil texture data. A *t*-test comparing the gravimetric soil moisture measurements against the volumetric soil moisture measurements showed these two data sets are not statistically different at a 95 percent confidence level with *P*-value of 0.3. The results from the *t*-test indicates that either of these data sets can be used for further calculations. Finally, the gravimetric soil moisture values from four flight dates were pooled representing the maximum, minimum and mean values of 30.6, 10.1 and 19.7, respectively, and used as model targets. Also, the spatial distribution of soil moisture from high to low values is in accordance with time after irrigation. The highest values occur immediately behind the center pivot lateral, and the driest spots were concentrated in front of the lateral.

**Table 1.** Temporal and spatial changes in normalized difference vegetation index (NDVI) values during the study period.

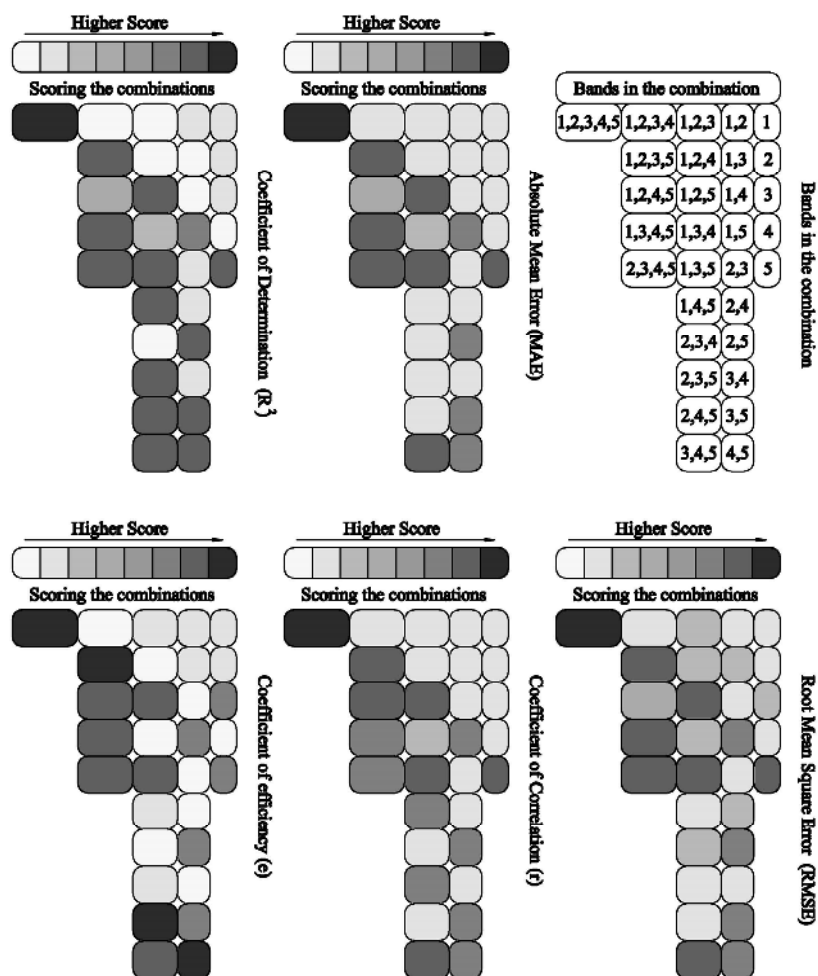
Crop Type/Date	NDVI(Mean)			
	5/17/2013	6/1/2013	6/9/2013	6/17/2013
Three way, Oat, Barley , Wheat	0.09 Planting	0.34 continued growth	0.43 continued growth	0.53 full growth
Alfalfa	0.42 continued growth	0.47 continued growth	0.53/0.08 full growth/ after cut	0.59/0.13 full growth/ after cut
Oat, Alfalfa	0.43 germination	0.48 continued growth	0.53 full growth	0.57 full growth

##### 3.1.2. Spatial Information of Vegetation Indices

Due to heterogeneity within the field because of the different crop types, an access road, wheel tracks, the center pivot station, and historic locations of fence lines and ditch banks that once occupied the modern field, spatial analysis was required. The significance of spatial information comes from the ability of the human brain to detect spatial patterns in a map or an image. Table 1 represents the temporal and spatial changes of NDVI during the study period. The same information is provided for other three VIs in the supplementary material.

### 3.2. Wrapper Selection Outcome

Goodness-of-fit statistics were used to test the degree of association between the observed and estimated data. As noted previously, root mean square error (RMSE), mean absolute error (MAE), coefficient of correlation (r), coefficient of performance (e) and coefficient of determination ( $R^2$ ) were calculated for the models to score their predictive power. In the scoring phase, the authors referred to RMSE and judged the models predictive power based on them, further MAE and  $R^2$  were considered and finally e and r came to account.



**Figure 5.** Schematic view of possible performance by 5-input dataset using wrapper selection method and artificial neural network (ANNs).

The models with high but similar predictive power were compared spatially against thermal, NIR and false color images. Also the research crew has collected a set of notes about their observations during the data collection procedure. The notes paid attention to crop types, crops growing stage, location of lateral, irrigation uniformity, wet and dry spots (created due to deficiencies in the irrigation sprinkler system), existence of wind (wind direction if it scatters the water) and weather condition. After the models with high but similar predictive power were developed, the best model was selected visually to accommodate the spatial distribution of above information. Figure 5 illustrates how schematically wrapper selection would evaluate the models for the inputs from AggieAir (RGB, NIR, and Thermal) as

an example of wrapper scoring. for this study, 1023 models in 10 sets for all possible combinations of 10 inputs were developed (10 combinations of 1, 45 combinations of 2, 120 combinations of 3, 210 combinations of 4, 252 combinations of 5, 210 combinations of 6, 120 combinations of 7, 45 combinations of 8, 10 combinations of 9 and 1 combination of 10 inputs), and the model results were compared. A trial-and-error approach was utilized to select those models that worked on different numbers of neurons (up to  $2 \times (\text{number of inputs}) + 1$  to avoid over-fitting issues), hidden layers, training functions, and division setups [32]. Finally, the model with 8 inputs (red, blue, NIR, thermal, NDVI, VCI, EVI, and field capacity) was selected because it had the best predictive power and best spatial pattern, which was checked visually. Table 2 shows the best model results for all 10 sets of combinations along with their highest predictive power statistics.

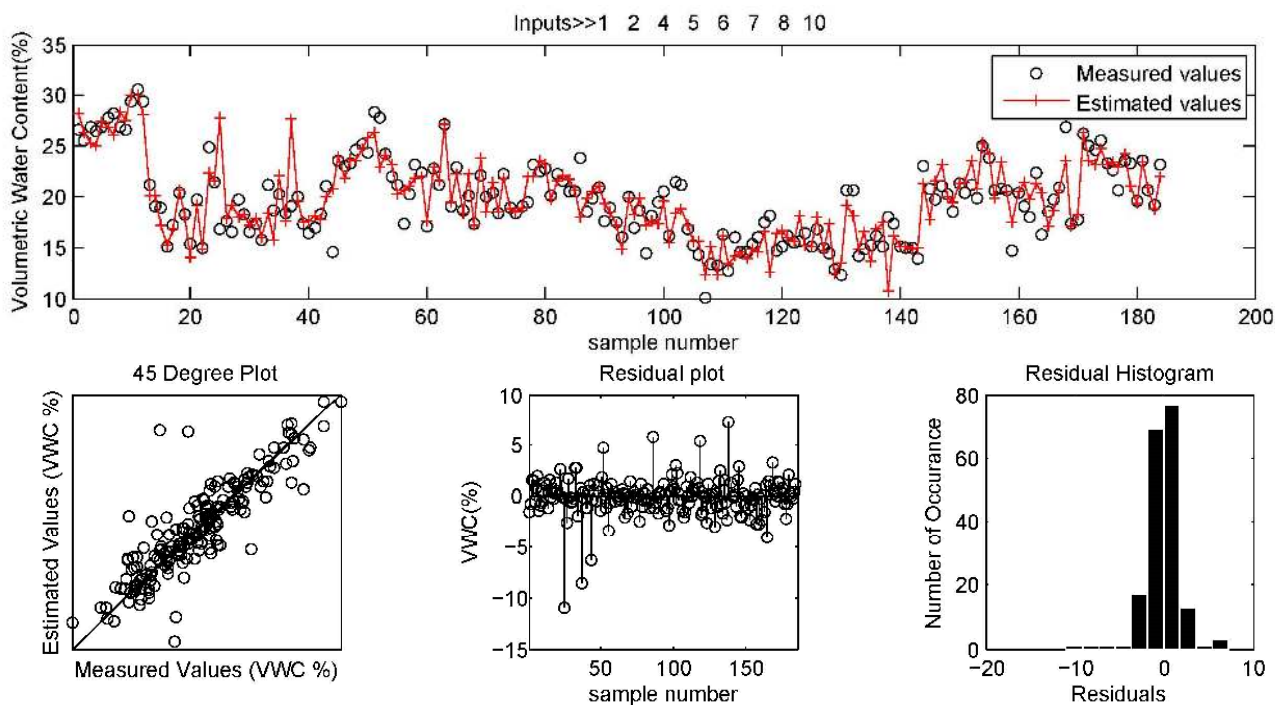
**Table 2.** Goodness-of-fit statistics from Wrapper selection results (1 to 10 inputs) with highest predictive power using ANN.

	<u>ANN Inputs</u>	<u>Division Set up</u>	<u># of Neurons</u>	<u>RMSE</u>	<u>MAE</u>	<u>r</u>	<u>ε</u>	<u>R<sup>2</sup></u>
<b>One Input</b>	Thermal	80/10/10	4	3.0	2.4	0.64	0.4	0.41
<b>Two Inputs</b>	Thermal, Field capacity	75/15/10	5	2.5	1.8	0.78	0.60	0.61
<b>Three Inputs</b>	Red, Blue, Thermal	70/15/15	7	2.7	2.1	0.74	0.54	0.55
<b>Four Inputs</b>	Red, NDVI, VCI, VHI	70/15/15	7	2.5	1.8	0.77	0.59	0.60
<b>Five Inputs</b>	Green, Thermal, VCI, EVI, Field Capacity	80/10/10	9	2.1	1.6	0.84	0.71	0.71
<b>Six Inputs</b>	NIR, Thermal, NDVI, EVI, VHI, Field Capacity	80/10/10	11	2.1	1.5	0.84	0.70	0.71
<b>Seven Inputs</b>	Red, Blue, NIR, Thermal, NDVI, VCI, Field Capacity	80/10/10	12	2.1	1.6	0.86	0.73	0.73
<b>Eight inputs</b>	<b>Red, Blue, NIR, Thermal, NDVI, EVI, VCI, Field Capacity</b>	<b>80/10/10</b>	<b>17</b>	<b>2.0</b>	<b>1.3</b>	<b>0.85</b>	<b>0.75</b>	<b>0.77</b>
<b>Nine Inputs</b>	Red, Green, Blue, Thermal, NDVI, EVI, VCI, VHI, Field Capacity	80/10/10	17	2.0	1.4	0.87	0.75	0.75
<b>Ten Inputs</b>	Red, Green, Blue, NIR, Thermal, NDVI, EVI, VCI, VHI, Field Capacity	80/10/10	19	2.0	1.3	0.85	0.73	0.73

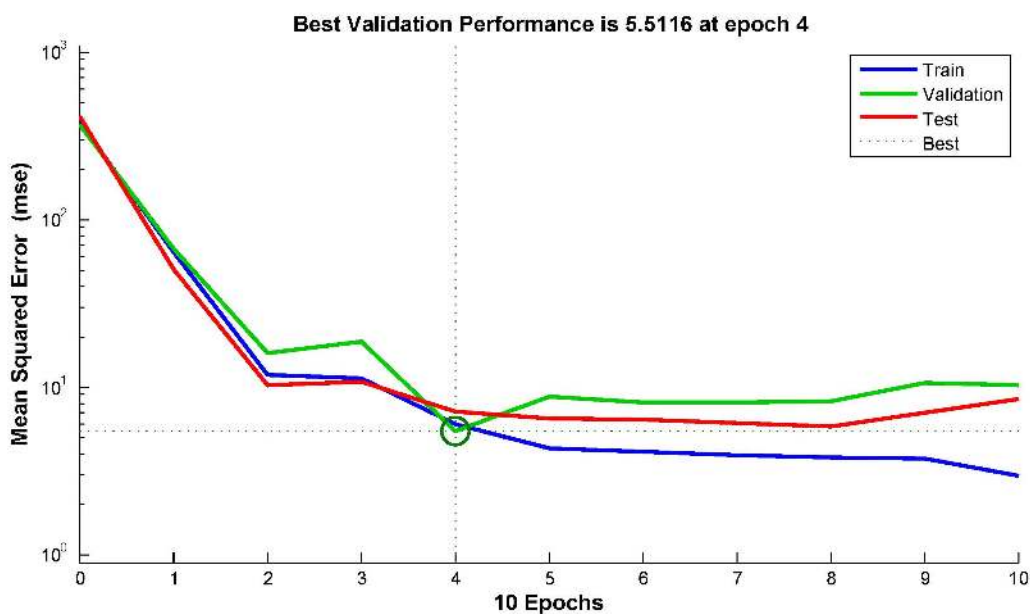
### 3.3. Results Extracted from Artificial Neural Networks (ANNs)

After the intensive trial and error selection procedure using cross validation procedure, a network architecture with one hidden layer and 17 nodes and a division set up of 80:10:10 with trainlm (Levenberg-Marquardt backpropagation) as a training function was selected. Figure 6 illustrates (a) the measured *versus* the estimated soil moisture values of the selected model; (b) the corresponding one-by-one scatter plot (showing that all the points are clustered along the 45° line); (c) residual plot; and (d) residual histogram to demonstrate the validation of the model in the sense of normality, linearity, and equality of variances. Figure 7 represents the performance plot of the selected ANN model with 8 inputs, showing Mean Squared Error (MSE) trend during the learning procedure for the training, testing and validation sets to evaluate possible over-fitting issues. The decreasing trend of the validation set confirms that there is no over-fitting in the model. Figure 8 shows estimated surface soil moisture maps (Volumetric Water Content (%)) (right column), at four different dates (a) 16 May, (b) 1 June, (c) 9 June

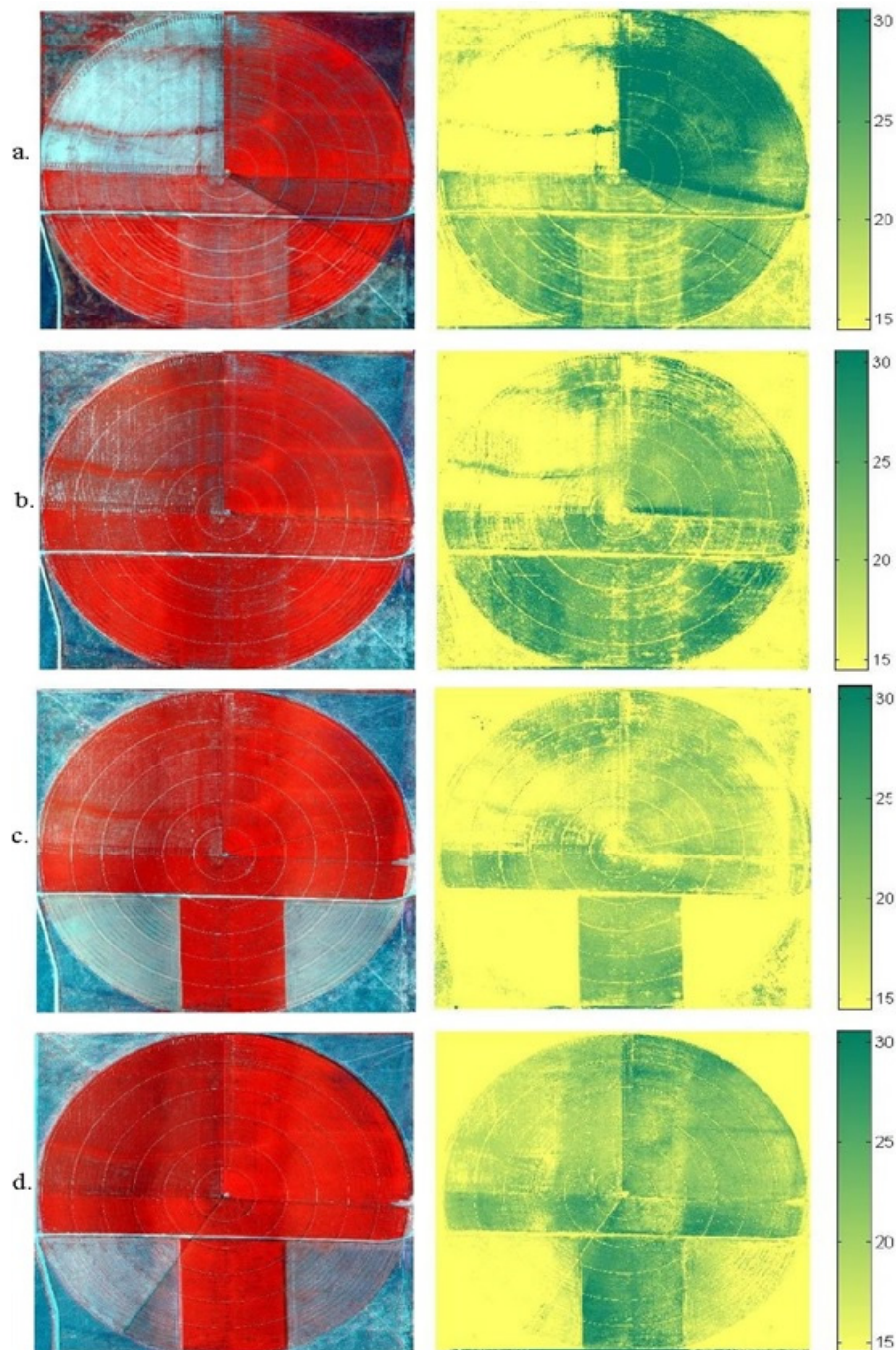
and (d) 17 June 2013 (Res. 15 cm) alongside false color composite images (NIR–Green–Blue) (left column). The false color map (NIR-Green-Blue) is related to the relative density of vegetation in the image. Exposed soil (bare) is expected to have lower soil moisture content while areas with high vegetation density the opposite. The concept of using false color composite images was taken from previously published studies [56,58].



**Figure 6.** (a) The measured soil moisture values *versus* the estimated values of the selected model; (b) one-by-one scatter plot; (c) residual plot; and (d) residual histogram.



**Figure 7.** Performance plot of the selected ANN model with 8 inputs showing Mean Squared Error (MSE) trend during the leaning procedure for three data sets (training, testing and validation).



**Figure 8.** Estimated soil moisture maps (Volumetric Water Content (%)), with ANNs model for four different dates (a) 16 May; (b) 1 June; (c) 9 June and (d) 17 June 2013 (Res. 15 cm) (right), false color images (left).

As shown in Figure 8, the soil moisture maps have a direct association with the false color composite maps. The field exterior area was not irrigated during the growing cycle and was expected to be less moist. Although the wheel tracks and the access road are located within the irrigation zone, they are expected to be drier since they are covered by bare soil, become more compacted due to traffic over them, and lose moisture rapidly. This assumption also applies to the zones where the crops have been cut.

Different crop types have different water demands and water up-take rates that cause surface soil moisture heterogeneity even after a uniform irrigation event. This heterogeneity appeared in the form of

cropping patterns in the soil moisture maps. According to the clockwise rotation of the lateral, the spots with the maximum soil moisture values are expected to fall near the lateral and in a counterclockwise direction. This status is clearest in Figure 7a where the field was under a heavy irrigation event at the time the aerial imagery was captured. Table 3 shows the comparison between measured and estimated soil moisture values for different crop type zones. Soil moisture numerical values are presented in the supplementary materials.

**Table 3.** Comparison between measured and estimated soil moisture values for different crop type zones.

Crop Type/Date	Soil Moisture (Volumetric Water Content (%)) (Zonal Mean)							
	5/17/2013		6/1/2013		6/9/2013		6/17/2013	
	Measured	Estimated	Measured	Estimated	Measured	Estimated	Measured	Estimated
Three way, Oat, Barley , Wheat	18.9	20.1	21.5	19.0	14.9	15.7	19.3	17.6
Alfalfa	27.6	25.9	25.0	23.5	18.4	18.9	21.2	20.4
Oat, Alfalfa	18.0	18.3	20.5	18.1	15.7	15.8	22.5	18.9

The main problem with such modelling procedures is being dependent to site and time. This implies that ground sampling and modeling will be required for every flight to ensure accurate and quality data. So far this is a handicap and should be strengthened with more studies over different types of crops, in different areas at different stages of growth. Having such a model (or a collection of models) makes this practical for routine use (independent of site and time). In addition, the current study was targeted toward showing the detailed information that can be interpreted from high resolution data. Even though such a high resolution might not be required for monitoring agricultural farm conditions that are cropped with inexpensive crops such as alfalfa and oats, this resolution presents its value for other crops that require high resolution data (e.g., vineyards, orchards). These results essentially help to justify future work to look at the value of high resolution data for precision farming activities.

One step forward in generalizing the presented modeling methodology in temporal scale could be the idea of pooling the soil moisture data collected from different dates. In the case of the current study, every single sampling location experiences four different conditions of soil moisture level, which provides a wide range of information about soil moisture status through time. This type of information makes the model more robust in its ability to simulate previously unseen soil moisture conditions through time.

#### 4. Conclusions

This paper demonstrates the application of a high resolution remote sensing technology (AggieAir) for estimating surface soil moisture as a key piece of information in irrigation water management. High-resolution multi-spectral imagery, in combination with ground sampling, provided enough information for the modelling approaches to accurately estimate spatially distributed surface soil moisture.

This paper presents the results of a modelling approach utilizing ANN in concert with time and site specific information. Parallel to other modeling approaches, such as data mining algorithms or linear regression, the ANN model is calibrated for this study within the conditions of the information collected including soil moisture measurements, soil texture, crop type information, and high resolution multi-spectral imagery. While the site-specific calibrated ANN in this study cannot be used immediately

in another location, the modeling procedure (identifying spatial information with the most significant contribution to soil moisture estimation (Table 2)) along with similar field measurements and high resolution multi-spectral imagery and the data mining algorithm, are transferable from this study.

Surface soil moisture estimation was accomplished with an ANN model (RMSE: 2.0, MAE: 1.3,  $r$ : 0.87,  $e$ : 0.75,  $R^2$ :0.77) for four dates in 2013 (16 May, 1 June, 9 June, and 17 June). These results show the capability of the model to accurately estimate surface soil moisture. Compared to the traditional soil moisture estimations that are based on a farmer's visual perceptions or a few soil moisture samples averaged across the farm, the modelling approach presented enables greater precision in the application of water and identifies dry/wet spots and water stressed crops.

AggieAir imagery, combined with appropriate analytic tools, allows spatial estimation of surface soil moisture. These estimates were made at much finer resolutions in space and time than those available from conventional remote sensing technologies (e.g., satellite or commercial aerial photography services). Also, the application of data mining algorithms to AggieAir aerial imagery allows for quantification of actionable information for precision agriculture (soil moisture values across the field). The soil moisture maps that are produced can then be related to irrigation water management for scheduling and application rates.

The results from the wrapper selection (Table 2) prove the significance of thermal imagery as the most relevant information in surface soil moisture estimations. In the case of one input, a model with thermal imagery can estimate the soil moisture values with RMSE of approximately 3% (thermal images are provided in supplementary material section).

Soil water holding capacity as a function of soil texture plays an important role in soil moisture values. This parameter was observed by utilizing field capacity as an input to the models. Table 2 shows that field capacity is a component of most of the models. The effect of this parameter is confounded by other important inputs in the spatial distribution of soil moisture in Figure 8. Based on the information presented in Table 2, among the available vegetation indices, NDVI and VCI have a greater explanatory contribution in surface soil moisture estimates. The soil moisture maps have a good association with false color composite maps that allows for distinction of agricultural features in the field.

## 5. Future Work

Further studies will involve the estimation of surface soil moisture using other data mining algorithms and its application as a boundary condition to produce remotely sensed estimates of root zone soil moisture. In addition, Pixel-wise estimation of soil moisture could also be applied in a water balance models.

## Acknowledgments

This project was financially supported by the Utah Water Research Laboratory through an MLF Seed Grant. The authors would like to thank the Utah Water Research Laboratory, Utah State University, and the Provo, Office of the US Bureau of Reclamation for their support of this research. The authors acknowledge the extraordinary efforts of the AggieAir flight team and Andres Ticlavilca, Roula Bachour, and Manal Elarab. The authors appreciate the support of Ivan Robins, the farm owner, whose cooperation greatly improved of the data collection procedure.



## Author Contributions

Austin Jensen guided the AggieAir team in preparing the high-resolution imagery as the required data. Mac McKee and Alfonso Torres-Rua developed the research plan and experiment design and supervised the work required for this paper. Leila Hassan-Esfahani completed the literature review, data acquisition, method selection, modelling, manuscript preparation, and discussions. All authors shared equally in the editing of the manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Grayson, R.B.; Western, A.W. Towards areal estimation of soil water content from point measurements: Time and space stability of mean response. *J. Hydrol.* **1998**, *207*, 68–82.
2. Gill, K.; Kemblowski, M.; McKee, M. Soil moisture data assimilation using support vector machines and ensemble Kalman filter. *JAWRA J. Am. Water Resour. Assoc.* **2007**, *43*, 1004–1015.
3. Zaman, B.; McKee, M.; Neale, C.M.U. Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines. *Int. J. Remote Sens.* **2012**, *33*, 6516–6552.
4. Liu, W.; Baret, F.; Gu, X.; Zhang, B.; Tong, Q.; Zheng, L. Evaluation of methods for soil surface moisture estimation from reflectance data. *Int. J. Remote Sens.* **2003**, *24*, 2069–2083.
5. Idso, S.; Jackson, R.; Reginato, R. Estimating evaporation: A technique adaptable to remote sensing. *Science* **1975**, *189*, 991–992.
6. Idso, S.; Jackson, R.; Reginato, R. Compensating for environmental variability in the thermal inertia approach to remote sensing of soil moisture. *J. Appl. Meteorol.* **1976**, *15*, 811–817.
7. Reginato, R.; Idso, S.; Jackson, R.; Vedder, J.; Blanchard, M.; Goettelman, R. Soil water content and evaporation determined by thermal parameters obtained from ground-based and remote measurements. *J. Geophys. Res.* **1976**, *81*, 1617–1620.
8. Reginato, R.; Jackson, R.; Pinter, P. Evapotranspiration calculated from remote multispectral and ground station meteorological data. *Remote Sens. Environ.* **1985**, *18*, 75–89.
9. Jackson, T. Soil water modeling and remote sensing. *IEEE Trans. Geosci. Remote Sens.* **1986**, *24*, 37–46.
10. Quattrochi, D.; Luvall, J. Thermal infrared remote sensing for analysis of landscape ecological processes: Methods and applications. *Landsc. Ecol.* **1999**, *14*, 577–598.
11. Kaleita, A.; Tian, L.; Hirschi, M. Relationship between soil moisture content and soil surface reflectance. *Trans. ASAE* **2005**, *48*, 1979–1986.
12. Humes, K.; Kustas, W.; Jackson, T.; Schmugge, T.; Moran, M. Combined use of optical and microwave remotely sensed data for the estimation of surface energy balance components over a semi-arid watershed. In Proceedings of the IEEE Topical Symposium on Combined Optical-Microwave Earth and Atmosphere Sensing, Albuquerque, NM, USA, 22–25 March 1993; pp. 86–89.

13. Njoku, E.G.; Jackson, T.J.; Lakshimi, V.; Chan, T.K.; Ngheim, S.V. Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 215–229.
14. Aritola, J.F.; Pepper, I.L.; Brusseau, M.L. *Environmental Monitoring and Characterization*; Elsevier Academic Press: San Diego, CA, USA, 2004.
15. Das, N.N.; Mohanty, B.P. Root zone soil moisture assessment using remote sensing and vadose zone modeling. *Vadose Zone J.* **2006**, *5*, 296–307.
16. Njoku, E.G.; Ashcroft, P.; Chan, T.K.; Li, L. Global survey and statistics of radio-frequency interference in AMSR-E land observations. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 938–947.
17. Crosson, W.L.; Limaye, A.S.; Laymon, C.A. Parameter sensitivity of soil moisture retrievals from airborne C- and X-band radiometer measurements in SMEX02. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2842–2853.
18. Crow, W.T.; Chan, T.K.; Entekhabi, D.; Houser, P.R.; Hsu, A.Y.; Jackson, T.J.; Njoku, E.G.; O'Neill, P.E.; Shi, J.; Zhan, X. An observing system simulation experiment for hydros radiometer-only soil moisture products. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1289–1303.
19. Fernandez-Galvez, J. Errors in soil moisture content estimates induced by uncertainties in the effective soil dielectric constant. *Int. J. Remote Sens.* **2008**, *29*, 3317–3323.
20. Cai, G.; Xue, Y.; Hu, Y.; Wang, Y.; Guo, J.; Luo, Y.; Wu, C.; Zhong, S.; Qi, S. Soil moisture retrieval from MODIS data in northern China plain using thermal inertia model. *Int. J. Remote Sens.* **2007**, *28*, 3567–3581.
21. Lu, S.; Ju, Z.; Ren, T.; Horton, R. A general approach to estimate soil water content from thermal inertia. *Agric. For. Meteorol.* **2009**, *149*, 1693–1698.
22. Verstraeten, W.W.; Veroustraete, F.; van der Sande, C.J.; Grootaers, I.; Feyen, J. Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. *Remote Sens. Environ.* **2006**, *101*, 299–314.
23. Gill, M.; Asefa, T.; Kemblowski, M.; McKee, M. Soil moisture prediction using support vector machines. *J. Am. Water Resour. Assoc.* **2006**, *42*, 1033–1046.
24. Jiang, H.; Cotton, W. Soil moisture estimation using an artificial neural network: A feasibility study. *Can. J. Remote Sens.* **2004**, *30*, 827–839.
25. Hassan-Esfahani, L.; Torres-Rua, A.; Jensen, A.; McKee, M. Topsoil moisture estimation for precision agriculture using unmanned aerial vehicle multispectral imagery. In Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 3263–3266.
26. Elshorbagy, A.; Parasuraman, K. On the relevance of using artificial neural networks for estimating soil moisture content. *J. Hydrol.* **2008**, *362*, 1–18.
27. Khalil, A.; Gill, M.K.; McKee, M. New applications for information fusion and soil moisture forecasting. In Proceedings of the 8th International Conference on Information Fusion, Philadelphia, PA, USA, 25–28 July 2005; pp. 1622–1628.
28. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; MacMillan: New York, NY, USA, 1999.
29. Khan, M.; Coulibaly, P. Streamflow forecasting with uncertainty estimate using Bayesian learning For ANN. In Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, Montreal, QC, Canada, 31 July–4 August 2005; Volume 5, pp. 2680–2685.

30. Tokar, S.A.; Johnson, P.A. Rainfall-runoff modeling using artificial neural networks. *J. Hydrol. Eng.* **1999**, *4*, 232–239.
31. Rumelhart, D.E.; McClelland, J.L.; PDP Research Group. *Parallel Distributed Processing. Volume 1: Foundations*; The MIT Press: Cambridge, MA, USA, 1986.
32. Hecht-Nielsen, R. *Neurocomputing*; Addison-Wesley Publishing Company: Reading, MA, USA, 1990.
33. Wasserman, P.D. *Neural Computing Theory and Practice*; Van Nostrand Reinhold: New York, NY, USA, 1989.
34. Gillies, R.R.; Carlson, T.N. Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *J. Appl. Meteorol.* **1995**, *34*, 745–756.
35. Wetzell, P.J.; Woodward, R.H. Soil moisture estimation using GOES-VISSR infrared data: A case study with a simple statistical method. *J. Clim. Appl. Meteorol.* **1987**, *26*, 107–117.
36. Kogan, F.N. Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. *Bull. Am. Meteorol. Soc.* **1995**, *76*, 655–668.
37. Mallick, K.; Bhattacharya, B.K.; Patel, N.K. Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI. *Agric. For. Meteorol.* **2009**, *149*, 1327–1342.
38. Liang, S.; Rui, S.; Xiaowen, Li.; Shunlin, L.; Renhua, Z. Monitoring surface soil moisture status based on remotely sensed surface temperature and vegetation index information. *Agric. For. Meteorol.* **2012**, *166*, 175–187.
39. Bhuiyan, C.; Singha, R.P.; Koganc, F.N. Monitoring drought dynamics in the Aravalli region (India) using different indices based on ground and remote sensing data. *Int. J. Appl. Earth Observ. Geoinf.* **2006**, *8*, 289–302.
40. Jensen, A.M. A Geospatial Real-Time Aerial Image Display for a Low-Cost Autonomous Multispectral Remote Sensing. Master's Thesis, Utah State University, Logan, UT, USA, 2009.
41. MosaicMill EnsoMOSAIC. Available online: <http://www.ensomosaic.com> (accessed on 6 May 2012).
42. Long, D.; Longuevergne, L.; Scanlon, B.R. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resour. Res.* **2014**, *50*, 1131–1151.
43. Decagon Devices, Inc. Available online: <http://www.decagon.com/> (accessed on 20 November 2014).
44. Israelson, O.W.; West, F.L. Water holding capacity of irrigated soils. *Utah State Agric. Exp. Station Bull* **1922**, *183*, 1–24.
45. Costa, A.; Albuquerque, J.A.; Costa, A.; Pértile, P.; Silva, F.R. Water retention and availability in soils of the State of Santa Catarina-Brazil: Effect of textural classes, soil classes and lithology. *Braz. Soil Sci. Soc.* **2013**, *37*, 1535–1548.
46. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Volume 12, pp. 1137–1143.
47. Geisser, S. *Predictive Inference*; Chapman and Hall: New York, NY, USA, 1993.
48. Guyon, I. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.

49. Kohavi, R.; John, G.H. Wrappers for feature subset selection. *Artif. Intell.* **1997**, *12*, 273–324.
50. Glover, D.M.; Jenkins, W.J.; Doney, S.C. *Least Squares and Regression Techniques, Goodness of Fit and Tests, Non-Linear Least Squares Techniques*; Woods Hole Oceanographic Institute: Woods Hole, MA, USA, 2008.
51. Bowden, G.J.; Maier, H.R.; Dandy, G.C. Optimal division of data for neural network models in water resources applications. *Water Resour. Res.* **2002**, doi:10.1029/2001WR000266.
52. Bowden, G.J.; Maier, H.R.; Dandy, G.C. Input determination for neural network models in water resources applications. Part 2. Case study: Forecasting salinity in a river. *J. Hydrol.* **2005**, *301*, 93–107.
53. Anctil, F.; Lauzon, N. Generalisation for neural networks through data sampling and training procedures, with applications to stream flow predictions. *Hydrol. Earth Syst. Sci.* **2004**, *8*, 940–958.
54. Jia, Y.; Culver, T.B. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *J. Hydrol.* **2006**, *331*, 580–590.
55. Lauzon, N.; Anctil, F.; Petrinovic, J. Characterization of soil moisture conditions at temporal scales from a few days to annual. *Hydrol. Process.* **2004**, *18*, 3235–3254.
56. Whalley, W.R.; Leeds-Harrison, P.B.; Bowman, G.E. Estimation of soil moisture status using near infrared reflectance. *Hydrol. Process.* **1991**, *5*, 321–327.
57. Hassan-Esfahani, L.; Torres-Rua, A.; McKee, M. Assessment of optimal irrigation water allocation for pressurized irrigation system using water balance approach, learning machines, and remotely sensed data. *Agric. Water Manage.* **2015**, *153*, 42–50.
58. Yin, Z.; Lei, T.; Yan, Q.; Chen, Z.; Dong, Y. A near-infrared reflectance sensor for soil surface moisture measurement. *Comput. Electron. Agric.* **2013**, *99*, 101–107.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).