



**US Army Corps  
of Engineers®**  
Engineer Research and  
Development Center



*ERDC Center-Directed Research Program*

## **Determining Soil Moisture at Moderately High Resolution via Soil Moisture Active Passive (SMAP) and Landsat Using Inferred Learning**

Michael G. Lewis, Andmorgan Fisher, and Clint Smith

September 2018



**The U.S. Army Engineer Research and Development Center (ERDC)** solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at [www.erdcl.usace.army.mil](http://www.erdcl.usace.army.mil).

To search for other technical reports published by ERDC, visit the ERDC online library at <http://acwc.sdp.sirsi.net/client/default>.

# **Determining Soil Moisture at Moderately High Resolution via Soil Moisture Active Passive (SMAP) and Landsat Using Inferred Learning**

Michael G. Lewis, Andmorgan Fisher, and Clint Smith

*Geospatial Research Laboratory (GRL)*  
*U.S. Army Engineer Research and Development Center (ERDC)*  
*7701 Telegraph Road*  
*Alexandria, VA 22315-3864*

Final report

Approved for public release; distribution is unlimited.

Prepared for U.S. Army Corps of Engineers  
Washington, DC 20314-1000

Under ERDC Center-Directed Research Project, "Army Terrestrial Environmental Modeling and Intelligence System (ARTEMIS)"

## Abstract

This paper reviews an approach to increase soil moisture resolution over a sample region over Australia using the Soil Moisture Active Passive (SMAP) sensor and Landsat 8 only. This approach uses an inductive localized approach, eschewing the need to obtain a resilient and definitive model in favor of a temporary model that utilizes current conditional inputs only. For the purposes of this analysis, the SMAP 36 km soil moisture product is considered fully valid and accurate. Landsat bands coinciding in collection date with a SMAP capture are down sampled to match the resolution of the SMAP product. A series of indices describing the Soil-Vegetation-Atmosphere Triangle (SVAT) relationship are then produced using the down sampled Landsat bands. These indices are then related to the local coincident SMAP values to identify a series of rules or trees to identify the local rules defining the relationship between soil moisture and the aforementioned indices. This paper uses a random forest due to its highly accurate learning against local ground truth data. The defined rules are then applied to the Landsat image in the native Landsat resolution to determine local soil moisture. Ground truth comparison is done via a series of grids using time dielectric impedance and airborne L-band Multi-beam Radiometer (PLMR) observations done under the SMAPEX-5 campaign (Panciera 2013). The predictive power of the inferred learning soil moisture algorithm (ILSMA) did well with a mean absolute error of 0.054 over an airborne L-band retrieved surface over the same region.

**DISCLAIMER:** The contents of this report are not to be used for advertising, publication, or promotional purposes. Citation of trade names does not constitute an official endorsement or approval of the use of such commercial products. All product names and trademarks cited are the property of their respective owners. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

**DESTROY THIS REPORT WHEN NO LONGER NEEDED. DO NOT RETURN IT TO THE ORIGINATOR.**

# Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Figures and Tables</b> .....	<b>iv</b>
<b>Preface</b> .....	<b>vi</b>
<b>Unit Conversion Factors</b> .....	<b>vii</b>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background.....	1
1.2 History .....	1
1.3 Purpose.....	2
<b>2 Data</b> .....	<b>3</b>
2.1 Optical .....	3
2.2 Microwave .....	5
2.3 Integration between microwave and VIS/NIR .....	6
<b>3 Methods</b> .....	<b>8</b>
<b>4 Results</b> .....	<b>18</b>
<b>5 Discussion</b> .....	<b>28</b>
<b>References</b> .....	<b>30</b>
<b>Report Documentation Page</b>	

# Figures and Tables

## Figures

Figure 1. Correlation of soil moisture to Hyperion spectral bands meeting the requirements of not being strongly contaminated by atmospheric absorption (1350–1450 nm, 1800–1900 nm, and 2400–2500 nm), nor high vegetation influence compared to soil (750–1300 nm), and responsive to soil moisture. (Wu et al. 2010). .....	3
Figure 2. Impact of vegetative cover on soil moisture retrieval. (Wu et al. 2010).....	4
Figure 3. Triangle method scatterplot (Carlson 2007).....	9
Figure 4 Selected SMAP pixels over the three joined Landsat scenes used to train the algorithm. ....	11
Figure 5. Process flowchart of localized downscaling process via the Inferred Learning Soil Moisture Algorithm (ILSMA).....	14
Figure 6. Scatter plot of training effectiveness using the SMAP values as the independent variable and the down sampled Landsat imagery input into the aforementioned variables as dependent variables, the red lines demarcate a 0.04 variance off of a perfect prediction.....	14
Figure 7. Field dielectric impedance measurement grids, in which dielectric impedance measurements were conducted in triplicate within the field of the airborne radiometers collection. The radiometer is within the SMAP sampling and Landsat aggregation region (see upper right inset).....	15
Figure 8. The soil moisture output from the localized soil moisture downscaling process. ....	17
Figure 9. Calculated ground surface temperature from Landsat 8, band 10. ....	20
Figure 10. The NDVI from Landsat 8, bands 5, and 6. The three sites of exhaustive dielectric impedance measurement denoted by the red and black clusters.....	21
Figure 11. Initial radiometer data showing a strong banding artifact in the direction of the flight path. ....	22
Figure 12. The six beam bands illustrated over the area of interest.....	23
Figure 13. Normalized L-band return adjusted for variation in the individual beams along flight path. ....	24
Figure 14. The Kriged surface of the L-band radiometer data. ....	25
Figure 15. A visual comparison of the L-band radiometer and the Landsat inferred soil moisture shows generally good agreement overall with areas of notable exception, such as the southern edge, just west of center. ....	26
Figure 16. Qualitative review of error between the L-band retrieval and the Landsat-SMAP downscaling algorithm. ....	27
Figure 17. The number of dielectric impedance samples on the x-axis compared with the error from the Landsat inferred soil moisture value suggests that dielectric impedance values are prone to high error in terms of the average value of the pixel when less than nine samples are obtained over the 30m <sup>2</sup> pixel area. ....	28

## Tables

Table 1. Cross correlation table between variables and the L-band retrieval as well as the ILSMA predicted value showing the covariance between variables and with the ground truth and ILSMA predicted value. Correlation coefficients have all been rendered in terms of absolute value. ....	10
Table 2. Mean and Standard Deviation of each individual flight path and the full population. ....	23
Table 3. The mean of the proposed algorithm is 0.26 while the mean of the L-band retrieval is 0.28; the mean absolute error (MAE) is 0.05. ....	26
Table 4. Magnitude of error within the algorithm based on the L-band surface. Currently 56.44% of the result meets NASA specifications of 0.04 providing the L-band retrieval was accurate. ....	27
Table 5. MAE values of data types from on-another over the naïve test region. ....	29

## Preface

This study was conducted for the Engineer Research and Development Center's Center-Directed Research Program under "Army Terrestrial Environmental Modeling and Intelligence System (ARTEMIS)." The technical monitor was Dr. John Eylander, U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory (ERDC-CRREL).

The development of this report was led by the Data Signature and Analysis Branch (DSAB) of the ERDC-Geospatial Research Laboratory (GRL). At the time of publication Dr. Michael G. Lewis was the branch chief, Ms. Martha Kiene was the division chief. Ms. Valerie Carney was the Deputy Director, and was Mr. Gary Blohm was the Director of ERDC-GRL.

COL Ivan P. Beckman was Commander of ERDC, and Dr. David W. Pittman was the Director.

## Unit Conversion Factors

Multiply	By	To Obtain
inches	0.0254	meters
Joules	0.000239006	kilocalorie
gigahertz	$1.0 \times 10^9$	hertz
nanometers	$1.0 \times 10^{-9}$	meters
watts	1	joule/sec

# 1 Introduction

## 1.1 Background

Soil moisture is a measurement of the hydrological component within a finite amount of soil. The variability of soil moisture is highly dependent on the soil properties (Cosby et al. 1984), the local geologic situation (Weizu and Freer 1995), vegetation density and draw (Denmead and Shaw 1962), and antecedent conditions. Soil moisture is a vital key to understanding the physiologic condition of the earth for many systems. Agriculture relies on soil moisture for plant vigor and growth, providing root zone moisture and a direct relationship to CO<sub>2</sub> respiration by soil microbial activity (Orchard and Cook 1983). The moisture itself acts as a primary nutrient to growing crop and plant life, carrying with it the organic and trace nutrients necessary for plant growth. Weather and climate models are strongly coupled to land-atmosphere interactions (Koster et al. 2004). Flood control is highly dependent of antecedent soil moisture conditions (De Michele and Salvadori 2002). According to the National Aeronautics and Space Administration (NASA), “reservoir management, early warning of droughts, irrigation scheduling, and crop yield forecasting” can all be influenced by soil moisture conditions (Soil Moisture 1999). Slope failure and mass movement can be linked to soil plasticity induced by elevated moisture content creating hazards to communities and travel (Crosta 1998). From a tactical standpoint, soil moisture affects trafficability in route planning, both in a mounted and dismounted mode (Shoop 1993). Road pavement can fail due to improper design or anomalous soil moisture content for the road construction design (Ahmed et al. 2012).

## 1.2 History

In situ soil moisture measurement first originated in a methodical and accurate manner in 1949. Following World War II, George Pieper described a method to utilize a neutron probe to measure moisture content (Evelt 2000). This technology was soon the primary means of soil moisture monitoring; however, it required site calibration and posed radiation hazards which were undesirable (Topp et al. 1980). Using the electrical properties of soil, a dielectric methodology was described by Topp et al. (1980). This permitted the use of unattended, automated collection sites (Ochsner et al. 2013). In roughly this same timeframe,

remotely sensed soil moisture research was occurring in multiple fronts including microwave radiometers, scatterometers, synthetic aperture RADAR (SAR), and combined RADAR-radiometer systems (Ochsner et al. 2013). The negative to using such frequencies is the requirement for a large antenna for even very low resolution. Using a classical orbiting arrangement without any novel approach, a 50 km resolution would require a 20 m antenna (Monerris and Schmugge, 2009). This has been the primary area of investigation into soil moisture, primarily due to soil moisture sensitivity but also due to the size of the wavelength (the resolution continues to remain rather coarse). The soil response to electromagnetic probing is based on the dielectric constant of the target. It is complex and has many models of description. The dielectric constant of dry soils exhibit properties independent of temperature and frequency unlike wet soils, which seem to exhibit properties based on complex interactions between the soil, air, and moisture fractions (Topp et al. 1980) (Monerris and Schmugge 2009). Further confounding interpretation of microwave based remote sensing is the response to variations of surface roughness. The response however is not constant, but variable based on moisture content and soil type (Panciera et al. 2009), with clay soils demonstrating noticeably more aberrant response from other types (Monerris and Schmugge 2009).

The Soil Moisture Active Passive (SMAP) mission, launched in 2015, attempted to produce a higher resolution product with a resolution better than 10 km. This is accomplished through the integration of the highly accurate and low resolution passive radiometer with higher resolution active microwave, which is prone to vegetation and surface roughness effects (Wu et al. 2015). Unfortunately, SMAP suffered an irrecoverable system failure in the active microwave portion of the sensor, leading to the passive radiometer being the only usable sensor upon the instrument.

### **1.3 Purpose**

While beneficial scientific investigation has been pursued in algorithms to link the SMAP sensor to ground condition soil moisture, the resolution issue still remains. SMAP has a functional received product at 36 km resolution, a very coarse pixel size that is problematic at the local scale. Below is proposed a method that may begin to accommodate that shortcoming via an inferred learning soil moisture algorithm (ILSMA), building on the work that has been completed in validating the SMAP soil moisture product.

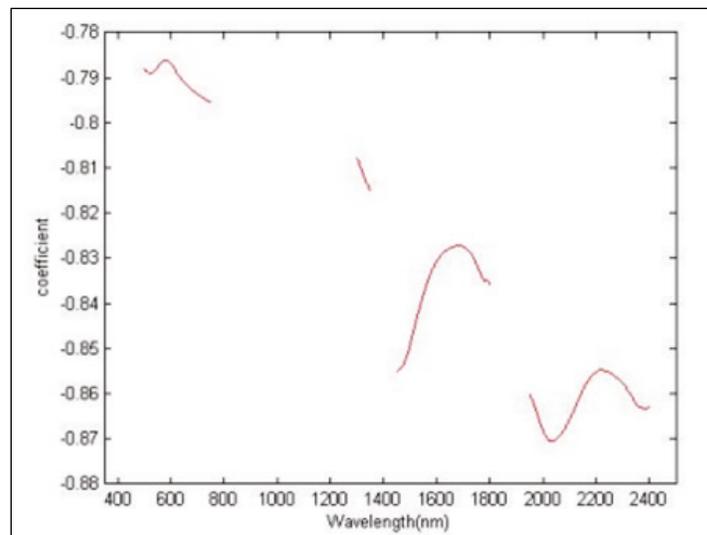
## 2 Data

### 2.1 Optical

Earth Observing-1 (EO-1), launched 21 November 2000, was a part of the NASA New Millennium Program designed to develop breakthrough technologies for the next generation of remote sensing systems. The original planned mission duration of 1 year was greatly exceeded with the primary sensors retrieving and relaying data until being decommissioned on 30 March, 2017. EO-1 trails the Landsat 7 imager by one minute for calibration and analysis purposes, putting the system on a 16 day repeat cycle on a sun-synchronous polar orbit at 705 km at the Equator (NASA 2011). Included aboard EO-1 is the hyperspectral imager Hyperion, and the Advanced Land Imager (ALI).

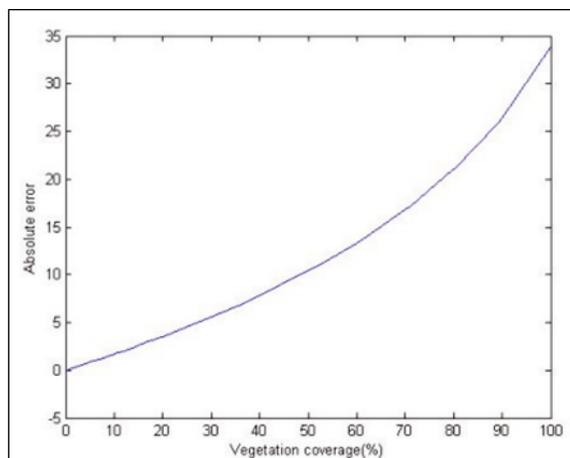
In a study of pure correlation metrics to soil moisture content, Hyperion band reduction best performance proved to be at 2036 nm (2.036  $\mu\text{m}$ ), well within the short wave infrared region, with a correlation coefficient of 0.8706 (Wu et al. 2010) (Figure 1). Short-wave infrared (SWIR) has the advantage of being able to penetrate haze and smoke, although thick cloud cover is still opaque.

Figure 1. Correlation of soil moisture to Hyperion spectral bands meeting the requirements of not being strongly contaminated by atmospheric absorption (1350–1450 nm, 1800–1900 nm, and 2400–2500 nm), nor high vegetation influence compared to soil (750–1300 nm), and responsive to soil moisture. (Wu et al. 2010).



The Wu et al. (2010) study found that the degree of vegetative cover affected the retrieval in an exponential manner (Figure 2) as the unmixing of the pixels began to increase in failure to accurately describe the coverage with increasing leaf area coverage.

Figure 2. Impact of vegetative cover on soil moisture retrieval. (Wu et al. 2010).



The ALI instrument on EO-1 carries bands similar to Landsat with a panchromatic of 10 m as opposed to the Landsat 15 m band, with multispectral bands of a 30 m resolution ranging from 433 nm to 2,350 nm (0.433  $\mu\text{m}$  to 2.35  $\mu\text{m}$ ). The high absorption water bands sit well inside the spectral limits of the sensor. While the coarser spectral resolution does not permit the degree of spectral signature slope analysis provided by its co-resident Hyperion sensor, the swath of 37 km provides a wider coverage. Spectral band coverage in the regions described according to Wu et al. (2010) is fit by bands 9 and 10, with band 8 providing additional discernment above a 0.80 coefficient.

Landsat 8, on an identical path and revisit time, provides only two bands meeting the requirements set forth by Wu et al. (2010), Bands 6 and 7, 1.57–1.65  $\mu\text{m}$  and 2.11–2.29  $\mu\text{m}$ , respectively. However, Landsat provides an even greater coverage and regular non-directed revisits every 16 days, providing a continuous bi-monthly record globally with open access. Local modifying conditions such as soil type, surface thermal radiance, and vegetation density are well described by Landsat in addition to possessing two key bands highly bound to soil moisture content. Furthermore, Landsat provides a continuous 30+ year record of data and will be carried forward through the foreseeable future through the Landsat Data

Continuity Mission (LDCM) and the announcement of Landsat 9. This analysis used three Landsat 8 scenes acquired on 14 September 2015.

## 2.2 Microwave

Remote sensing of soil moisture is most sensitive in the electromagnetic spectrum in the range of 1 to 5 GHz due to the maximum stretching of the dielectric effect of dry soil to water. In the range of 1.4–1.427 GHz, the L-band is not only the signal most sensitive to soil moisture, but the atmosphere is nearly transparent and vegetation is semi-transparent allowing nearly universal viewing at the soil surface (Monerris and Schmugge 2009). The negative to using such frequencies is the requirement for a large antenna for even very low resolution; using a classical orbiting arrangement without any novel approach, a 50 km resolution would require a 20 m antenna (Monerris and Schmugge 2009). Further confounding interpretation of microwave based remote sensing is the response to variations of surface roughness. The response however is not constant, but variable based on moisture content and soil type (Panciera et al. 2009), with clay soils demonstrating noticeably more aberrant response from other types (Monerris and Schmugge 2009).

This has been the primary area of investigation into soil moisture due to the soil moisture sensitivity in the L-band however, due to the wavelength size, resolution will remain rather coarse. The Soil Moisture and Ocean Salinity (SMOS) mission was the first soil moisture dedicated mission. Launched in 2009, the mission has a return period of three days and a pixel size of roughly 40km (Merlin et al. 2008), based on latitude. The SMAP mission launched in 2015 attempted to produce a higher resolution product (with a resolution better than 10 km), through the integration of the highly accurate and low resolution passive radiometer with higher resolution active microwave, which is prone to vegetation and surface roughness effects (Wu et al. 2015). Unfortunately, SMAP suffered an irrecoverable system failure in the active microwave portion of the sensor leading to the passive radiometer being the only usable sensor upon the instrument. This analysis used a subsample of SMAP and airborne L-band (PLMR) data with a center frequency of 1.413 GHz, and a band width of 24 MHz, acquired on 13 September 2015 as part of the SMAPEX-5 campaign.

## 2.3 Integration between microwave and VIS/NIR

In 2014, Song and Massimo (2014) proposed a hybrid methodology for obtaining soil moisture at a 1 km resolution in which microwave soil moisture brightness data is downscaled through the use of land surface temperature and NDVI data. While the method does appear to hold some validity, additional research should be conducted with increased ground truth to support and ensure the results are not spurious, owing the collinear relationship between NDVI and soil moisture. Furthermore, the purported reasoning of use in lieu of other methods due to the lack of global high resolution data doesn't seem mitigated due to the necessity of this method, also requiring that same high resolution data to produce the NDVI layer. However, it is a worthy alternative due to the relative lack of subjectivity over SVAT methodologies. The key to this method is the downscaling of microwave brightness using the exponential relationship between the microwave polarization difference index (MPDI) and NDVI:

$$NDVI = E_0 + E_1 * \exp(E_2 * MPDI) \quad (1)$$

where  $E_0$ ,  $E_1$ , and  $E_2$  are constants. The necessity for downscaling microwave data is due to the resolution being limited by the antennae length of onboard sensors. This is due to the dependent system parameter of beam width which is determined by the wavelength over the antennae length. Microwave satellite sensors would be far too large to gather high resolution data using current technology or budgetary constraints.

Combining this equation with the MPDI, a standard difference index between the vertical and horizontal polarizations of either 18.7 or 36.5 GHz to mitigate the influence of soil moisture in the temperature analysis, brings the equation:

$$T_{18.7H} = \frac{1 - F_1 * \ln(F_2 * (NDVI - F_0))}{1 + F_1 * \ln(F_2 * (NDVI - F_0))} * T_{18.7V} \quad (2)$$

In which the constants can be solved when upscaling the NDVI to the same resolution as the original microwave data. Once the constants and NDVI are known, the downscaled horizontal polarization can be determined. Although not explicitly explained, it appears this can be done as follows:

If:

$$B = \frac{1 - F_1 * \ln(F_2 * (NDVI - F_0))}{1 + F_1 * \ln(F_2 * (NDVI - F_0))} \quad (3)$$

Then

$$A_i = \frac{B_i - D_0 + B_i D_1 T_s}{2A_1 - B_i D_2}; C = D_0 + D_1 T_s + D_2 A_i \quad (4)$$

where  $D_0$ ,  $D_1$ ,  $D_2$ , and  $F_0$ ,  $F_1$ , and  $F_2$  are constants that can be determined, although it appears to do so, at least three pixels would need to have a known surface soil moisture. Song and Massimo (2014) showed a distinct link between an earth observing sensor (EO) and a microwave L-band sensor, demonstrating the EO sensors can be successfully used as a dependent variable set against an independent variable soil moisture set to identify a higher resolution result.

However, the use of machine learning can bypass much of the localized uncertainty due to seasonal and optical vegetation depth as well as reduce computation time significantly by establishing a series of variables that describe the relationship between imagery data and the SMAP soil moisture values and apply those rules over a single scene or local set of scenes in a single collect. Furthermore, machine learning can incorporate and identify soil moisture via optical and infrared easily in the same algorithm, not only using the earth observing sensors as a proxy for microwave, but gaining valuable additional skill in determining true soil moisture distribution.

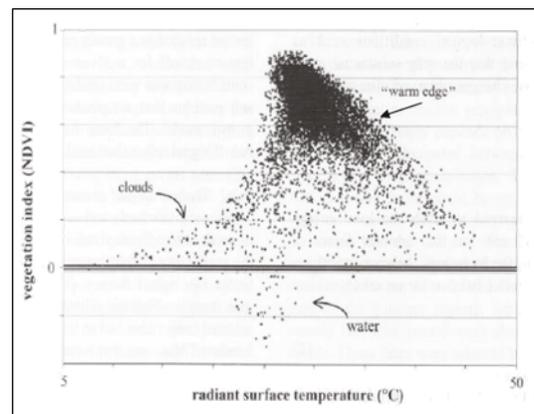
### 3 Methods

Beginning with three consecutive Landsat scenes over a region of interest, digital number (DN) values are extracted for each pixel over a sample set over the region to conduct relationship testing and pairing of variables. As the algorithm reevaluates a new random forest for each evaluation (scene and date time) and all variables are in terms of ratios, the DN values is appropriate in an uncorrected form. Variables are generated through the iterative ratio of each band with every other band. In the following iteration, the ratio is conducted with the initial bands as well as the new ratios, continuing through four iterations. Working exclusively with ratios largely nullifies atmospheric effects due to each band value being affected in a comparative fashion and all values being in a single set of scenes taking over a similar region on the same day. Since ratios look at total response in comparison to another total response, no other manipulation (such as atmospheric correction) is required. The sampled values are then compared via a simple correlation and those ratios performing best are brought forward for further scrutiny. This was done completely naively using multiple high resolution soil moisture surfaces over the United States and Australia, inductively learning the highest information variables in accommodating the soil moisture surface. Many initial variables were chosen in addition to the NDVI. Variable's neighborhood association was evaluated to understand what key factors appeared to be the best predicting variable types. Test of prediction skill, multicollinearity, and leave one out variable testing provided a base set of eight new variables to be utilized in three neighborhoods.

Neighborhood 1, being clearly grounded in the thermal range is attempting to identify moisture via temperature variation in the soil. Neighborhood 2 appears to be tightly bound directly to detecting moisture while neighborhood 3 is more aligned with using a vegetation proxy to identify moisture and therefore is associated with plant growth. These neighborhoods of variables are representative proxies for the key components of the Soil-Vegetation-Atmosphere Transfer (SVAT) "triangle" model. The triangle method is independent of readings from a site outside those that can be obtained from the red, NIR, and thermal bandwidths; however, it requires multiple data points to establish the triangle of the method.

The triangle is a scatterplot of NDVI plotted against thermal surface data (Figure 3). The assumption is that given a sufficiently large sampling, the pixels will exhibit a sufficiently large enough variance to plot the full range of possible scenarios from wet to dry and bare earth to fully vegetated (Carlson 2007). This scatter plot will form a triangle that delineates a shape “warm edge” where surface temperature is higher than other pixels similarly vegetated and indicating a lack of available water.

Figure 3. Triangle method scatterplot (Carlson 2007).



The assumption made is that vegetation temperature does not vary in space and that surface temperature alone accounts for the variation seen, at least within the margin of error for detection (Carlson 2007), which is somewhat in disagreement with work done earlier by Thomas Jackson of the USDA on the moisture availability of foliage and temperature; however, the leaf area sourced radiation is minor in comparison to the ground temperature effect. The “cool edge,” somewhat less defined, represents pixels with adequate or excess moisture for cooling. The lower surface trails off into points which can be accounted for by surficial anomalies such as standing bodies of water and clouds, and should be removed to create a cluster in tight arrangement.

Unfortunately, this model requires some subjectivity in defining the triangle edges, which poses a problem for cross site and cross experiment comparisons. A sufficiently large number of instances are also required, assuming soil characteristics remain overall rather uniform. Despite the apparent subjectivity of the method, Carlson (2007) states the errors in estimating the evapotranspiration function (EF), and thus available soil moisture, from such a method to be typically  $\pm 0.1-0.2$  (Carlson 2007), near the theoretical minimum.

Supplanting created variables with established variables such as the Modified Normalized Difference Water Index (MNDWI) and the Normalized Soil Moisture Index (NMSI) demonstrated better performance over other neighborhood 2 variables while neighborhoods 1 (thermal) was best served by two new variables referred to as variable 2 and variable 10. Neighborhood 3 is well served by the NDVI alone. A third variable, variable 16, has also shown promise in discriminating soil moisture and is brought forward based on its ability to identify strongly with soil moisture in past analysis.

The final variable set for Landsat 8 is:

$$\begin{aligned} \text{Var10} &: (\text{Band8}/\text{Band11})/(\text{Band9}/\text{Band10}) \\ \text{Var2} &: \text{BANDand1}/\text{BANDand10} \\ \text{Var16} &: (\text{BAND4} * \text{BAND7}^3)/(\text{BAND5} * \text{BAND6}^2) \\ \text{NDVI} &: (\text{BAND5}-\text{BAND4})/(\text{BAND5}+\text{BAND4}) \\ \text{NMSI} &: (\text{BAND7}-\text{BAND6})/(\text{BAND7}+\text{BAND6}) \\ \text{MNDWI} &: (\text{BAND3}-\text{BAND6})/(\text{BAND3}+\text{BAND6}) \end{aligned}$$

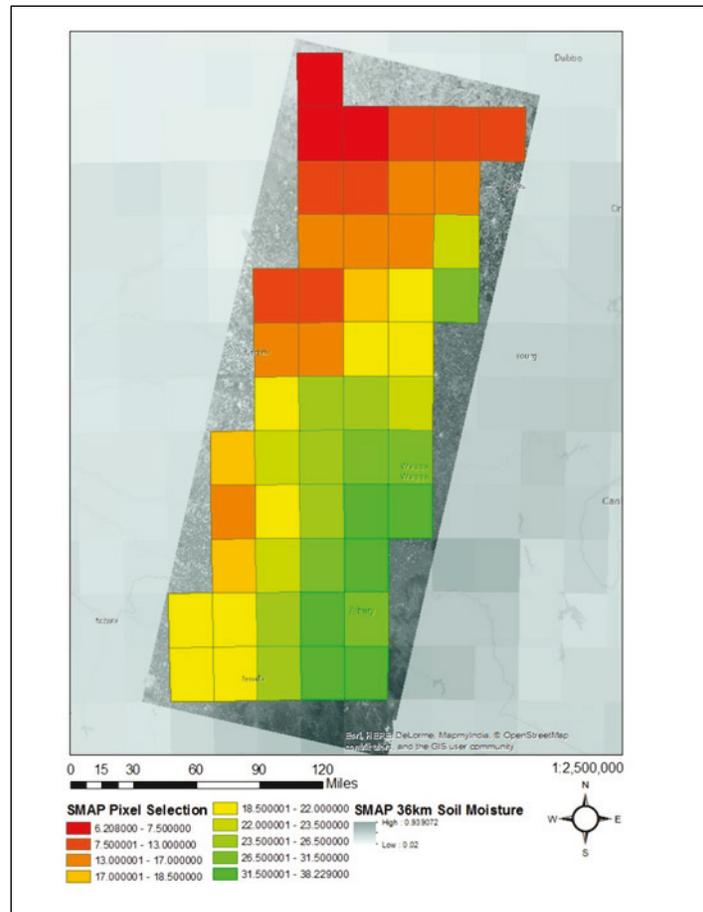
In the dielectric impedance measured subsamples of the entire population, variable 10 is poorly correlated with NDVI at -0.53, NMSI at -0.48, and MNDWI at -0.21 but well correlated with L-band soil moisture readings at 0.89. Variable 2, which can be described as an image adjusted thermal band is less correlated to soil moisture at 0.73 while the next highest correlation to Variable 2 is Variable 10 at 0.34. Variable 16 is highly inversely correlated to soil moisture, NDVI and NMSI at -0.82, -0.89, and -0.92 respectively. Broadening to the entire population of L-band measured samples, Correlations between variables drop considerably (Table 1).

**Table 1. Cross correlation table between variables and the L-band retrieval as well as the ILSMA predicted value showing the covariance between variables and with the ground truth and ILSMA predicted value. Correlation coefficients have all been rendered in terms of absolute value.**

	Variable 10	Variable 2	Variable 16	NDVI	NMSI	MNDWI	L-Band	ILSMA
Variable 10		0.377	0.609	0.501	0.466	0.213	0.234	0.343
Variable 2	0.377		0.13	0.084	0.03	0.224	0.301	0.236
Variable 16	0.609	0.13		0.875	0.914	0.425	0.408	0.426
NDVI	0.501	0.084	0.875		0.93	0.22	0.326	0.277
NMSI	0.466	0.03	0.914	0.93		0.547	0.38	0.399
MNDWI	0.213	0.224	0.425	0.22	0.547		0.269	0.454
L-Band	0.234	0.301	0.408	0.326	0.38	0.269		0.579
ILSMA	0.343	0.236	0.426	0.277	0.399	0.454	0.579	

Now that the variables are identified, the obvious issue is the dependent variable since ample ground truth data is rarely available for the date, time, and location required. Here the SMAP values (Figure 4) can provide a local training set, the independent variable to the machine learning algorithm.

Figure 4. Selected SMAP pixels over the three joined Landsat scenes used to train the algorithm.



The majority of machine learning is, at its core, a classification exercise, often of complex relationships with incomplete information. The applicability of machine learning is in the ability to make reasonably accurate classifications of expected outcomes without attempting to encapsulate and explain each outlying error. Instead of being the end goal, the machine learned model is the initialization of the exploration of causality. When a sufficiently strong predictive power or classification is discovered, then the researcher has a point from which to begin describing the interaction and proxy effects causing the relationship. As such, the model should be kept as simple as possible using as few initialization

parameters as possible to attain the desired outcome, this will help minimize generalization errors (Cannon and Whitfield 2002). While a multilayer perceptron may be appropriate to provide an ongoing predictive application, the researcher should continue beyond such an implementation to discover the causal relationships driving those connections. Some models provide a considerably more intuitive construct from which to begin constructing further research understanding the underlying principles.

A more intuitive model that performs often as well, if not better than an artificial neural network (ANN) such as the multilayer perceptron, is the tree regression model. These models are newer than the artificial neural network and unlike the black box response seen from ANN, the method of decomposition is easily understood in the model formulation (Onyari and Ilunga 2013). The tree regression model is clearly read and denotes by its very structure, the import of variables and their inherent weights. An exploration of variable importance down the length of the tree, exploring possible proxy and spurious relationships should foster a movement from vague approximation of the import of variables to the full numerical model describing the system (i.e., data to knowledge).

The model tree iteratively splits a training set of variables in order to minimize regression error until the deviation is only a small fraction of a standard deviation of the original instance, at which point the splitting process ceases and the model is pruned back one node. At the new tree branch termination, the leaf, a regression model is constructed to describe the data reaching that leaf. This construct creates a non-linear piecewise function that describes the data. Model trees deliver better compactness and prediction accuracy in comparison to classical regression trees (Deepa et al. 2010). However, a fault often seen in model trees is the tendency to over fit (Hastie et al. 2008). In order to avoid such problems, an ensemble method of generalization is used. The random forest corrects for this by assimilating multiple decision trees based on selections of datapoints, building multiple decision trees, and using a voting model based on aggregated results, a method similar to bootstrap aggregating. Machine learning algorithms can be inherently unstable with small changes in the training data producing vastly different models. Bootstrap aggregation works by resampling with replacement on the training data multiple times and then averaging the mean of the member models (Cannon and Whitfield 2002). This averaging process effectively controls model

variance, error due to instability of the model and limited training data, without increasing overall bias (Breiman 1996). In this lies the strength of the random forest algorithm. Noted imitations of random forests are the inability of regression prediction beyond the range of the training data.

This algorithm (see Figure 5) is a form of transfer learning in which rules are learned at the native SMAP resolution using coarsened Landsat data and then applied at the native Landsat resolution to achieve a higher resolution product. The Landsat values must be aggregated up to the SMAP pixel size to determine the localized ruleset, in this case a random forest due to its superior performance in local testing. However regression, an ANN, or other inductive method may be applied as well. This is an agnostic step and the choice of random forest was purely based on performance results and the afore mentioned benefits to avoid over fitting. Random Forest are excellent at classification tasks (see Figure 6) which essentially is what downscaling is, due to the sample being in the same space and time as the requested output however it's continuous nature is described in terms of a regression. Doing so allows the user to identify the specific ruleset, in this case a set of decision trees, applicable to the region of inquiry. Those rules are then utilized at the Landsat native resolution of 30 m to provide the down-sampled soil moisture surface. The rules are variable across regions and not transferable across Landsat scene sets due to changes in the land surface model and the ability of the spectral and thermal data to fully encompass the full range of periodic and localized events. The term "scene set" refers to a single scene as well as its previous and following scene in a single path. However, the training data provides a strong localized approximation of the local soil moisture condition.

A total of four datasets were utilized in the training, testing, and validation of this method. The SMAP radiometer data was used as the source of soil moisture at low resolution; it is the source we wish to downscale. Landsat 8 imagery was acquired the same day and is the source of the variables in which the SMAP data is to be downscaled. The dielectric impedance and aircraft acquired PLMR are the validation ground truth measurements used to compare our results to high fidelity external measurements (see Figure 7).

Figure 5. Process flowchart of localized downscaling process via the Inferred Learning Soil Moisture Algorithm (ILSMA).

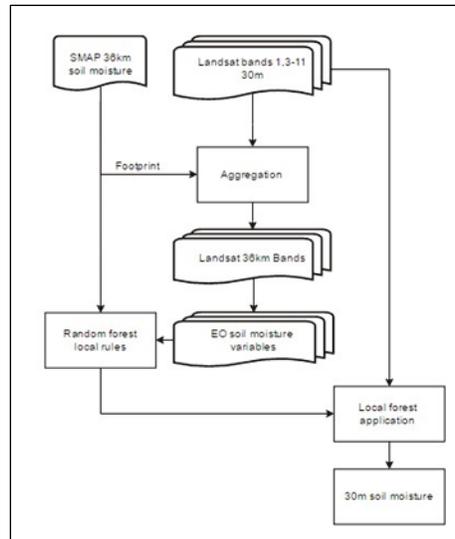


Figure 6. Scatter plot of training effectiveness using the SMAP values as the independent variable and the down sampled Landsat imagery input into the aforementioned variables as dependent variables, the red lines demarcate a 0.04 variance off of a perfect prediction.

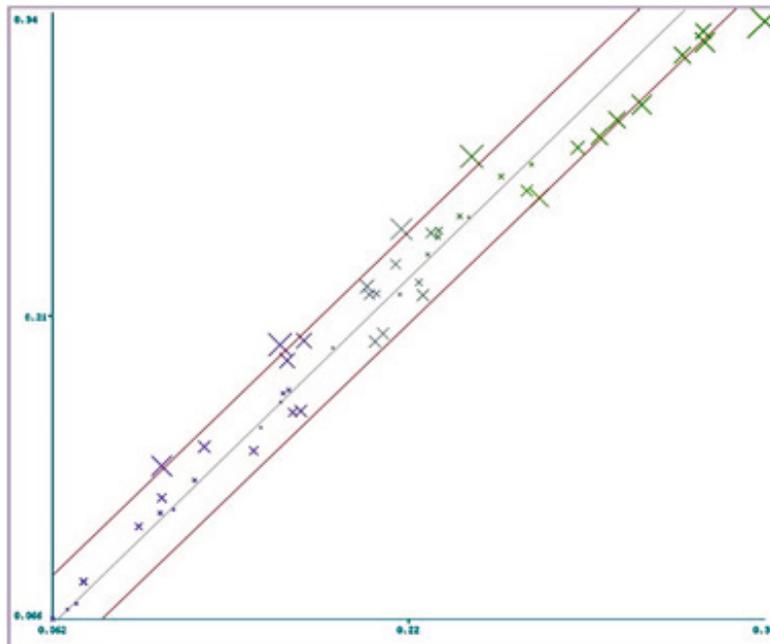
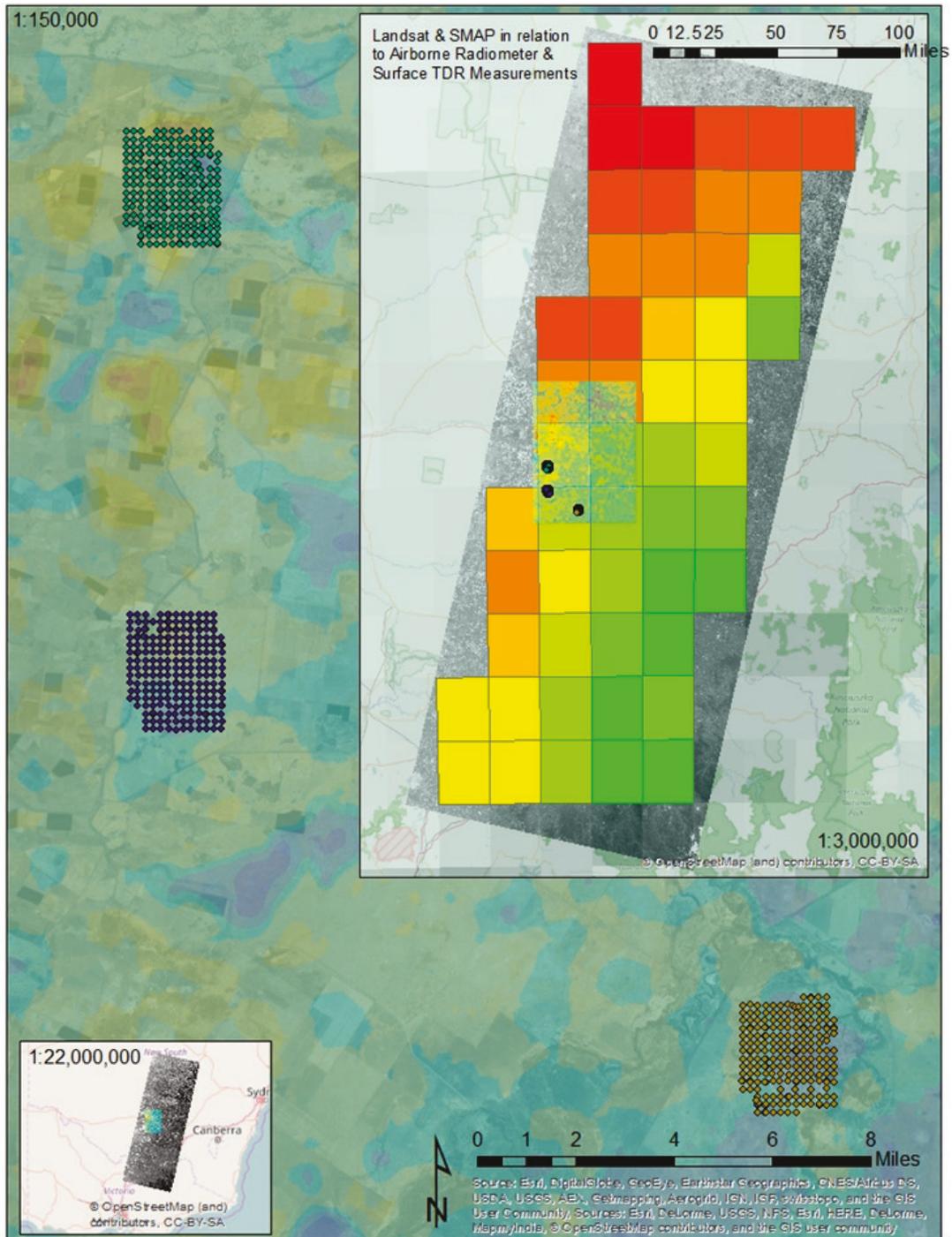
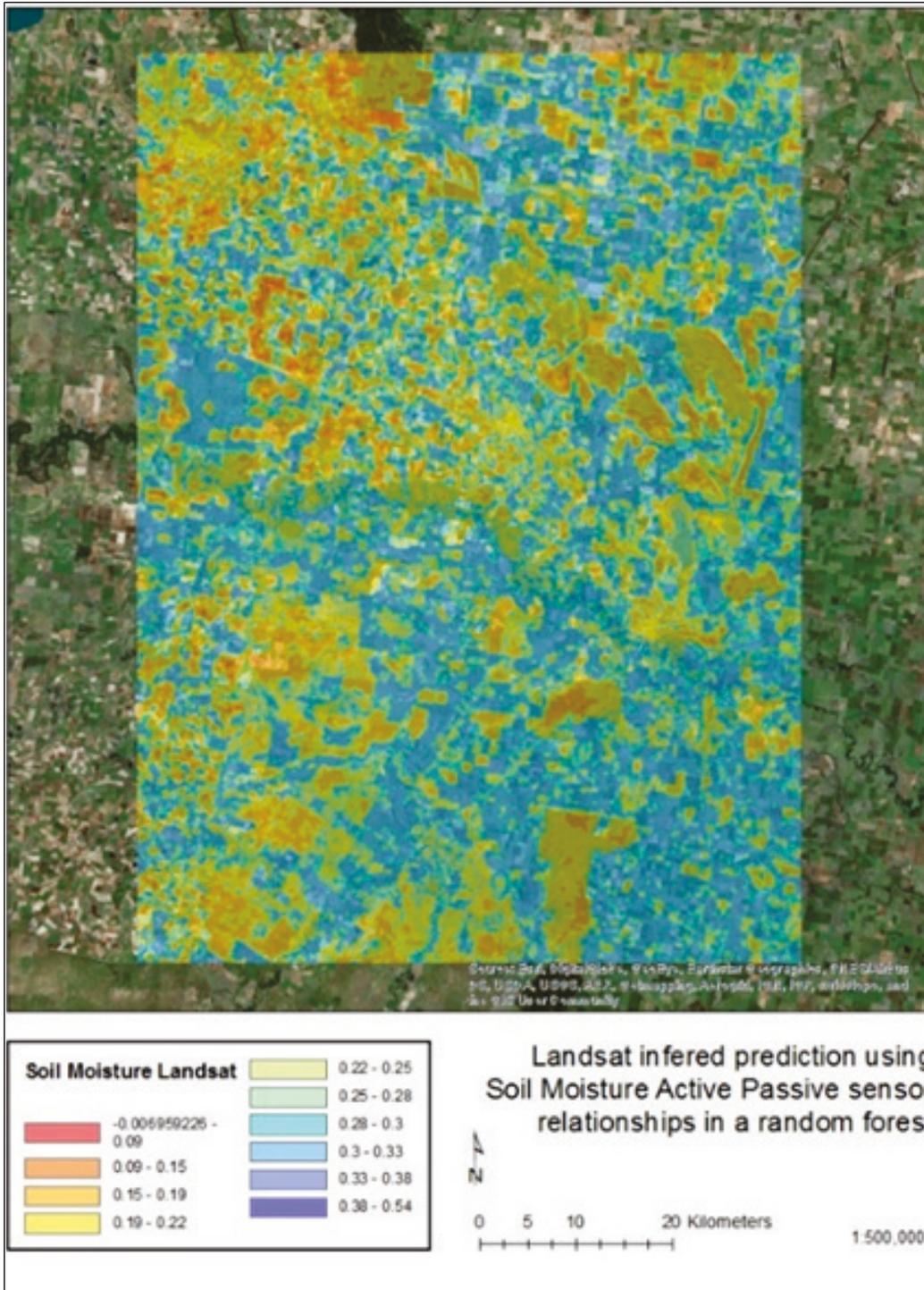


Figure 7. Field dielectric impedance measurement grids, in which dielectric impedance measurements were conducted in triplicate within the field of the airborne radiometers collection. The radiometer is within the SMAP sampling and Landsat aggregation region (see upper right inset).



The final output of the algorithm can be seen in Figure 8. Noticeable is the wetting of agricultural fields while natural woodlands and shrub regions remain dryer. This is likely due to the reliance on irrigation in the region but analysis against the PLMR and dielectric impedance datasets will provide accuracy measurements.

Figure 8. The soil moisture output from the localized soil moisture downscaling process.



## 4 Results

Ground truth over the area is available via an excellent dataset of time domain reflectometry; however, the coverage of this dataset is minimal compared to the entire range of the tested region. An airborne L-band radiometer (PLMR) also took readings on the same day and will be the primary source used to validate the localized downscaling method. Utilizing the PLMR microwave radiometer flown over the site requires conversion of the radiometric brightness to the dielectric constant. This can be done via the Fresnel reflection equations (Ulaby et al. 1986) using the formulas:

$$R_0^H = \left| \frac{\cos\theta - (\epsilon_r - \sin^2\theta)^{\frac{1}{2}}}{\cos\theta + (\epsilon_r - \sin^2\theta)^{\frac{1}{2}}} \right|^2 ; R_0^V = \left| \frac{\epsilon_r \cos\theta - (\epsilon_r - \sin^2\theta)^{\frac{1}{2}}}{\epsilon_r \cos\theta + (\epsilon_r - \sin^2\theta)^{\frac{1}{2}}} \right|^2 \quad (6) \ \& \ (7)$$

However, because a relationship is known to exist via the microwave polarization difference index between the polarizations (H and V), and NDVI via the Microwave Polarization Index:

$$\text{MPDI} = (T_V - T_H) / (T_V + T_H) \quad (8)$$

$$\text{NDVI} = E_0 + E_1 * \exp(E_2 * \text{MPDI}) \quad (9)$$

and combined with surface temperature, and all data can be reported to the same resolution negating the need for rescaling, soil moisture can easily be identified using reported values in localized areas for training via the function:

$$f(SM) == (T_V, T_H, T_S, \text{NDVI}) \quad (10)$$

To find Land surface temperature, the DN value of Band 10 must be converted to spectral radiance. Band 10 is used over band 11 due to an ongoing investigation into stray light boosting reported temperatures from band 11 by up to 8 kelvin, while band 10 has reported an increase of half that value or less.

$$L_\lambda = M_L * Q_{\text{cal}} + A_L \quad (11)$$

Where:

- $L_\lambda$  = Spectral radiance (W/(m<sup>2</sup> \* sr \* μm))
- $M_L$  = Radiance multiplicative scaling factor for the band  
(RADIANCE\_MULT\_BAND\_n from the metadata).
- $A_L$  = Radiance additive scaling factor for the band  
(RADIANCE\_ADD\_BAND\_n from the metadata).
- $Q_{cal}$  = L1 pixel value in DN. (USGS 2016)

$$L_\lambda = 3.3420E-04 * Q_{cal} + 0.10000 \quad (12)$$

This can then be transferred to the Top of Atmosphere Brightness Temperature via the formula:

$$T = \frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)} \quad (13)$$

where:

- T = TOA Brightness Temperature, in Kelvin.
- $L_\lambda$  = Spectral radiance (Watts/(m<sup>2</sup> \* sr \* μm))
- $K_1$  = Thermal conversion constant for the band  
( $K_1\_CONSTANT\_BAND\_n$  from the metadata)
- $K_2$  = Thermal conversion constant for the band  
( $K_2\_CONSTANT\_BAND\_n$  from the metadata) (USGS 2016)

Land surface temperature (Figure 9) can be determined via the single window algorithm via the formula

$$T/1 + w * (T/p) * \ln(e) \quad (14)$$

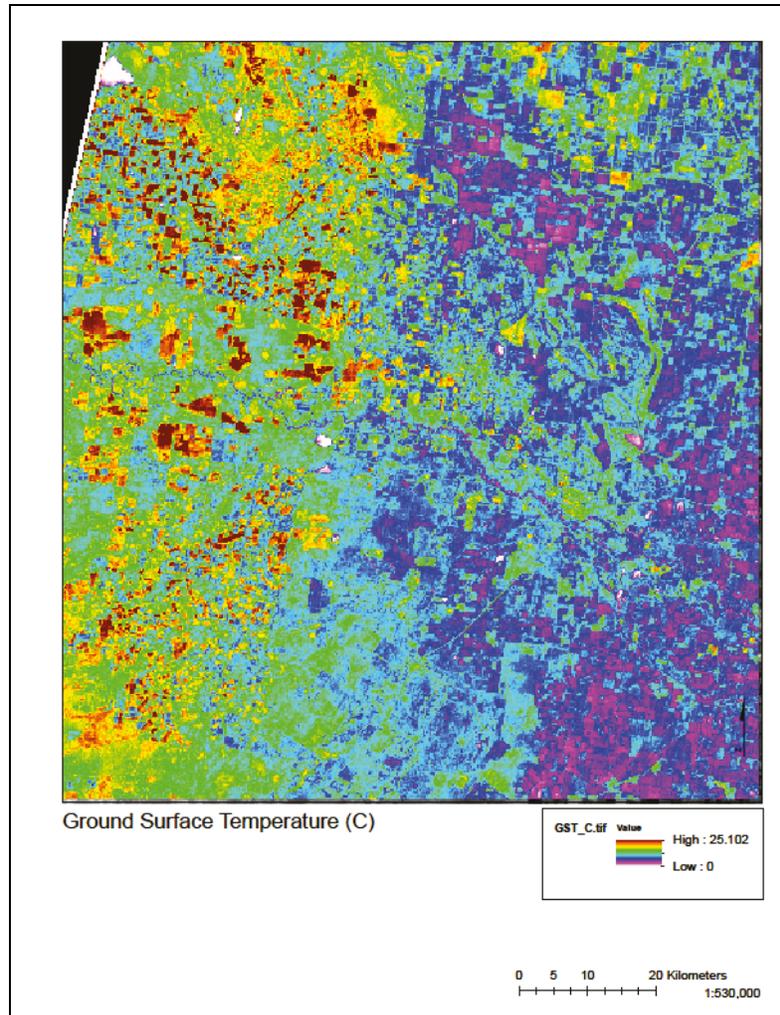
Where:

- T = Top of atmosphere brightness for Band 10
- w = wavelength of emitted radiance (11.5μm)
- p =  $h * c/s$  (1.438E-2 mK)
- h = Plank's constant (6.626E-34 Js)
- s = Boltzmann constant (1.38E-23 J/K)
- c = velocity of light 2.998E8 m/s)
- e = land surface emissivity

Where:

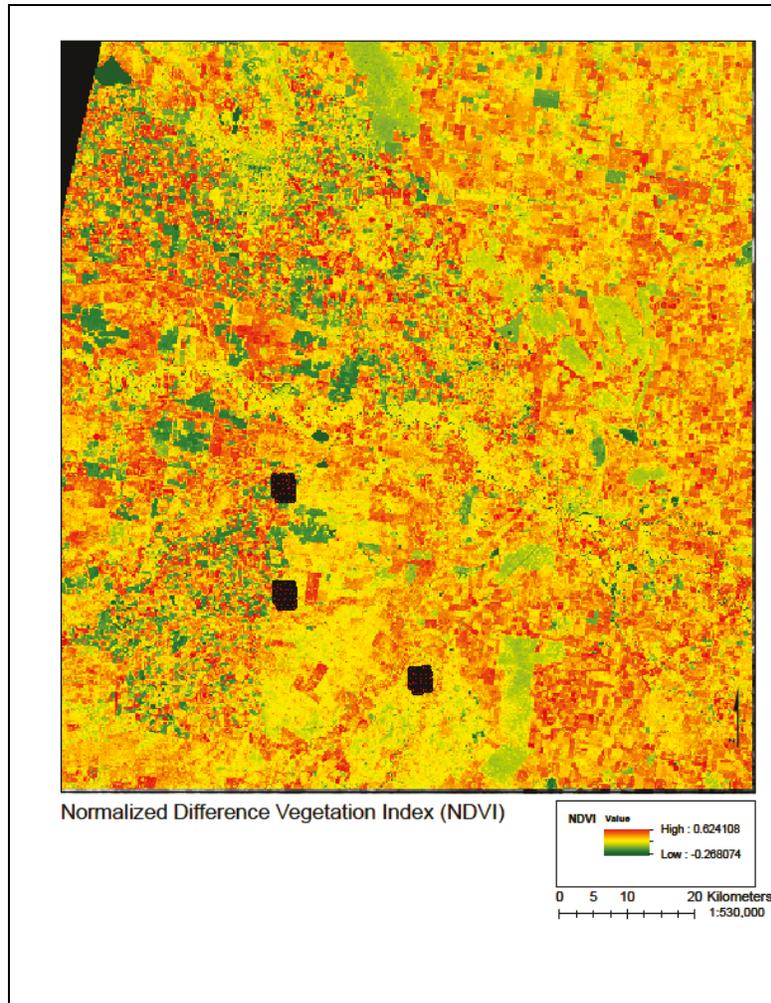
$$P_v = (\text{NDVI} - \text{NDVImin} / \text{NDVImax} - \text{NDVImin})^2 \& e = 0.004P_v + 0.986 \text{ Equation (15)}$$

Figure 9. Calculated ground surface temperature from Landsat 8, band 10.



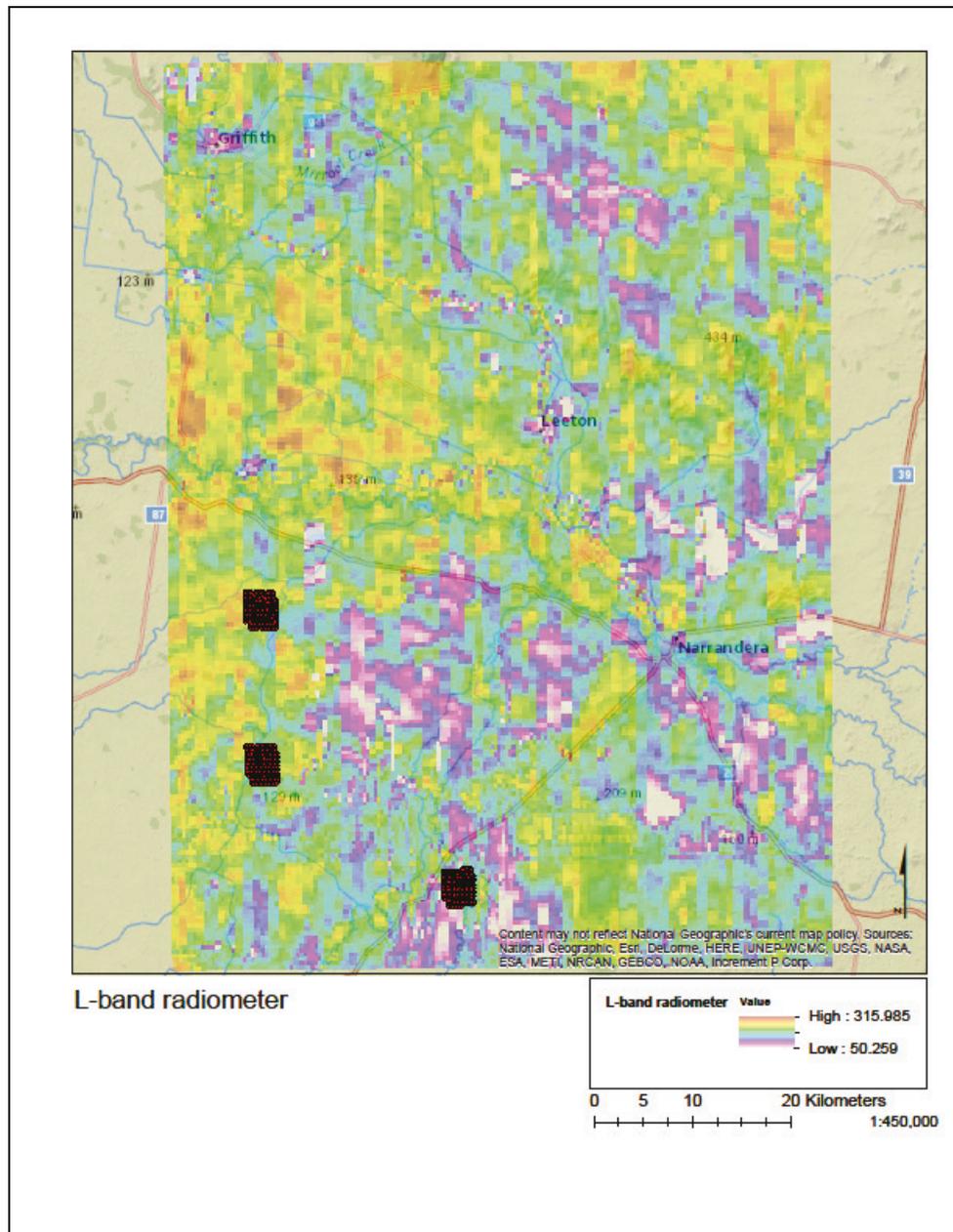
When combined with NDVI (Figure 10), all variables are now available that are a function of soil moisture,  $f(SM) = (T_v, T_h, T_s, \text{NDVI})$ . Using the measured values in the high resolution grids, the site specific soil moisture equation should be easily determinable from the polarized microwave radiometer datasets.

Figure 10. The NDVI from Landsat 8, bands 5, and 6. The three sites of exhaustive dielectric impedance measurement denoted by the red and black clusters.



Evaluating for the f-channel radiometer was conducted via an empirical Bayesian kriging as it provides a better fit to the data over standard kriging methods with a mean error of 0.03, a Root Mean Square (RMS) of 3.4, and an average standard error of 3.6532. Initial analysis of the radiometer data shows considerable banding in the direction of the flight path, while a filter function may be appropriate, the first step is to standardize all the data from the various beams (Figure 11).

Figure 11. Initial radiometer data showing a strong banding artifact in the direction of the flight path.



As the beams cover a relatively equal area and are over a wide enough region, the assumption can be made that all six should be from the same population and share the same distribution of values. Therefore, all six individual beams are normalized to the population of the entire dataset (Figure 12, Table 2).

Figure 12. The six beam bands illustrated over the area of interest.

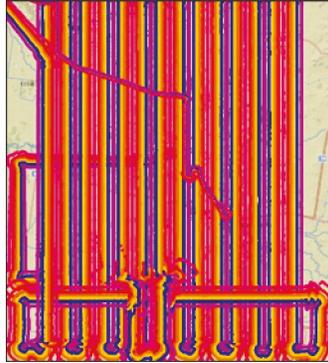


Table 2. Mean and Standard Deviation of each individual flight path and the full population.

	Beam1	Beam2	Beam3	Beam4	Beam5	Beam6	All
Mean	231.914	236.82003	245.7915	242.4373	241.3469	236.4652	239.1292
Standard Deviation	27.09663	27.196371	25.90451	27.82851	27.94064	26.48179	27.46527

Normalizing each band to that of the total population reduces the mean error of prediction in the Kriged surface to 0.027 and the Average Standard Error to 3.48. The final output of the normalized L-band data shows the majority of banded effects removed, which can now be used to determine volumetric water content in concert with the vegetation optical depth via the NDVI, and ground surface temperature.

Application of Land Surface Temperature, NDVI, and L-band radiometric brightness temperature was averaged over an area of 0.5 km pixels and trained against the aggregation of several hundred dielectric impedance measured soil moisture values (Figure 6) in the same resolution. The training and application algorithm utilized was a random forest. The ground truth metric upon which evaluation of the ILSMA will be measured is the normalized L-band which derived values from the aggregated dielectric impedance measurements (Figure 13). The radiometer derived soil moisture had a correlation with the radiometer raw return values of 0.9674 with a mean absolute error of 0.0207. Variation is due primarily to input of ground thermal data and vegetation density. Numerical values were then transferred to a surface via an empirical Bayesian Krig with an error of 1.6 E-5 (Figure 14).

Figure 13. Normalized L-band return adjusted for variation in the individual beams along flight path.

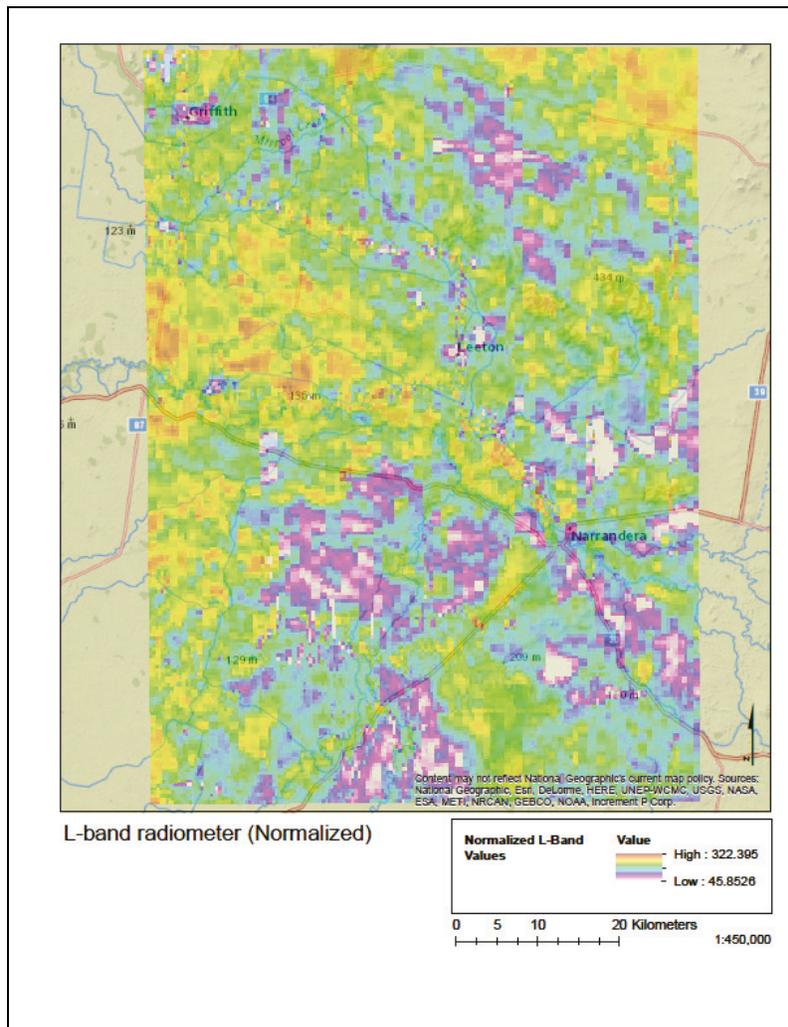
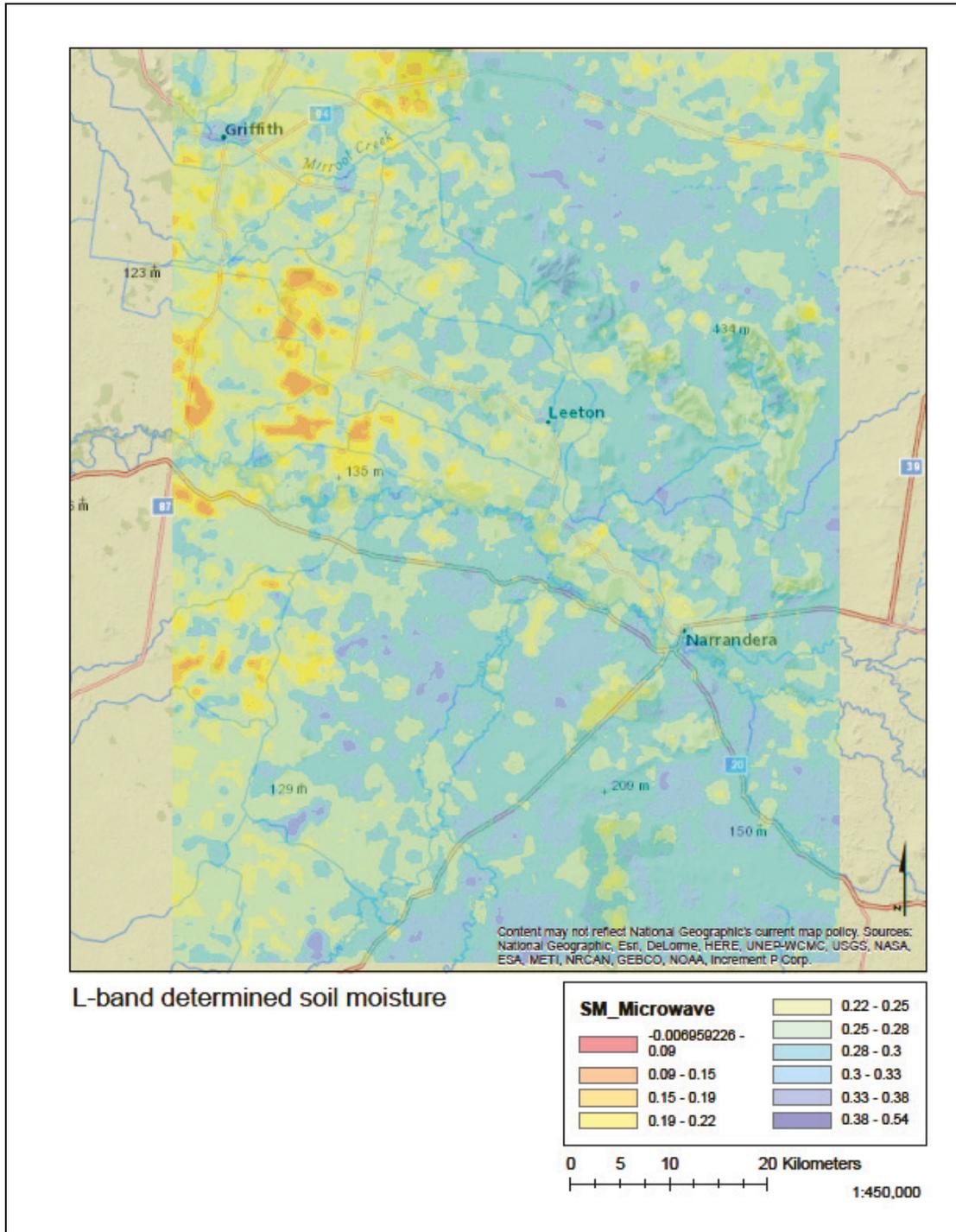
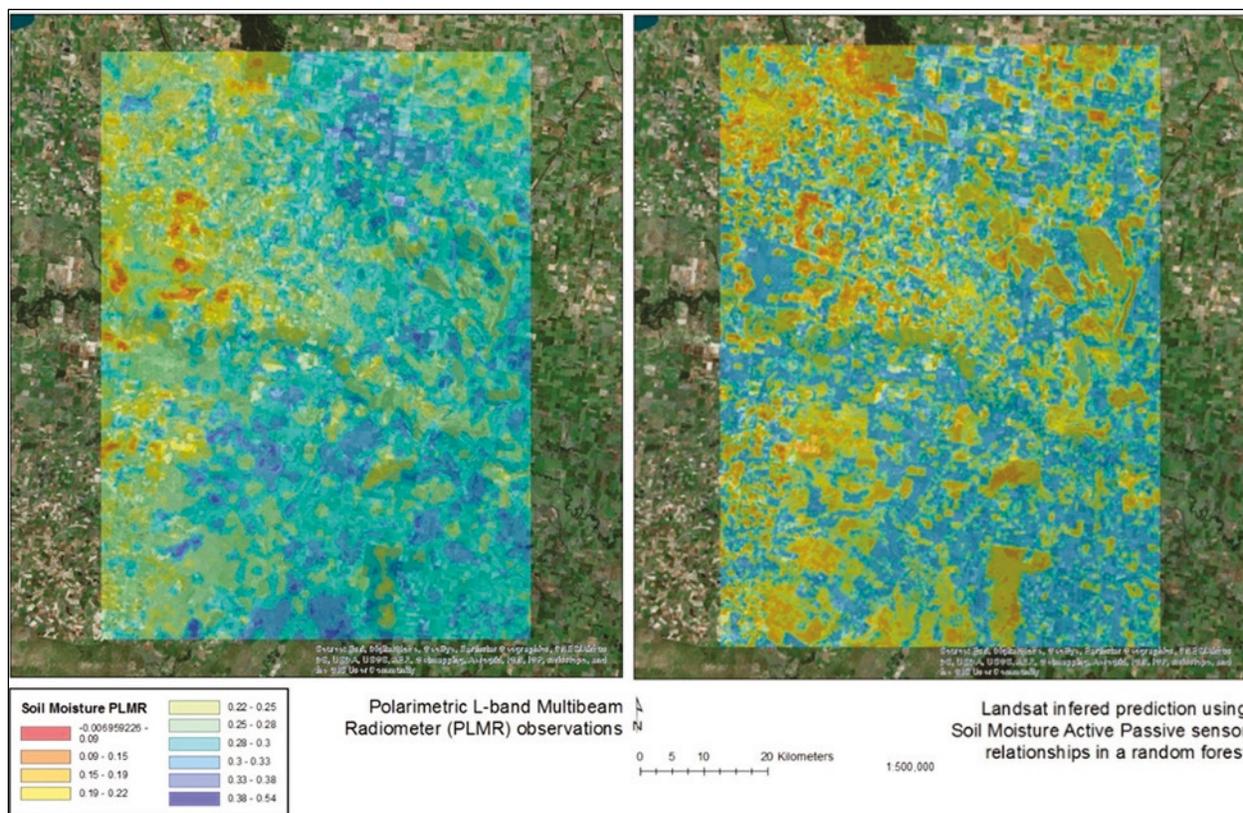


Figure 14. The Kriged surface of the L-band radiometer data.



Comparing this against the radiometer data (Figure 15):

Figure 15. A visual comparison of the L-band radiometer and the Landsat inferred soil moisture shows generally good agreement overall with areas of notable exception, such as the southern edge, just west of center.



Overall, the predictive power of the downscaled soil moisture did quite well, although just outside the NASA standard of 0.04 for the SMAP validation program. The full population of 3,097 pixels by 2,257 pixels, for a total of 6,989,929 samples with an absolute error of 0.054 over the L-band retrieved surface (see Tables 3 and 4, Figure 16).

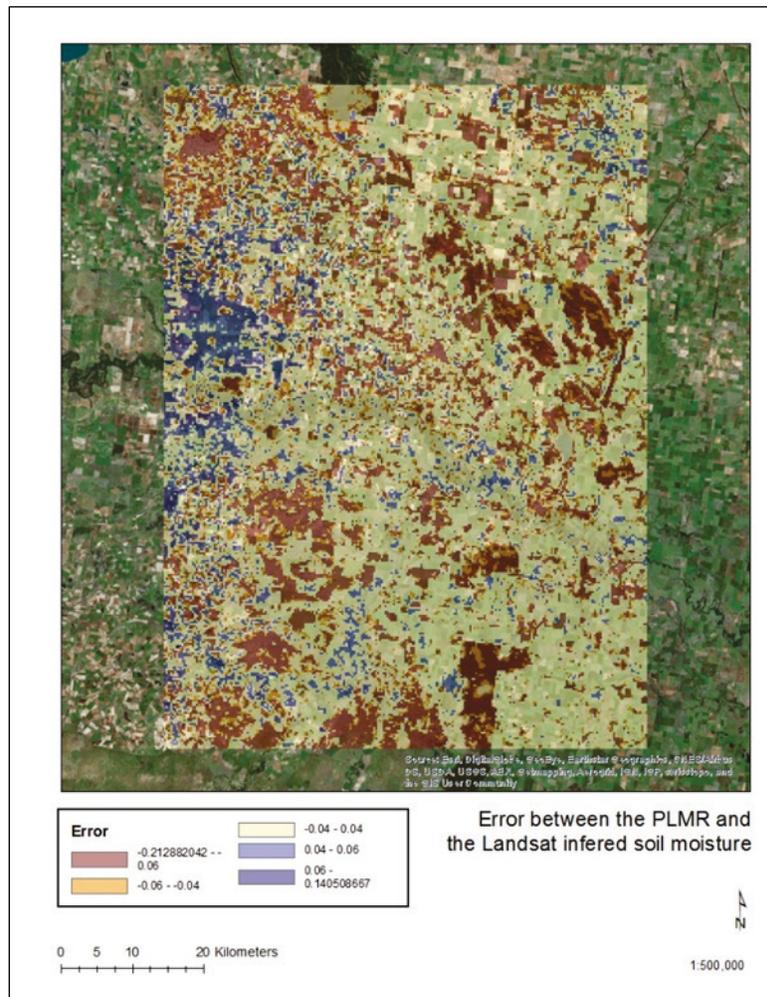
Table 3. The mean of the proposed algorithm is 0.26 while the mean of the L-band retrieval is 0.28; the mean absolute error (MAE) is 0.05.

	LS Moisture Pred	L-Band	Error	Abs Error
Mean	0.258620513	0.275642403	0.017021891	0.053538068
StdDev	0.066202027	0.028492341	0.062812946	0.036998462
Skew	-0.459527334	-1.409925752	0.375371997	0.734758179

Table 4. Magnitude of error within the algorithm based on the L-band surface. Currently 56.44% of the result meets NASA specifications of 0.04 providing the L-band retrieval was accurate.

Pct of pop within each error bin	
< -0.06	6.56%
-0.06: -0.04	11.82%
-0.04: 0.00	33.15%
0.04: 0.00	11.47%
0.04: 0.06	6.67%
> 0.06	30.34%

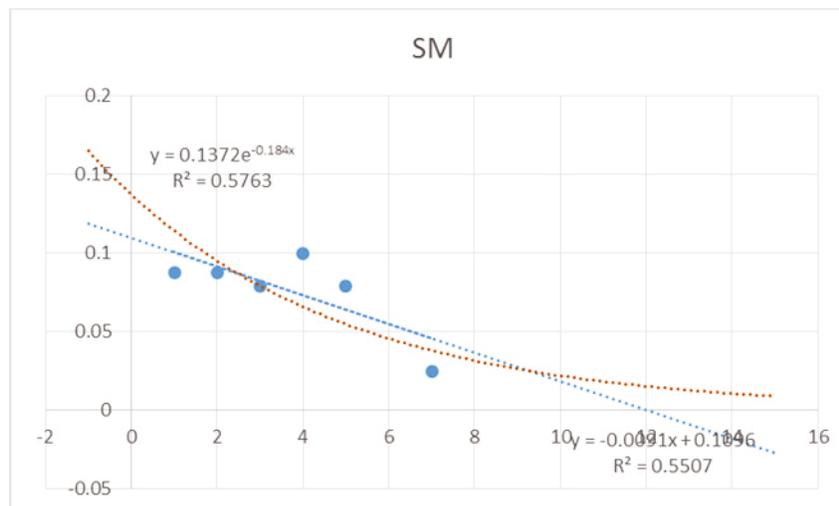
Figure 16. Qualitative review of error between the L-band retrieval and the Landsat-SMAP downscaling algorithm.



## 5 Discussion

Applying the previously described methodology has shown considerable fitness in describing soil moisture distribution at significantly higher resolution than currently available using SMAP alone. The method utilizes the refinement of the SMAP algorithm, ingraining in the results the description of the soil moisture relationship to the spectral conditions on the ground. While the method did not prove an accuracy within the 0.04 range as required by the NASA SMAP mission in this experiment, it may be possible with additional variable refinement or additional validation. Time domain reflectometry, like all point based in-situ soil measurements, observe ground conditions in a highly localized area, no more than 1.2 inches beyond the waveguides (Munoz-Carpena et al. 2004). Since such a small sample of the pixel space is tested, the dielectric impedance values assigned to a pixel in the original interpretation of the L-band may not have always been accurate in describing the true value of the pixel and instead fell within the tail of the distribution of soil moisture values within a pixel. This can be demonstrated by evaluating the number of dielectric impedance sample points within a 30 m Landsat pixel and the error of that averaged dielectric impedance value with the downscaled Landsat inferred algorithm results (Figure 17).

**Figure 17.** The number of dielectric impedance samples on the x-axis compared with the error from the Landsat inferred soil moisture value suggests that dielectric impedance values are prone to high error in terms of the average value of the pixel when less than nine samples are obtained over the 30m<sup>2</sup> pixel area.



While the sheer number of observations aid the alleviation of some of the training errors, a large number of them can affect the training algorithm that allocated the transition variables from raw L-band return to soil moisture content. Among the three sampling regions of soil moisture dielectric impedance collection, two were used for training the L-band radiometer in transitioning return to soil moisture content, the 3rd was used for testing. That region experienced a 0.0858 MAE, not too dissimilar from the slightly lower 0.0823 MAE seen over the same test region for the Landsat inferred data and the dielectric impedance measurements. However, the Landsat inferred data and the radiometer demonstrated a significantly higher coupling in their returns with a 0.0535 MAE suggesting the Landsat/SMAP inferred algorithm produced a dataset with a stronger description of the L-Band radiometer data than that of the dielectric impedance and thus is a viable option for downscaling SMAP into a local resolution (see Table 5).

**Table 5. MAE values of data types from on-another over the naïve test region.**

Mean Average Error		
	ILSMA	L-Band
L-Band	0.0535	
TDR	0.0823	0.0858

While obvious weakness exists in that visual remote sensing cannot obtain imagery through cloud cover as opposed to microwave, this algorithm does allow for widespread moderately high resolution soil moisture capture over large regions. This could be a viable option for detecting moderately high resolution soil moisture in arid and semi-arid regions as well as for soil moisture pattern distribution analysis.

## References

- Ahmed, M. Y., A. H. Nury, F. Islam, and M. J. Alam. 2012. Evaluation of geotechnical properties and structural strength enhancing road pavement failure along Sylhet-Sunamganj highway, Bangladesh. *Journal of Soil Science and Environmental Management* 3(5): 110–117.
- Breiman, L. 1996. Bagging Predictors. *Machine Learning* 24(2):123–140. doi: 10.1007/BF00058655.
- Cannon, A. J., and P. H. Whitfield. 2002. Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *Journal of Hydrology* 259(1–4):136–151.
- Carlson, T. 2007. An Overview of the "Triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors* 7(8): 1612–1629. doi: [10.3390/s7081612](https://doi.org/10.3390/s7081612).
- Cosby, B. J., C. M. Hornberger, R. B. Clapp, and T. R. Ginn. 1984. A statistical exploration of the relationship of soil moisture characteristics to the physical properties of soils. *Water Resources Research* 20(6):682–690. doi: 10.1029/WR020i006p00682.
- Crosta, G. 1998. Regionalization of rainfall thresholds: an aid to landslide hazard evaluation. *Environmental Geology* 35(2–3):131–145. doi: 10.1007/s002540050300.
- De Michele, C., and G. Salvadori 2002. On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture conditions. *Journal of Hydrology* 262(1–4):245–258. [https://doi.org/10.1016/S0022-1694\(02\)00025-2](https://doi.org/10.1016/S0022-1694(02)00025-2).
- Deepa, C., K. Sathiyakumari, and V. Pream Sudha. 2010. Prediction of the compressive strength of high performance concrete mix using tree based modeling. *International Journal of Computer Applications* 6(5):18–24.
- Denmead, O. T., and R. H. Shaw. 1962. Availability of soil moisture to plants as affected by soil moisture content and meteorological conditions. *Agronomy Journal* 54(5): 385–390. doi:10.2134/agronj1962.00021962005400050005x.
- Evet, S. R. 2000. Exploits and endeavors in soil water management and conservation using nuclear techniques. Nuclear techniques in integrated plant nutrient, water, and soil management: In *Proceedings of an International Symposium (151–177)*. Vienna, Austria: International Atomic Energy Agency.
- Hastie, T., R. Tibshirani, and J. Friedman. 2008. *The elements of statistical learning* (2<sup>nd</sup> Ed.). Springer.

- Koster R. D., P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C. H. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada. 2004. Regions of strong coupling between soil moisture and precipitation. *Science* 305(5687): 1138–1140. doi: [10.1126/science.1100217](https://doi.org/10.1126/science.1100217).
- Merlin, O., J. P. Walker, A. Chehbouni, and Y. Kerr. 2008. Towards deterministic downscaling of SOM soil Moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment* 112(10):3935–3946. doi: <https://doi.org/10.1016/j.rse.2008.06.012>.
- Monerris, A., and T. Schmugge. 2009. Soil moisture estimation using L-band radiometry. Ed. G. Jedlovec. In *Advances in Geoscience and Remote Sensing*. Shanghai: InTech.
- Munoz-Carpena, R., S. Shukla, and K. Morgan. 2004. *Field devices for monitoring soil water content. southern regional water program*. University of Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, EDIS.
- National Aeronautics and Space Administration (NASA). 2011. NASA. Retrieved Feb. 20, 2015, from Landsat 7 Science Data Users Handbook: [http://landsathandbook.gsfc.nasa.gov/orbit\\_coverage/](http://landsathandbook.gsfc.nasa.gov/orbit_coverage/).
- Ochsner, T. E., M. H. Cosh, R. H. Cuenca, W. A. Dorigo, C. A. Draper, Y. Hagimoto, and M. Zreda. 2013. State of the art in large-scale moisture monitoring. *Soil Science Society of America Journal* 77(6):1888–1910. doi: 10.2136/sssaj2013.03.0093.
- Onyari, E. K., and F. M. Ilunga. 2013. Application of MLP neural network and M5P model tree in predicting streamflow: A case study of luvuvhu catchment, South Africa. *International Journal of Innovation, Management and Technology* 4(1):11–15.
- Orchard, V. A., and F. J. Cook. 1983. Relationship between soil respiration and soil moisture. *Soil Biology and Biochemistry* 15(4):447–453. [https://doi.org/10.1016/0038-0717\(83\)90010-X](https://doi.org/10.1016/0038-0717(83)90010-X).
- Panciera, R. W. 2013. The soil moisture active passive experiments (SMAPEX): Towards soil moisture retrieval from the SMAP mission. *IEEE Transaction on Geoscience and Remote Sensing* 52(1):490–507. doi: [10.1109/TGRS.2013.2241774](https://doi.org/10.1109/TGRS.2013.2241774).
- Panciera, R., J. P. Walker, and O. Merlin. 2009. Improved understanding of soil surface roughness parameterization for L-Band passive microwave soil moisture retrieval. *IEEE Geoscience and Remote Sensing Letters* 6(4):625–629. doi: [10.1109/LGRS.2009.2013369](https://doi.org/10.1109/LGRS.2009.2013369).
- Shoop, S. A. 1993. Terrain characterization for trafficability. CRREL Report 93-6 . Hanover, NH: U.S. Army Corp of Engineers. Cold Regions Research and Engineering Laboratory.
- Soil Moisture. 1999. National Aeronautical and Space Administration (NASA) earth science office. [http://wwwghcc.msfc.nasa.gov/landprocess/lp\\_home.html](http://wwwghcc.msfc.nasa.gov/landprocess/lp_home.html). (Retrieved January 15, 2015)

- Song, C., J. Li, and M. Massimo. 2014. Retrieving high-resolution surface soil moisture by downscaling AMSR-E Brightness Temperature Using MODIS LST and NDVI data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 7(3):935–942.
- Topp, G. C., J. L. Davis, and A. P. Annan. 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resource Research* 16(3):574–582. doi: 10.1029/WR016i003p00574.
- Ulaby, F. T., Moore, R. K., & Fung, A. K. (1986). Microwave remote sensing: active and passive. Volume III: from theory to applications. In *Microwave remote sensing: active and passive. Volume III: from theory to applications*. Artech House; Remote Sensing Series, 4.
- U.S. Geological Survey (USGS). 2011. Earth Observing 1 (EO-1). <http://eo1.usgs.gov/sensors/hyperion>. Retrieved 20 Feb 2015.
- U.S. Geological Survey (USGS). 2016. Landsat 8 (L8) data users handbook. USGS Home: [http://landsat.usgs.gov//l8handbook\\_section5.php](http://landsat.usgs.gov//l8handbook_section5.php). Retrieved 01-05-16.
- Weizu, G., and J. Freer. 1995. Patterns of surface and subsurface runoff generation. *IAHS Publications-Series of Proceedings and Reports-Intern Association Hydrological Sciences* 229: 265–273.
- Wu, D., B. Yan, Y. Cui, and X. Xu. 2010. Retrieve soil moisture from mixed-pixels based on scale transformation using hyperspectral data. *International Geoscience and Remote Sensing Symposium* 3875–3878.
- Wu, X., J. P. Walker, C. Rüdiger, and R. Panciera. 2015. Effect of land-cover type on the SMAP Active/Passive Soil Moisture Downscaling Algorithm Performance. *IEEE Geoscience and Remote Sensing Letters* 12(4):846–850. doi: [10.1109/LGRS.2014.2364049](https://doi.org/10.1109/LGRS.2014.2364049).
- Ye, N., Walker, J. P., Wu, X., de Jeu, R., Gao, Y., Jackson, T. J., Jonard, F., Kim, E., Merlin, O., Pauwels, V., Renzullo, L. J., Rüdiger, C., Sabaghy, S., von Hebel, C., Yueh, S. H., and Zhu, L. 2014. The Soil Moisture Active Passive Experiments: Towards Calibration and Validation of the SMAP Mission. *IEEE Transactions on Geoscience and Remote Sensing* 52(1):490–507.

# REPORT DOCUMENTATION PAGE

*Form Approved*  
**OMB No. 0704-0188**

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE (DD-MM-YYYY)</b> September 2018	<b>2. REPORT TYPE</b> Final report	<b>3. DATES COVERED (From - To)</b>		
<b>4. TITLE AND SUBTITLE</b>  Determining Soil Moisture at Moderately High Resolution via Soil Moisture Active Passive (SMAP) and Landsat Using Inferred Learning		<b>5a. CONTRACT NUMBER</b>		
		<b>5b. GRANT NUMBER</b>		
		<b>5c. PROGRAM ELEMENT NUMBER</b>		
<b>6. AUTHOR(S)</b>  Michael G. Lewis, Andmorgan Fisher, and Clint Smith		<b>5d. PROJECT NUMBER</b> ERDC GRL ARTEMIS		
		<b>5e. TASK NUMBER</b>		
		<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  U.S. Army Engineer Research and Development Center Geospatial Research Laboratory 7701 Telegraph Road Alexandria, VA 22315-3864		<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  ERDC/GRL TR-18-1		
		<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  Headquarters, U.S. Army Corps of Engineers Washington, DC 20314-1000		<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
		<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for public release; distribution unlimited.		
<b>13. SUPPLEMENTARY NOTES</b>				
<b>14. ABSTRACT</b>  This paper reviews an approach to increase soil moisture resolution over a sample region over Australia using the Soil Moisture Active Passive (SMAP) sensor and Landsat 8 only. This approach uses an inductive localized approach, eschewing the need to obtain a resilient and definitive model in favor of a temporary model that utilizes current conditional inputs only. For the purposes of this analysis, the SMAP 36 km soil moisture product is considered fully valid and accurate. Landsat bands coinciding in collection date with a SMAP capture are down sampled to match the resolution of the SMAP product. A series of indices describing the Soil-Vegetation-Atmosphere Triangle (SVAT) relationship are then produced using the down sampled Landsat bands. These indices are then related to the local coincident SMAP values to identify a series of rules or trees to identify the local rules defining the relationship between soil moisture and the aforementioned indices. This paper uses a random forest due to its highly accurate learning against local ground truth data. The defined rules are then applied to the Landsat image in the native Landsat resolution to determine local soil moisture. Ground truth comparison is done via a series of grids using time dielectric impedance and airborne L-band Multi-beam Radiometer (PLMR) observations done under the SMAPEX-5 campaign (Panciera 2013). The predictive power of the inferred learning soil moisture algorithm (ILSMA) did well with a mean absolute error of 0.054 over an airborne L-band retrieved surface over the same region.				
<b>15. SUBJECT TERMS</b>		Soil moisture Environmental monitoring—Remote sensing Microwave remote sensing		
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>
<b>a. REPORT</b> UNCLASSIFIED	<b>b. ABSTRACT</b> UNCLASSIFIED	<b>c. THIS PAGE</b> UNCLASSIFIED		42
			<b>19a. NAME OF RESPONSIBLE PERSON</b>	
			<b>19b. TELEPHONE NUMBER (include area code)</b>	