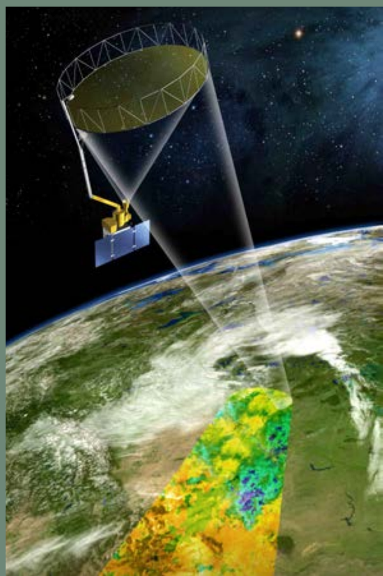
High Res Soil Moisture Inference via Remote Sensing using SMAP

Michael Lewis

Physical Scientist

Geospatial Research Laboratory

ERDC – US Army Corps of Engineers



U.S.ARMY

US Army Corps of Engineers®

ERDC

Engineer Research and Development Center

The problem

How do we find a really great and easy to obtain dataset for validation of soil moisture testing methods?

The concept origin

The origin of the idea arose from Song & Massimo, 2014.The team's approach to answering this question utilized an upscaled L-band surface to a MODIS NDVI resolution as a path to downscale NDVI.

In short:

An established relationship exists between the Microwave readings of soil moisture and remote sensing in the VIS/NIR range in the form of the Microwave polarization difference Index:

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

where E_0, E_1, and E_2 are constants. These constants can be solved through upscaling microwave to the MODIS resolution and sampling 3 times to develop the curve to fit all three constants.

Song, C., Li, J., & Massimo, M. (2014). Retrieving High-Resolution Surface Soil Moisture by Downscaling AMSR-E Brightness Temperature Using MODIS LST and NDVI Dat. *IEEEJOURNALOFSELECTEDTOPICSINAPPLIEDEARTHOBSERVATIONSANDREMOTESENSING*, 935-942.

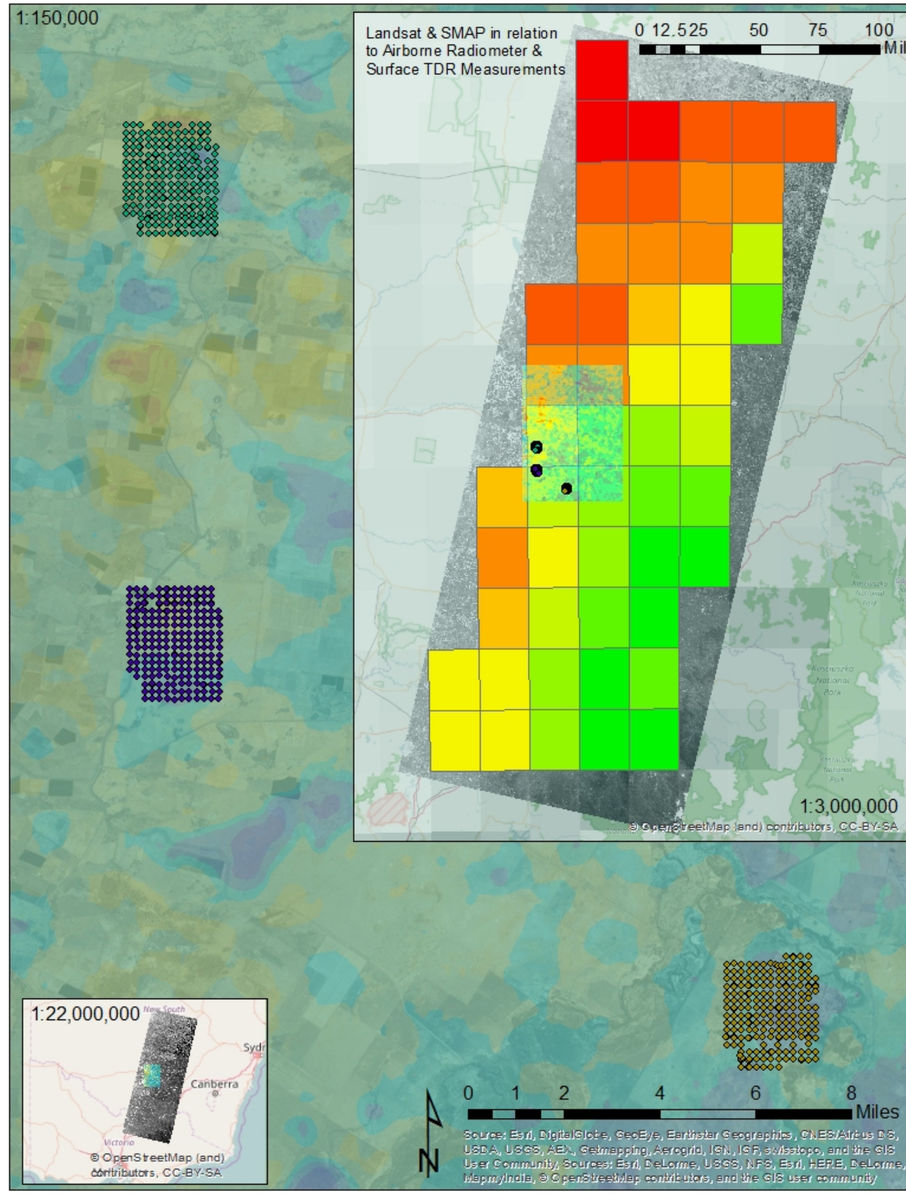
Why limit at NDVI and Temperature

Song’s work made me realize I had a plentiful training set available for soil moisture in the SMAP data itself, regardless of the scale. If soil moisture distribution is fractal then local relationships would be reasonable to apply at differing scales within the same locale.

Rules defining soil moisture at the macro scale apply to the micro scale (mostly).

Finding the perfect field sample set

Working with Jeff Walker of Monash U. under the NASA SMAPex over the Murrumbidgee River Basin and surrounding areas as well as a close-up of TDR test sites, each point is 3 readings. TDR, Microwave, (clear) Landsat, and SMAP. (as well as low veg)



Panciera, R., Walker, J. P., Jackson, T. J., Gray, D., Tanase, M. A., Ryu, D., Monerris, A., Yardley, H., Rüdiger, C., Wu, X., Gao, Y., Hacker, J. (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.

Finding the perfect variables

Johnny Nunez, a graduate student from the University of Puerto Rico, Mayaguez and I spent several weeks finding clean Landsat 8 images (and a few Landsat 7) that coincided with Microwave overflights, and TDR field studies.

We then extracted all band values (1, 3-11) and took a completely naïve approach. We evaluated the total population of points we had relatively certain soil moisture values for against the band values for any strong correlation, unsurprisingly we found not too much of a correlation. We then ratioed each band by every other band and looked again. We iterated this two more times (although due to the growing number we only chose specific areas for the second iteration) where we not only ratioed against products of that iteration but all those before it.

Finding the perfect variables

With a full set of very quickly expanding potential indices

(10 + 55 + 2,145 + a lot … Regardless to say, we were in the thousands.)

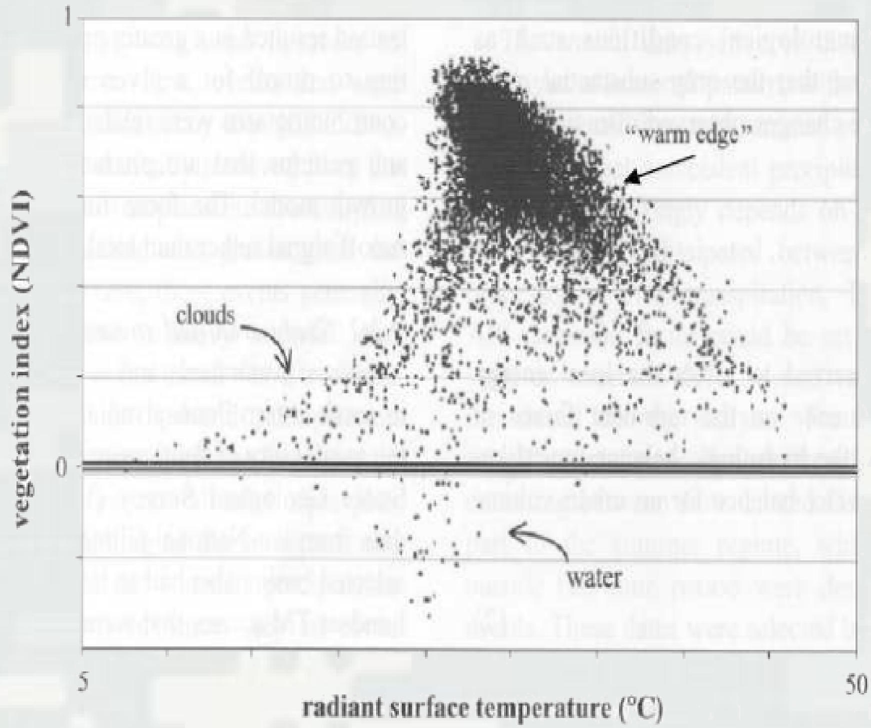
we ran multiple machine learning algorithms and naïve attribute selection functions, identifying in which ones were holding the most information linking to soil moisture.

We also added many vegetation, moisture, and temperature relevant indices as well prior to testing.

Initially we settled on about 30 that performed quite well, then we iteratively paired down the variables. Using PCA as a guide, we continued removing variables until performance routinely declined across multiple tests. We eventually settled on 6 variables that retained a disproportionate amount of the information.

The final 6 variables describe the SVAT triangle (and moisture)

Evaluation of remotely sensed indices and ratios identified those describing the SVAT triangle were most important to identifying soil moisture. NDVI and temperature proxies as well as some that are sensitive directly water.



Carlson, T. (2007). An Overview of the "Triangle Method" for Estimating Surface Evapotranspiration and Soil Moisture from Satellite Imagery. *Sensors, 7*, pp. 1612-1629.

The final 6 variables Describe the SVAT triangle

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, therefore is associated with plant growth.

Neighborhood 1 (Thermal)
Var2 - Band1/Band10

Neighborhood 2 (Moisture)
*NMSI – (B7-B6)/(B7+B6)
**MNDWI – (B3-B6)/(B3+B6)

Neighborhood 3 (Vegetation)
Var10 - (B8/B11)/(B9/B10)
NDVI – (B5-B4)/(B5+B4)

From Parts Unknown
Var16 – (B4 * B7^3)/(B5 * B6^2)

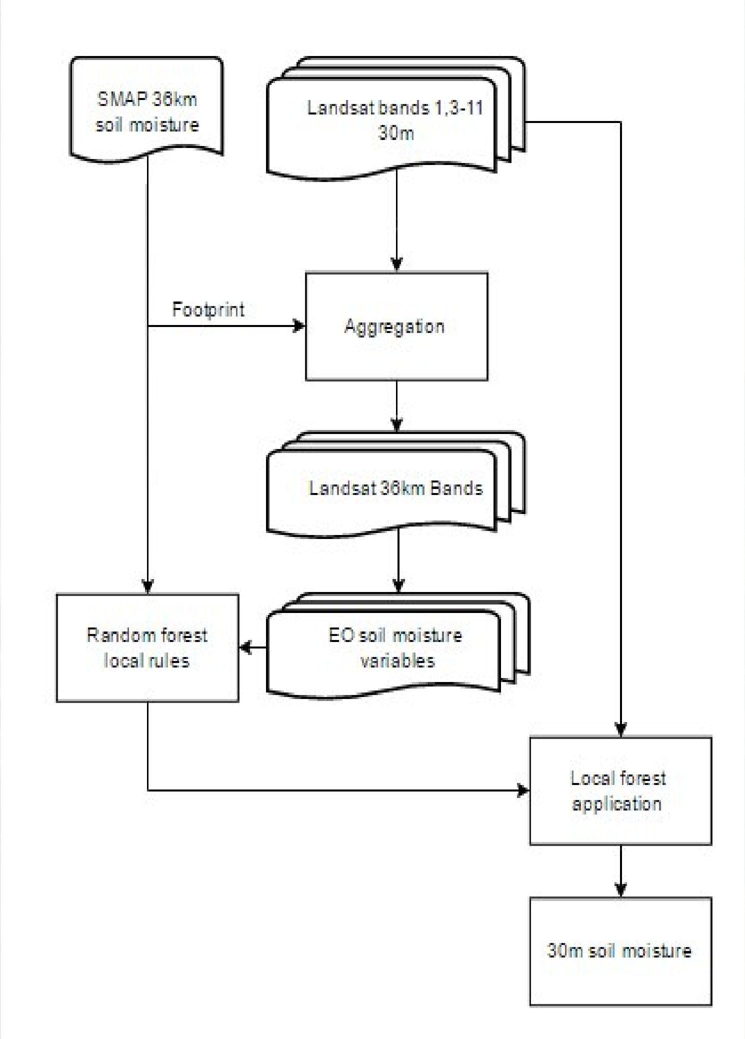
* Normalized Soil Moisture Index
** Modified Normalized Difference Water Index

The process

Using SMAP to build local rules based on the aggregated constants, then transferring those rules to higher resolution imagery indices creates a higher resolution product.

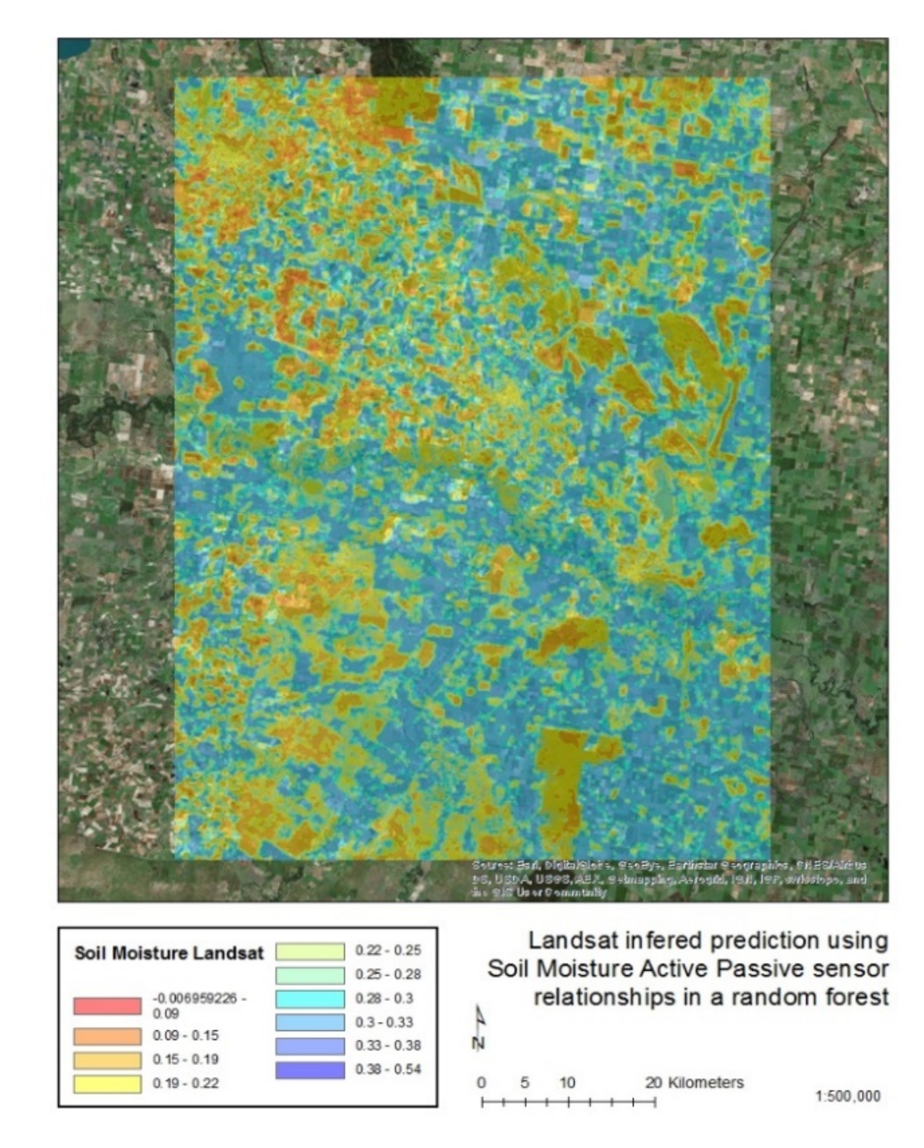
Why Random Forest?

- It's simpler than physically modelling each region's unique interactions.
- Because we are working within our training region, it is essentially classification over prediction.
- Random Forest (AFAIK) had the best results.

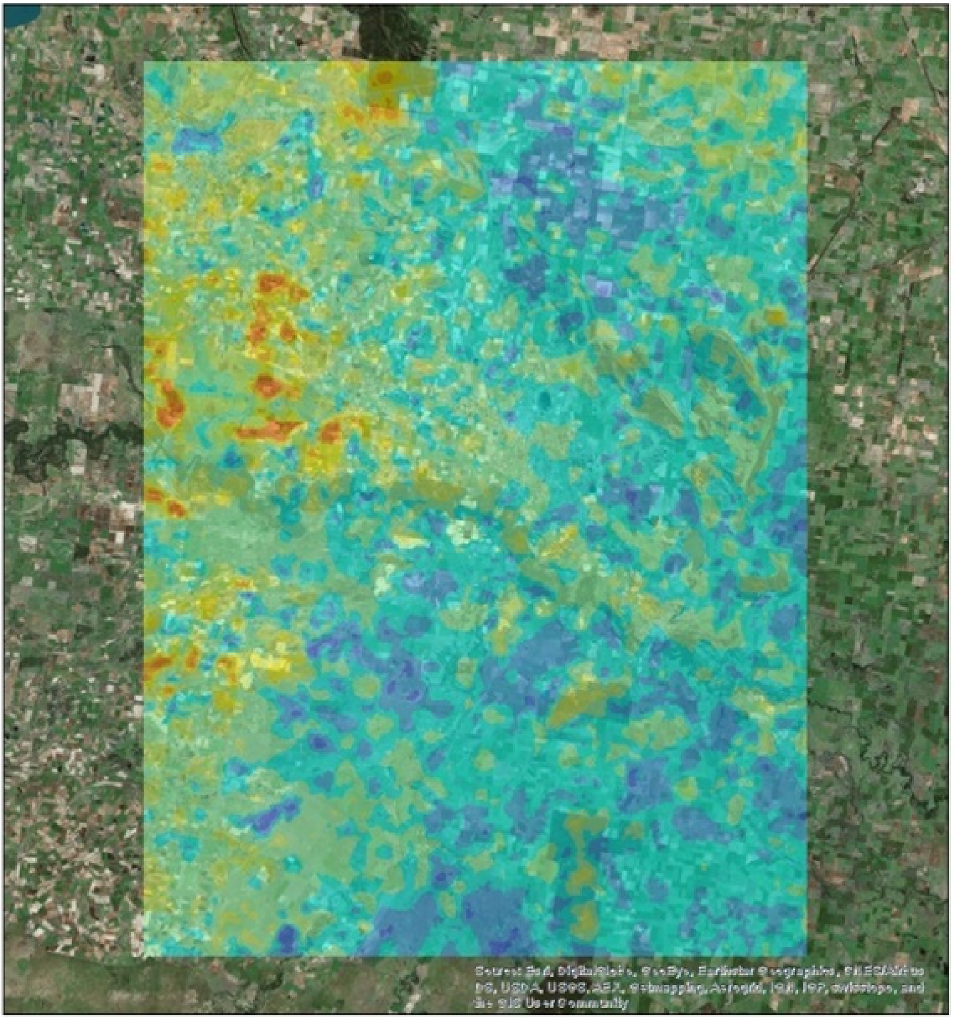


Visualizing Results

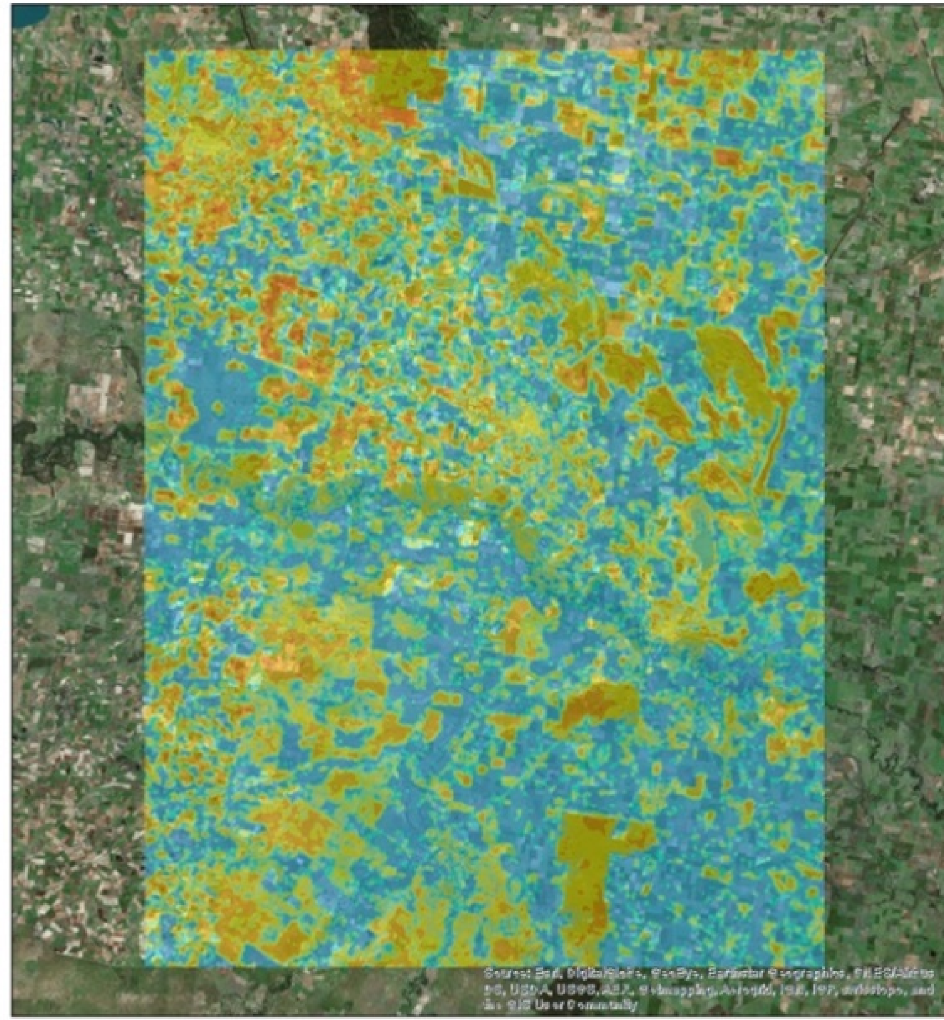
Providing this downscaled product at 30m Landsat resolution.



Comparing to L-band



Polarimetric L-band Multibeam Radiometer (PLMR) observations



Landsat infered prediction using Soil Moisture Active Passive sensor relationships in a random forest

Results

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.

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

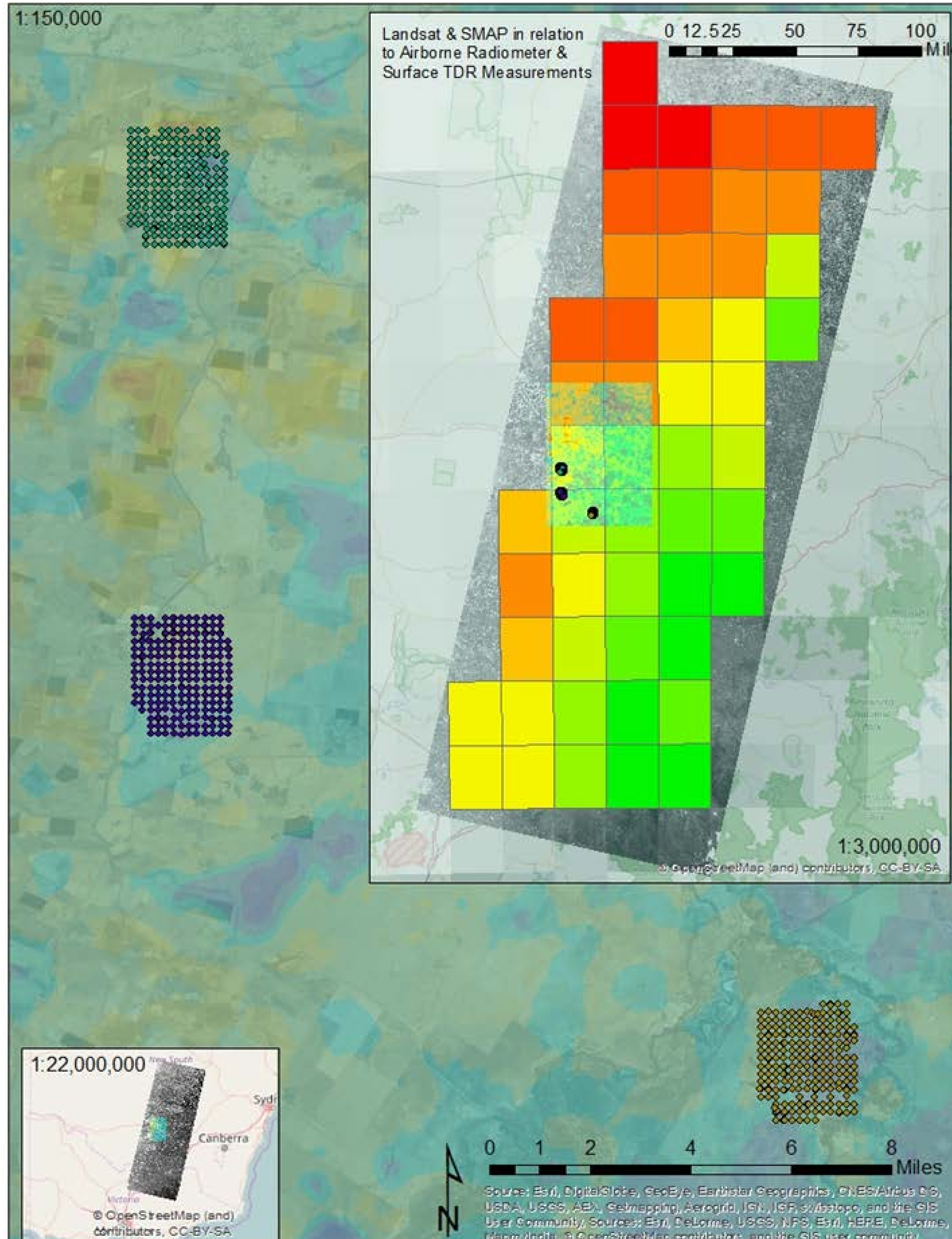
Over 600+ TDR sample points. [Side note – In sub-samples, the more TDR points in a pixel, the higher the correlation.]

Landsat inferred rules seem to hold up!

One more look at the testbed

You can see the local rulesets for this was created using 3 Landsat scenes, 51 SMAP pixels. These rules will be variable for each “region” due to soil types, moisture conditions, vegetation reflectance, seasonal variation, etc. The rules change but are discoverable, the variables and method should be able to discern soil moisture in arid and semi-arid regions. Testing in a high vegetation region has not yet been attempted.

Evaluation against PLMR was done at 30m postings over the entirety of the overlap. PLMR was aggregated to 30m for evaluation. Evaluation was done against the average of the TDR readings within each pixel (for both PLMR and Landsat based).



Conclusion

- Problem: Need large high(er) resolution soil moisture surface for validation and modeling.

- Ans: Using VIS/NIR variables that seem to well describe the local soil moisture regimen, a unique random forest can be generated to demonstrate distribution of soil moisture using SMAP as the independent variable for any time in a 3-scene set. Application of the learned random forest can then be applied at higher resolution.

- Result: Adequately describes soil moisture in initial test comparable to both TDR and PLMR, additional test sites needed to confirm.