Ground Network Design and Dynamic Operation for Validation of Space-Borne Soil Moisture Measurements

M. Moghaddam, M. Liu, D. Teneketzis (U of Michigan), D. Entekhabi (MIT)

Students: M. Burgin, Y. Goykhman, K. Li, A. Nayyar, D. Shuman, X. Wang

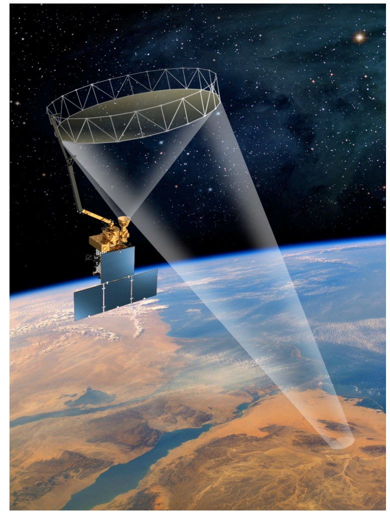
AIST-08-0017

ESTF, 22 June 2010

Technology Relevance: SMAP

SMAP Primary Science Objectives:

- Global high-resolution mapping of soil moisture and its freeze-thaw state to:
  - Link terrestrial water, energy, and carbon cycle processes
  - Estimate global water and energy fluxes at the land surface
  - Quantify net carbon flux in boreal landscapes
  - Extend weather and climate forecast skill
  - Develop improved flood and drought prediction capability



Mission Approach:

- GSFC L-band radiometer
- JPL L-band radar
- Common 6m rotating antenna for 3-day global repeat coverage
- Merged radar and radiometer data for high-accuracy, mid-resolution, soil moisture
- 670 km polar sun-sync

Development Status:

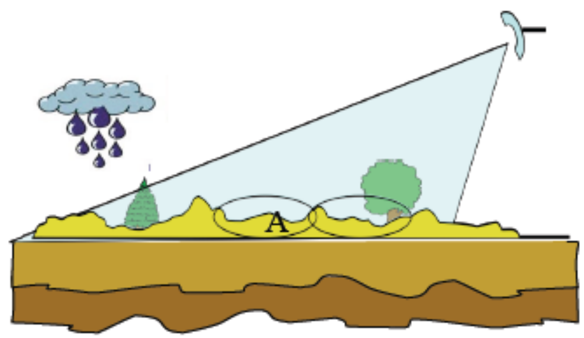
- Entered Phase B in January 2010
- Science Definition Team (SDT) selected in 2008
- Algorithms and Cal/Val workshop held in June 2009 and March 2010; Applications workshop September 2009
- Now completing mission trade studies

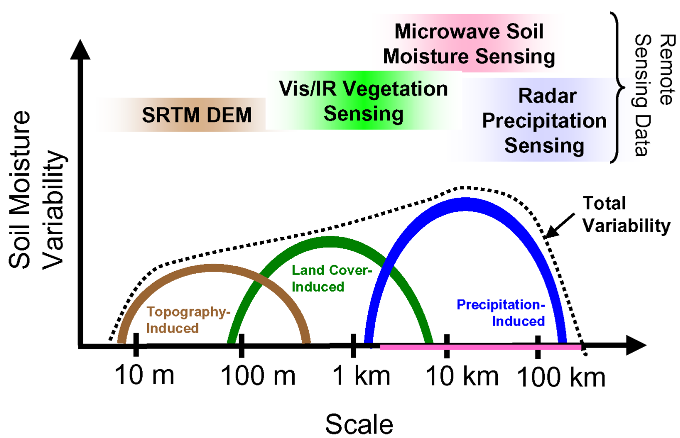
Development Objectives:

- Just completed KDP-B; now in Phase B
- Phase B will focus on further trade studies, risk reduction, requirements and interface maturation
- 2nd Algorithms Workshop March 2010
- Cal/Val Field Campaigns Summer 2010 (Oklahoma, Canada, Australia)
- Launch planned for late 2014

Technology Background

Soil moisture Sensing Controller And oPtimal Estimator (SoilSCAPE): develop technologies for near real-time validation of spaceborne soil moisture estimates





The Challenge for SMAP Validation:

- SMAP's radar and radiometer will measure soil moisture with different spatial resolutions
- Soil moisture varies on multiple spatial scales
  - ⇨ O(10 m) due to vegetation cover and topography
  - ⇨ O(100 m) due to topography and soil type
  - ⇨ O(1000m) due to cloud cover and precipitation
- Deploying validation sensors at all scales and with high density is infeasible
  - Old paradigm doesn't work
- Need smart and adaptive time and space sampling
  - Balance cost and accuracy
- This problem is at the boundary of the conventional instrument domain and information technologies domain

Objectives

SoilSCAPE Objectives are:

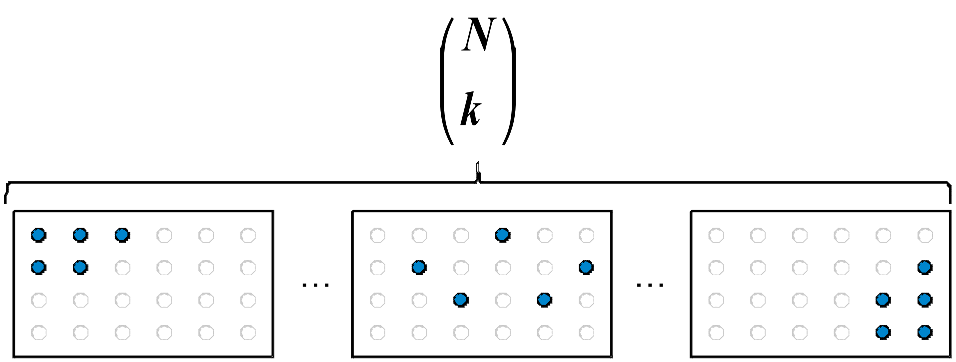
1. Optimal design of sensor node **placement and scheduling controller** based on modeled and measured soil moisture spatial and temporal statistics

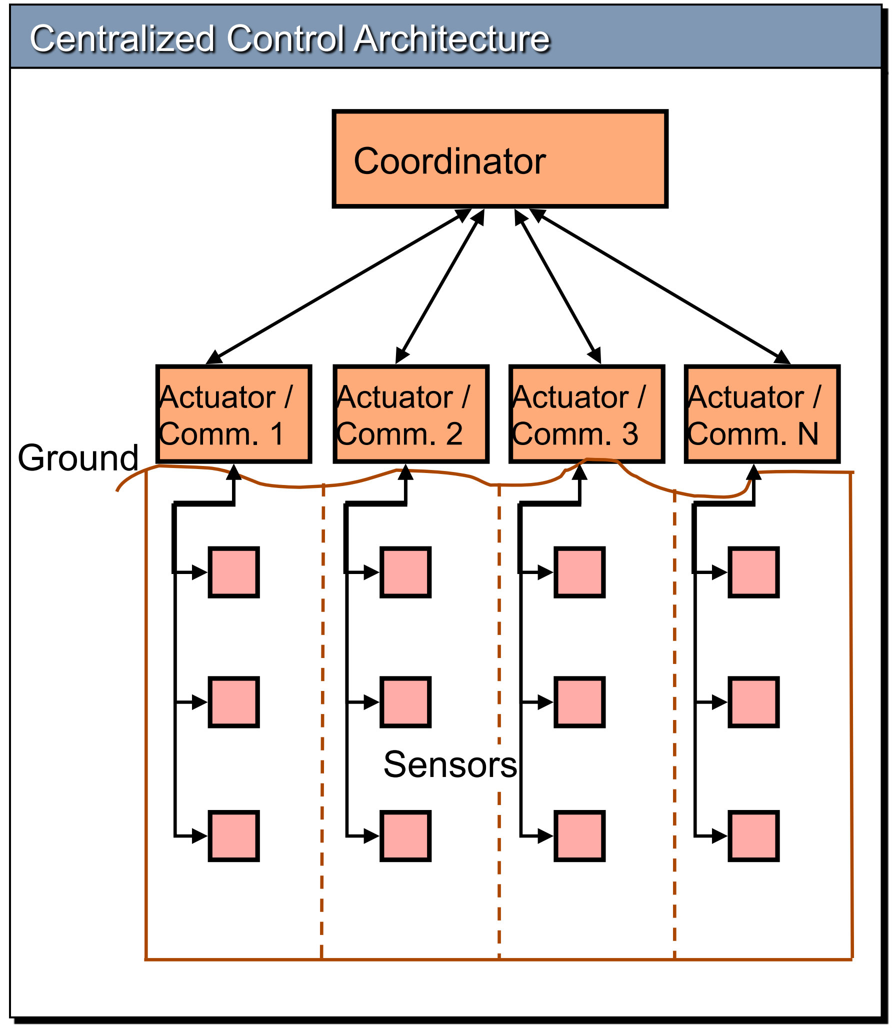
2. Derivation of large-scale remote sensing estimates of heterogeneous soil moisture, compatible with ground sensor network estimates of true mean of soil moisture field via a **landscape simulator**

3. Design and implementation of large-scale wireless **communication & actuation system** to configure sampling within the in-situ sensor network and to produce estimates of the soil moisture field mean

Control System overview:

- Design of sensor node placement and scheduling based on soil moisture spatial and temporal statistics
  - ⇨ Implemented through a “centralized control” architecture
  - ⇨ Initially will decouple sensor placement solution from sensor scheduling solution





Control System overview:

- Sensor placement assuming continuous-time sampling
  - Conducted studies on simulated data
  - Developed a cluster-based placement scheme
- Field mean estimation problem assuming a fixed placement
  - With a fixed placement, computed scheduling policies for sensors
  - Modified the estimation policy to estimate the mean value of soil moisture over the field of interest
- Methodology to address the joint placement and scheduling problem
  - Had previously developed scheduling controller independent of placement
  - Placement and scheduling problem are inter-related
  - Optimal placements should take into account the dynamic scheduling costs
  - Identified a methodology that incorporates the dynamic aspects of scheduling into the static placement problem

Sensor placement algorithms using simulated data

- tRIBS (TIN-based real-time integrated basin simulator; TIN: triangulated irregular network) is a landscape hydrology simulation tool developed at MIT; has been used here to investigate space/time soil moisture dynamics
- Assuming perfect measurements in time, address sensor placement as a stand-alone problem
- Exploit special properties of data
- Investigate a cluster-based placement scheme to better exploit the data features
- Experimental results

The stand-alone sensor placement problem

- A field with N possible locations to place sensors $V = \{v_1, v_2 \ldots v_N\}$
- Signal to be sensed assumed to be a random process; random variable X_i at location v_i
- If we place a sensor at v_i we observe perfectly X_i; otherwise we need to provide an estimate $\hat{X}_i$

- Want to select K locations to place sensors

$$\text{(P)} \qquad A^* = \arg\min_{A \subset V, |A|=K} E[\mathrm{err}(X_V, \hat{X}_V)]$$

- *err*() is some error measure; most commonly used is the MSE

$$E[\mathrm{err}(X_V, \hat{X}_V)] = \| X_V - \hat{X}_V \|^2$$

- This is a joint optimization: simultaneously determine the best subset and the best estimate
- Can limit the solution space, e.g., only consider linear estimates

One very commonly used approach

- Assume the underlying spatial random process is Gaussian
  - The best estimate for an unobserved location is the conditional mean of a Gaussian random variable, a linear estimator

$$\hat{x}_V = u_V + \Sigma_{VA} \Sigma^{-1}_{AA} (x_A - u_A)$$

  - Can use metrics like entropy and mutual information as alternative objective functions (though an approximate one to MSE) for subset selection

(MaxEN)

$$A^* = \arg\min_{A \subset V, |A|=K} H(X_{V \setminus A} \mid X_A)$$

$$= \arg\max_{A \subset V, |A|=K} H(X_A)$$

(MaxMI)

$$A^* = \arg\max_{A \subset V, |A|=K} \mathrm{MI}(X_{V \setminus A}, X_A)$$

$$= \arg\max_{A \subset V, |A|=K} H(X_{V \setminus A}) - H(X_{V \setminus A} \mid X_A)$$

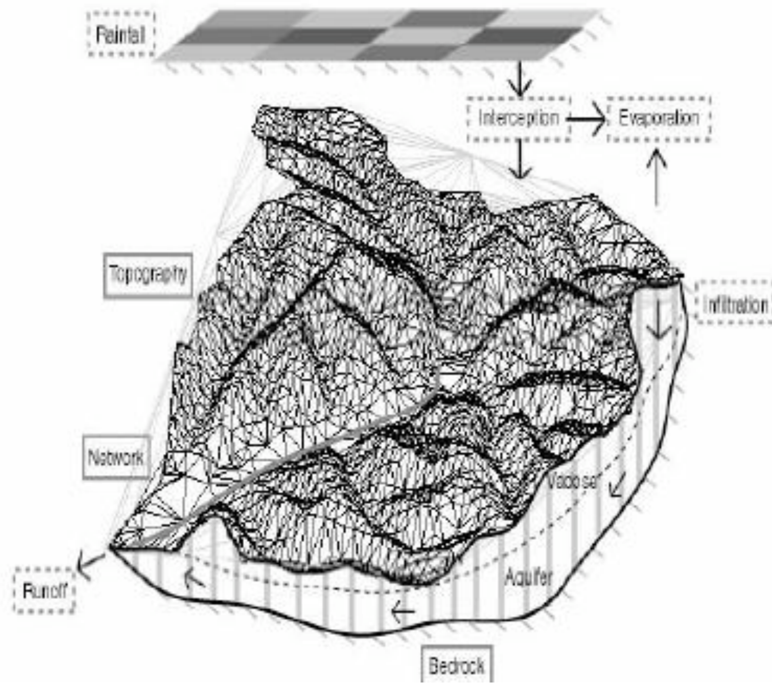
- They remain NP-hard
- Simple greedy algorithms shown to have good performance

How the greedy algorithms work

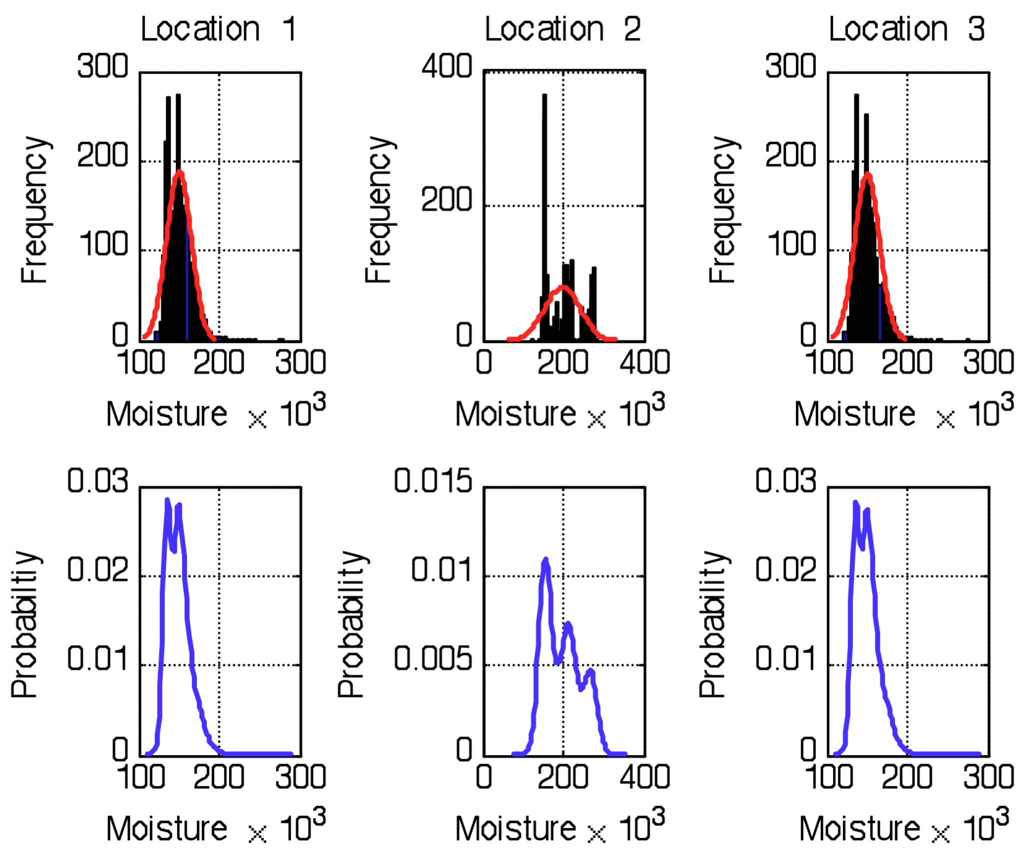
- Use certain training data to compute the mean, variances, and covariances at (and among) all locations
- Greedy placement:
  - At each step *t*, select one location that maximizes EN/MI (or minimizes MSE) given the set of locations already selected
  - Repeat till we have selected *K* locations
- Estimation/Prediction:
  - Conditional mean, also known as Gaussian regression
- How well does the Gaussian assumption hold for soil moisture data?
- How does this affect the design of good sensor placement and field estimation algorithms?

tRIBS simulation of soil moisture fields

- A 2km x 2km basin with 2400 locations (9 depths each) on a regular square grid
- Over a three-month period (simulated time), one snapshot per hour, a total of 2208 snapshots used in our experiments



Properties of the data: is the (surface) soil moisture process Gaussian?

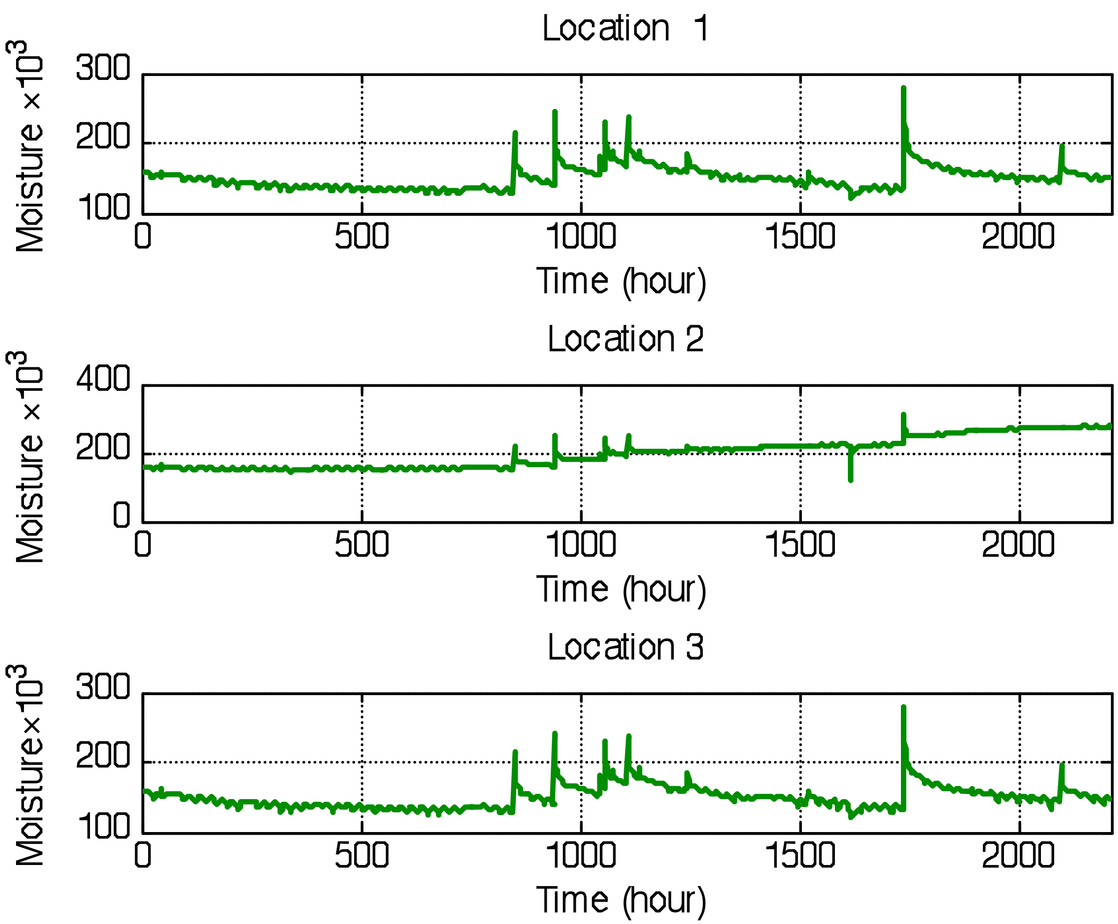


- Three randomly selected locations
- Surface soil moisture only
- Moisture readings amplified 1000x

- Top: histograms of moisture at these locations (black) and the estimated Gaussian kernel (red)
- Bottom: estimated pdf

- Observation: these are clearly non-Gaussian

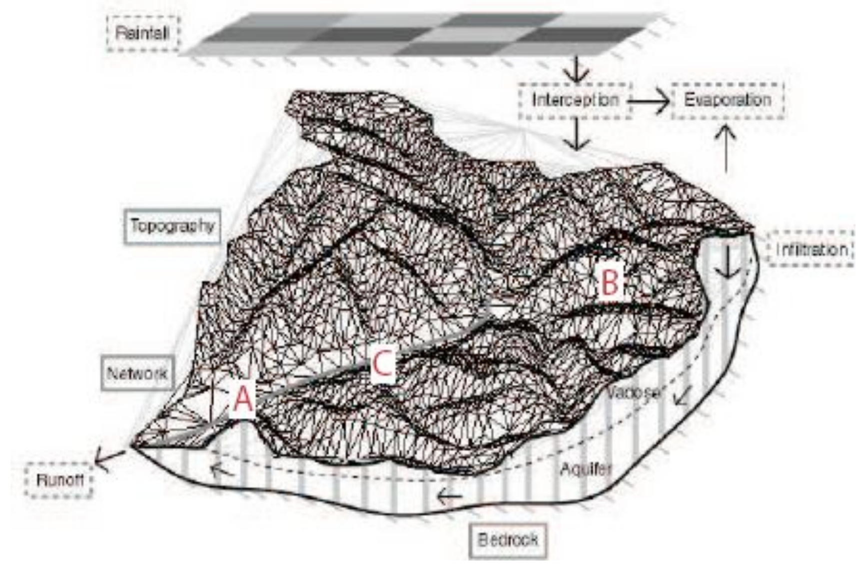
Temporal changes at these locations

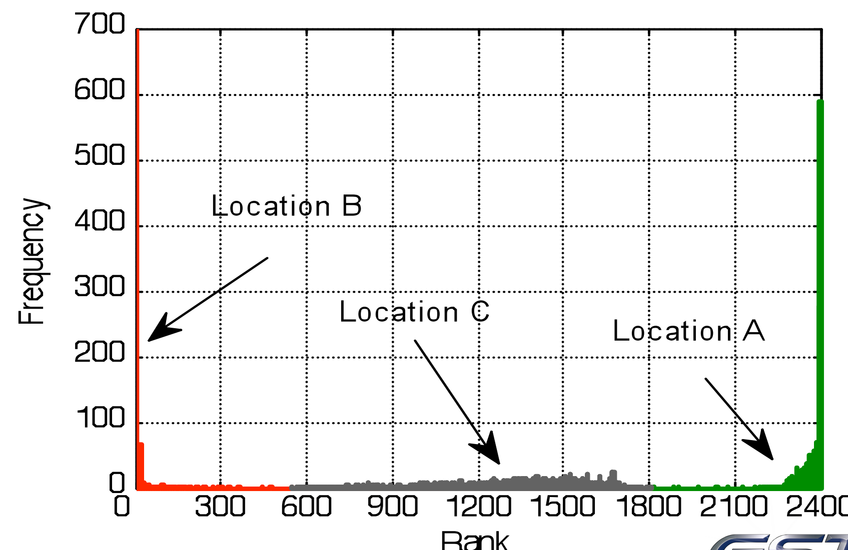


- The same three locations as in previous slide
- Figures show the change over time at these locations
- Figures show, qualitatively, how soil moisture is correlated between them
- Spikes correspond to rain events

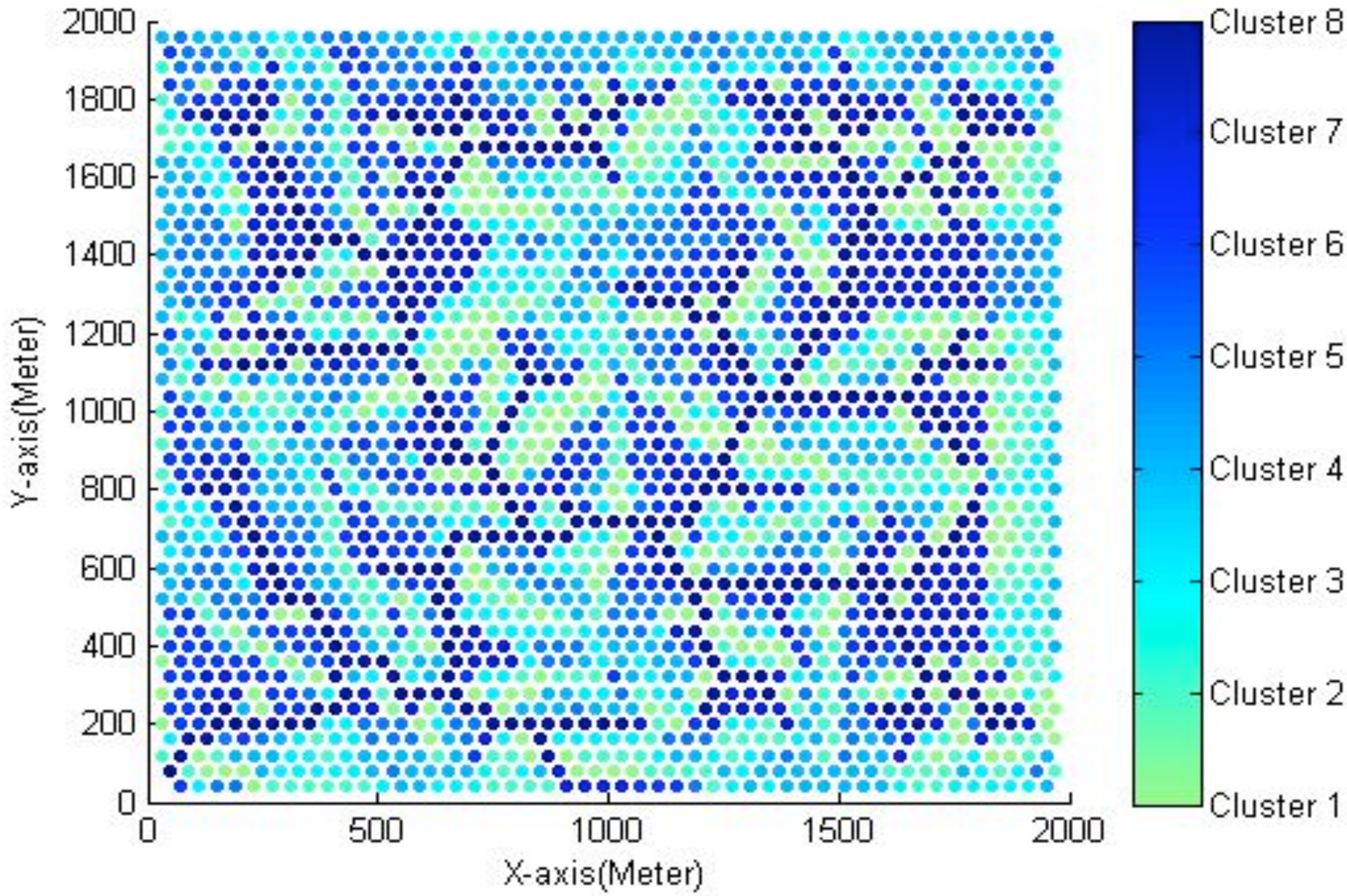
Observations

- Locations with similar features (soil type, vegetation cover, etc.) will show high correlation
- Most of these are relatively stable features over time
- May expect relative soil moisture values to hold steady even as absolute values vary over time
- Shown in bottom figure: histogram of a location's numerical rank in each snapshot





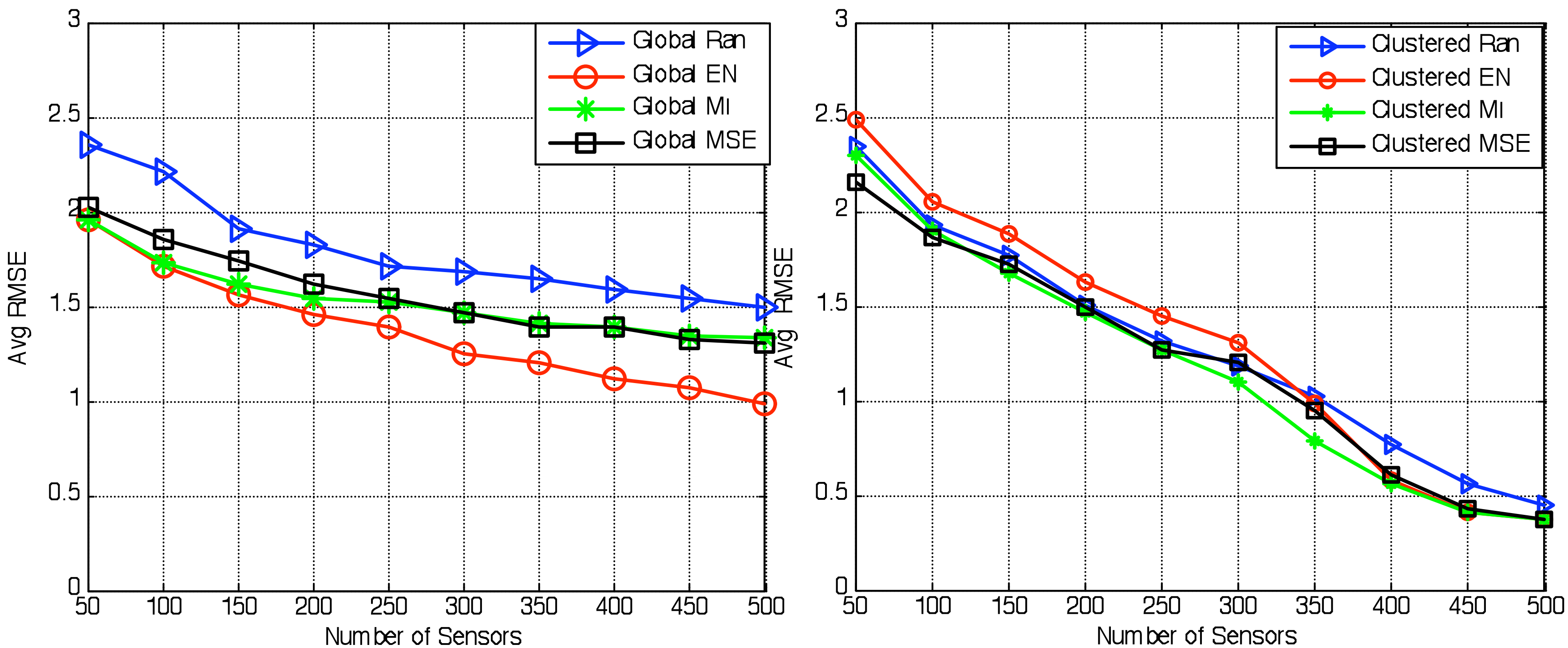
What does the clustering look like (W=8)



Placement using coarse-grained ordering

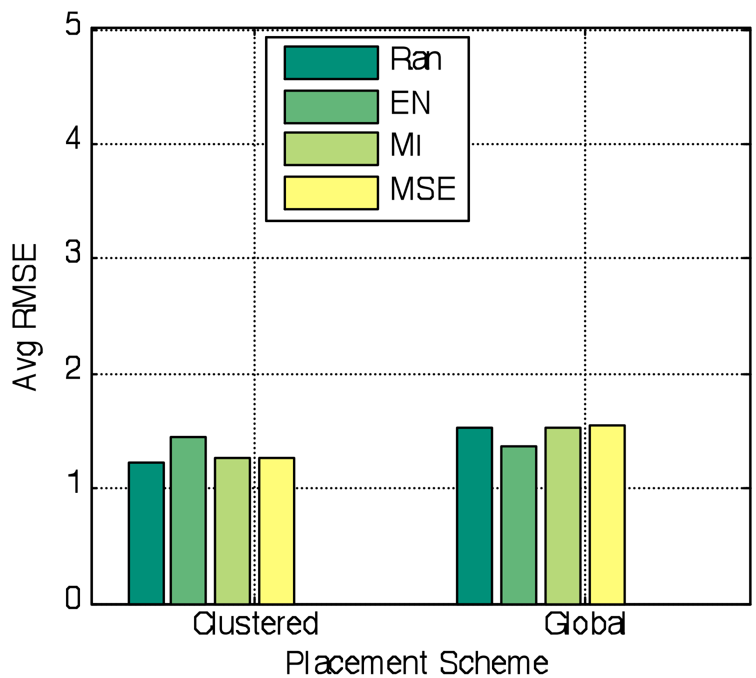
- Use a subset of simulated data for training, compute the mean, variances, and covariances at (and among) all locations
- Sensor placement:
  - Solve the placement problem independently for each cluster
  - Can allocate more sensors to clusters with higher average moisture (variance) levels
  - Place K_{i} sensors in cluster *i*, with $\sum_{i=1}^{W} K_i = K$
  - Within each cluster can use any existing scheme (e.g., max EN/MI)
- Field estimation:
  - Will use Gaussian regression

Clustered vs. global placement: size of the selected subset

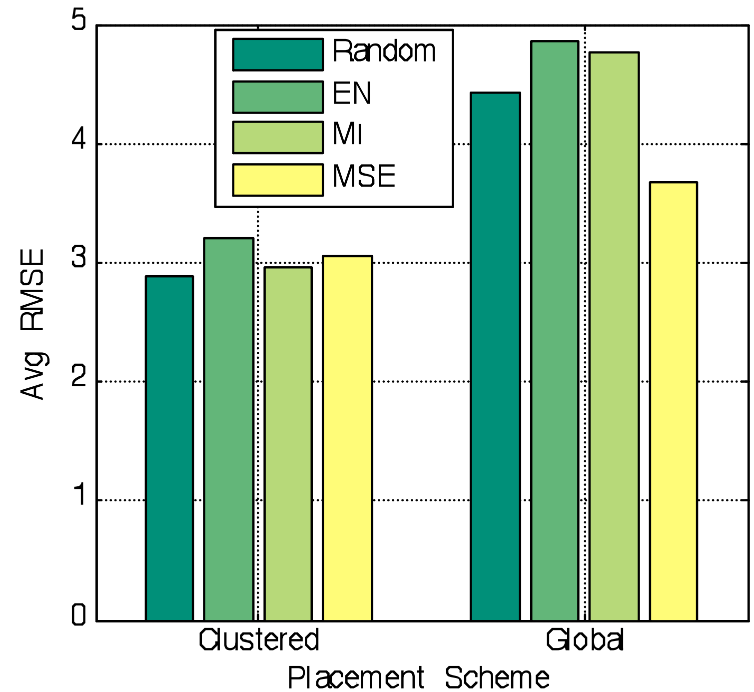


- Using the first 1500 snapshots for training and the last 700 for testing (out of 2208)
- Clustered schemes show advantage when placing sensors K>=200

Clustered vs. global placement: performance improvement



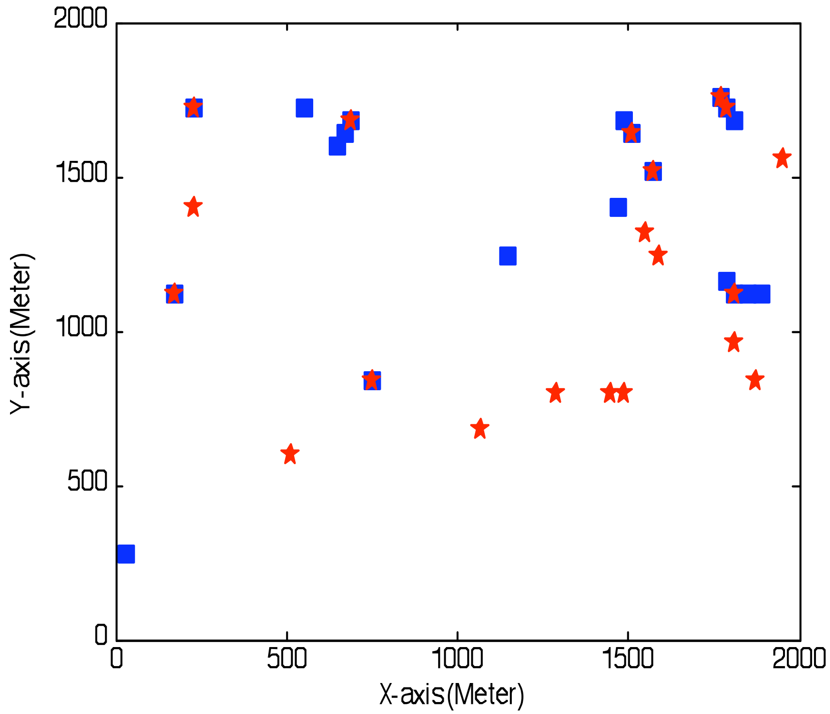
1500 training snapshots



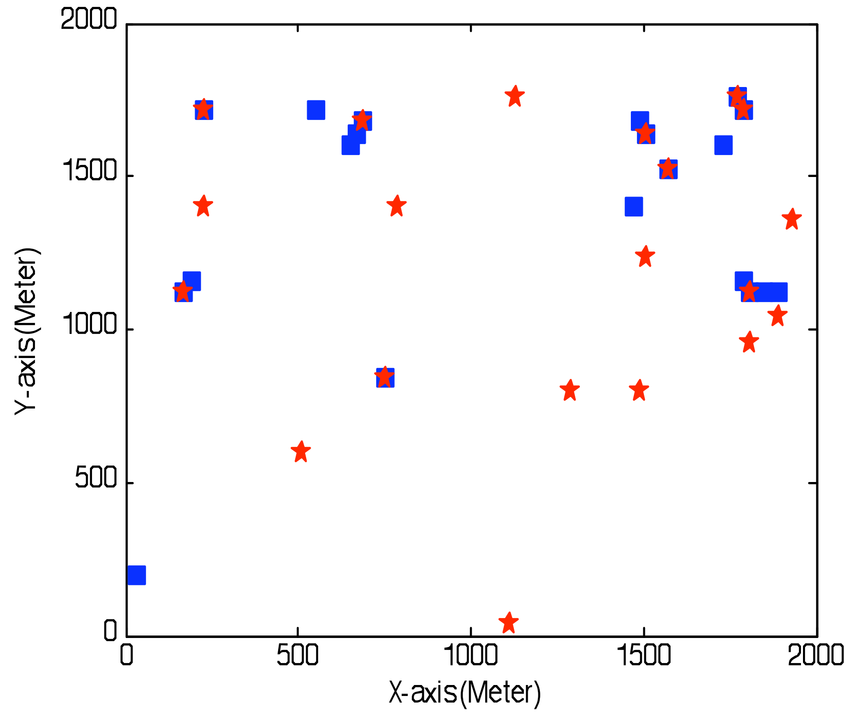
1000 training snapshots

- 250 sensors are placed
- Regardless of training amount, clustering results in better performance

A quick glance at the actual placement (20 sensors) under different schemes: Many features are similar between EN and MI approaches



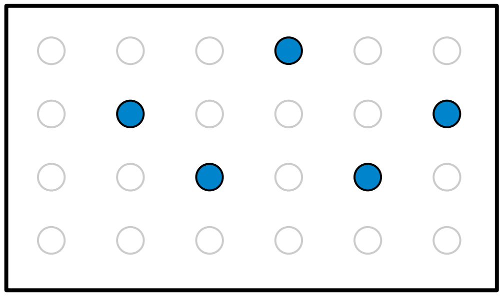
- Global MaxEN (blue)
- Clustered MaxEN (red)



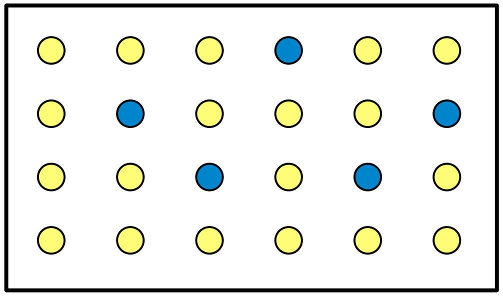
- Global MaxMI (blue)
- Clustered MaxMI (red)

Non-Gaussian Dynamic Mean Estimation with a Fixed Placement

- Currently, the scheduling objective is to estimate soil moisture evolution at *K* fixed lateral locations using sensors placed at those locations



- However, we are also interested in using the *K* sensors to estimate a mean value of the soil moisture over the area of interest



- Statistics are not Gaussian; estimation costs are dynamic

Mean Estimation : Solution Methodology

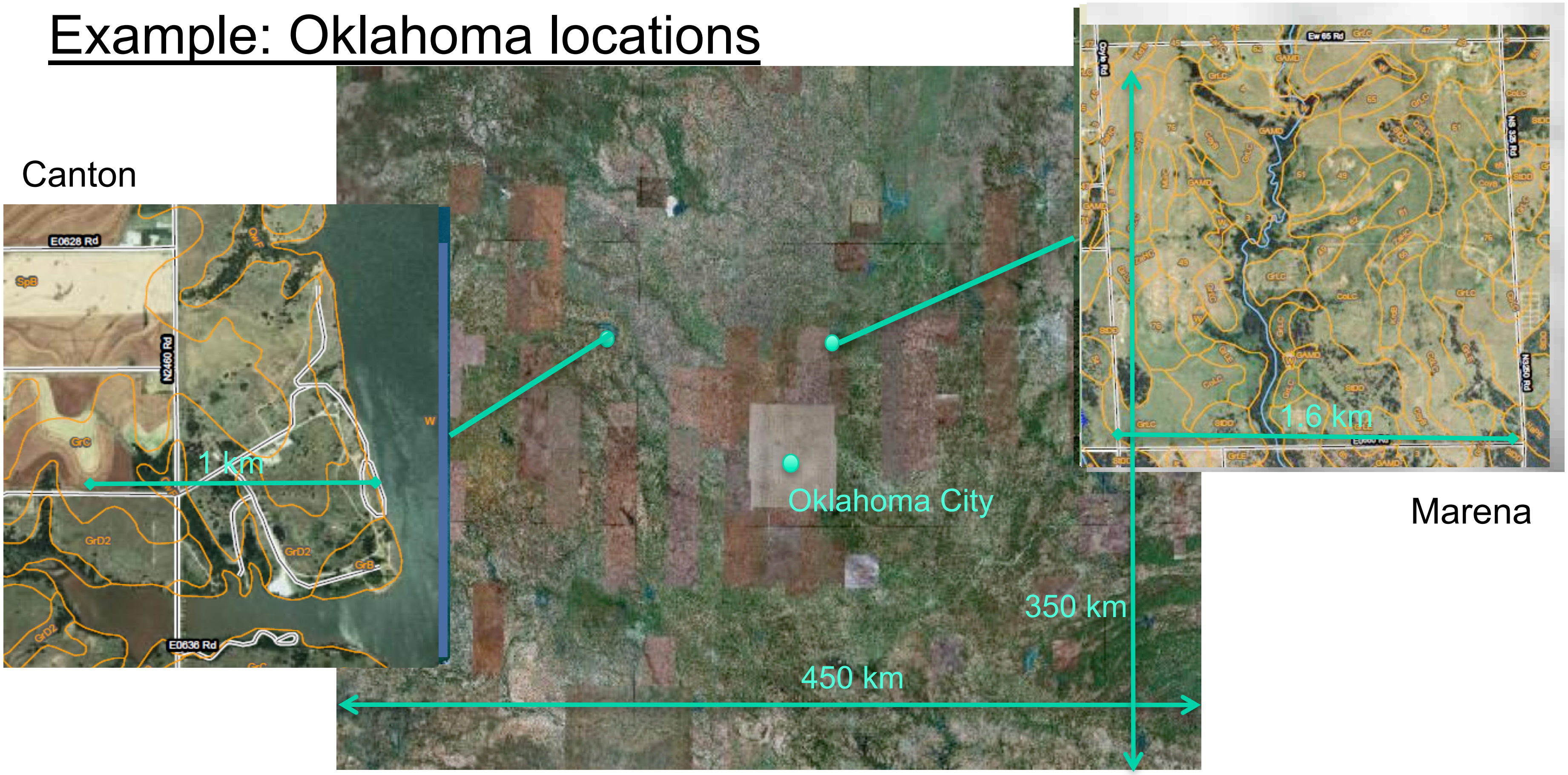
- Solve the scheduling problem independently for each sensor location
- Exploit the correlation among soil moisture values at different locations to find local estimates, $\hat{X}_t(location)$ using past measurements from all locations (Joint Estimation)
- Find the joint statistics of the field mean *M* and the local values of soil moisture at sensor locations
- Use the joint statistical model of the soil moisture at sensor locations and the field mean to convert the local estimates to a mean field estimate $\hat{M}_t$
- Performance of scheduler (dynamic problem) depends on placement (static problem), and vice versa
- Currently developing solution of joint placement and scheduling problems

Landscape Simulator Overview

- Proof-of-concept heterogeneous landscape simulator
  - Developed architecture of simulator
  - Implemented unified multi-layered multi-species vegetation model adaptable to various land cover types
  - Created a data base of input files using land cover types of NLCD 2001
- Visualization in Google Earth
  - Layers of information co-registered on whole Earth
- Preliminary aggregation studies
  - Investigated how coarse-resolution remote sensor measurements relate to finer-resolution measurements

Example: Oklahoma locations

Canton



Marena

Area around Oklahoma City, OK

1. Google Earth
2. Land cover type from NLCD 2001
3. Digital Elevation Map (DEM) / National Elevation Dataset (NED)
4. Soil type from USDA

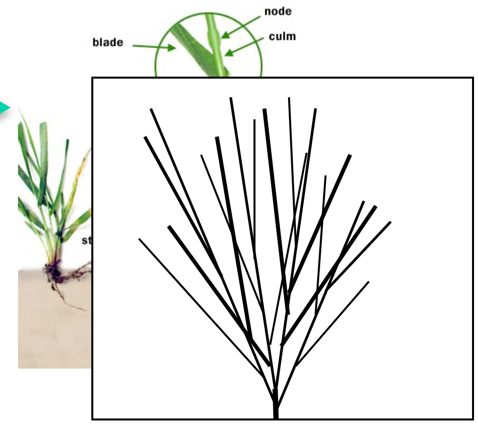
From land cover type to model

NLCD 2001 Land Cover Classification Legend

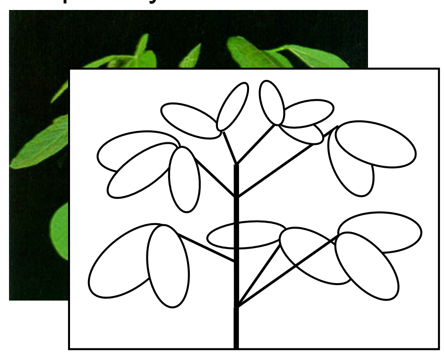
- 11 Open Water
- 12 Perennial Ice/Snow
- 21 Developed, Open Space
- 22 Developed, Low Intensity
- 23 Developed, Medium Intensity
- 24 Developed, High Intensity
- 31 Barren Land
- 41 Deciduous Forest
- 42 Evergreen Forest
- 43 Mixed Forest
- 51 Dwarf Scrub*
- 52 Shrub/ Scrub
- 71 Grassland/ Herbaceous
- 72 Sedge/ Herbaceous *
- 74 Moss *
- 81 Pasture Hay
- 82 Cultivated Crops
- 90 Woody Wetlands
- 95 Emergent Herbaceous Wetlands

* Alaska Only

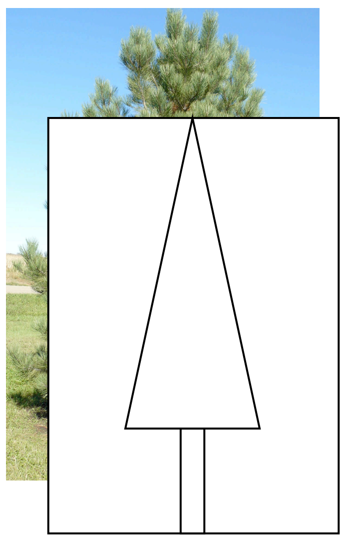
Grass/ Grassland



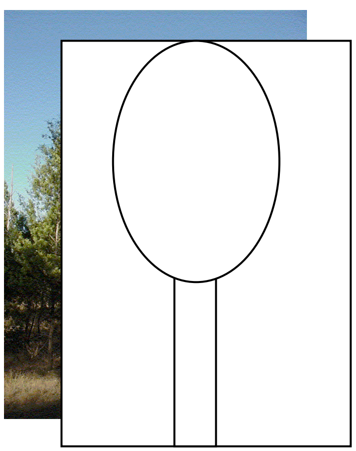
Crop – soybean



Evergreen tree



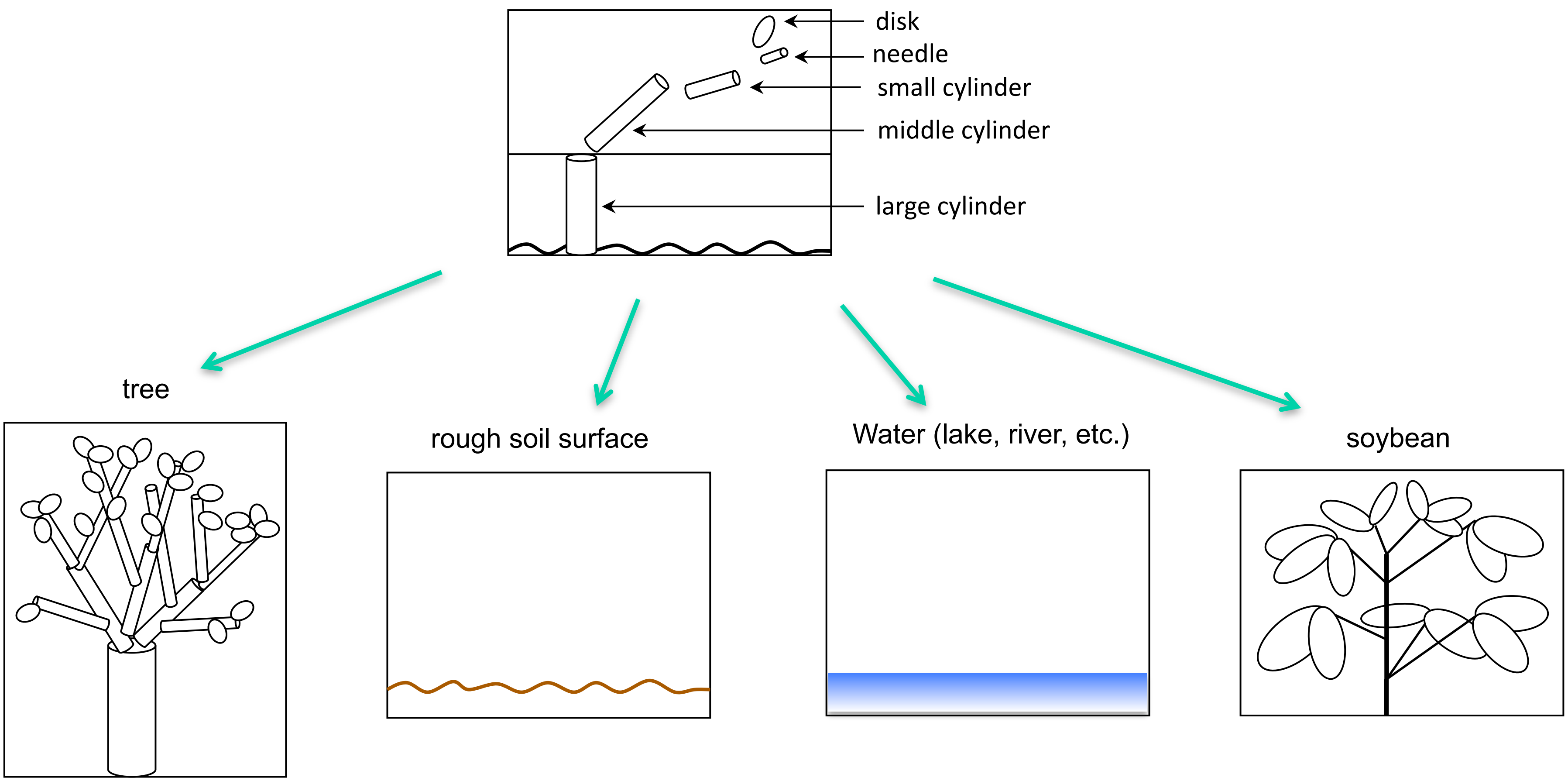
Deciduous tree



Available information & ancillary data is used to adapt model to specific landscape (via parameter input file)

- **Current**: selection of pre-determined input files based on land cover type
- **Prospective**: generate input file based on more diverse combinations of ancillary data

From land cover type to model: One model



Model is general enough to represent various land cover types

Model:

- Can simulate multi-layer vegetation and multiple vegetation types simultaneously
- Builds on existing single species forest model (Durden et al., 1989)
  - Scattering from layers of arbitrarily oriented dielectric cylinders above a rough dielectric surface
  - Extended to multi-layer multi-species discrete scatterer model with rough surface representing ground
  - Simulation of full Stokes matrix and polarization signature
- Analysis is based on wave theory; distorted Born approximation
  - Scattering and transmission matrices are formed, from which Stokes matrices are calculated

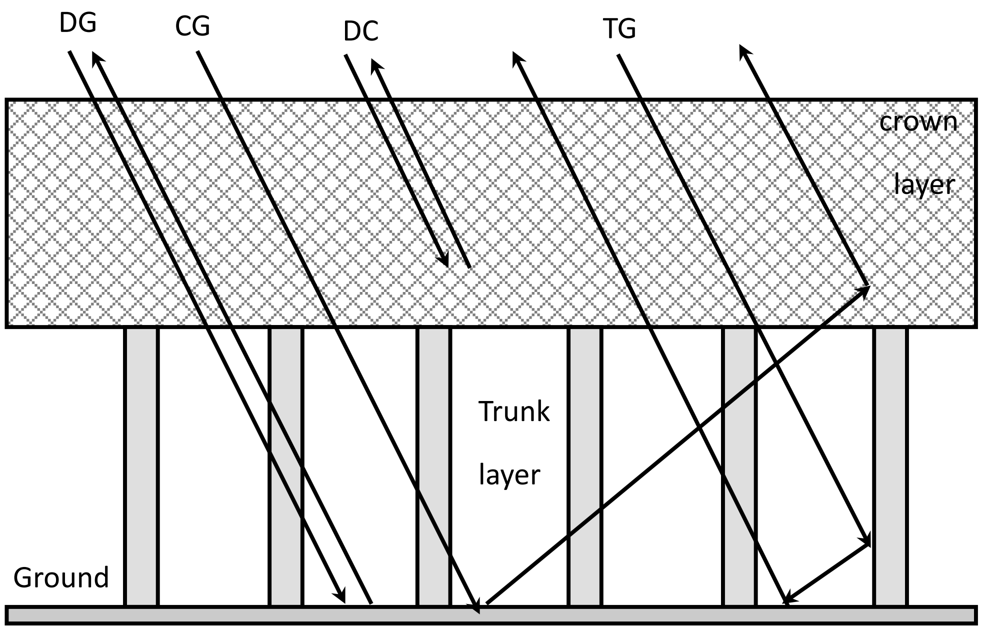
Model (2):

Species-specific parameters:

- Set of 27 parameters define geometry and structure of single species:
  - Soft- or hardwood
  - Dielectric characteristics of leaves, branches, trunks, soil
  - Densities, lengths, radii
  - Probability density function (pdf's) for orientation of branches, trunks

- Allometric relations exist for different species, ideally unique relationships
  - Knowledge of plant anatomy results in species-specific relations for modeling

Single species geometry:



Model considers four scattering mechanisms:

- Direct backscatter from crown layer (DC)
- Direct backscattering from ground (DG)
- Specular crown scattering followed by ground reflection (CG) modified for surface roughness
- Specular trunk scattering followed by ground reflection (TG)

Single species setup:

- Total Stokes matrix found by summing the matrices of the different scattering mechanisms

$$M_{Total} = M_b + T_b T_t M_g T_t T_b + T_b T_t M_{bg} T_t T_b + T_b T_t M_{tg} T_t T_b$$

M: Stokes matrix for backscattering

T: Stokes matrix for transmission through layer

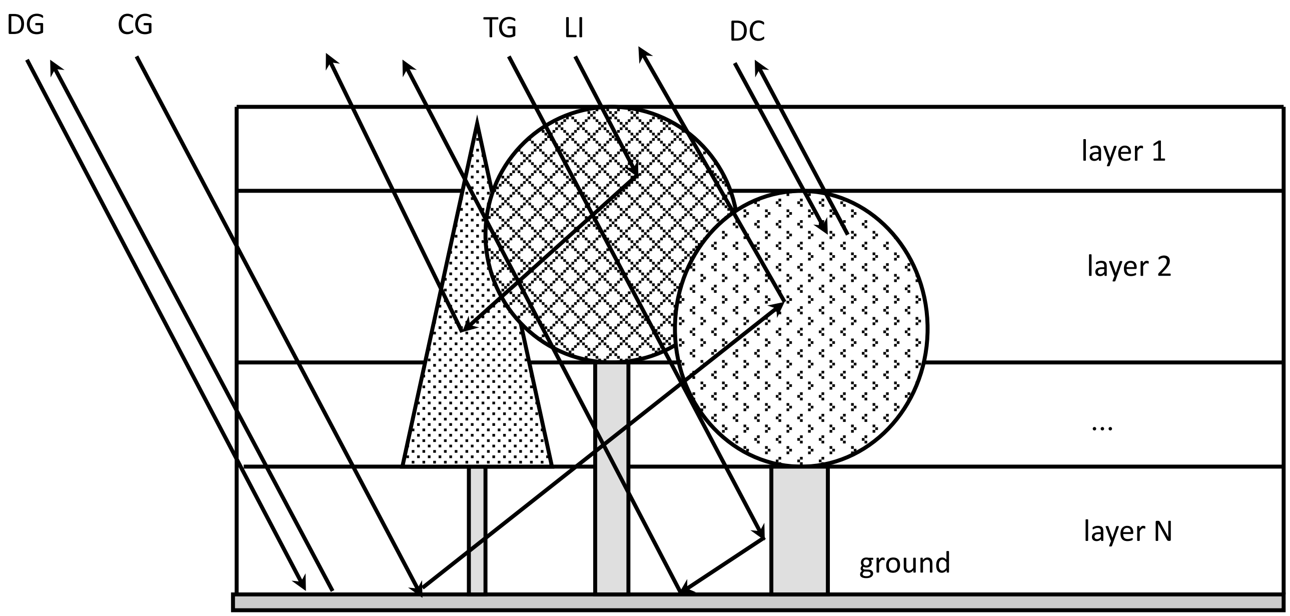
b: Branch layer

t: Trunk layer

g: Ground

Multi-species geometry:

- Introduction of more species will result in N layers
- Determination of layer composition, including that of overlapping layers is an important step
- Model proceeds with methodical calculation of layer scattering, attenuation and interaction



Multi-species setup:

M: Stokes matrix for backscattering
T: Stokes matrix for transmission through layer
b: Branch layer
t: Trunk layer
g: Ground

- Total Stokes matrix for multi-species model:

$$M_{Total} = M_b + M_g + M_{bg} + M_{tg} + M_{li}$$

$$M_b = M_{bL1} + T_{L1} M_{bL2} T_{L1} + T_{L1} T_{L2} \ldots T_{LN-1} M_{bLN} T_{LN-1} \ldots T_{L2} T_{L1}$$

$$M_g = T_{L1} T_{L2} \ldots T_{LN-1} M_{gLN} T_{LN-1} \ldots T_{L2} T_{L1}$$

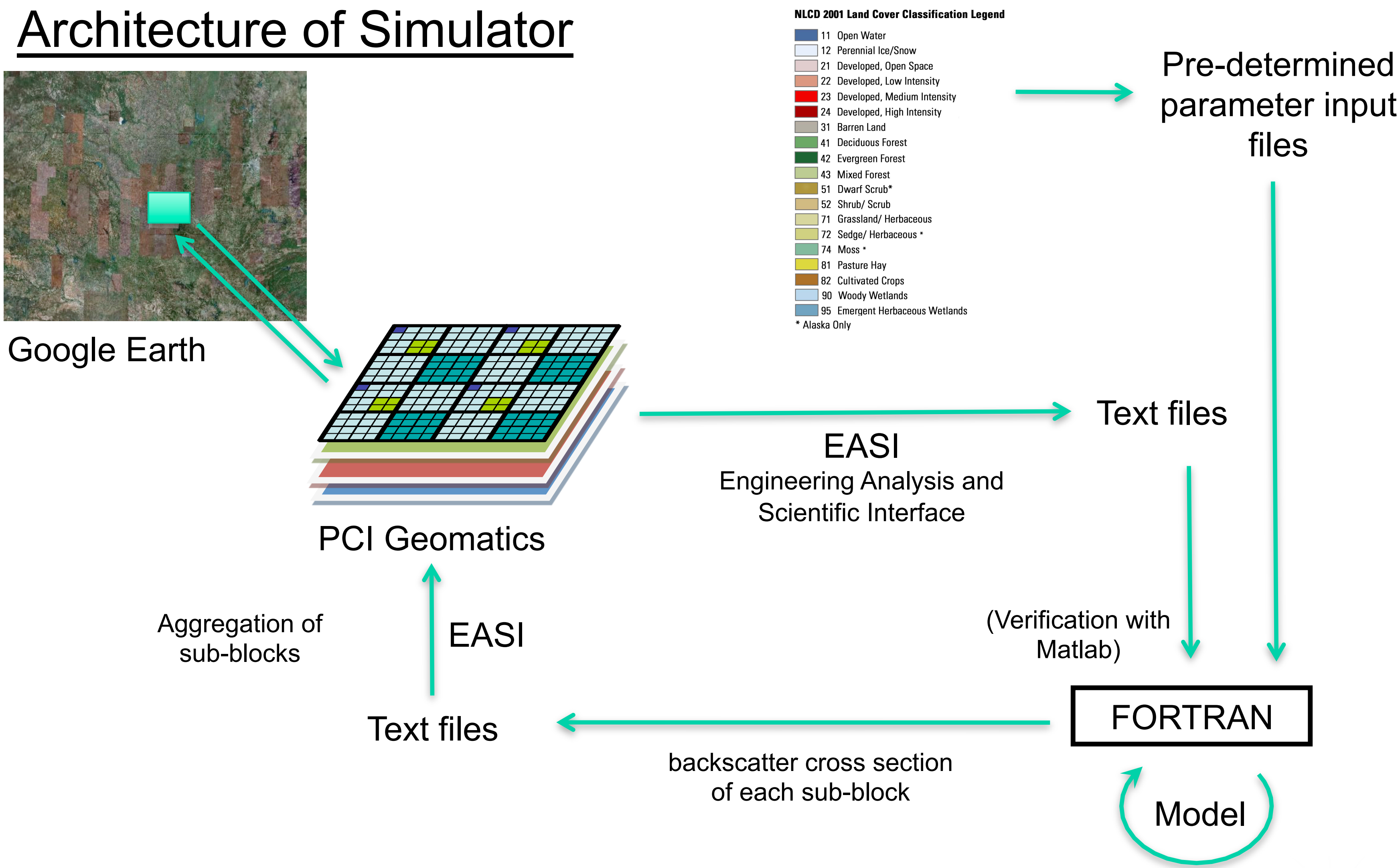
$$M_{bg} = T_{L1} T_{L2} \ldots T_{LN-1} M_{bg1LN} T_{LN-1} \ldots T_{L2} T_{L1} + T_{L1} T_{L2} \ldots T_{LN-1} M_{bg2LN} T_{LN-1} \ldots T_{L2} T_{L1} + \cdots$$

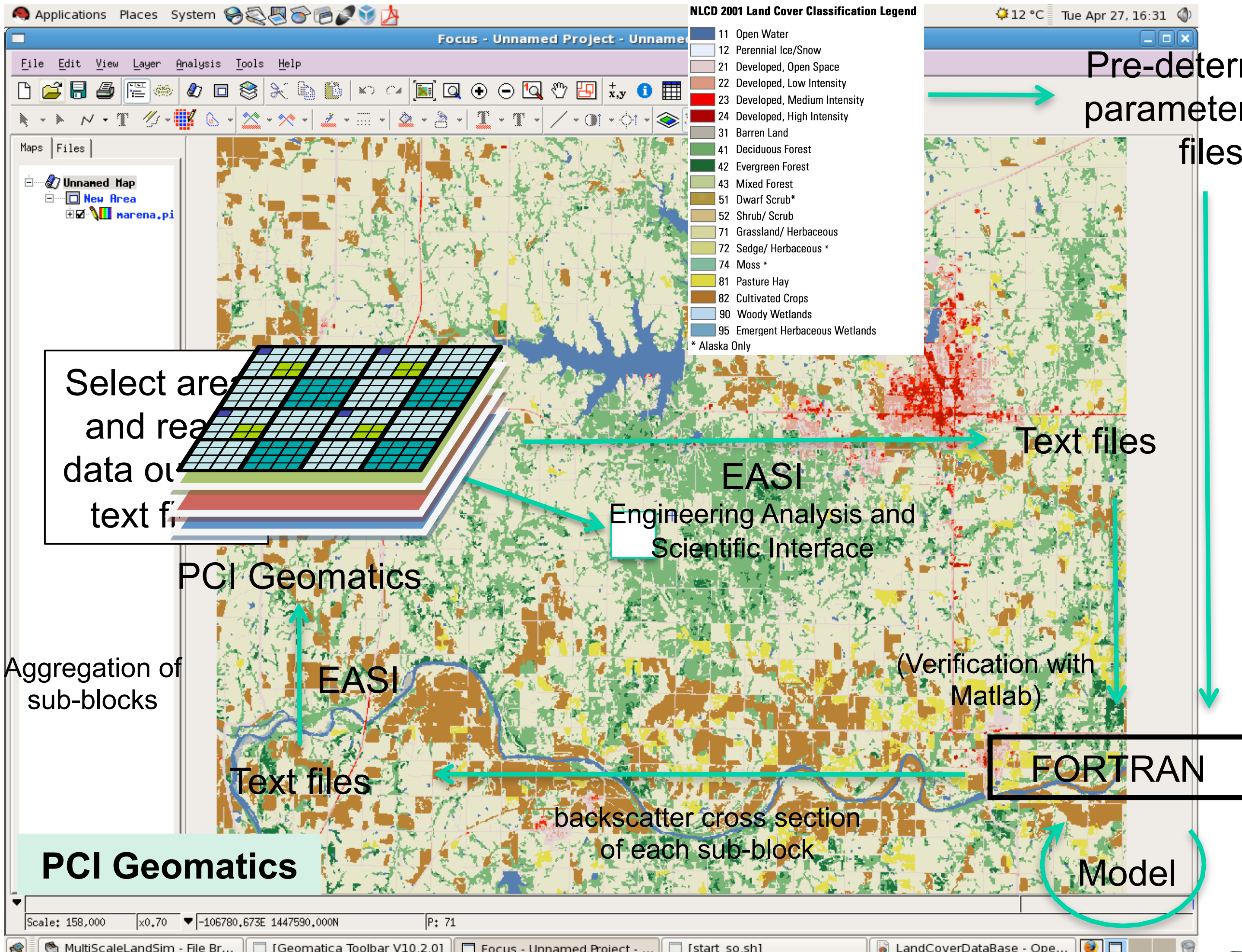
$$M_{tg} = T_{L1} T_{L2} \ldots T_{LN-1} M_{tg1LN} T_{LN-1} \ldots T_{L2} T_{L1} + T_{L1} T_{L2} \ldots T_{LN-1} M_{tg2LN} T_{LN-1} \ldots T_{L2} T_{L1} + \cdots$$

$$M_{li} = M_{liL1s1L2s2} T_{L1} + M_{liL1s1L3s2} T_{L2} T_{L1} + \ldots + T_{L1} M_{liL2s1L3s2} T_{L2} T_{L1} + \ldots + \ldots + T_{L1} M_{liL2L3} T_{L2} T_{L1} + \cdots$$

TL1, ...TLN: can contain combination of crown and trunk layer, depending on geometry

Architecture of Simulator

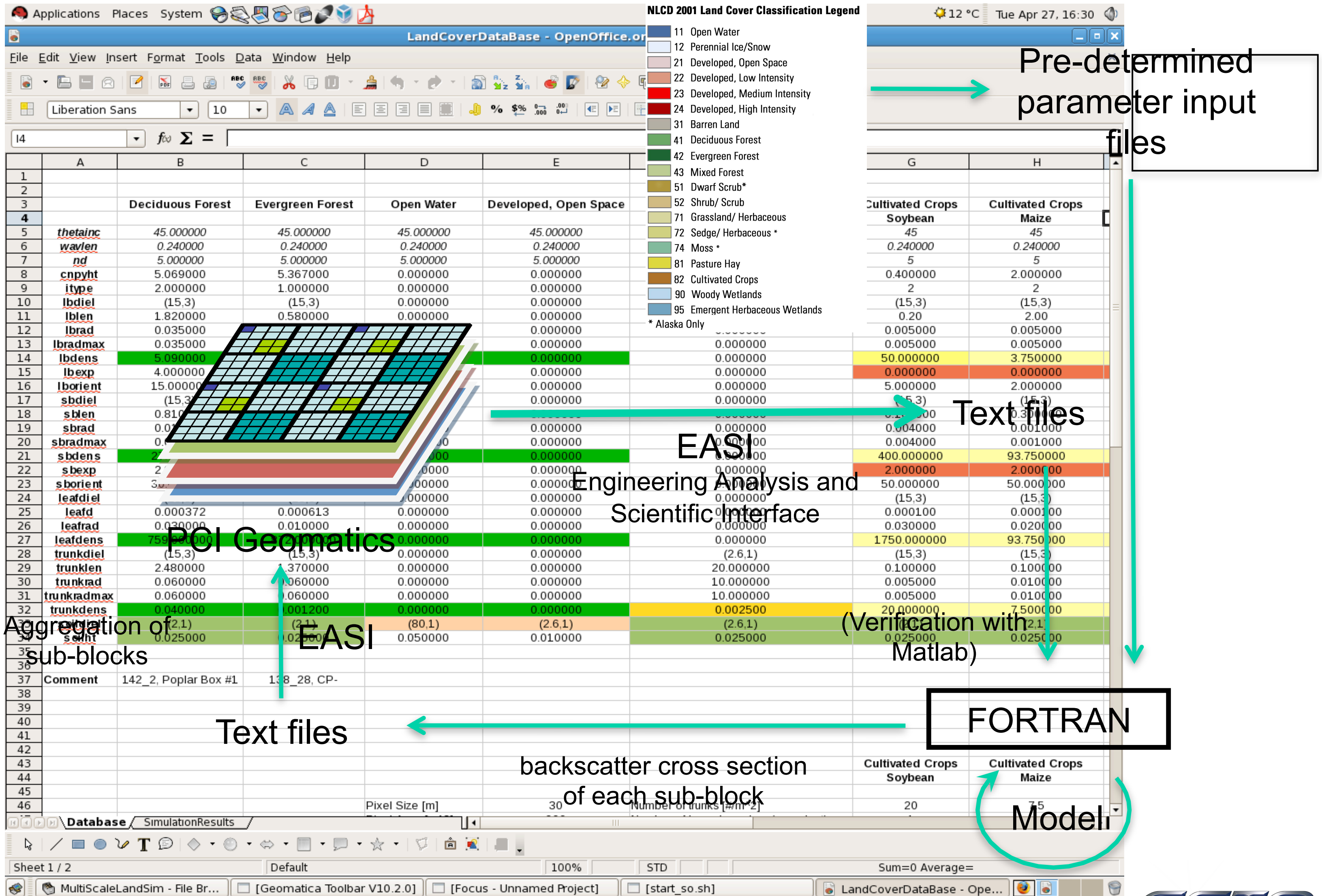



NLCD 2001 Land Cover Classification Legend
11 Open Water
12 Perennial Ice/Snow
21 Developed, Open Space
22 Developed, Low Intensity
23 Developed, Medium Intensity
24 Developed, High Intensity
31 Barren Land
41 Deciduous Forest
42 Evergreen Forest
43 Mixed Forest
51 Dwarf Scrub*
52 Shrub/ Scrub
71 Grassland/ Herbaceous
72 Sedge/ Herbaceous *
74 Moss *
81 Pasture Hay
82 Cultivated Crops
90 Woody Wetlands
95 Emergent Herbaceous Wetlands
* Alaska Only
Focus - Unnamed Project
Pre-determined parameter input files
Select area and read data out to text files
Text files
EASI
Engineering Analysis and Scientific Interface
PCI Geomatics
Aggregation of sub-blocks
EASI
(Verification with Matlab)
FORTRAN
Text files
backscatter cross section of each sub-block
Model
PCI Geomatics

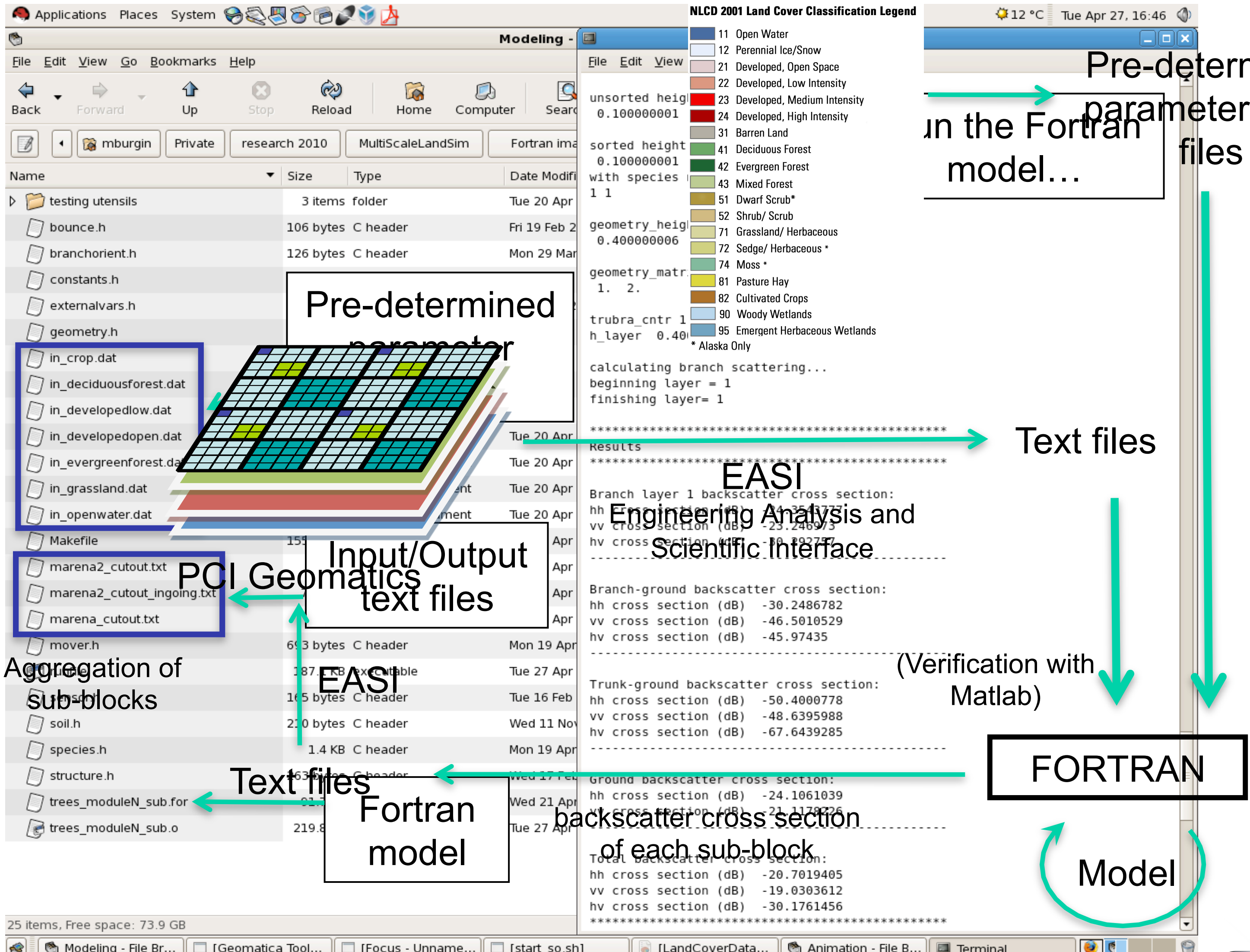
Control System

Landscape Simulator

Communication & Actuation

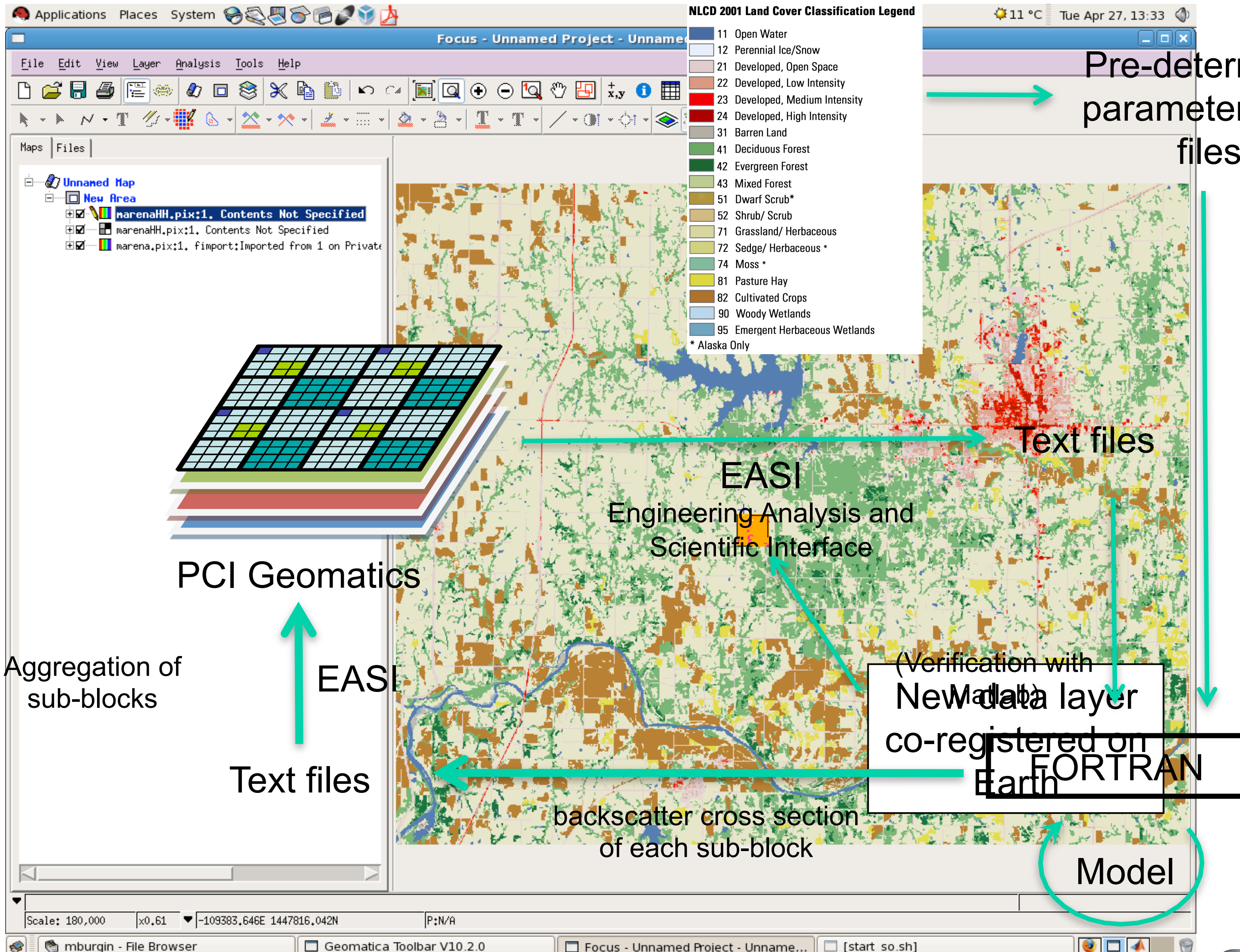


ESTO Earth Science Technology Office

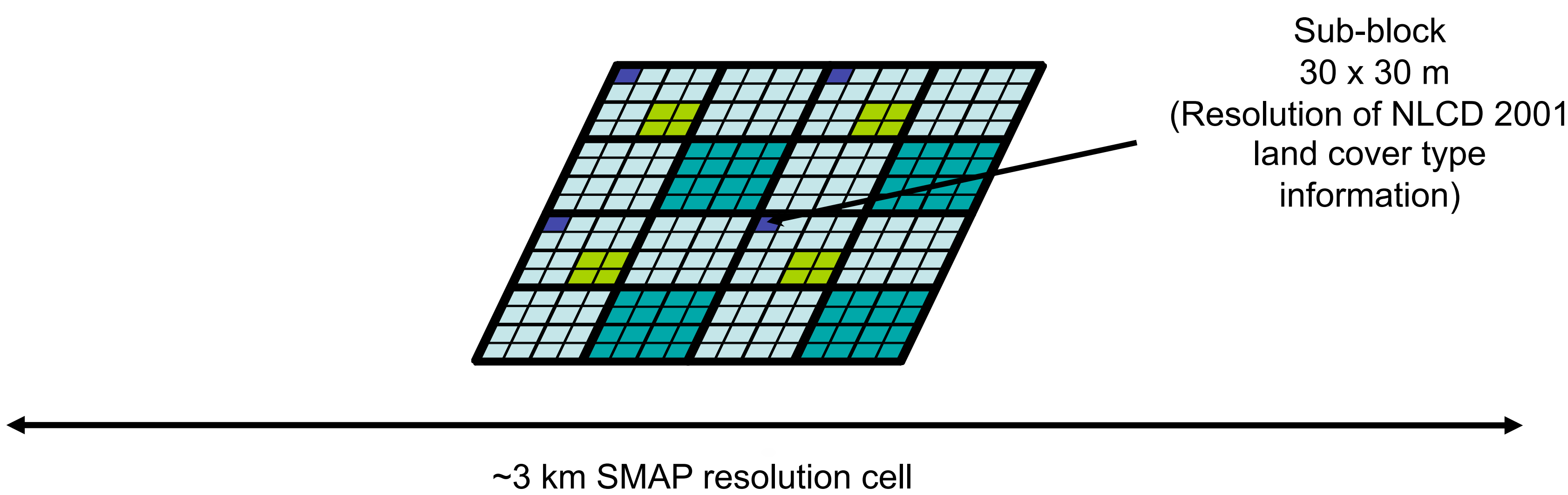


Earth Science Technology Office

Control System | **Landscape Simulator** | Communication & Actuation



Architecture, con't.

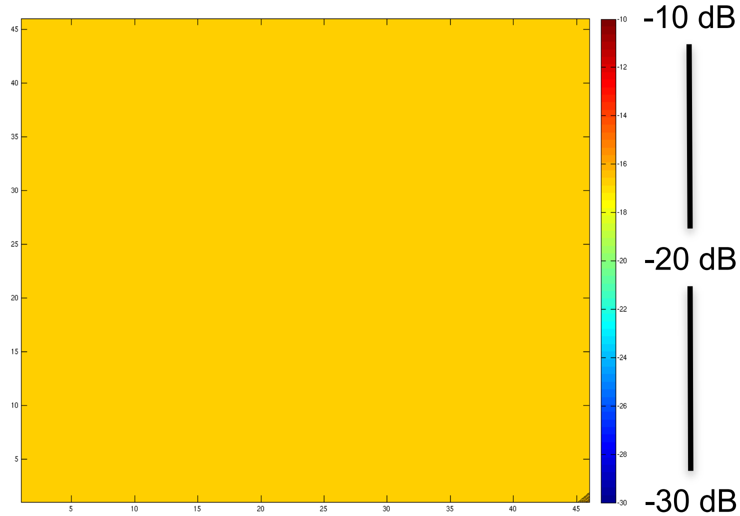


- Backscatter cross section from each sub-block is calculated
- Aggregation types can be investigated: blocks of 4 (light green), 16 (dark green), 64, etc., to achieve a statistically representative mean value for the backscattering cross section of the scene (e.g., for one SMAP pixel)
- Final result can be exported to PCI & Google Earth for visualization
- Will be used in the future as basis for *disaggregation* analysis for SMAP retrievals

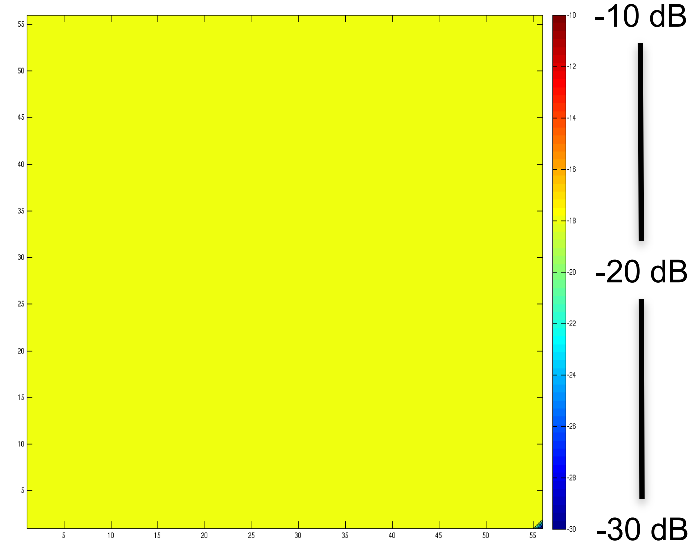
First results:

HH backscatter coefficient in dB for L-band

Canton



Marena

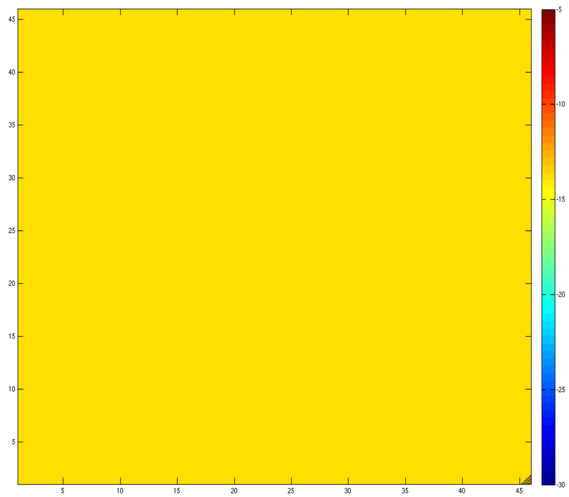


1. Google Earth
2. Land cover type from NLCD 2001
3. Backscatter coefficient
   a) Block of 1 x 1 sub-blocks
   b) Block of 2 x 2 sub-blocks
   c) Block of 4 x 4 sub-blocks
   d) Block containing all sub-blocks

- The final aggregation stage is similar to what SMAP radar sees
- Landscape detail is lost
- Current aggregation simply shows linear averaging; in reality SMAP data might correspond to some other (nonuniform or nonlinear) aggregation

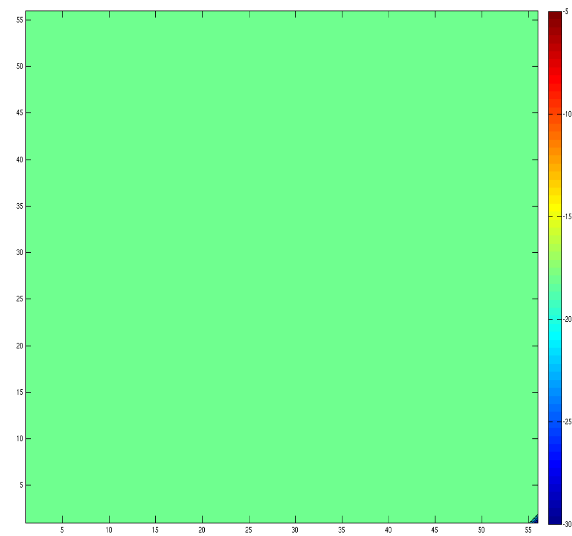
First results: VV backscatter coefficient in dB for L-band

Canton



-5 dB
-17.5 dB
-30 dB

Marena



-5 dB
-17.5 dB
-30 dB

1. Google Earth
2. Land cover type from NLCD 2001
3. Backscatter coefficient
   a) Block of 1 x 1 sub-blocks
   b) Block of 2 x 2 sub-blocks
   c) Block of 4 x 4 sub-blocks
   d) Block containing all sub-blocks

- VV results are similar to HH
- This aggregation simply shows linear averaging; in reality SMAP data might correspond to some other (nonuniform or nonlinear) aggregation

Ongoing Activities

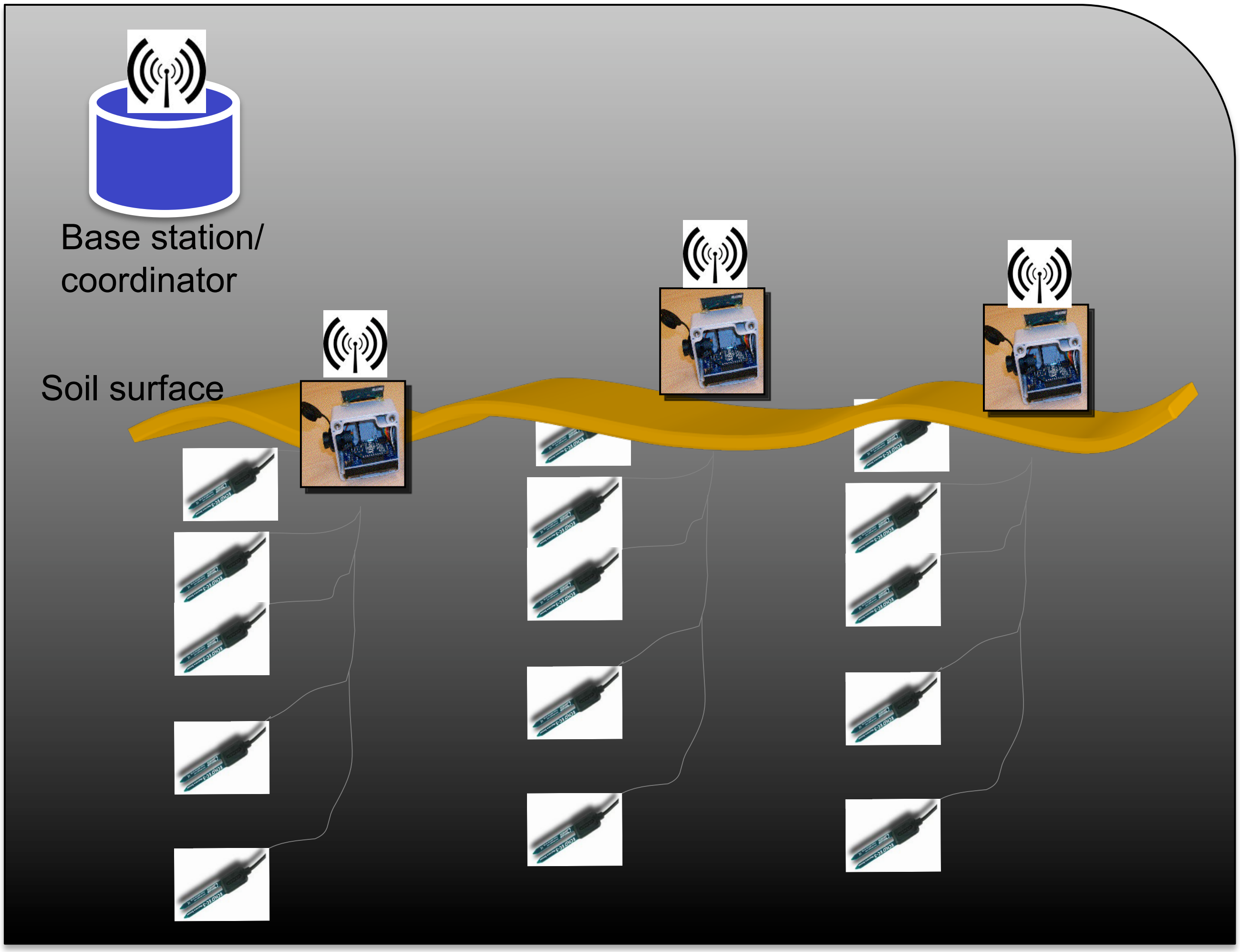
- Improve modeling and finalize landscape simulator:
  - Search for and integrate more available information
  - Use more ancillary data to build input files of sub-blocks to make modeling as realistic as possible
  - Integrate topography (slope, etc.)

- Investigate forward-mode multi-scale aggregation/disaggregation
  - Can a coarse resolution measurement be represented as a weighted sum of the fine-resolution ones? What statistical rules apply?
  - Study sensitivity of answer to above question to perturbations in soil moisture

Wireless Comm and Actuation System Overview

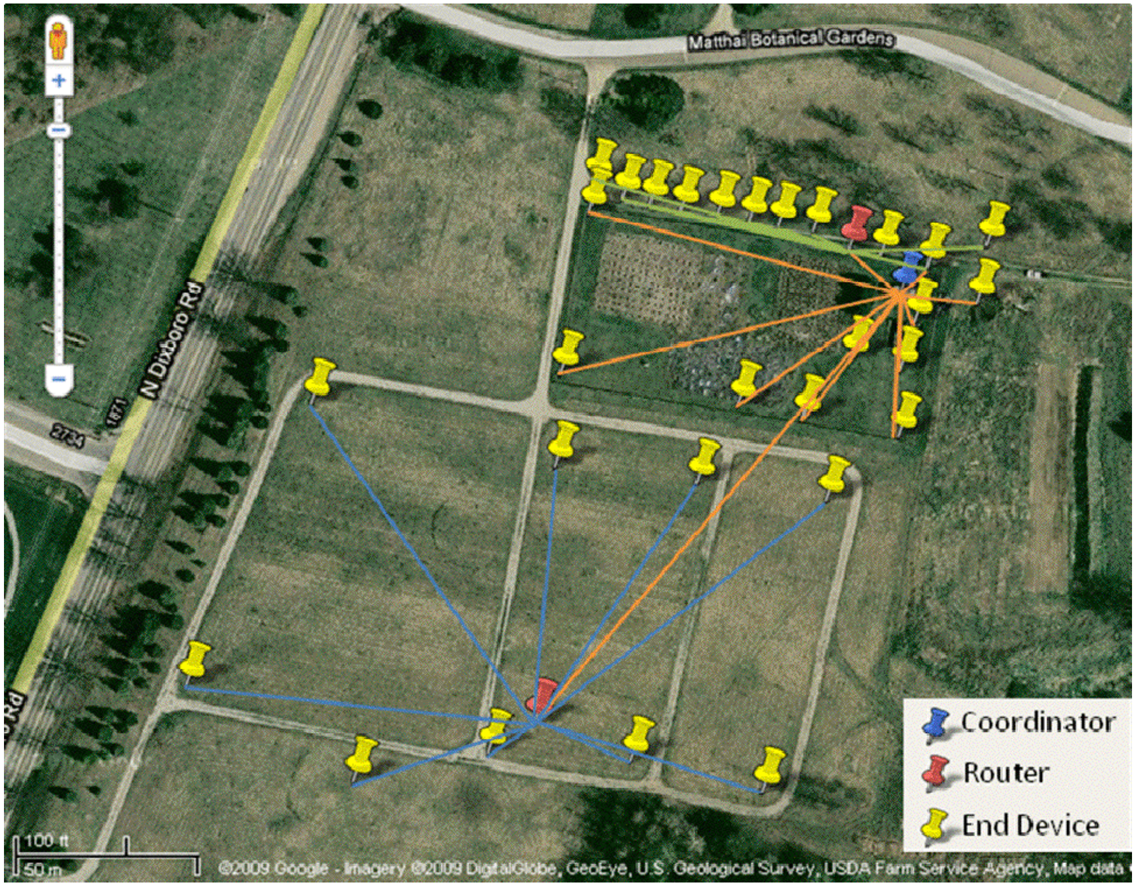
- Developed and successfully tested “Ripple-1” wireless sensor nodes for field deployment at U of M Matthaei Botanical Gardens
- Started on the design of “Ripple-2” ground unit platform to provide better energy efficiency for router nodes
- Webpage released, with a backend database to store and retrieve real-time soil moisture data collected at the Botanical Gardens

Functional view



- Soil moisture observations provided by sparse set of sensors in the field
- Each sensor sends data to coordinating center via a wireless network
- Base station assimilates data, generates control, and sends that to actuators at sensor locations

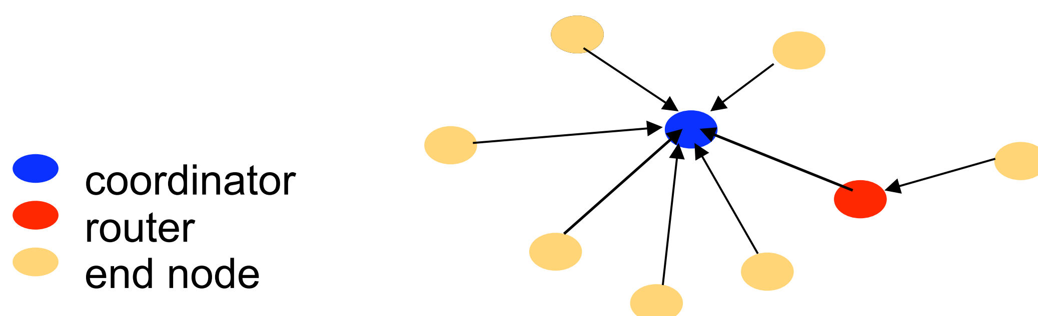
Field Deployment: Matthaei Botanical Gardens



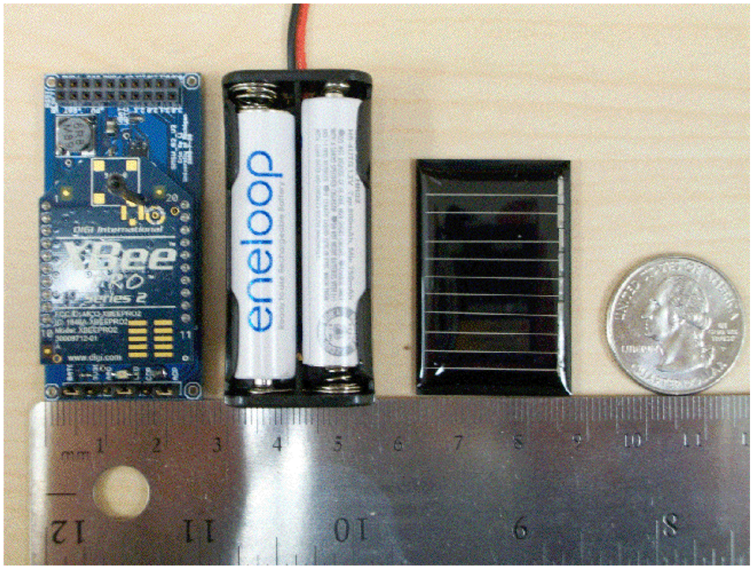
Thirty Ripple-1 sensor node were built and successfully deployed at Matthaei during our AIST-05 project. Figure shows the Zigbee network formed by these sensor nodes.

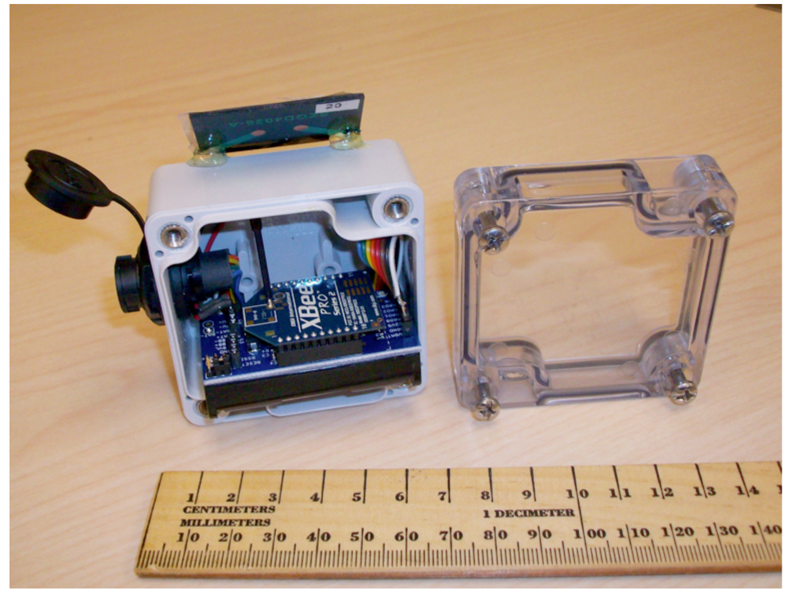
Ripple-1 ZigBee Network

- A multi-hop network consisting of three types of logic devices
- An end device is heavily duty cycled; the base station is plugged in; the router currently needs more power than desired
- Can remotely access sensors and control data collection
- Can remotely set the sensor and transceiver on cyclic sleep mode to better control energy consumption



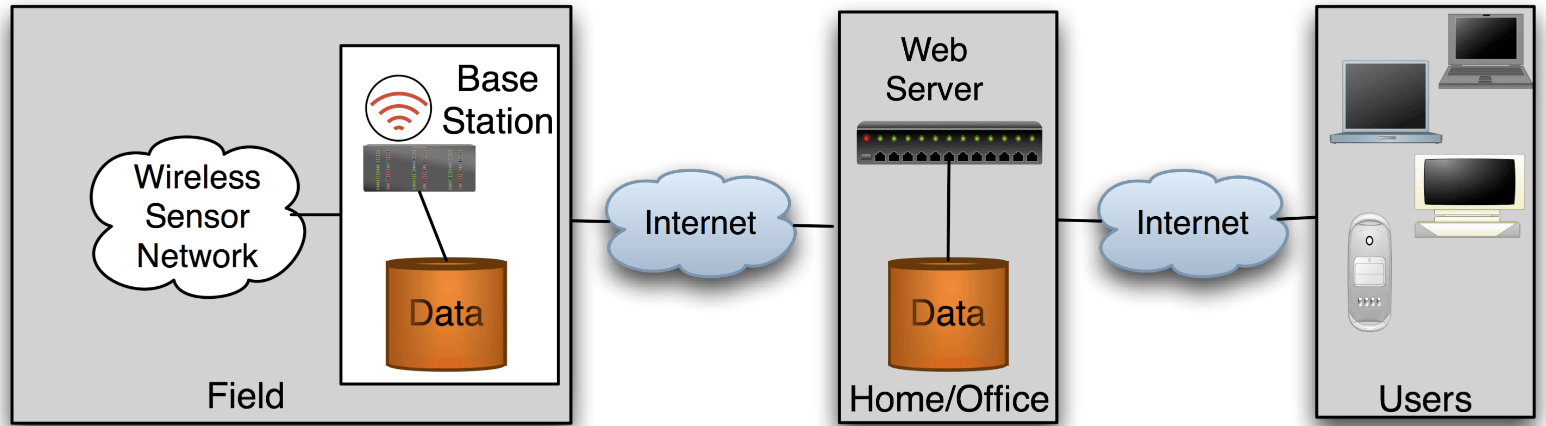
AIST-05: The Ripple-1 Node





- Xbee Pro SOC module serves as MCU and radio
  - Long Communication Range: up to 1 mile (1600 m)
  - Low Power Consumption
    - ✓ 295mA @3.3 V (TX)
    - ✓ 45 mA @3.3 V (RX)
    - ✓ < 10 uA (Sleep)

Global architecture of the Ripple system



- In target field
  - A sensor network consisting of multiple wireless ground units and sensors
  - An indoor base station; data stored in a database
  - Scheduling policy run on the base station
- On UM campus
  - Web site hosted on a server (soilscape.eecs.umich.edu)
  - Real-time data query and display
  - Enable mobile access

Plans for Deploying at SMAP 2010 Field Sites

- SMAP has deployed a network of ground sensors in Marena, OK, this spring
  - Primary goal is to benchmark various in-situ sensors against each other
  - May also investigate issues related to scaling of soil moisture measurements, but limited scope
- We plan to deploy networks concurrently
  - Marena, OK (interleaved with SMAP sensors)
  - Canton, OK (our network only)
  - Sites are within ~ 100 miles of each other, but Canton has more heterogeneity and allows better testing of optimal placement strategies

Summary

- Tested several candidate approaches for sensor placement optimization
  - Implemented and verified empirical placement strategies; started developing analytical joint placement/scheduling methodology
- Built landscape simulator
  - Simulator architecture developed, implemented, and tested; preliminary scaling studies performed
- Developed wireless sensor actuation and communication nodes
  - Multihop architecture investigated
- Ongoing collaborations with SMAP algorithms and cal/val team
  - Continually collaborating with team; will install in one or two of SMAP cal/val nodes (one in Oklahoma, another TBD)
- Various project elements at TRLs of 3 to 4