Cyberinfrastructure Tools for Precision Agriculture in the 21st Century

Michela Taufer

The University of Tennessee Knoxville

THE UNIVERSITY OF TENNESSEE KNOXVILLE

BIG ORANGE. BIG IDEAS.®

GCLab

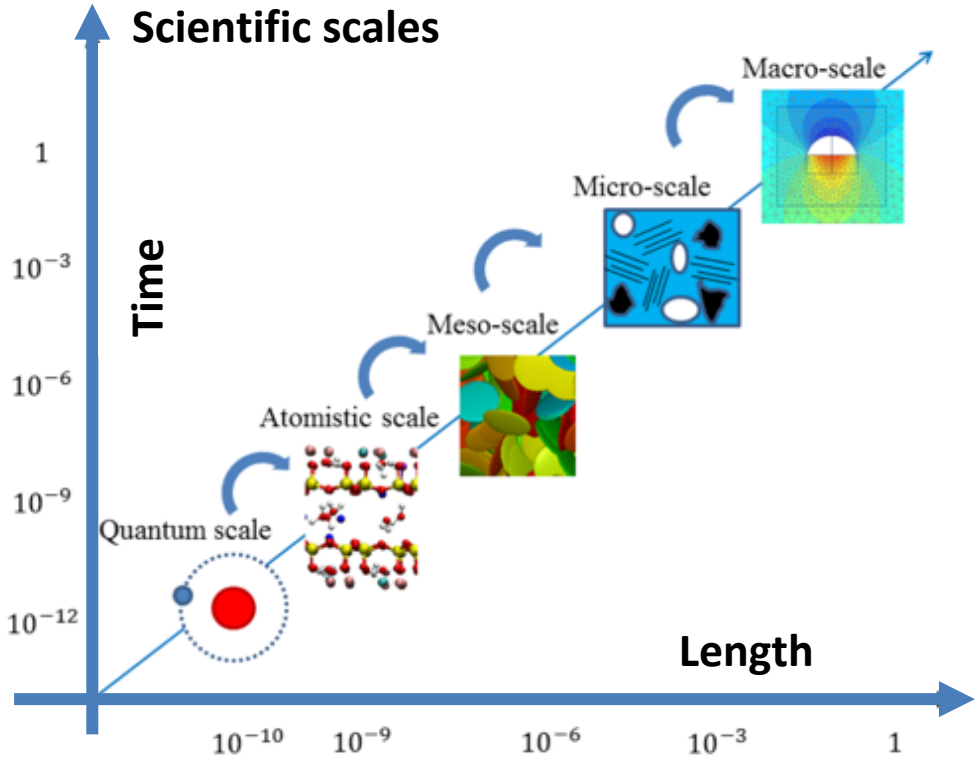
Contributors and collaborators

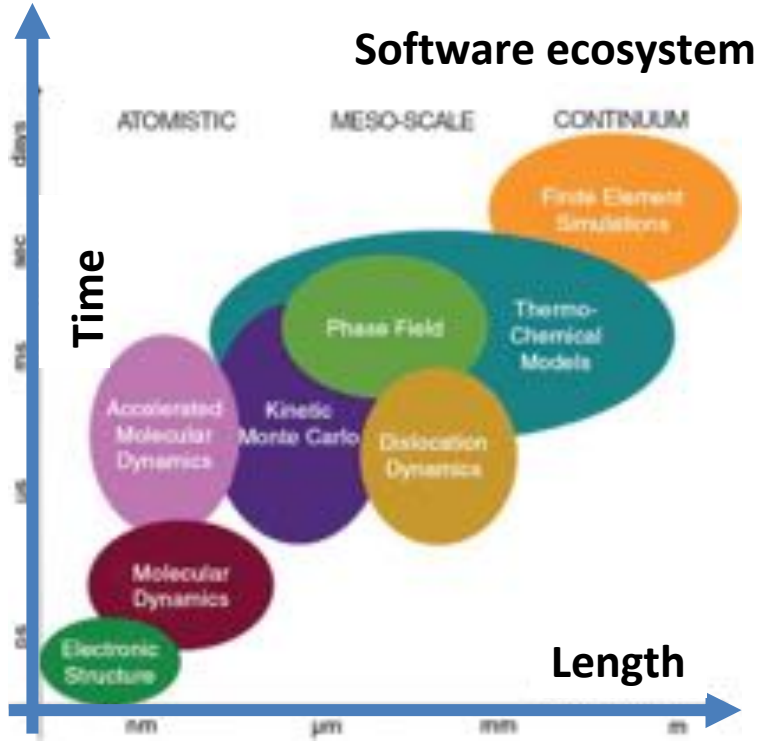
- U. Delaware: Ricardo Llamas, Mario Guevara, and Rodrigo Vargas
- UTK: Danny Rorabaugh, Kae Suarez, Leobardo Valera, Ria Patel, and David Icove
- ORNL: Jimmy Landmesser
- UIUC: Craig Willis and Victoria Stodden

Sponsors and supporters

- NSF OAC 1854312 CIF21 DIBBs: PD: Cyberinfrastructure Tools for Precision Agriculture in the 21st Century (PIs: Taufer and Vargas)
- NSF OAC 1941443 EAGER: Reproducibility and Cyberinfrastructure for Computational and Data-Enabled Science (PIs: Stodden and Taufer
- IBM Shared University Research (SUR) Award
- NSF XSEDE JetStream: Allocations EAR180011 and TRA180041
  - Many thanks to Jeremy Fischer, IU

Multiscale computational modeling

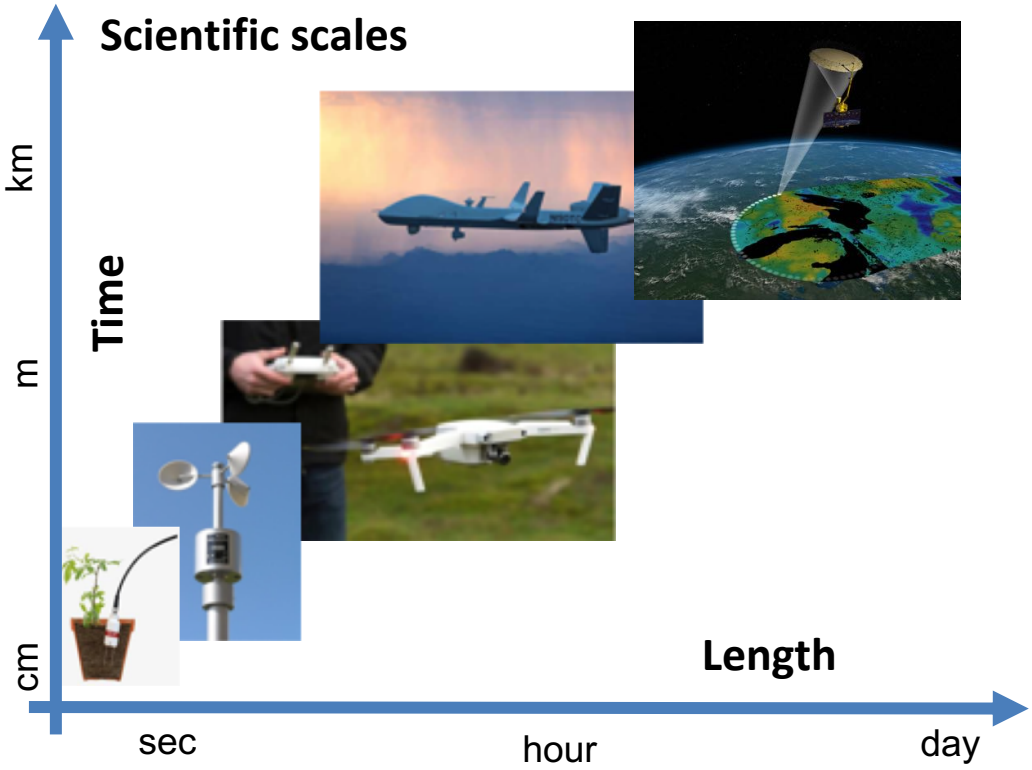




https://ajw-group.mit.edu/multiscale-modeling-clays

M Stan, Material Today, 12, 2009, 20-28

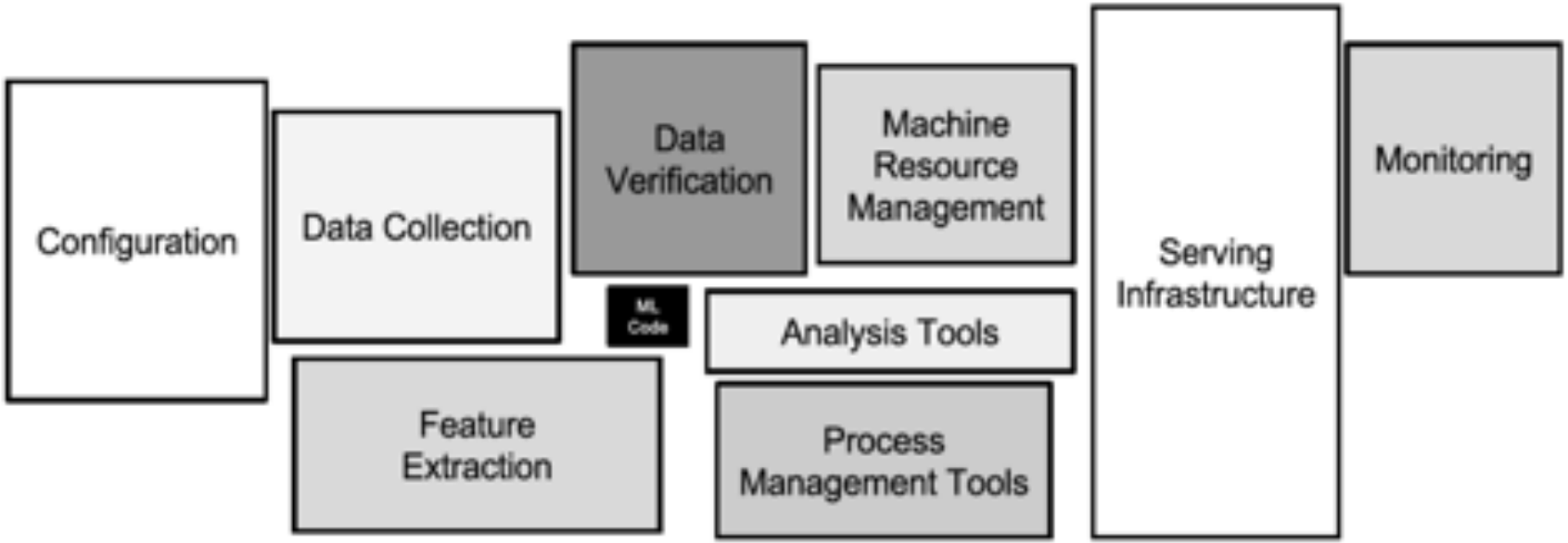
Multiscale data modeling (MSDM)



Software ecosystem

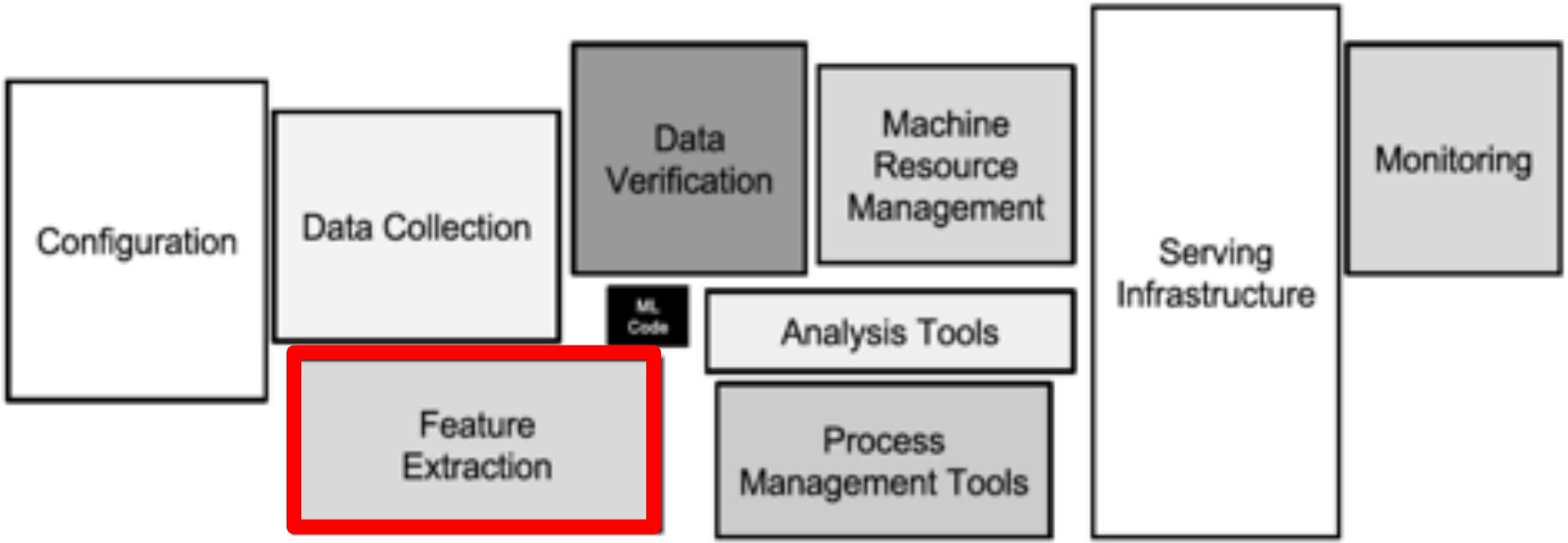
?

Hidden (forgotten?) software ecosystem



"**Only a small fraction of real-world ML systems is composed of the ML code**" D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips () Hidden Technical Debt in Machine Learning Systems

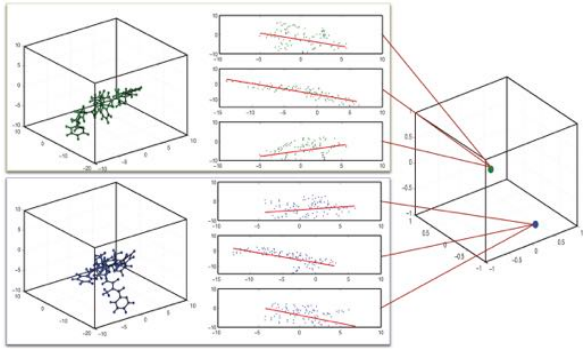
Hidden (forgotten?) software ecosystem



"**Only a small fraction of real-world ML systems is composed of the ML code**" D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips () Hidden Technical Debt in Machine Learning Systems

Feature extraction

Protein-ligand docking

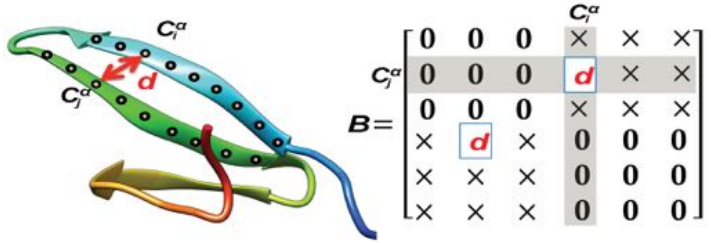


Linear regression: map ligands into **3-D point** representation

M. Taufer, T. Estrada, and T. Johnston. Algorithms for In Situ Data Analytics of Next Generation Molecular Dynamics Workflows. Numerical algorithms for high-performance computational science. *Issue of Philosophical Transactions A*., 2019.

Feature extraction

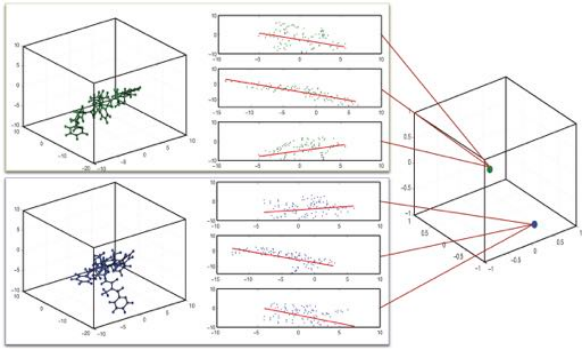
Protein folding



$$B = \begin{bmatrix} 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & d & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ \times & d & \times & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} & & & C_i^\alpha & & \\ & & & & & \\ C_j^\alpha & & & & & \end{matrix}$$

Numerical analyses: map secondary structures into **eigenvalues**

Protein-ligand docking

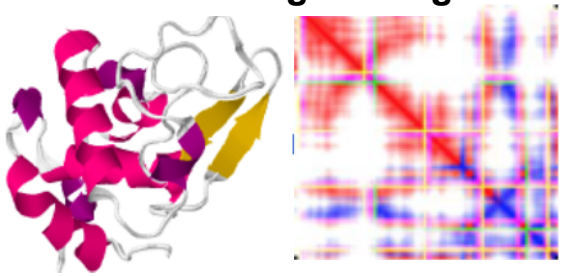


Linear regression: map ligands into **3-D point** representation

M. Taufer, T. Estrada, and T. Johnston. Algorithms for In Situ Data Analytics of Next Generation Molecular Dynamics Workflows. Numerical algorithms for high-performance computational science. *Issue of Philosophical Transactions A.*, 2019.

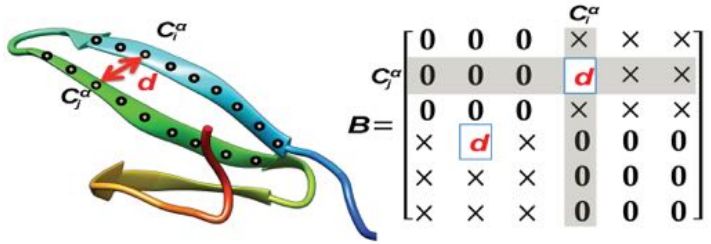
Feature extraction

Protein engineering



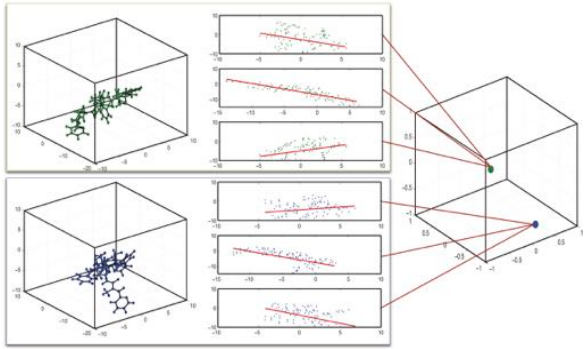
Deep leaning: map both secondary and ternary structures into **tensors**

Protein folding



Numerical analyses: map secondary structures into **eigenvalues**

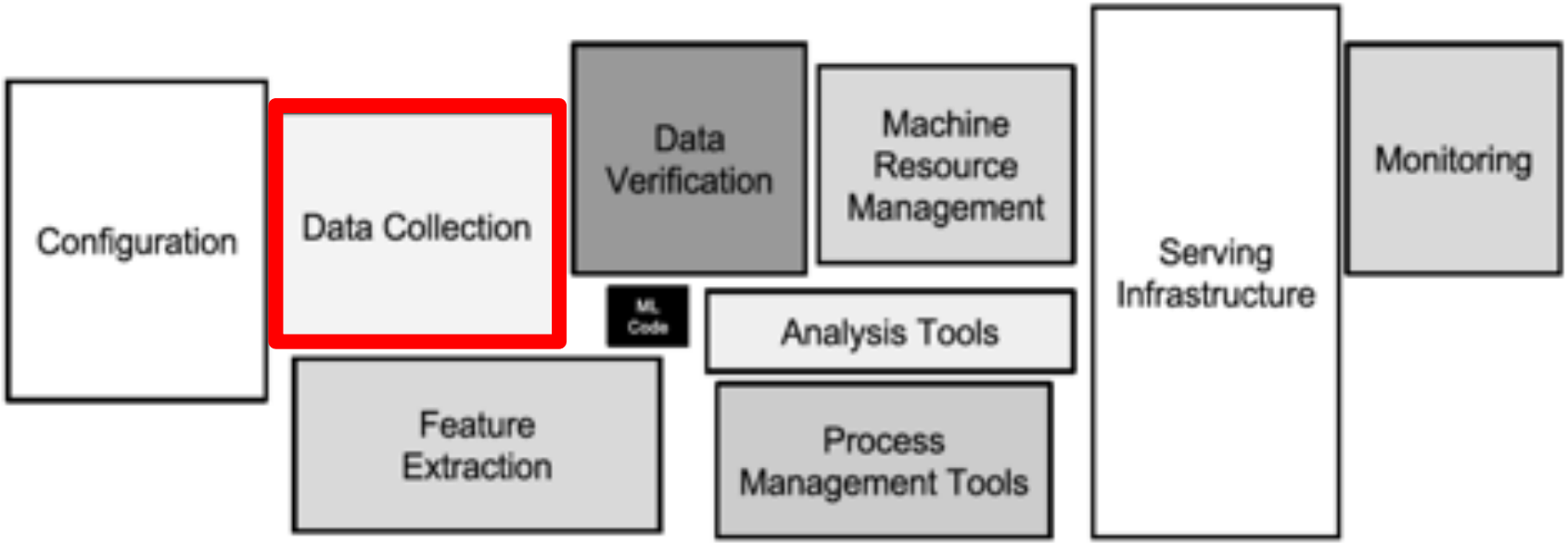
Protein-ligand docking



Linear regression: map ligands into **3-D point** representation

M. Taufer, T. Estrada, and T. Johnston. Algorithms for In Situ Data Analytics of Next Generation Molecular Dynamics Workflows. Numerical algorithms for high-performance computational science. *Issue of Philosophical Transactions A.*, 2019.

Hidden (forgotten?) software ecosystem



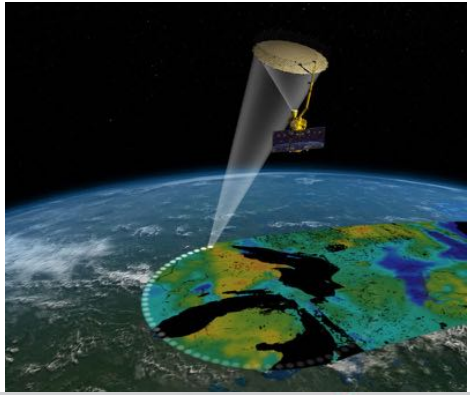
"**Only a small fraction of real-world ML systems is composed of the ML code**" D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips () Hidden Technical Debt in Machine Learning Systems

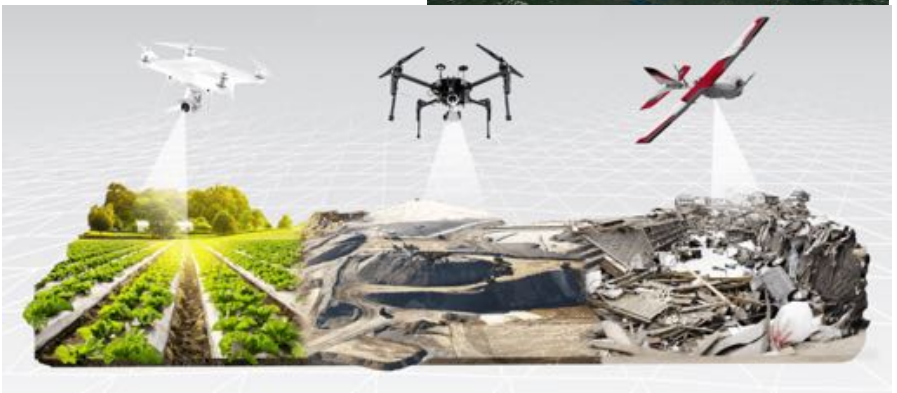
Data collection at the edge

Point Field
Measurements

Data collection at the edge

Remote Sensor Measurements



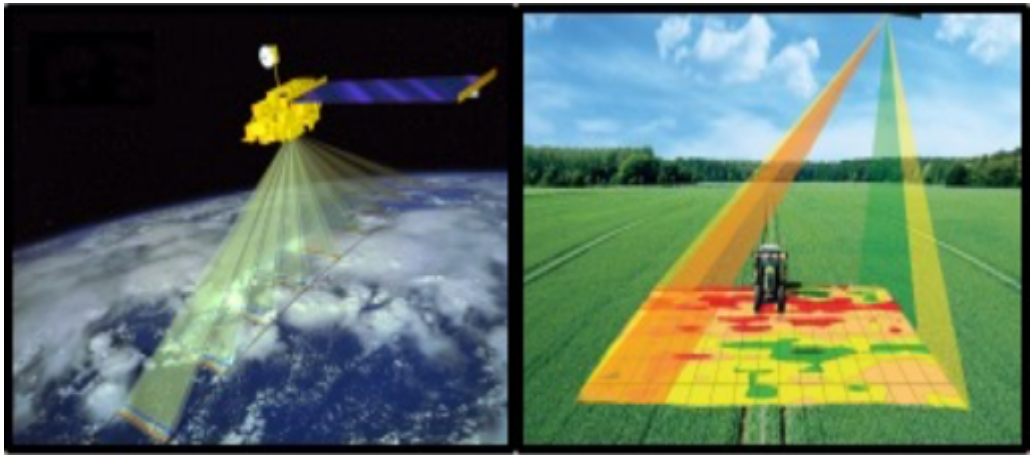


Point Field Measurements

Challenges in MMDM

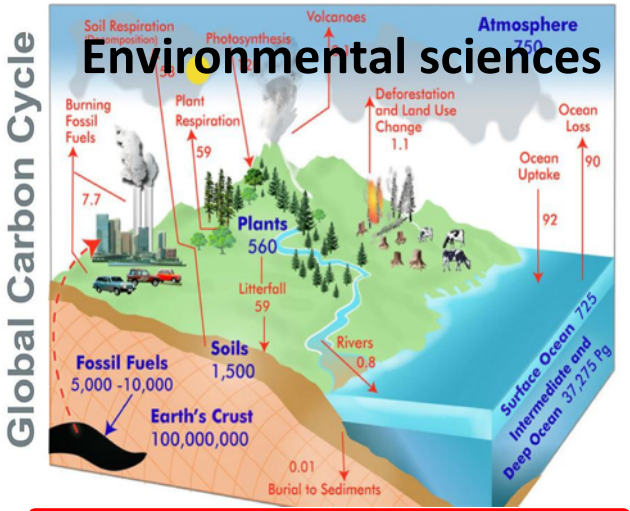
- Design and implement robust and sustainable software ecosystems
- Combine analytics and computing across heterogenous platforms (i.e., HPC, Cloud, and edge computing)
- Build trust in results through reproducibility, replicability, and transparency (RRT)

Relevance of soil moisture data



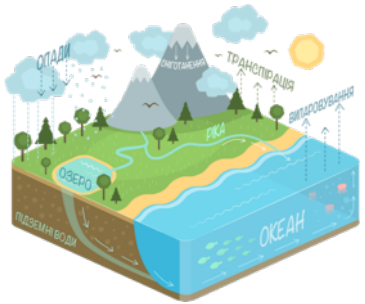
- Satellite-borne remote sensing technology
  - Infrared to radio
  - Active and passive

Environmental sciences



Workflows for precision agriculture

Application



Soil moisture leveraged for:
- Environmental sciences
- Precision agriculture

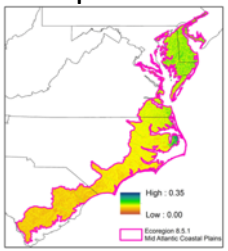
Computation

Data Feedback

Fine-grained Soil Moisture
Landscape Surface DSM
Weather Data NOAA

Analytics representations + algorithms

Data prediction



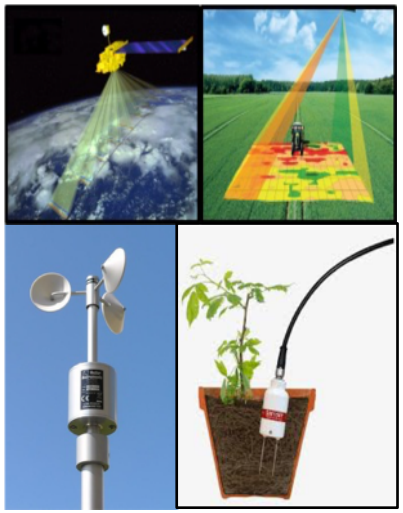
Data Analytics

Soil Moisture ESA-CCI
Landscape Surface DSM
Weather Data NOAA

Fine-grained, complete data

Coarse-grained, incomplete data

Multiscale data

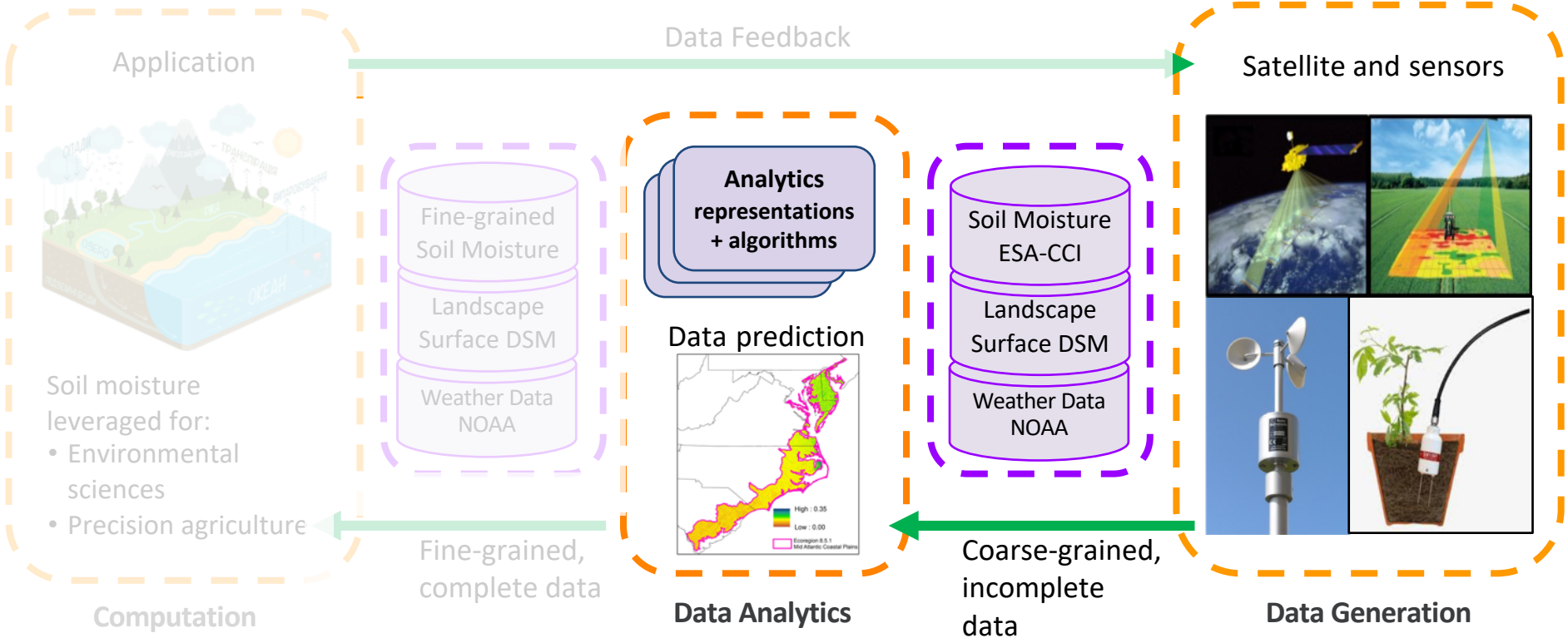


Data Generation

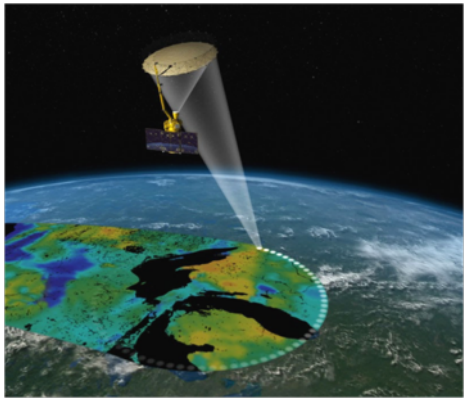
Design and implement a software ecosystem for precision agriculture

Collaborators: Rodrigo Varga's Group (UD)
Platform: NSF XSEDE Jetstream
NSF OAC 1854312 CIF21 DIBBs: PD: Cyberinfrastructure Tools for Precision Agriculture in the 21st Century

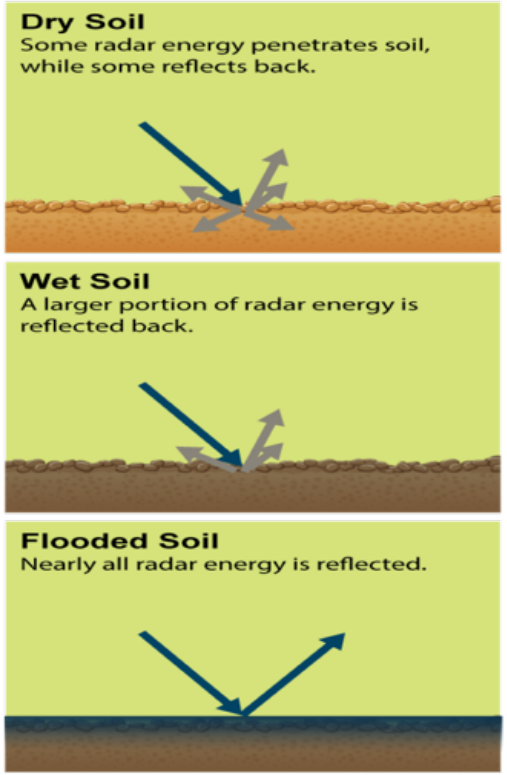
Data analytics for soil moisture

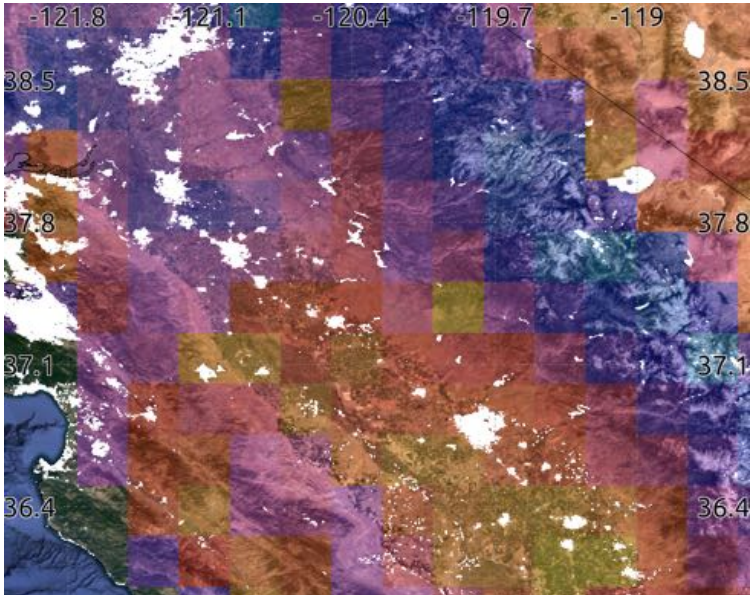


Challenge 1: incomplete soil moisture data (I)



Satellites collect raster data across the surface of the Earth

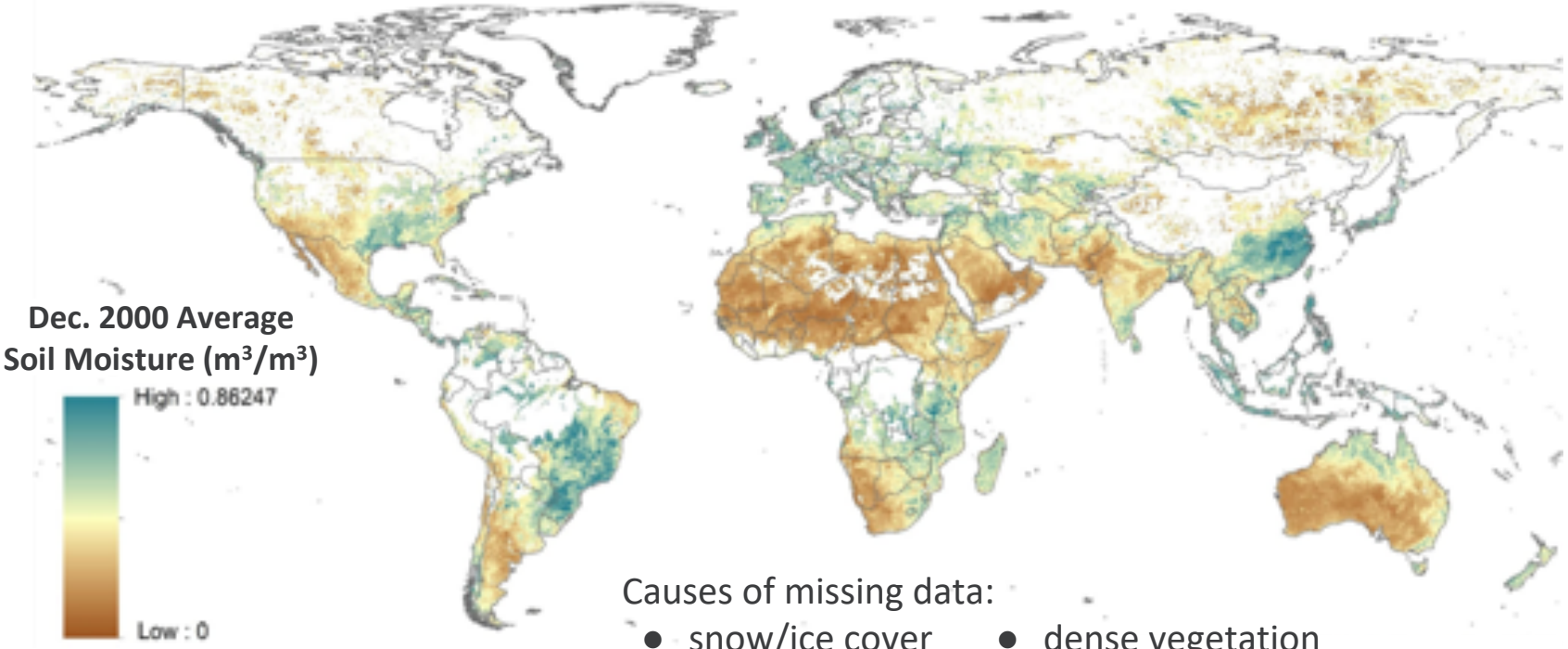




Visualization example of the ESA-Climate Change Initiative Soil Moisture database with a coarse pixel size of 27x27km

(Liu et al. 2011 HESS, Liu et al. 2012 RSE)

Challenge 1: incomplete soil moisture data (II)

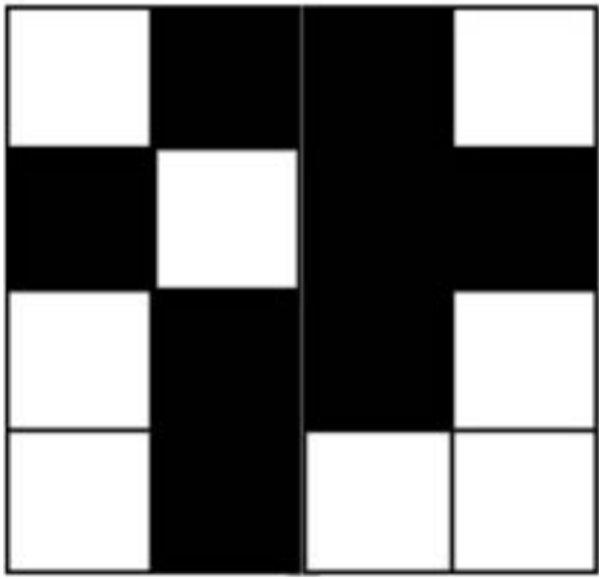


Causes of missing data:

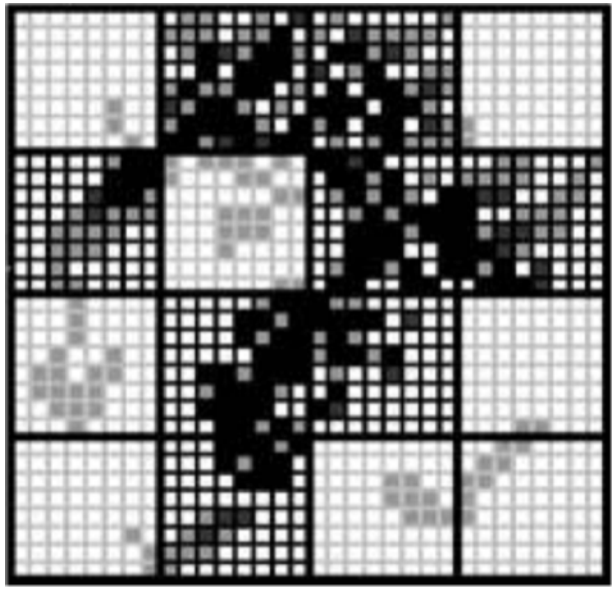
- snow/ice cover
- frozen surface
- dense vegetation
- extremely dry surface

ESA-CCI soil moisture database, http://www.esa-soilmoisture-cci.org

Challenge 2: coarse-grained soil moisture data (I)



Original Resolution
27 km × 27 km

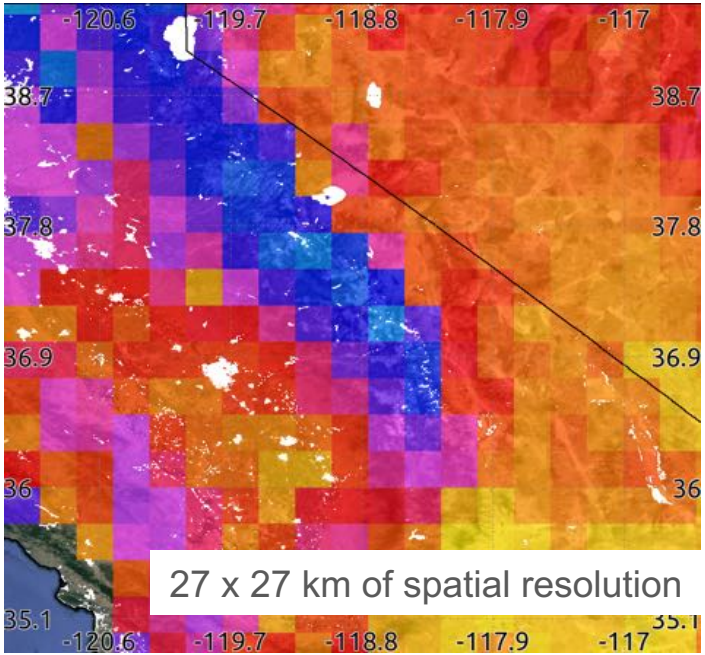


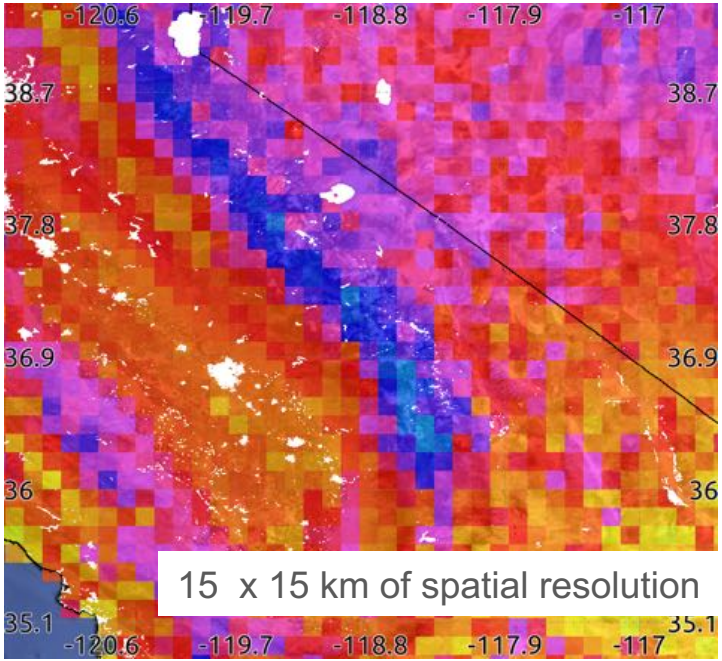
Desired Resolution
1 km × 1 km

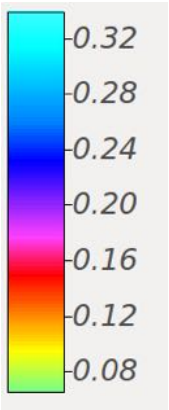
Image source: McPherson et al., *Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions—possibilities and limitations*, Ecological Modeling **192**:499–522, 2006.

Challenge 2: coarse-grained soil moisture data (II)

Original product ESA CCI (m^3 m^{-3}, mean 2013)

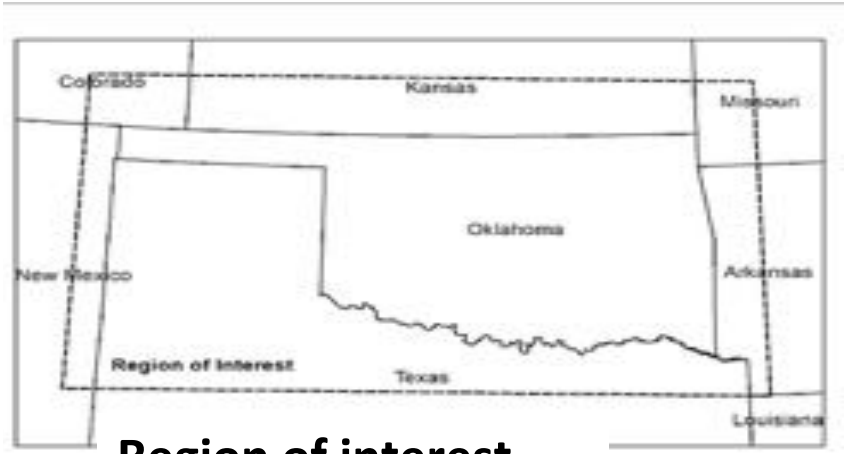




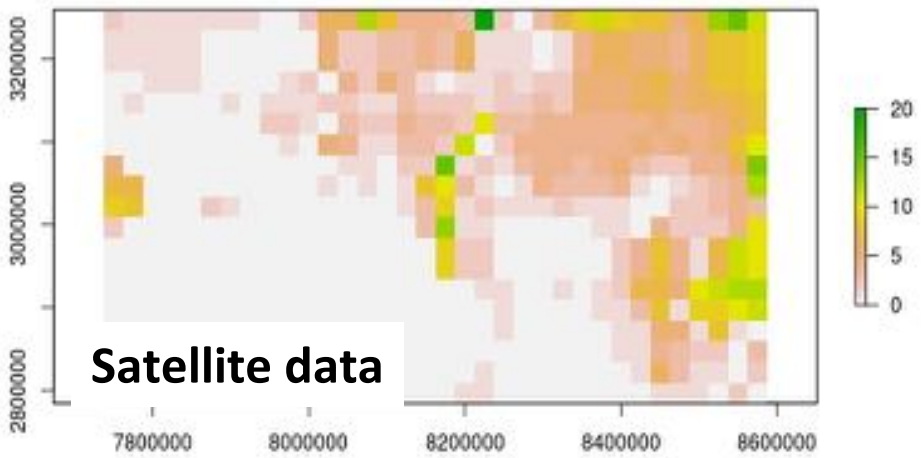


M. Guevara , M. Taufer, and R. Vargas. Gap-Free Annual Soil Moisture Global across 15km Grids: 1991-2016. Earth System Science Data, 2019.

Integration of multiscale data: from satellites …



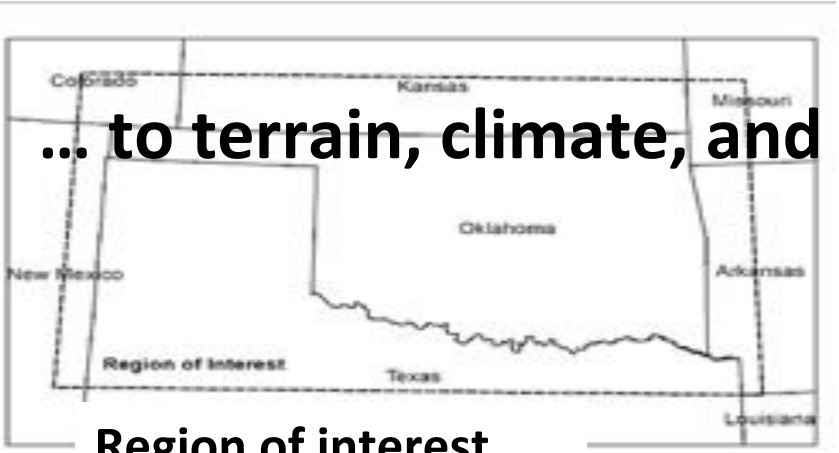
Region of interest



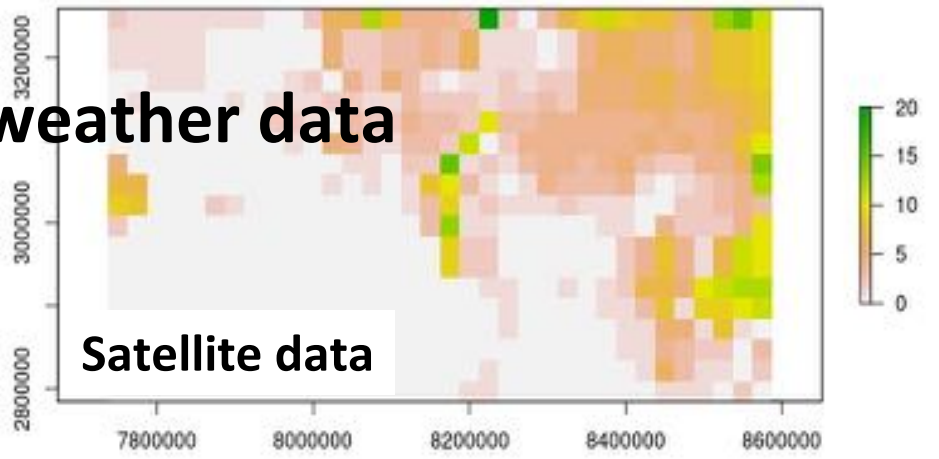
Satellite data

R Llamas, M Guevara, D Rorabaugh, M Taufer, R Vargas. Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture based on Geostatistical Techniques and Multiple Regression. Remote Sensing, 2020.

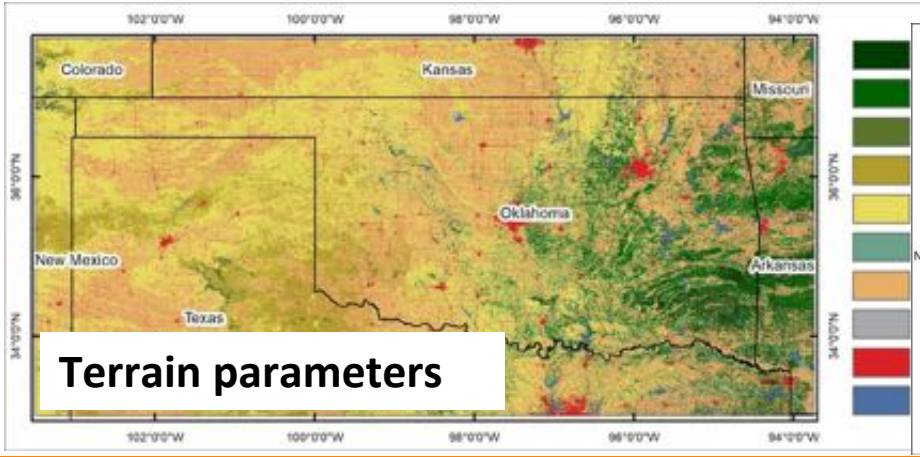
… to terrain, climate, and weather data



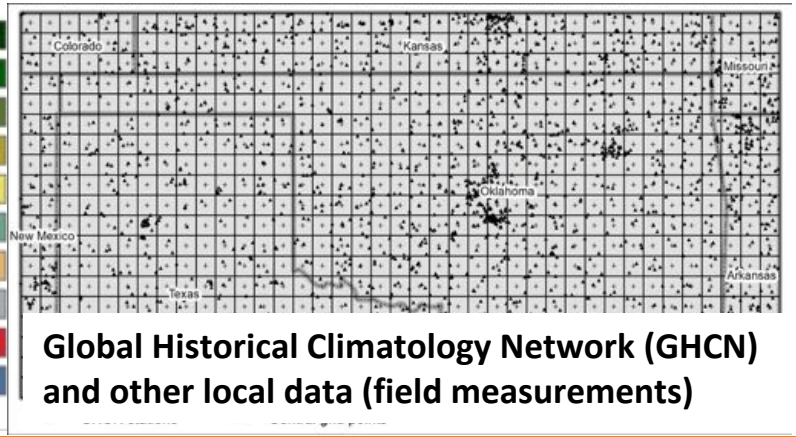
Region of interest



Satellite data



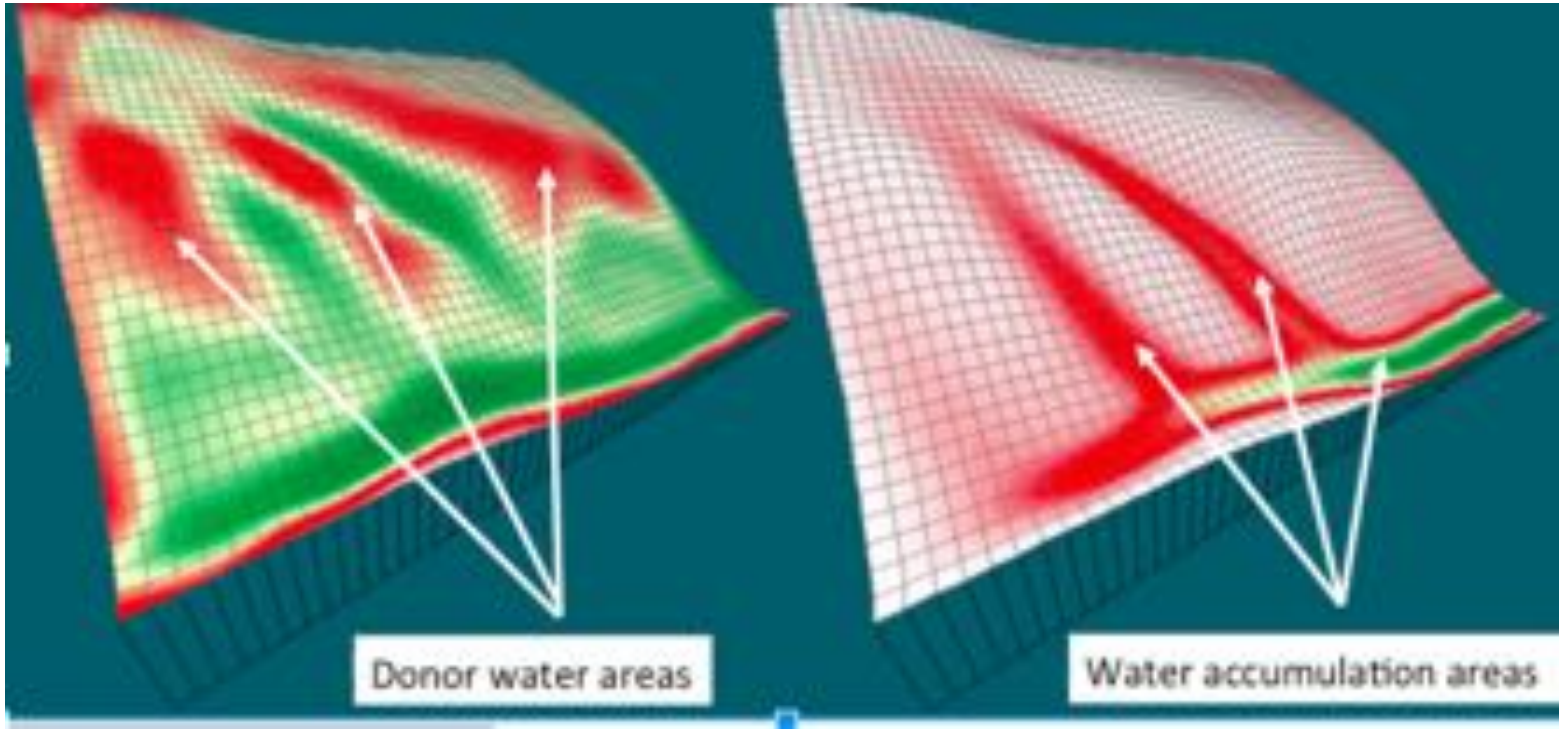
Terrain parameters



Global Historical Climatology Network (GHCN) and other local data (field measurements)

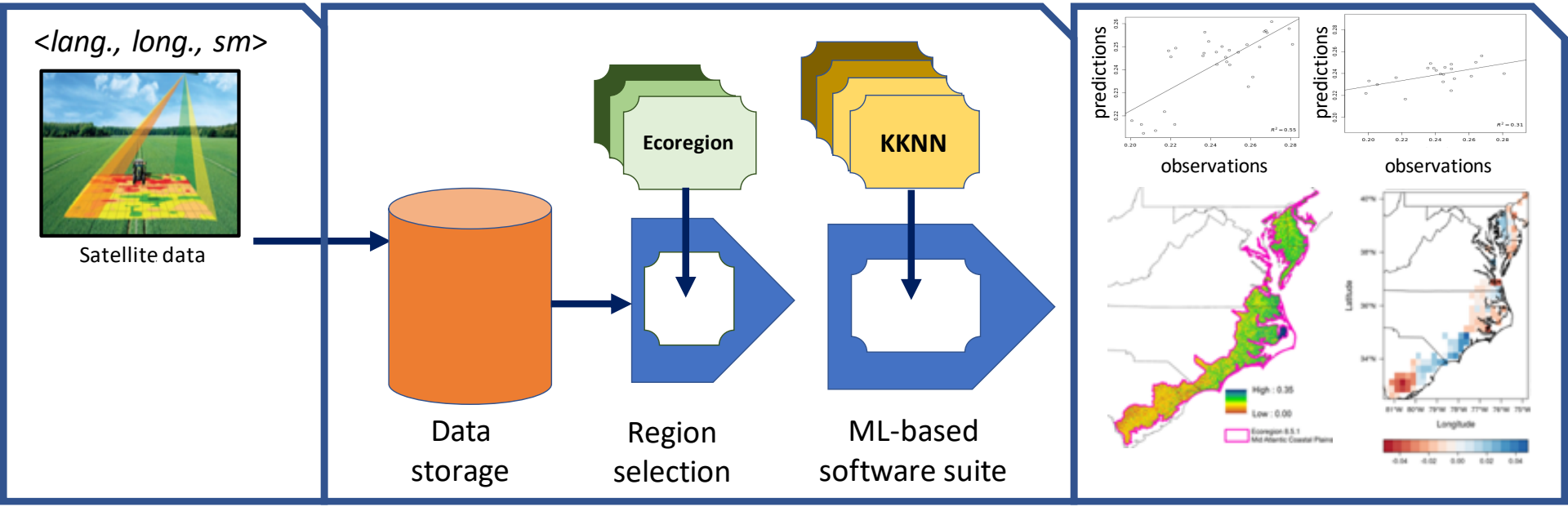
R Llamas, M Guevara, D Rorabaugh, M Taufer, R Vargas. Spatial Gap-Filling of ESA CCI Satellite-Derived Soil Moisture based on Geostatistical Techniques and Multiple Regression. Remote Sensing, 2020.

Example of terrain parameters: water wetness index



Shaw et al., 2016 GRL, Moore 2012, Geomorphology.

SOMOSPIE: SOil MOisture SPatial Inference Engine

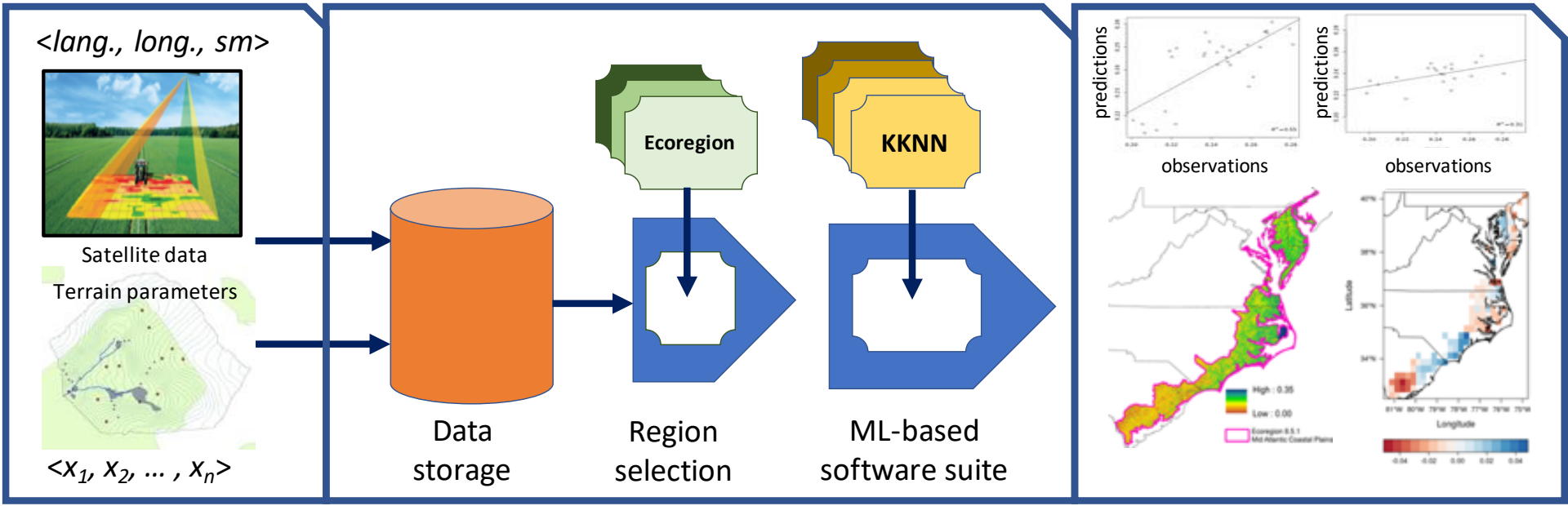


Data collection **Feature extraction** **Predictions** **Analysis tools**

D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas, and **M. Taufer**. SOMOSPIE: A Modular SOil MOisture SPatial Inference Engine based on Data Driven Decisions. eScinece 2019.

BIG ORANGE BIG IDEAS

SOMOSPIE: SOil MOisture SPatial Inference Engine

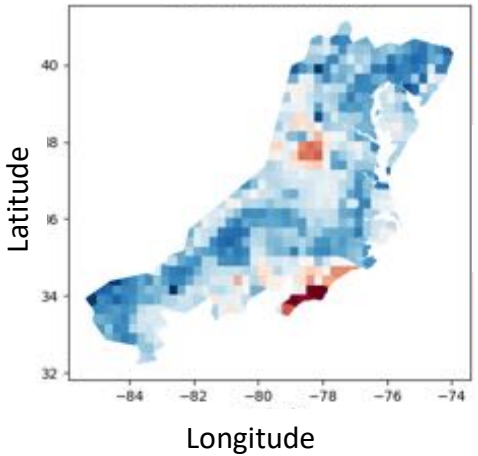


Data collection **Feature extraction** **Predictions** **Analysis tools**

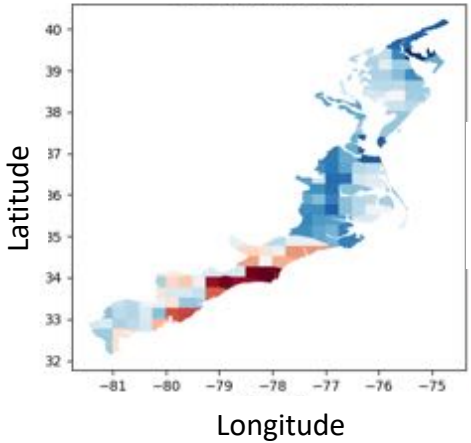
D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas, and **M. Taufer**. SOMOSPIE: A Modular SOil MOisture SPatial Inference Engine based on Data Driven Decisions. eScinece 2019.

Region selection: format of regions of interest

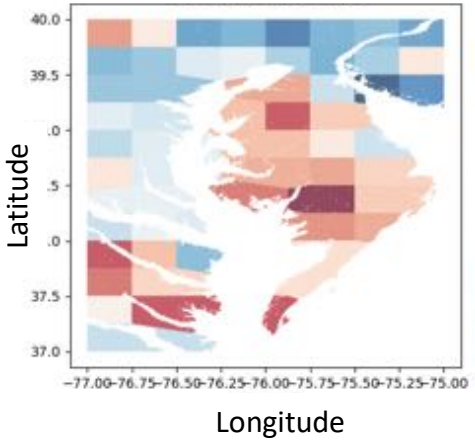
("NEON", "Mid Atlantic")



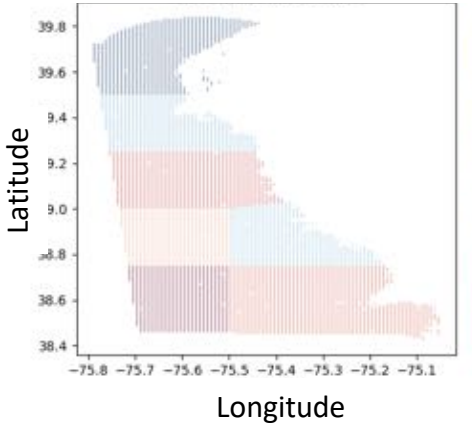
("CEC", "8.5.1")



("BOX", "-77_-75_37_40")



("STATE", "Delaware")



Algorithmic solutions: ML-based software suite

Random Forest

→ Compute weighted mean of 500 prediction trees

KKNN:

→ Use **local data**

→ Compute k and distance kernel using cross validation automatically

→ Compute weighted means with the kernel **(*many values*)**

Surrogate based model (SBM):

→ Use **all sampled data**

→ Use regression to generate one single polynomial model **(*single polynomial model*)**

Algorithmic solutions: ML-based software suite

Random Forest

→ Compute weighted mean of 500 prediction trees

KKNN:

→ Use **local data**

→ Compute k and distance kernel using cross validation automatically

→ Compute weighted means with the kernel (***many values***)

Surrogate based model (SBM):

→ Use **all sampled data**

→ Use regression to generate one single polynomial model (***single polynomial model***)

HYPPO (Hybrid Piecewise Polynomial Modeling):

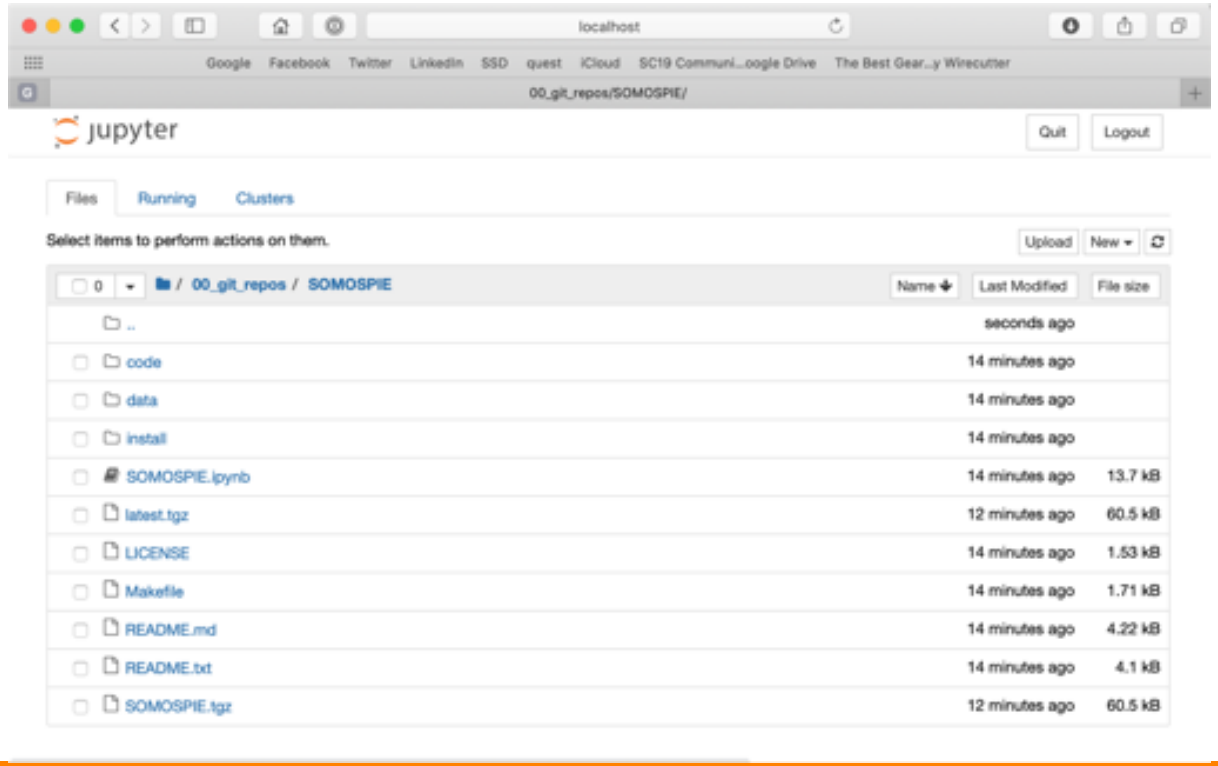
→ Use **local data**

→ Determine local polynomial degree using cross validation

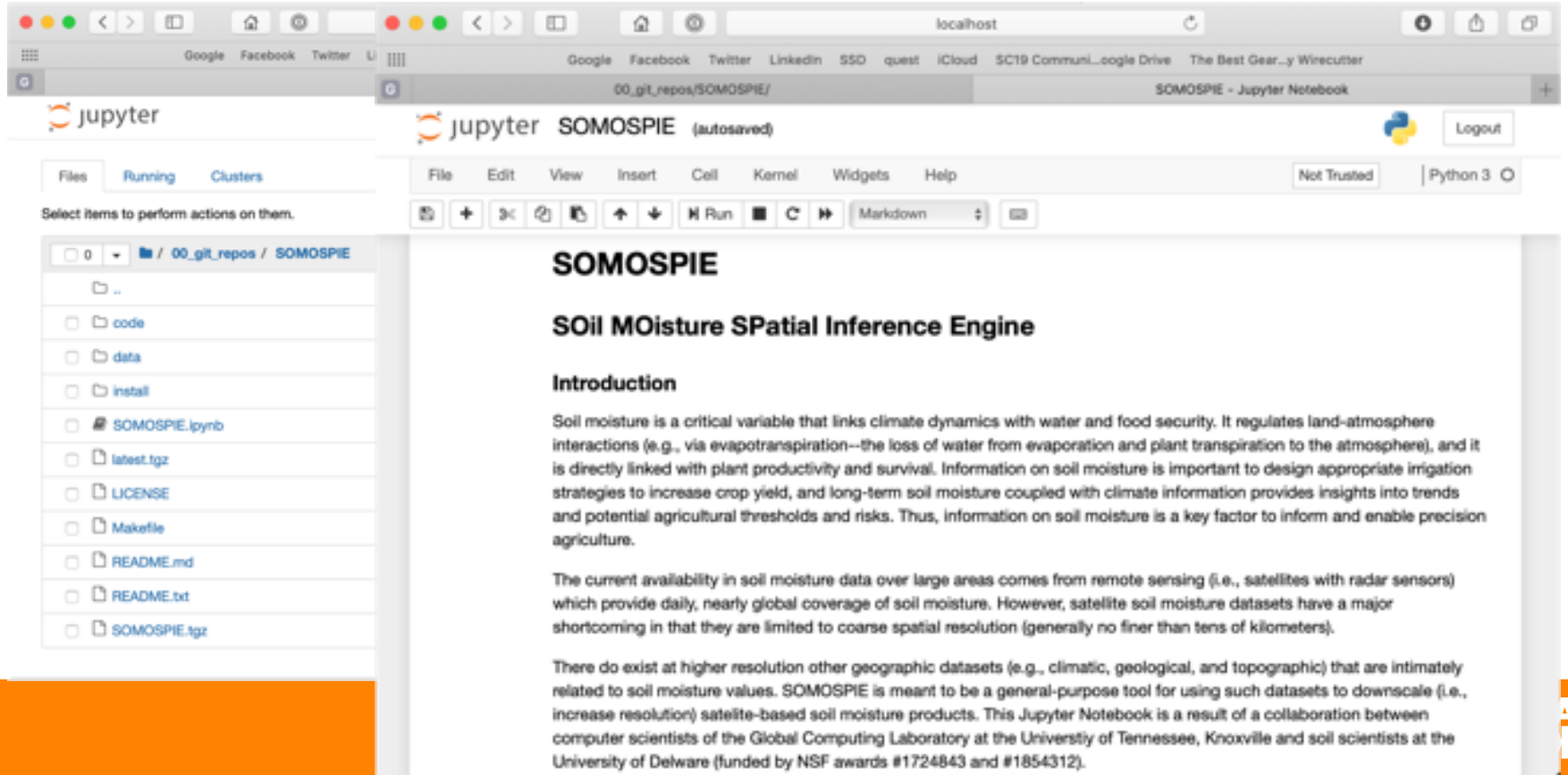
→ Use regression to generate local polynomial model (***many* polynomial models**)

D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas, and **M. Taufer**. SOMOSPIE: A Modular SOil MOisture SPatial Inference Engine based on Data Driven Decisions. eScinece 2019.

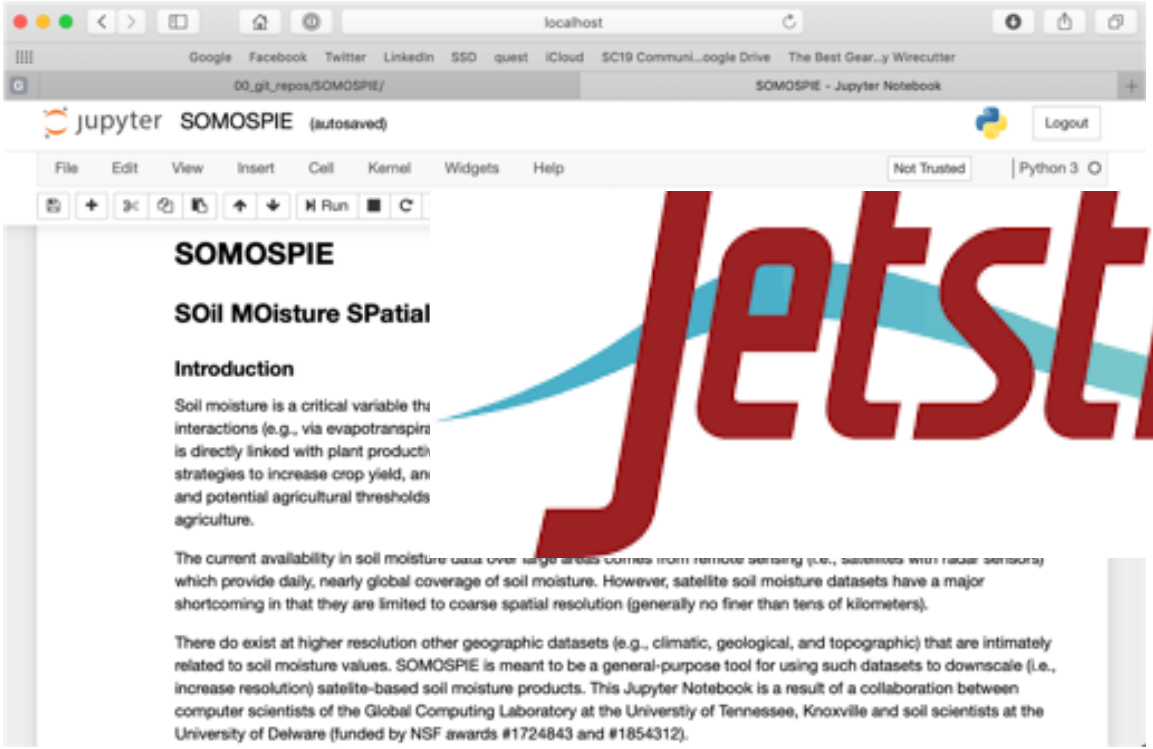
Computational solutions: Jupyter + XSEDE Jetstream



Computational solutions: Jupyter + XSEDE Jetstream



Computational solutions: Jupyter + XSEDE Jetstream

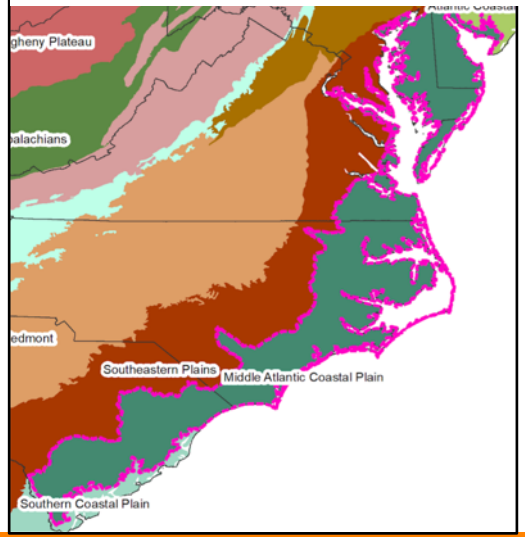


Use case I: from 27x27km to 1x1km

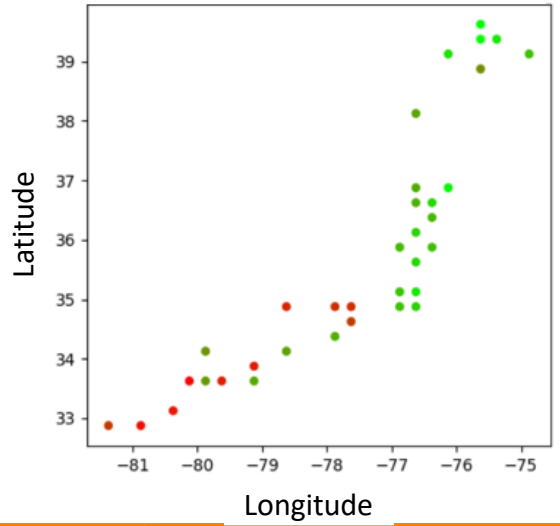
Fine-grained modeling of Mid-Atlantic region in April 2017:

- Terrain parameters: Elevation, Slope, and Wetness Index

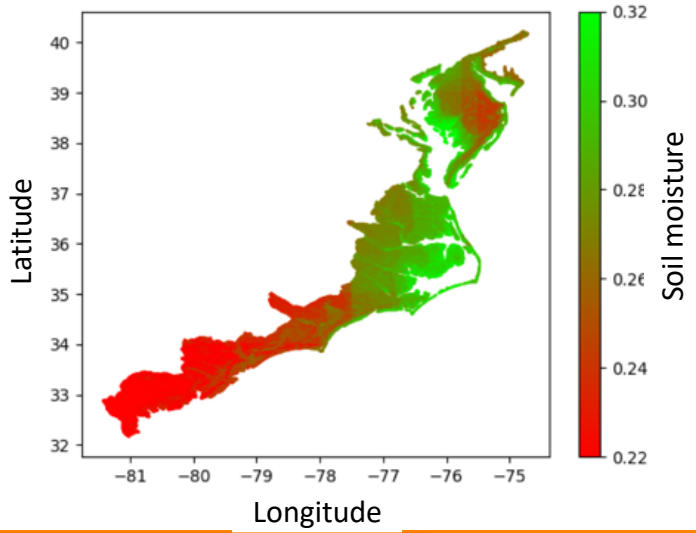
Level III Ecoregions of the Continental United States (CEVLv3)



Original satellite data (27x27km)



Random Forest fine-grained predictions 1x1km)

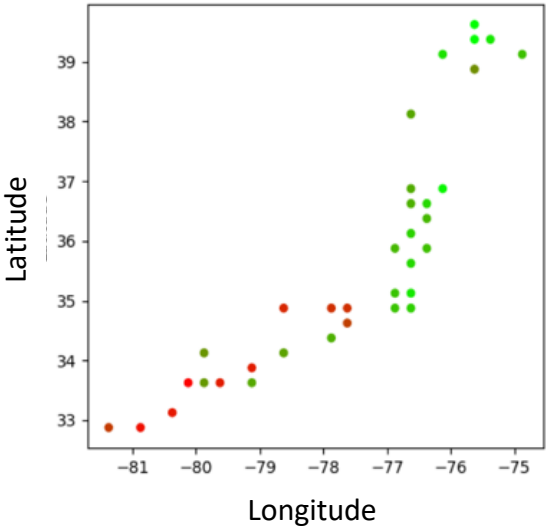


Use case I: from 27x27km to 1x1km

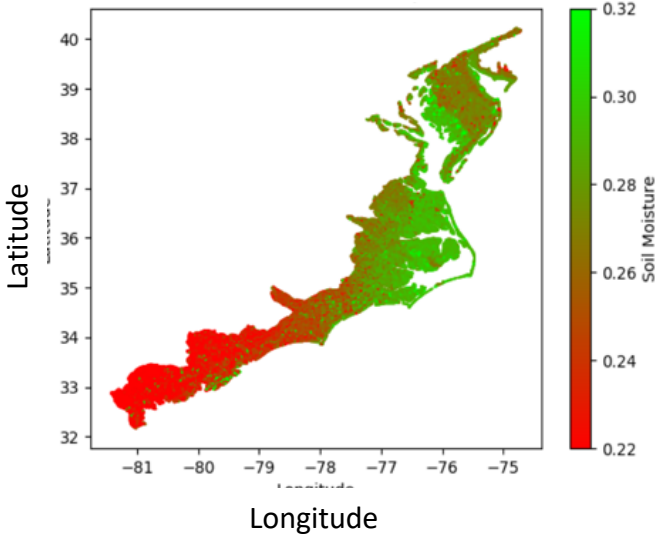
Fine-grained modeling of Mid-Atlantic region in April 2017:

- Terrain parameters: Elevation, Slope, and Wetness Index

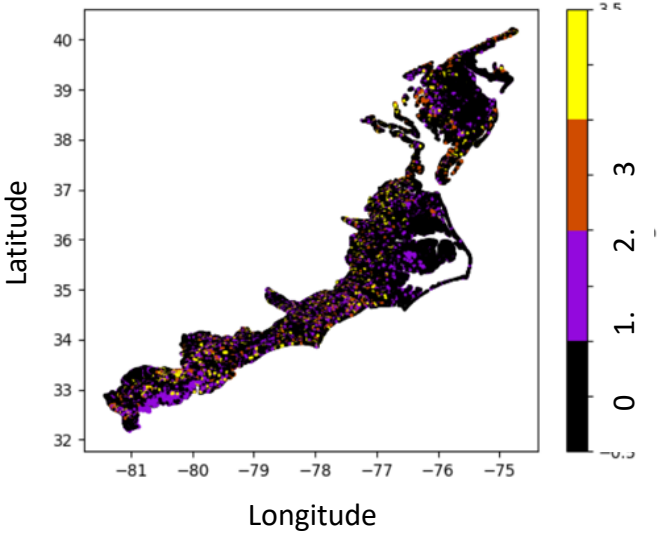
Original
satellite data



HYPPO
fine-grained predictions

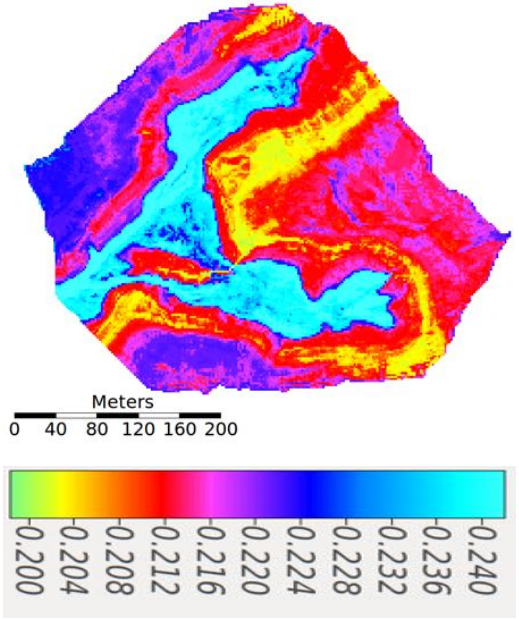


HYPPO
polynomial degrees



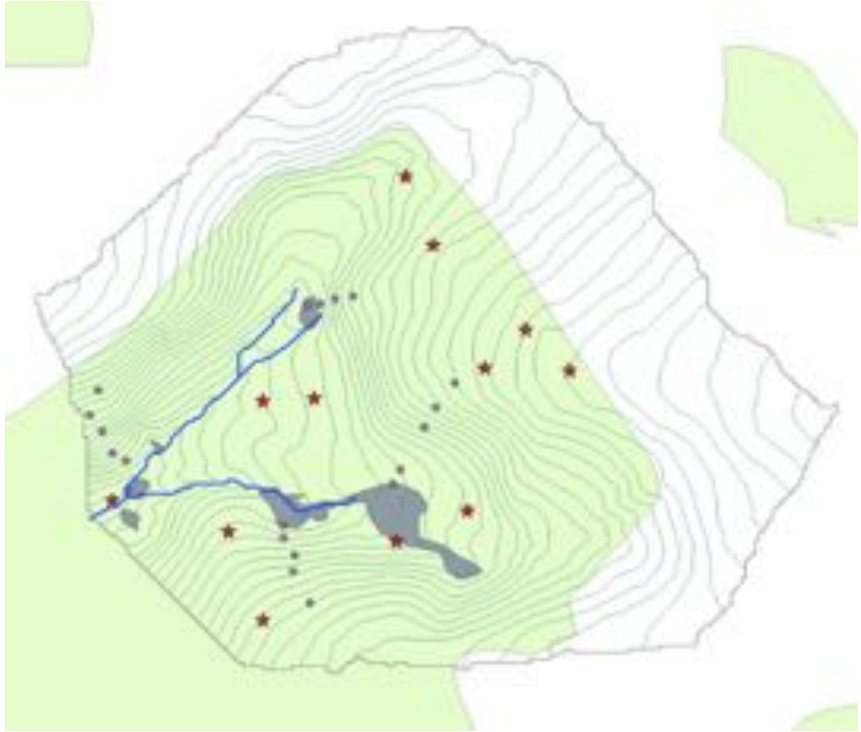
Use case II: **Local scale predictions** - 1x1m resolution

Predictions

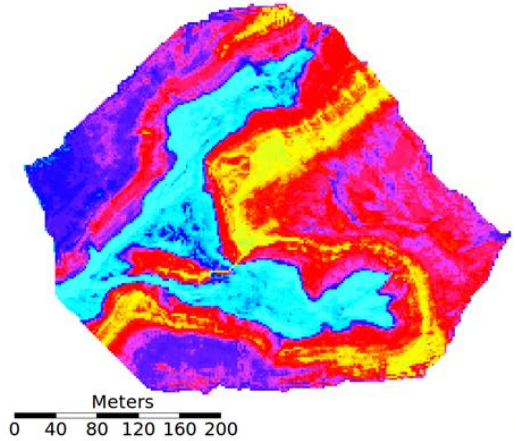


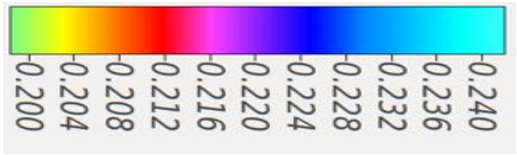
M. Guevara , M. Taufer, and R. Vargas. Gap-Free Annual Soil Moisture Global across 15km Grids: 1991-2016. Earth System Science Data, 2019.

Use case II: **Local scale predictions** - 1x1m resolution

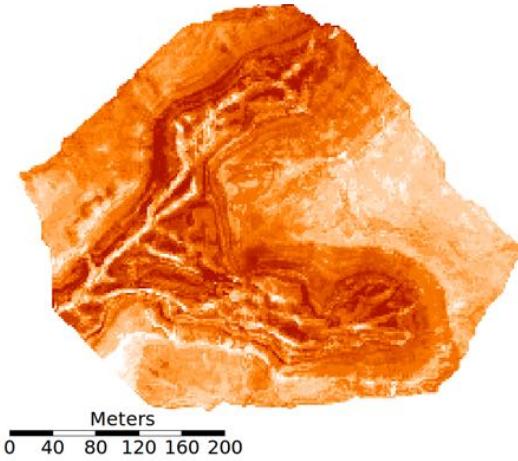


Predictions





Uncertainties



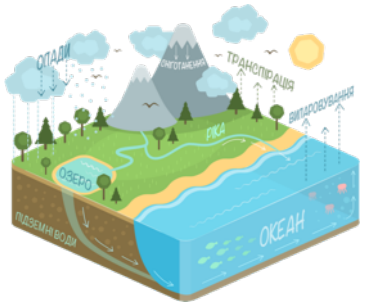
M. Guevara , M. Taufer, and R. Vargas. Gap-Free Annual Soil Moisture Global across 15km Grids: 1991-2016. Earth System Science Data, 2019.

Combine computing and analytics: integration of soil moisture predictions into controlled (or prescribed) burn

Collabororator: David Icove's group (UTK)
Planform: Tellico cluster (IBM Power9 system) – supported by 2019 IBM Shared University Research (SUR) Award

Soil moisture data for simulating controlled burn

Application



Soil moisture leveraged for:

- **Controlled or prescribed burn**

Computation

Data Feedback

Fine-grained Soil Moisture

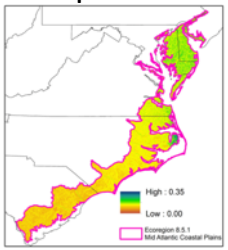
Landscape Surface DSM

Weather Data NOAA

Fine-grained, complete data

Analytics representations + algorithms

Data prediction



Data Analytics

Soil Moisture ESA-CCI

Landscape Surface DSM

Weather Data NOAA

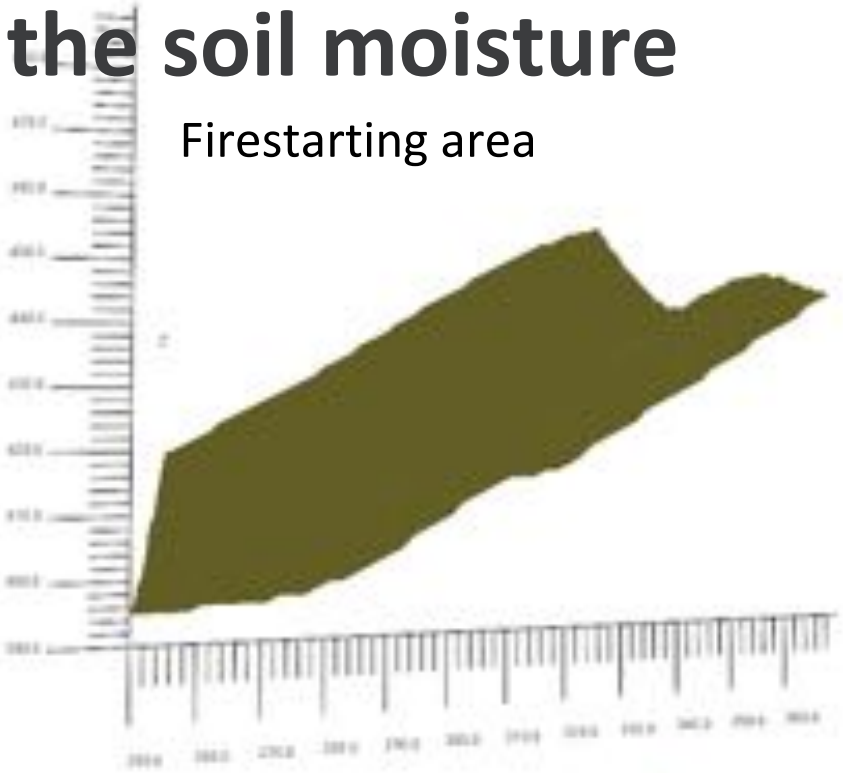
Coarse-grained, incomplete data

Multiscale data

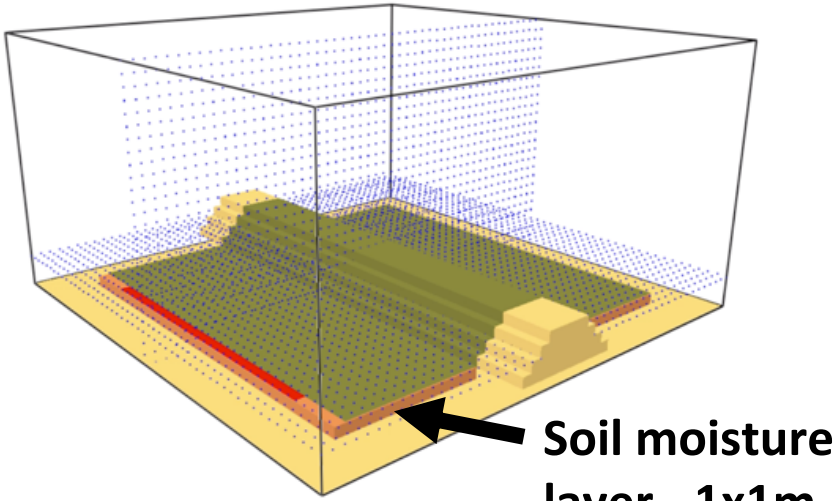
Data Generation

Elephant in the room: the soil moisture

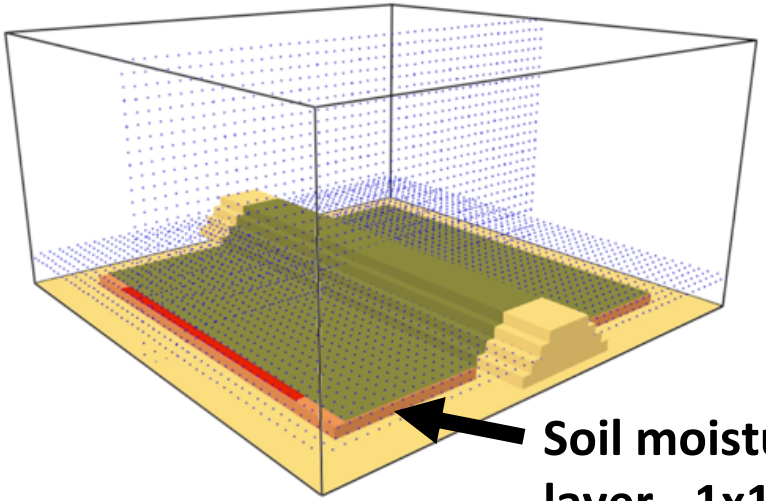
- Simulation of the 2016 Gatlinburg wildfire
- Software:
  - Fire Dynamics Simulator (FDS) - large-eddy simulation (LES) for low-speed flows
- Platform:
  - IBM Power9 cluster at UTK
- Simulation specs:
  - 120m×120m×100m domain
  - 5 frames/sec temp. resolution

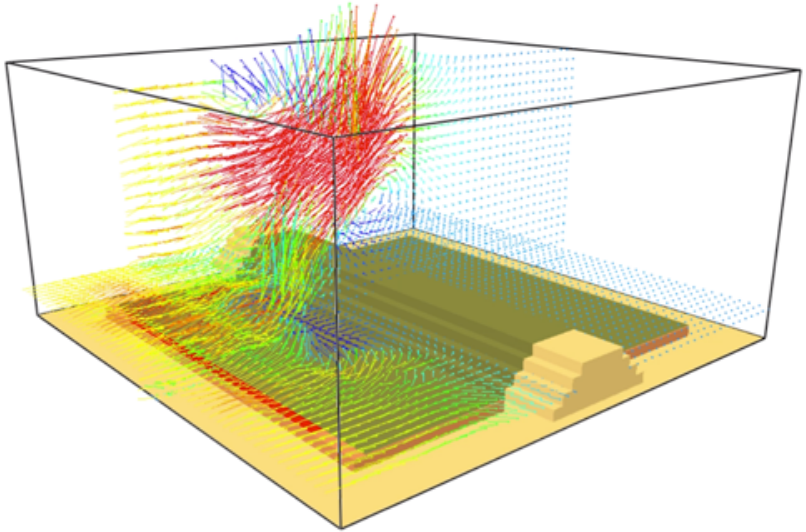


Soil moisture is missing in FDS

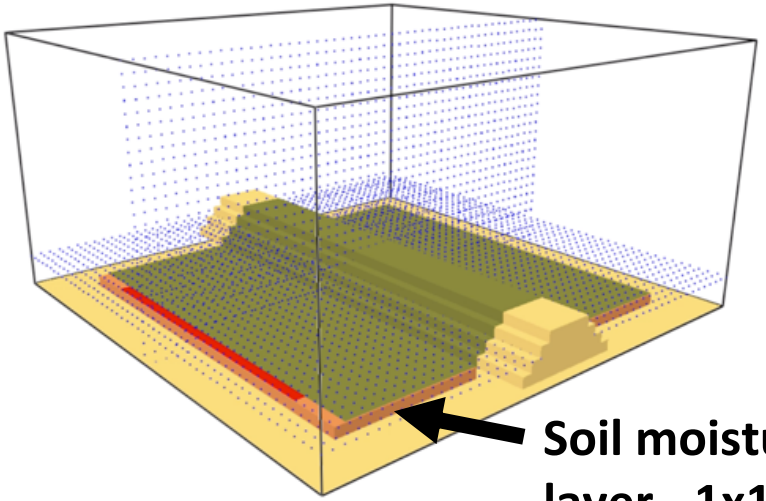


FDS simulations

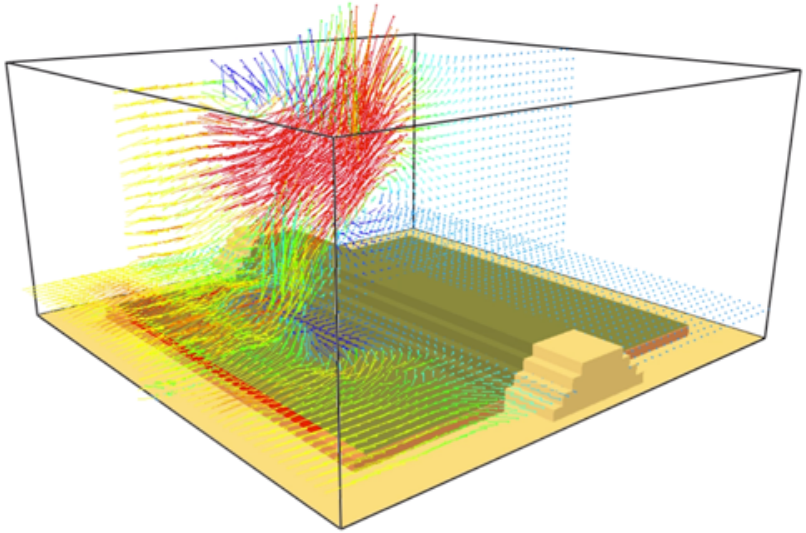


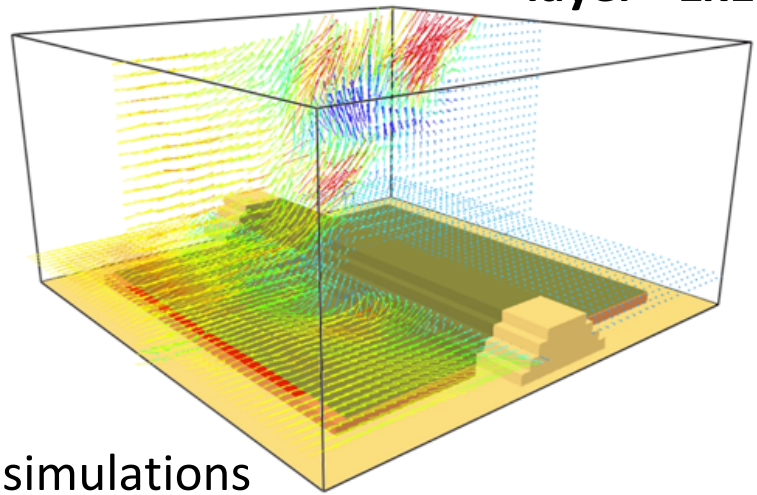


FDS simulations

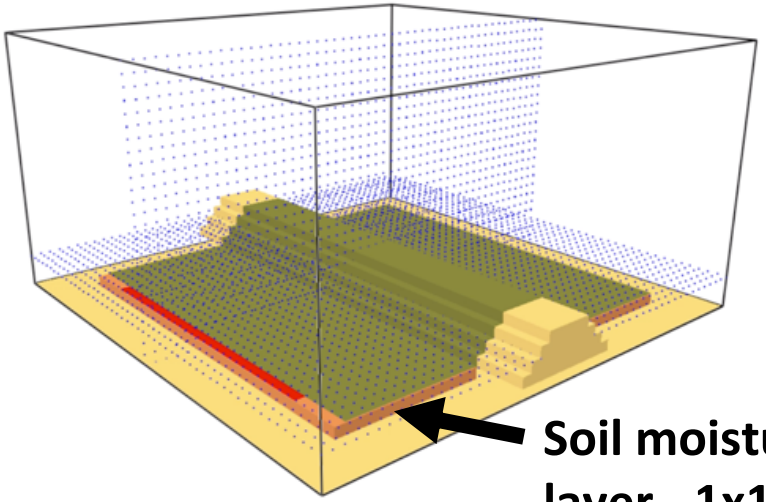


Soil moisture layer - 1x1m

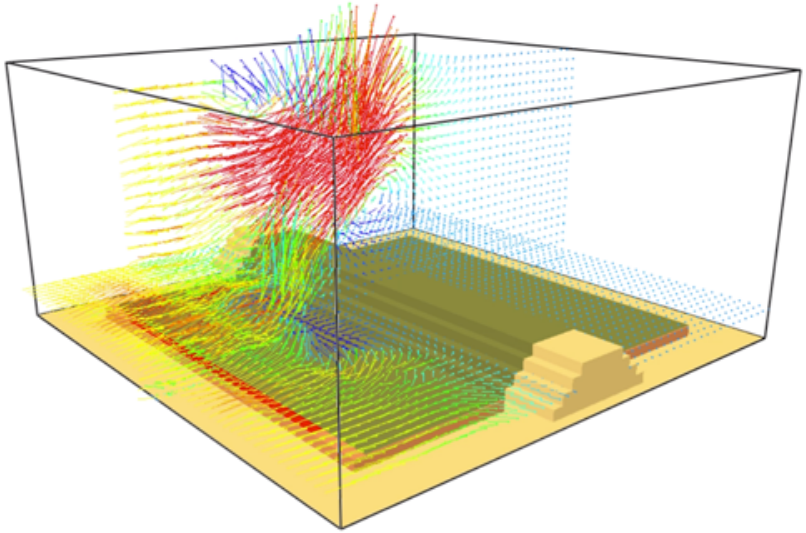


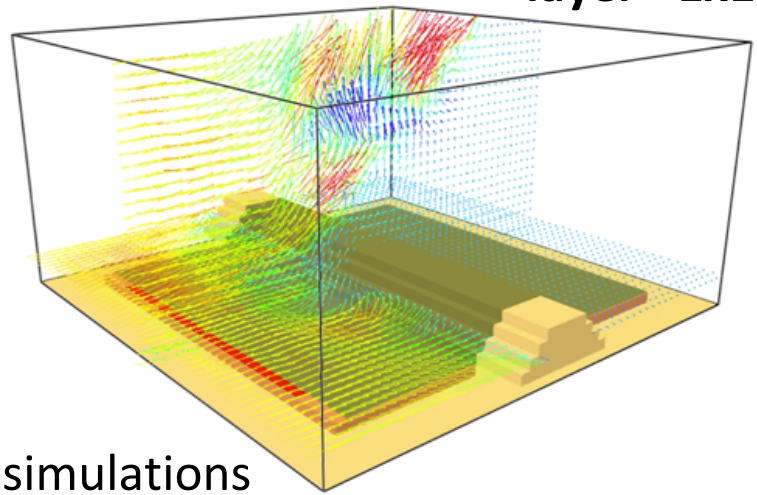


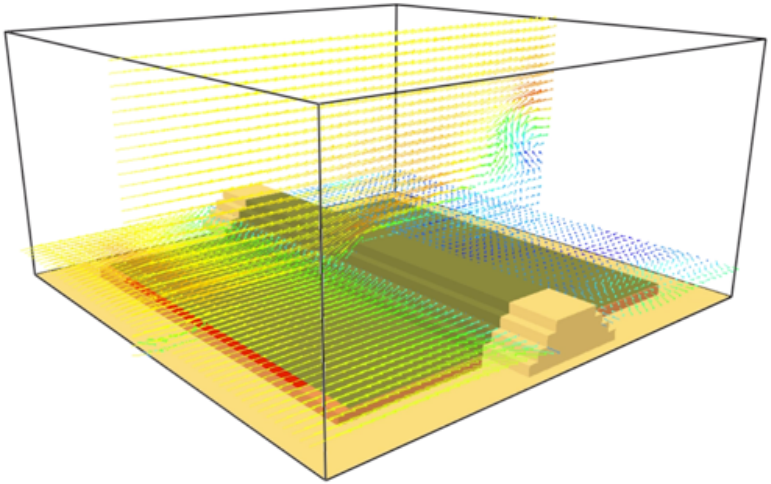
FDS simulations



Soil moisture layer - 1x1m







FDS simulations

Build trust in results through reproducibility, replicability, and transparency

Collaborator: Victoria Stodden's Group (UIUC)
Platform: NSF XSEDE Jetstream
NSF OAC 1941443 EAGER: Reproducibility and Cyberinfrastructure for Computational and Data-Enabled Science

Leveraging other NSF projects: Whole Tale

- Building an **open platform for computational reproducibility**
  - Create and publish **executable research objects** ("*Tales*")
- Simplify process of creating & verifying reproducible computational artifacts for scientific discovery

Easy-to-access **cloud-based** computing **environments**

Transparent access to research **data**

Export and **publish executable research objects**

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
THE UNIVERSITY OF CHICAGO
UCSB UNIVERSITY OF CALIFORNIA SANTA BARBARA
TEXAS The University of Texas at Austin
UNIVERSITY OF NOTRE DAME

This material is based upon work supported by the National Science Foundation under Grant No. OAC-1541450

BIG ORANGE BIG IDEAS

WHOLE TALE DASHBOARD BROWSE MANAGE

WHOLE TALE Dashboard

New to WholeTale? Learn more about using the Dashboard...

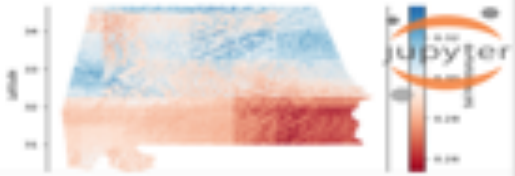
Create New...

Tales

All Tales My Tales

Search tales...

Original Tale PRIVATE



MACHINE LEARNING

SOMOSPIE: A modular SOil MOisture S...

by Danny Rorabaugh, Mario Guevara, Ricardo Llamas, Joy Kitson, Rodrigo Vargas, Michela Taufer

Run Tale

Original Tale PUBLIC

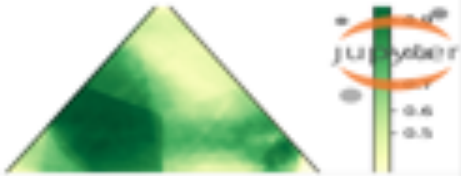
MATERIALS SCIENCE

Accelerated discovery of metallic g...

by Logan Ward

Run Tale

Original Tale PUBLIC



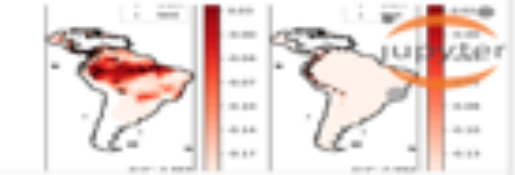
MATERIALS SCIENCE

Predicting the Properties of Inorga...

by Logan Ward

Run Tale

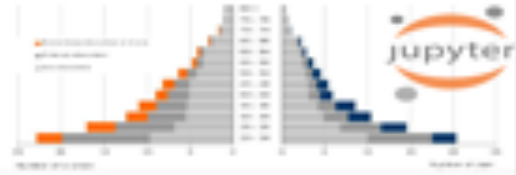
Original Tale PUBLIC



ECOLOGY

Replication of a classical ecologic...

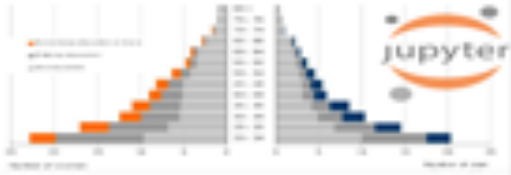
Original Tale PUBLIC



SCIENCE

Concept mapping

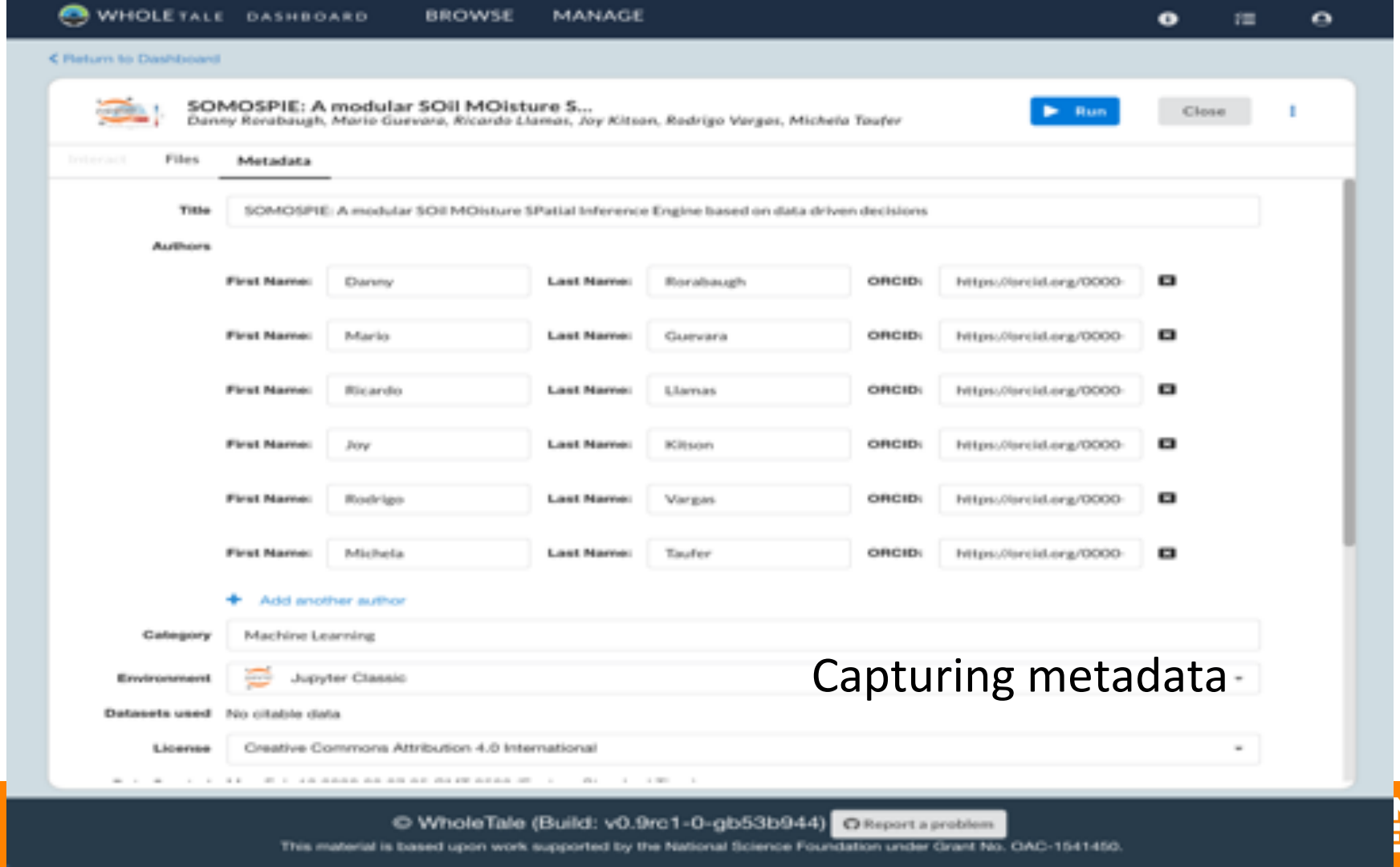
Original Tale PUBLIC



SCIENCE

Species co-occurrence distribution ...

This material is based upon work supported by the National Science Foundation under Grant No. OAC-1541450.


WHOLETALE DASHBOARD BROWSE MANAGE
Return to Dashboard
SOMOSPIE: A modular SOil MOisture S...
Danny Rorabaugh, Mario Guevara, Ricardo Llamas, Joy Kitson, Rodrigo Vargas, Michela Taufer
Run
Close
Files
Metadata
Title
SOMOSPIE: A modular SOil MOisture SPatial Inference Engine based on data driven decisions
Authors
First Name: Danny
Last Name: Rorabaugh
ORCID:
First Name: Mario
Last Name: Guevara
ORCID:
First Name: Ricardo
Last Name: Llamas
ORCID:
First Name: Joy
Last Name: Kitson
ORCID:
First Name: Rodrigo
Last Name: Vargas
ORCID:
First Name: Michela
Last Name: Taufer
ORCID:
Add another author
Category
Machine Learning
Environment
Jupyter Classic
Datasets used
No citable data
License
Creative Commons Attribution 4.0 International
Capturing metadata
© WholeTale (Build: v0.9rc1-0-gb53b944)
Report a problem
This material is based upon work supported by the National Science Foundation under Grant No. OAC-1541450.

jupyter SOMOSPIE Last Checkpoint: 01/28/2020 (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

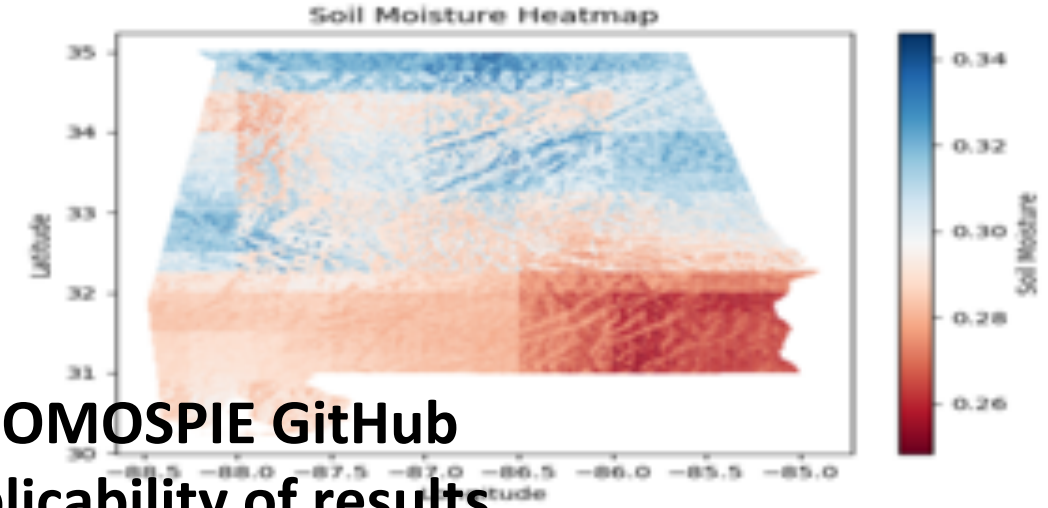
Not Trusted | Python 3

Run Markdown

VBox(children=(H…

In [22]:
```
try:
    visualizer.display_output()
except ValueError:
    print("You must select an output folder in the widget above before running this cell.")
```

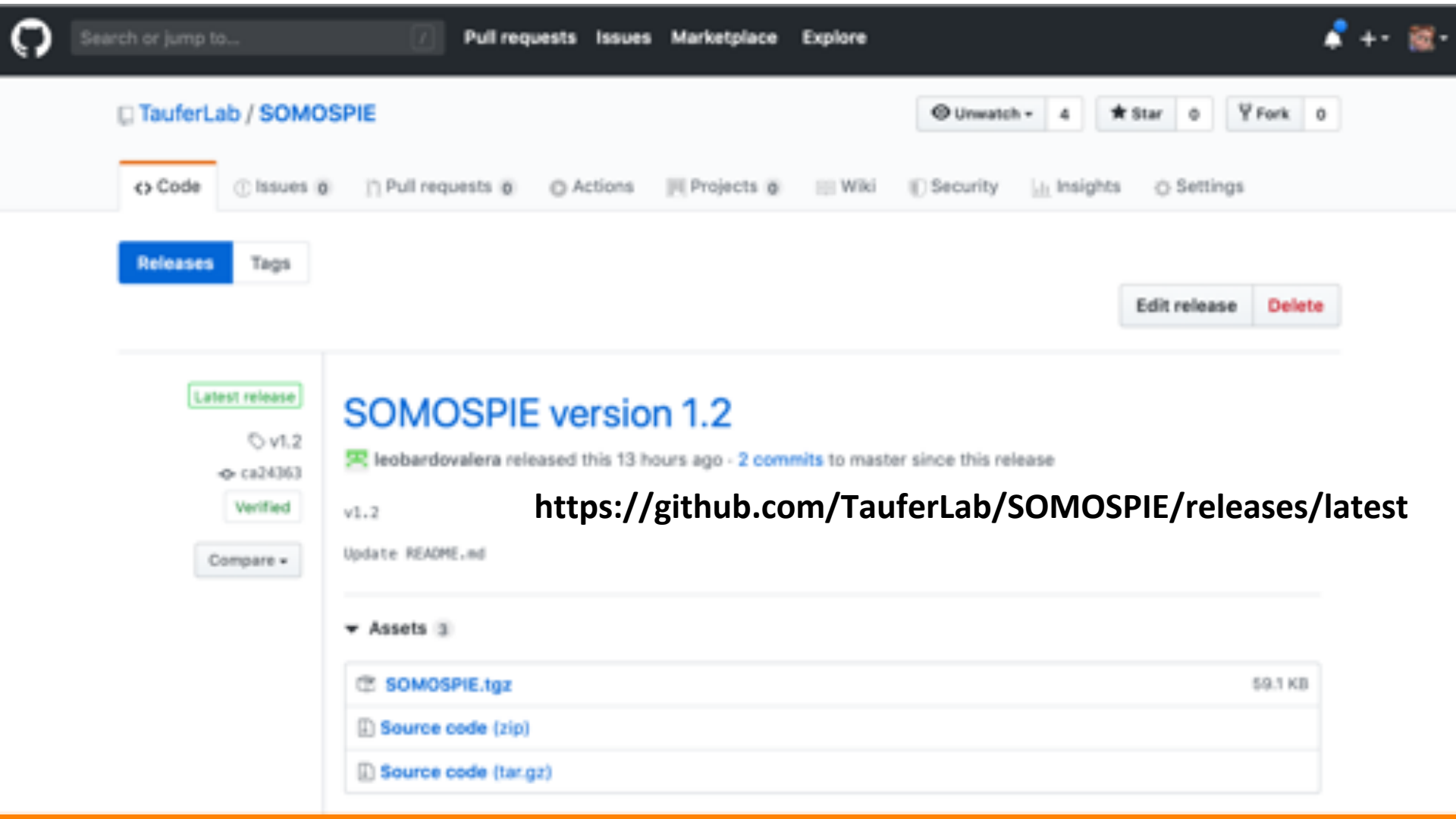
```
{0: {'START': '../', 'CODE': '../code/', 'DATA': '../data/', 'OUTPUT': '../out/', 'CO
V_FILE': 'topo_predictors/stack.tif', 'SM_FILE': 'ESA_CCI/2017_ESA_monthly.rds', 'EVAL_FI
LE': '', 'MAKE_T_E': 1, 'USE_PCA': 0, 'RAND_SEED': 0, 'USE_VIS': 1, 'BUFFER': 0, 'SUPER':
0, 'MIN_TEST_POINTS': 11, 'VALIDATE': 1.3, 'REG_LIST': [{'STATE', 'Alabama'}], 'MONTHS': [1],
'MODICT': {'RF': {}, 'UNMODEL': {}}, 'YEAR': [2017], 'COV_LAYERS': ['Aaspect'], 'TRAIN_DIR':
'2017/t', 'EVAL_DIR': '2017/e'}, 1: {1: {'seed': 65050}}, 2: 1580838903.2455065, 3: 22.56
0522794723511}
{1: {'Alabama': '/home/jovyan/work/workspace/out/job_2020_02_04_17_54_41/2017/month1-0.30_650
50/Alabama'}}
```



Method completion time = t=8.003055095672607

Plug into SOMOSPIE GitHub

Enable replicability of results


Search or jump to...
Pull requests
Issues
Marketplace
Explore
TauferLab / SOMOSPIE
Unwatch 4
Star 0
Fork 0
Code
Issues 0
Pull requests 0
Actions
Projects 0
Wiki
Security
Insights
Settings
Releases
Tags
Edit release
Delete
Latest release
v1.2
ca24363
Verified
Compare
SOMOSPIE version 1.2
leobardovalera released this 13 hours ago · 2 commits to master since this release
v1.2
Update README.md
https://github.com/TauferLab/SOMOSPIE/releases/latest
Assets 3
SOMOSPIE.tgz
59.1 KB
Source code (zip)
Source code (tar.gz)