Webinar 5: Spatial Prediction Using Remote Sensing

Ian Breckheimer, PhD
RMBL Staff Research Scientist
@ibreckhe | ikb@rmbl.org

Webinar Schedule

Tuesday September 22nd 2020
Introduction to the RMBL Spatial Data Platform,
How to access RMBL SDP data in GIS and programming environments, and where we are going with the platform.

Tuesday October 20th 2020 *Designing Robust Field Studies using Geospatial Tools*
How to optimize site selection using GIS and the RMBL SDP.

Tuesday January 26th, 2021 *Successful UAV Data Collection in Mountain Environments*
How to design and execute UAV flights for high-quality scientific data in challenging environments.

Tuesday February 23rd, 2021 *Leveraging Point Cloud Data from Lidar and UAV Photogrammetry*
Mapping vegetation structure and function using 3D data from lidar and drones.

Tuesday March 23rd, 2021 *Linking Field Data with Remote Sensing for Spatial Prediction*
How to leverage high-resolution remote sensing from imaging spectroscopy and lidar to map species, traits, and processes.

Tuesday April 20th, 2021 *What's New in the RMBL Spatial Data Platform*
Introduction to new snow and phenology datasets that form part of the SDP Release 2 and Release 3.

Outline

- **Why?**

  Why should you use your field data to make maps?

  What makes spatial prediction a unique challenge?

- **What?**

  What core concepts do I need to know to succeed?

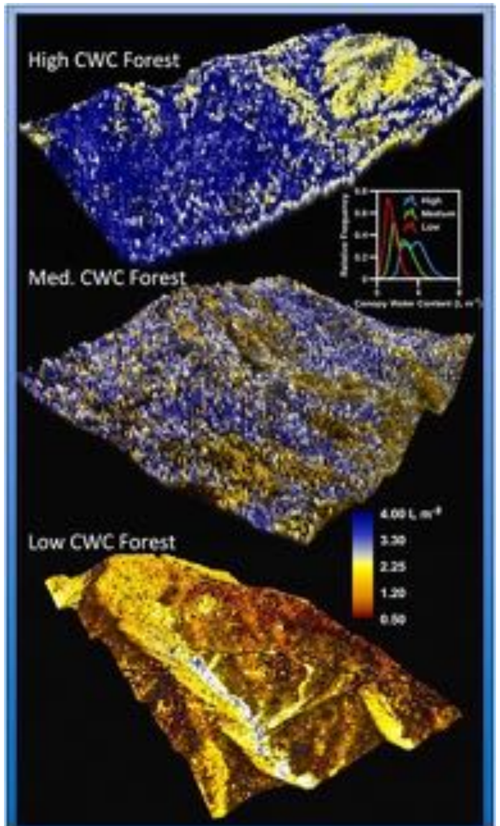
  What is the basic workflow for spatial prediction?

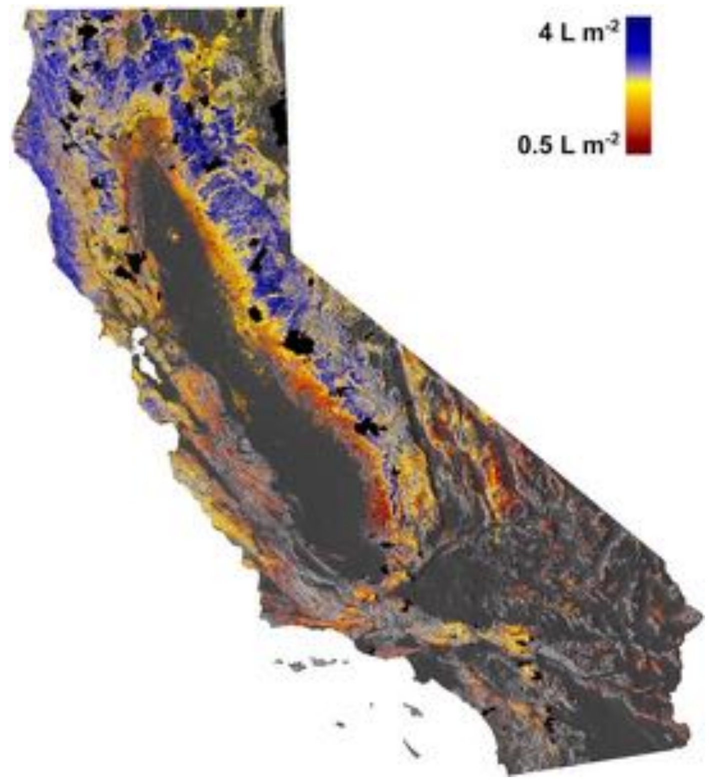
  What are some common missteps?

- **How?**

  Case Study: vegetation deciduousness from imagery and LiDAR data.

Why? Scaling up



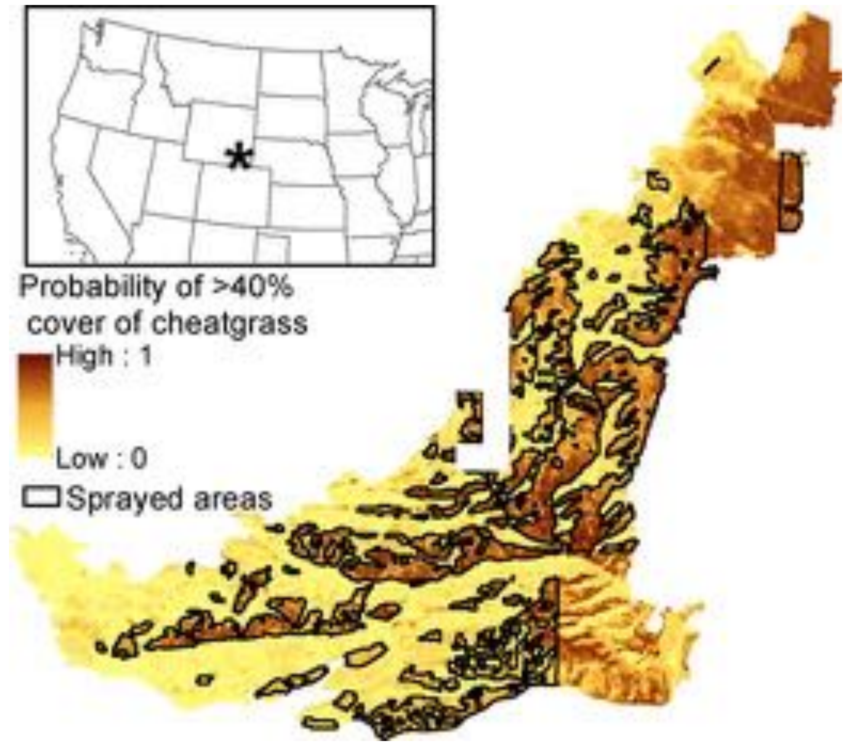


Figures: Asner et al. 2016

Why? Filling in missing covariates

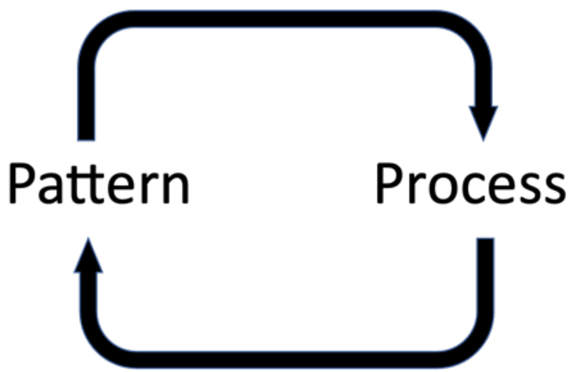
Figure: USA NPN

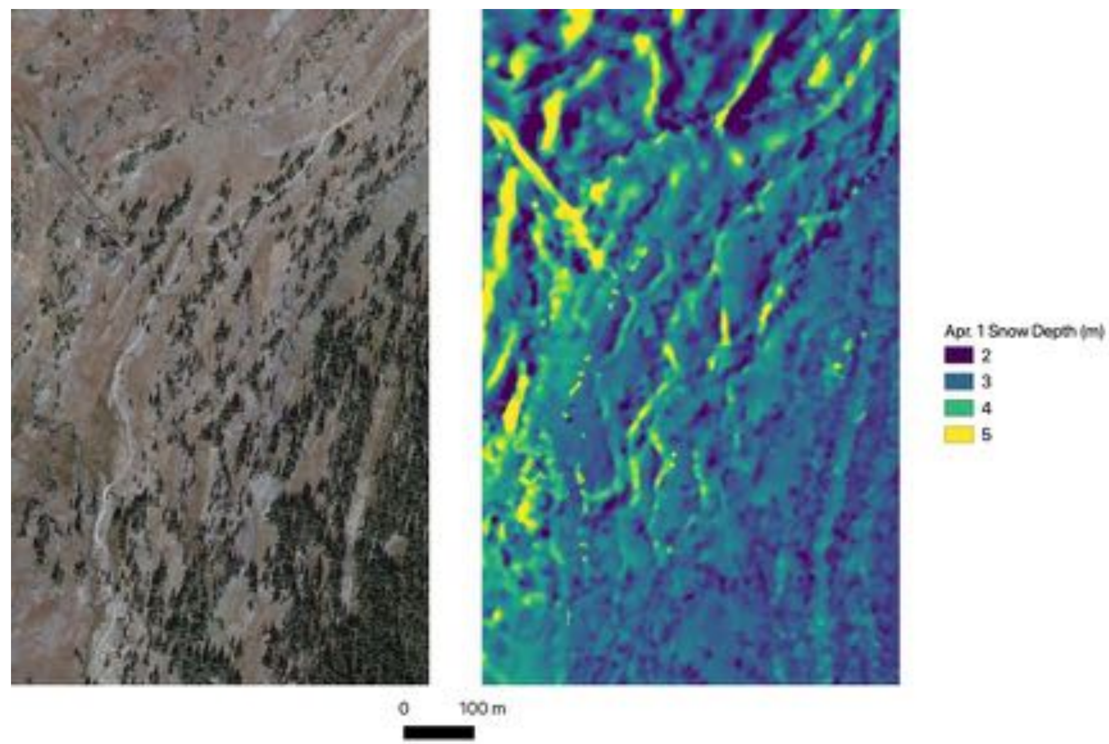
Why? Guiding field work



Images: Sofaer et al. 2019, CSU

Why? Pattern-process relationships





Outline

- **Why?**

Why should you use your field data to make maps?

What makes spatial prediction a unique challenge?

What core concepts do I need to know to succeed?

- **What?**

What is the basic workflow for spatial prediction?

What are some common missteps?

Case Study: vegetation deciduousness from imagery and LiDAR data.

- **How?**

Challenges: data volume

~47 million values

3448 rows ->

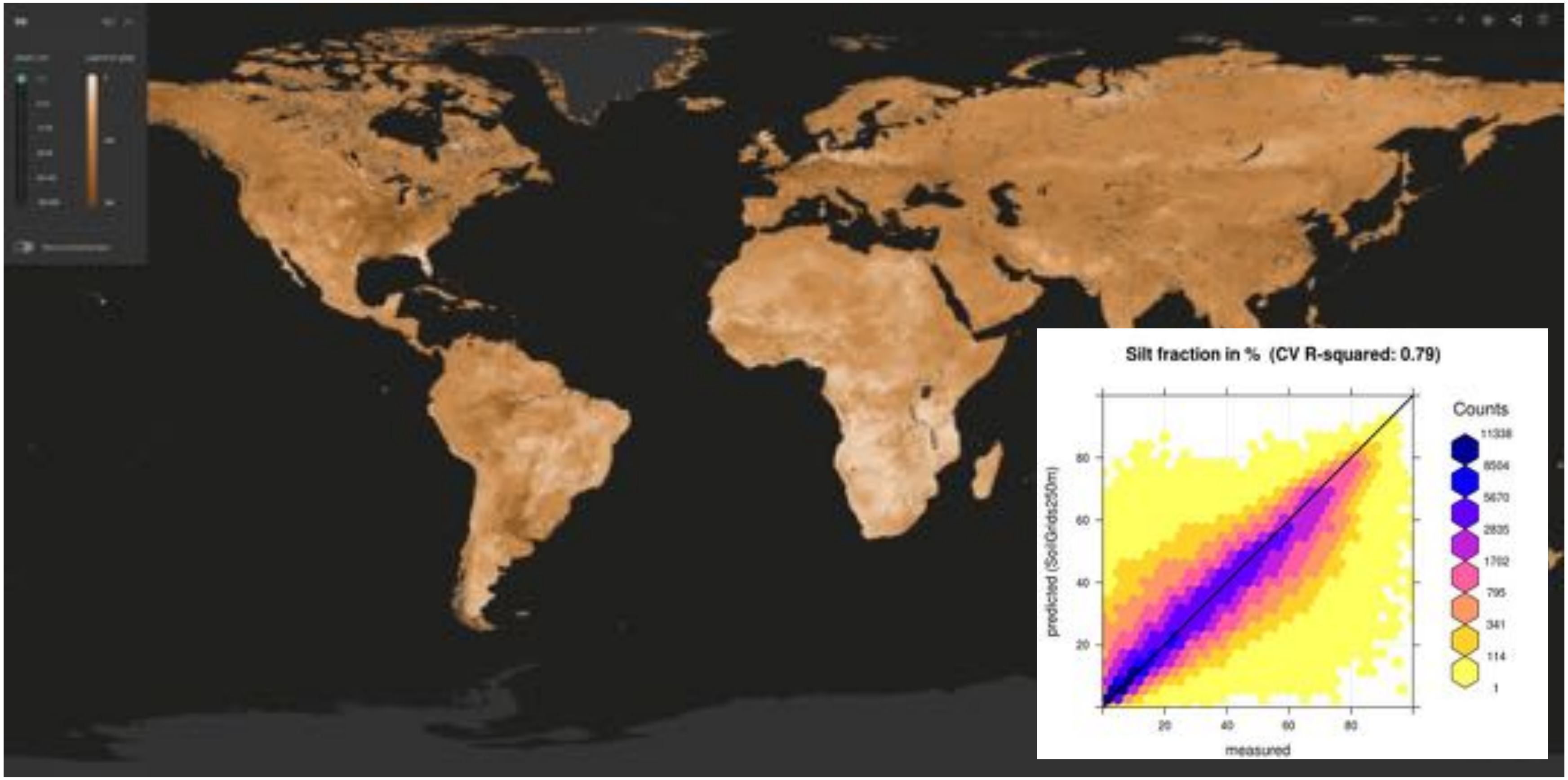
4592 columns ->

x 3 bands

Challenges: interpolation and extrapolation

Image: NPR

Challenges: dealing with uncertainty



Images: soilgrids.org, Hengl et al. 2017

Outline

- **Why?**

Why should you use your field data to make maps?

What makes spatial prediction a unique challenge?

- **What?**

What core concepts do I need to know to succeed?

What is the basic workflow for spatial prediction?

What are some common missteps?

- **How?**

Case Study: vegetation deciduousness from imagery and LiDAR data.

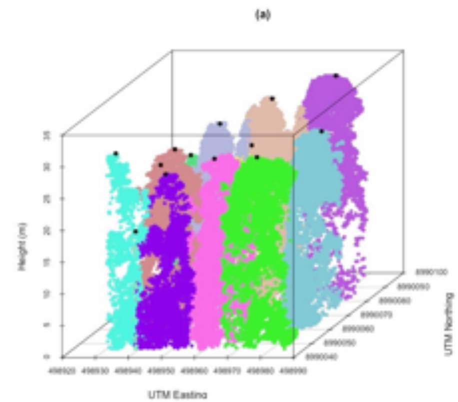
Core concept: inference vs prediction

Inference

Core concept: features (predictors or covariates)

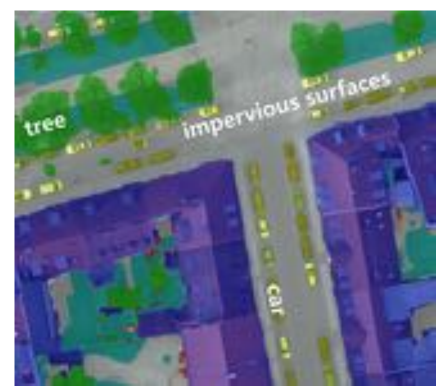
Spectral Features

Structural Features



Temporal Features

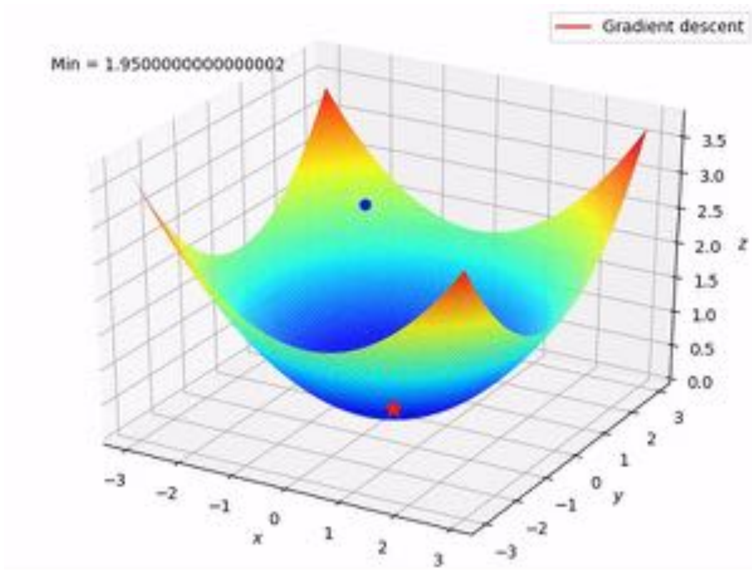
Semantic Features



Images: Lisa Kimmorly, Millikan et al. 2019, GroundAI

Core concept: loss function

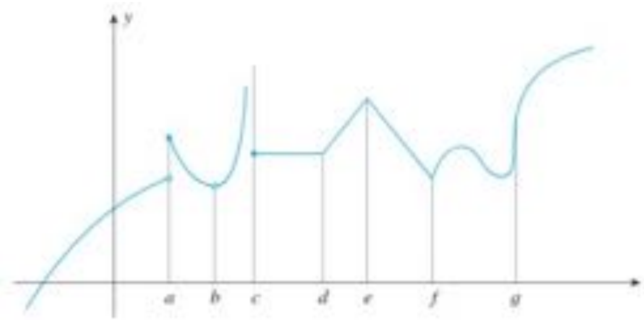
Loss = "badness"



Loss functions encode goals

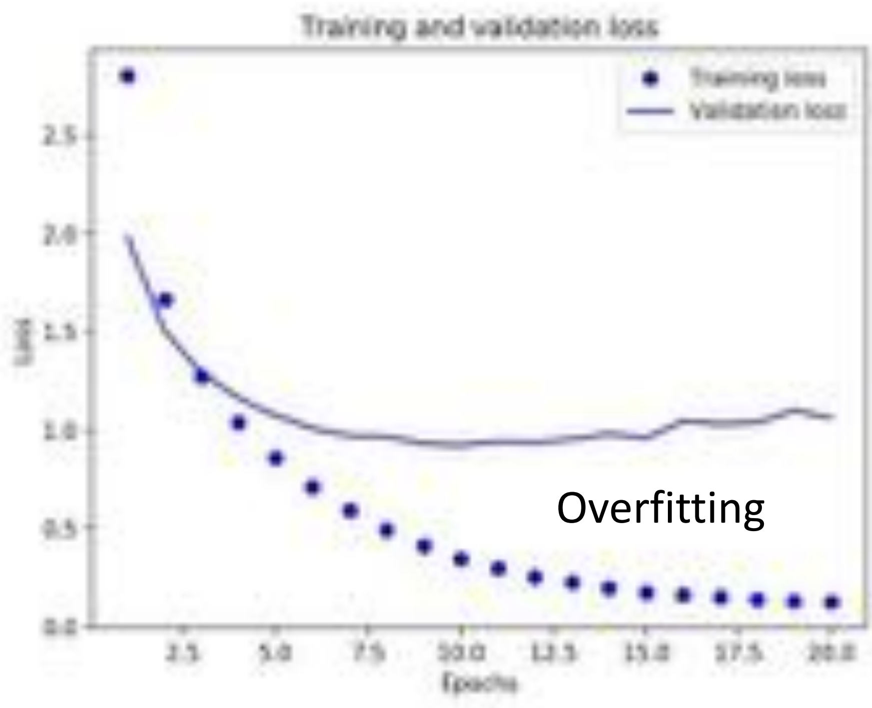
$$CE = -\sum_{i=1}^{C'=2} t_i log(s_i) = -t_1 log(s_1) - (1 - t_1) log(1 - s_1)$$

Loss functions must be differentiable

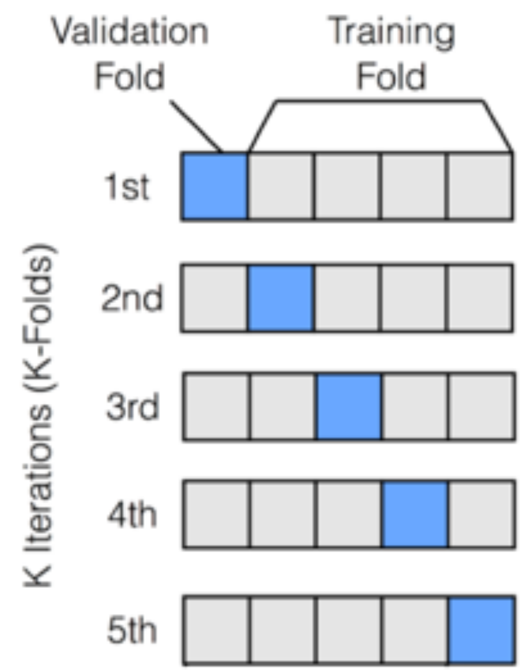


Images: Athimung S, Applied Calculus, 8th ed

Core concept: fitting and overfitting



Cross-validation



Images: Yash Kandewahl

Core concept: precision and bias

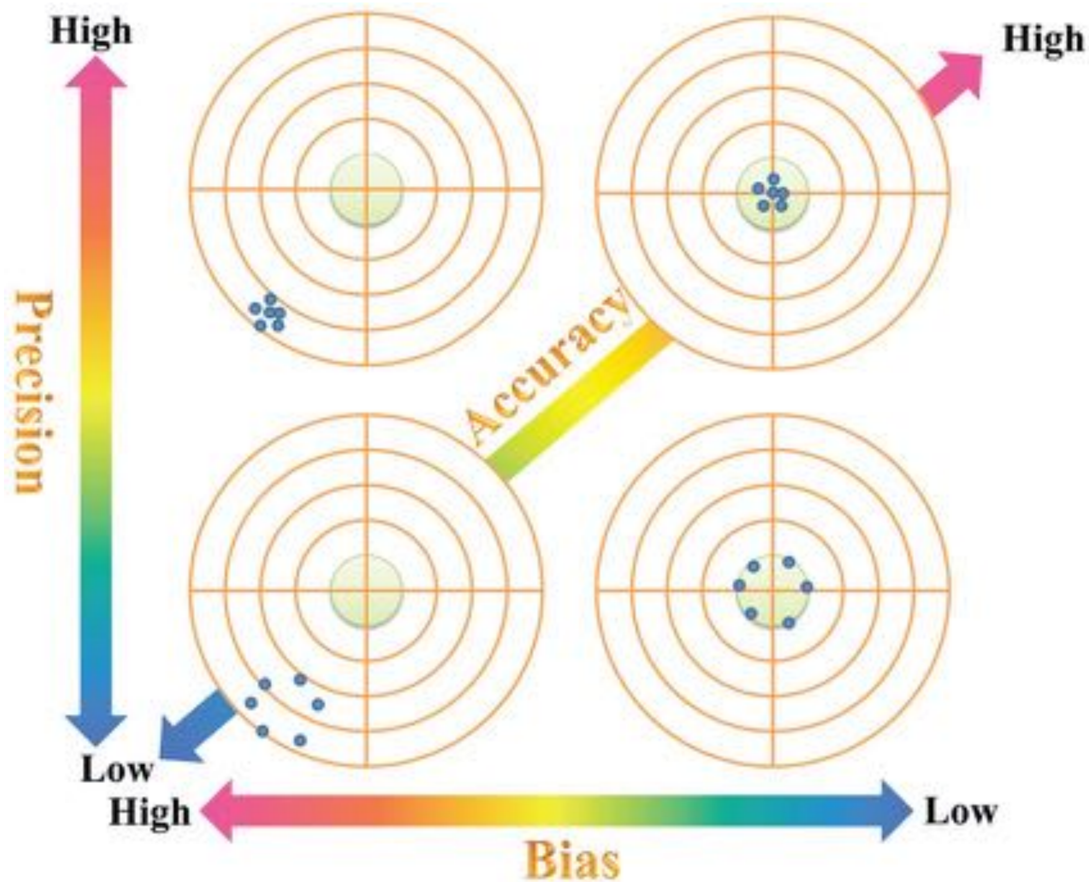


Image: Chen et al. 2018

Core concepts summary

- Machine learning models use **features** to learn a task and make **predictions** by minimizing a **loss function**. They don’t care about statistical significance.

- These models are really good at making predictions on training data, but can **overfit**. This means you always want to evaluate a model on independent data, sometimes by **cross-validation**.

- An **accurate** model gives **precise** and **unbiased** predictions.

Outline

- **Why?**

Why should you use your field data to make maps?

What makes spatial prediction a unique challenge?

- **What?**

What core concepts do I need to know to succeed?

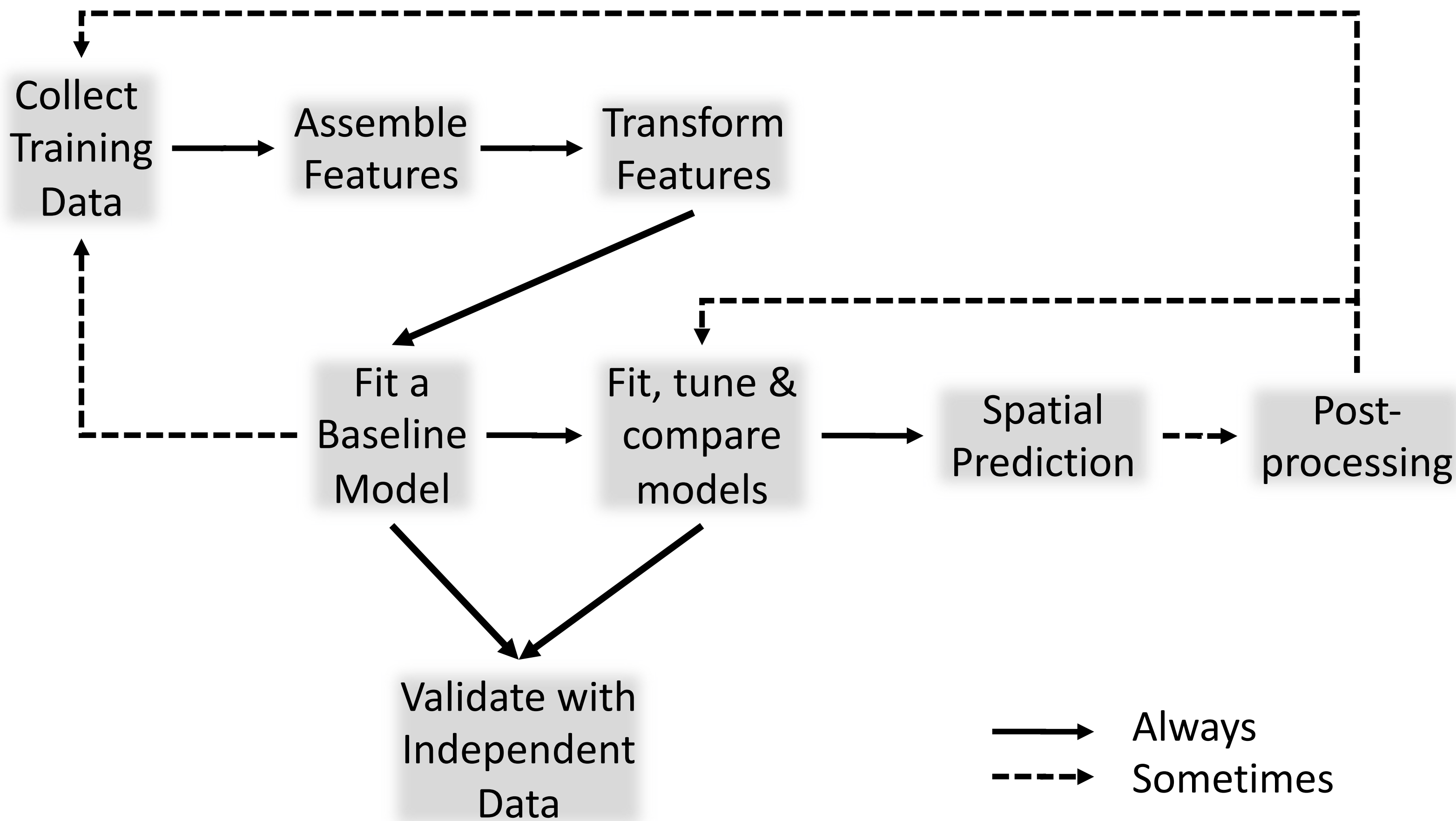
$\boxed{\text{What is the basic workflow for spatial prediction?}}$

What are some common missteps?

- **How?**

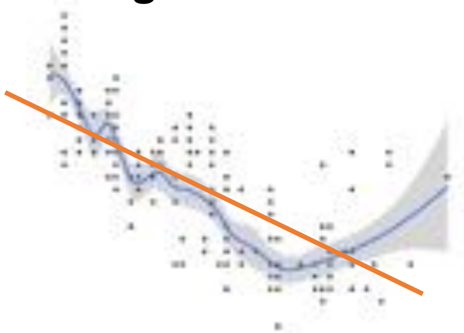
Case Study: vegetation deciduousness from imagery and LiDAR data.

General spatial prediction workflow



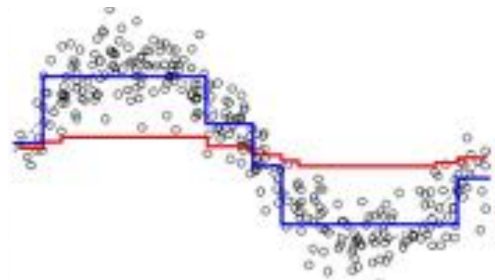
Approaches to explore

Regression-based



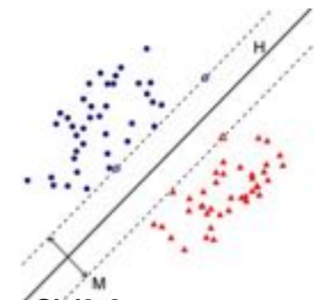
GLM, GLMM, GAM

Tree-based



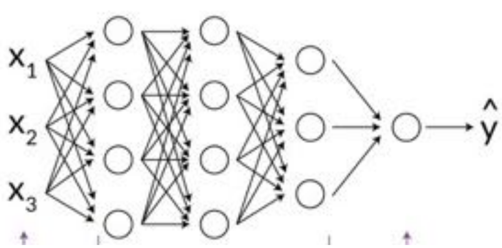
Random Forest, BRT

Vector-based



SVM

Neural networks



MLP, DNNs

Geostatistical

Kriging, IDW Interpolation

Approaches to explore

Regression-based

+structure
+uncertainty

-interactions
-overfits

GLM, GLMM, GAM

Tree-based

+fast
+thresholds

-artifacts
-overfits

Random Forest, BRT

Vector-based

+interactions

-artifacts
-uncertainty

SVM

Neural networks

+big data
+complexity

-big data
-can be slow
-reproducibility

MLP, DNNs

Geostatistical

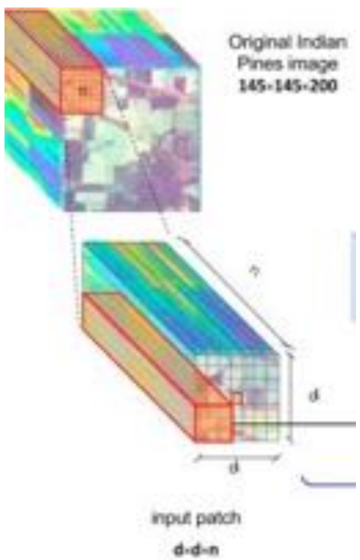
+geodata
+sparse

-slow
-oversmooth

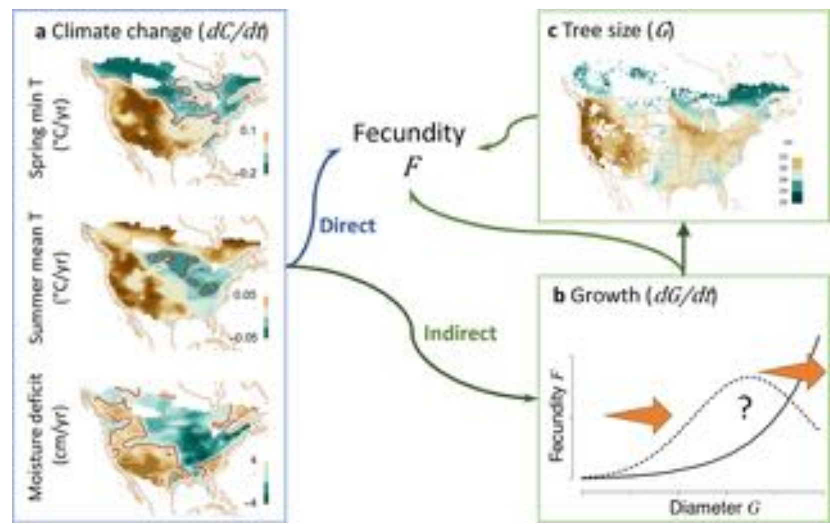
Kriging, IDW
Interpolation

On the rise

Fast ConvNets for multispectral data (e.g. Hyper3Dnet, Morales et al. 2020)



Combining statistical and process models (Clark et al. 2021)



Outline

- **Why?**

Why should you use your field data to make maps?

What makes spatial prediction a unique challenge?

- **What?**

What core concepts do I need to know to succeed?

What is the basic workflow for spatial prediction?

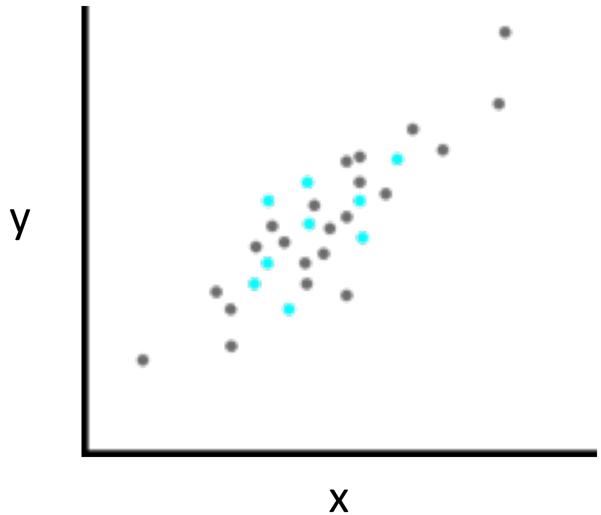
What are some common missteps?

- **How?**

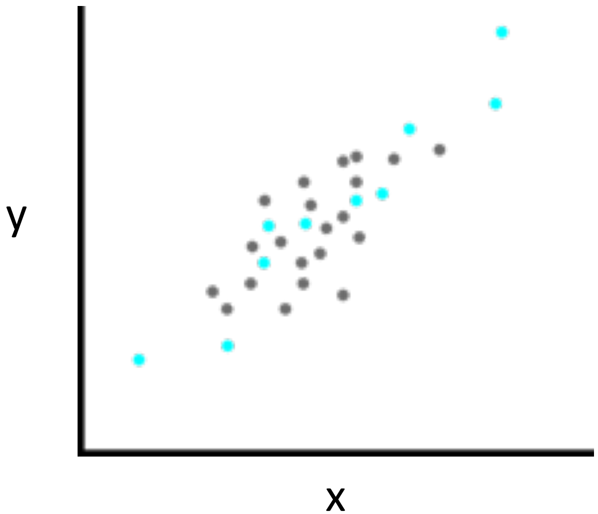
Case Study: vegetation deciduousness from imagery and LiDAR data.

Mis-step: inadequate training data

Feature coverage matters more than sample size

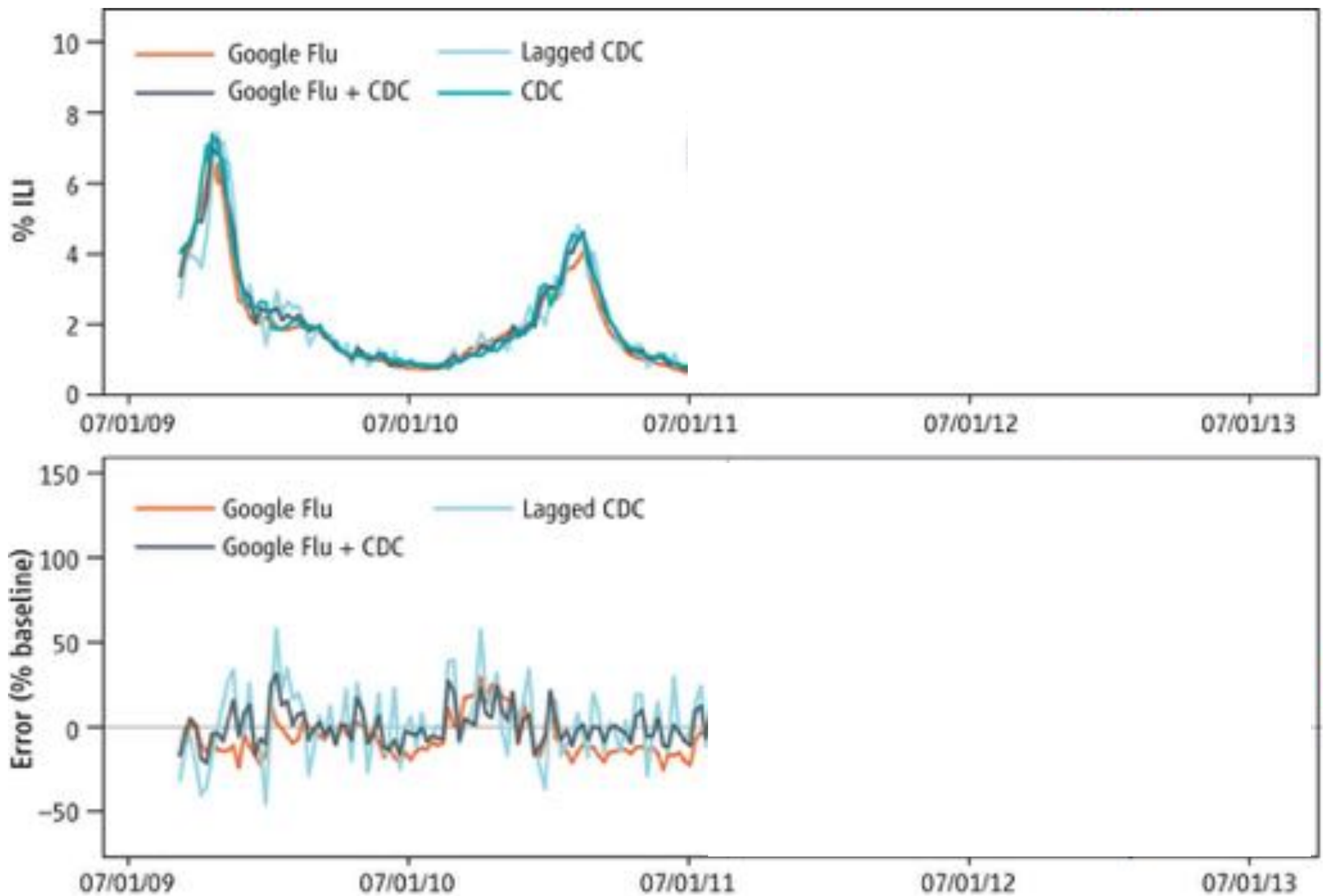


Good representativeness, low coverage



Good coverage, poor representativeness

Mis-step: ignoring non-stationarity



Miss-step: scope and scale mismatch

Images: soilgrids.org

Outline

- **Why?**

Why should you use your field data to make maps?

What makes spatial prediction a unique challenge?

What core concepts do I need to know to succeed?

- **What?**

What is the basic workflow for spatial prediction?

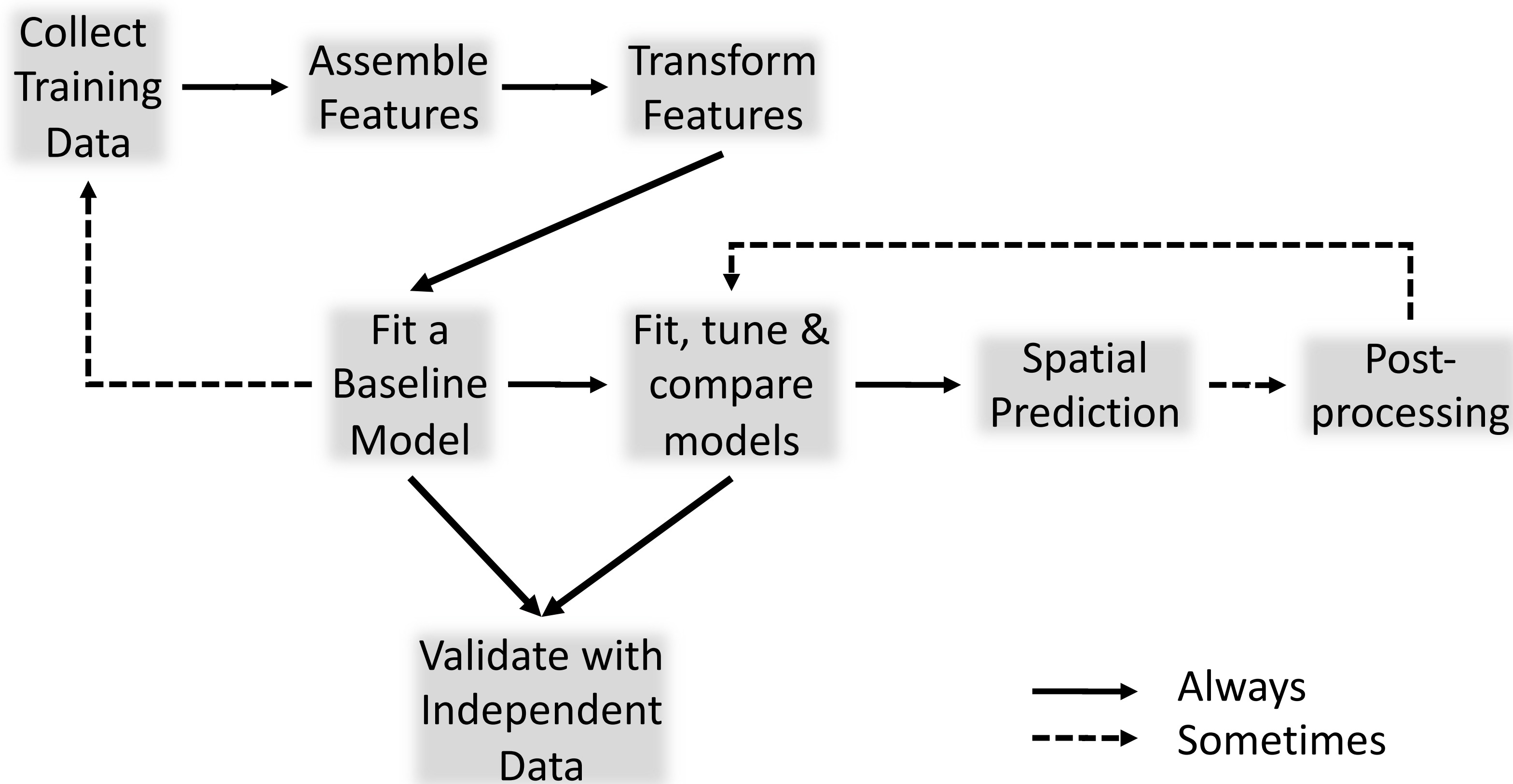
What are some common missteps?

- **How?**

Case Study: vegetation deciduousness from imagery and LiDAR data.

Case Study: mapping deciduousness

Case Study: mapping deciduousness



Case Study: mapping deciduousness

tidymodels.org

Thanks!

Contact Me:

Ian Breckheimer
ikb@rmbl.org
Twitter, Github: @ibreckhe

References

- Asner, G. P., Brodrick, P. G., Anderson, C. B., Vaughn, N., Knapp, D. E., & Martin, R. E. (2016). Progressive forest canopy water loss during the 2012–2015 California drought. *Proceedings of the National Academy of Sciences*, *113*(2), E249-E255.
- Sofaer, H. R., Jarnevich, C. S., Pearse, I. S., Smyth, R. L., Auer, S., Cook, G. L., ... & Hamilton, H. (2019). Development and delivery of species distribution models to inform decision-making. BioScience, 69(7), 544-557.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLoS one, 12(2), e0169748.
- Chen, Y., Wu, Y., & Shen, T. J. (2018). Evaluation of the estimate bias magnitude of the Rao’s quadratic diversity index. PeerJ, 6, e5211.
- Morales, G., Sheppard, J. W., Scherrer, B., & Shaw, J. A. (2020). Reduced-cost hyperspectral convolutional neural networks. Journal of Applied Remote Sensing, 14(3), 036519.
- Clark, J. S., Andrus, R., Aubry-Kientz, M., Bergeron, Y., Bogdziewicz, M., Bragg, D. C., ... & Zlotin, R. (2021). Continent-wide tree fecundity driven by indirect climate effects. Nature communications, 12(1), 1-11.