University of Idaho
College of Natural Resources

FOREST MANAGEMENT WITH AI AND MACHINE LEARNING

IFC REGIONAL AND NATIONAL SDImax MODELING

JASLAM POOLAKKAL, MARK KIMSEY

GMUG
Growth Model User's Group

APRIL 16, 2025

Intermountain Forestry Cooperative

RESEARCH SUPPORT

PRIVATE-PUBLIC PARTNERSHIP

INDUSTRY	COUNTY	STATE	FEDERAL
PotlatchDeltic; Manulife; MOLPUS WOODLANDS GROUP; STIMSON LUMBER COMPANY; NewForests; Bennett Lumber Products, Inc.; IDAHO FOREST GROUP	HOOD RIVER COUNTY	IDAHO DEPARTMENT OF LANDS; WASHINGTON DEPARTMENT OF NATURAL RESOURCES	FOREST SERVICE U.S. DEPARTMENT OF AGRICULTURE; U.S. DEPARTMENT OF THE INTERIOR BUREAU OF LAND MANAGEMENT; Center for Advanced Forestry Systems

SETTING THE CONVERSATION

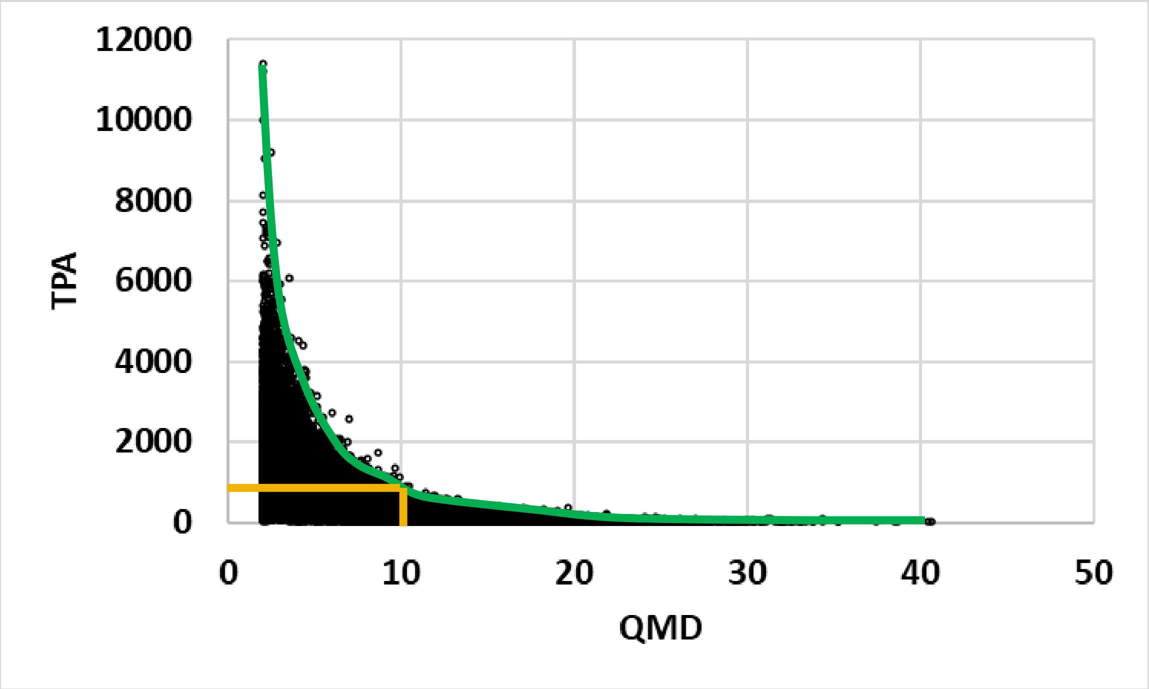
HOW DOES SDIMAX MODELS INFORM MANAGEMENT DECISIONS?

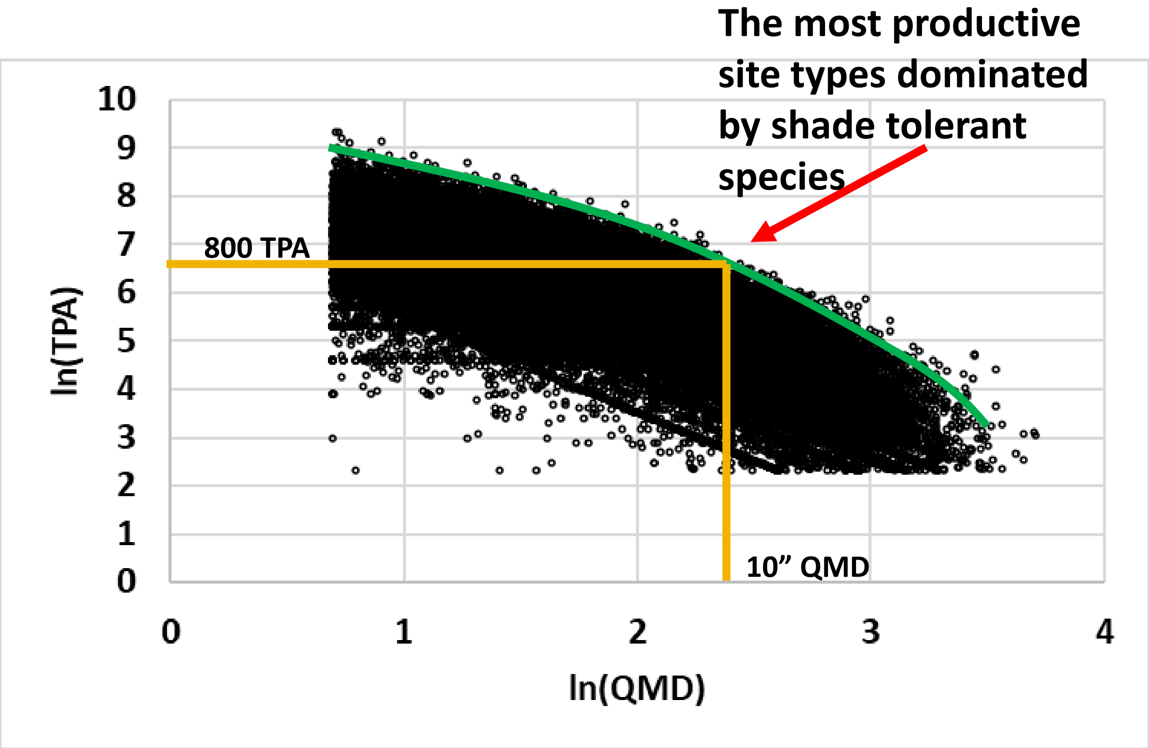
- *We need tools that spatially quantify expert knowledge that can address questions like:*
  - *How many trees are too many?*
  - *How is "too many" related to stand development stage, species composition, site type (climate, soil, geology, topography)?*
  - *How can we optimize #trees/acre to meet management objectives through initial planting density, PCT or CT, site-species relationships, while minimizing density induced growth reduction/mortality*
  - *Does projected climate patterns suggest modifying current management strategies?*

LET'S VISUALLY DEFINE SDI$_{MAX}$

STAND DENSITY INDEX (SDI) = ƒ(QUADRATIC MEAN DIAMETER, TREES PER ACRE)

$$SDI = TPA \times (QMD/10)^{1.605}$$

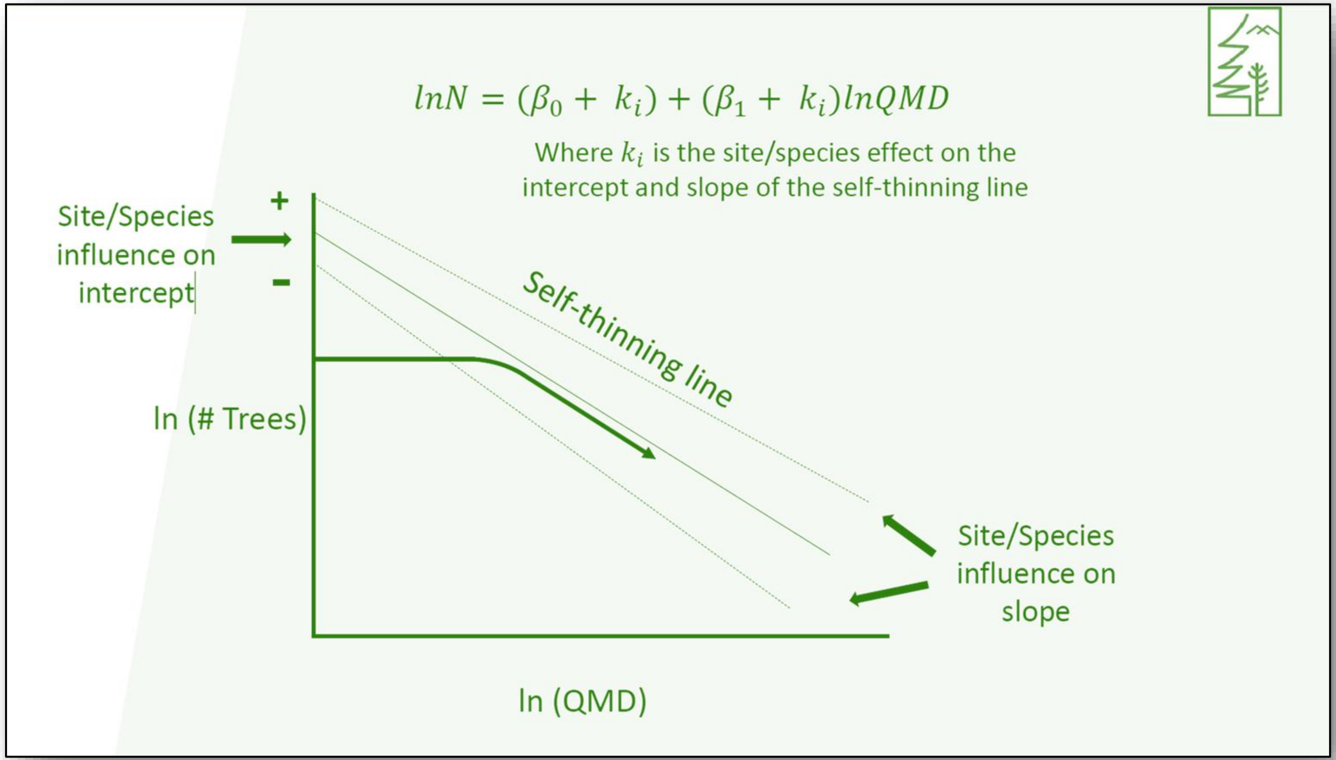




Maximum SDI referenced to a stand with a QMD of 10 inches = ~800 TPA

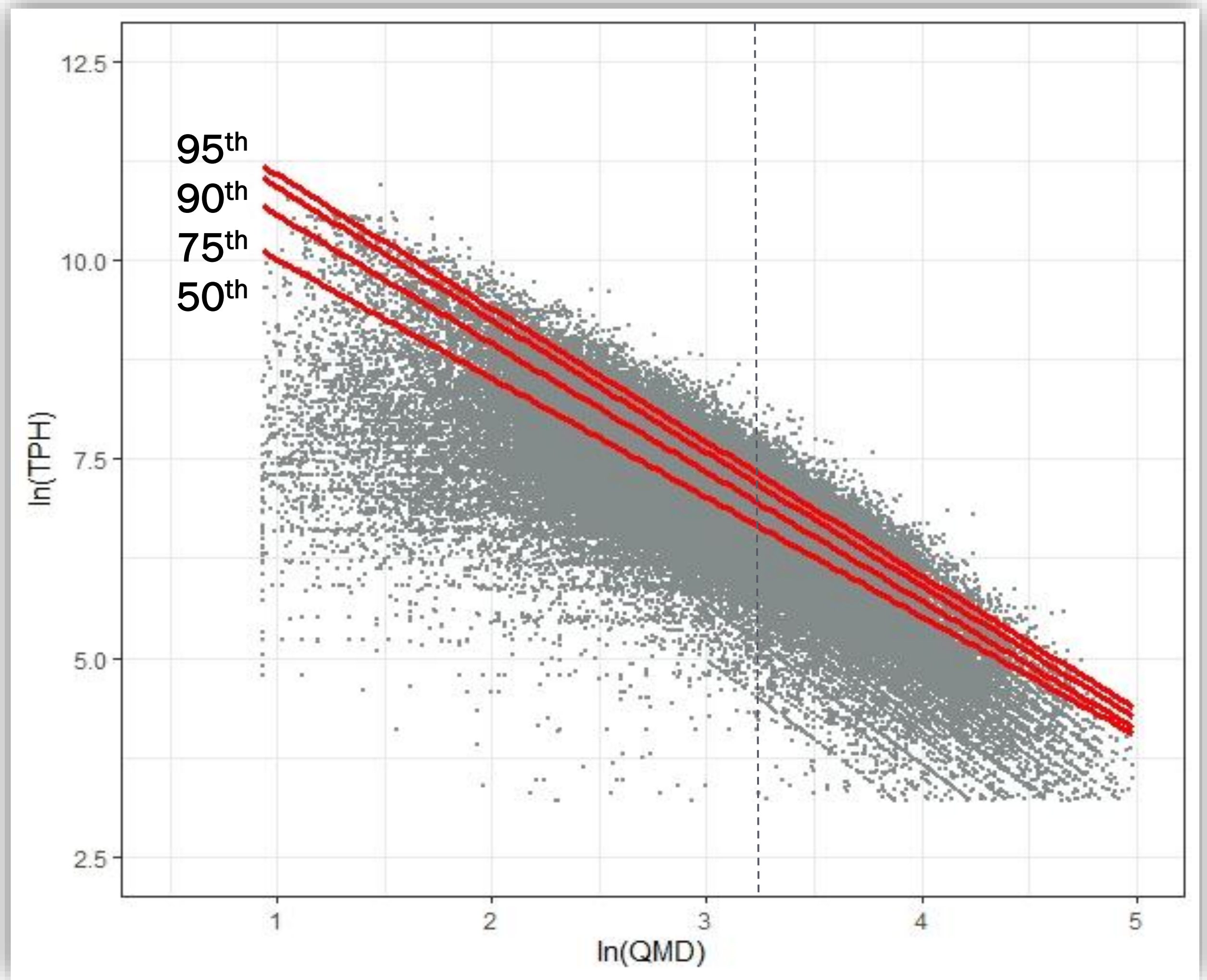
SDImax Modeling - Model Versions

- ❑ Reineke (1933) first described the self-thinning line utilizing log-log paper and visually fit a reference curve to the outer boundary.
- ❑ Ordinary least squares
- ❑ Linear mixed models
- ❑ Quantile regression
- ❑ Linear quantile mixed models
- ❑ **Frontier analysis**



HOW DOES SITE QUALITY IMPACT SDI?

WHERE IS SDIMAX IN THIS CLOUD OF DATA POINTS?



- ***SDI – a function of QMD & TPA***
- How can we pull out site?
  - For example, if you plot a stand's QMDxTPA and if falls within the 75th percentile of all plots, is it nearing a maximum density for its suite of site-stand features,
  - or is it simply a point along its size-density stand development continuum?

Looking Back: IFC's Work on SDImax

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

ELSEVIER

Site sensitive maximum stand density index models for mixed conifer stands across the Inland Northwest, USA

doi.org/10.1016/j.foreco.2018.11.013

Mark J. Kimsey Jr.*, Terry M. Shaw, Mark D. Coleman

University of Idaho, Department of Forestry, Rangeland and Fire Sciences – Intermountain Forestry Cooperative, 875 Perimeter Dr MS1133, Moscow, ID 83844-1133, USA

1

Canadian Science Publishing

ARTICLE

A species-specific, site-sensitive maximum stand density index model for Pacific Northwest conifer forests[1]

Ryan R. Heiderman and Mark J. Kimsey, Jr.

https://doi.org/10.1139/cjfr-2020-0426

Received: 2 March 2023 | Accepted: 19 June 2023

https://doi.org/10.1111/nrm.12381

DOI: 10.1111/nrm.12381

Natural Resource Modeling WILEY

Pacific Northwest conifer forest stand carrying capacity under future climate scenarios

Ryan R. Heiderman | Mark J. Kimsey Jr.

Development Drivers:
Forest Carrying Capacity Calculator v2.0

- **Regional Discrepancies**:
  Feedback reveals: V1.0 overpredicting in South Central Oregon, East Slopes of Cascades, underpredicting in North Central Idaho.

- **Methodological Constraints**:
  Current V1.0 model relies on Stochastic Frontier Analysis (SFA), which poses limitations due to specific form assumptions and linearity in input-output relationships.

- **Species Coverage Gap:**
  V1.0 model: Few species, unclear mix impact on carrying capacity (is mix of shade tolerants, or drought tolerants?)

Why Machine Learning? A Path Forward

Limitations of Current Approach

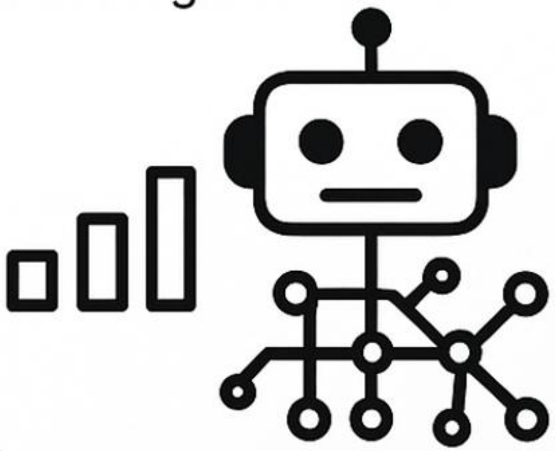
- SFA and traditional regression assume fixed model forms and relationships,
- Difficulty handling:
  - Non-linearity
  - High-dimensional interactions
  - Regional and species-specific variations

Enter Machine Learning (ML)

- **Definition: ML** models *learn* patterns from data without assuming specific forms.
- **Flexibility:** Captures complex, non-linear, and interactive effects
- **Adaptability:** Can be trained separately for different regions, species, or ecoregions

ML for Carrying Capacity Modeling

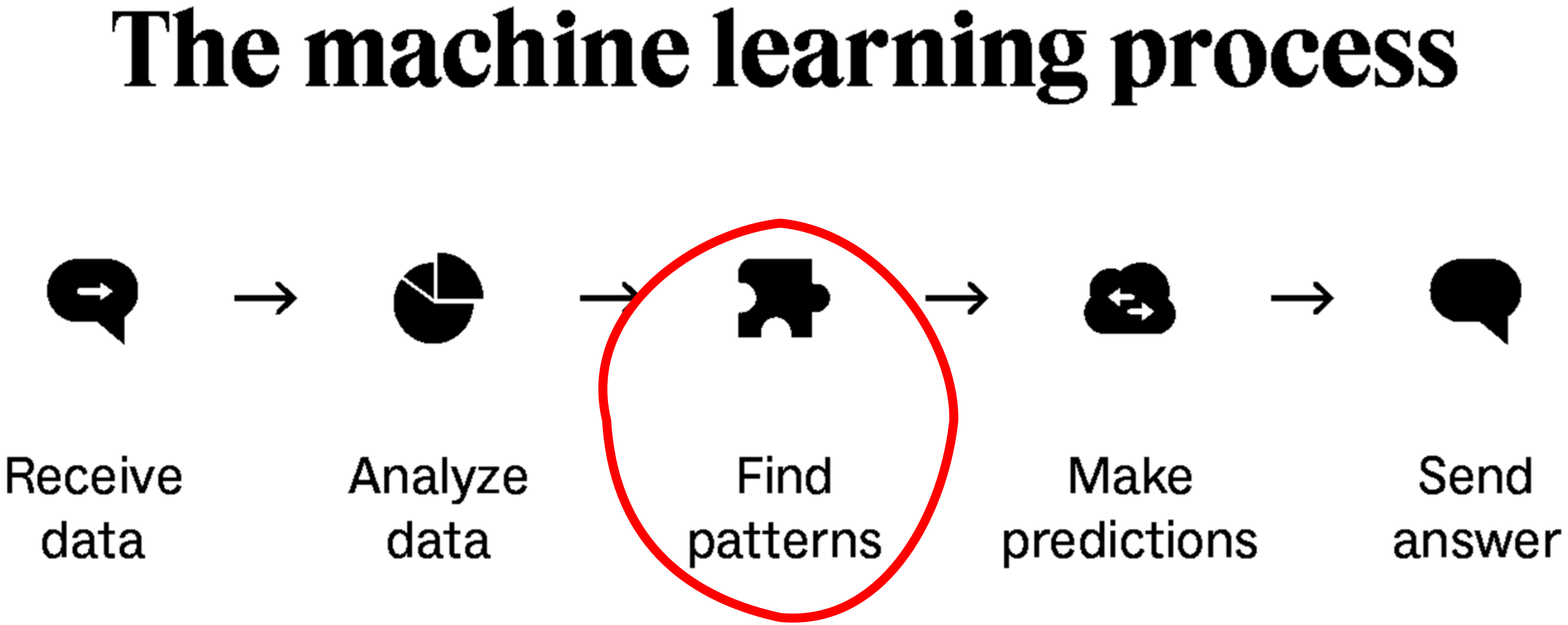
- Integrates species composition, site productivity, and climatic interactions
- Handles **imbalanced datasets** (e.g. more data from some regions/species)
- Enables **uncertainty quantification** and error **tracking**

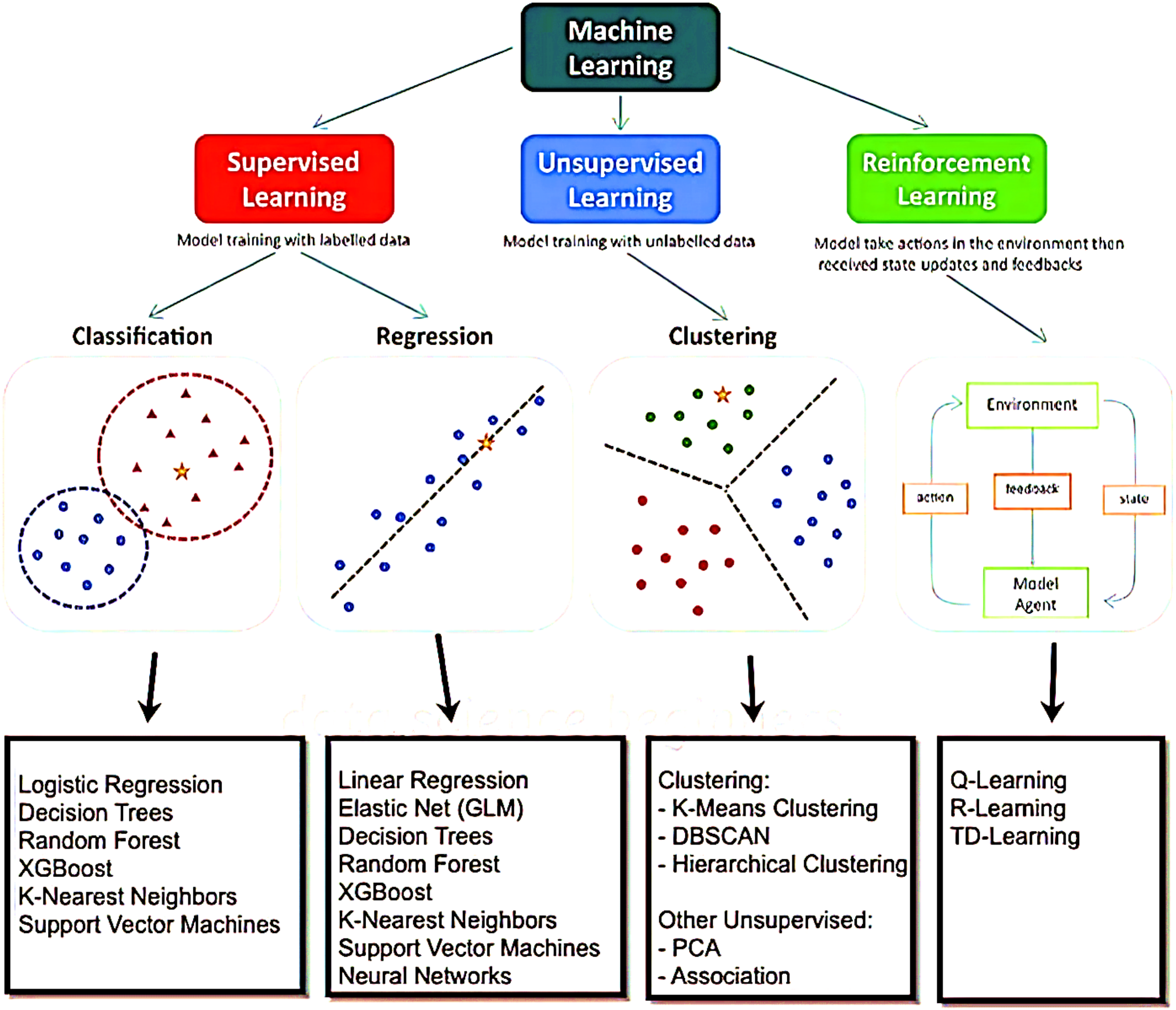


What is machine learning?

- Machine Learning is the art of turning data into decisions using algorithms and optimization
- ML is an automated hypothesis testing framework at scale — powered by data, driven by loss, refined by iterations.
- Machine Learning is how we teach machines to learn like humans — by experience, from examples, and through correction.
- Machine Learning Is Not Magic: It’s All About Math, Stats, Data, and Programming

What is machine learning?





Choosing the Best Supervised ML Model

Why Data Scientists Test Multiple Models

No One-Size-Fits-All

- No Free Lunch Theorem"

  Every algorithm performs differently on different data. Data-specific characteristics drive model performance

Bias-Variance Trade-Off

Model	Bias	Variance
Logistic Regression	High	Low
Random Forest	Medium	Medium
XGBoost/LightGBM	Low	High

Find the sweet spot by experimenting

Practical Strategy

- Cross-validation is your best friend
- Test on actual data, not just assumptions

AutoML for Speed

- Auto-sklearn
- H2O.ai
- Google AutoML

Then Optimize

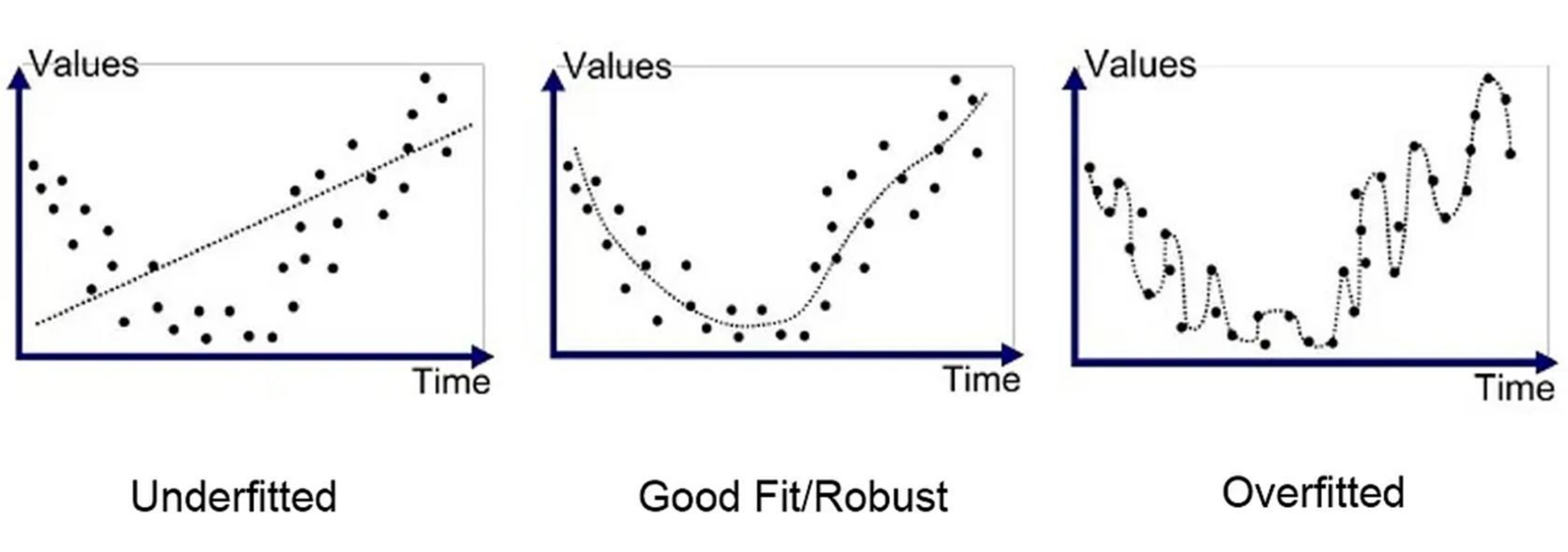
- Tune hyperparameters
- Compare performance metrics

Model choice should be data-driven, not guesswork. Experiment. Evaluate. Evolve.

Preventing Overfitting in Machine Learning

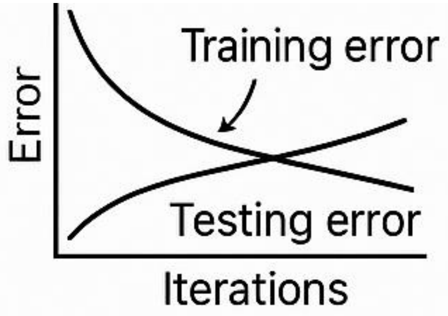
Understanding Overfitting

Bias-Variance Tredeoff



- If performance is much better on the training data than on testing data, overfitting is likely

- Train with More Data
- Cross-Validation
- Early Stopping
- Regularization
- Ensembling



Beyond Accuracy: The Case for Explainability

Great accuracy ≠ Great model

Model explainability (or interpretability) refers to our ability to understand and trust the predictions made by a machine learning model. It's about opening up the "black box."

Technique	Description	Best for
Feature Importance	Shows which features contributed most to a prediction	Tree models like Random Forest, XGBoost
SHAP (SHapley Additive exPlanations)	Provides a unified measure of feature impact across models	Any model (very popular)
LIME (Local Interpretable Model-Agnostic Explanations)	Explains individual predictions by approximating the model locally	Any model
Partial Dependence Plots (PDP)	Shows the relationship between features and predictions	Tabular data models

Inland Northwest Data

- 145,041 plots/stands info post pre-processing and EDA.

I.e., TPA, QMD, Species BA proportions

- Weighted Shade Tolerance & Drought Tolerance

(Niinemets and Valladares (2006))

- Species diversity indices

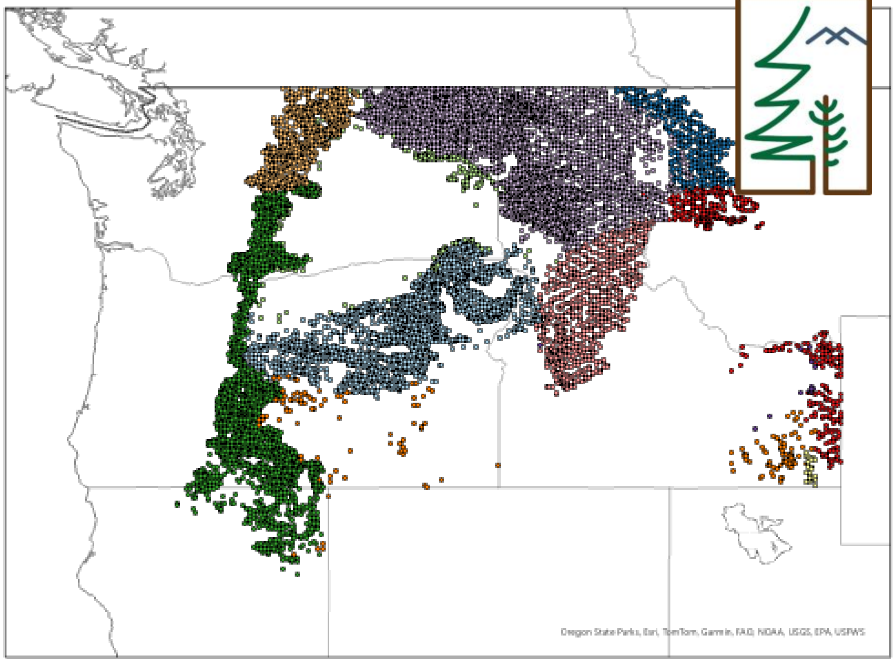
(e.g., Shannon Diversity Index, Species Richness, Pielou evenness, Berger Parker Index)

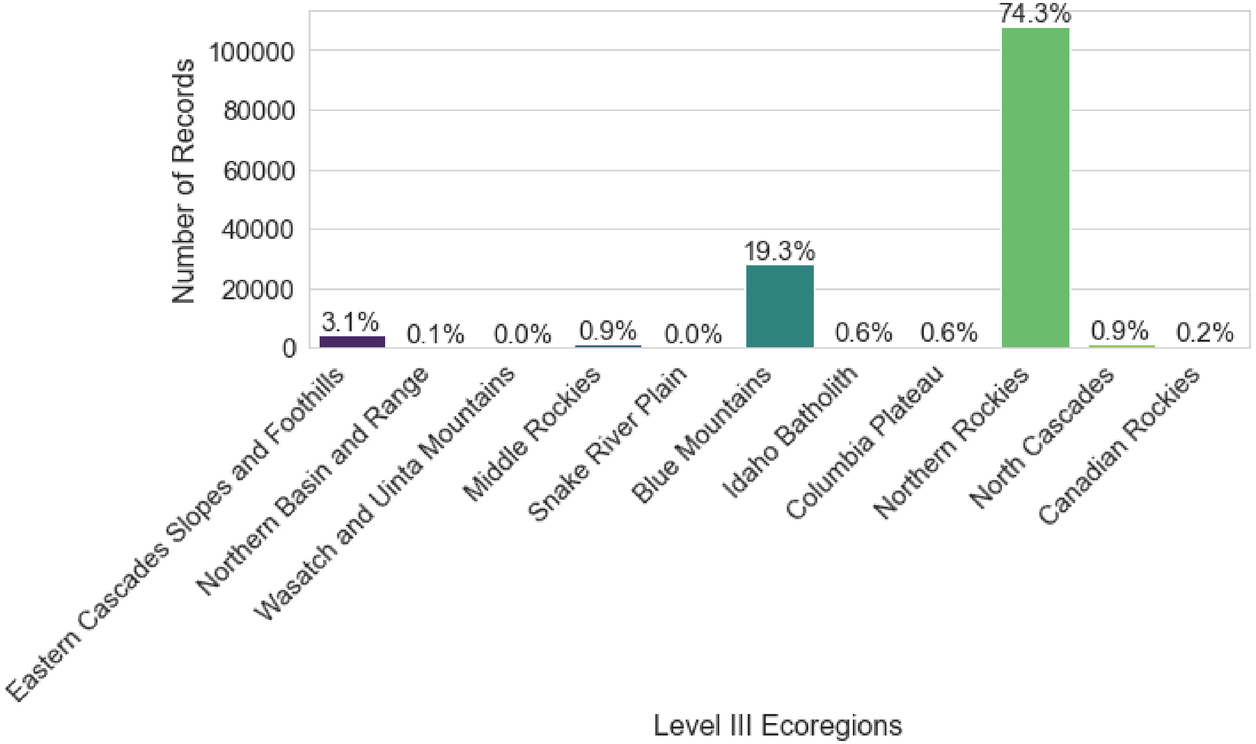
- Topography extraction from 30m DEM

(e.g., Slope, Aspect, Topographic wetness index, Solar radiation)

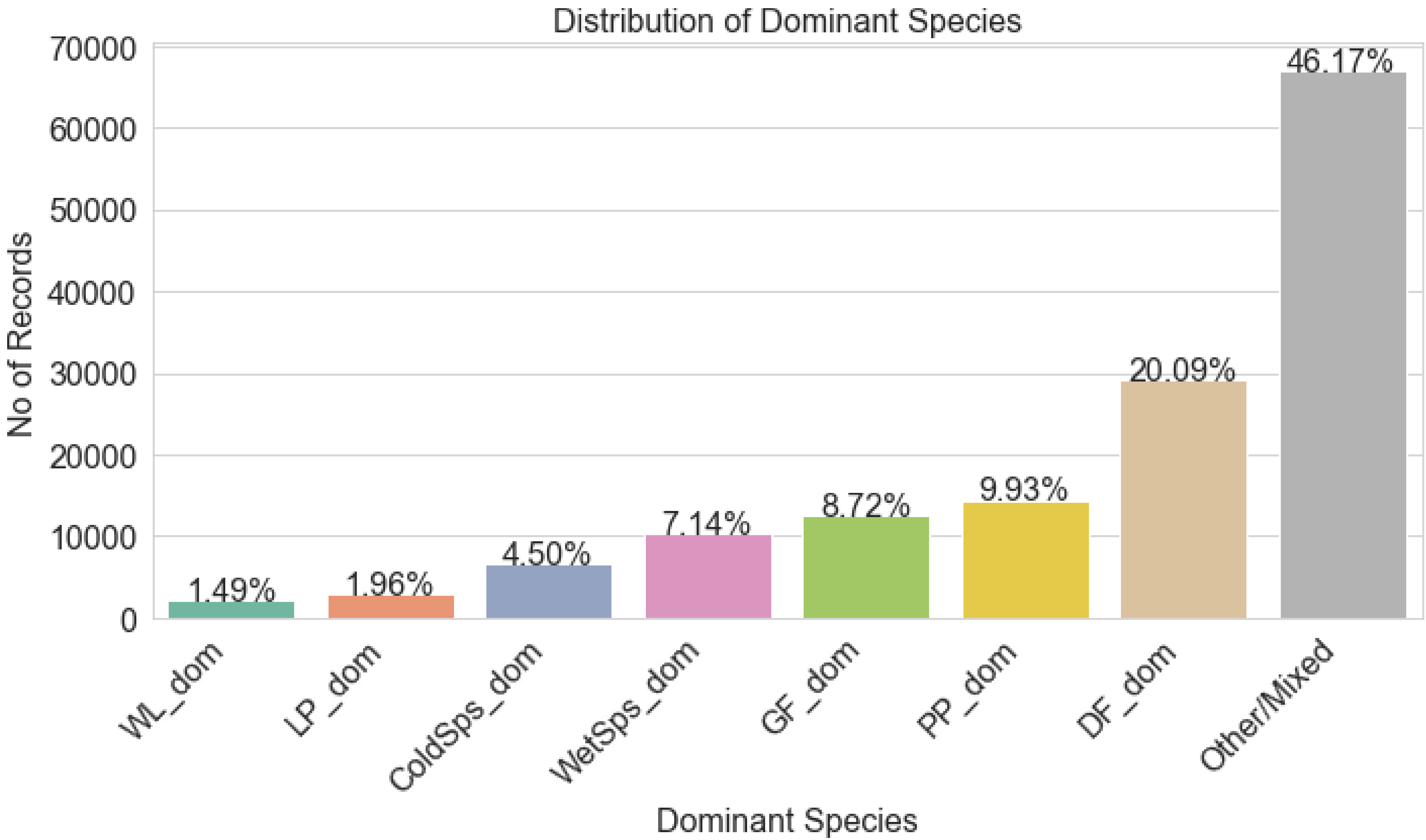
- ClimateNA – (Annual, Month, Season)

- Geology and Soil layer - (SGMC & gSSURGO geodatabase

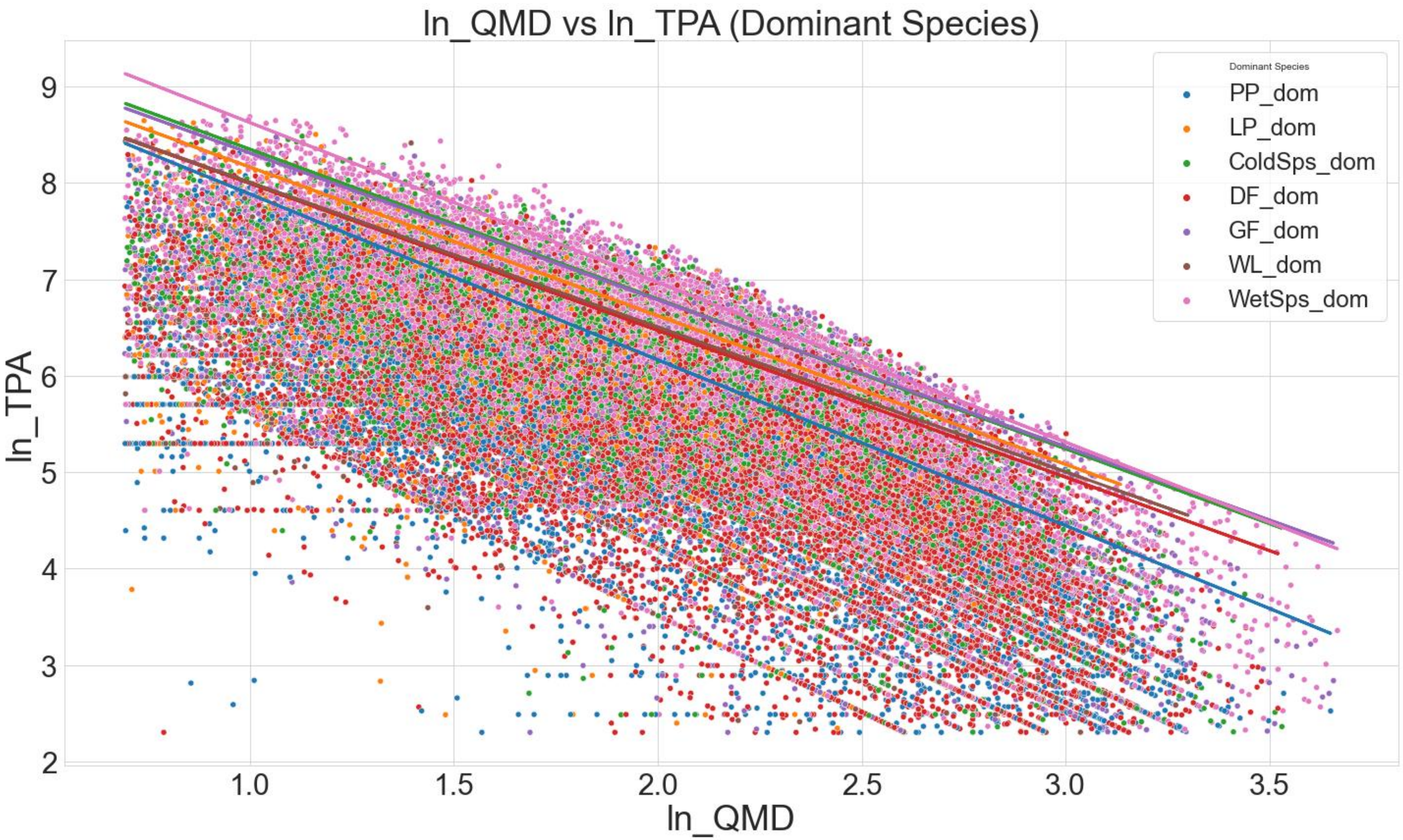




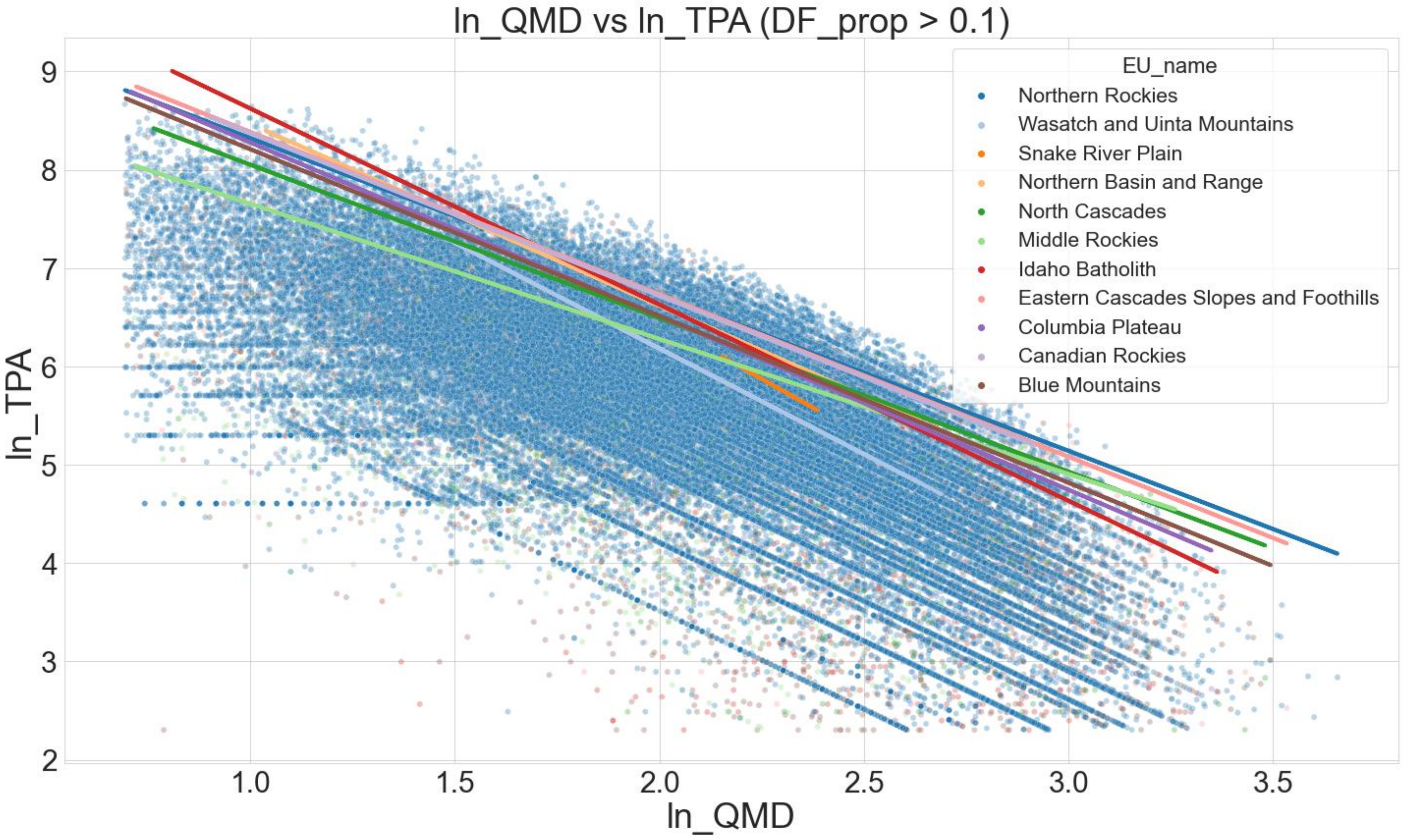
Insights into the Data



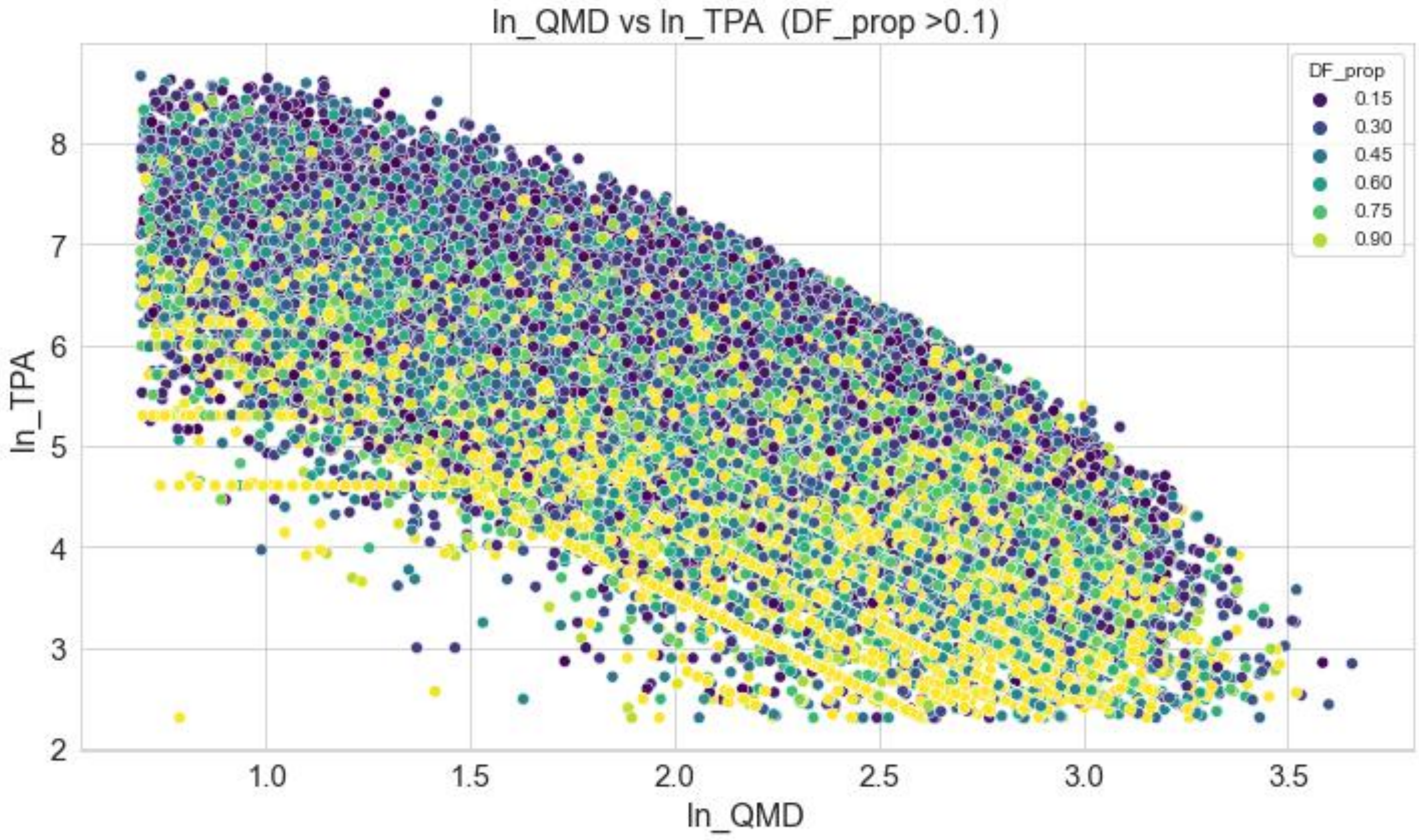
Insights into the Data



Insights into the Data



Insights into the Data



Modeling Approach

Data Cleaning:

Missing expansion factors, at least 10 TPA, QMD at least 2-inch, questionable & missing data

Feature Selection: Correlation + Mutual Information Score (MIS)

Step 1: Find Highly Correlated Pairs

- Compute Pearson Correlation between all feature pairs
- Select pairs with |correlation| > 0.7

Step 2: Calculate MIS

- Compute Mutual Information Score between each feature and the target
- MIS is non-parametric → supports all data type

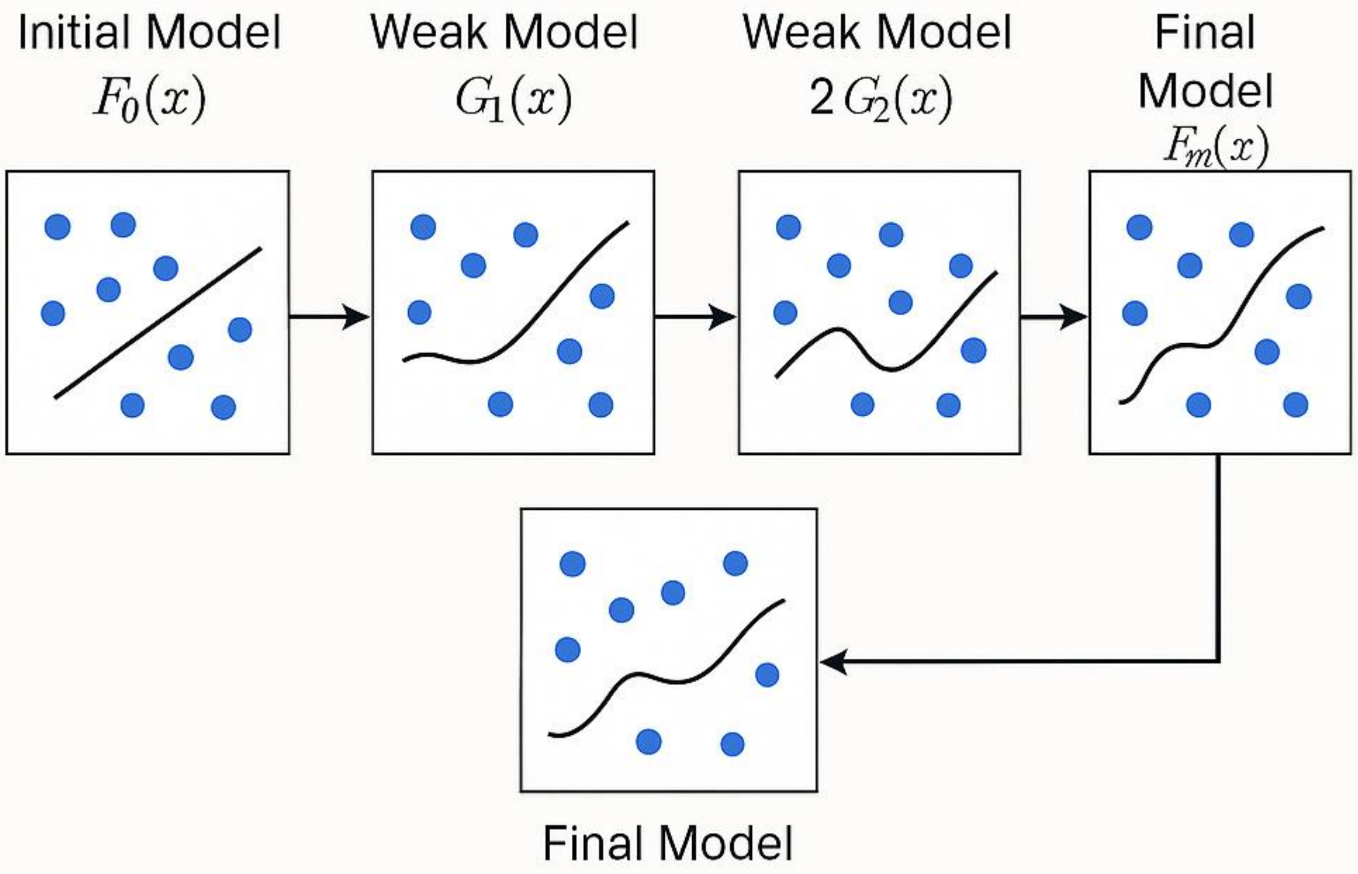
Step 3: Eliminate Redundant Features

- For each correlated pair:
  - Compare their MIS values
  - Remove the one with lower MIS
- Repeat until no correlated pairs remain

Final Result:

- High relevance to target (via MIS)
- Low redundancy between features (via correlation

Gradient Boosting



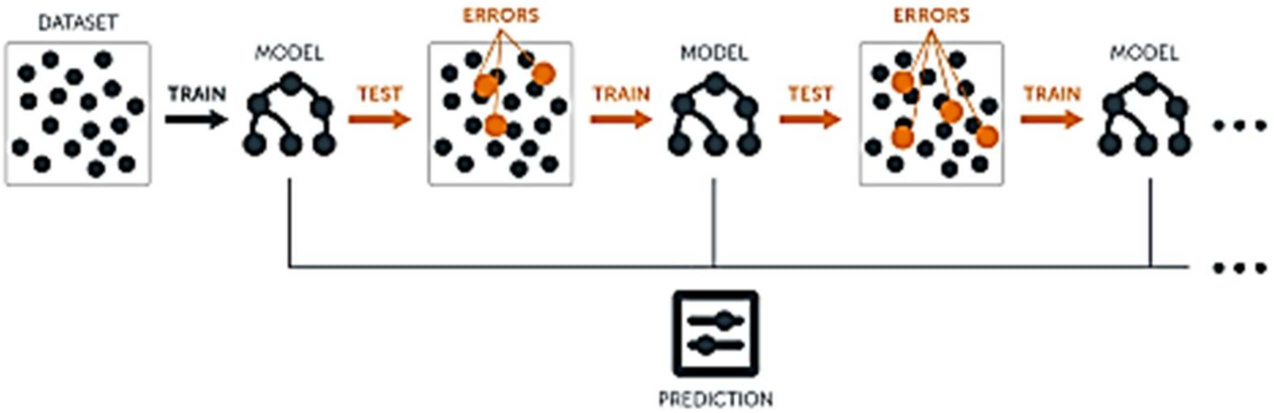
Final Model $= F_0(x) + \gamma_1 G_1(x) + \gamma_2 G_2(x) + \gamma_3 G_3(x)$

- ❑ GBM is an ensemble learning method that builds models sequentially, where each new model corrects the errors of the previous one.
- ❑ Workflow: Start with a base model (like a decision tree).
- ❑ Train new models on the residuals (errors) of the previous model.
- ❑ Combine models to minimize the loss function using gradient descent.

Addressing Challenges: The Analytical Approach

Quantile Gradient Boosting Machine

Quantile GBM predicts upper quantiles of size-density or carrying capacity distributions, targeting extreme values and high-density scenarios, providing insights into the self-thinning relationship.



Flexibility

Accommodating complex relationships and interactions in stand density data.

Accuracy

Capturing non-linearities and handling high-dimensional predictor spaces effectively

Robustness

Robust to model misspecification, ensuring reliable estimation of maximum stand density even in the presence of outliers and complex data patterns.

Model Summary:

Quantile regression loss function can predict a specified percentile

Quantile

$$f = \begin{cases} w \times QuantileAlpha \times (y - f) & \text{for } y > f \\ w \times (1 - QuantileAlpha) \times (f - y) & \text{for } y \leq f \end{cases}$$

Where:

- y is a true response
- f is a predicted response
- w is weight

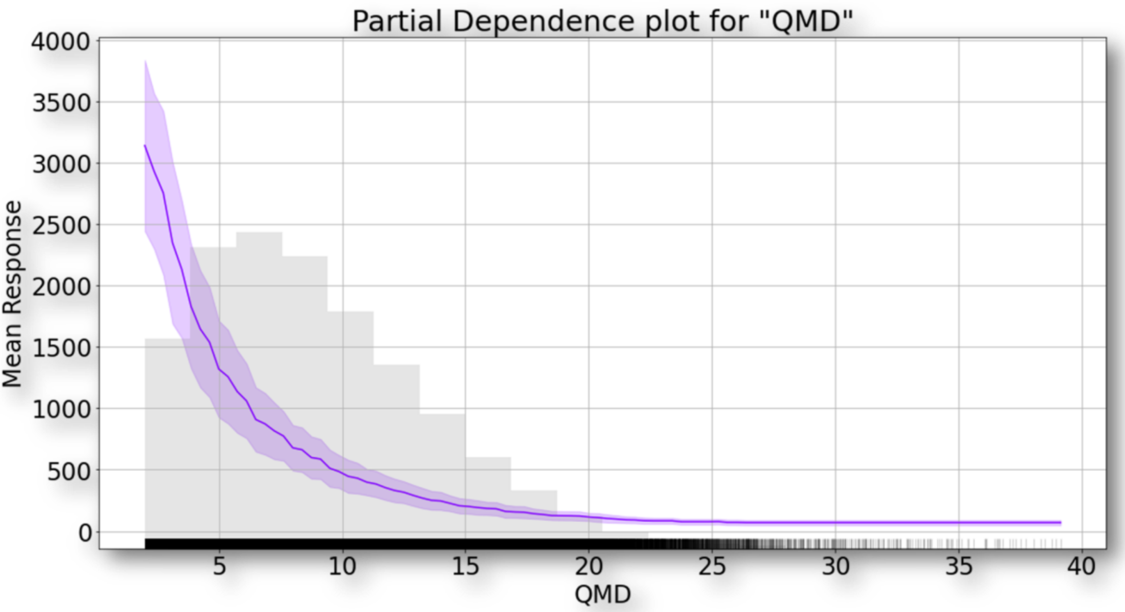
distribution = "quantile", quantile_alpha =.95

number_of_trees	number_of_internal_trees	model_size_in_bytes	min_depth	max_depth	mean_depth	min_leaves	max_leaves	mean_leaves
75	75	32206	5	5	5	19	32	29.493334

Consistent pinball loss (~32) in training and cross-validation signifies robust quantile estimation.

KEY PREDICTORS OF SDIMAX

QMD	RC_prop	TD	Cons_LITH
W_ShadeTol	WH_prop	PRATIO	soc0_20
W_DroughtTol	GF_prop	MAPMCMT	DEP2RES
	LP_prop	Rad_sm	aws0_100
	WL_prop	HeatLoad	Slope
	DF_prop		
	PP_prop		



TD - Temperature difference between mean warmest month temperature and mean coldest month temperature

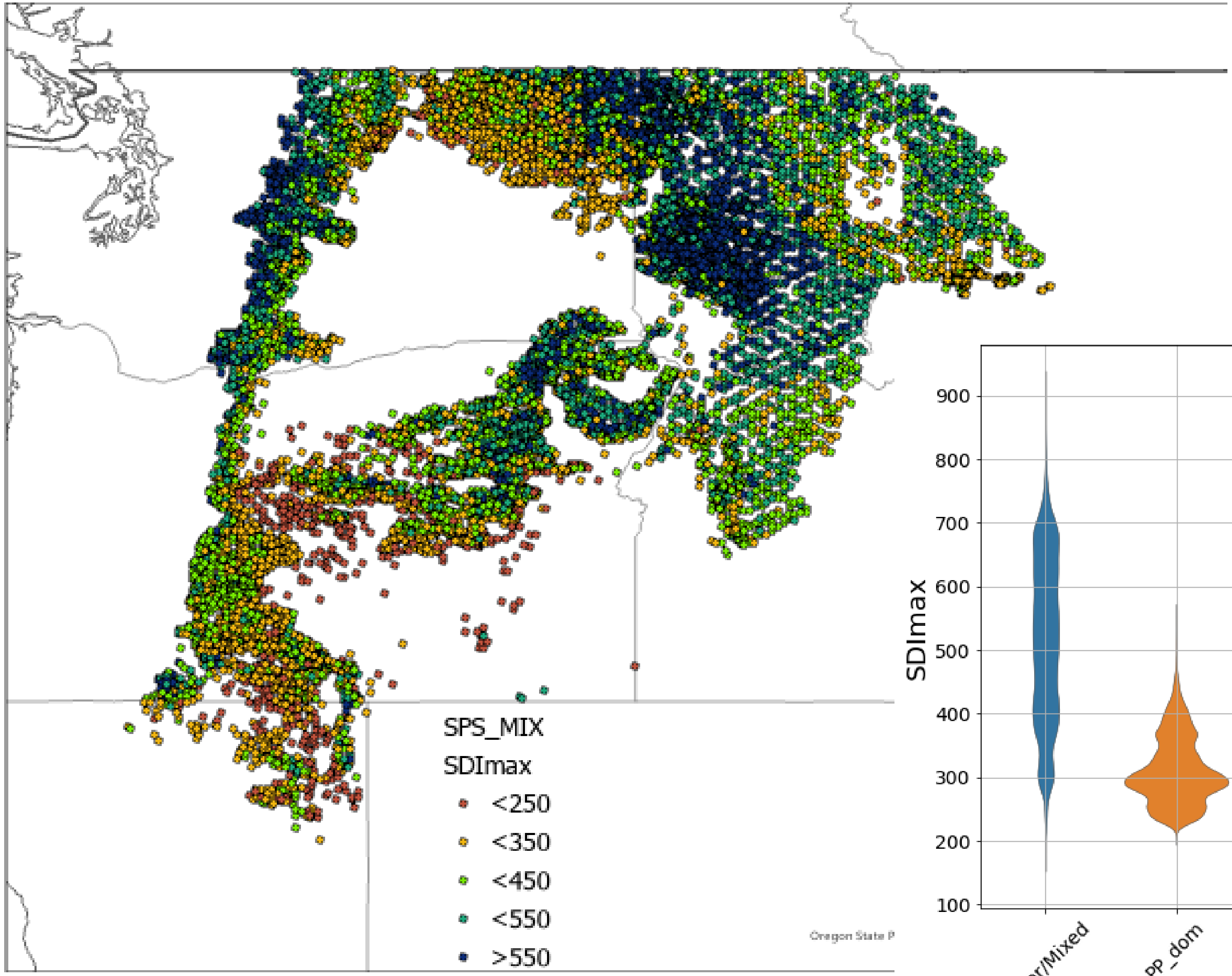
PRATIO - Ratio of growing season precipitation to total precipitation

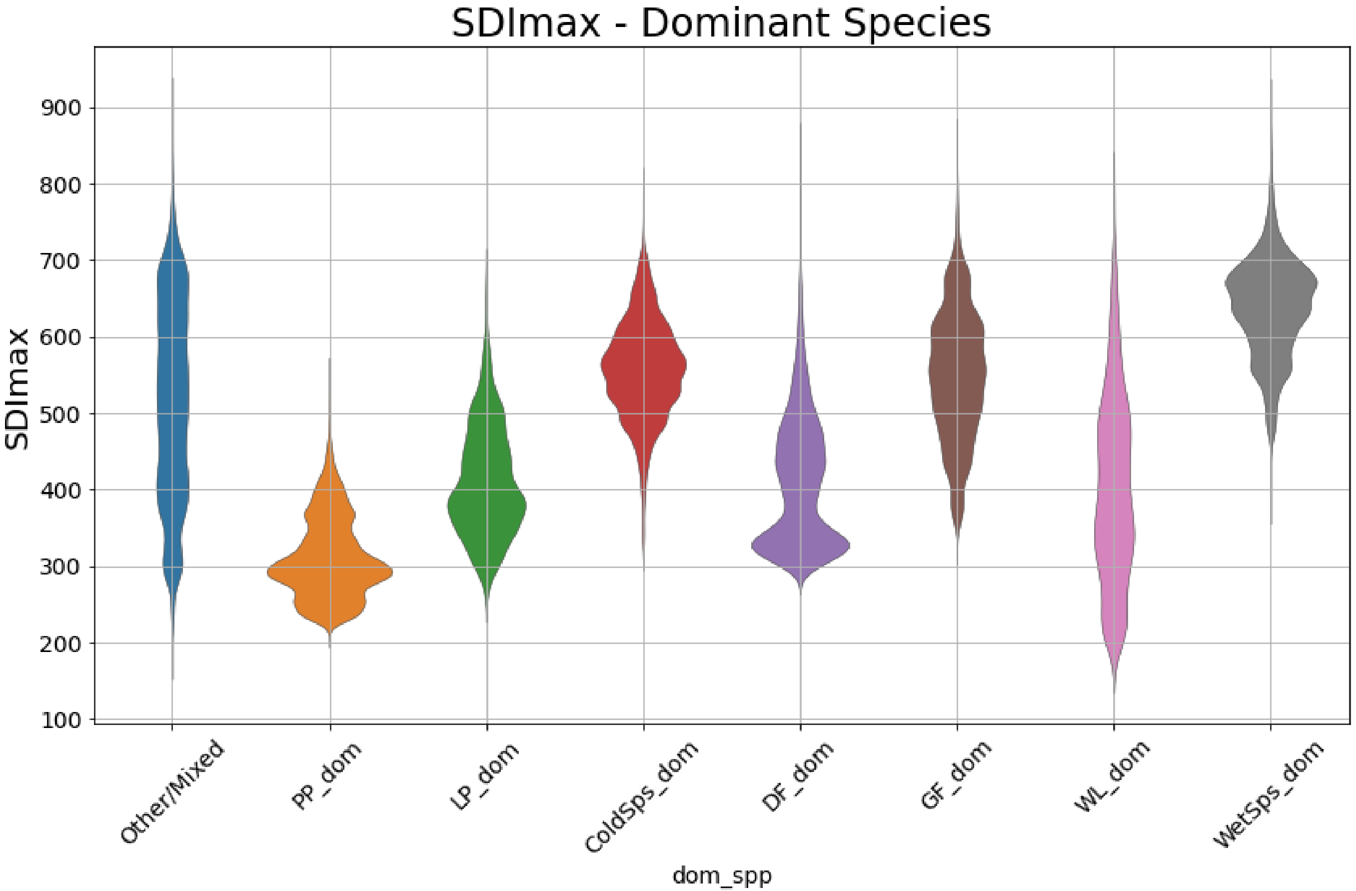
MAPMCMT - Interaction between mean annual precipitation and mean coldest month temperature

RAD_sm - Summer solar radiation

Heatload - Interaction between Annual Solar Radiation and Degree Days between 10 and 40 degrees Celsius

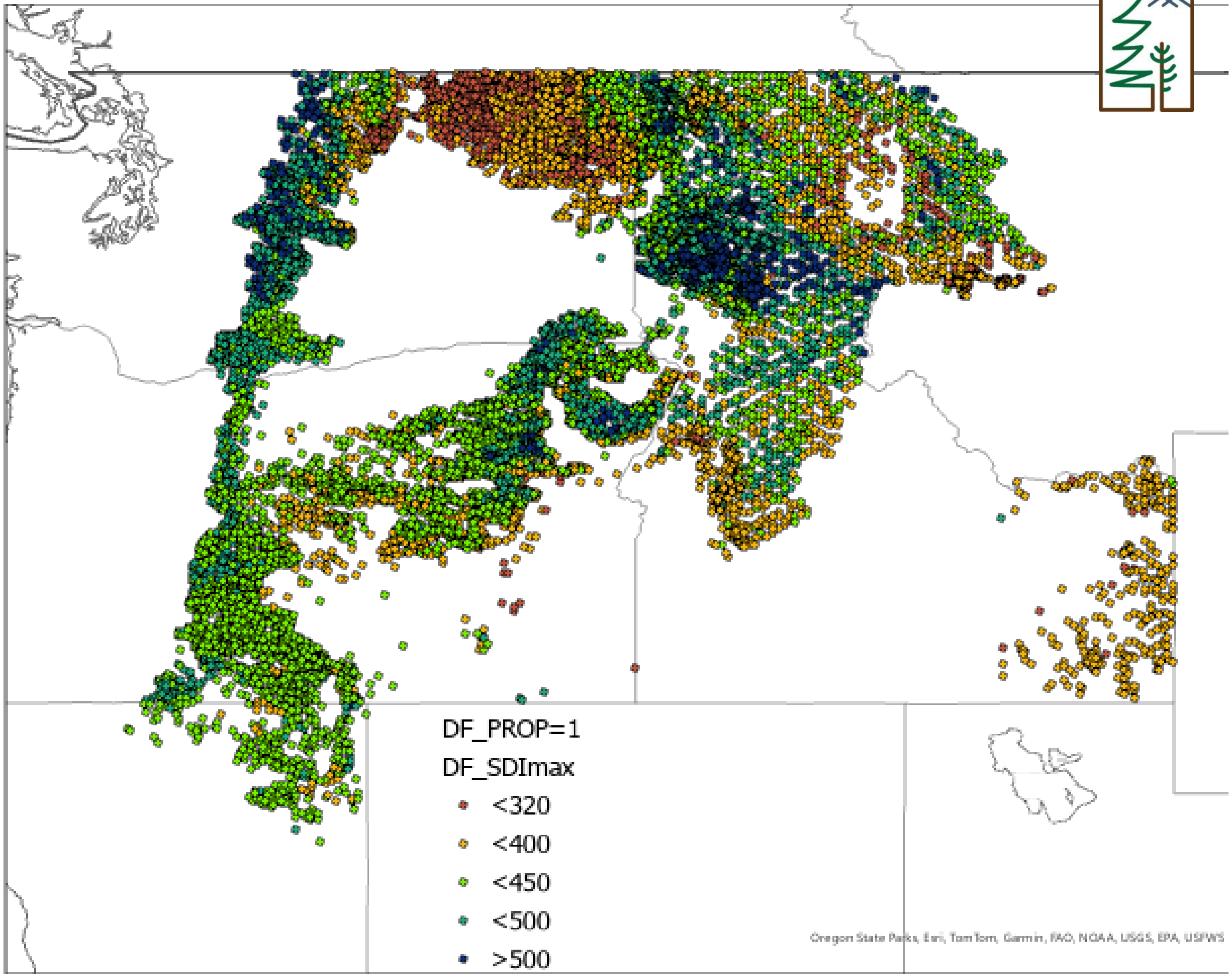
SDImax estimates with Species info



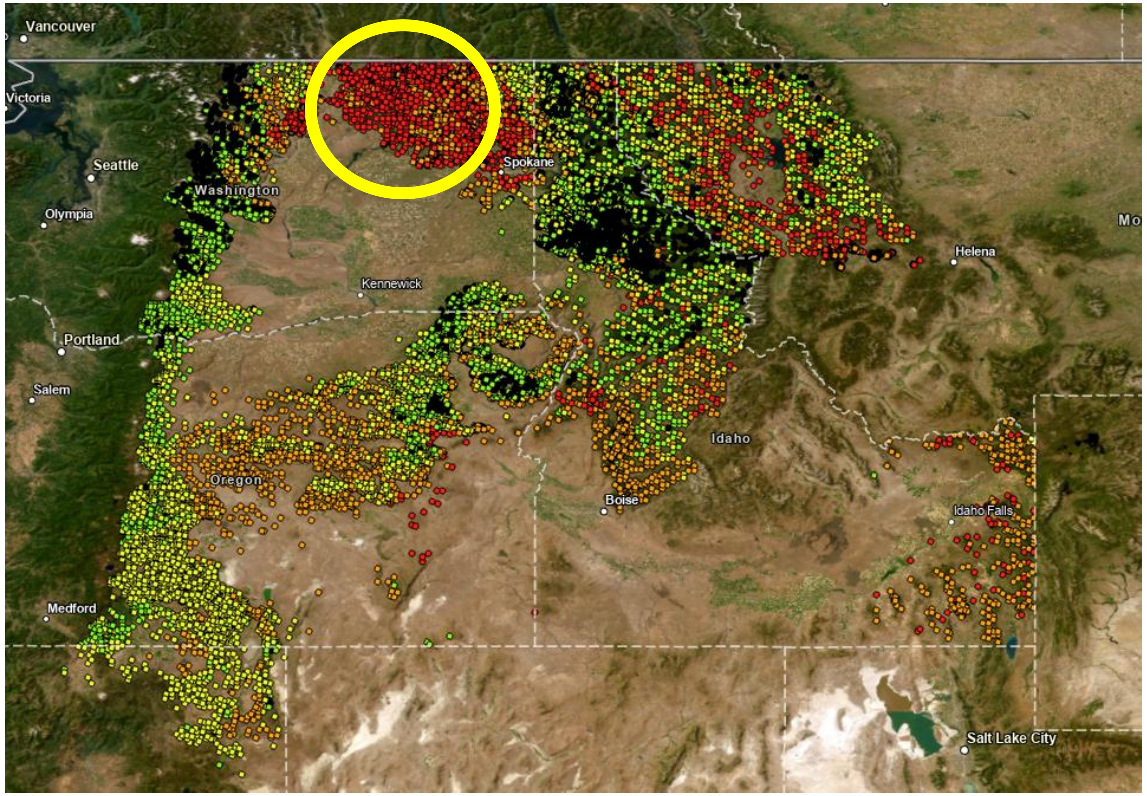


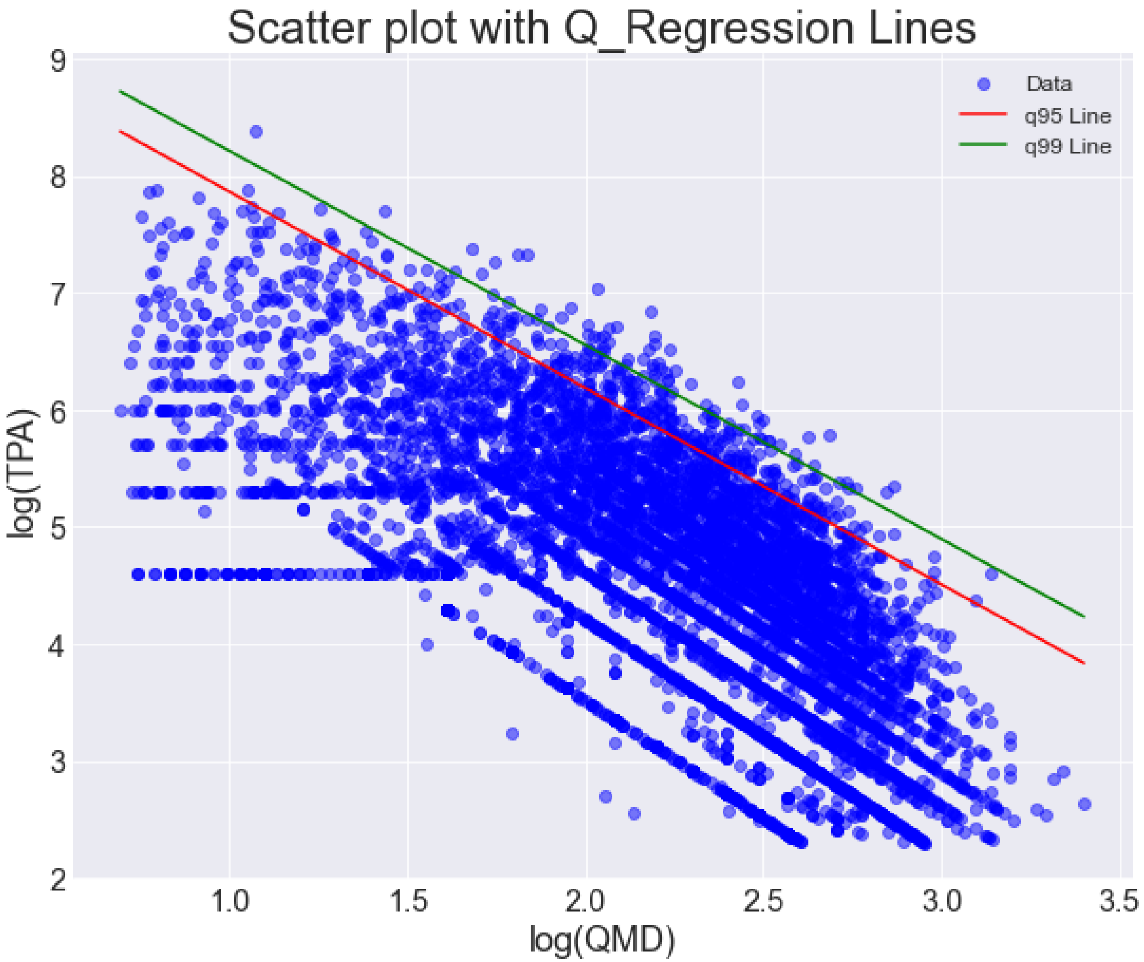
SDImax estimates for 100% Douglas-fir

min	290
25%	348
mean	423
50%	434
75%	476
95%	527
99%	548
max	668

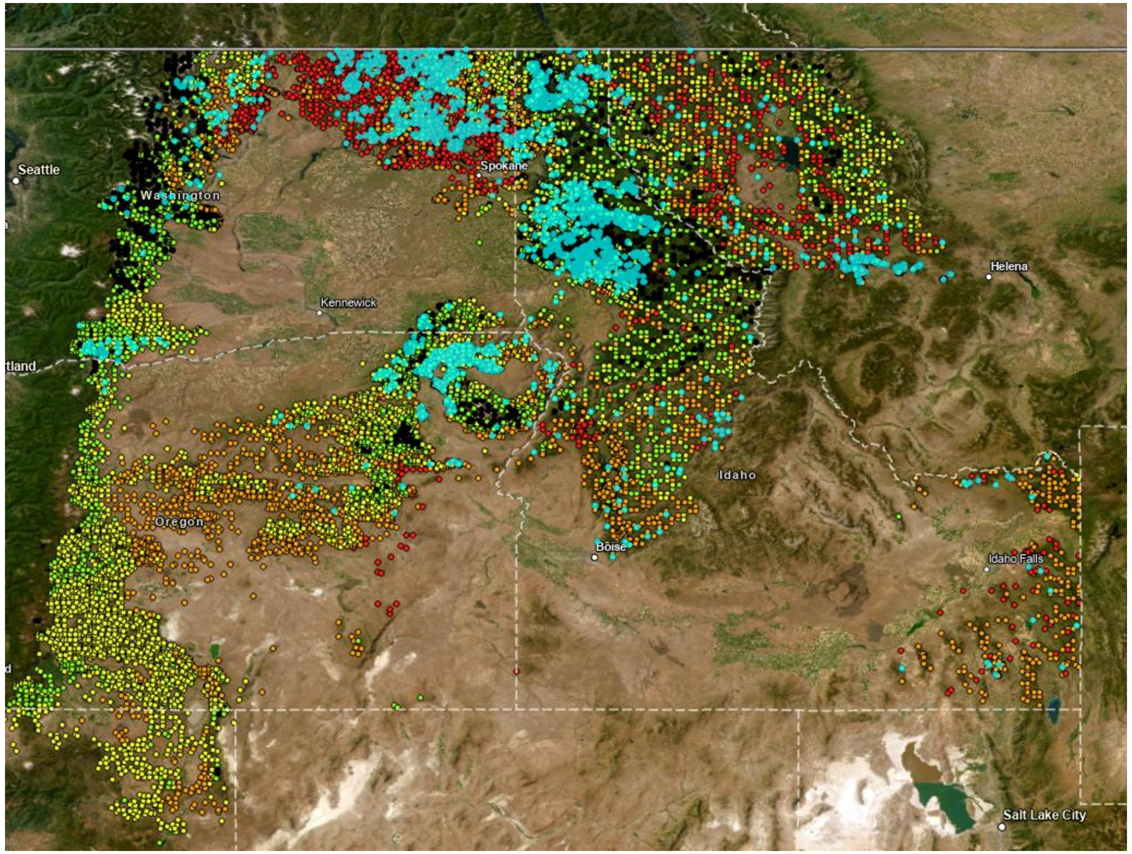


6747 pure stands of Douglas-fir (20% of total)

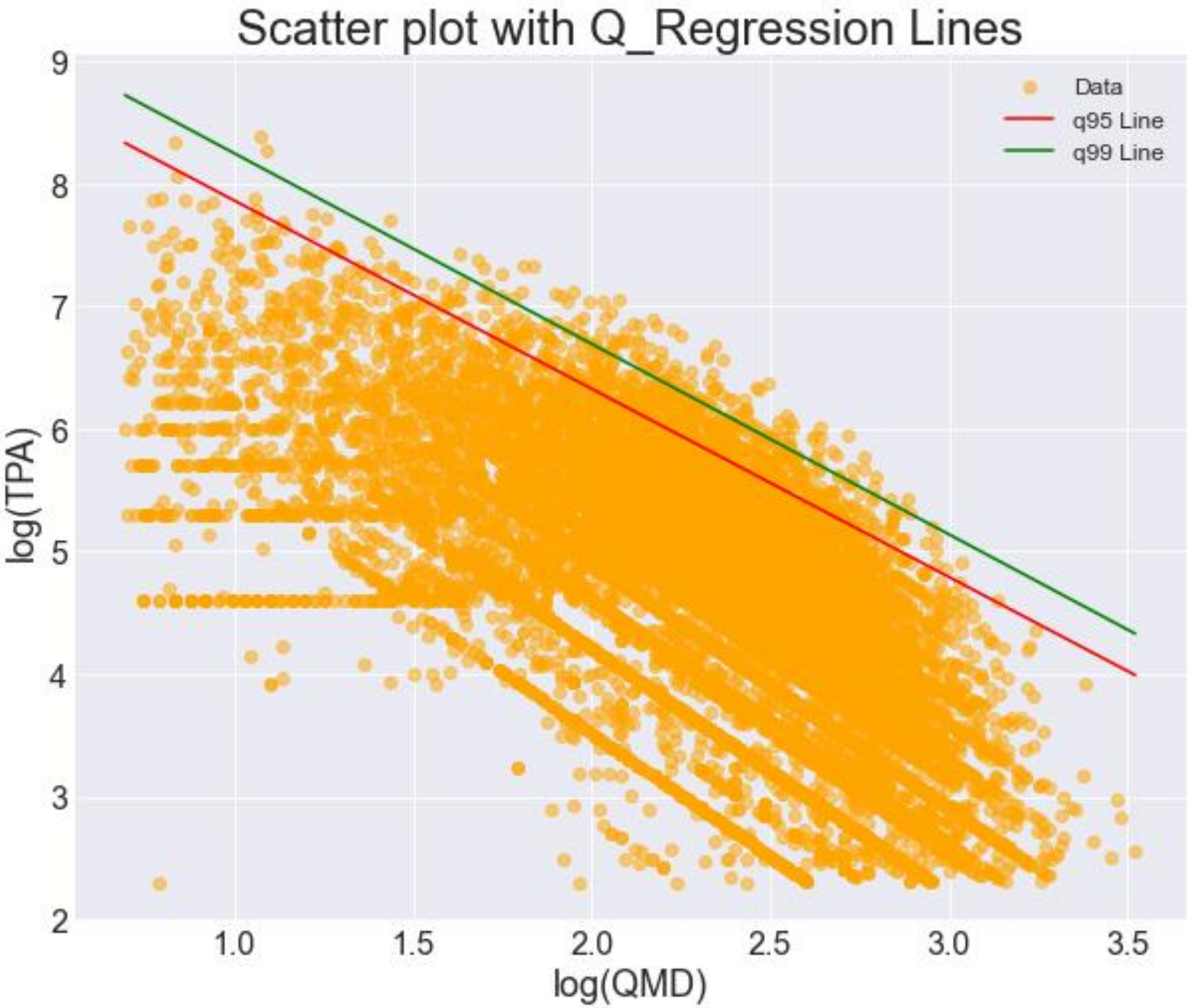




@QMD 10-inch Q95 of TPA = 292 and Q99 of TPA =405

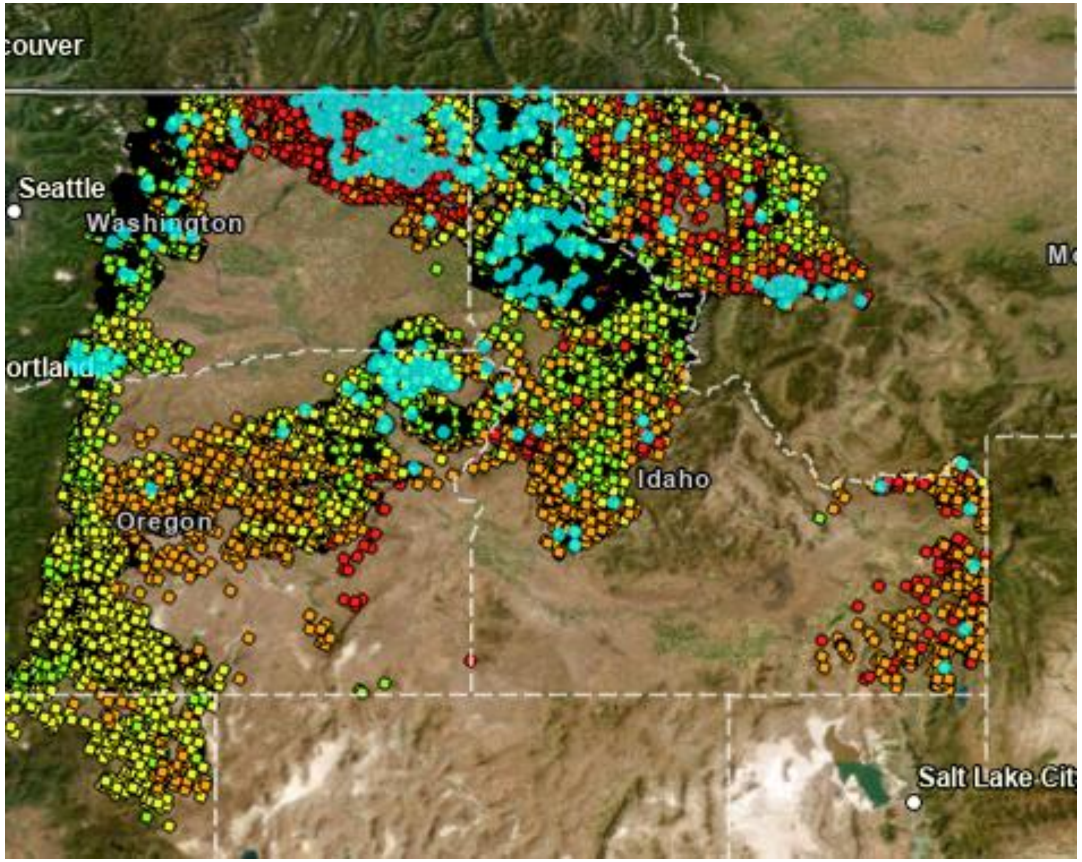


13622 pure stands of Douglas-fir (10% of total records)



@QMD 10-inch Q95 of TPA = 350 and Q99 of TPA =504

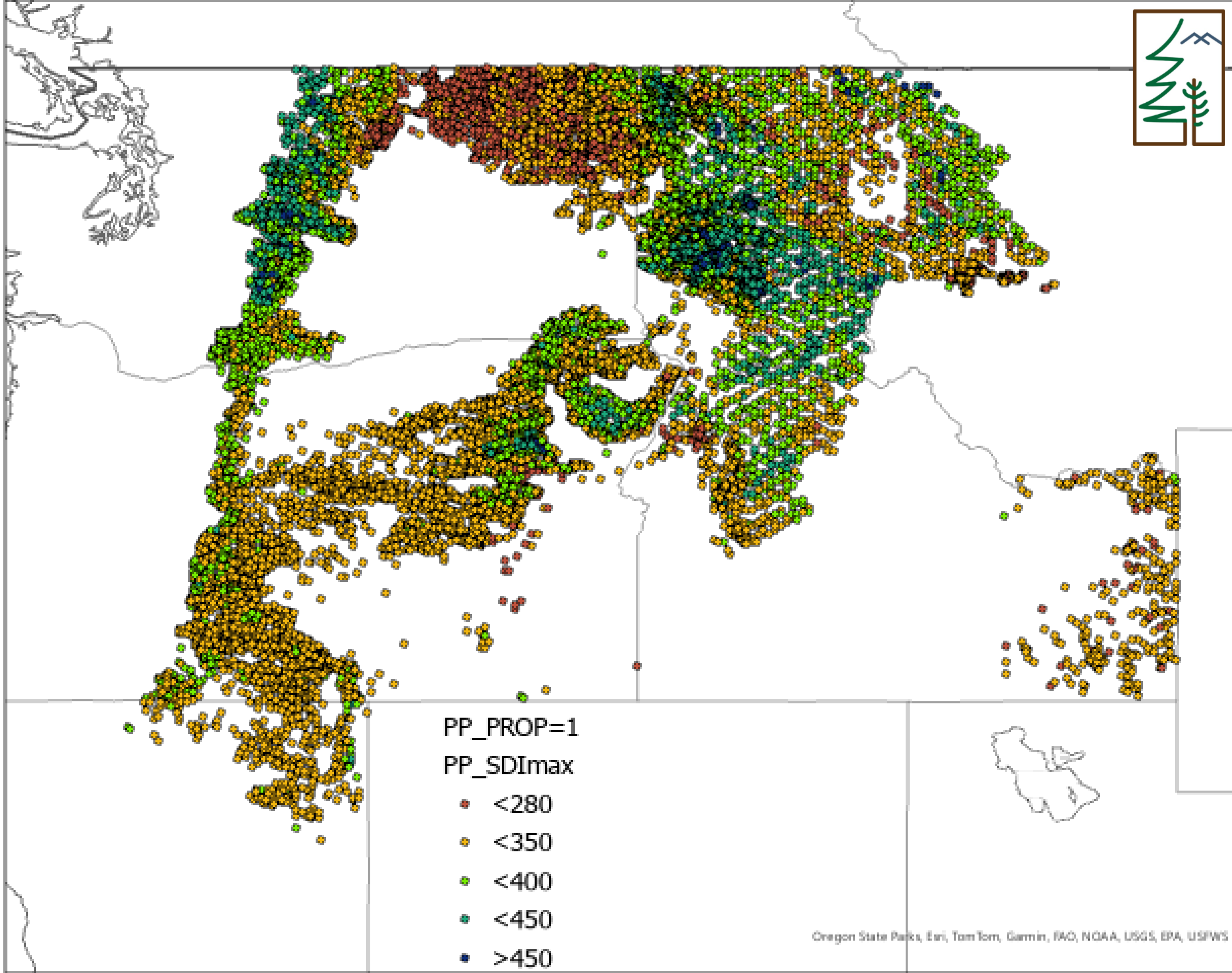
824 stands of Douglas-fir with 90%BA



@QMD 10-inch Q95 of TPA = 469 and Q99 of TPA = 670

SDImax estimates for 100% Ponderosa Pine

min	250
25%	300
mean	350
50%	343
75%	400
95%	439
99%	458
max	561

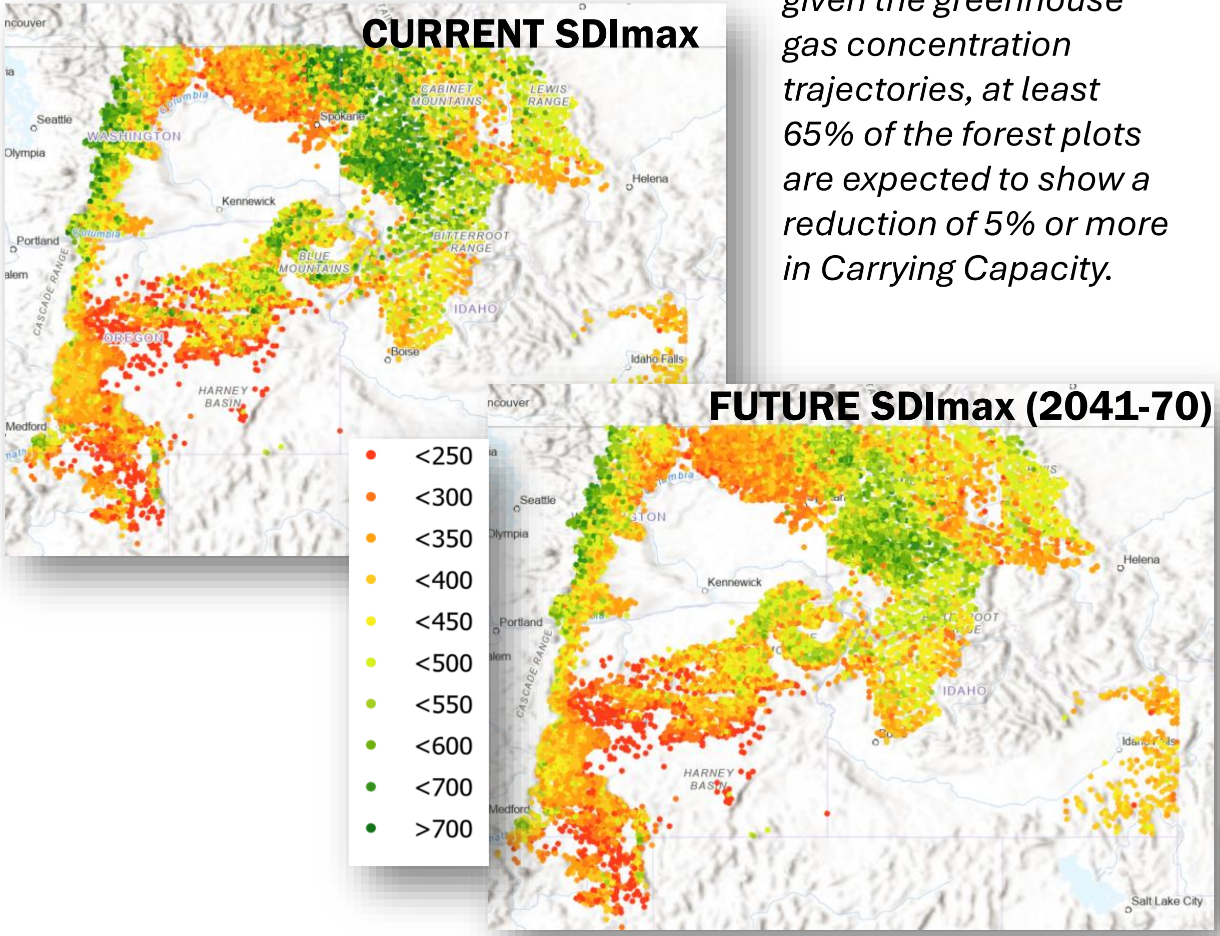


Model showed significant increases in maximum SDI when shifting from pure stands of Douglas-fir or Ponderosa pine to mixed stands (Other Species BA prop >10%)

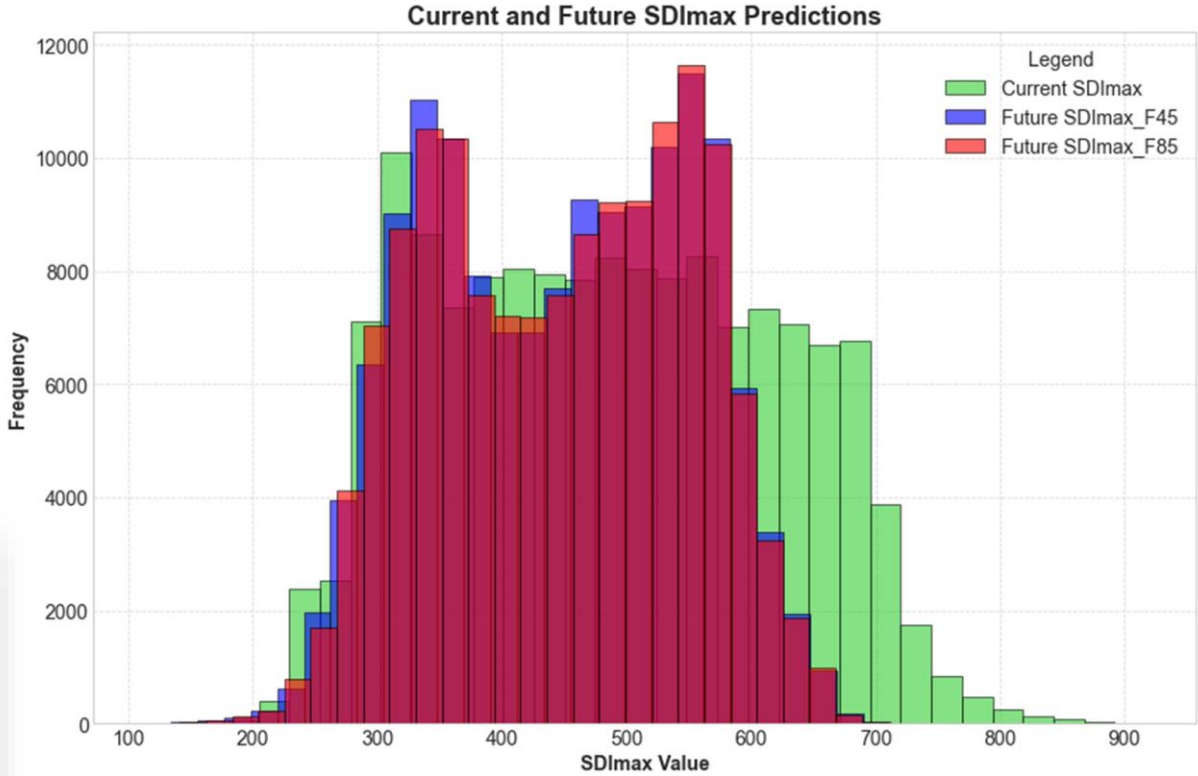
- This increase is due to a mixing of more shade tolerant species as shown throughout the literature
- Models reflective of habitat/forest type based SDImax thresholds
  - Powell's Green Book

Climate projections – impact on FCC

INLAND EVALUATION



This study revealed that, given the greenhouse gas concentration trajectories, at least 65% of the forest plots are expected to show a reduction of 5% or more in Carrying Capacity.



RCP4.5

WCRP CMIP5 World Climate Research Programme

SDI IN A ML ENVIRONMENT

SOLUTION NEEDED

Reineke based SDI Equation

$$SDI = TPA \times \left(\frac{QMD}{10}\right)^{\mathbf{1.605}}$$

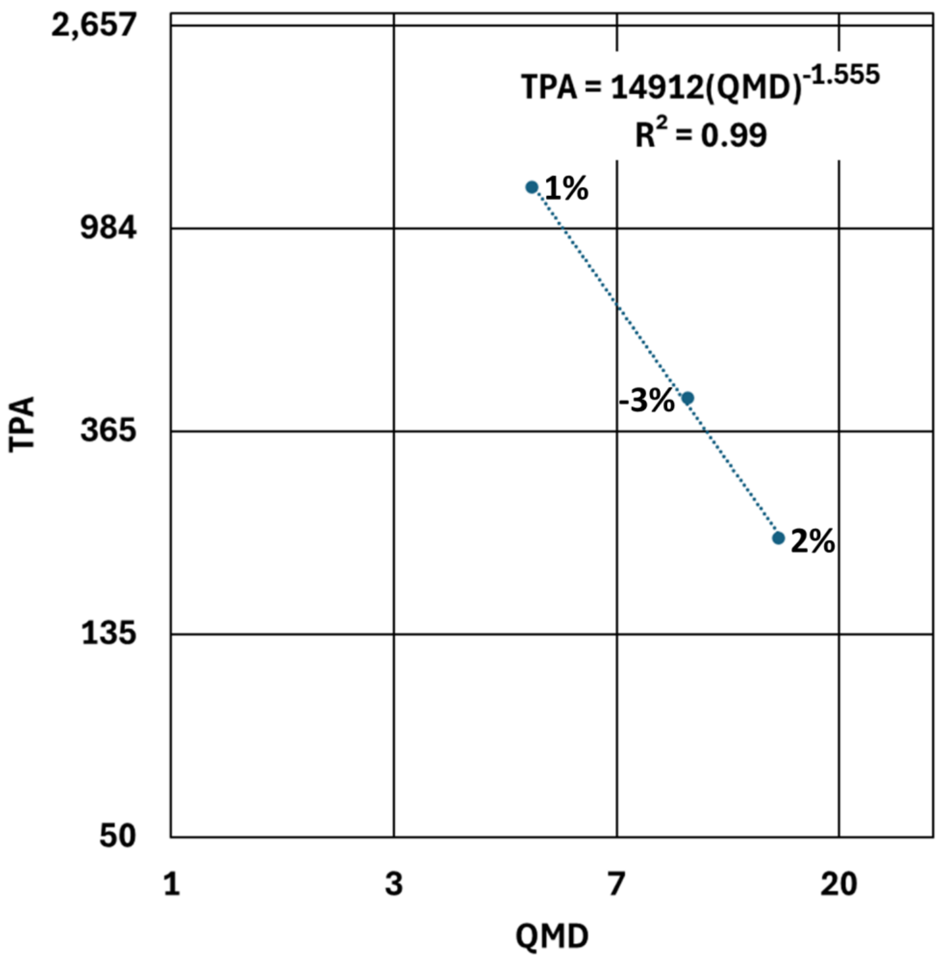
$$RD = {SDI}/{SDImax}$$

- 1.605 exponent a legacy of Reineke's MaxSDI modeling and subsequent species, linear based modeling efforts
- IFC V1.0 models created using Stochastic Frontier Regression (linear models) created a revised species model exponent that could be substituted for **1.605** to compute Relative Density (RD)
- New Machine Learning approach to SDImax modeling is not linear, exponent is unique to each user input – *f*(stand, site conditions)
- Exponent critical to compute Relative Density as a function of predicted IFC SDImax after inventory updates

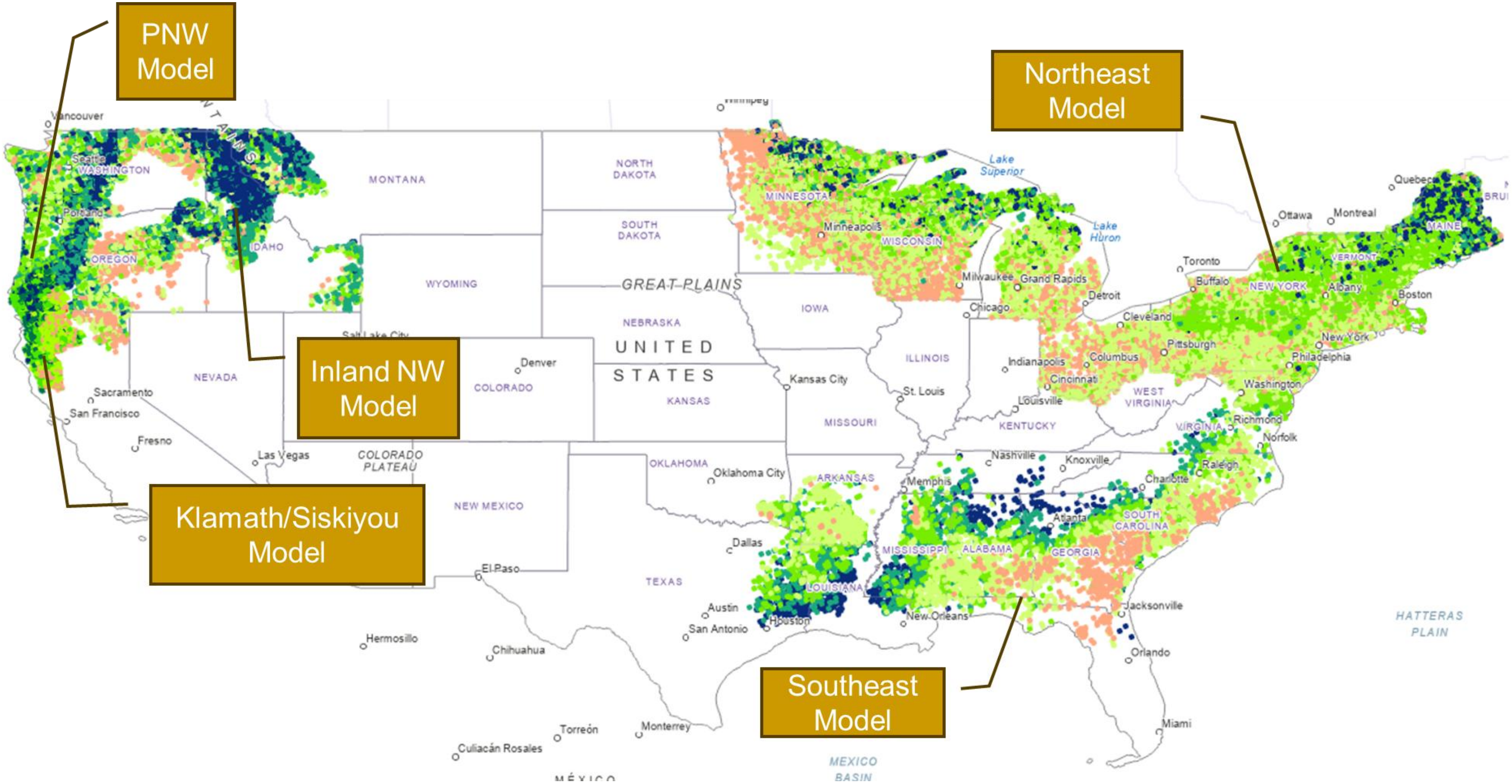
INTRODUCING IFC SDI FORMULATION

CUSTOM EXPONENT COMPUTATION

- One time computation of SDI max for a range of QMDs and User Inputs (initial stand metrics, location)
- Fit power functions to the X-Y pairs of predicted max TPA and associated QMD
- Power function exponent is the slope required to compute SDI
- Integrate slope coefficients into User database for future RD computation after inventory updates
  - No need to re-run SDImax model to update RD



SDI$_{MAX}$ MODELS AVAILABLE



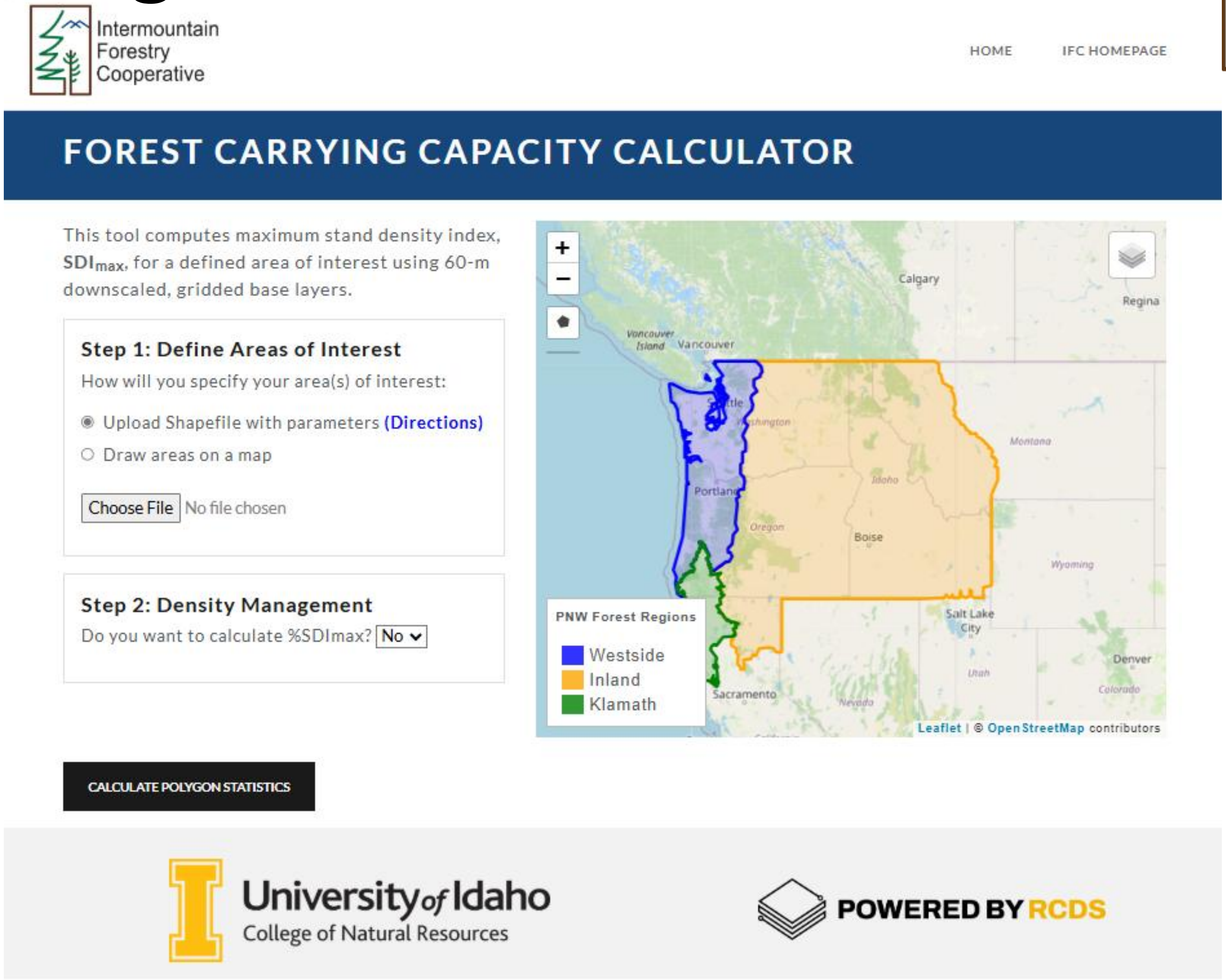
Toolsets for managers

IFC WEB APP

V2.0 INLAND,

V1.0 K-S,

V1.0 PACIFIC



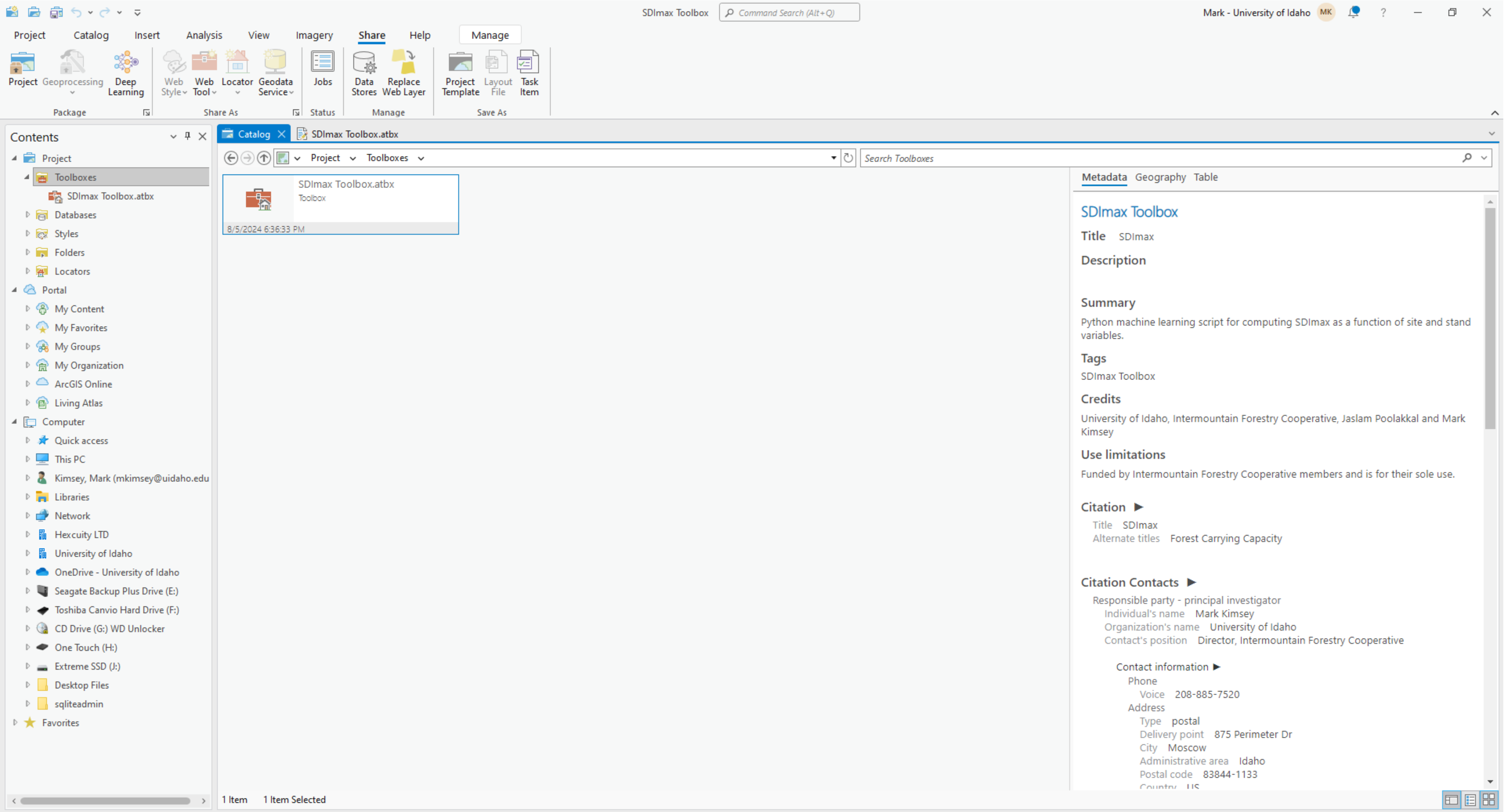
TOOLSETS FOR MANAGERS

JUPYTER NOTEBOOK

TOOLSETS FOR MANAGERS

Available Summer 2025

ARC TOOLBOX



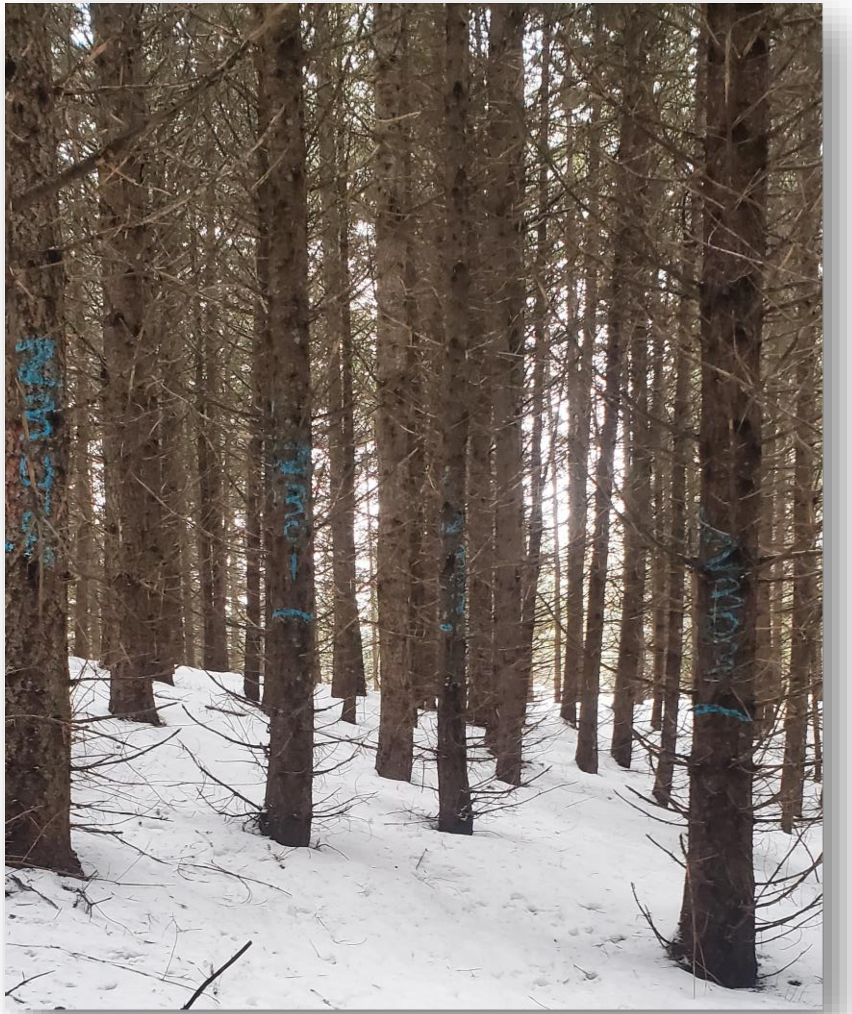
Take home points

CUSTOM SOLUTIONS FOR DENSITY MANAGEMENT

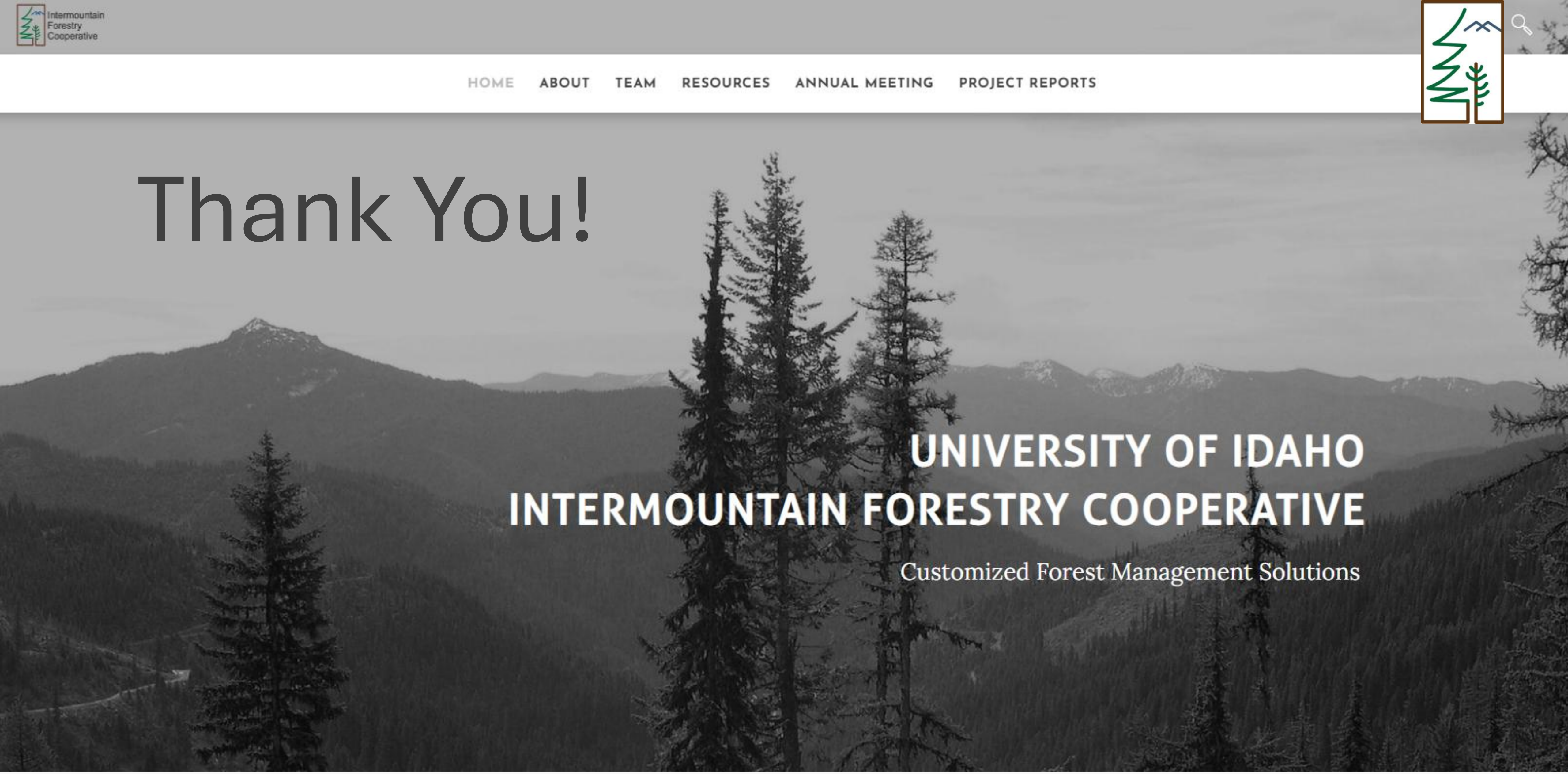
- SDImax predictions customized to your ownership, and stand inventory – leveraging the power of machine learning

- Climate projections for both the broader Pacific Northwest indicate a range of 5-20% reduction in carrying capacity on average (RCP 4.5 models) – summary report to be provided

- Customized SDI exponent solution incoming for computing Relative Density as a function of SDImax models



Thank You!



SPECIALIZING IN APPLIED SILVICULTURE RESEARCH

[IFC - Home (intermtnforestcoop.com)](intermtnforestcoop.com)