Imputation of Continuous Tree Suitability over the Continental United States from Sparse Measurements

Jitendra Kumar[α], **Forrest M. Hoffman**[α], **William W. Hargrove**[β], **Kevin Potter**[γ]

[α] Oak Ridge National Laboratory,
[β] USDA Forest Service,
[γ] North Carolina State University

98th Annual Ecological Society of America Meeting, Minneapolis, MN, USA

August 06, 2013

U.S. DEPARTMENT OF ENERGY

Climate Change Science Institute
AT OAK RIDGE NATIONAL LABORATORY

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE U.S. DEPARTMENT OF ENERGY

Why map trees?

- Mapping the range and suitability of tree species is important for the management of forest resources
- Species of economic and ecological importance
- Understand and assess the response of forests to climate change
- Conservation, restoration and diversity
- Observations avaiable are few and sparse
- Upscaling of point measurement is important and challenging problem

What do we know about tree ranges?

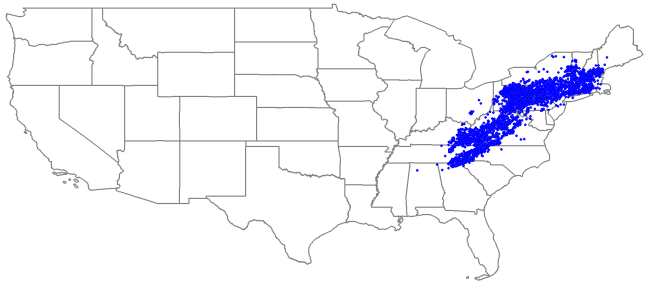
Traditionally we have used expert drawn maps of tree ranges and ecoregions.

What do we know about tree ranges?

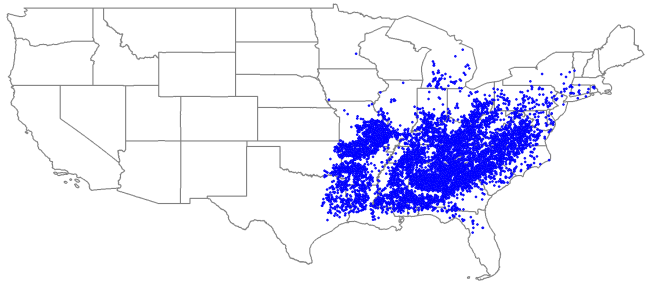
Traditionally we have used expert drawn maps of tree ranges and ecoregions. Tree species specific Little's range maps.

- Critchfield, W.B., and Little, E.L., Jr., 1966, Geographic distribution of the pines of the world: U.S. Department of Agriculture Miscellaneous Publication 991, p. 1-97.
- Little, E.L., Jr., 1971, Atlas of United States trees, volume 1, conifers and important hardwoods: U.S. Department of Agriculture Miscellaneous Publication 1146, 9 p., 200 maps.
- Little, E.L., Jr., 1976, Atlas of United States trees, volume 3, minor Western hardwoods: U.S. Department of Agriculture Miscellaneous Publication 1314, 13 p., 290 maps.
- Little, E.L., Jr., 1977, Atlas of United States trees, volume 4, minor Eastern hardwoods: U.S. Department of Agriculture Miscellaneous Publication 1342, 17 p., 230 maps.
- Little, E.L., Jr. 1978, Atlas of United States trees, volume 5, Florida: U.S. Department of Agriculture Miscellaneous Publication 1361, 262 maps.

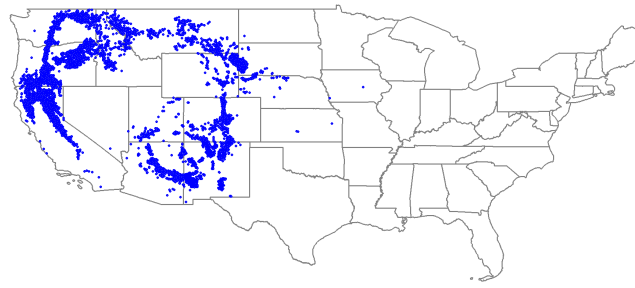
Little's range maps



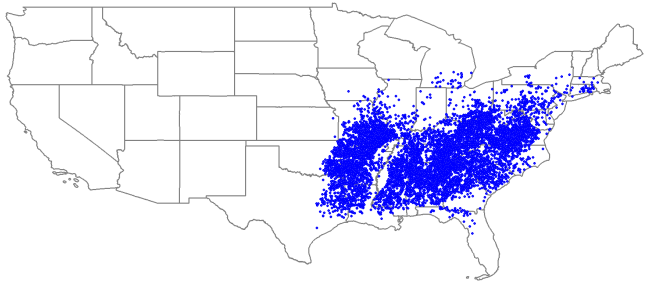
Betula lenta



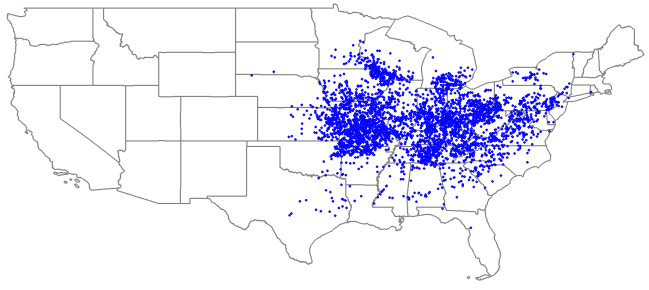
Cornus florida



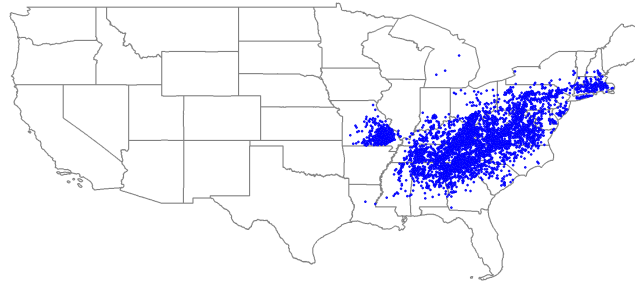
Pinus ponderosa



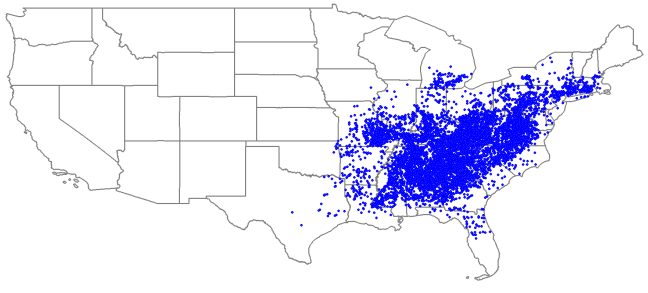
Carya alba



Juglans nigra

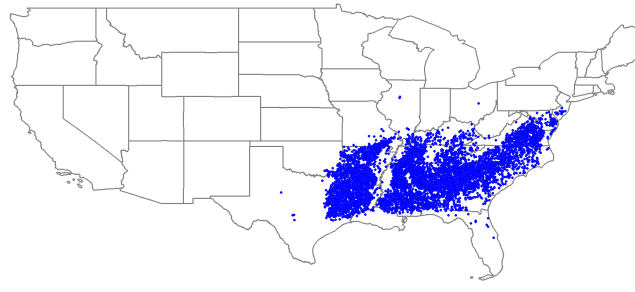


Quercus coccinea



Carya glabra

Pinus lambertiana



Quercus falcata

http://esp.cr.usgs.gov/data/little/

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Data and observations

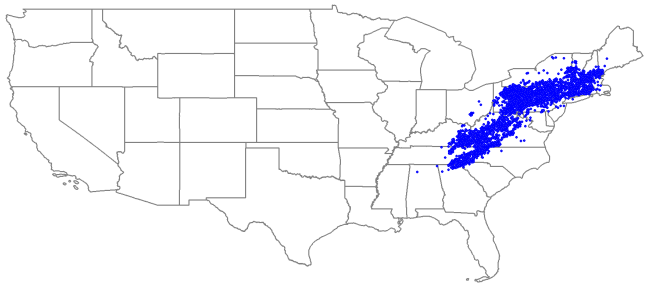
Forest Inventory and Analysis (FIA) program run by USDA Forest Service provides a very long term rich data for forest health and productivity.

Data and observations

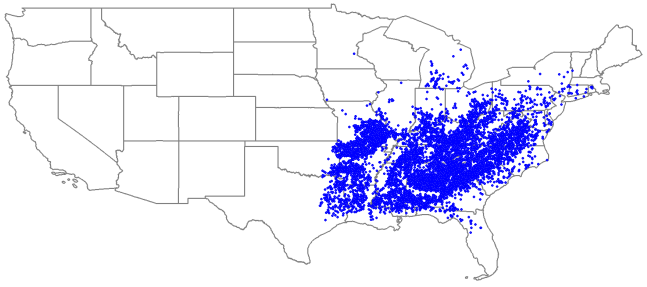
Forest Inventory and Analysis (FIA) program run by USDA Forest Service provides a very long term rich data for forest health and productivity. Observations from the Forest Inventory and Analysis (FIA) plots were employed in the study to include 325 species for CONUS. Some common species ...

- *Betula lenta*: Sweet Birch (2976)
- *Carya alba*: Mockernut Hickory (8158)
- *Carya glabra*: Pignut Hickory (7405)
- *Cornus florida*: Flowering Dogwood (7473)
- *Juglans nigra*: Black Walnut (3857)
- *Pinus lambertiana*: Sugar Pine (904)
- *Pinus ponderosa*: Ponderosa Pine (6099)
- *Quercus coccinea*: Scarlet Oak (4593)
- *Quercus falcata*: Southern Red Oak (6665)

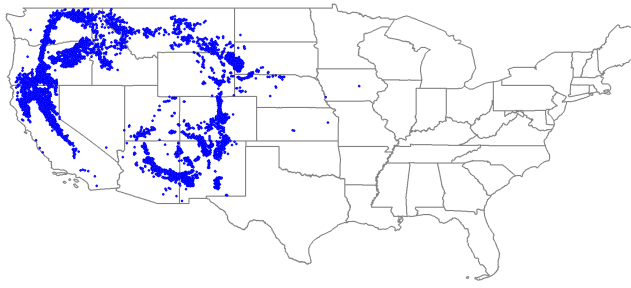
Forest Inventory and Analysis Data



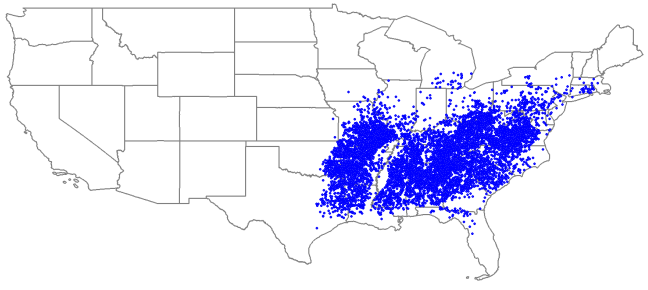
Betula lenta



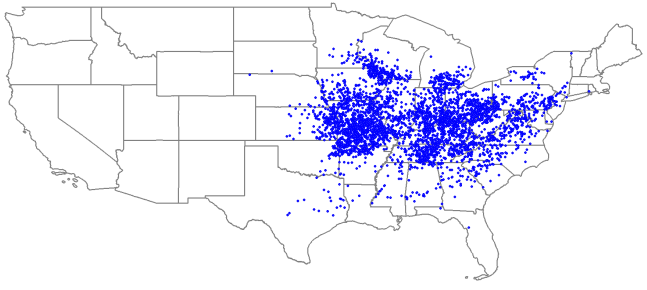
Cornus florida



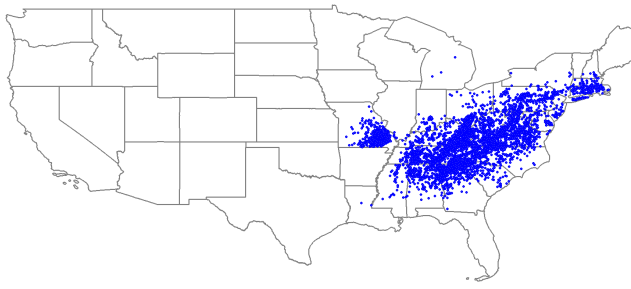
Pinus ponderosa



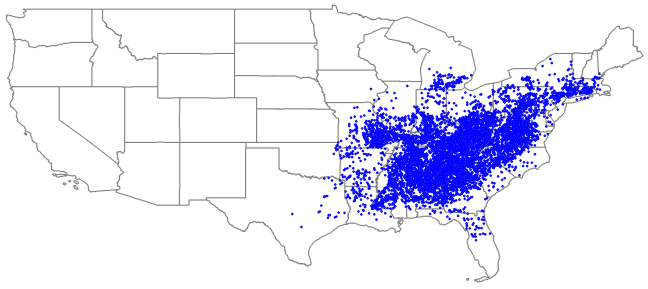
Carya alba



Juglans nigra

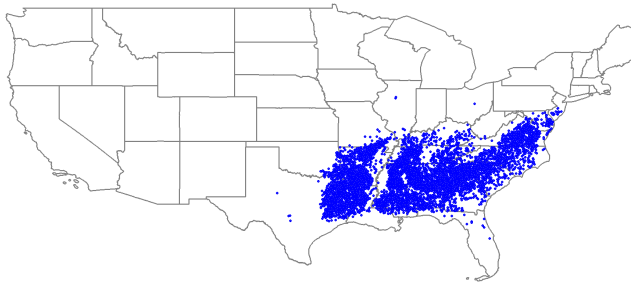


Quercus coccinea



Carya glabra

Pinus lambertiana



Quercus falcata

Objective

- How can we map tree species range for CONUS using sparse observations?
- Can we quantify the productivity of these species?
- Impact of climate change on tree ranges?
- Can we do this in automated, consistent and objective way?

Two key questions...

1. Species range: Where do they grow?
2. Species productivity: How well do they grow?

It's important to get both of these right.

Growing conditions as the surrogates

- Climate, topographic and edaphic factors determines the suitability of a tree species in a location
- These factors are surrogates/indicators of growing conditions and productivity

Data sets used in the study

Models: PCM and Hadley GCMs
Scenarios: Present conditions (WorldClim), A1FI, B1
Resolution: 4 km^2
Variables: 17

1. Precipitation during the hottest quarter
2. Precipitation during the coldest quarter
3. Precipitation during the driest quarter
4. Precipitation during the wettest quarter
5. Ratio of precipitation to potential evapotranspiration
6. Temperature during the coldest quarter
7. Temperature during the hottest quarter
8. Sum of monthly Tavg where Tavg >=5 deg C
9. Integer number of consecutive months where Tavg >= 5 deg C (Length of potential growing season)
10. Available water holding capacity of soil
11. Bulk density of soil
12. Carbon content of soil
13. Nitrogen content of soil
14. Compound topographic index (relative wetness)
15. Solar interception
16. Day/night diurnal temperature difference
17. Elevation

Methodology

- Statistical imputation of suitability (Importance Value/Basal Area)
- All imputations are done in the data space (not geographical space)
- 1.6M cells in CONUS at 4 km^2
- 48.6M cells on the globe (9 globes)
- Analysis carried out for Present and two Futures (A1FI, B1) time periods (2050, 2100)

Productivity Metric

Importance Value is measure of the relative dominance of species in a forest community. Importance values rank species within a site based upon three criteria:

- **Frequency:** the percentage of inventory points occupied by a given species, a measure of species distribution across the site
- **Density:** the average number of individuals per unit area (per acre or hectare)
- **Dominance:** the average dominance each species within the study area is estimated by its total basal area per unit area (ft^2 per acre or m^2 per hectare)

Importance value = Relative frequency + Relative density + Relative dominance

Imputation Schemes

Point based:

- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Imputation Schemes

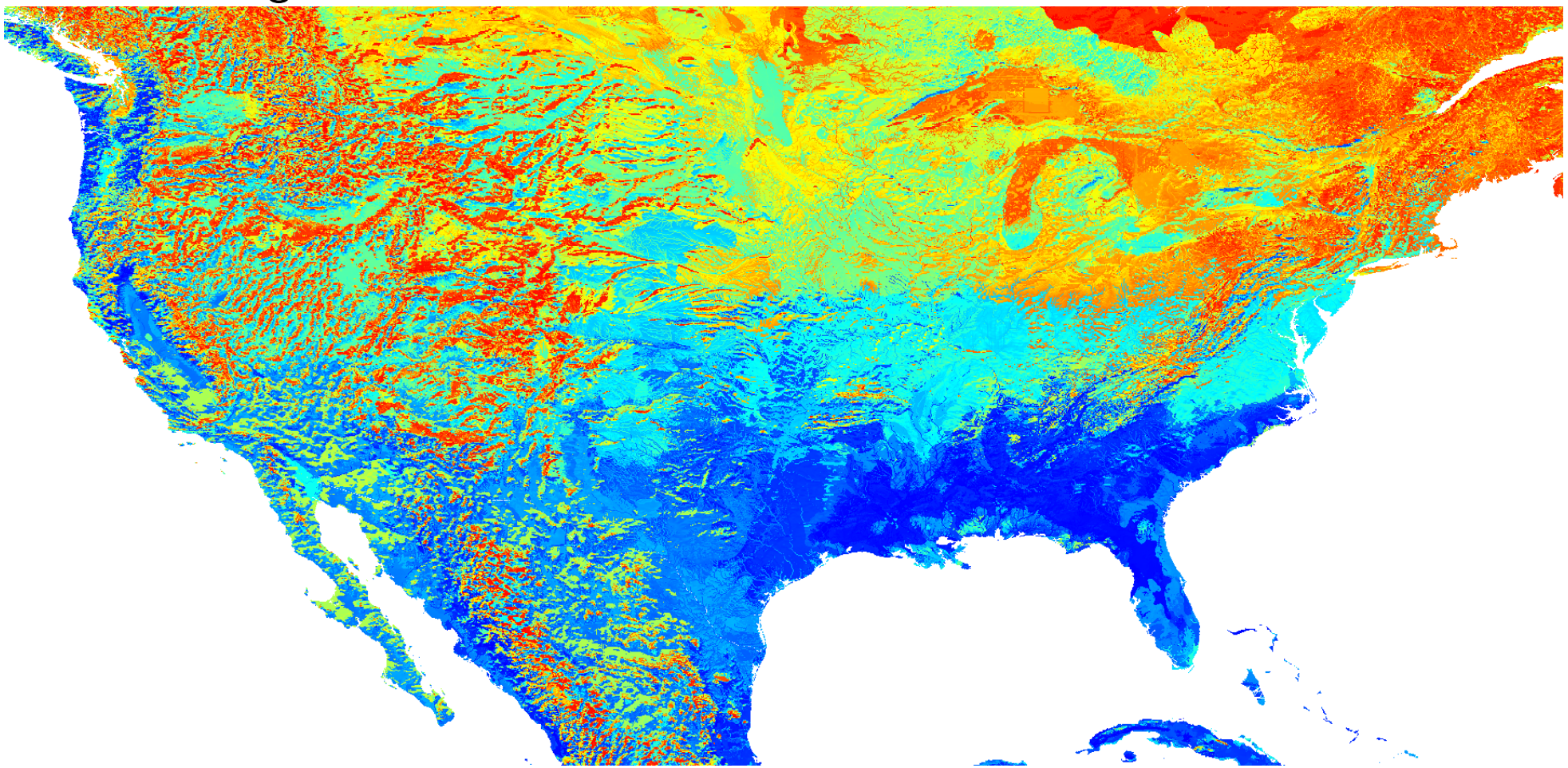
Point based:

- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Validation for extent using existing range maps and for magnitude using FIA measurements.

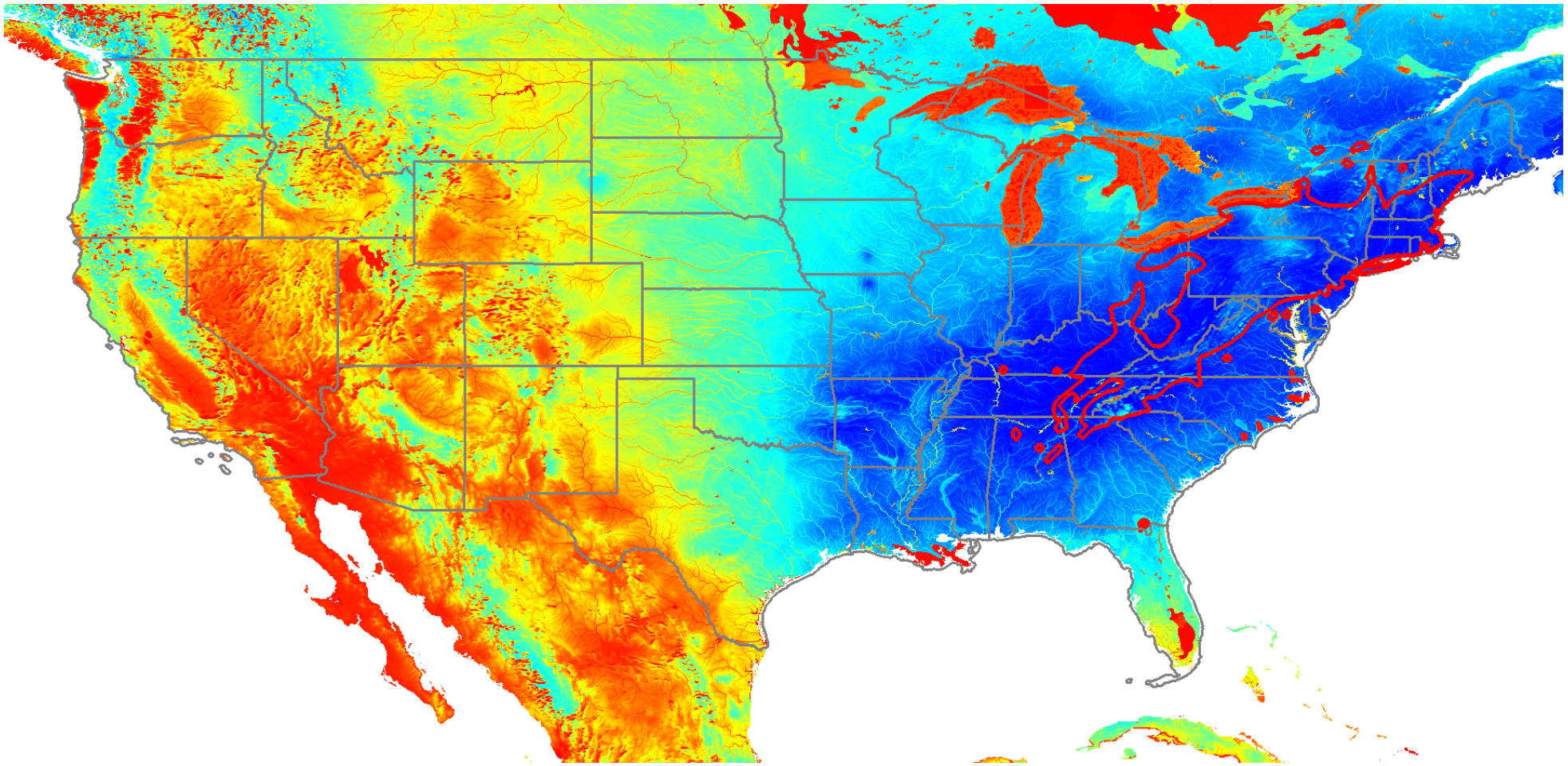
Point based nearest neighbor approach: Inverse distance weighted mean

Betula Lenta: Continuous map of productivity.. Where to draw the tree range borders???



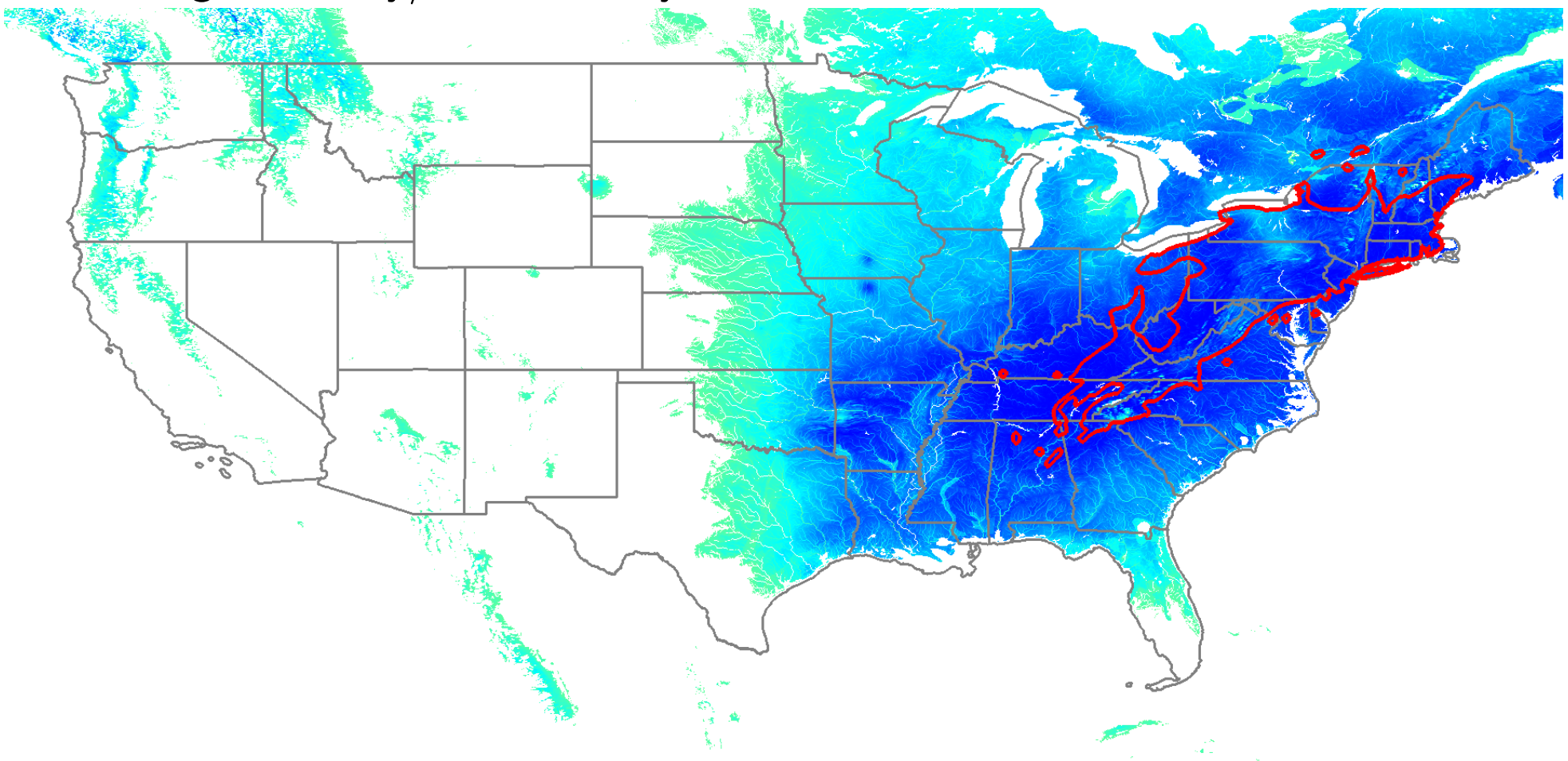
Point based nearest neighbor approach: Inverse distance weighted mean

Distance in the N-dimensional (climate conditions) space..



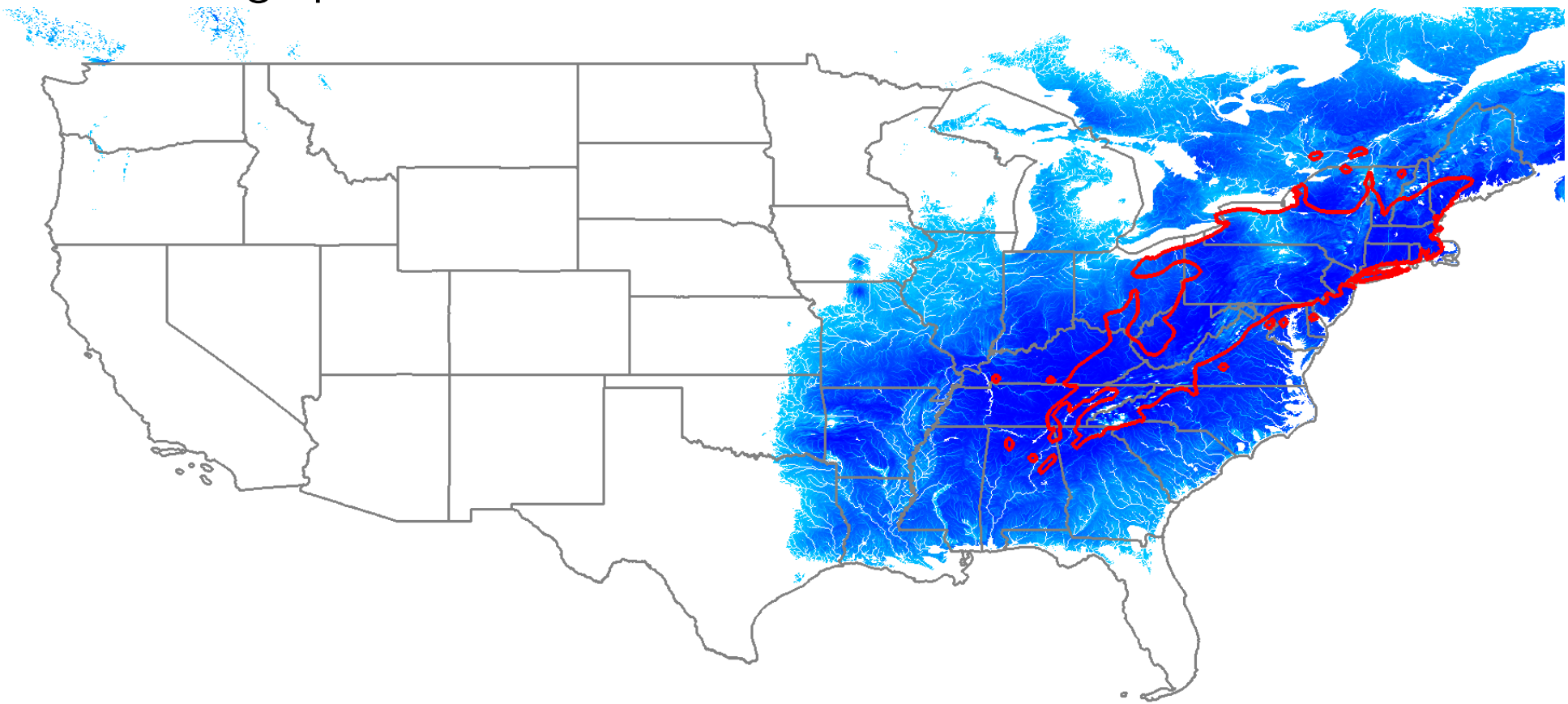
Point based nearest neighbor approach: Inverse distance weighted mean

Probably we can apply some threholds to restrict the tree range using similarity/dis-similarity...



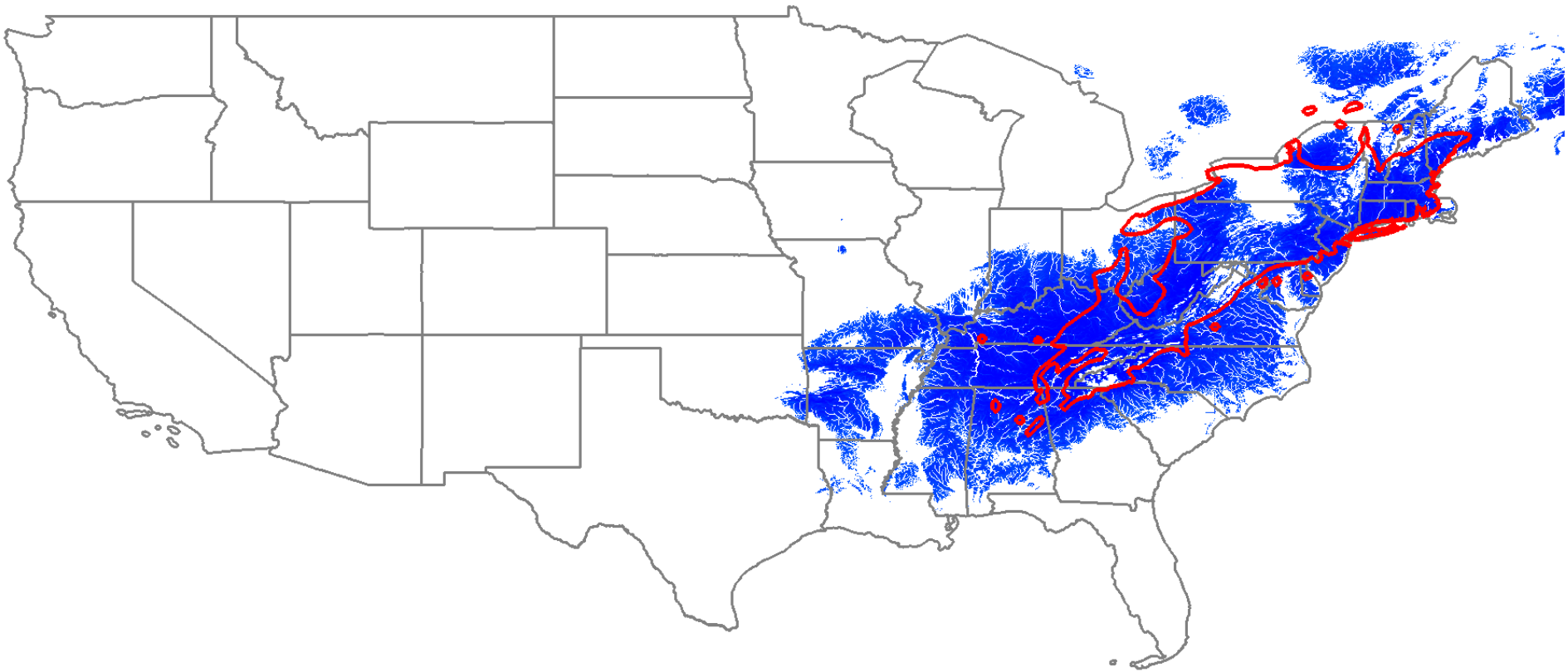
Point based nearest neighbor approach: Inverse distance weighted mean

Cranking up the threshold..



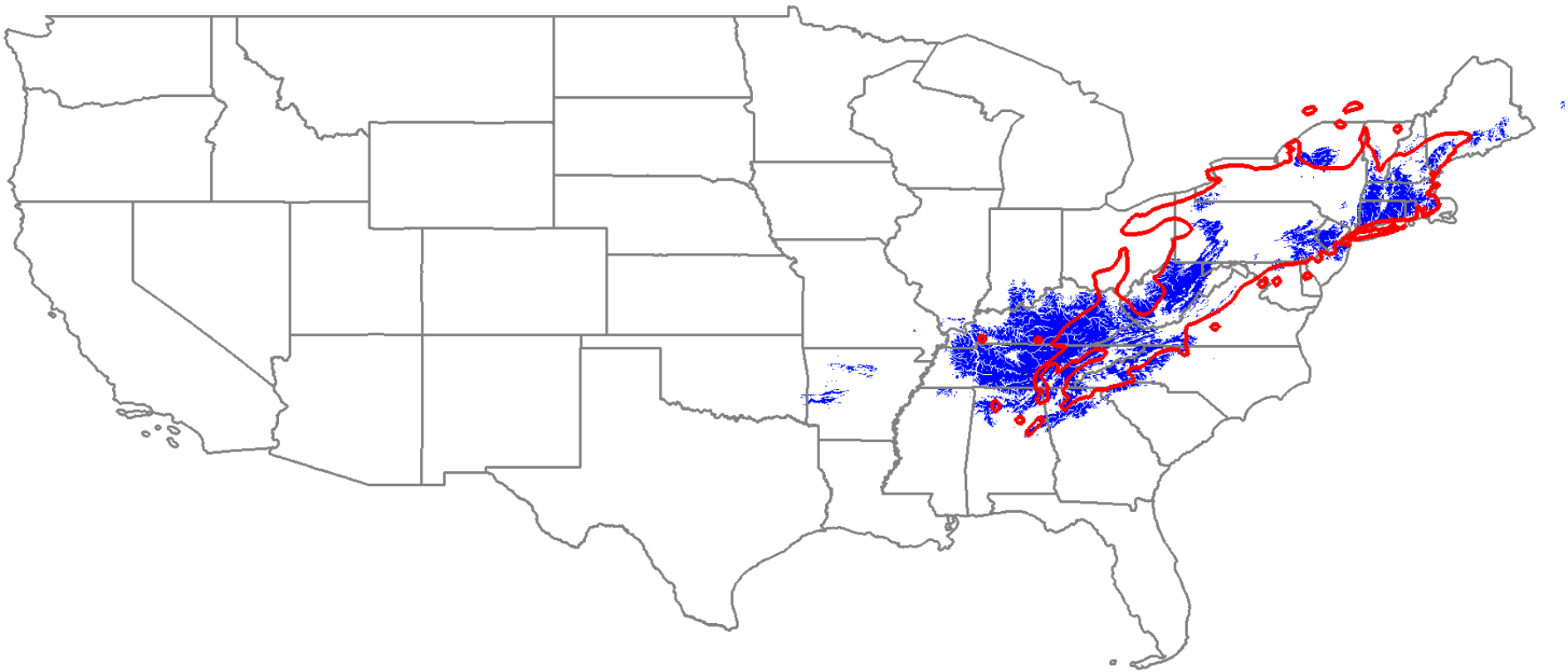
Point based nearest neighbor approach: Inverse distance weighted mean

Cranking up the threshold.. little more!



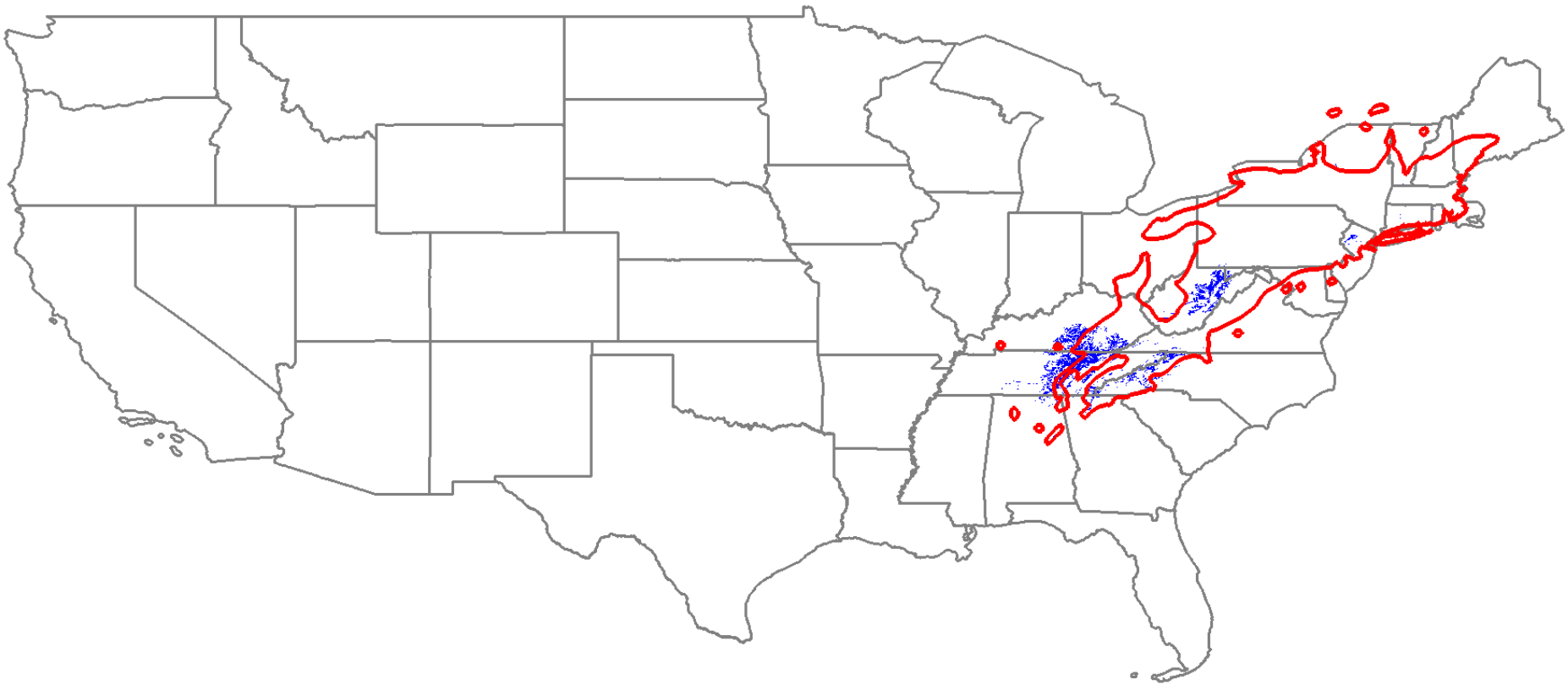
Point based nearest neighbor approach: Inverse distance weighted mean

Cranking up the threshold.. too much!



Point based nearest neighbor approach: Inverse distance weighted mean

Cranking up the threshold.. too much!



Imputation Schemes

- We can calculate productivity using various regression/kriging approaches

Imputation Schemes

- We can calculate productivity using various regression/kriging approaches
- But that solves one part of the problem. Ecoregion boundaries are difficult to draw.. expert ecologists are good at this.. but its subjective

Imputation Schemes

- We can calculate productivity using various regression/kriging approaches
- But that solves one part of the problem. Ecoregion boundaries are difficult to draw.. expert ecologists are good at this.. but its subjective
- We have developed and applied a "Associative Clustering" methodology

Imputation Schemes

- We can calculate productivity using various regression/kriging approaches
- But that solves one part of the problem. Ecoregion boundaries are difficult to draw.. expert ecologists are good at this.. but its subjective
- We have developed and applied a "Associative Clustering" methodology
- We delineate the Continental United States in climatic ecoregions using high resolution data sets (k-means clustering)
- Associate dependent variables (species level data) to identify the ecoregions suitable for any given species

Imputation Schemes

Point based:

- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Clustered ecoregions (centroid) based:

- Nearest neighbor
- Mean value
- Maximum value
- Inverse weighted distance mean
- Multi-linear regressions

Clustered ecoregions (point) based:

- Nearest neighbor
- Inverse weighted mean
- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Imputation Schemes

Point based:

- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Clustered ecoregions (centroid) based:

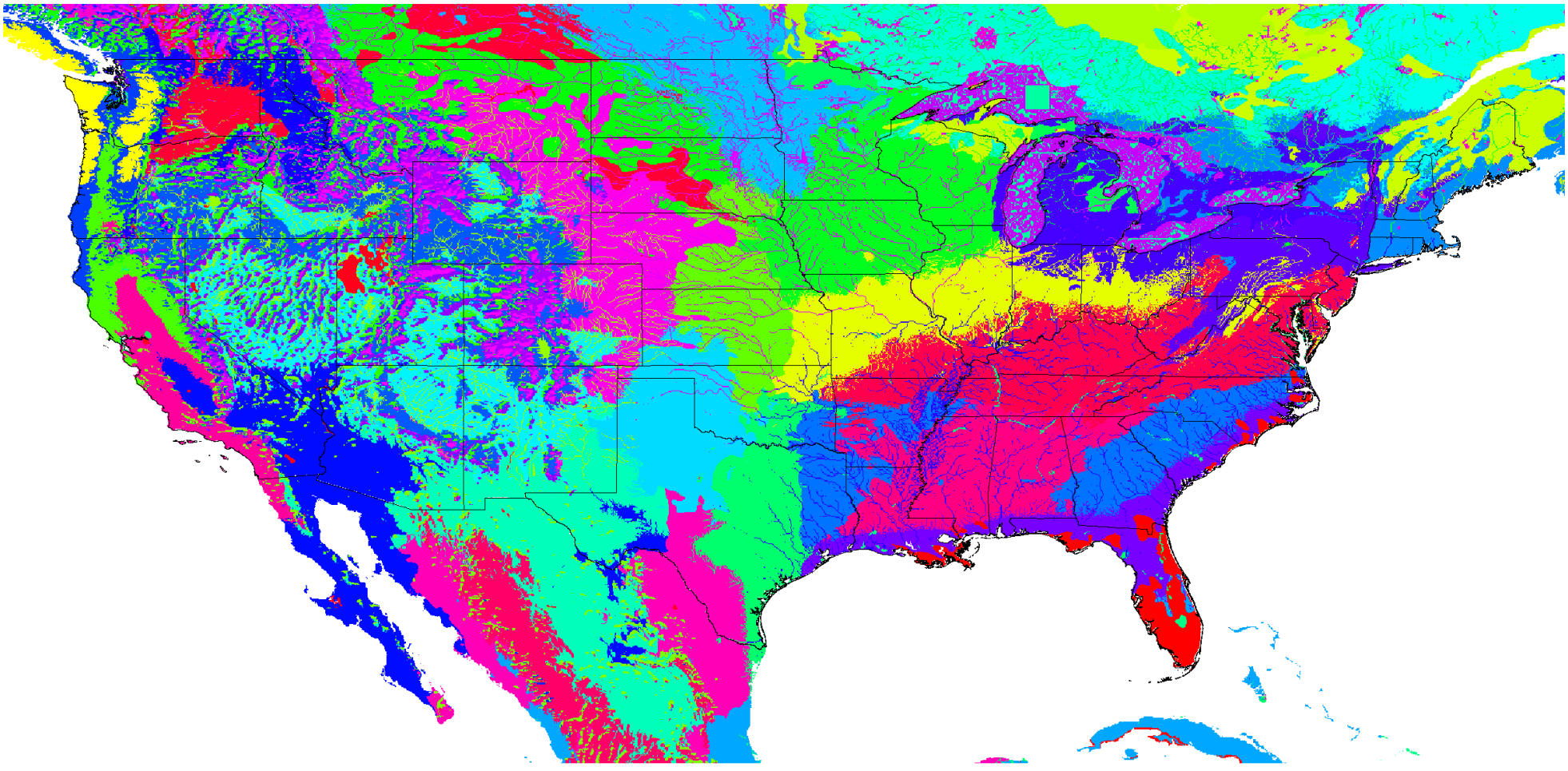
- Nearest neighbor
- Mean value
- Maximum value
- Inverse weighted distance mean
- Multi-linear regressions

Clustered ecoregions (point) based:

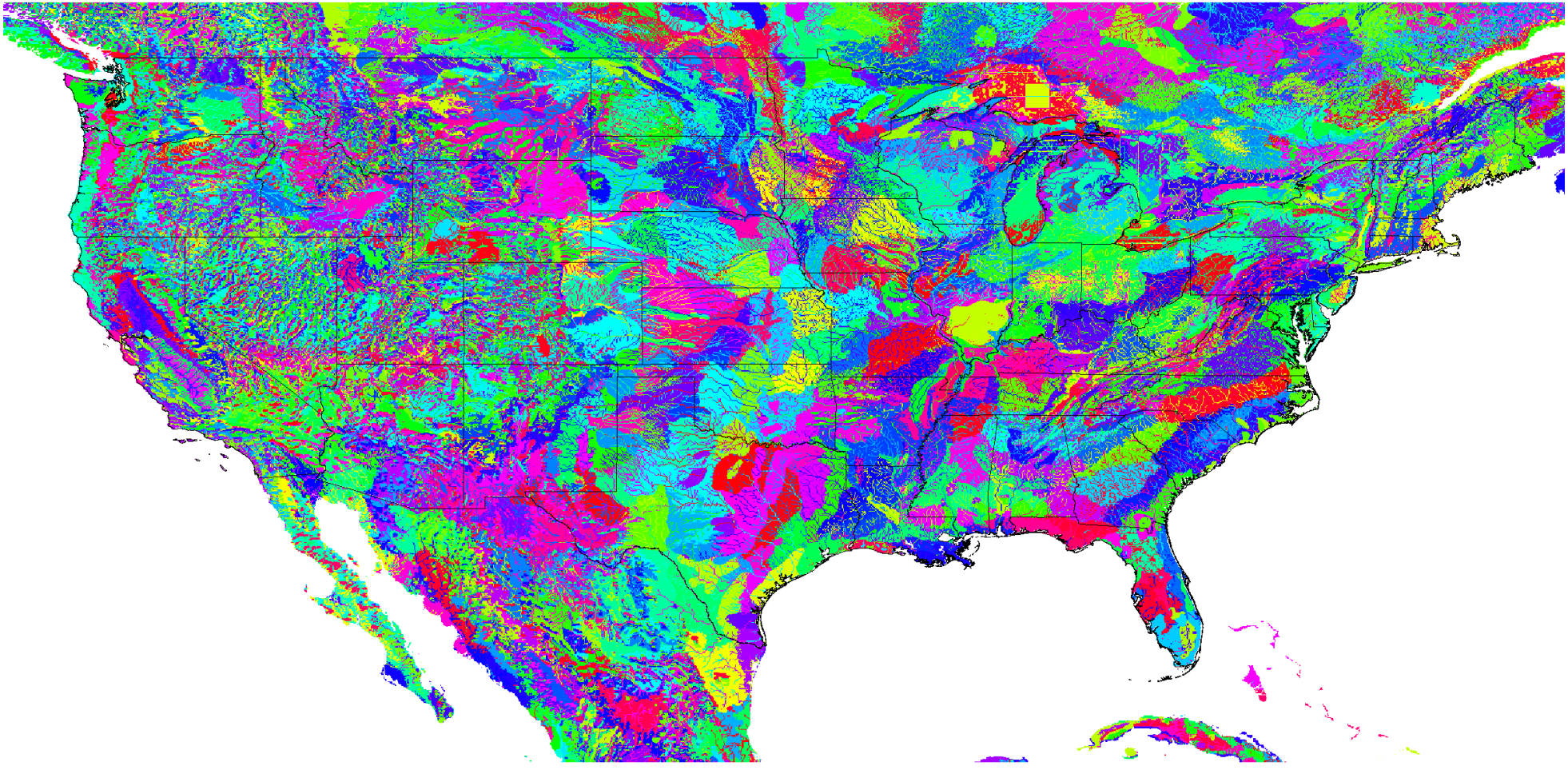
- Nearest neighbor
- Inverse weighted mean
- N-nearest neighbor inverse weighted distance mean
- N-nearest neighbor percentiles
- Multi-linear regressions

Validation for extent using existing range maps and for magnitude using FIA measurements.

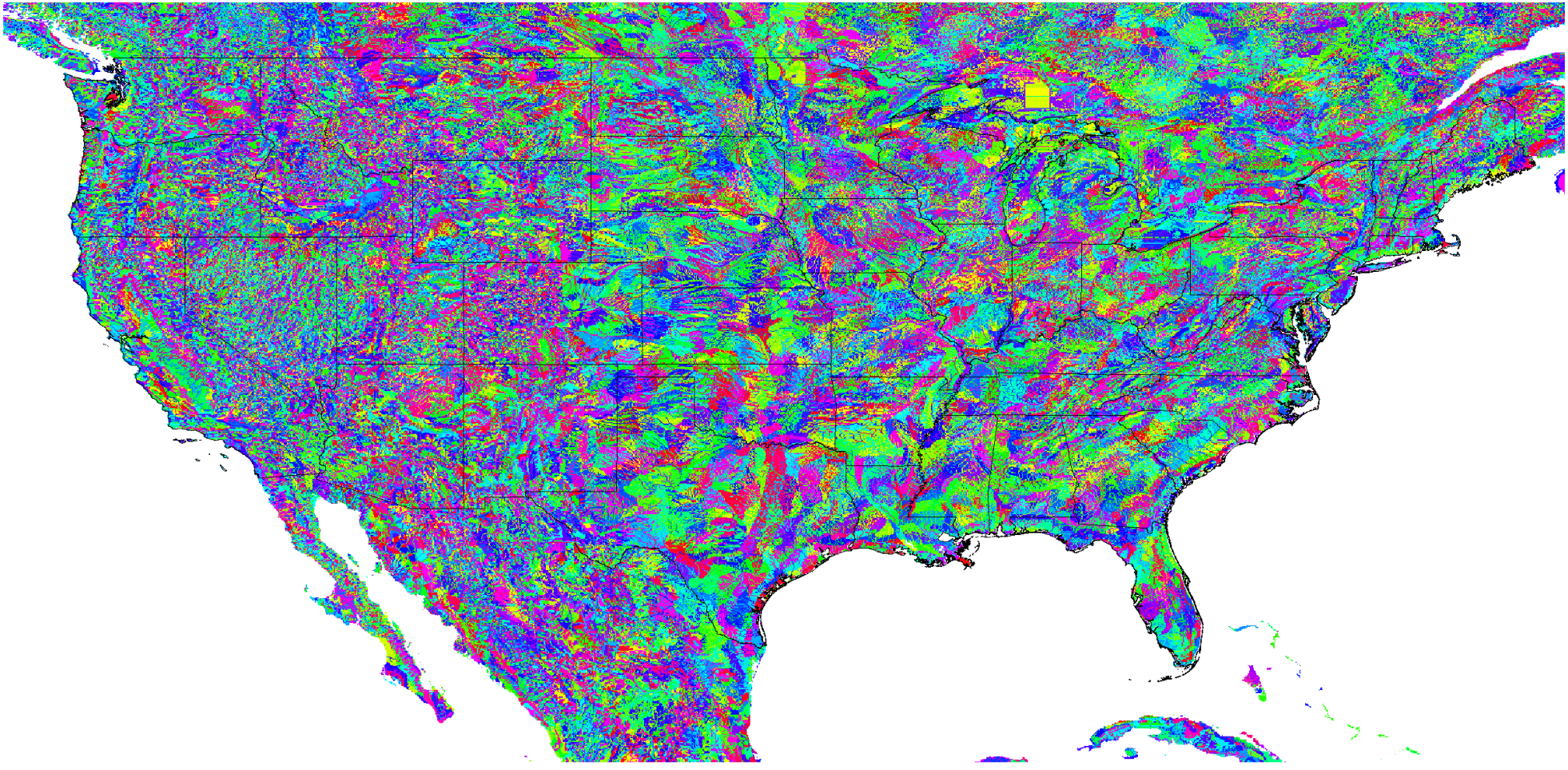
Clustered Ecoregions (K=50)



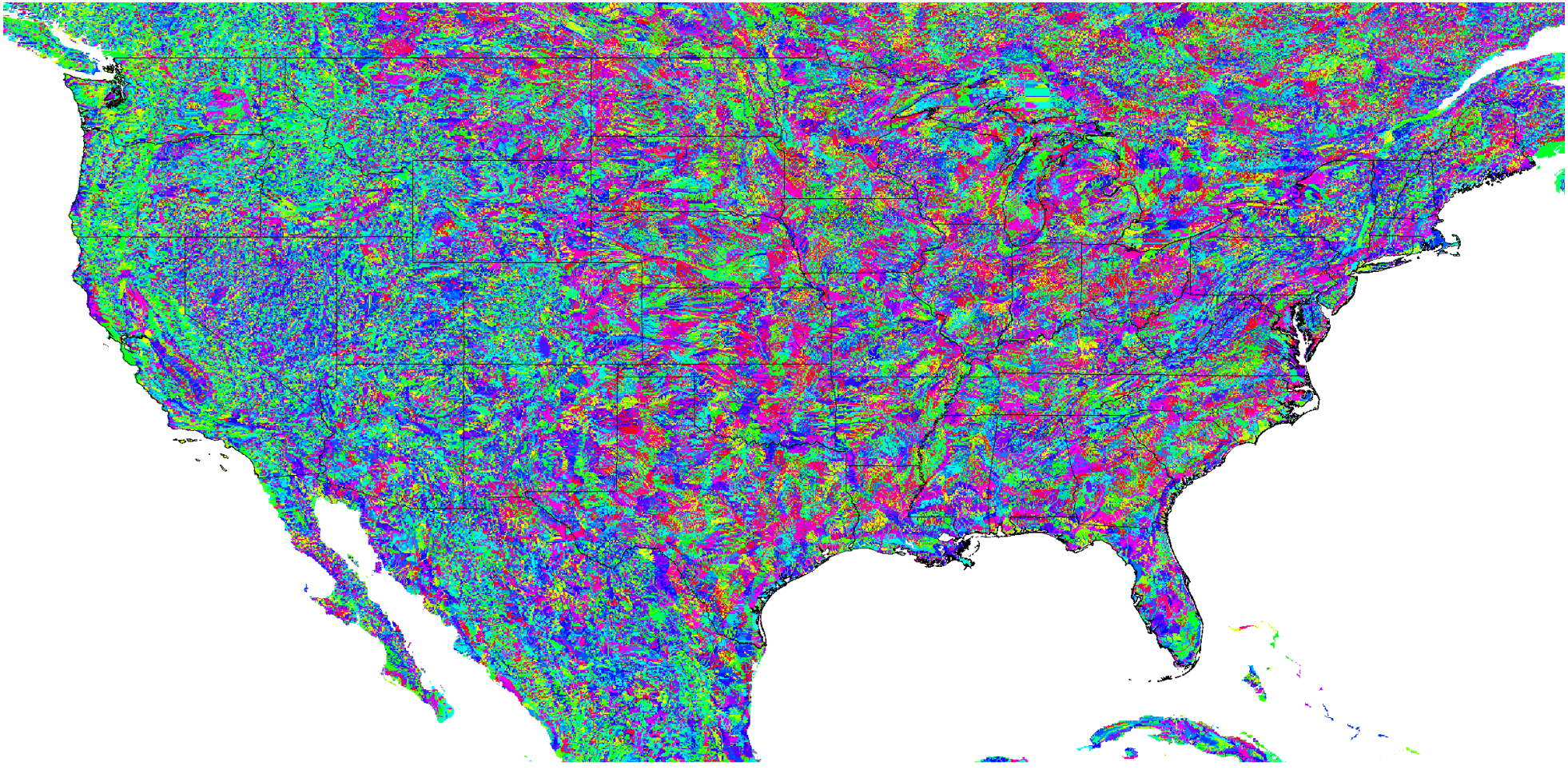
Clustered Ecoregions (K=1000)



Clustered Ecoregions (K=10000)

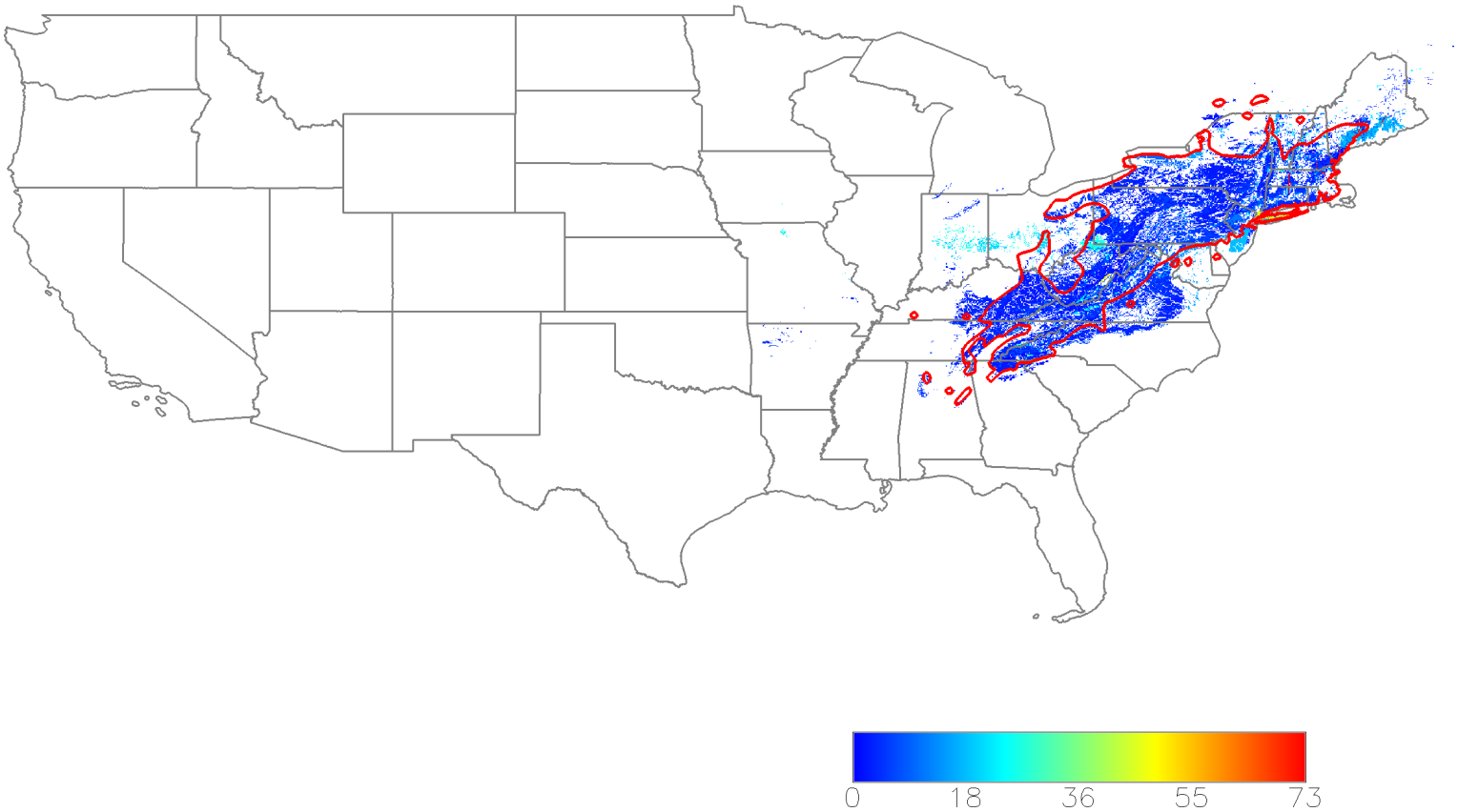


Clustered Ecoregions (K=20000)



Imputed Productivity Maps

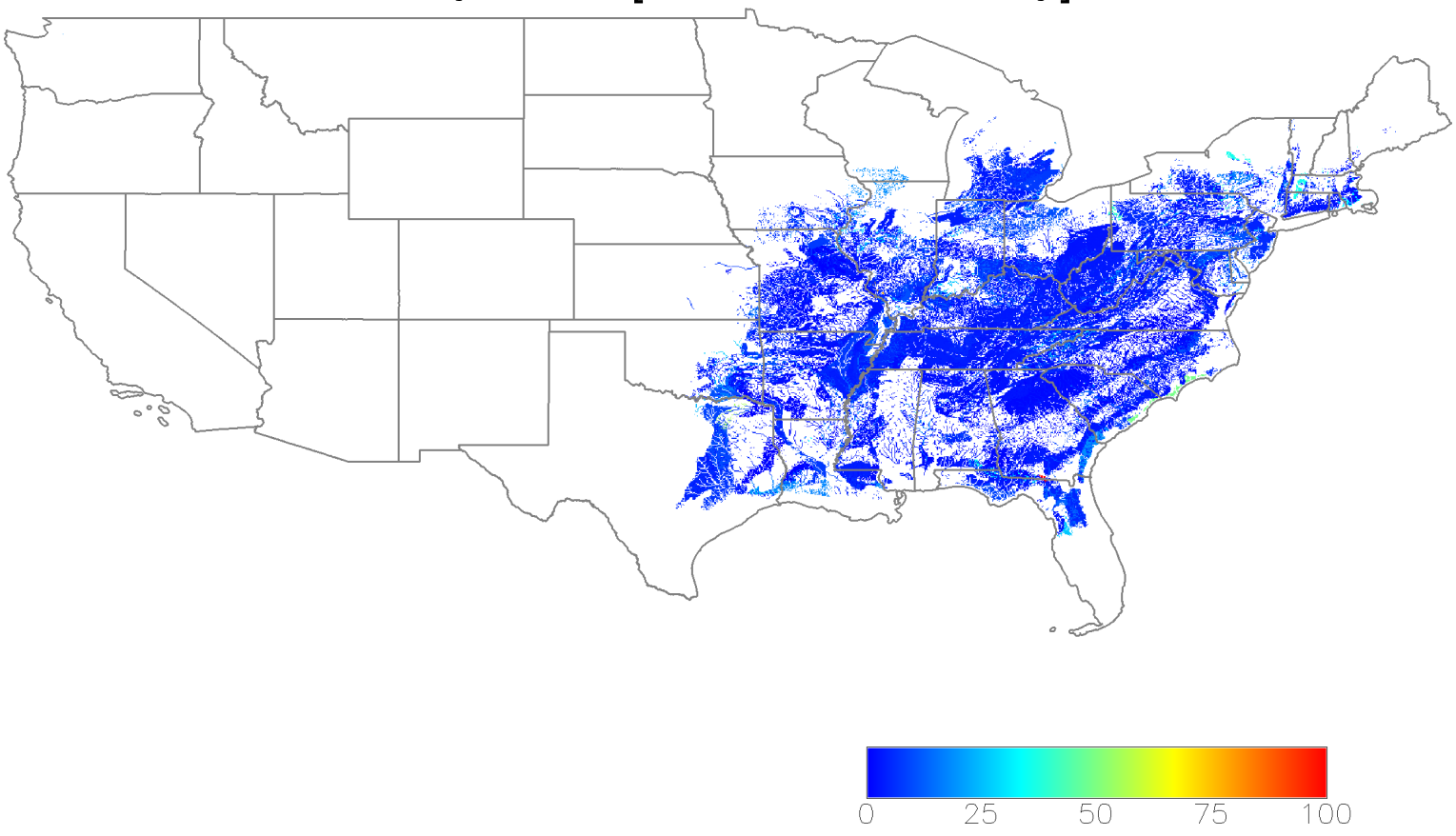
Betula lenta [Sweet Birch]



Polygons in red are Little's range map.

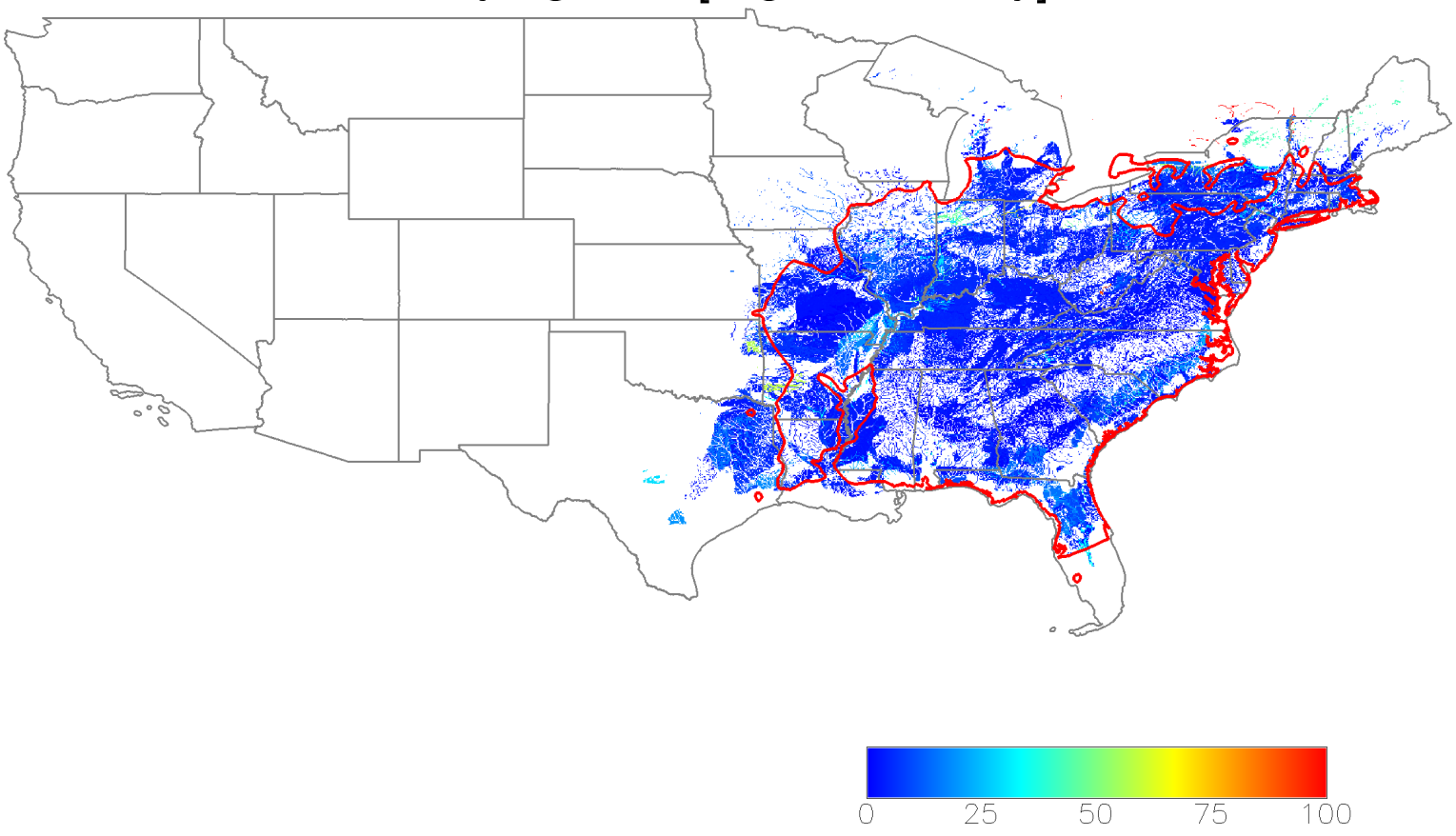
Imputed Productivity Maps

Carya alba [Mockernut Hickory]



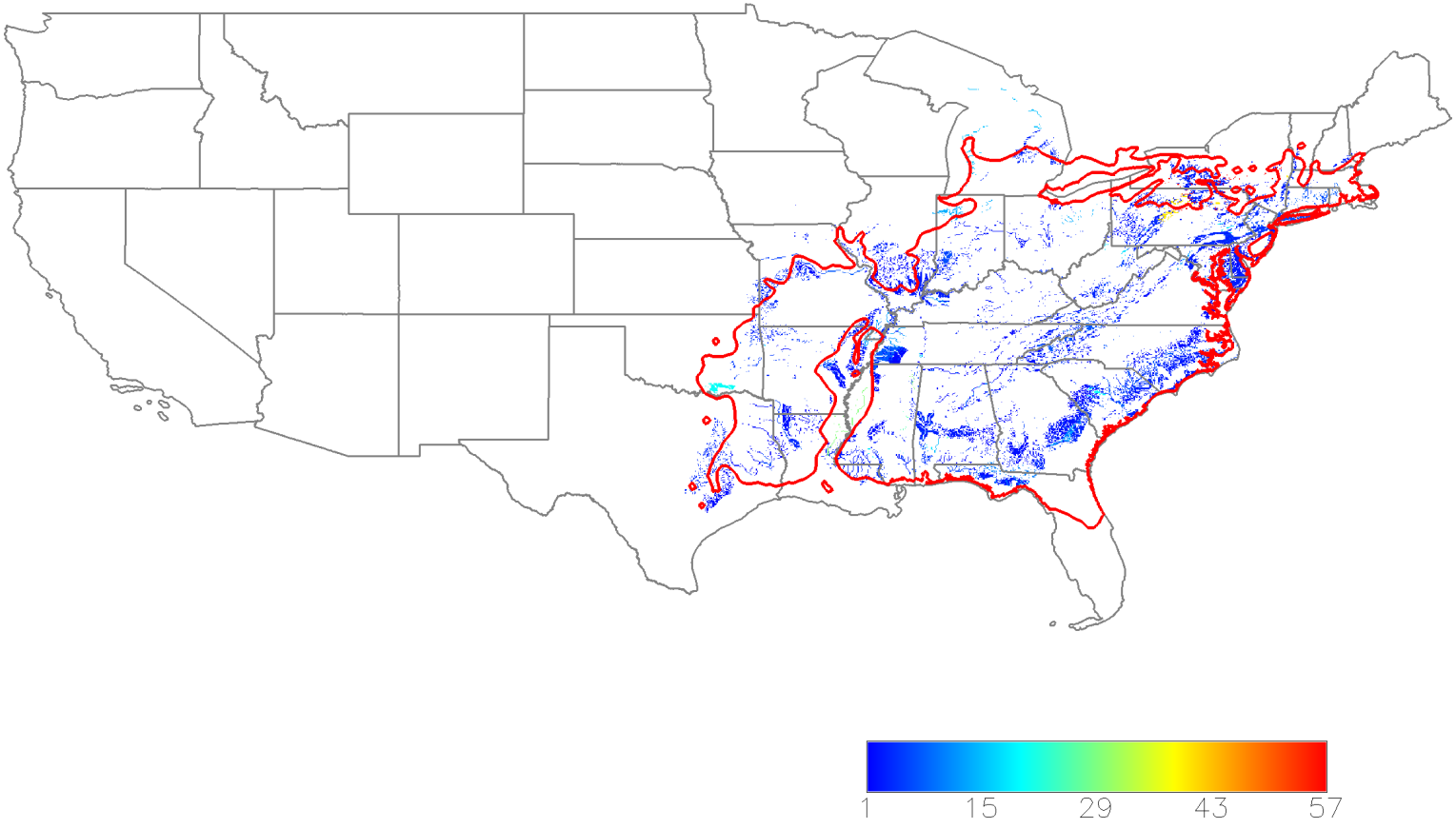
Imputed Productivity Maps

Carya glabra [Pignut Hickory]



Imputed Productivity Maps

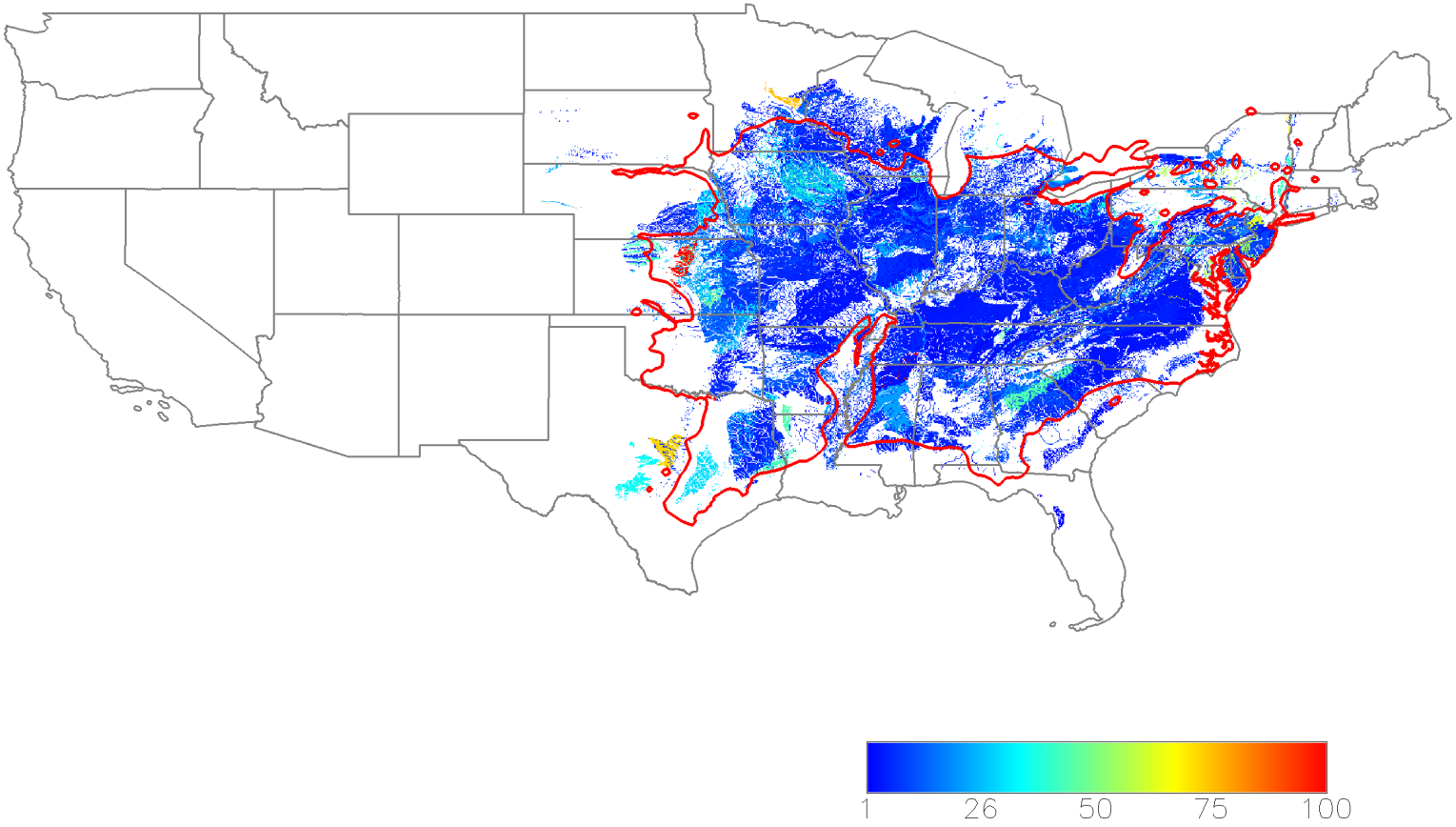
Cornus florida [Flowering Dogwood]



Polygons in red are Little's range map.

Imputed Productivity Maps

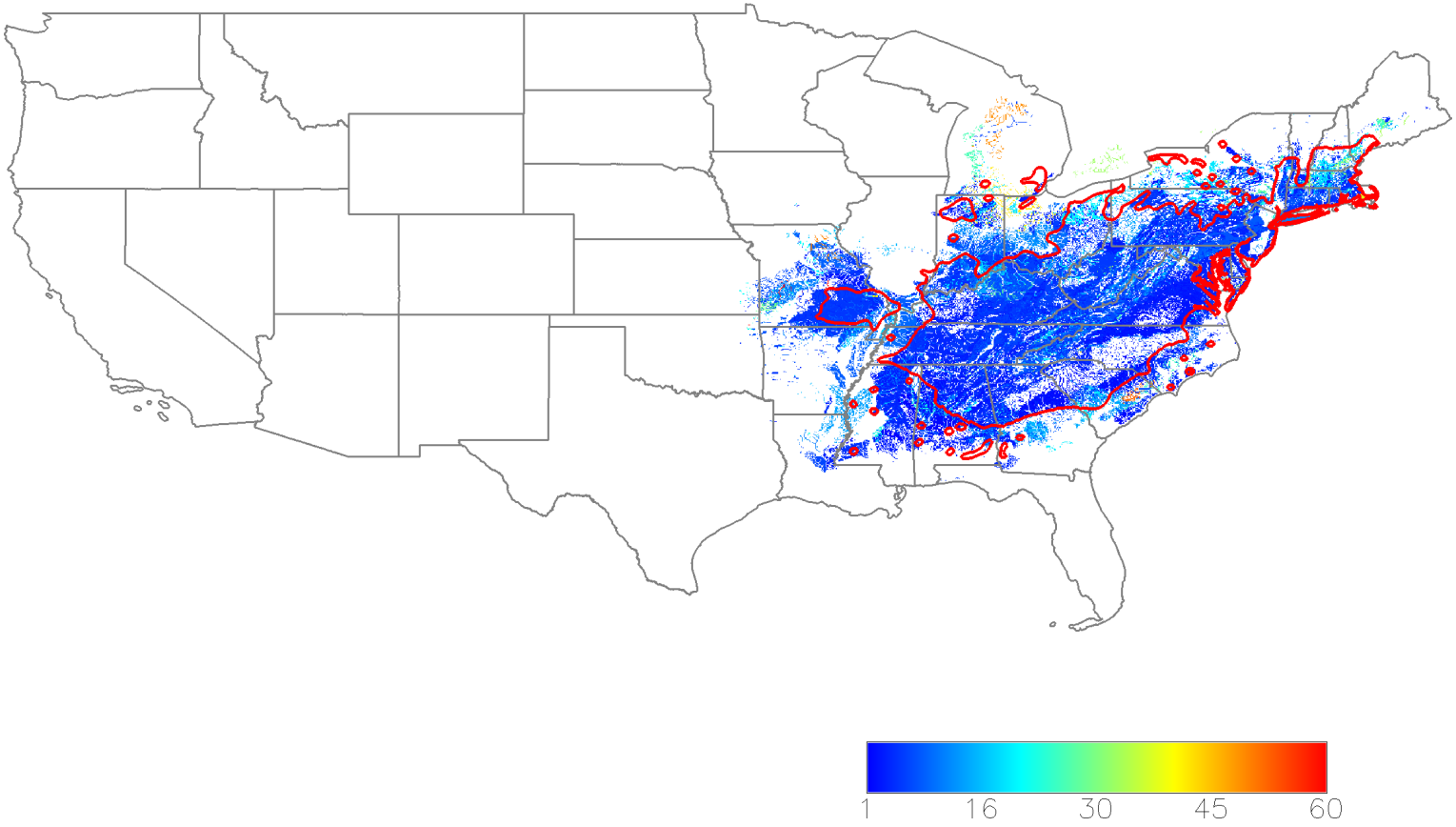
Juglans nigra [Black Walnut]



Polygons in red are Little's range map.

Imputed Productivity Maps

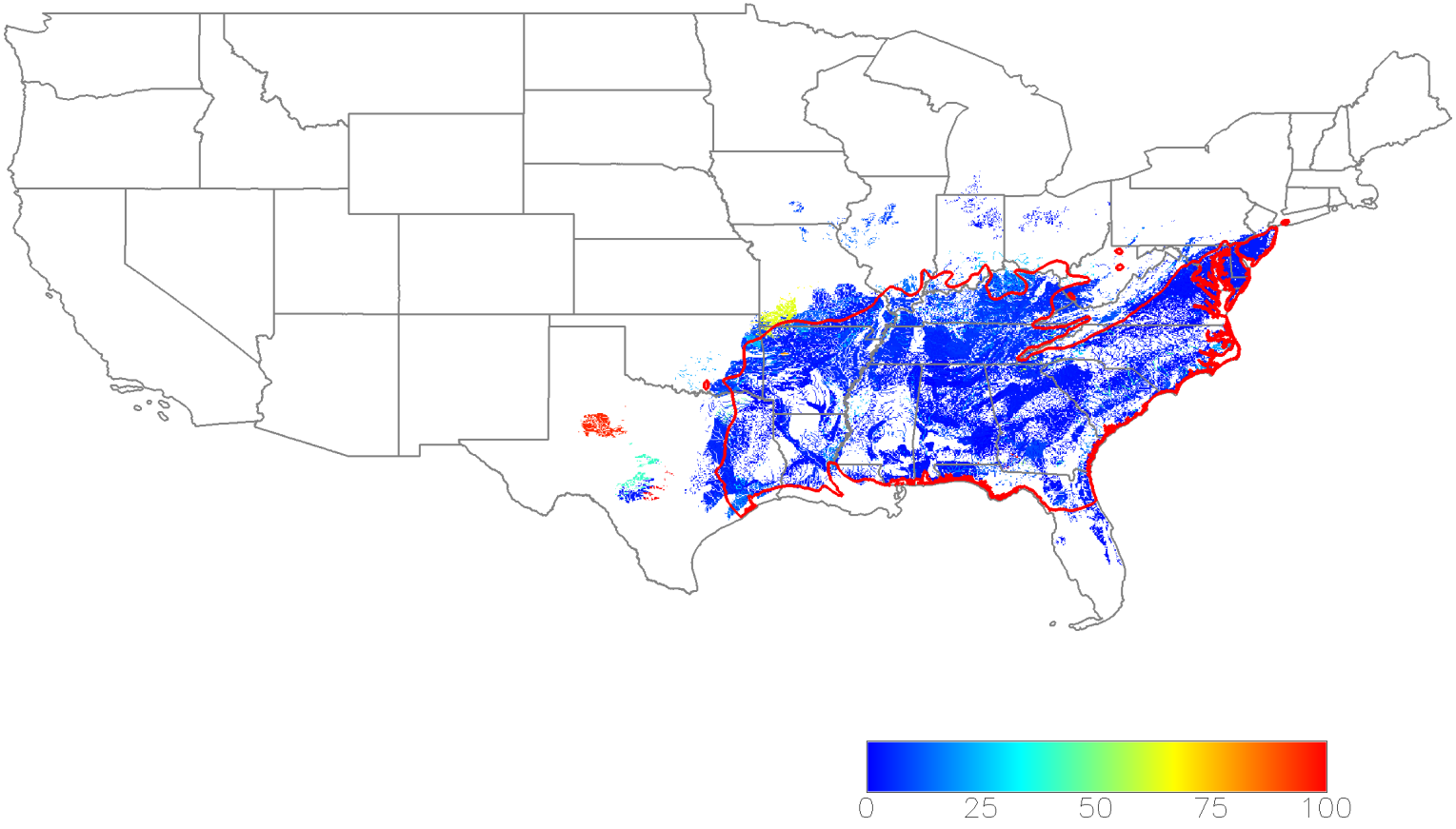
Quercus coccinea [Scarlet Oak]



Polygons in red are Little's range map.

Imputed Productivity Maps

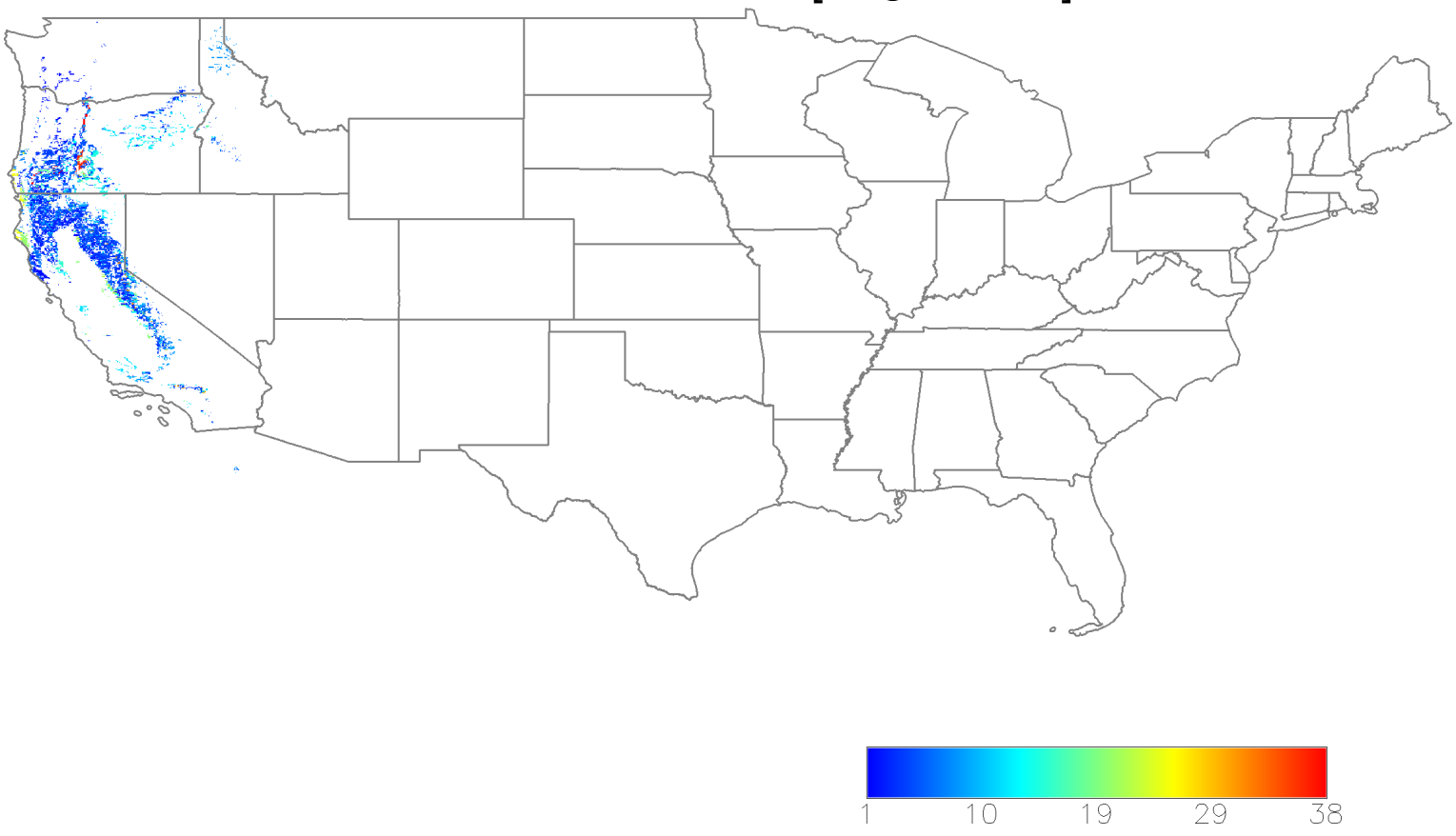
Quercus falcata [Southern Red Oak]



Polygons in red are Little's range map.

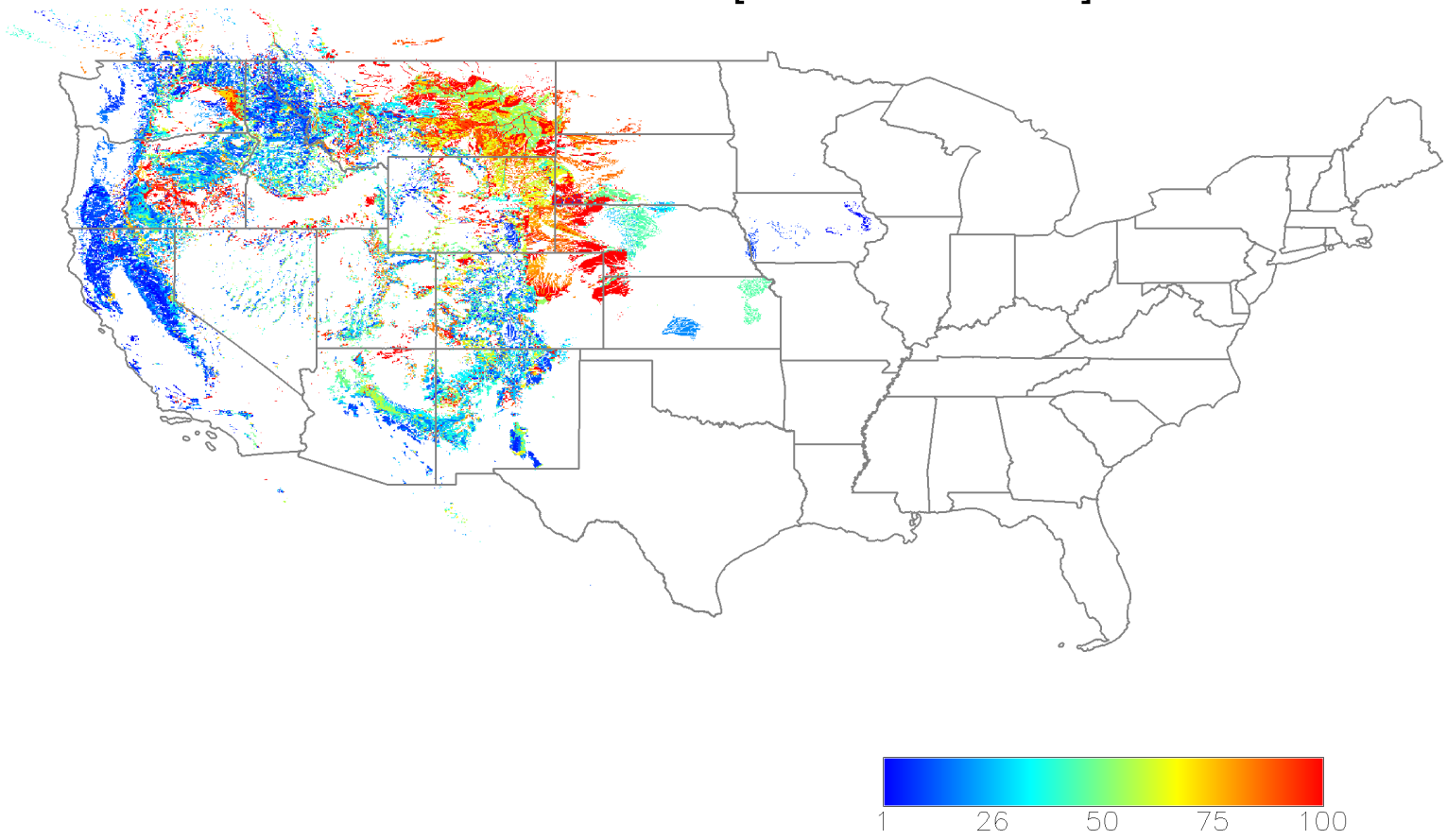
Imputed Productivity Maps

Pinus Lambertiana [Sugar Pine]



Imputed Productivity Maps

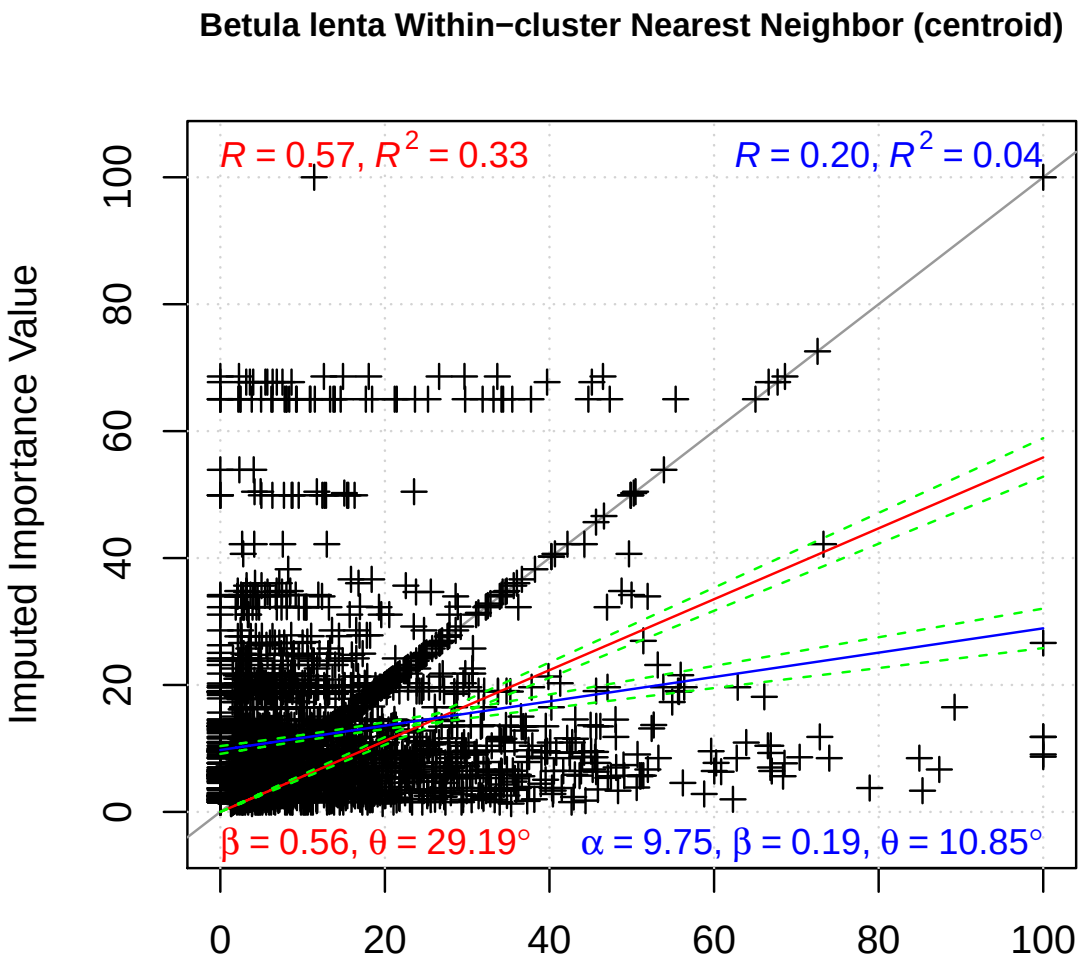
Pinus Ponderosa [Ponderosa Pine]



How well are we doing?
Validation using FIA observations.

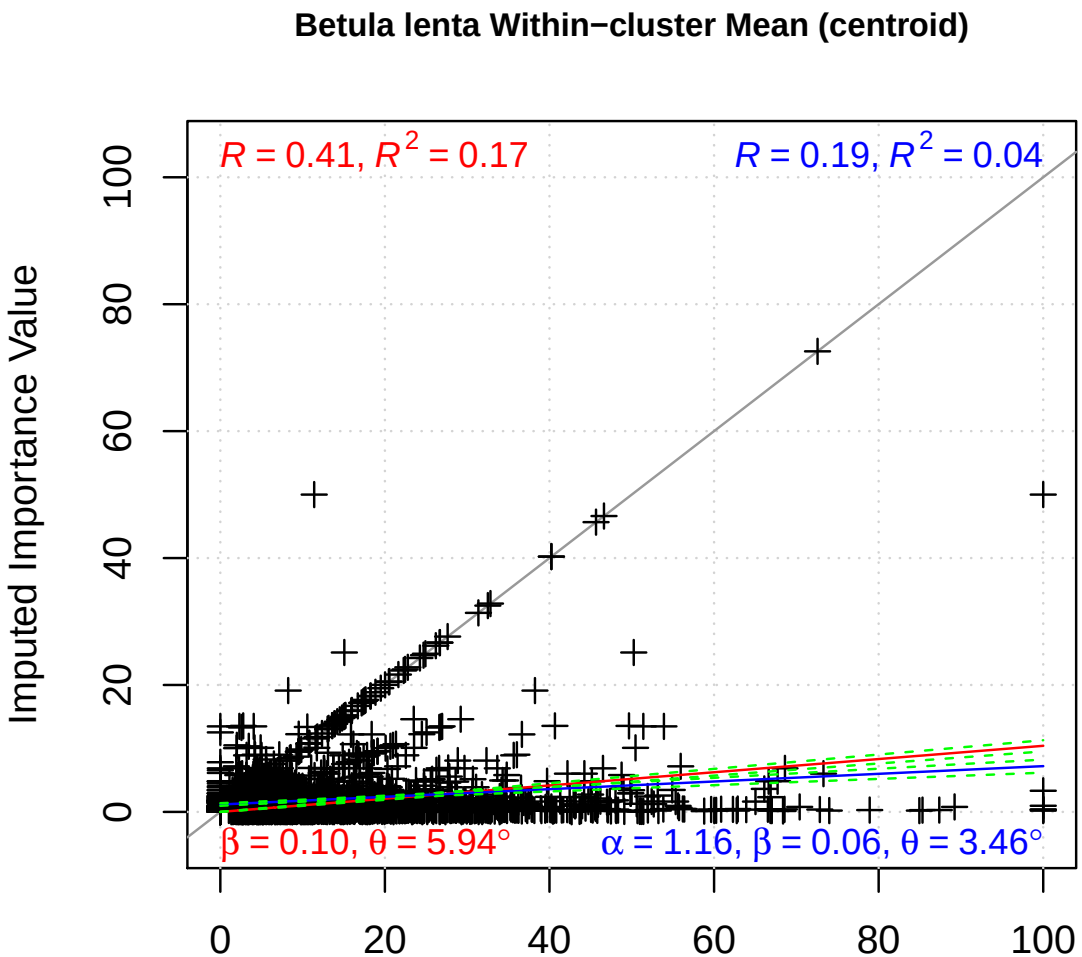
Validation of imputed importance values

Nearest neighbor within the cluster (centroid)



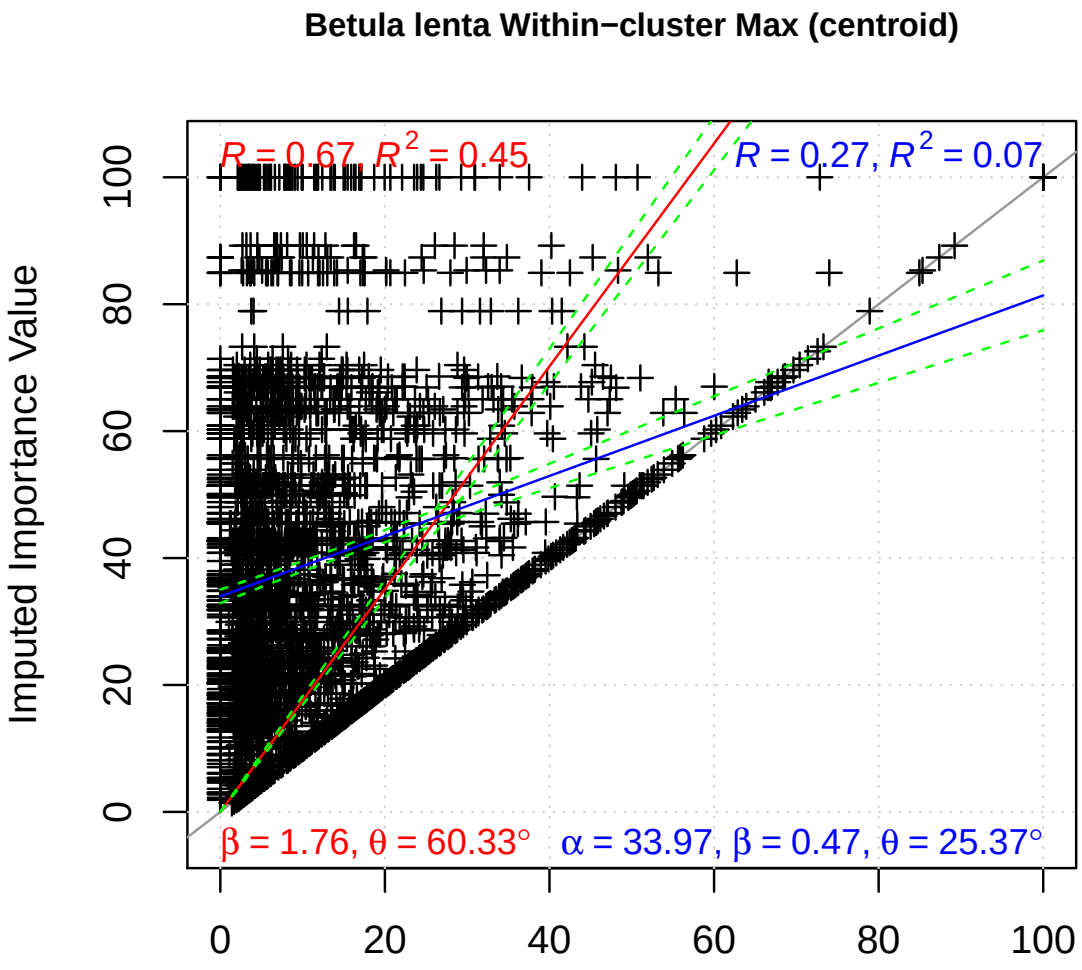
Validation of imputed importance values

Mean of points in the cluster (centroid)



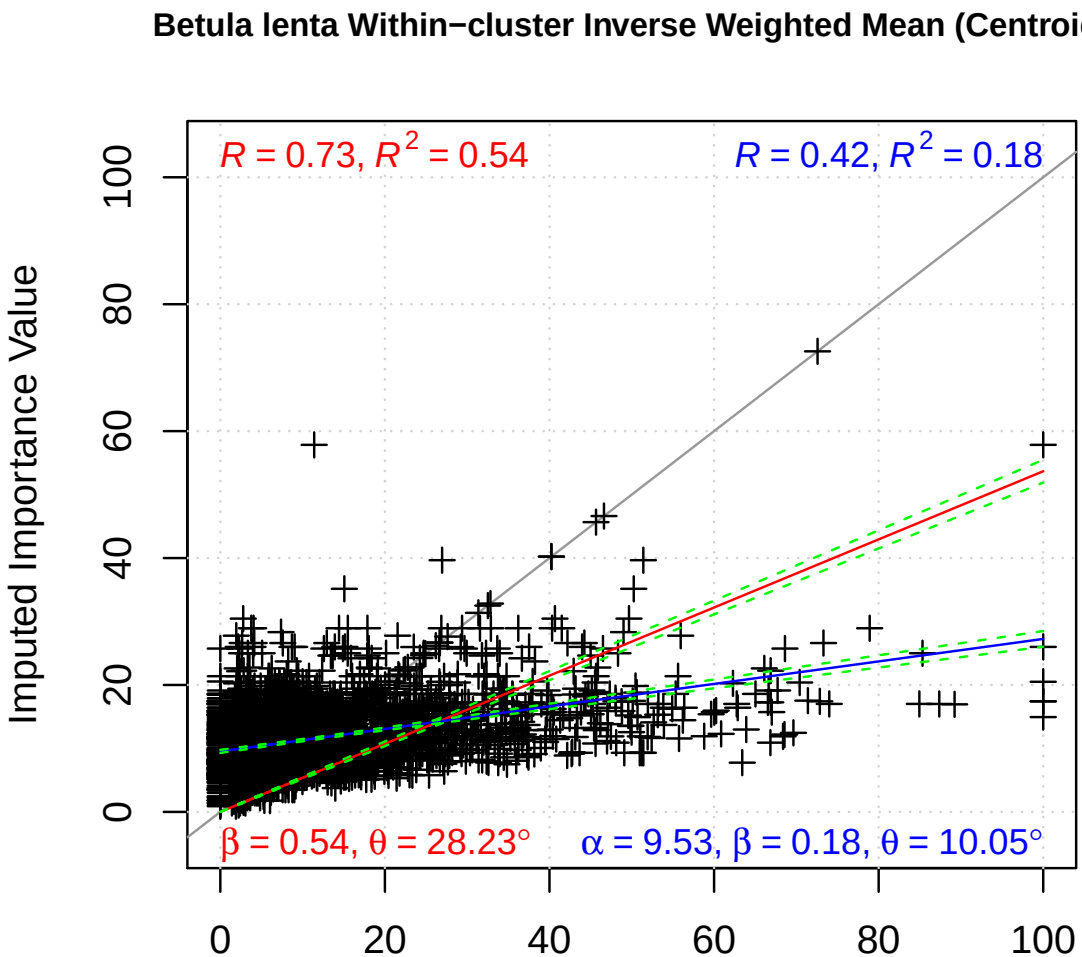
Validation of imputed importance values

Max of points in the cluster (centroid)

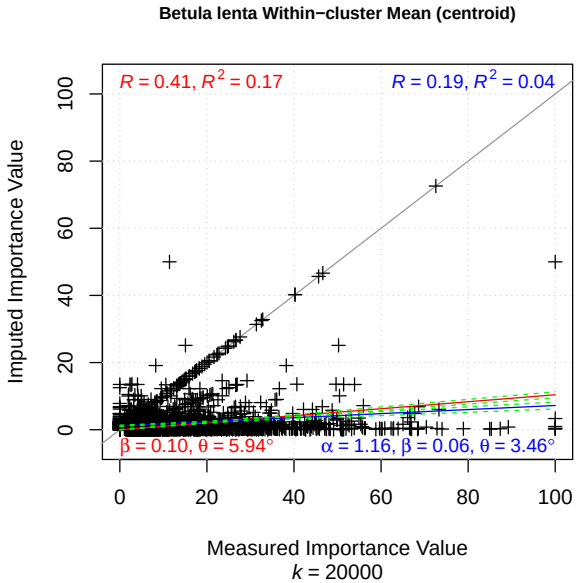


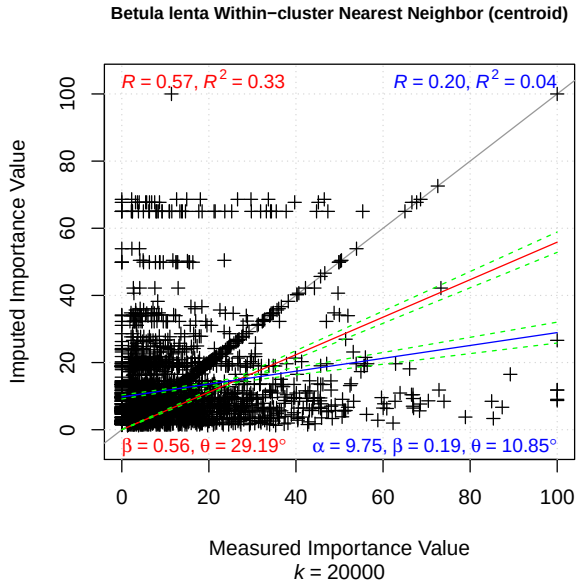
Validation of imputed importance values

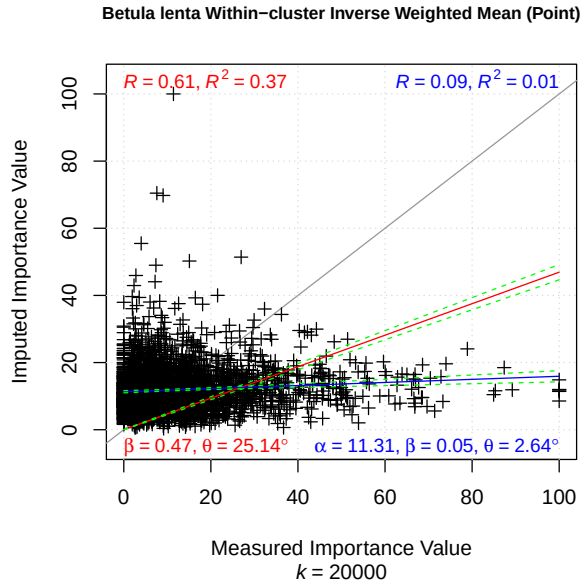
Inverse weighted distance mean of points in the cluster (centroid)

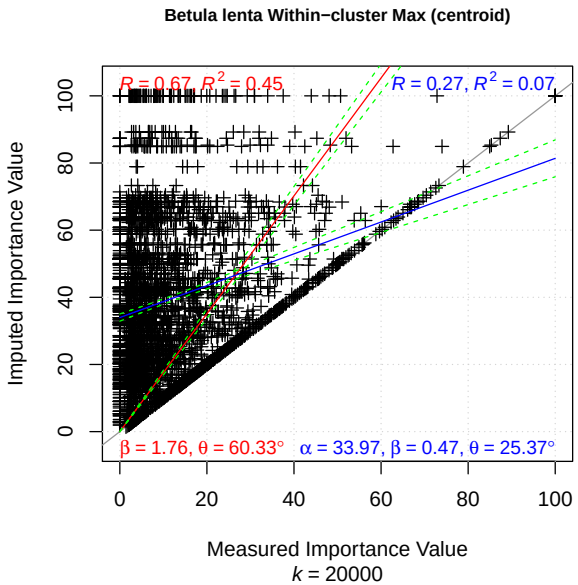


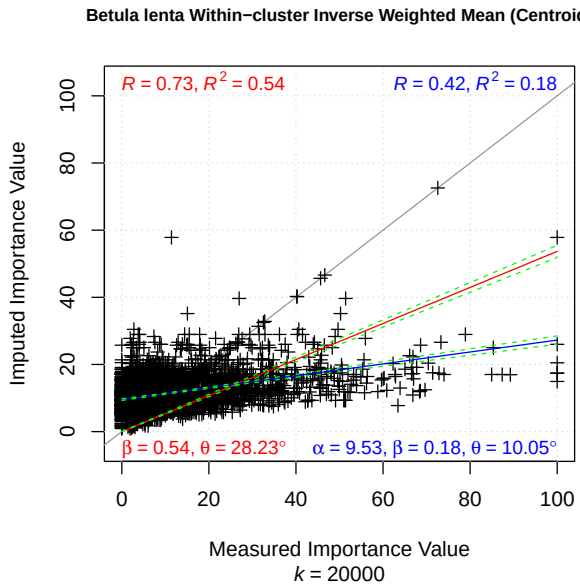
Betula Lenta: Performance of imputation schemes

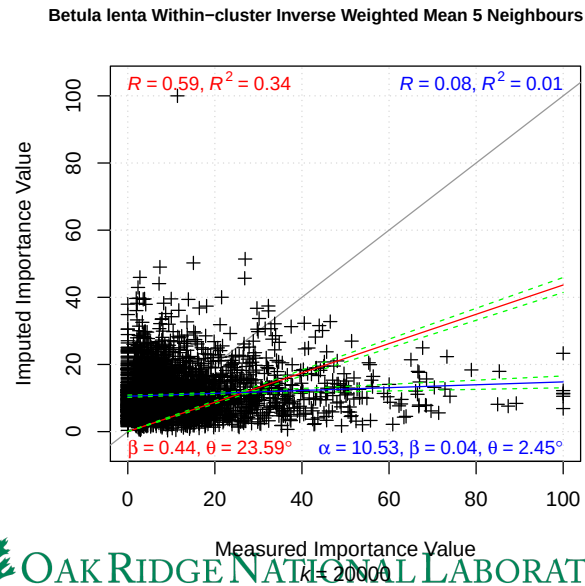




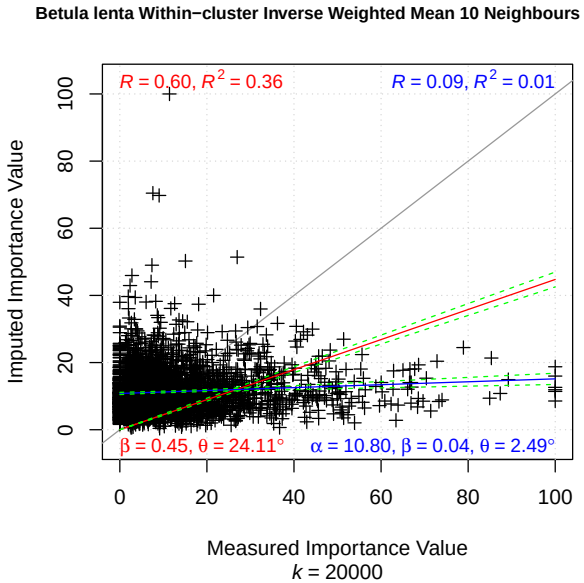


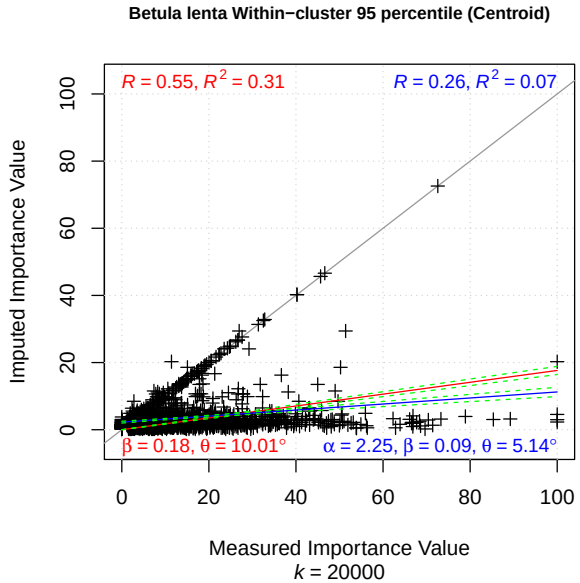


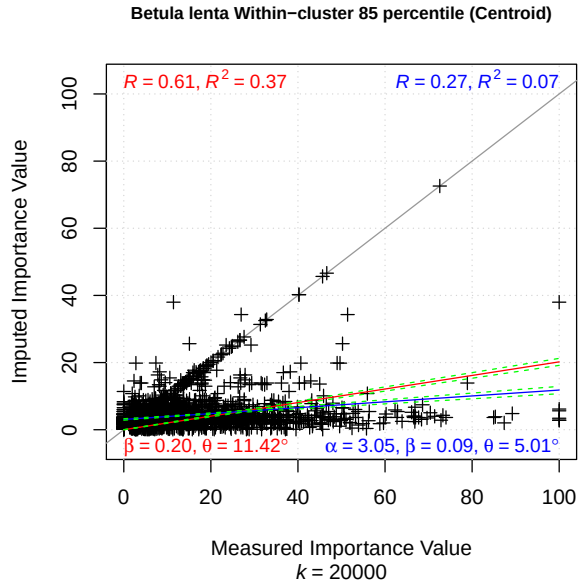


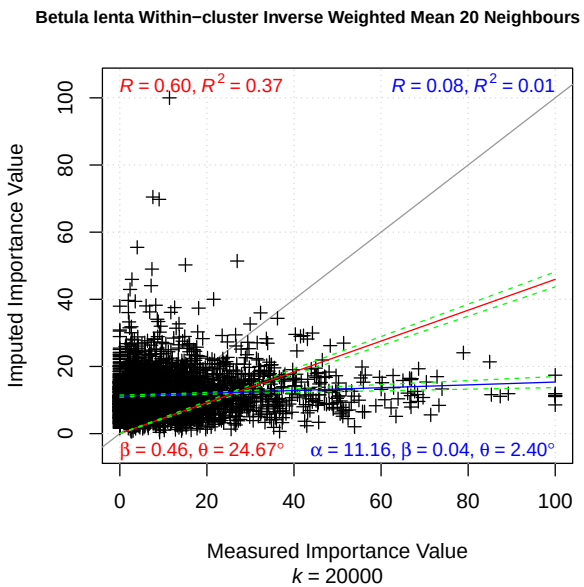


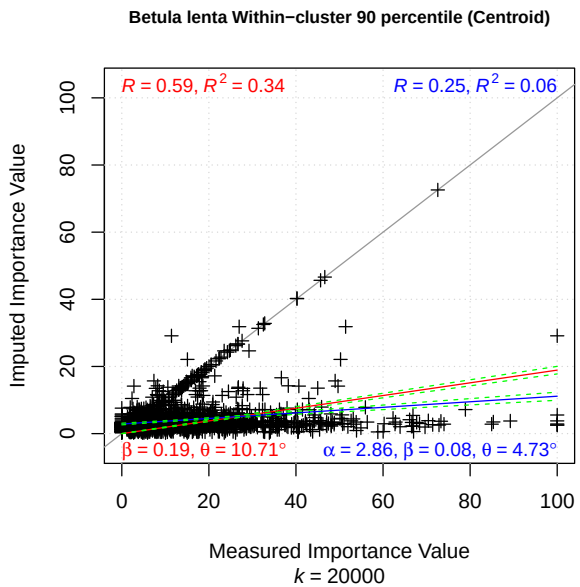
Betula Lenta: Performance of imputation schemes

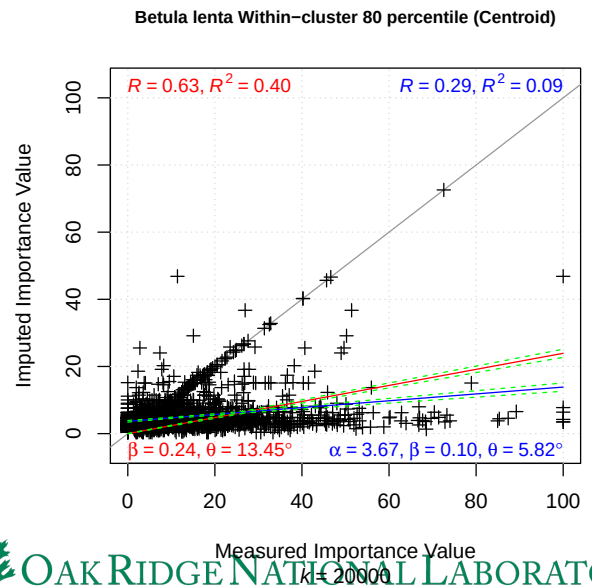








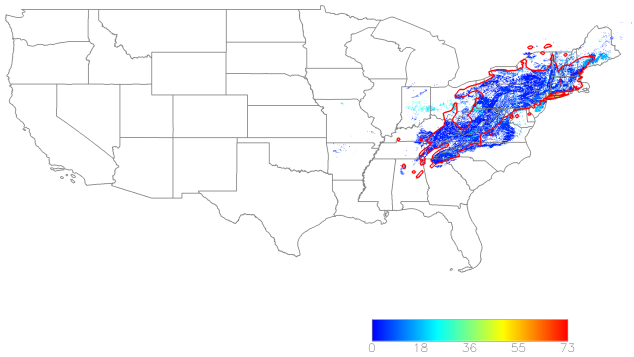


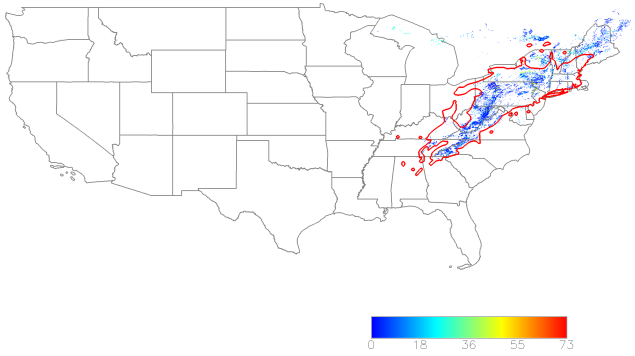


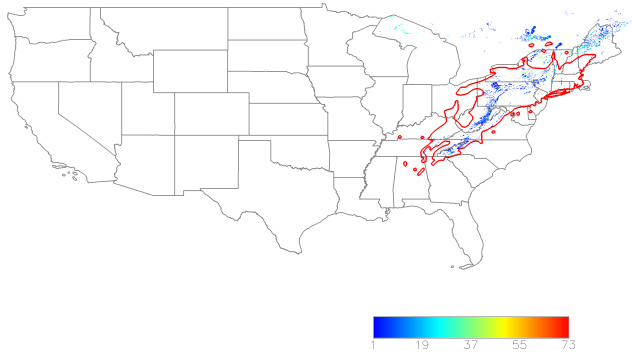
Time travel!!

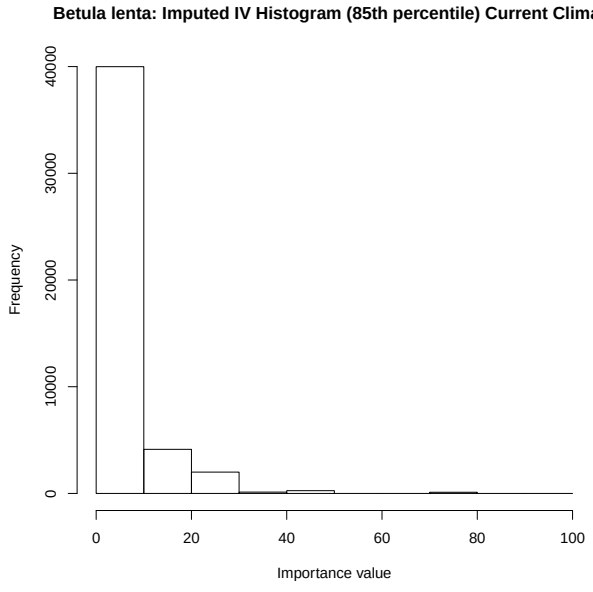
What can we do about projecting these to the future under various climate change scenarios!!

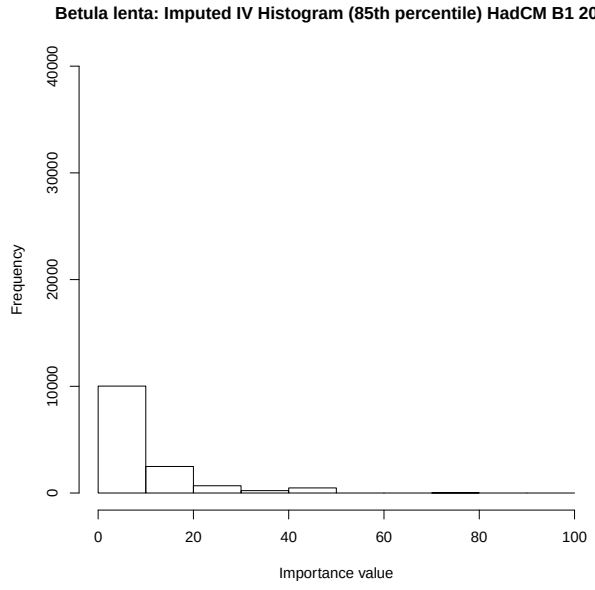
Present/Future ranges under B1 scenario: Betula lenta

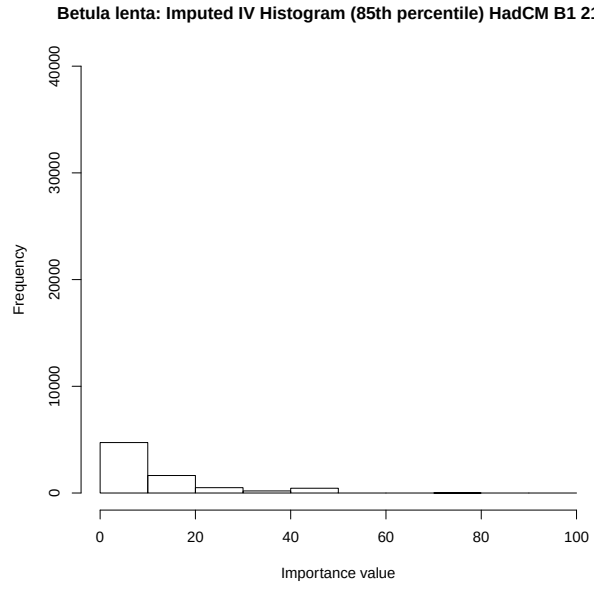










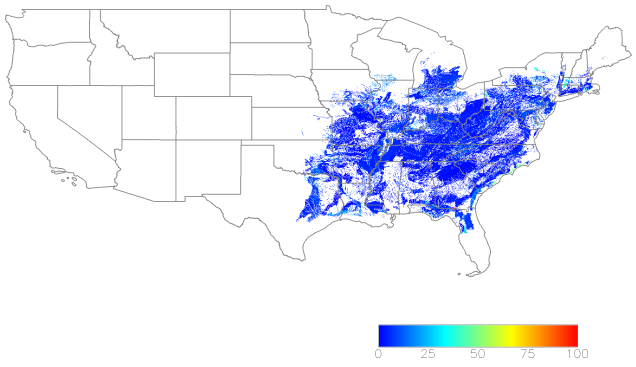


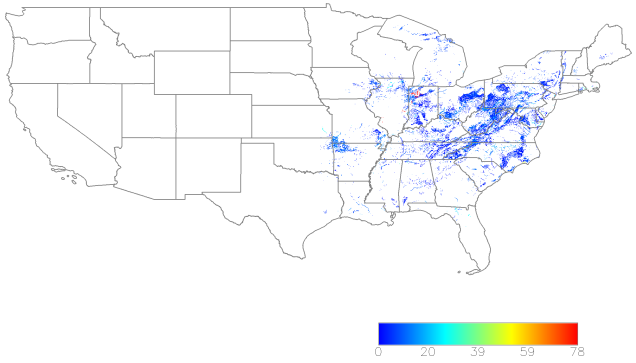
Present period

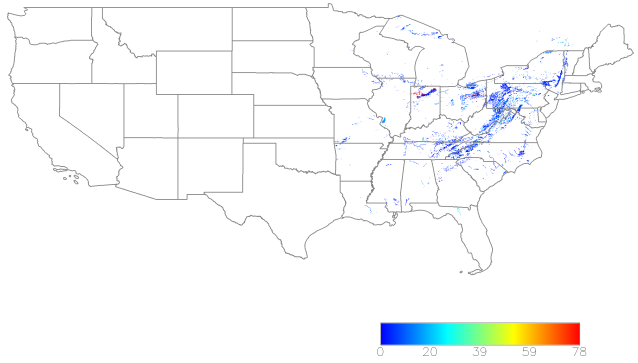
HadCM B1 2050

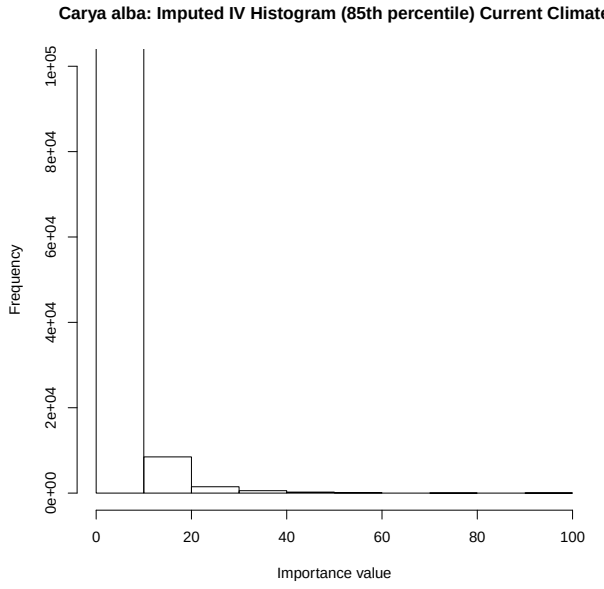
HadCM B1 2100

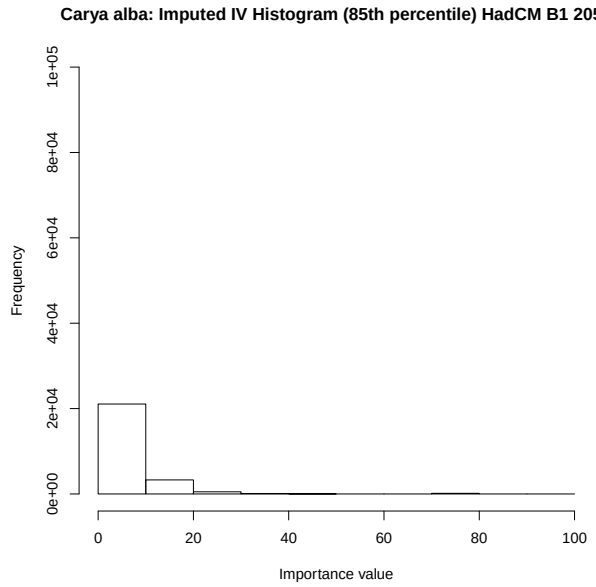
Present/Future ranges under B1 scenario: Carya alba

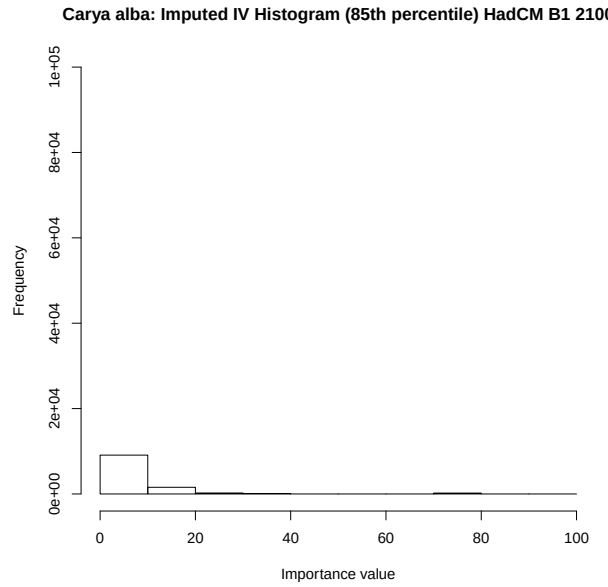










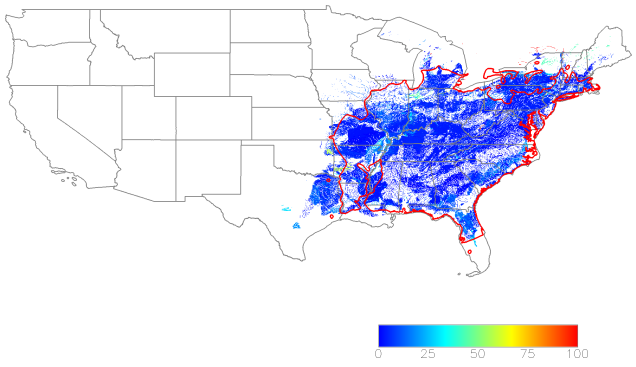


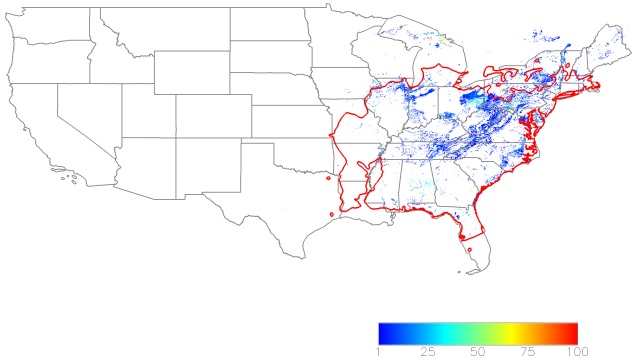
Present period

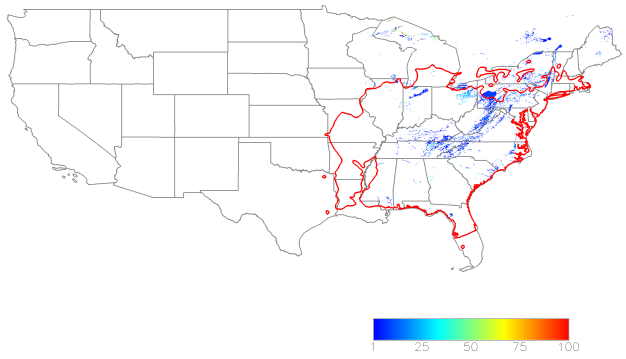
HadCM B1 2050

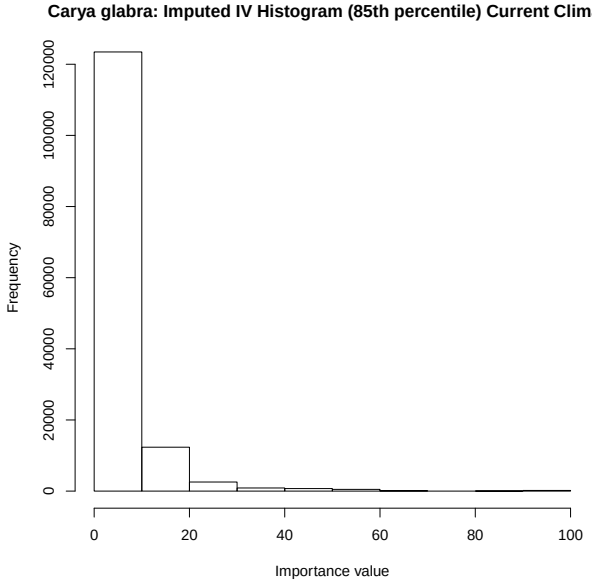
HadCM B1 2100

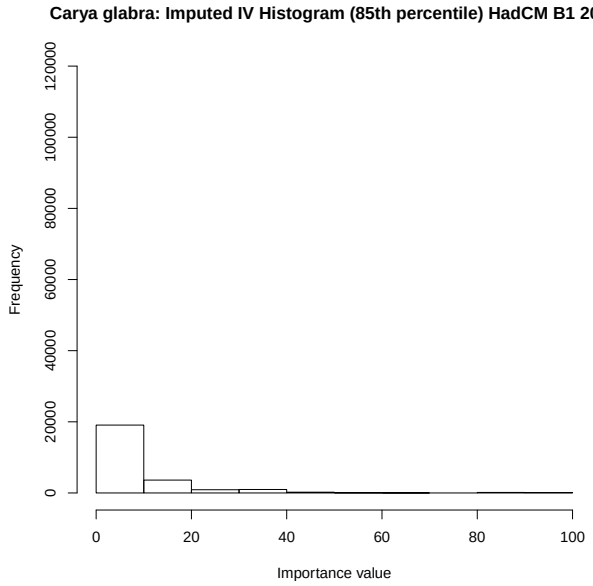
Present/Future ranges under B1 scenario: Carya glabra

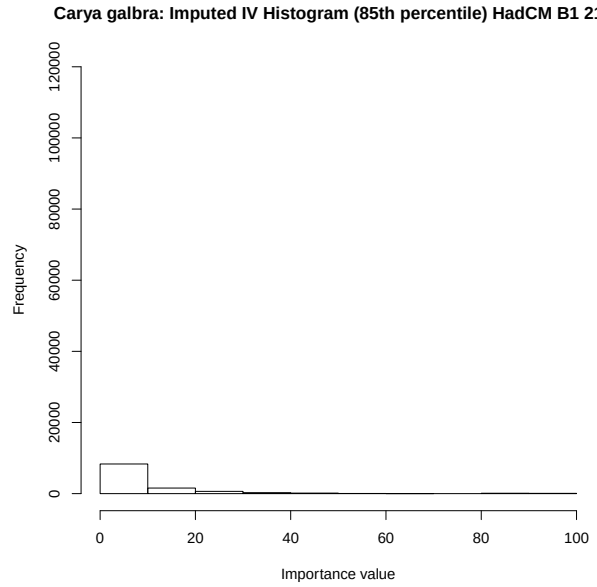












Present period

HadCM B1 2050

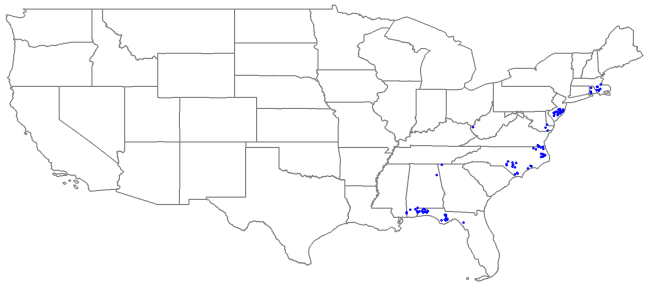
HadCM B1 2100

Observations are often available for common species, but can we map rare species?

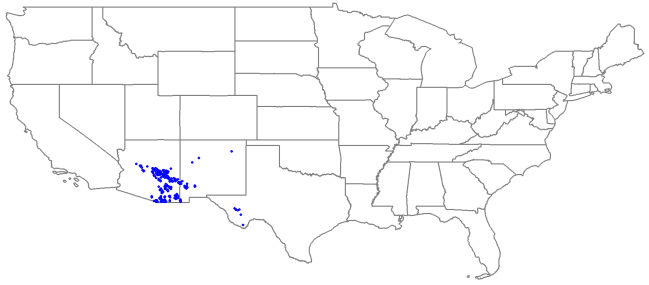
Rare species in CONUS

- *Chamaecyparis thyoides*: Atlantic White Cedar (106)
- *Magnolia macrophylla*: Bigleaf Magnolia (171)
- *Pinus strobiformis*: Southwestern White Pine (164)
- *Quercus emoryi*: Emory Oak (295)
- *Quercus minima*: Dwarf Live oak (141)
- *Sequoia sempervirens*: Coastal Redwood (254)

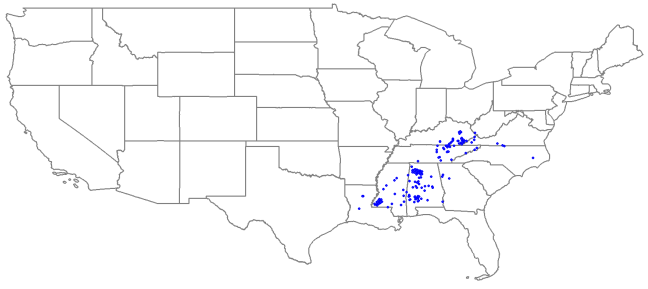
Forest Inventory and Analysis Data



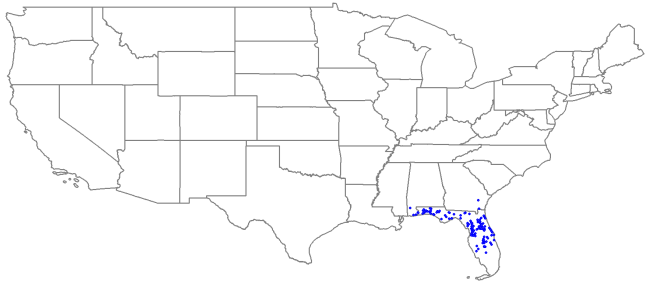
Chamaecyparis thyoides



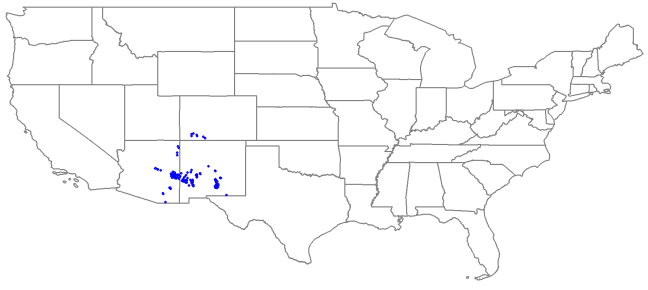
Quercus emoryi



Magnolia macrophylla



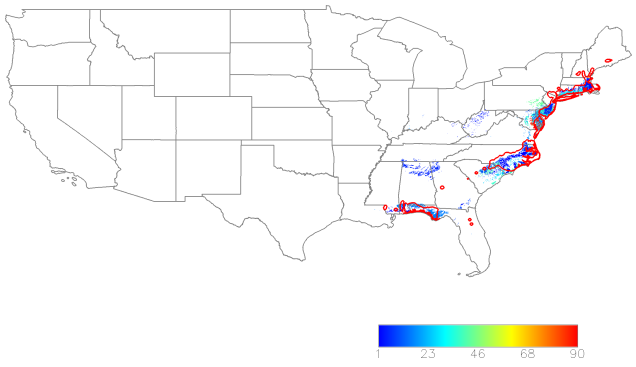
Quercus minima

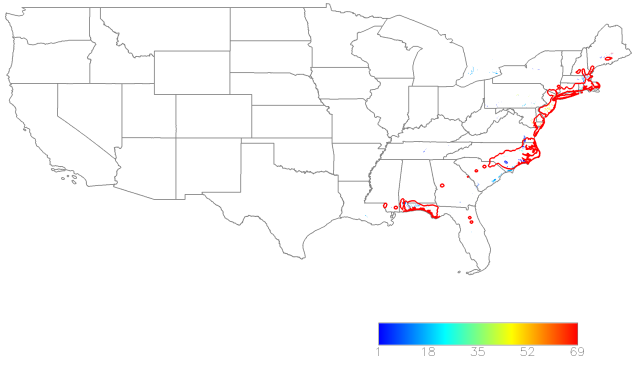


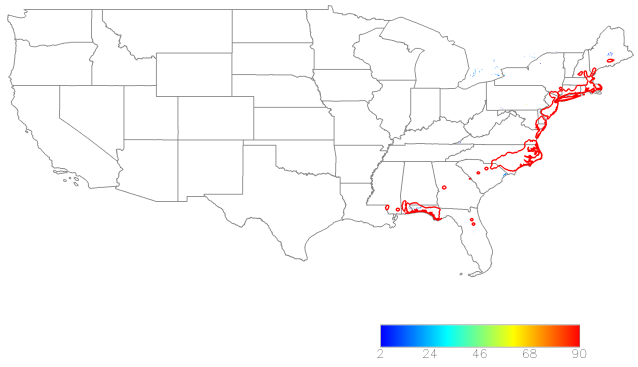
Pinus strobiformis

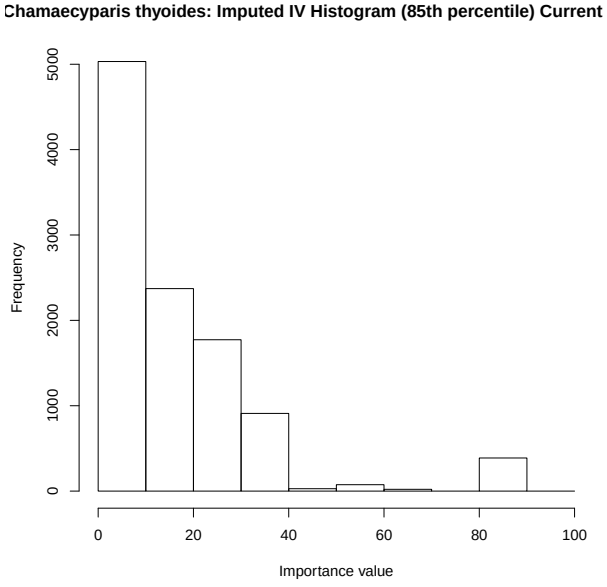
Sequoia sempervirens

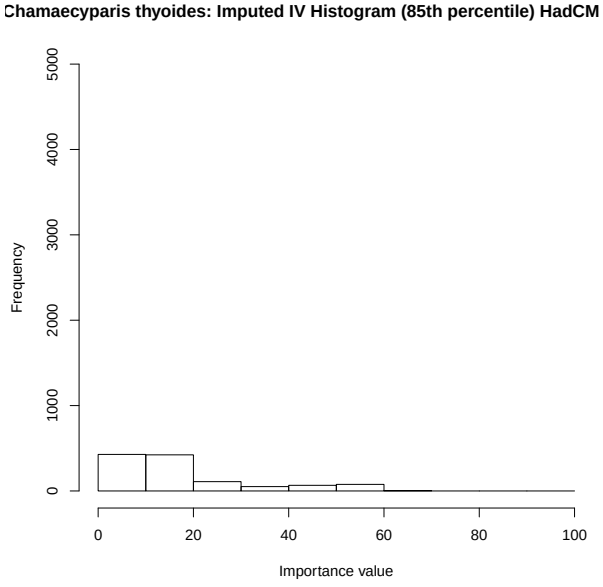
Present/Future ranges under B1 scenario: Chamaecyparis thyoides

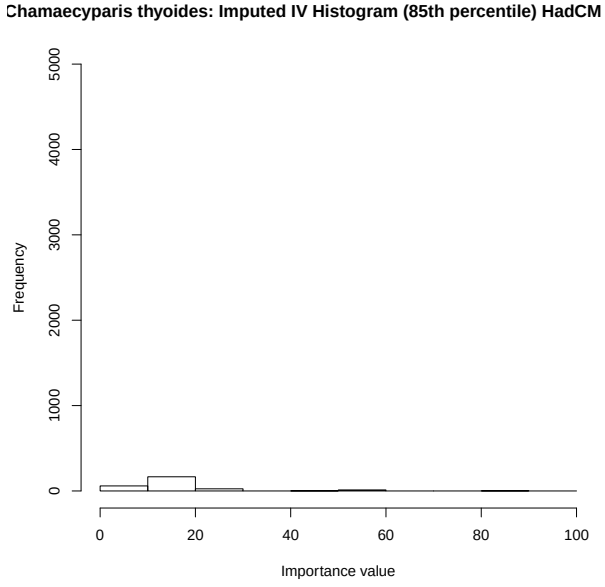










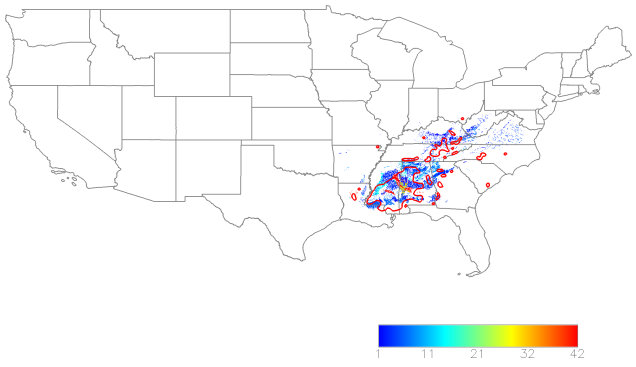


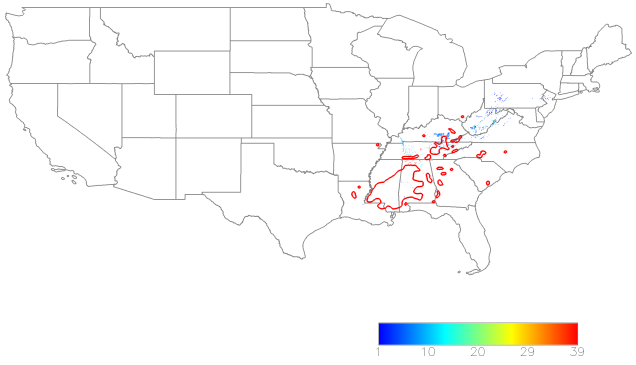
Present period

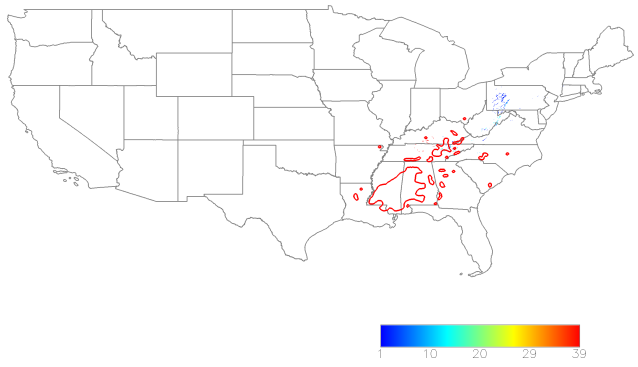
HadCM B1 2050

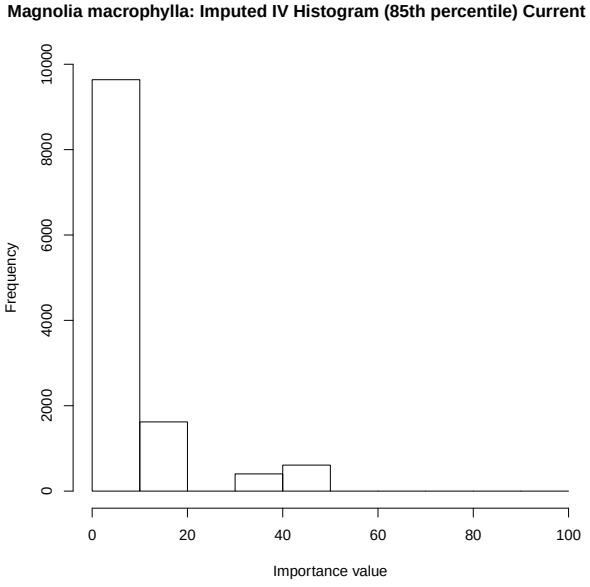
HadCM B1 2100

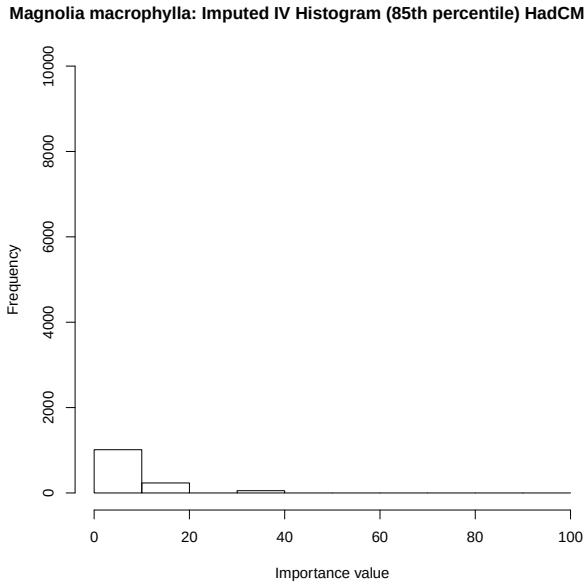
Present/Future ranges under B1 scenario: Magnolia macrophylla

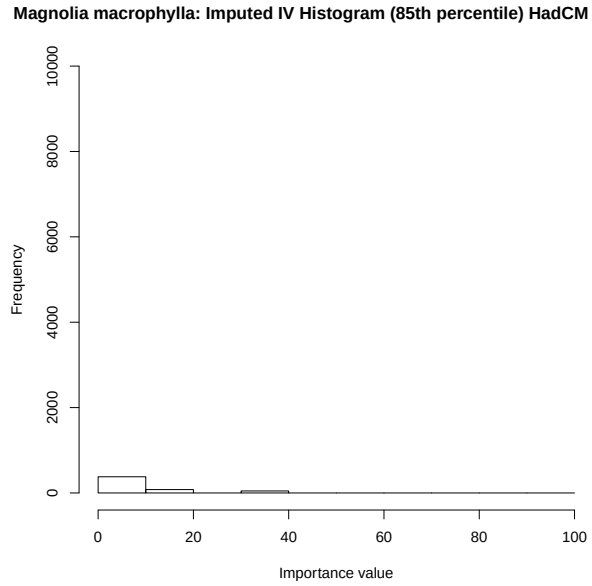












Present period

HadCM B1 2050

HadCM B1 2100

Summary and future directions

Summary:

- Based on preliminary results most of the tree species (except few) are predicted to loose their suitable habitat
- Prediction of future suitable habitat and productivity would aid in forest management planing and conservation
- Results from associative clustering based imputation approach are promising
- Species distribution maps were developed (good agreement with existing range maps)
- Automated statistical approach using sparse measurement
- Generic upscaling tool for scaling point based measurement to broader landscape

Summary and future directions

Ongoing/Future directions:

- Improve the schemes for imputation of productivity
- Include more tree species in the analysis
- Further analysis of changes/shifts in the tree species habitat under climate change scenarios..
- Implications for forest resource management, carbon budget...

Acknowledgments

Thanks for your attention!

jkumar@climatemodeling.org

This research was sponsored by the U.S. Department of Agriculture Forest Service, Eastern Forest Environmental Threat Assessment Center (EFETAC). This research used resources of the National Center for Computational Science at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.