Landscape Characterization and Representativeness Analysis for Tropical Sampling Network Design

Forrest M. Hoffman*†, Damian M. Maddalena†, Jitendra Kumar†, William W. Hargrove‡, and James T. Randerson*

*University of California–Irvine, †Oak Ridge National Laboratory, and ‡USDA Forest Service

BGC Feedbacks SFA Tropical Modeling Update

November 26, 2014

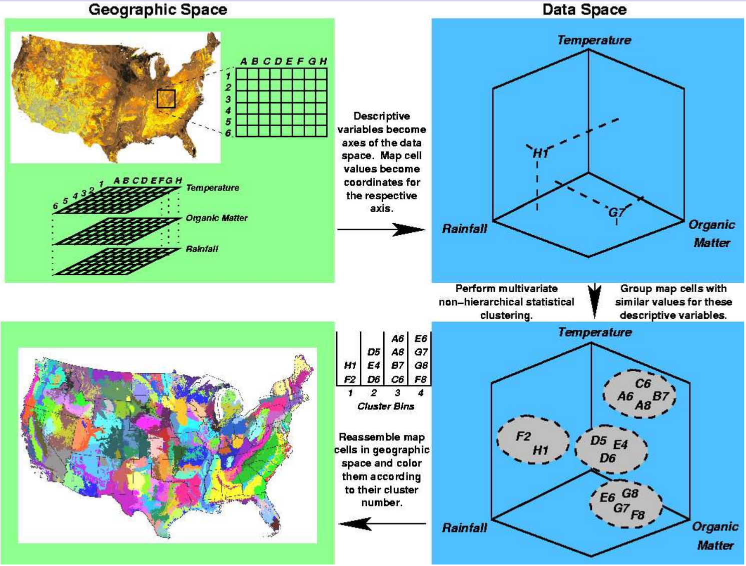
Research Context

- Tropical vegetation is poorly represented in current Earth system models (ESMs).
- Spatial heterogeneity of highly diverse tropical forests is absent from ESMs.
- Understanding potentially vulnerable tropical systems is important under a changing climate (*Research Priorities for Tropical Ecosystems Under Climate Change*, DOE, 2012).
- Logistics and resource constraints limit where and when measurements can be made.
- NGEE Tropics Phase 1 will require upscaling methods and quantitative quality assessment of currently available data.

Methods

- Classify ecoregions using quantitative cluster analysis.
- Label unsupervised classification with expert-derived ecoregion names.
- Quantify representativeness of single and combined network coverage.

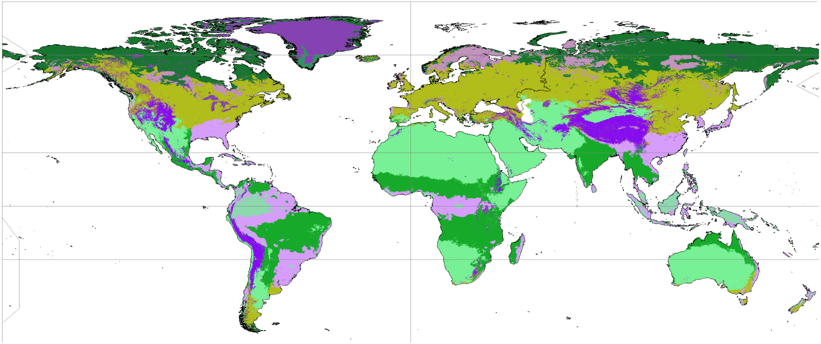
Multivariate Spatiotemporal Clustering (MSTC)



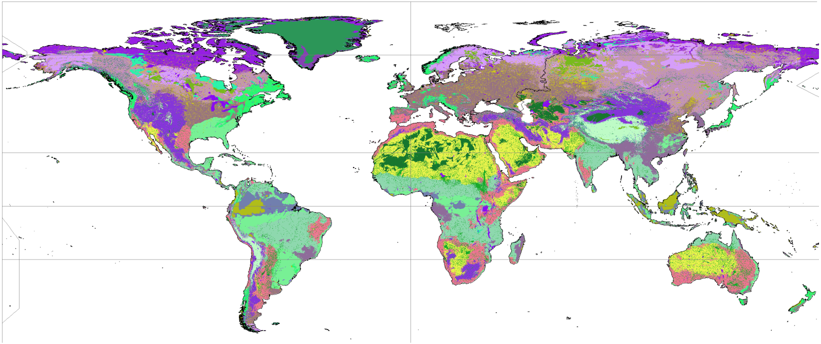
Data: 17 Variables

Variable Description	Units
Bioclimatic Variables	
Precipitation during the hottest quarter	mm
Precipitation during the coldest quarter	mm
Precipitation during the driest quarter	mm
Precipitation during the wettest quarter	mm
Ratio of precipitation to potential evapotranspiration	
Temperature during the coldest quarter	°C
Temperature during the hottest quarter	°C
Day/night diurnal temperature difference	°C
Sum of monthly T_{avg} where $T_{avg} \geq 5°C$	°C
Integer number of consecutive months where $T_{avg} \geq 5°C$	
Edaphic Variables	
Available water holding capacity of soil	
Bulk density of soil	g/cm^3
Carbon content of soil	g/cm^2
Nitrogen content of soil	g/cm^2
Topographic Variables	
Compound topographic index (relative wetness)	
Solar interception	(kW/m^2)
Elevation	m

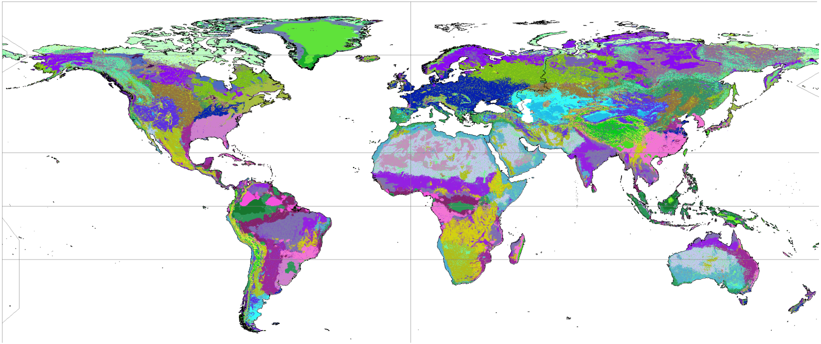
Clustered Regions, k=10



Clustered Regions, k=25



Clustered Regions, k=50

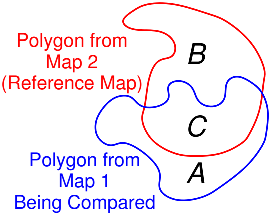


Informed Labeling of Quantitatively Derived Regions

- Clustering is an unsupervised classification technique...no descriptive labels
- Informed labeling: Consult the experts
- Employ Mapcurves to select the best ecoregion labels from ecoregionalizations drawn by human experts.
- Consider a library of ecoregion and land cover maps, and choose the label with the highest GOF score for every ecoregion.

Mapcurves: A Method for Comparing Categorical Maps

- Hargrove et al. (2006) developed a method for quantitatively comparing categorical maps that is
  - independent of differences in resolution,
  - independent of the number of categories in maps, and
  - independent of the directionality of comparison.



Goodness of Fit (GOF) is a unitless measure of spatial overlap between map categories:

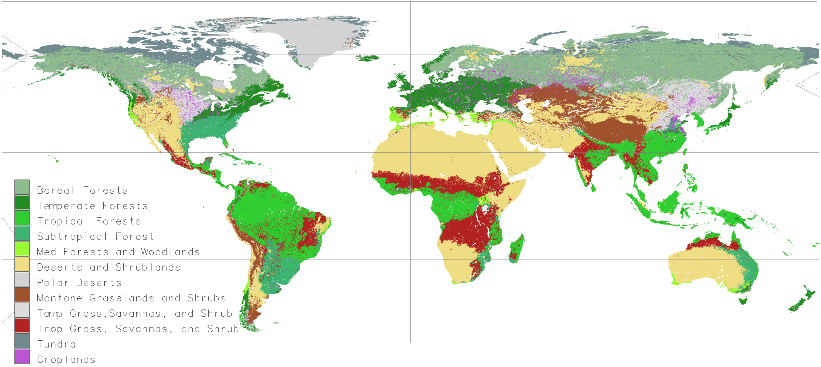
$$\mathrm{GOF} = \sum_{\text{polygons}} \frac{C}{B+C} \times \frac{C}{A+C}$$

- GOF provides “credit” for the area of overlap, but also “debit” for the area of non-overlap.
- Mapcurves comparisons allow us to reclassify any map in terms of any other map (*i.e.*, color Map 2 like Map 1).
- A greyscale GOF map shows the degree of correspondence between two maps based on the highest GOF score.

Expert Maps used with Mapcurves Algorithm

Map	# Categories
Foley Land Cover	14
Holdridge Life Zones	25
IGBP Land Cover	16
European Space Agency Global Land Cover Map	23
Olson's Ecoregions of the World	14

Clustered Regions, k=50, Manually Reclassified by Combining Like Classes



Representativeness Analysis of Points and Networks

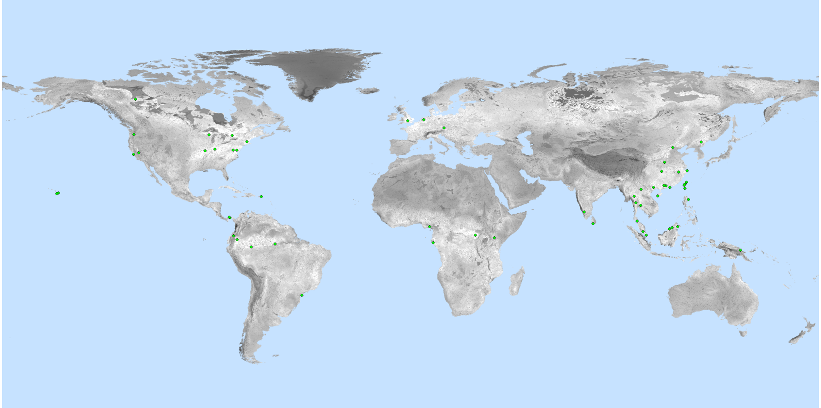
1. Representativeness analysis compares a single point to all other points in data space.
2. Euclidean distance in data space is mapped as a dissimilarity score in XY space, where darker colors indicate high degrees of dissimilarity
3. A single map is created from all maps (sites) in a set by selecting the minimum values for each grid cell from the collection of maps (network of sites)

Monitoring Networks Used in Rep Analysis

1. RAINFOR – 368
2. CTFS-ForestGEO – 59
3. Fluxnet – 240

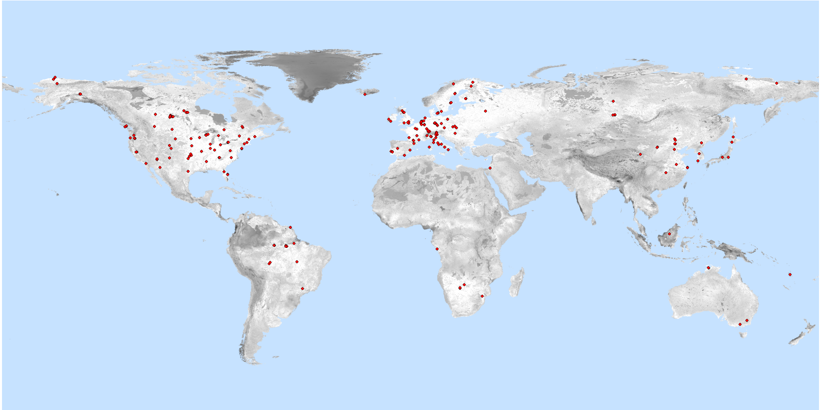
Single Point Representativeness: CTFS-ForestGEO, Mpala, Kenya

Global Representativeness of CTFS-ForestGEO Network



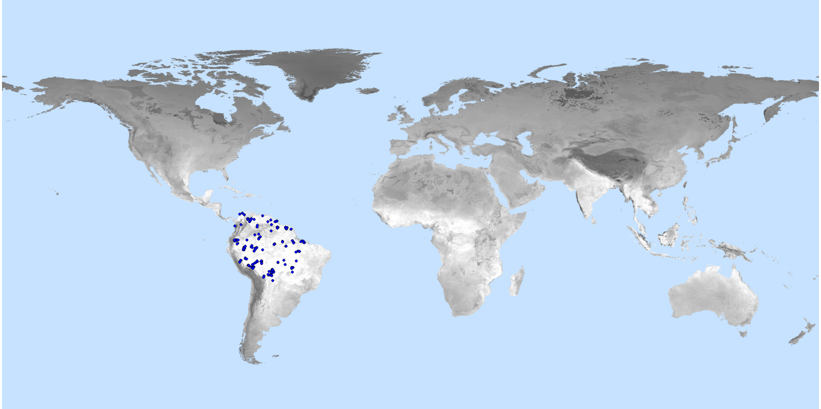
● - CTFS sites

Global Representativeness of Fluxnet Network



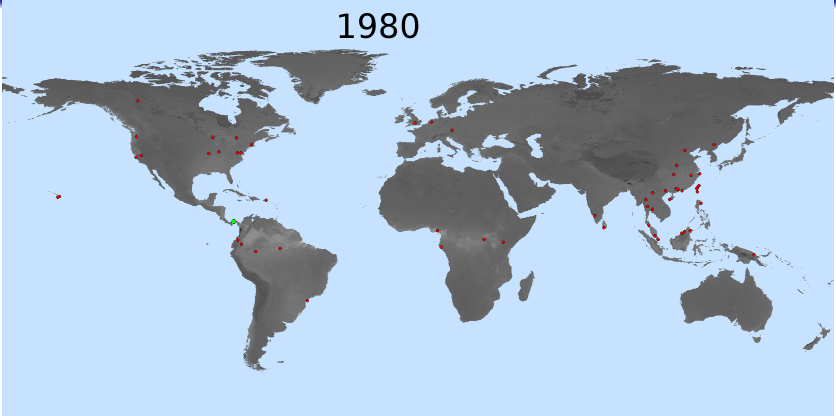
● - Fluxnet sites

Global Representativeness of RAINFOR Network



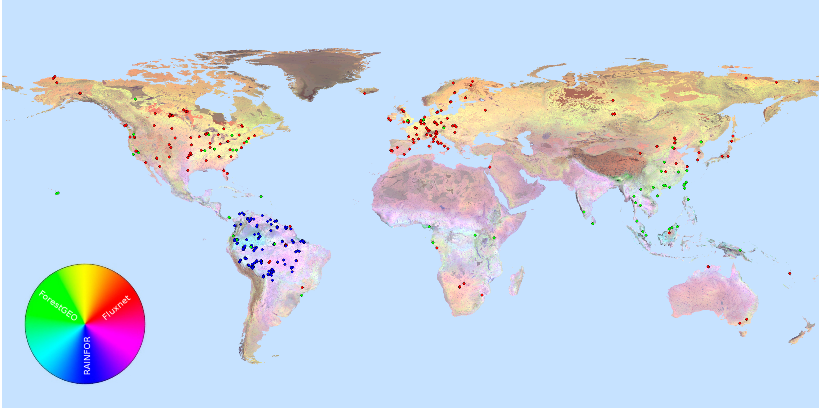
● - RAINFOR sites

CTFS-ForestGEO Network through time



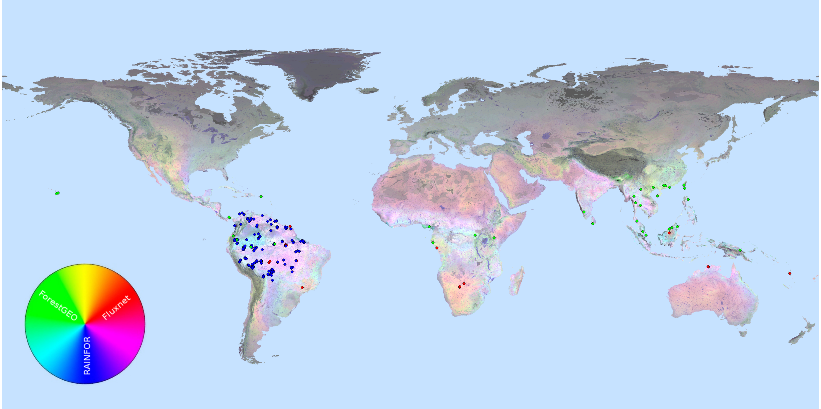
ForestGEO network coverage through time. Analysis allows us to understand how the improved availability of observations help constrain the model and highlights data gaps in space and time.

Combined Representativeness of Fluxnet, CTFS-ForestGEO, and RAINFOR



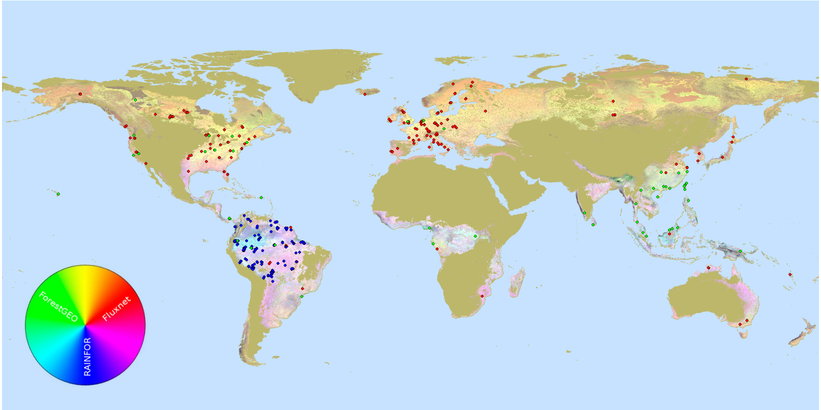
● - Fluxnet sites; ● - CTFS sites; ● - RAINFOR sites

Combined Representativeness of Fluxnet, CTFS-ForestGEO, and RAINFOR: Tropical Sites Only



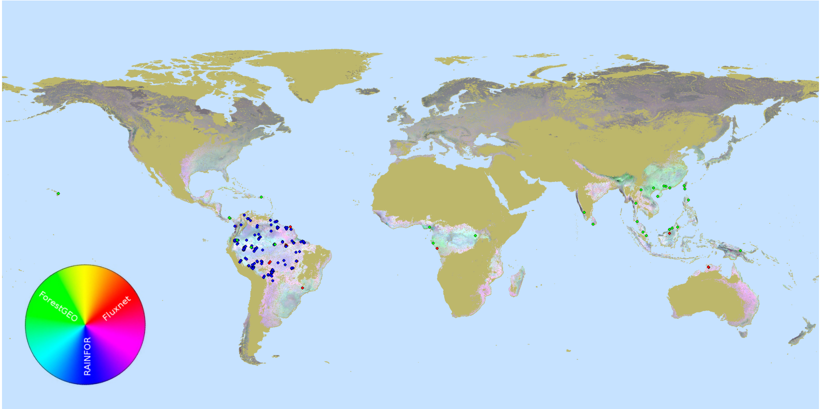
● - Fluxnet sites; ● - CTFS sites; ● - RAINFOR sites

Global Combined Representativeness of Fluxnet, CTFS-ForestGEO, and RAINFOR: Forested Sites Only



● - Fluxnet sit es; ● - CTFS sites; ● - RAINFOR sites

Combined Representativeness of Fluxnet, CTFS-ForestGEO, and RAINFOR: Tropical Forest Sites Only



● - Fluxnet sites; ● - CTFS sites; ● - RAINFOR sites

Conclusions

1. Landscape classification with MSTC, Mapcurves, and representative analysis are a suite of tools suitable for quantitative assessment of data from monitoring networks through space and time
2. Poorly covered regions revealed by using these tools are potential sites for future network sites
3. These methods are part of a larger effort of data upscaling and quantifying uncertainty in model data assimilation

References

William W Hargrove, Forrest M Hoffman, and Paul F Hessburg. Mapcurves: a quantitative method for comparing categorical maps. *Journal of Geographical Systems*, 8(2):187–208, 2006.