Landscape Characterization and Representativeness Analysis for Understanding Sampling Network Coverage

Damian M. Maddalena, Forrest M. Hoffman, Jitendra Kumar, William Hargrove

Introduction

- 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 (DOE, 2012).
- Logistics and resource constraints limit where and when measurements can be made.
- Tropical forest research will require upscaling methods and quantitative quality assessment of currently available data.

Methods

1. Classify ecoregions using Multivariate Spatiotemporal Clustering (MSTC)
2. Label unsupervised classification with ecoregion type names using Mapcurves
3. Quantify representativeness of single and combined network coverage using distance in a hyper-volume data space

(See Hoffman et al. (2013).)

Ecoregion Delineation

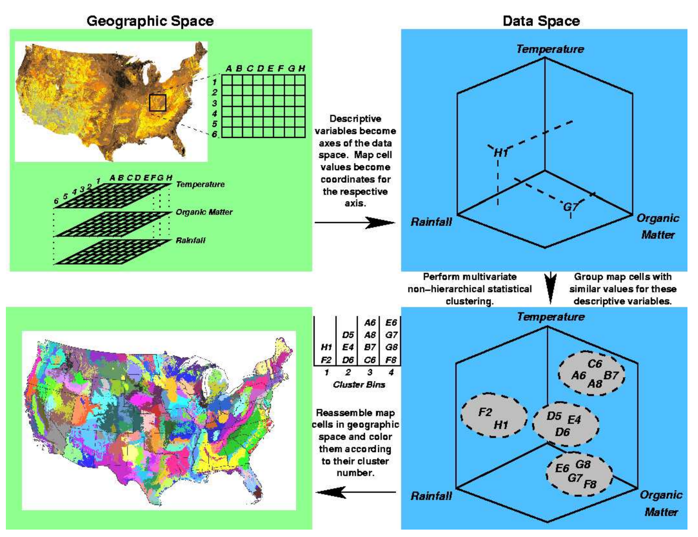


Figure 1: *Multivariate Spatiotemporal Clustering (MSTC)*

Table 1: *Variables used in MSTC for ecoregion delineation and representativeness analysis. These data are raster grids with a 4km² resolution.*

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 $T_{\text{avg}} \geq 5°$C	°C
Integer number of consecutive months where $T_{\text{avg}} \geq 5°$C	
Edaphic Variables	
Available water holding capacity of soil	
Bulk density of soil	g/cm³
Carbon content of soil	g/cm²
Nitrogen content of soil	g/cm²
Topographic Variables	
Compound topographic index (relative wetness)	
Solar interception	(kW/m²)
Elevation	m

Mapcurves

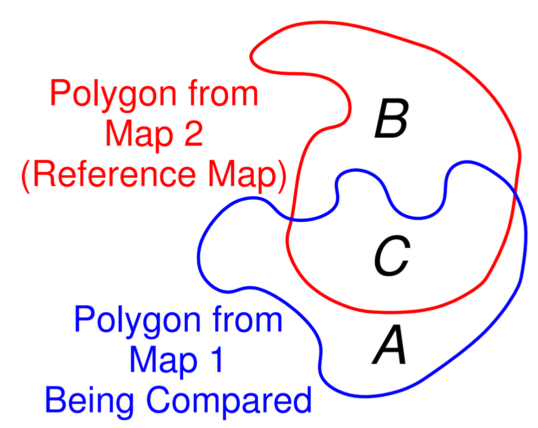


Figure 2: *Mapcurves compares the agreement and disagreement of categorical maps in a way that is independent of differences in resolution, the number of categories, or the direction of comparison (Hargrove et al., 2006).*

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

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

Table 2: *Expert maps used with the mapcurves algorithm to assign labels to the MSTC results.*

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

Representativeness Analysis of Points and Networks

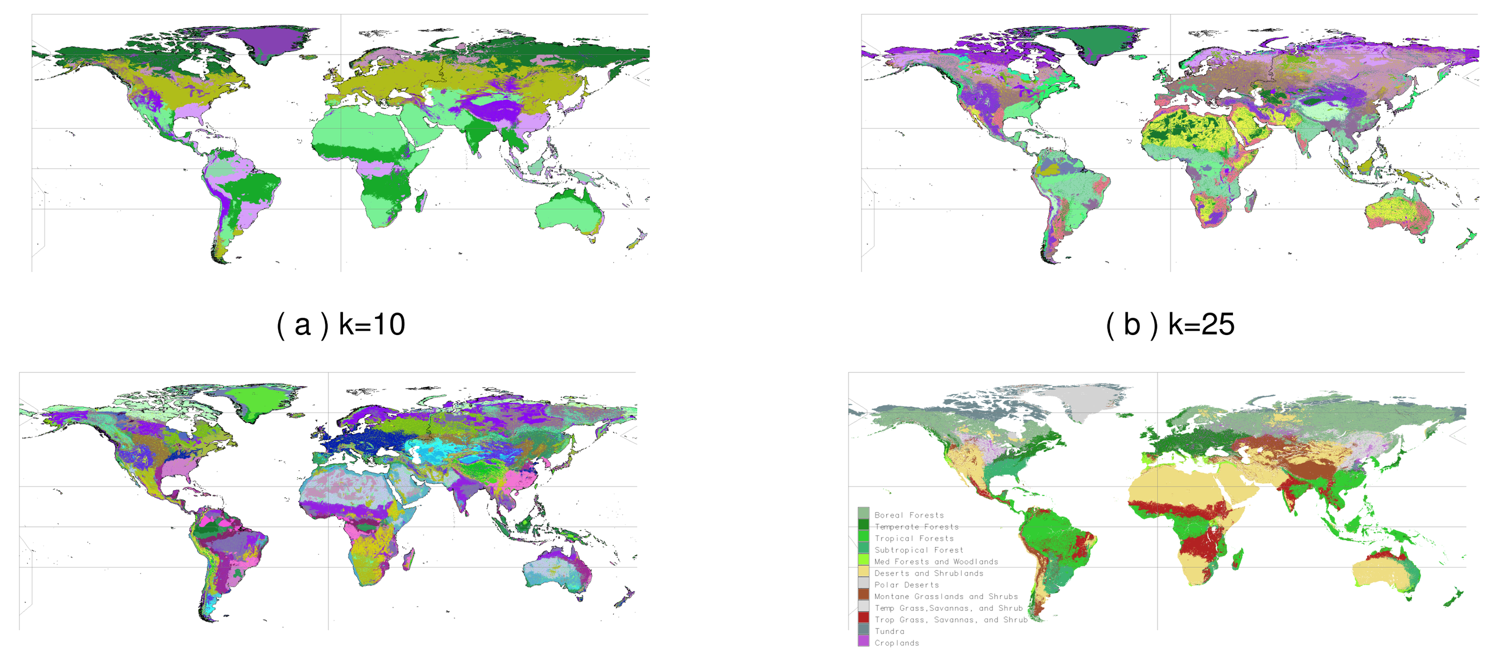
1. Representativeness analysis compares a single point to all other points in data space.
2. Euclidean distance in data space is mapped as a dissimilarity score in geographic space, where darker colors indicate high degrees of dissimilarity.
3. A single map is created from all maps (sites) in a set by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

Table 3: *Total number of sites for each network used in the representativeness analysis.*

Network	Number of Sites
RAINFOR	368
CTFS-ForestGEO	59
Fluxnet	240

Results

Ecoregion Classification

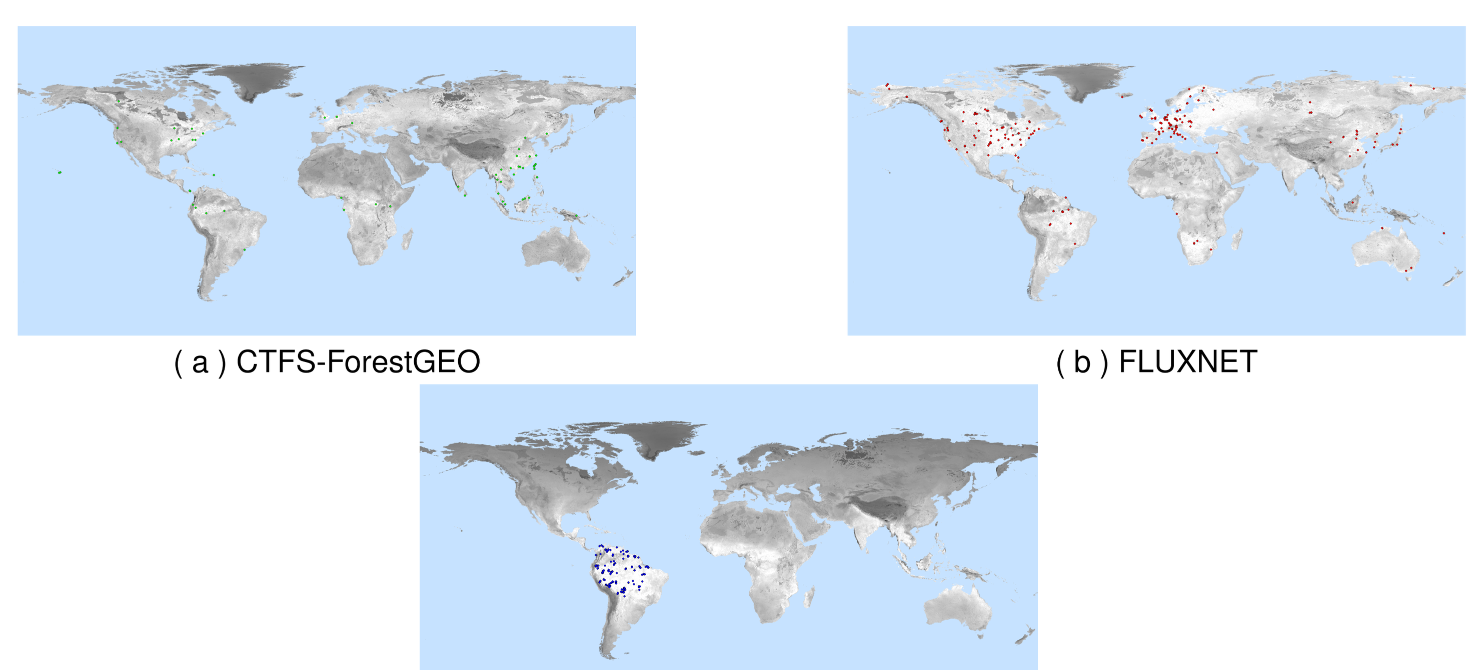


(a) k=10 (b) k=25 (c) k=50 (d) k=50 reclass

Figure 3: *The MSTC algorithm was used to group the 17 observation vectors (Table 1) into regions with equal variance across all clusters. Clustering was repeated until 0.05% of all observations changed cluster membership between iterations (Kumar et al., 2011). The results are categorical maps of k regions. The derived ecoregions were then identified for type using the mapcurves algorithm developed by Hargrove et al. (2006) with a suite of expert ecoregion maps (Table 2). The k=50 map was manually reclassified to group similarly labeled regions (d).*

Representativeness Analysis for Individual Monitoring Networks

Figure 4: *Single Point Representativeness: CTFS-ForestGEO, Mpala, Kenya. Representativeness of the entire globe with respect to an individual sampling point quantified in data space. Euclidean distance in data space is mapped as a dissimilarity score in geographic space, where darker colors indicate high degrees of dissimilarity.*



(a) CTFS-ForestGEO (b) FLUXNET (c) RAINFOR

Figure 5: *Network Representativeness: Total representativeness for the CTFS-ForestGEO, Fluxnet, and RAINFOR networks. Each network representativeness map was created from all single point maps (sites) in a set by selecting the minimum values for each grid cell from the collection of maps (network of sites). Table 3 lists the number of points in each sampling network. Darker colors indicate high degrees of dissimilarity.*

Representativeness Analysis for Combined Monitoring Networks

Individual network representativeness maps (Figure 5) were combined as an RGB map where color combinations represent combinations of network coverage.

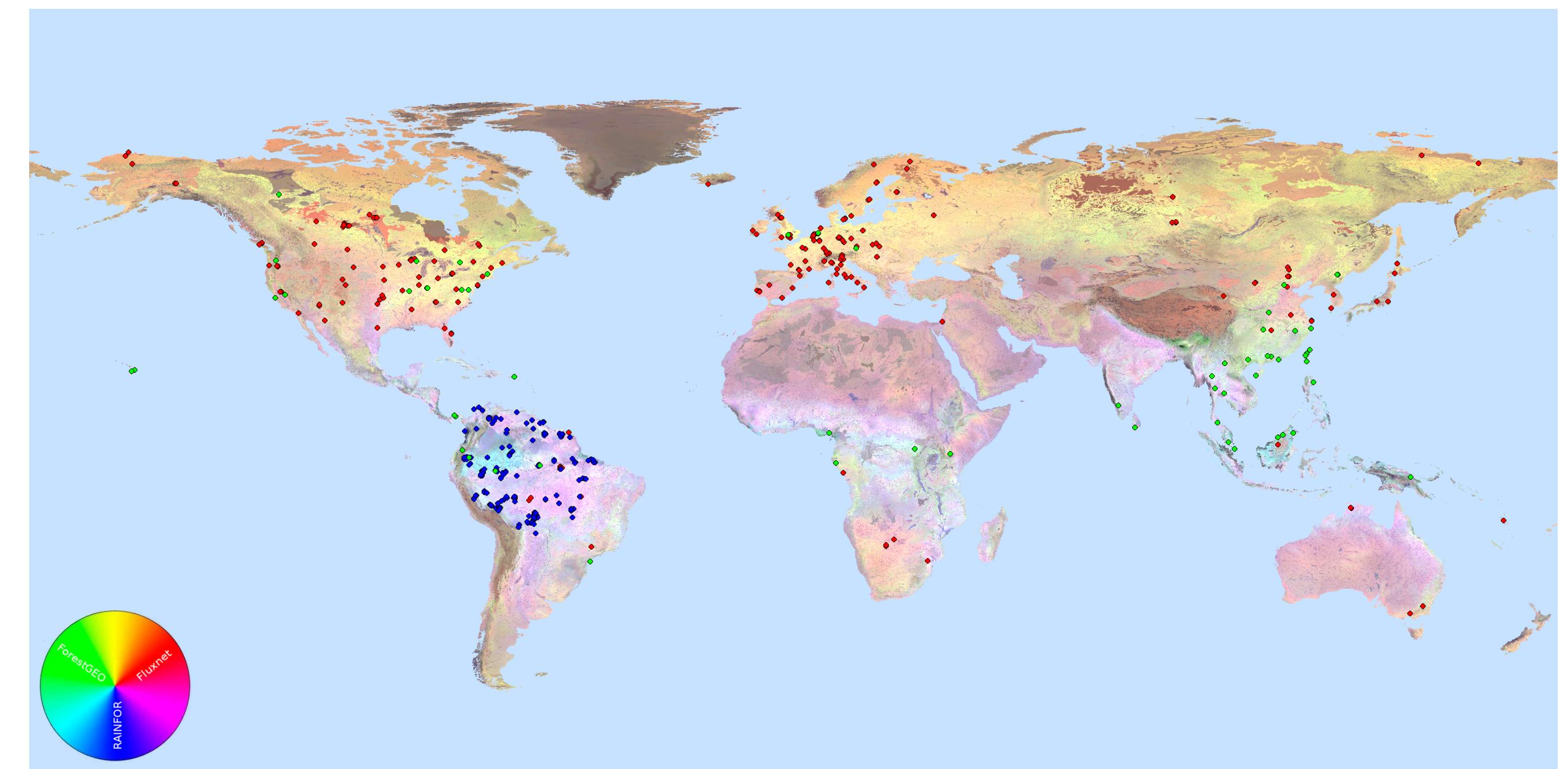


Figure 6: *Combined Representativeness of Fluxnet (●), CTFS-ForestGEO (●), and RAINFOR (●). Color combinations of RGB represent the combined coverage of the three networks where white areas are combinations of all three and dark areas lack coverage of any network.*

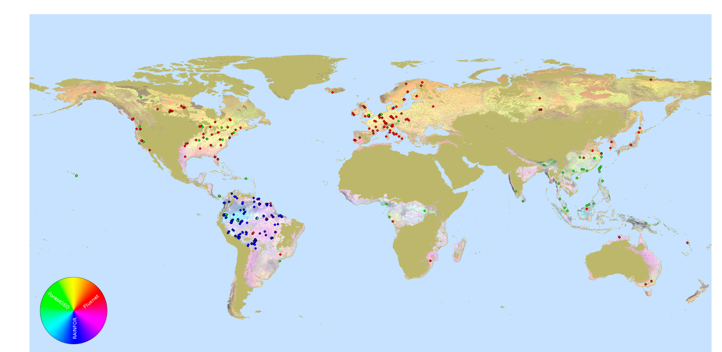


Figure 7: *Combined Representativeness of Fluxnet (●), CTFS-ForestGEO (●), and RAINFOR (●) for forested regions globally. Forested regions were defined using MSTC (Figure 3(d)). Color combinations of RGB represent the combined coverage of the three networks where white areas are combinations of all three and dark areas lack coverage of any network.*

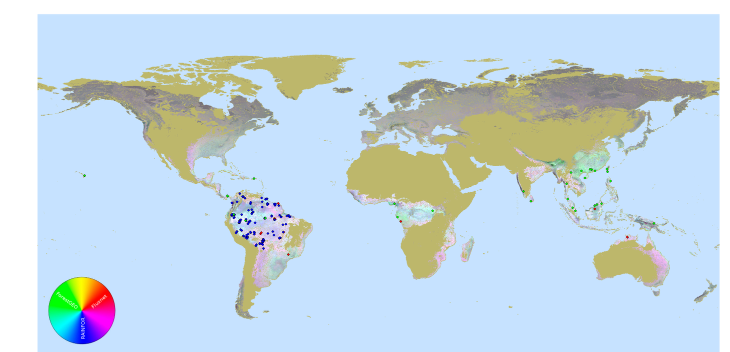


Figure 8: *Combined Representativeness of Fluxnet (●), CTFS-ForestGEO (●), and RAINFOR (●) for tropical forested regions globally. Forested regions were defined using MSTC (Figure 3(d)). Color combinations of RGB represent the combined coverage of the three networks where white areas are combinations of all three and dark areas lack coverage of any network.*

Conclusions

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

References

DOE. Research priorities for tropical ecosystems under climate change. Technical report, Department of Energy, 2012.

Forrest M Hoffman, Jitendra Kumar, Richard T Mills, and William W Hargrove. Representativeness-based sampling network design for the State of Alaska. *Landscape Ecology*, 28(8):1567–1586, 2013.

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.

Jitendra Kumar, Richard T Mills, Forrest M Hoffman, and William W Hargrove. Parallel k-means clustering for quantitative ecoregion delineation using large data sets. *Procedia Computer Science*, 4:1602–1611, 2011.

Contact Info

Damian M. Maddalena
email: damian@climatemodeling.org