Characterizing Forest Representativeness for Optimizing Sampling Network Coverage

Forrest M. Hoffman[1,2], Jitendra Kumar[1,2],
James T. Randerson[3], and Stuart J. Davies[4,5]

[1]Oak Ridge National Laboratory, [2]University of Tennessee Knoxville,
[3]University of California Irvine, [4]Smithsonian Tropical Research Institute, and
[5]National Museum of Natural History

August 11, 2016

Ecological Society of America (ESA) Annual Meeting

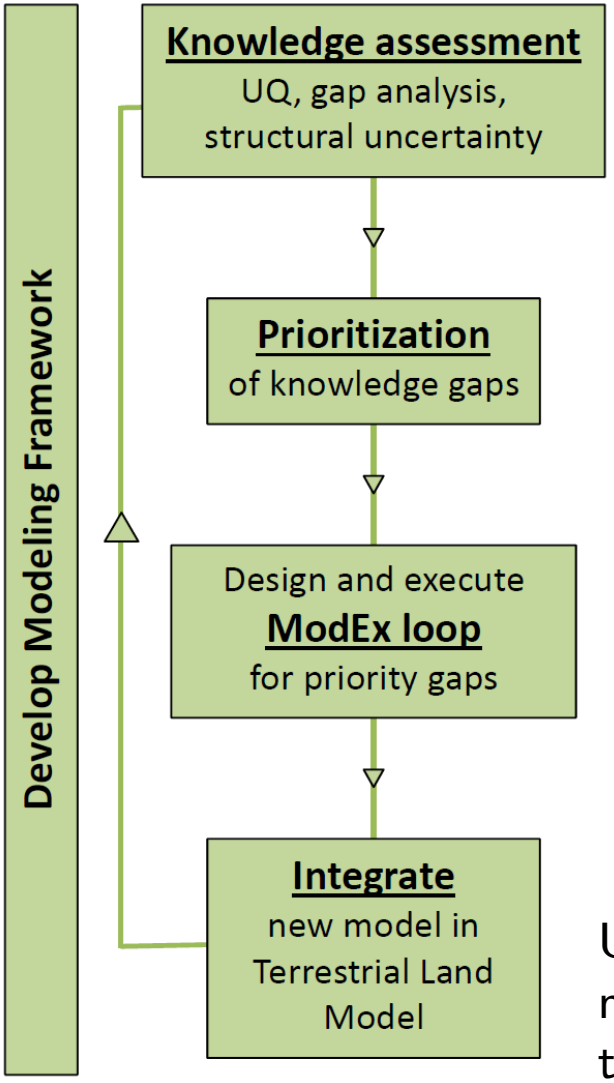
Fort Lauderdale, Florida, USA
August 7–12, 2016

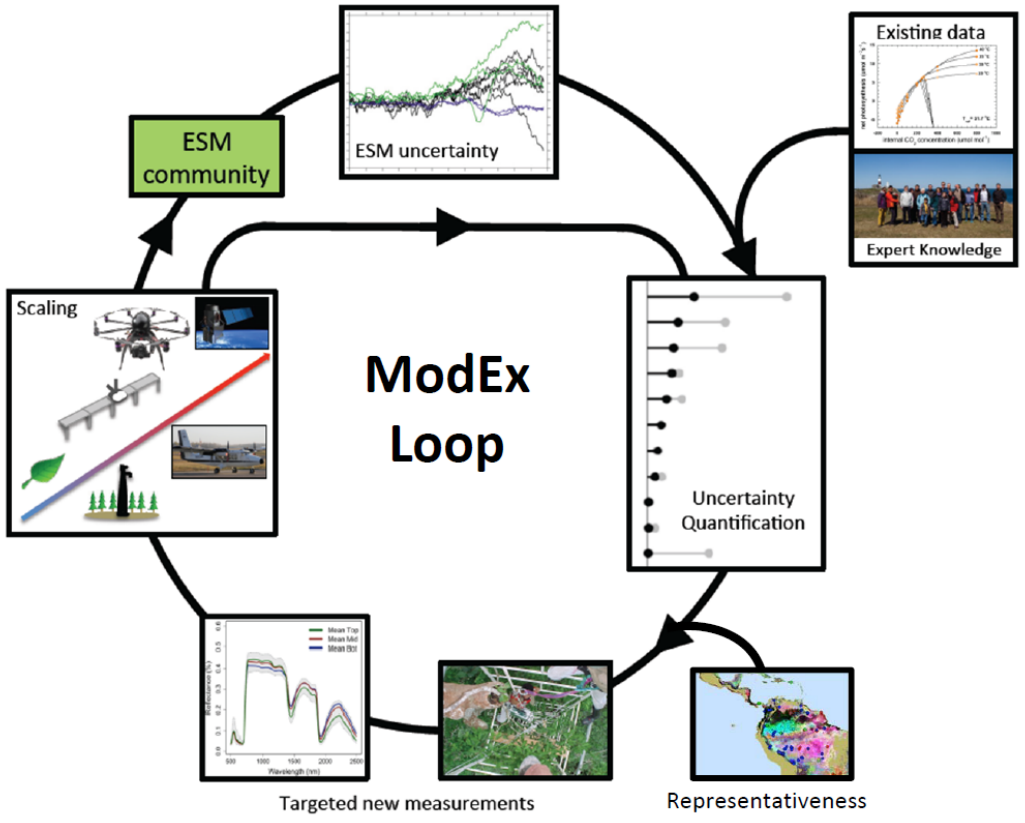
Next Generation Ecosystem Experiments (NGEE Tropics)

- **Overarching Goal:** Determine if tropical forests will continue to act as a large net carbon sink throughout the 21st century.
- **Grand Deliverable:** A representative, process-rich tropical forest ecosystem model, extending from bedrock to the top of the vegetative canopy, in which the evolution and feedbacks of tropical ecosystems in a changing climate can be modeled at the scale/resolution of a next generation Earth System Model grid cell.
- **Key Science Questions:**
  - How do tropical forest ecosystems respond to changing temperature, precipitation, and atmospheric CO_2 concentration?
  - How do disturbance and landuse change in tropical forests affect carbon, water and energy fluxes?
  - How will the response of tropical forests to climate change be modulated by spatial and temporal heterogeneity in belowground processes?

BERKELEY LAB
BROOKHAVEN NATIONAL LABORATORY
Los Alamos NATIONAL LABORATORY EST. 1943
OAK RIDGE National Laboratory
Pacific Northwest NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY | Office of Science
Smithsonian Tropical Research Institute
FOREST SERVICE US DEPARTMENT OF AGRICULTURE
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Incorporating Models and Experiments to Advance Knowledge



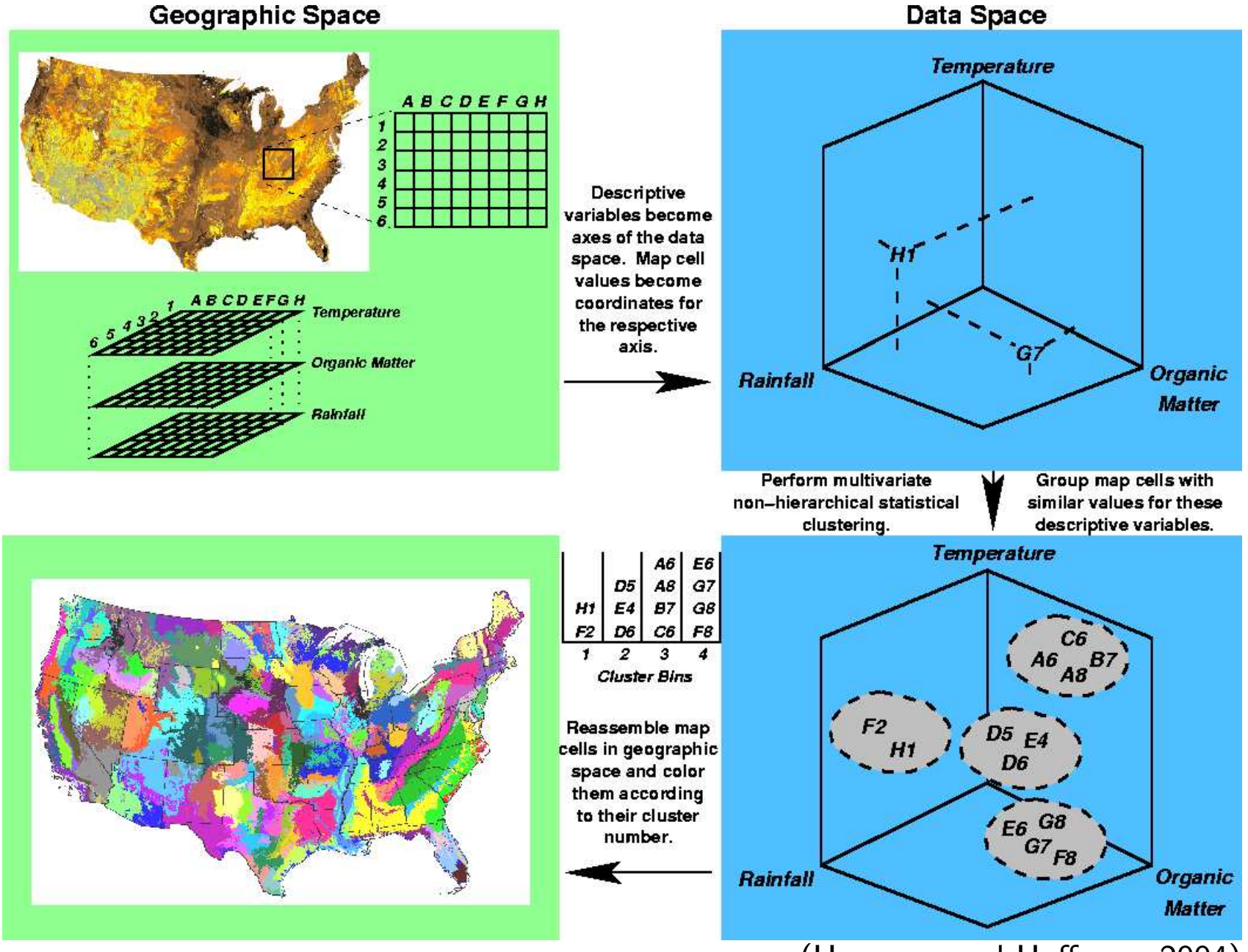


Understanding the representativeness of measurements helps to minimize sampling biases that may influence interpretation of observations.

Quantitative Sampling Network Design

- Resource and logistical constraints limit the frequency and extent of observations, necessitating the development of a systematic sampling strategy that objectively represents environmental variability at desired spatial scales.
- Required is a methodology that provides a quantitative framework for informing site selection and determining the representativeness of measurements.
- Multivariate spatiotemporal clustering (MSTC) was applied at the landscape scale (4 km $\times$ 4 km) globally to demonstrate its utility for representativeness and scaling.
- Method recently used to quantify representativeness of candidate sampling sites for the State of Alaska (Hoffman et al., 2013).
- An extension of the method applied by Hargrove and Hoffman for design of National Science Foundation's (NSF's) National Ecological Observatory Network (NEON) domains (Schimel et al., 2007; Keller et al., 2008).

Multivariate Spatiotemporal Clustering (MSTC)



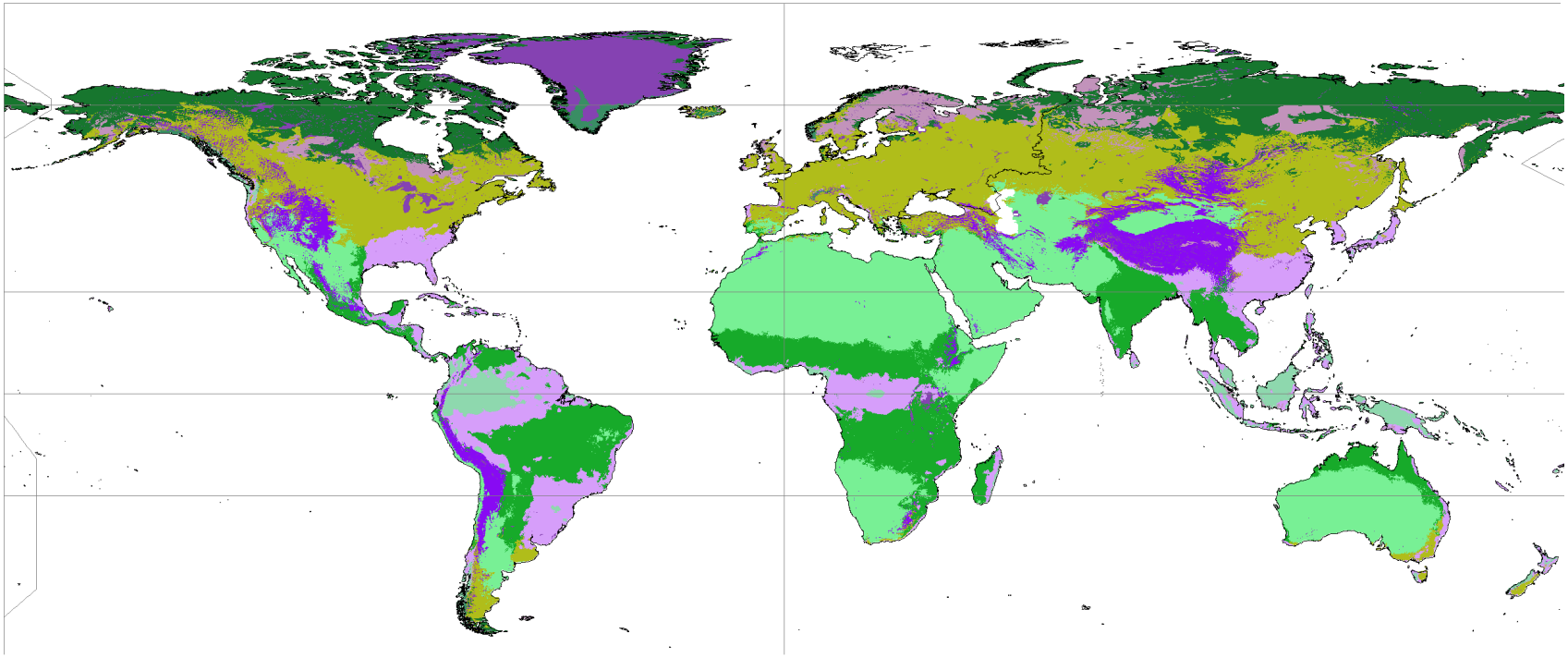
(Hargrove and Hoffman, 2004)

17 Data Layers

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	unitless
Temperature during the coldest quarter	$^\circ$C
Temperature during the hottest quarter	$^\circ$C
Day/night diurnal temperature difference	$^\circ$C
Sum of monthly T_{avg} where $T_{avg} \geq 5^\circ$C	$^\circ$C
Integer number of consecutive months where $T_{avg} \geq 5^\circ$C	unitless
Edaphic Variables	
Available water holding capacity of soil	unitless
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)	unitless
Solar interception	kW/m^2
Elevation	m

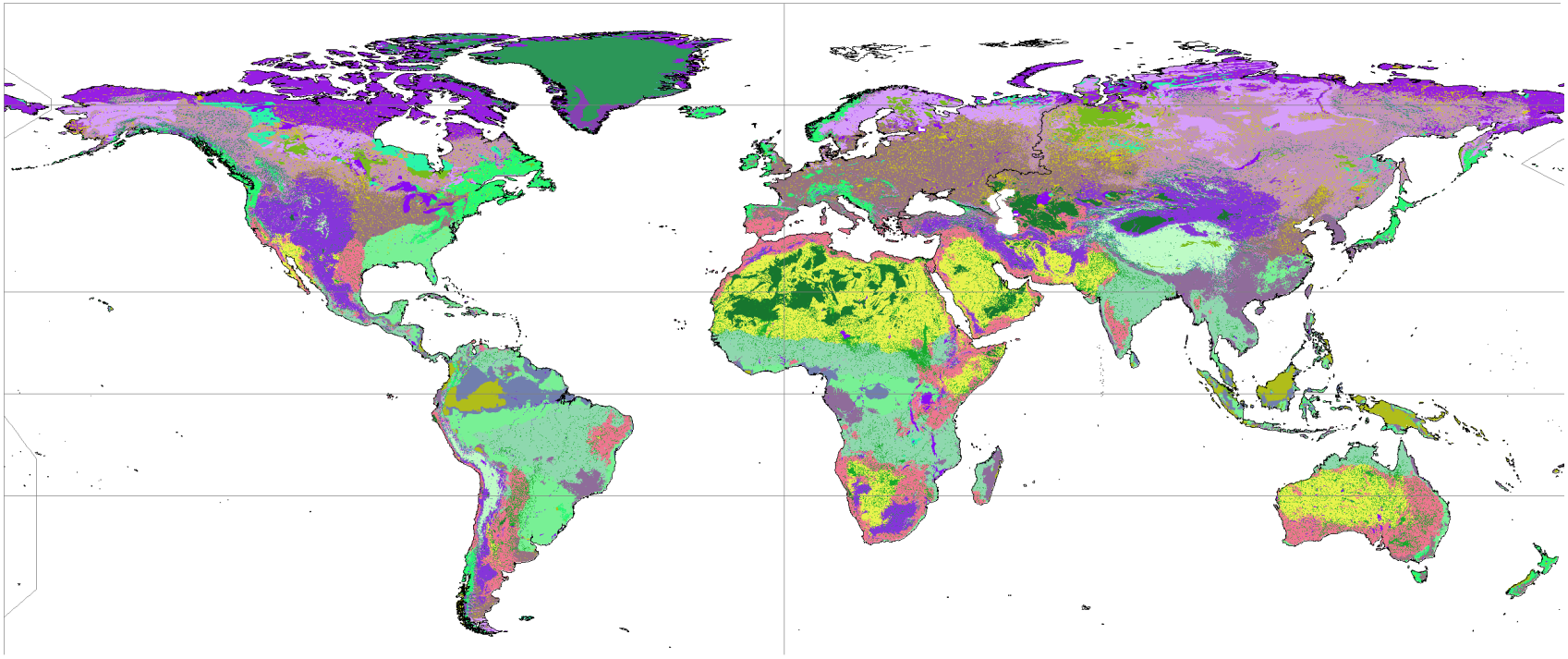
(Potter and Hargrove, 2013)

10 Global Ecoregions, Random Colors



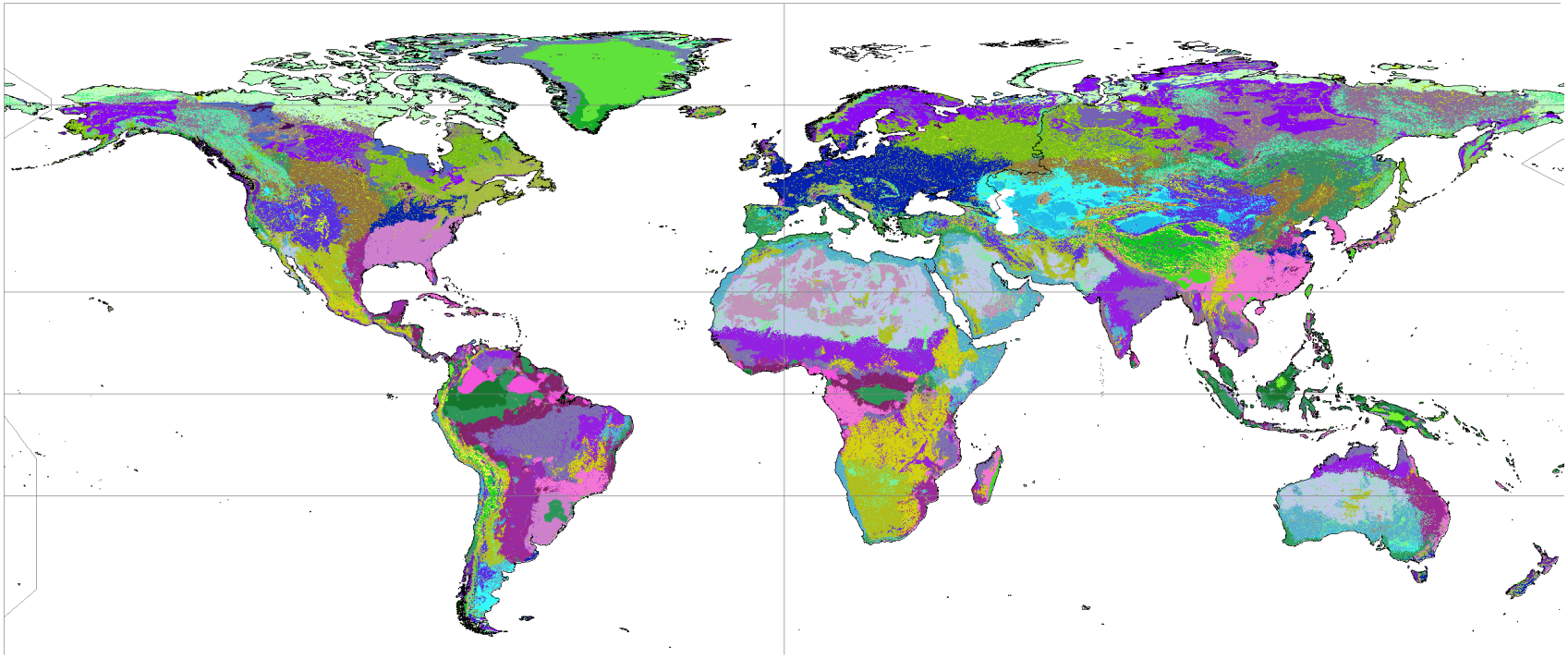
The 10 most different ecoregions globally are shown in random colors. Notice that areas with similar environmental characteristics are colored the same no matter where they occur on Earth.

25 Global Ecoregions, Random Colors



The 25 most different ecoregions globally are shown in random colors. Notice that areas with similar environmental characteristics are colored the same no matter where they occur on Earth.

50 Global Ecoregions, Random Colors



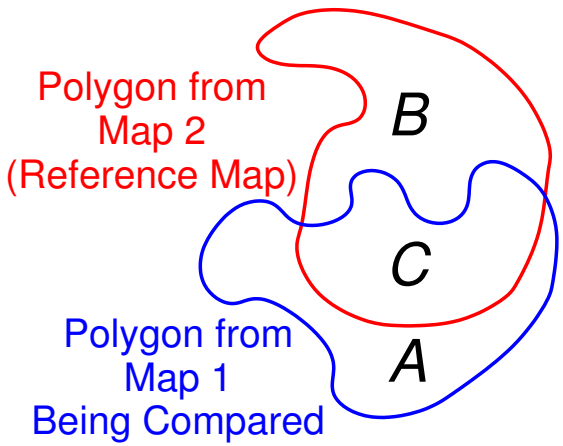
The 50 most different ecoregions globally are shown in random colors. Notice that areas with similar environmental characteristics are colored the same no matter where they occur on Earth.

Automated Supervision for Unsupervised Classification

- Clustering is an unsupervised classification technique, so ecoregions have no descriptive labels (e.g., **Eastern Deciduous Forest Biome**).
- **Label stealing** allows us to perform automated "supervision" by "stealing" the best corresponding human-created descriptive labels to assign to ecoregions.
- We employ a tool called **Mapcurves** to select the best ecoregion labels from ecoregionalizations delineated by human experts.
- We consider an entire library of ecoregion and land cover maps, and choose the label with the highest **goodness-of-fit (GOF)** score for every ecoregions polygon.

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.

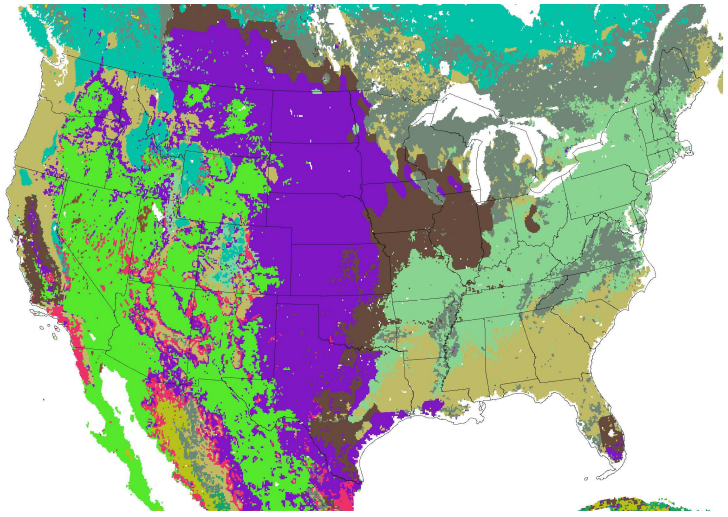


Goodness of Fit (GOF) is a unitless measure of spatial overlap between map categories:

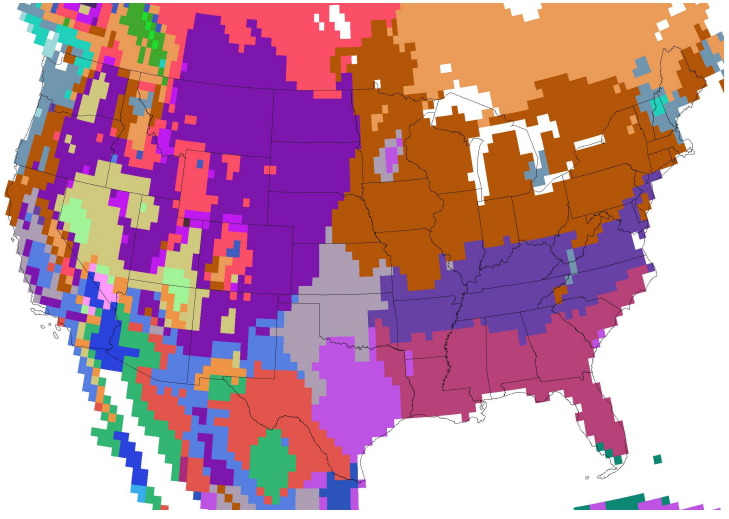
$$\text{GOF} = \sum_{\text{polygons}} \frac{C}{B+C} \times \frac{C}{A+C}$$

- 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-Derived Land Cover/Vegetation Type Maps



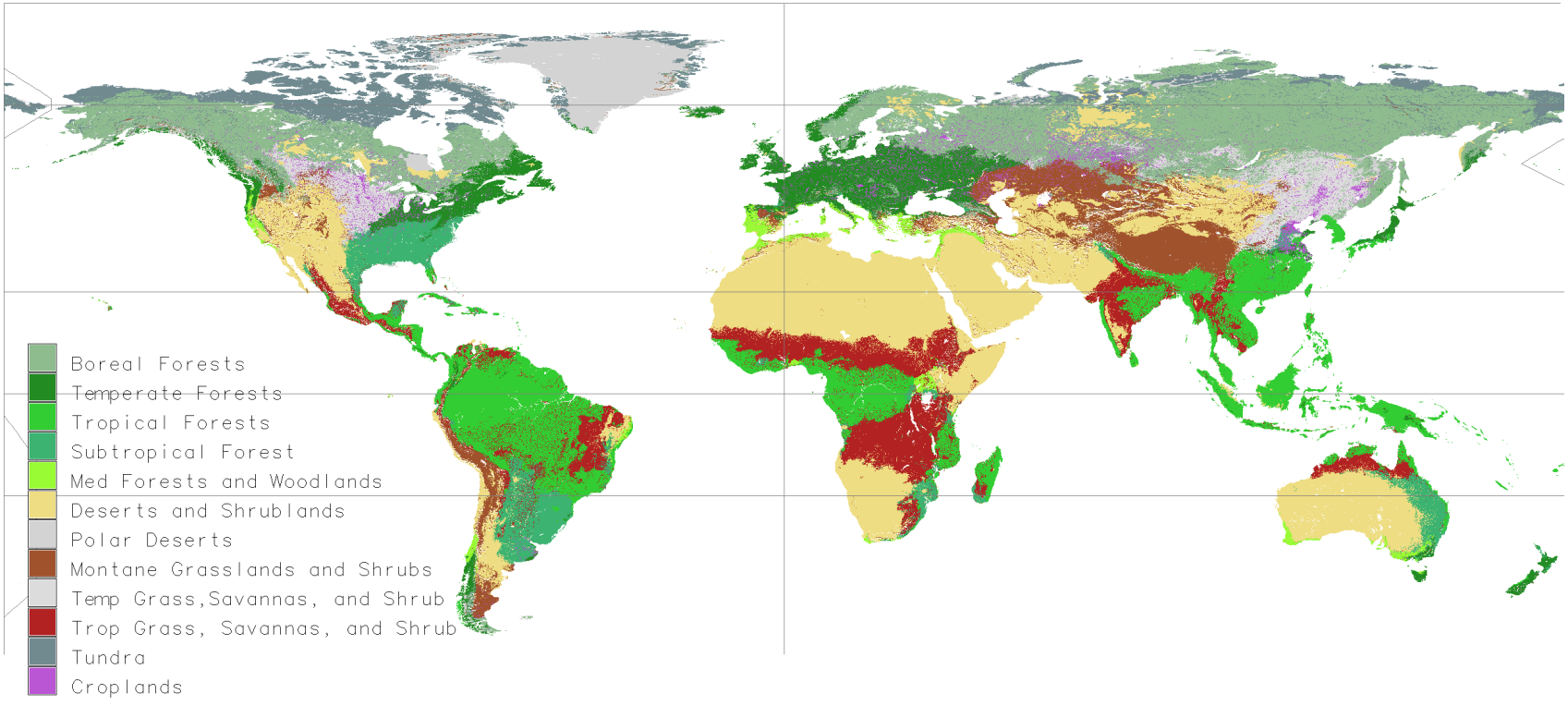
Foley Land Cover



Holdridge Life Zones

	Expert Map	# Cats
1.	DeFries UMd Vegetation	12
2.	Foley Land Cover	14
3.	Fedorova, Volkova, and Varlyguin World Vegetation Cover	31
4.	GAP National Land Cover	578
5.	Holdridge Life Zones	25
6.	Küchler Types	117
7.	BATS Land Cover	17
8.	IGBP Land Cover	16
9.	Olson Global Ecoregions	49
10.	Seasonal Land Cover Regions	194
11.	USGS Land Cover	24
12.	Leemans-Holdridge Life Zones	26
13.	Matthews Vegetation Types	19
14.	Major Land Resource Areas	197
15.	National Land Cover Database 2006	16
16.	Wilson, Henderson, & Sellers Primary Vegetation Types	23
17.	Landfire Vegetation Types	443
18.	ESA Global Land Cover	23

50 Ecoregions Reclassified by Label Stealing



The 50 quantitatively derived global ecoregions are reduced to 12 broadly defined land cover classes through the Label Stealing process.

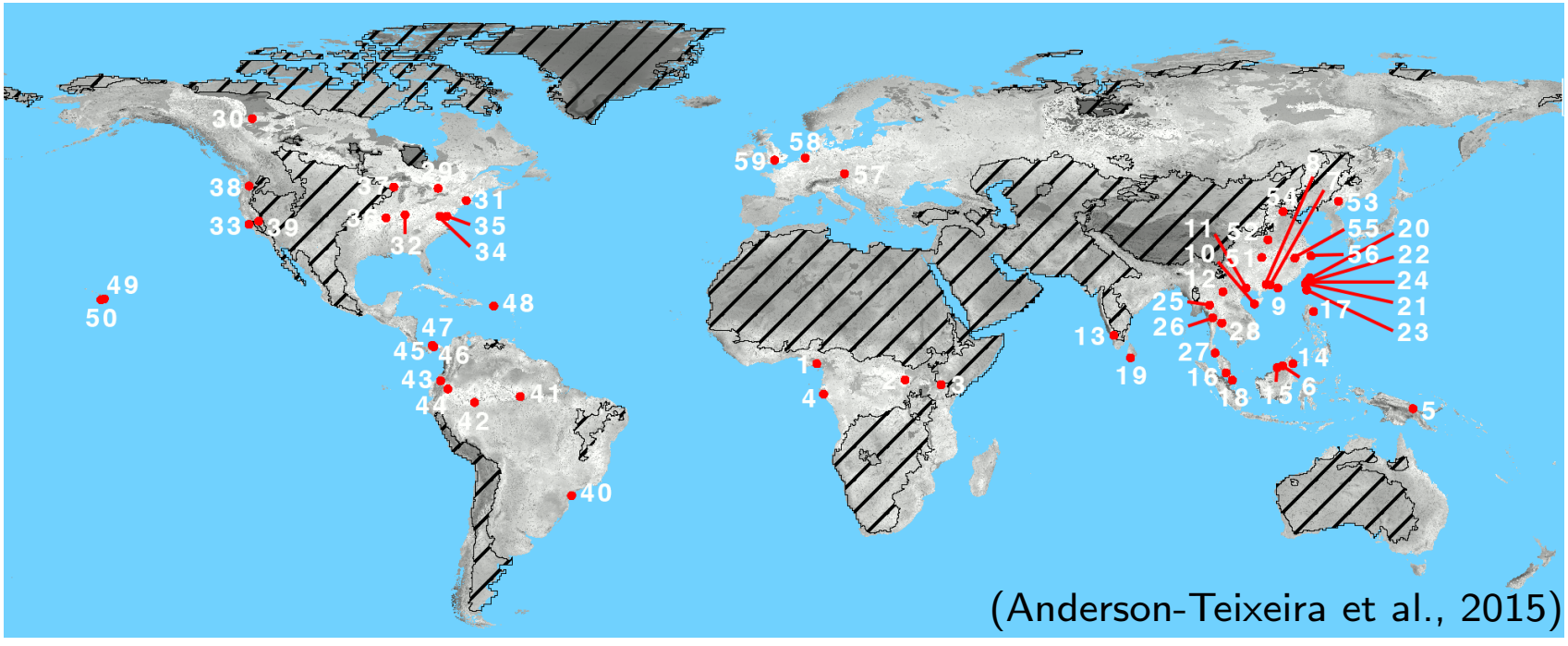
Global Forest Site Representativeness

- *Representativeness analysis* uses the standardized *n*-dimensional data space formed from all 17 input data layers.
- In this data space, the Euclidean distance between a sampling location (like Manaus, Brazil) and every other point is calculated.
- These data space distances are then used to generate grayscale maps showing the similarity, or lack thereof, of every location to the sampling location.
- In the subsequent maps, white areas are well represented by the sampling location or network, while dark and black areas as poorly represented by the sampling location or network.
- This analysis assumes that the climate surrogates maintain their predictive power and that no significant biological adaptation occurs in the future.

Site Representativeness: CTFS-ForestGEO, Mpala, Kenya

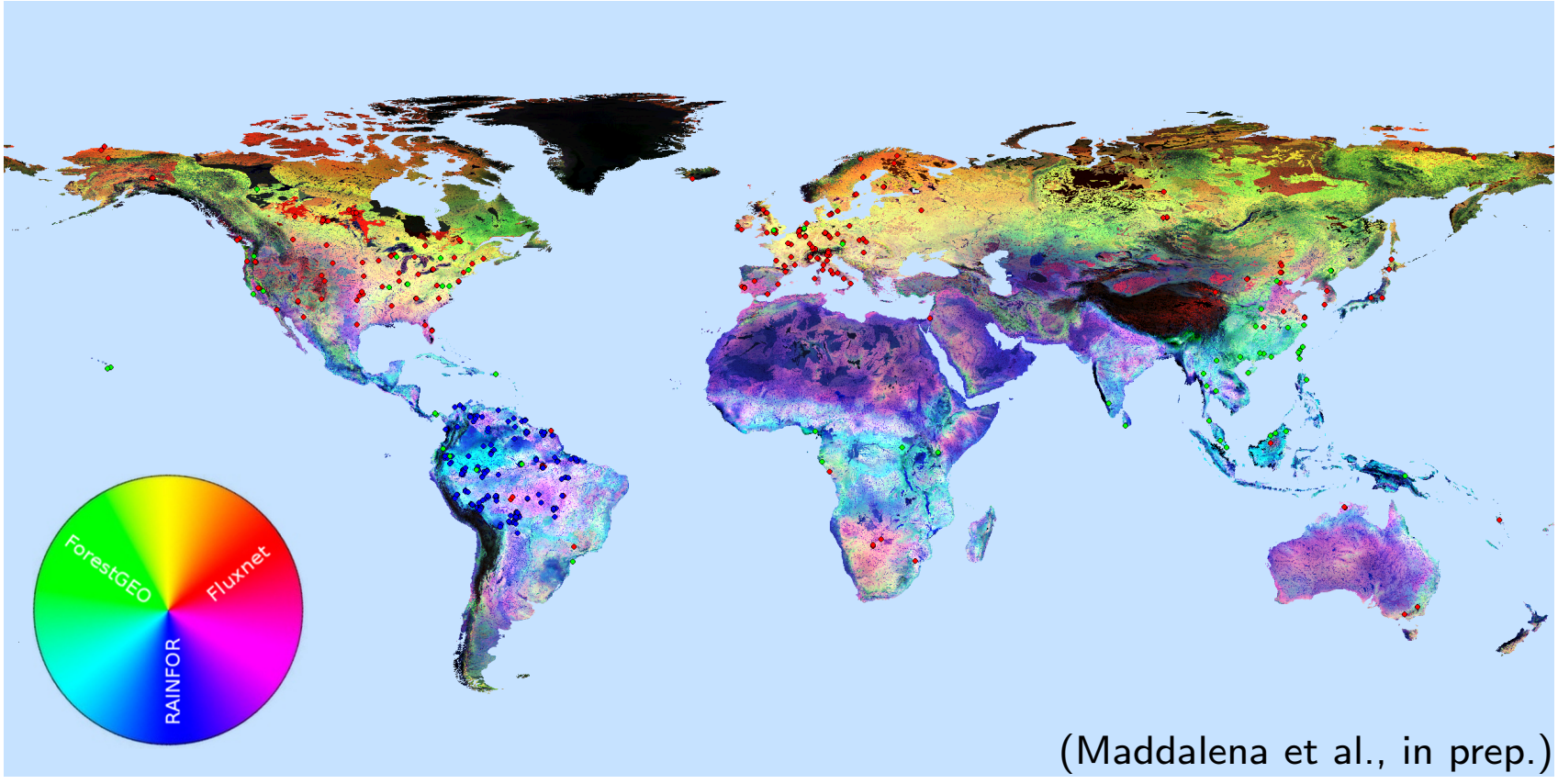
Light-colored regions are well represented and dark-colored regions are poorly represented by the sampling location shown in red.

ForestGEO Network Global Representativeness



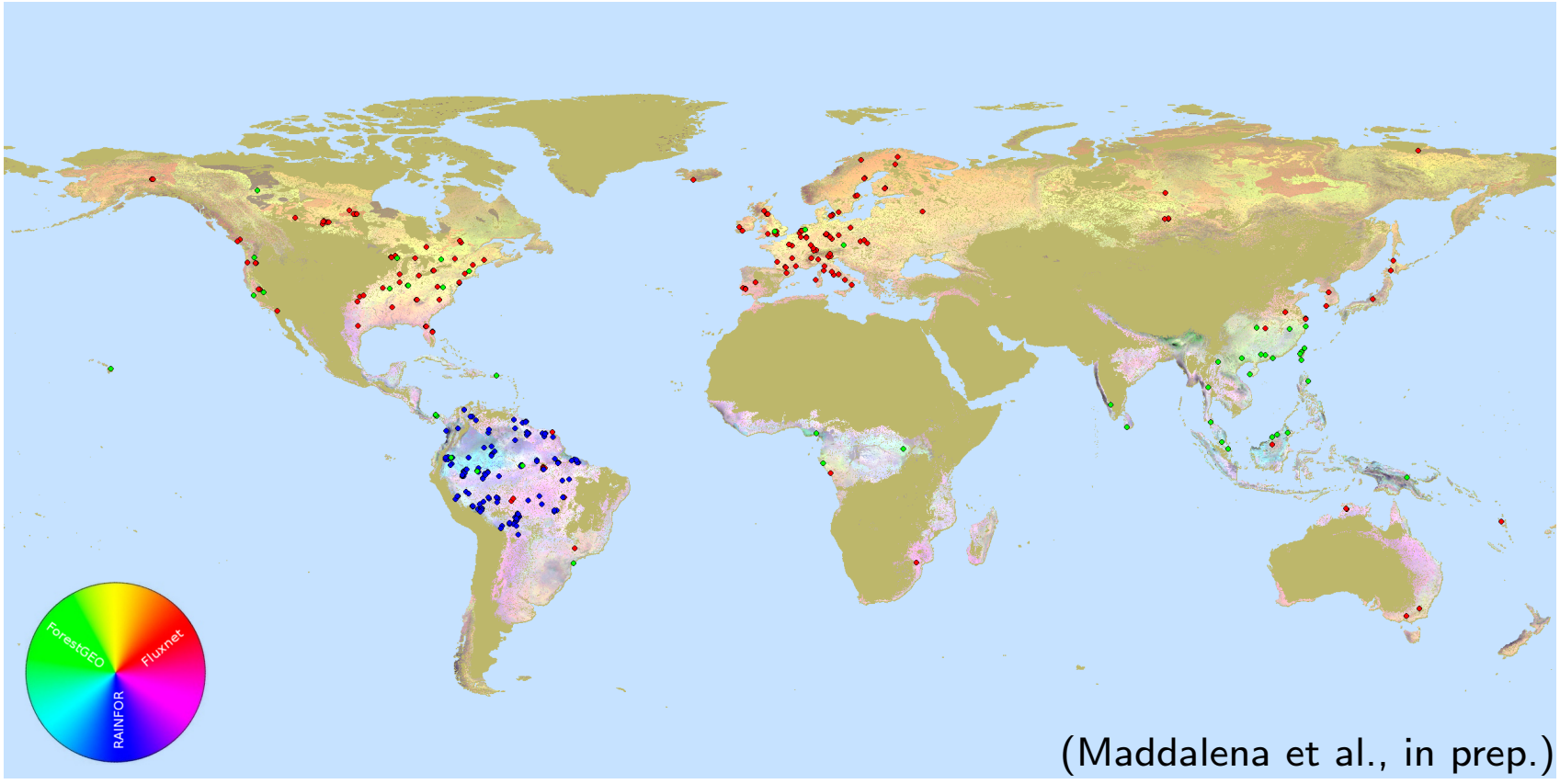
Map illustrating ForestGEO network representation of 17 bioclimatic, edaphic, and topographic conditions globally. Light-colored regions are well represented and dark-colored regions are poorly represented by the ForestGEO sampling network. Stippling covers non-forest areas.

Triple-Network Global Representativeness



Map indicates the sampling networks that offer the most representative coverage for any location. Every location is made up of a combination of three primary colors from Fluxnet (red), ForestGEO (green), and RAINFOR (blue).

Triple-Network Global Representativeness of Forests



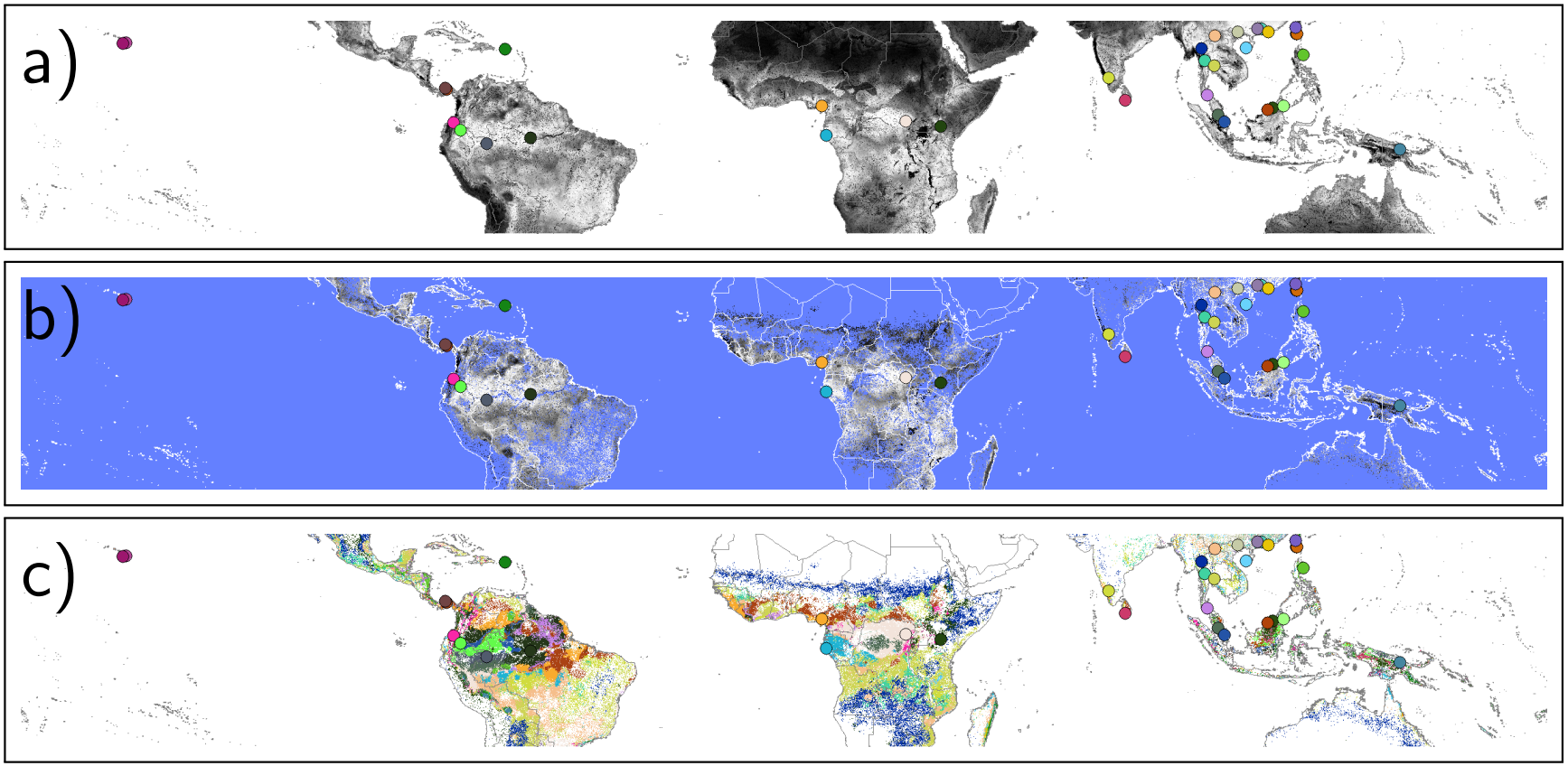
Map indicates the sampling networks that offer the most representative coverage at any forest location. *Only sites within forests were included and forest area was determined by Label Stealing.* Every location is made up of a combination of three primary colors from Fluxnet (red), ForestGEO (green), and RAINFOR (blue).

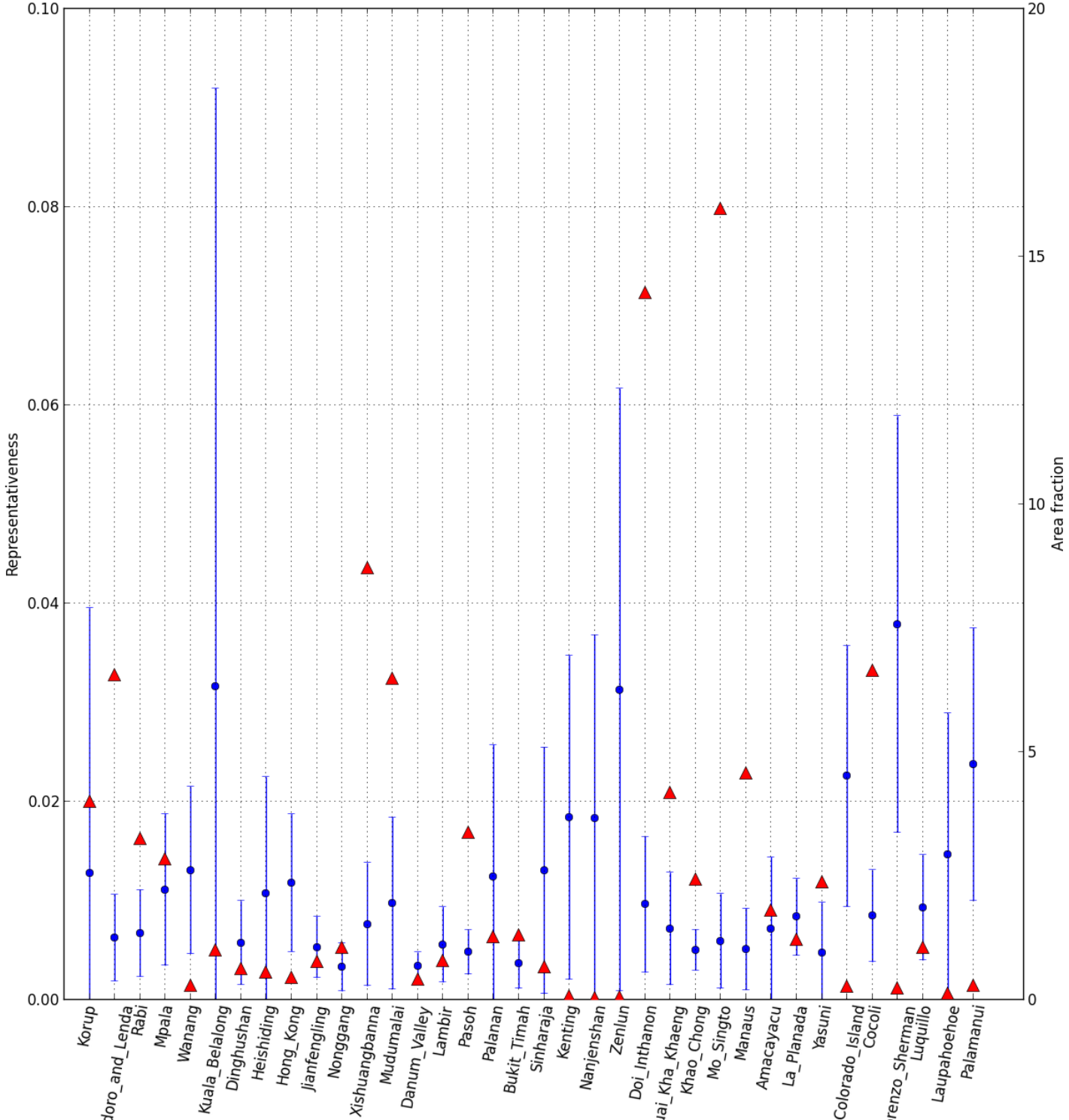
Global Forest Site Constituency

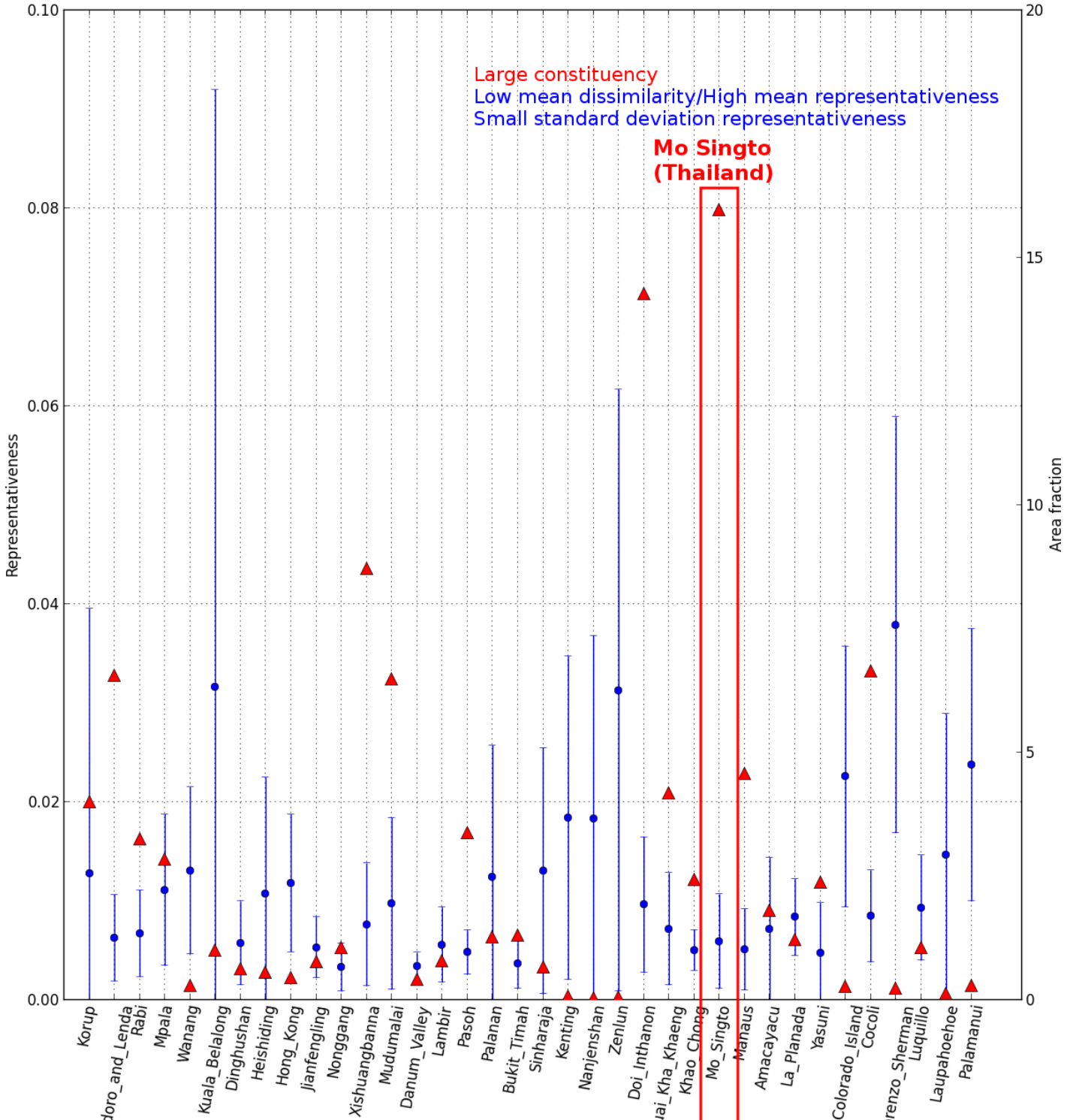
- For a fixed network of sampling sites, *constituency analysis* yields the spatial area represented best by any given site based on Euclidean distance in data space.
- For a given constituency, we can calculate a mean and standard deviation site representativeness.
- Thus,
  - a site with a large **constituency** provides broad spatial coverage;
  - a site with high **mean representativeness** (low dissimilarity) is a strong archetype of its constituency; and
  - a site with a large **standard deviation representativeness** provides broad data space coverage and is, therefore, the best (possibly poor) representative of a diverse constituency.
- These three metrics are (mostly) independent measures of network optimality.

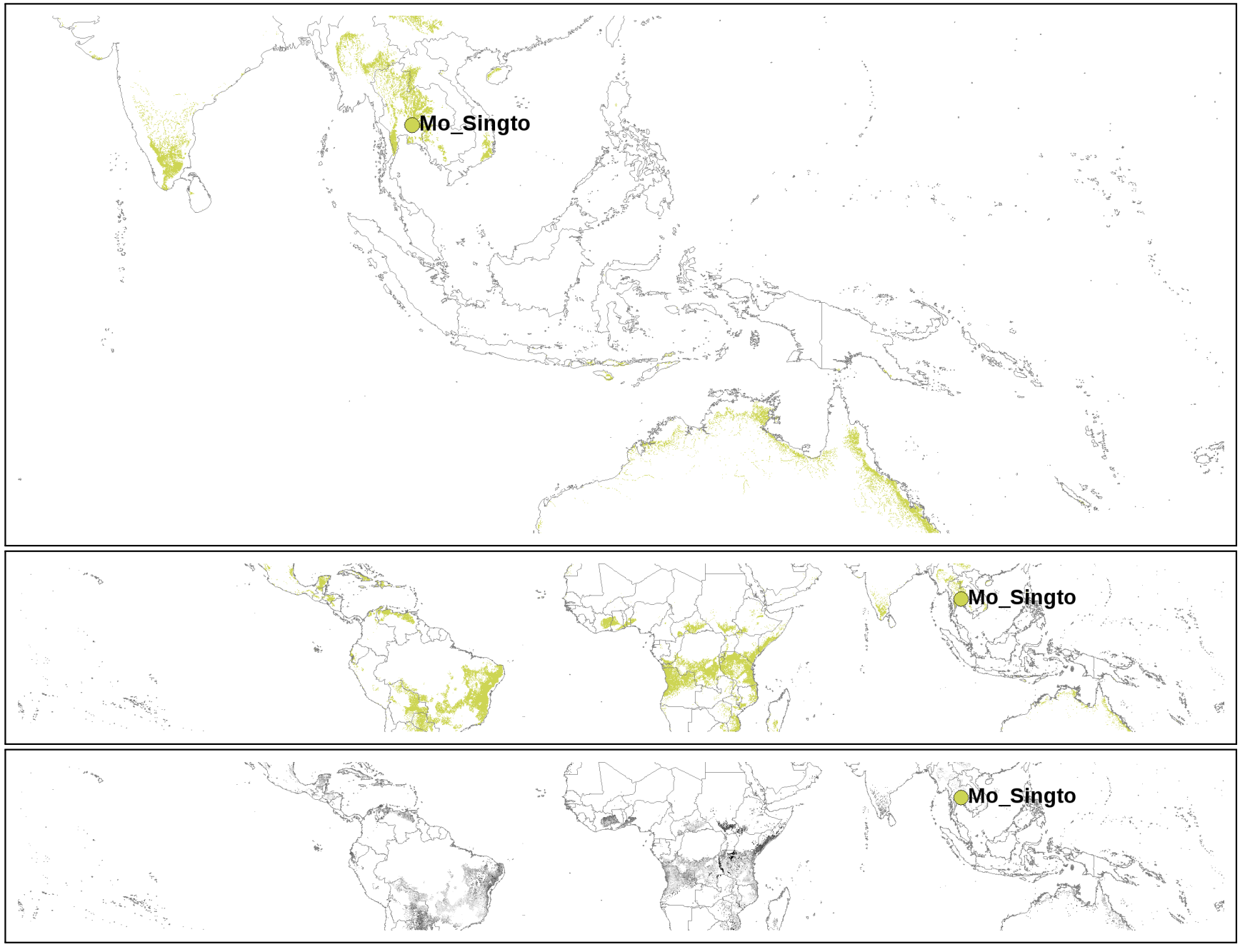
Tropical Forest Site Constituency

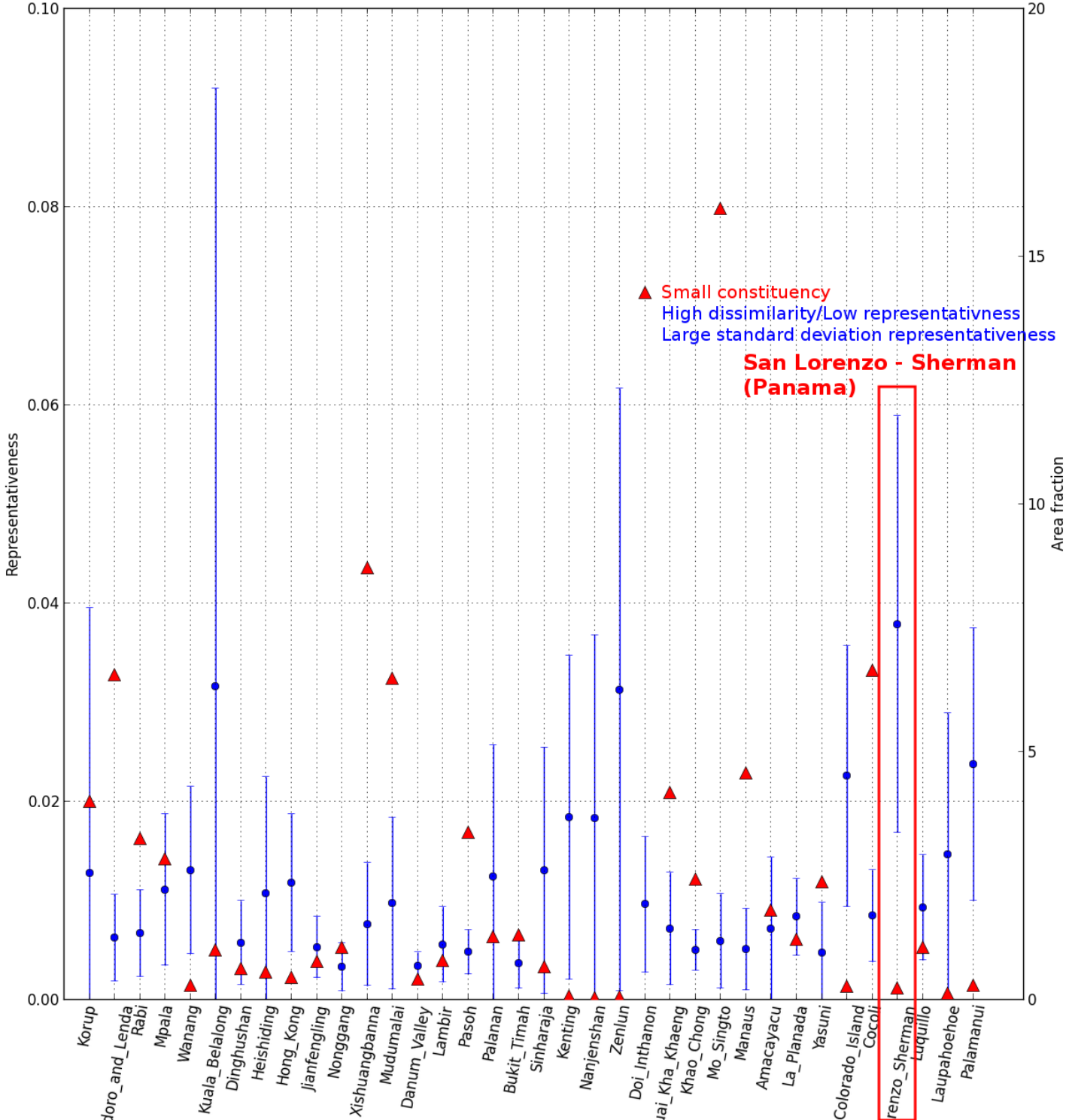
For insights into tropical forest network design, we perfomed representativeness and constituency analysis using the 36 CTFS-ForestGEO tropical sites to compute network a) representativeness, b) representativeness for tropical forests, and c) constituency for tropical forests.

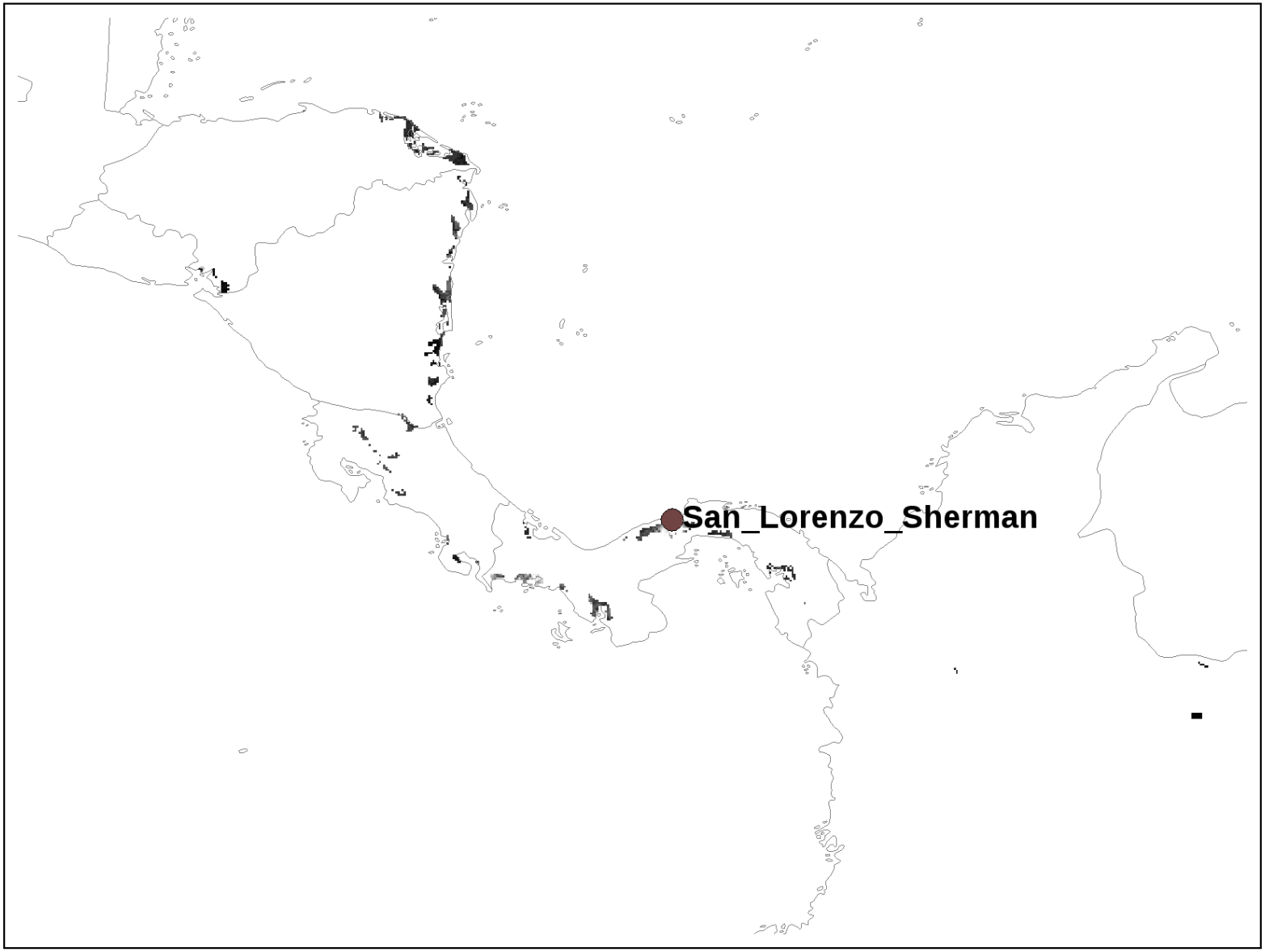


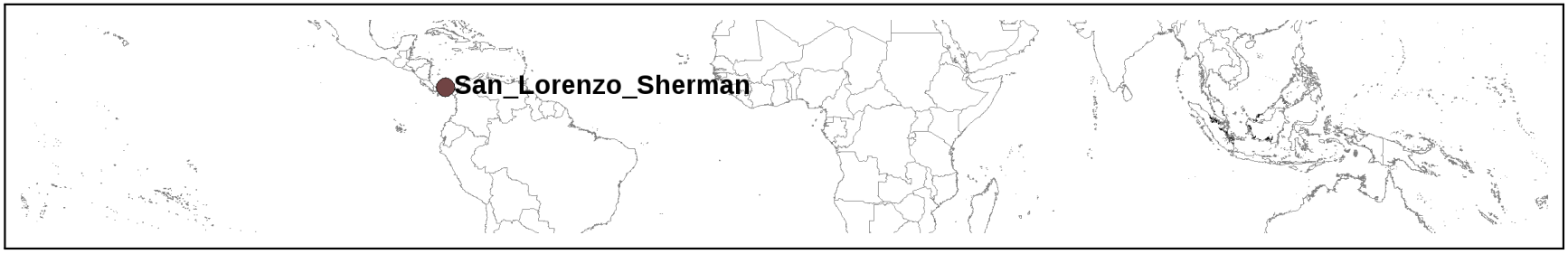

Representativeness
Area fraction
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto
Manaus
Amacayacu
La_Planada
Yasuni
Colorado_Island
Cocoli
Luquillo
Laupahoehoe
Palamanui


Large constituency
Low mean dissimilarity/High mean representativeness
Small standard deviation representativeness
Mo Singto
(Thailand)
Representativeness
Area fraction
0.10
0.08
0.06
0.04
0.02
0.00
20
15
10
5
0
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto
Manaus
Amacayacu
La_Planada
Yasuni
Colorado_Island
Cocoli
Luquillo
Laupahoehoe
Palamanui


Mo_Singto
Mo_Singto
Mo_Singto


Representativeness
Area fraction
0.10
0.08
0.06
0.04
0.02
0.00
20
15
10
5
0
Small constituency
High dissimilarity/Low representativness
Large standard deviation representativeness
San Lorenzo - Sherman
(Panama)
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto
Manaus
Amacayacu
La_Planada
Yasuni
Colorado_Island
Cocoli
Luquillo
Laupahoehoe
Palamanui


San_Lorenzo_Sherman


San_Lorenzo_Sherman

Conclusions

- **Multivariate Spatiotemporal Clustering (MSTC)** provides a quantitative framework for stratifying sampling domains, informing site selection, and determining representativeness of measurements.
- **Label Stealing** offers a useful means for interpreting and understanding ecoregion or sampling domain delineation.
- **Representativeness Analysis** and **Constituency Analysis** provide a systematic approach for optimizing site selection and up-scaling point measurements to larger domains.
- Methodology is *independent of resolution and surrogate data*, thus can be applied from site/plot scale to landscape/global scale with site measurements, remote sensing, and models.
- Paper describing analysis method applied for State of Alaska:

Hoffman, F. M., J. Kumar, R. T. Mills, and W. W. Hargrove (2013), Representativeness-based sampling network design for the State of Alaska, *Landscape Ecol.*, 28(8):1567–1586, doi:10.1007/s10980-013-9902-0.

Next Steps for Tropical Site Selection

- Input data layers must be selected to capture all the important environmental gradients related to carbon cycle drivers.
- A more careful analysis of existing sampling sites should consider type, frequency, and protocol of measurements.
- Observation data may be paired with projected changes in climate and atmospheric CO_2 to estimate how ecoregions may reorganize in the future.
- Method will be used to develop an optimized network of tropical forest sampling sites to answer key science questions.

For more information or to contribute ideas, please contact:

Forrest M. Hoffman
forrest@climatemodeling.org

References

K. J. Anderson-Teixeira, S. J. Davies, et al. CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Global Change Biol.*, 21(2):528–549, Feb. 2015. doi:10.1111/gcb.12712.

W. W. Hargrove and F. M. Hoffman. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environ. Manage.*, 34(Supplement 1):S39–S60, Apr. 2004. doi:10.1007/s00267-003-1084-0.

W. W. Hargrove, F. M. Hoffman, and P. F. Hessburg. Mapcurves: A quantitative method for comparing categorical maps. *J. Geograph. Syst.*, 8(2):187–208, July 2006. doi:10.1007/s10109-006-0025-x.

F. M. Hoffman, J. Kumar, R. T. Mills, and W. W. Hargrove. Representativeness-based sampling network design for the State of Alaska. *Landscape Ecol.*, 28(8): 1567–1586, Oct. 2013. doi:10.1007/s10980-013-9902-0.

M. Keller, D. Schimel, W. Hargrove, and F. Hoffman. A continental strategy for the National Ecological Observatory Network. *Front. Ecol. Environ.*, 6(5):282–284, June 2008. doi:10.1890/1540-9295(2008)6[282:ACSFTN]2.0.CO;2. Special Issue on Continental-Scale Ecology.

K. M. Potter and W. W. Hargrove. Quantitative metrics for assessing predicted climate change pressure on North American tree species. *Math. Comput. For. Nat.-Res. Sci.*, 5(2):151–169, Sept. 2013. URL `http://mcfns.com/index.php/Journal/article/view/MCFNS_169`.

D. Schimel, W. Hargrove, F. Hoffman, and J. McMahon. NEON: A hierarchically designed national ecological network. *Front. Ecol. Environ.*, 5(2):59, Mar. 2007. doi:10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2.

Acknowledgements

U.S. DEPARTMENT OF ENERGY

Office of Science

This research was sponsored by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the U. S. Department of Energy Office of Science and the U. S. Department of Agriculture Forest Service, and the National Science Foundation (AGS-1048890). This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.