Understanding the Representativeness of FLUXNET for Upscaling Carbon Flux from Eddy Covariance Measurements

Forrest M. Hoffman[1,2], Jitendra Kumar[1,2],
William W. Hargrove[3], and Nathaniel Collier[1]

[1]Oak Ridge National Laboratory, [2]University of Tennessee Knoxville, and
[3]USDA Forest Service Eastern Forest Environmental Threat Assessment Center

2017 US-IALE Annual Meeting
Inner Harbor, Baltimore, Maryland, USA
April 10, 2017

CLIMATE CHANGE SCIENCE INSTITUTE
OAK RIDGE NATIONAL LABORATORY

US - IALE

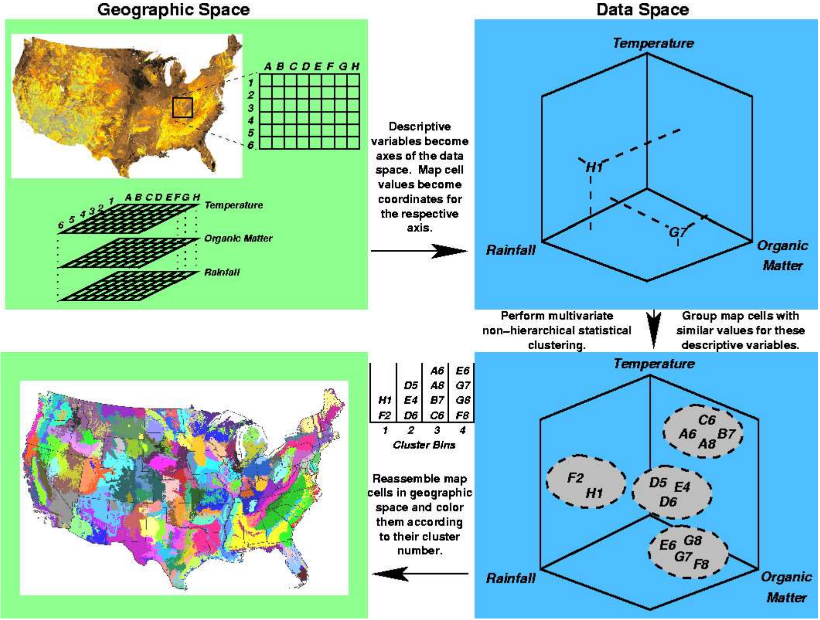
FOREST SERVICE U S DEPARTMENT OF AGRICULTURE

OAK RIDGE National Laboratory

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)



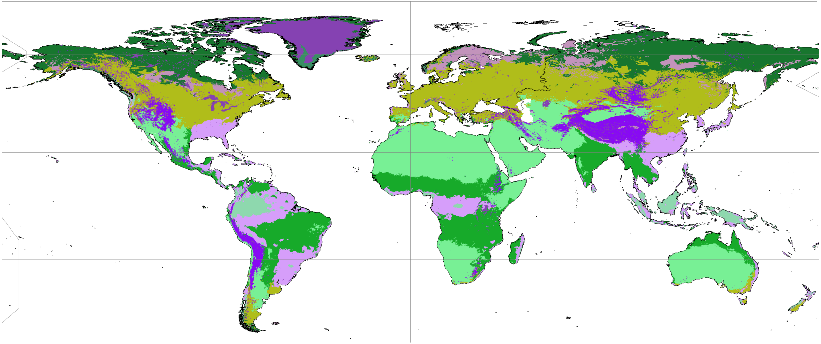
(Hargrove and Hoffman, 2004)

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 $T_{\text{avg}} \geq 5^\circ$C	$^\circ$C
Integer number of consecutive months where $T_{\text{avg}} \geq 5^\circ$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

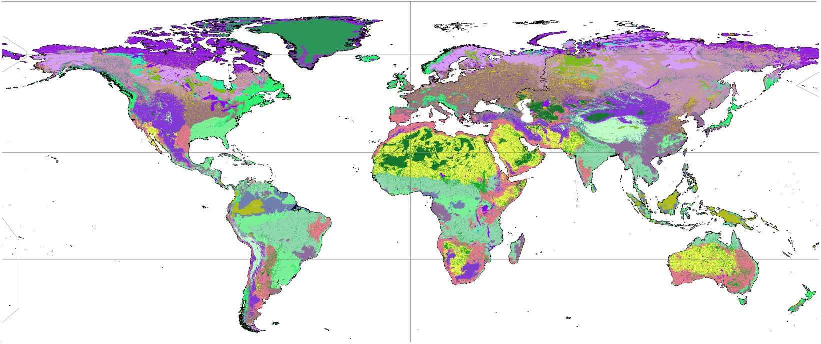
(Potter and Hargrove, 2013)

10 Global Ecoregions, Random Colors



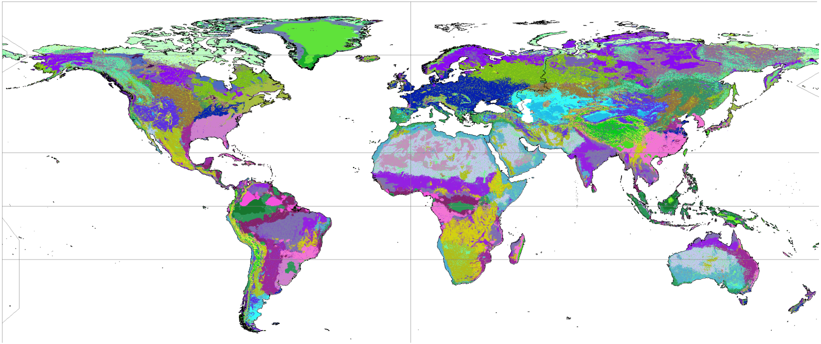
The 10 most different ecoregions globally are shown in random colors. Notice that areas with similar environmental characteristics are colored the same no matter where they occur on Earth.

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.

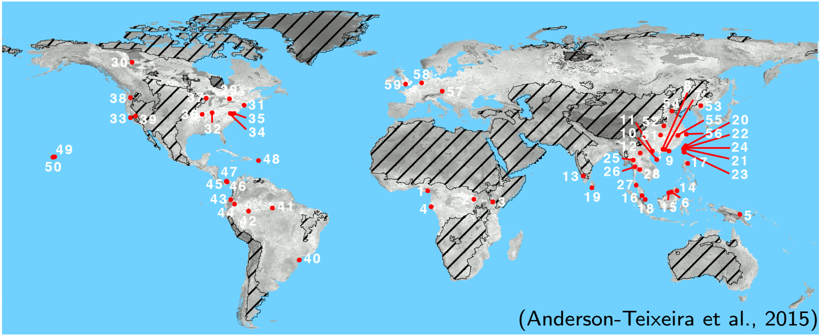
Global Forest Site Representativeness

- *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



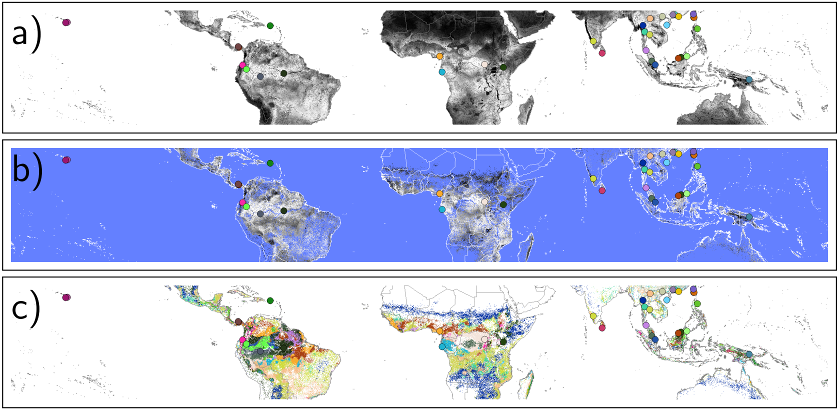
Map illustrating ForestGEO network representation of 17 bioclimatic, edaphic, and topographic conditions globally. Light-colored regions are well represented and dark-colored regions are poorly represented by the ForestGEO sampling network. Stippling covers non-forest areas.

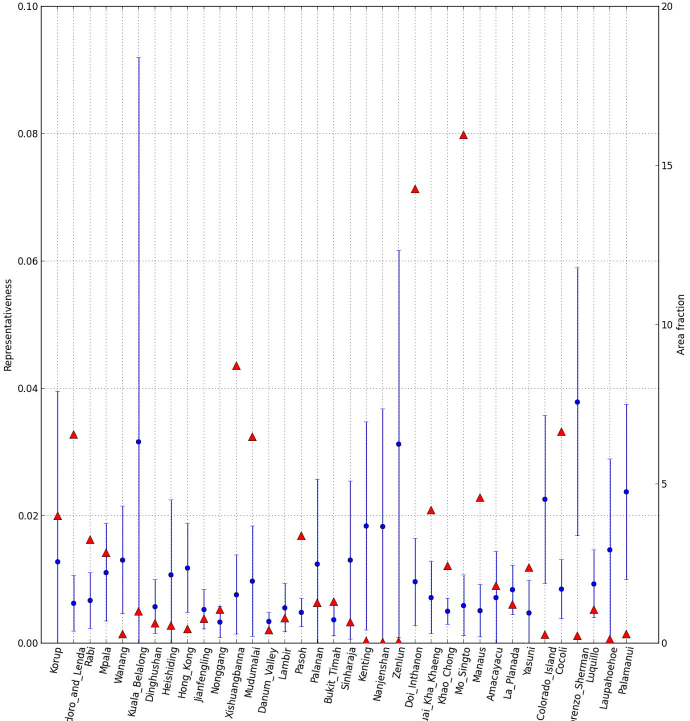
Global Forest Site Constituency

- 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.

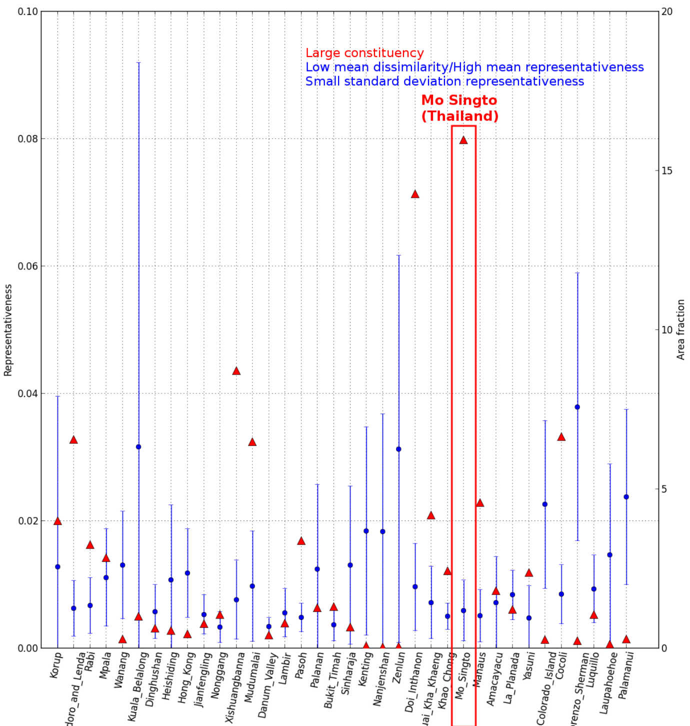
Tropical Forest Site Constituency

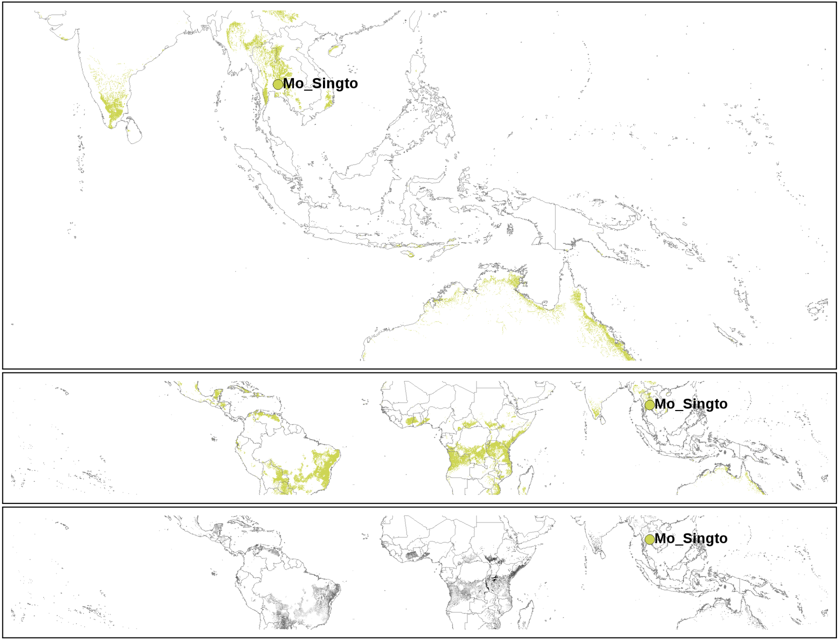
For insights into tropical forest network design, we perfomed representativeness and constituency analysis using the 36 CTFS-ForestGEO tropical sites to compute network a) representativeness, b) representativeness for tropical forests, and c) constituency for tropical forests.

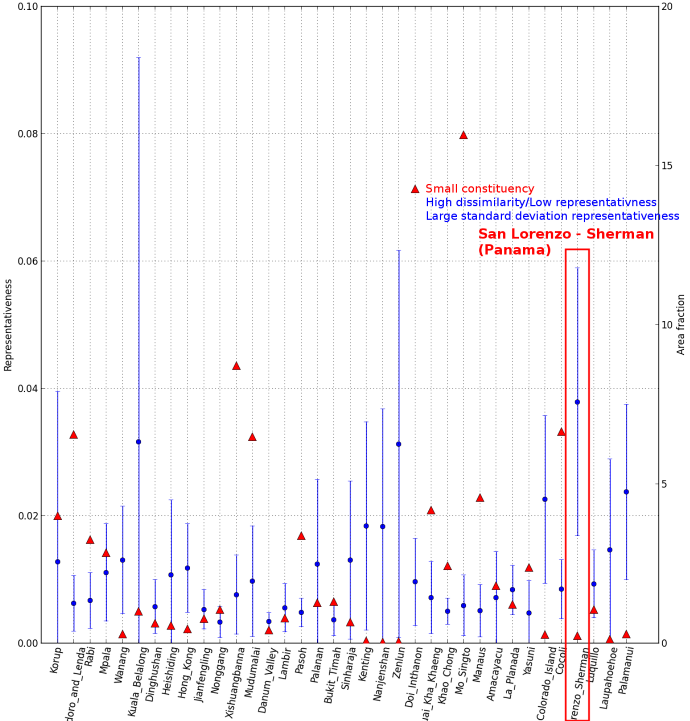


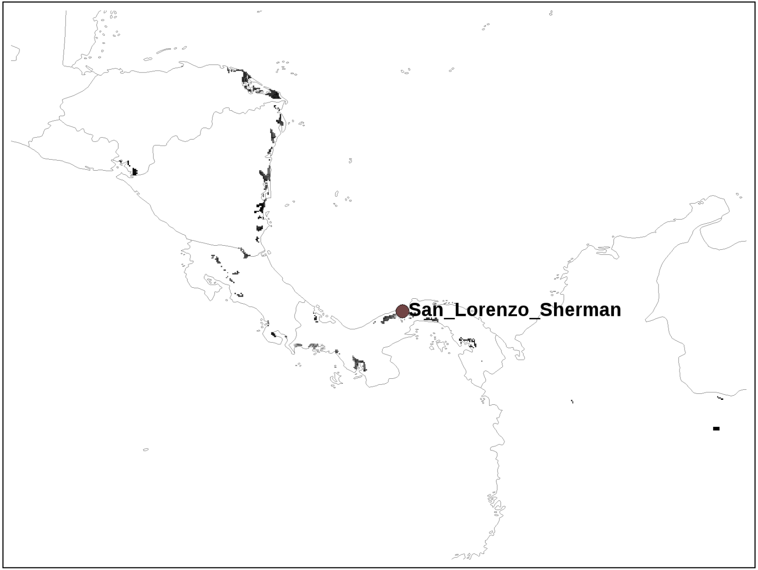


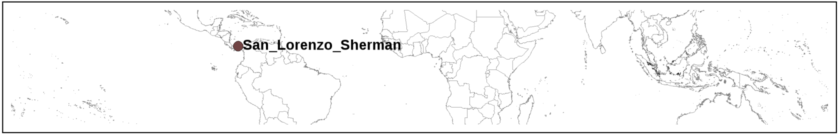
Representativeness
Area fraction
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
Korup
Rabi
Mpala
Wanang
Kuala_Belalong
Dinghushan
Heishiding
Hong_Kong
Jianfengling
Nonggang
Xishuangbanna
Mudumalai
Danum_Valley
Lambir
Pasoh
Palanan
Bukit_Timah
Sinharaja
Kenting
Nanjenshan
Zenlun
Doi_Inthanon
Khao_Chong
Mo_Singto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

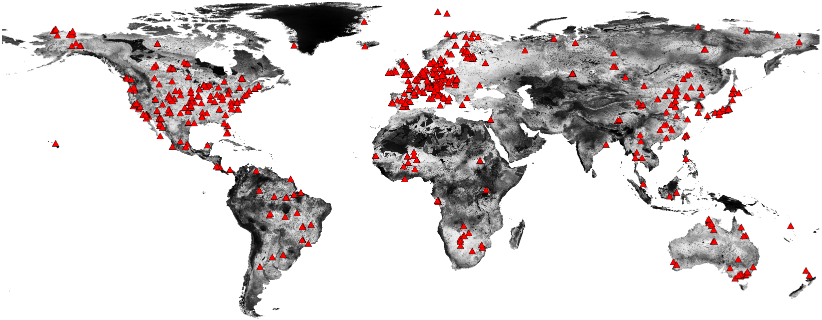

Mo_Singto
Mo_Singto
Mo_Singto


Small constituency
High dissimilarity/Low representativness
Large standard deviation representativeness
San Lorenzo - Sherman
(Panama)
Representativeness
Area fraction
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

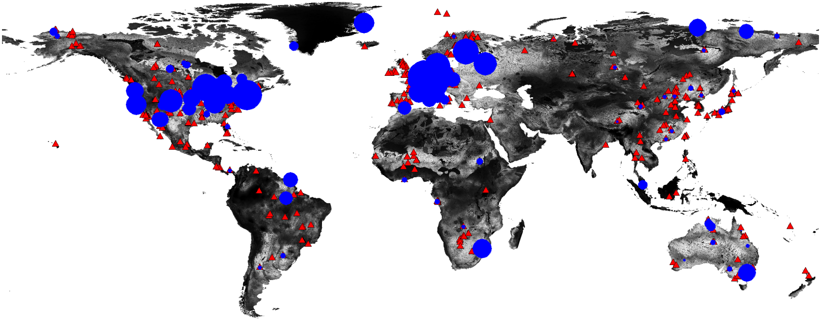
Network Representativeness of 770 Fluxnet Sites



(Kumar et al., 2016)

Red triangles indicate locations of all 770 Fluxnet sites. Light regions are well represented by this collection of sites, while dark regions are poorly represented.

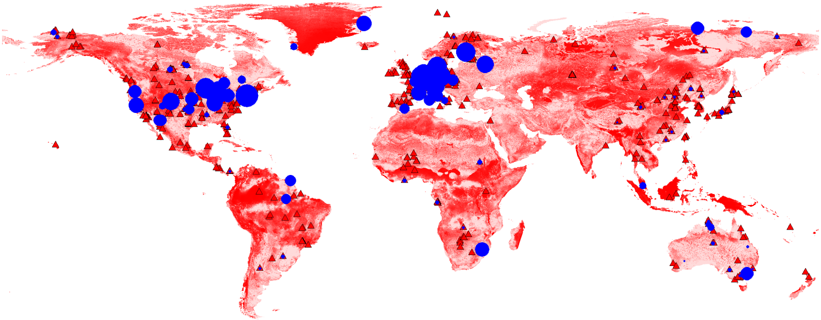
Network Representativeness of 164 FLUXNET2015 Sites



(Kumar et al., 2016)

Blue circles indicate locations and data years of the 164 sites in the FLUXNET2015 data set. Light regions are well represented by this collection of sites, while dark regions are poorly represented.

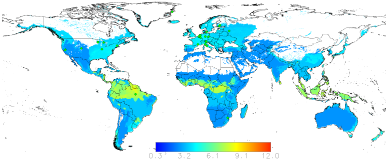
Network Representativeness Difference for Fluxnet Sites



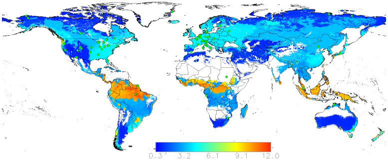
(Kumar et al., 2016)

Map shows the spatial distribution of representativeness lost by the unavailability of data for the 606 sites not included in the FLUXNET2015 data set.

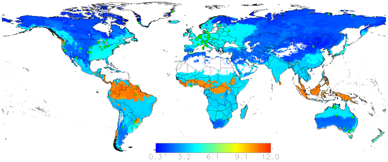
Upscaled Integrated Annual Mean GPP



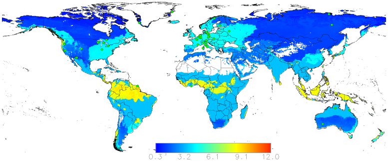
2000



2006



2011

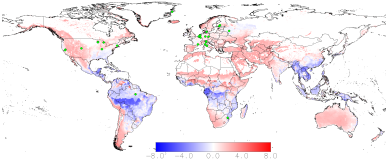


2014

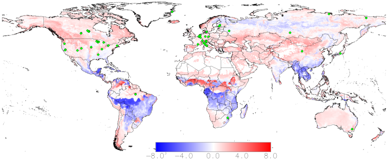
(Kumar et al., 2016)

Maps show the spatial distribution of the annual mean GPP ($\mathrm{g\,C\,m^{-2}\,d^{-1}}$) upscaled using clustering and inverse distance weighting (IDW).

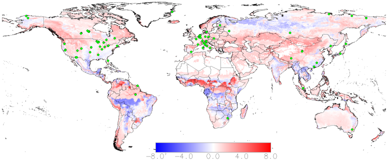
Comparison of FLUXNET2015 IDW with FLUXNET-MTE



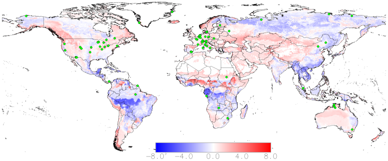
2000



2002



2004

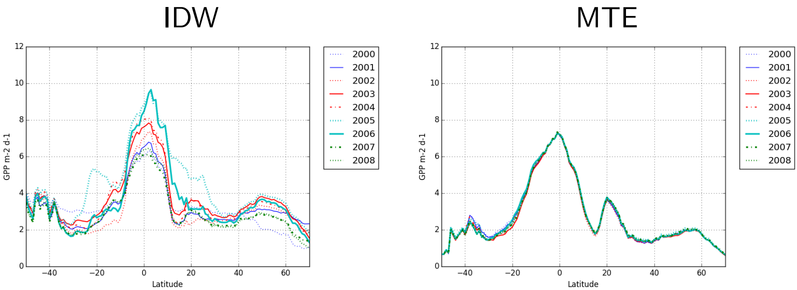


2008

(Kumar et al., 2016)

Maps show the difference between the IDW-upscaled GPP with FLUXNET-MTE. IDW estimated higher GPP in red regions and lower GPP in blue regions than FLUXNET-MTE.

Zonal mean comparison with FLUXNET-MTE



(Kumar et al., 2016)

The IDW scaled GPP has much greater zonal interannual variability than MTE.

Summary

- **Multivariate Spatiotemporal Clustering (MSTC)** provides a quantitative framework for stratifying sampling domains, informing site selection, and determining representativeness of measurements.
- **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.

- Paper describing FLUXNET2015 GPP scaling in discussion:

Kumar, J., F. M. Hoffman, W. W. Hargrove, and N. Collier (2016), Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements, *Earth Syst. Sci. Data Discuss.*, doi:10.5194/essd-2016-36.

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.

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.

J. Kumar, F. M. Hoffman, W. W. Hargrove, and N. Collier. Understanding the representativeness of FLUXNET for upscaling carbon flux from eddy covariance measurements. *Earth Syst. Sci. Data Discuss.*, 2016:1–25, Aug. 2016. doi:10.5194/essd-2016-36.

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

This research was sponsored by the Climate and Environmental Sciences Division (CESD) of the Biological and Environmental Research (BER) Program in the US Department of Energy Office of Science and the US 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 US Department of Energy under Contract No. DE-AC05-00OR22725.