EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 84 NUMBER 48
2 DECEMBER 2003
PAGES 529–544

New Analysis Reveals Representativeness of the AmeriFlux Network

PAGES 529, 535

The AmeriFlux network of eddy flux covariance towers was established to quantify variation in carbon dioxide and water vapor exchange between terrestrial ecosystems and the atmosphere, and to understand the underlying mechanisms responsible for observed fluxes and carbon pools. The network is primarily funded by the U.S. Department of Energy, NASA, the National Oceanic and Atmospheric Administration, and the National Science Foundation. Similar regional networks elsewhere in the world—for example, CarboEurope, AsiaFlux, OzFlux, and Fluxnet Canada—participate in synthesis activities across larger geographic areas [*Baldocchi et al.*, 2001; *Law et al.*, 2002].

BY WILLIAM W. HARGROVE, FORREST M. HOFFMAN, AND BEVERLY E. LAW

The existing AmeriFlux network will also form a backbone of "Tier 4" intensive measurement sites as one component of a four-tiered carbon observation network within the North American Carbon Program (NACP). The NACP seeks to provide long-term, mechanistically detailed, spatially resolved carbon fluxes across North America [*Wofsy and Harriss*, 2002]. For both of these roles, the AmeriFlux network should be ecologically representative of the environments contained within the geographic boundaries of the program. A new ecoregion-scale analysis of the existing AmeriFlux network reveals that, while central continental environments are well-represented, additional flux towers are needed to represent environmental

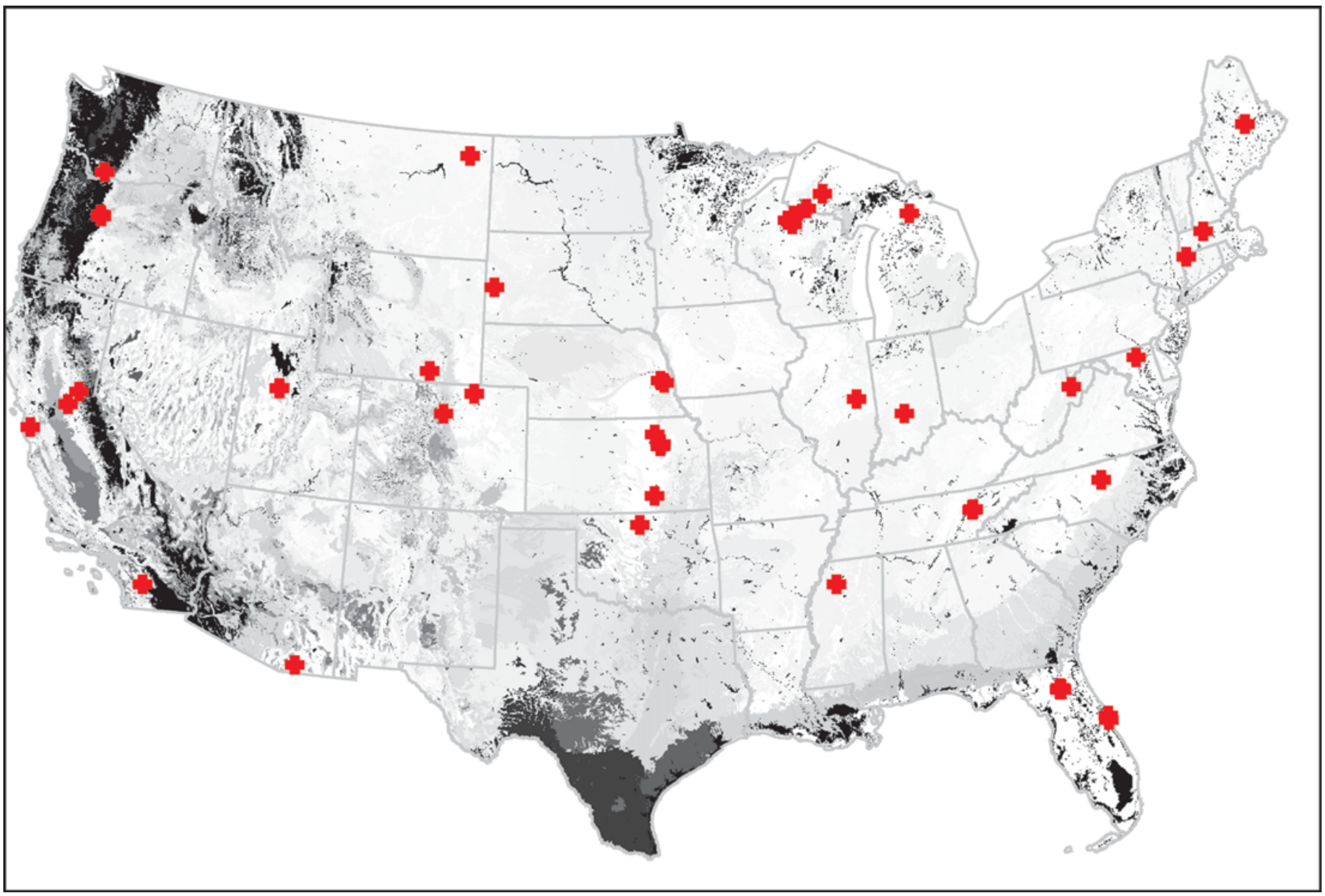

Fig. 1. The representativeness of an existing spatial array of sample locations or study sites—for example, the AmeriFlux network of carbon dioxide eddy flux covariance towers—can be mapped relative to a set of quantitative ecoregions, suggesting locations for additional samples or sites. Distance in data space to the closest ecoregion containing a site quantifies how well an existing network represents each ecoregion in the map. Environments in darker ecoregions are poorly represented by this network.

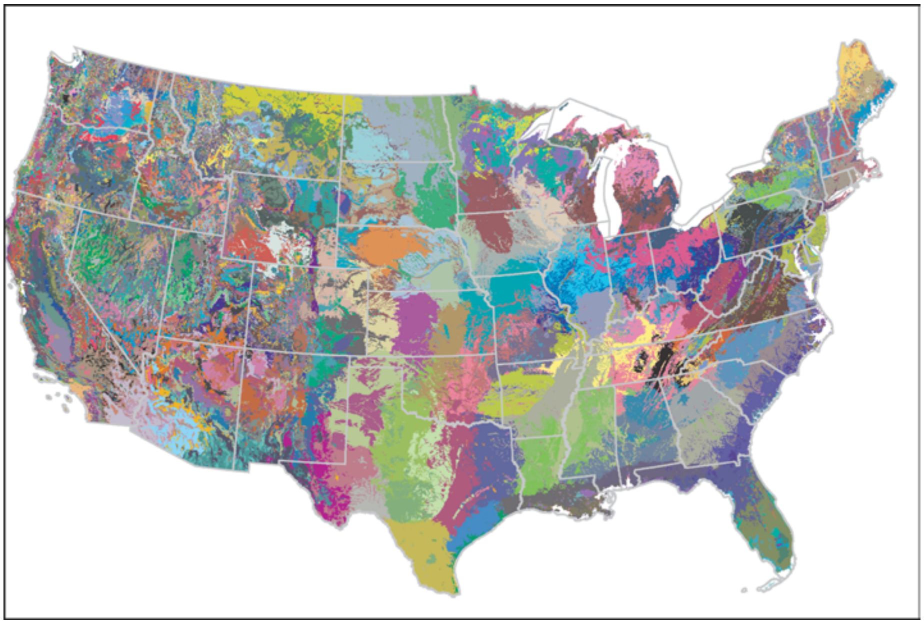

Fig. 2. Dividing the United States into the 2000 most different multivariate ecoregions based on 25 climatic, physiographic, and edaphic factors provides a foundation for the analysis of networks. The user can specify the number of clustered ecoregions that are produced. All ecoregions within the map have roughly equal environmental heterogeneity.

combinations in south Texas, the Sonoran Desert, and the Pacific Northwest (Figure 1).

Analysis of the representativeness of a network of sites entails a comparison of distances among centroids in a multivariate environmental "data space." A network in this sense consists of a geographic constellation of installations or facilities, or can simply represent locations where samples have been taken. To quantify network coverage, we determine how different each ecoregion in the map is from the ecoregion containing the most similar network site or sample.

The new analysis is based on a set of statistically derived ecoregions produced for the lower 48 United States using a multivariate clustering process. Twenty-five environmental conditions form a multivariate description of the environment present within each 1-km map raster cell. Because of the small cell size and the large number of cells, the cluster analysis was performed in parallel on a supercomputer [*Hargrove et al.*, 2001]. Normalized variable values from each map cell are used as coordinates to plot the cells in an environmental space with as many axes as there are multivariate environmental dimensions. Two map cells that are plotted close to one another in data space will have similar mixtures of environmental conditions, and are likely to be classified in the same homogeneous ecoregion cluster.

The user can specify the number of clustered ecoregions that result from the process, making it possible to divide the map into a few large, coarsely defined ecoregions, or a larger number of small, highly specified ecoregions (see http://geobabble.ornl.gov/cgi-bin/pzs for a spectrum of quantitative U.S. ecoregions). Such statistically generated ecoregions can be produced based on user-selected continuous variables, allowing customized regions to be generated for any specific problem [*Hargrove and Hoffman*, 2003].

An iterative k-means clustering procedure assigns each map cell to the closest of k centroids. At the end of each iteration, centroid positions are recomputed and another assignment iteration begins. After convergence, the map cells, with their new ecoregion assignments, are re-assembled in geographic space. All final quantitative ecoregions contain roughly equal environmental heterogeneity. Coordinates of the final centroids quantitatively define the synoptic conditions for each ecoregion.

For each ecoregion in the map, we find the Euclidean distance in data space to the single closest ecoregion that contains a site from the network. This distance is coded to a gray level, so that darker areas represent areas that are poorly represented by the existing network (Figure 1). Network analysis shows how well the sampled environments represent the rest of the map and identifies the best locations for new sites or installations. The best location for an additional site will be in places that are the least well-represented (darkest) by the network of existing sites.

The 2000 most different multivariate ecoregions within the conterminous United States (Figure 2) were statistically delineated using 25 primary environmental forcing factors, including elevation, mean and extremes of annual temperature, mean monthly precipitation, soil nitrogen, organic matter, water capacity, frost-free days, soil bulk density and depth, and solar aspect and insolation. No direct vegetation parameters were included in these analyses, only primary climatic and physiographic drivers. This fine division into many tightly defined ecoregions is far more than can be delineated traditionally using human expertise. This fine-scale ecoregion map was used as the basis for the quantitative analysis of networks.

Environments in the central, midwestern, and northeastern portions of the United States are well-represented by existing AmeriFlux tower sites (Figure 1). Southern, southwestern, and Pacific Northwest environments are less well represented by existing tower sites. As a result of a May 2003 proposal solicitation (DE-FG01-03ER03-22), DOE is currently selecting additional tower sites in the "upper Midwest region of the USA, bounded by Minnesota/Wisconsin on the north, Missouri/Oklahoma on the south, Indiana on the east and Nebraska on the west." This region has comparatively simple terrain, a large range of seasonal fluxes from agriculture, modest and widely distributed anthropogenic emissions, and dense pre-existing meteorological measurements. In terms of representing environmental conditions within the lower 48 states of the United States, however, our regional analysis suggests that additional sites in the Pacific Northwest or south Texas would contribute significantly more marginal representation than would additional Midwestern sites.

Showing Geographical Representation

The geographic representation contributed by each site can be shown by repeating the network analysis with each site alone. Importance values can be calculated for each site based on the marginal representation it adds to the network. Were it necessary to remove a site, quantifying the contribution of each site to network representation could minimize the impact of site elimination on network representation. Finally, for a network that has not yet been deployed (e.g., NSF's proposed National Ecological Observation Network (NEON)), a theoretically optimal network can be designed that has the highest possible representation of environmental conditions on a map, given a specified number of sites and underlying ecoregions.

Network analysis is ecoregion-based and operates at the scale of the entire sampling network, considering how well a sampling network represents the map that contains it. It does not consider specific local conditions, land uses, disturbance history, or anthropogenic treatments—that is, clear-cuts, forest plantations, or agriculture—unless data about such fine-scale land conditions are available, and are included as inputs. Network analysis depends on judicious selection of the environmental variables being considered. Results are calculated with respect to the selected input variables only and depend on the quality of the input data.

Results showing the national representativeness of NSF's Long-Term Ecological Research (LTER) network and the network of U.S. National Parks can be seen at http://research.esd.ornl.gov/~hnw/networks/presentation. Some poorly represented ecoregions appear in many of the network analysis maps; these may be outlying cluster ecoregions that are far from the center of occupied data space. These "unusual" ecoregions are generally located around the geographic

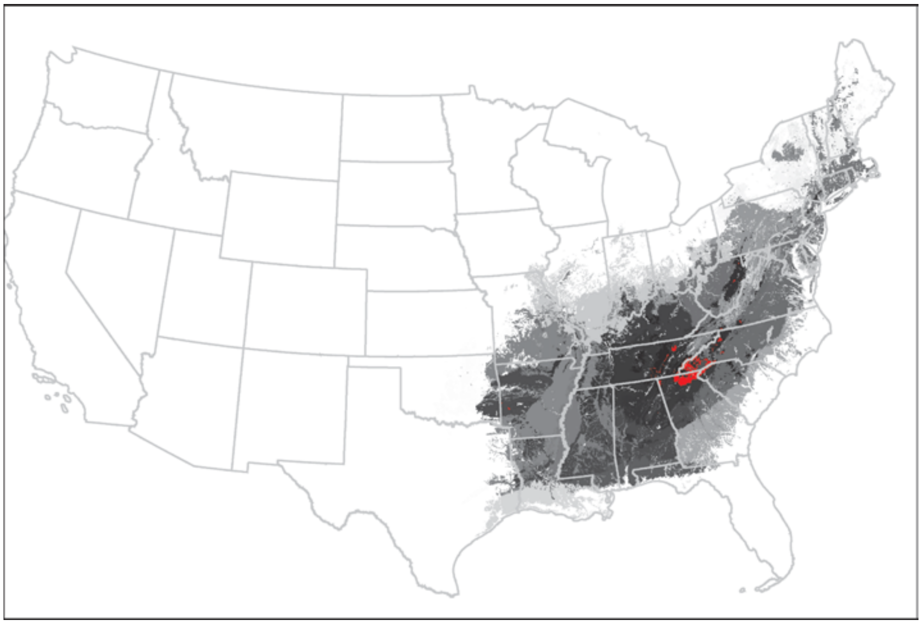

Fig. 3. Maps of similarity to any selected quantitative ecoregion can be produced. The Euclidean distance in data space from the centroid of each statistical ecoregion to the centroid of the chosen region is calculated. Ecoregions with closer centroids are more similar and are colored darker gray. The multivariate ecoregion containing the Great Smoky Mountains (shown in red) was selected from this 1000-ecoregion map based on 25 primary environmental variables, so that the map shows the quantitative degree of "Smoky Mountains-ness" across the map. The temperate deciduous forest biome is revealed to be high in multivariate "Smoky Mountains-ness."

periphery of the United States. The more "average" continental environments are easier to sample and represent than environments around the edges of the nation. The Pacific Northwest, for example, is finely divided into many small ecoregions, making this complex area harder to capture with network sites [*Law and Waring*, 1994].

Quantifying the Similarity of Ecoregions

Like a network, each single ecoregion in the map bears some resemblance to all other ecoregions. Similarity of any ecoregion to all other ecoregions can be quantified and displayed as a "representativeness" map. One can select a single ecoregion of interest, and then produce a sorted list of the similarity of all other ecoregions to the one selected. The chosen ecoregion establishes an origin in data space, and, using the Euclidean distance from this origin to the centroid of every other ecoregion, pair-wise similarity measures can be calculated. Coding these pair-wise similarity values as gray levels, the degree of similarity of all ecoregions to the selected ecoregion can be mapped.

Maps showing the innate similarity between a particular ecoregion and the rest of the map can be produced. For example, starting with a slightly coarser map of the 1000 most different ecoregions based on the 25 primary environmental factors described earlier, a map of "Smoky Mountains-ness" was produced that shows how similar other regions are to the Great Smoky Mountains of Tennessee and North Carolina (Figure 3). The multivariate ecoregion containing the Great Smoky Mountains is shown in red. Darker areas are most similar to the selected ecoregion.

The map of "Smoky Mountains-ness" statistically rediscovers the U.S. portion of the temperate deciduous forest biome. One small spot in the Ozarks and one spot in the Monongahela National Forest of West Virginia, though spatially disjoint, are actually a part of the same ecoregion that contains the Smoky Mountains, and, at this level of division, consist of pure "Smoky Mountains-ness." The Adirondack Mountains of New York are also relatively high in multivariate "Smoky Mountains-ness." These comparative representation maps quantify the resemblance among ecoregions.

Representation can be quantified within single ecoregions as well. The representation of individual map cells, as indicated by their distances from the centroid of the ecoregion to which they are assigned, could be used to pinpoint the best location of an additional network site within the most poorly represented ecoregion. Contours of quantitative representation were used to characterize borders between ecoregions as gradual, sharp, or of changing character along their length [*Hargrove and Hoffman*, 1999].

We consider this analysis of the AmeriFlux network to be a preliminary proof-of-concept; many improvements could be made. A custom ecoregionalization that is based on variables chosen as specifically relevant to carbon flux, perhaps including direct vegetation characteristics, should be used for best results. Expanding the extent of the analysis to cover the North American continent or the globe would show how well the existing AmeriFlux network represents even larger geographic areas. For example, inclusion of the single AmeriFlux site in Mexico and the numerous Canadian AmeriFlux sites might improve the representation of the network, even within the lower 48 states. Representation could be weighted for the estimated magnitude of the carbon flux expected from a region, or area weighting could be used, so that additional sites are not indicated as necessary in ecoregions that are unusual, but do not occupy much area. Using a "paint-by-the-number" approach, a set of flux-relevant ecoregions might be used as the basis for extrapolating measurements made at existing AmeriFlux towers into a continuous grid of seasonal carbon fluxes across the United States.

Networks of installations like LTER, AmeriFlux, and NEON represent significant investments of research capital. Network analysis, as an outgrowth of the quantitative treatment of ecoregions, provides the first objective guidance for the design, evaluation, and growth of such networks.

References

Baldocchi, D. D., et al., FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux densities, *Bull. Am. Geol. Soc.*, *82*, 2415–2434, 2001.

Hargrove, W. W., and F. M. Hoffman, Using multivariate clustering to characterize ecoregion borders, *Comp. Sci. Eng.*, *1*, 18–25, 1999.

Hargrove, W. W., and F. M. Hoffman, Quantitative methods to extend the ecoregion concept beyond the limits of human subjectivity, *Environ. Manage.*, in press.

Hargrove, W. W., F. M. Hoffman, and T. L. Sterling, The do-it-yourself supercomputer, *Sci. Am.*, *256*, 72–79, 2001.

Law, B. E., and R. H. Waring, Combining remote sensing and climatic data to estimate net primary production across Oregon, *Ecol. Appl.*, *4*, 717–728, 1994.

Law, B. E., et al., Carbon dioxide and water vapor exchange of terrestrial vegetation in response to environment, *Agric. Forest Meteorol.*, *113*, 97–120, 2002.

Wofsy, S. C., and R. C. Harriss, The North American Carbon Program (NACP). Report of the NACP Committee of the U.S. Interagency Carbon Cycle Science Program. 56 pp., 2002. (Available at http://www.esig.ucar.edu/nacp.)

Author Information

William W. Hargrove and Forrest M. Hoffman, Oak Ridge National Laboratory, Tenn; and Beverly E. Law, Oregon State University, Corvallis

For additional information, contact William W. Hargrove, Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, M.S. 6407, Oak Ridge, TN 37831-6407 USA; E-mail: hnw@fire.esd.ornl.gov