Tidy NEON data for biodiversity research

Daijiang Li[1,2†], Sydne Record[3†], Eric Sokol[4†], Matthew E. Bitters, Melissa Y. Chen, Anny Y. Chung, Matthew Helmus, Ruvi Jaimes, Lara Jansen, Marta A. Jarzyna, Michael G. Just, Jalene M. LaMontagne, Brett Melbourne, Wynne Moss, Kari Norman, Stephanie Parker, Natalie Robinson, Bijan Seyednasrollah, Colin Smith, Sarah Spaulding, Thilina Surasinghe, Sarah Thomsen, Phoebe Zarnetske

04 December, 2020

Abstract: Authors of this paper are all interested in using NEON data for biodiversity research. We have spent lots of time reading the documentations and cleaning up the data for our own studies. We believe that we can document our data cleaning process and provide the tidy NEON data for the community so that others can use the data readily for biodiversity research.

Key words: NEON, Biodiversity, Data

Introduction (or why tidy NEON data)

A central goal of ecology is to understand the patterns and processes of biodiversity, which is particularly important in an era of rapid global environmental change (Midgley and Thuiller 2005, Blowes et al. 2019). Such understanding comes from addressing questions like: How is biodiversity distributed across large spatial scales, ranging from ecoregions to continents? What mechanisms drive spatial patterns of biodiversity? Are spatial patterns of biodiversity similar among different taxonomic groups, and if not, why do we see variation? How does community composition vary across geographies? What are the local and landscape scale drivers of community structure? How and why do biodiversity patterns change over time? Answers to such questions are essential to understanding, managing, and conserving biodiversity and the ecosystem services it influences.

Biodiversity research has a long history (Worm and Tittensor 2018), beginning with major scientific expeditions (e.g., Alexander von Humboldt, Charles Darwin) that were undertaken to explore global biodiversity after the establishment of Linnaeus's Systema Naturae (Linnaeus 1758). Modern biodiversity research dates back to the 1950s (Curtis 1959, Hutchinson 1959) and aims to quantify patterns of species diversity and describe mechanisms underlying its heterogeneity. Since the beginning of this line of research, major theoretical breakthroughs (MacArthur and Wilson 1967, Hubbell 2001, Brown et al. 2004) have advanced our understanding of potential mechanisms causing and maintaining biodiversity. Modern empirical studies, however, have been largely constrained to local or regional scales, and focused on one or a few specific taxonomic groups. Despite such constraints, field ecologists have compiled unprecedented numbers of observations, which support research into generalities through syntheses and meta-analyses (Vellend et al. 2013, Blowes et al. 2019, Li et al. 2020). Such work is challenged, however, by the difficulty of bringing together data from different studies and with varying limitations, including: differing collection methods (methodological uncertainties); varying levels of statistical robustness; inconsistent handling of missing data; spatial bias; publication bias; and design flaws (Martin et al. 2012, Nakagawa and Santos 2012, Koricheva and Gurevitch 2014). Additionally, it has historically been challenging for researchers to obtain and collate data from a diversity of sources, for use in syntheses and/or meta-analyses (Gurevitch and Hedges 1999). This has been remedied in recent years by large efforts to digitize museum and herbarium specimens (e.g., iDigBio), successful community science programs (e.g., iNaturalist, eBird), and advances in technology (e.g., remote sensing, automated acoustic recorders) that together bring biodiversity research into the big data era (Hampton et al. 2013, Farley et al. 2018). Yet, each of these comes with its own limitations. For example, museum/herbarium specimens and community science records are incidental (thus, unstructured in terms of the sampling design) and show obvious geographic and taxonomic biases (Martin et al. 2012, Beck et al. 2014, Geldmann et al. 2016); remote sensing approaches can cover large spatial scales, but may be of low spatial resolution and unable to reliably penetrate vegetation canopy (Palumbo et al. 2017, G Pricope et al. 2019). Overall, our understanding of biodiversity is currently limited by the lack of standardized high quality and open-access data across large spatial scales and long time periods.

There is currently a major effort underway to overcome the issues above. For example, the Long Term Ecological Research Network (LTER) consists of 28 sites that provide long term datasets for a diverse set of ecosystems. However, there is no standardization in the design and data collections across LTER sites. The National Ecological Observatory Network (NEON) is a continental-scale observatory network that collects long-term, standardized, and open access datasets broadly aimed at enabling better understanding of how U.S. ecosystems change through time (Keller et al. 2008). Data collected include observations and field surveys, automated instrument measurements, airborne remote sensing surveys, and archival samples that characterize plants, animals, soils, nutrients, freshwater and atmospheric conditions. Data are collected at 81 field sites across both terrestrial and freshwater ecosystems across the United States and will continue for 30 years. These data provide a unique opportunity for advancing biodiversity research because consistent data collection protocols and the long-term nature of the observatory ensure sustained data availability and directly comparable measurements across locations. Spatio-temporal patterns in biodiversity, and the causes of changes to these patterns, can thus be confidently assessed and analyzed using NEON data.

Reference

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. Spatial bias in the gbif database and its effect on modeling species' geographic distributions. Ecological Informatics 19:10–15.

Blowes, S. A., S. R. Supp, L. H. Antão, A. Bates, H. Bruelheide, J. M. Chase, F. Moyes, A. Magurran, B. McGill, I. H. Myers-Smith, and others. 2019. The geography of biodiversity change in marine and terrestrial assemblages. Science 366:339–345.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West. 2004. Toward a metabolic theory of ecology. Ecology 85:1771–1789.

Curtis, J. T. 1959. The vegetation of wisconsin: An ordination of plant communities. University of Wisconsin Pres.

Farley, S. S., A. Dawson, S. J. Goring, and J. W. Williams. 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. BioScience 68:563–576.

Geldmann, J., J. Heilmann-Clausen, T. E. Holm, I. Levinsky, B. Markussen, K. Olsen, C. Rahbek, and A. P. Tøttrup. 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Diversity and Distributions 22:1139–1149.

G Pricope, N., K. L Mapes, and K. D Woodward. 2019. Remote sensing of human–environment interactions in global change research: A review of advances, challenges and future directions. Remote Sensing 11:2783.

Gurevitch, J., and L. V. Hedges. 1999. Statistical issues in ecological meta-analyses. Ecology 80:1142–1149.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11:156–162.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography (mpb-32). Princeton University Press.

Hutchinson, G. E. 1959. Homage to santa rosalia or why are there so many kinds of animals? The American Naturalist 93:145–159.

Keller, M., D. S. Schimel, W. W. Hargrove, and F. M. Hoffman. 2008. A continental strategy for the national ecological observatory network. The Ecological Society of America: 282-284.

Koricheva, J., and J. Gurevitch. 2014. Uses and misuses of meta-analysis in plant ecology. Journal of Ecology 102:828–844.

Li, D., J. D. Olden, J. L. Lockwood, S. Record, M. L. McKinney, and B. Baiser. 2020. Changes in taxonomic and phylogenetic diversity in the anthropocene. Proceedings of the Royal Society B 287:20200777.

Linnaeus, C. 1758. Systema naturae. Stockholm Laurentii Salvii.

MacArthur, R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton university press.

Martin, L. J., B. Blossey, and E. Ellis. 2012. Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. Frontiers in Ecology and the Environment 10:195–201.

Midgley, G. F., and W. Thuiller. 2005. Global environmental change and the uncertain fate of biodiversity. The New Phytologist 167:638–641.

Nakagawa, S., and E. S. Santos. 2012. Methodological issues and advances in biological meta-analysis. Evolutionary Ecology 26:1253–1274.

Palumbo, I., R. A. Rose, R. M. Headley, J. Nackoney, A. Vodacek, and M. Wegmann. 2017. Building capacity in remote sensing for conservation: Present and future challenges. Remote Sensing in Ecology and Conservation 3:21–29.

Vellend, M., L. Baeten, I. H. Myers-Smith, S. C. Elmendorf, R. Beauséjour, C. D. Brown, P. De Frenne, K. Verheyen, and S. Wipf. 2013. Global meta-analysis reveals no net change in local-scale plant biodiversity over time. Proceedings of the National Academy of Sciences 110:19456–19459.

Worm, B., and D. P. Tittensor. 2018. A theory of global biodiversity (mpb-60). Princeton University Press.