Finding the signal in the noise: predicting variability across the scales of life

Matthew Gage (matthew_gage@uml.edu, protein biochemistry and biophysics), Catherine Hulshof (cmhulshof@vcu.edu, plant ecology), Michele Johnson (mjohnso9@trinity.edu, evolution and animal behavior), Danielle Levesque (danielle.l.levesque@maine.edu, eco-evo physiology), Kira McEntire (kmcentir@trinity.edu, ecological modeling)
Authors listed in alphabetical order

Variability is a fundamental property of life. Proteins, cells, organisms, populations, species, communities, and ecosystems all exhibit variation. Yet as biologists, we most often focus on the mean. While this is a convenient way to summarize a group, it is one that ignores potentially meaningful variation. Each biological discipline is grappling with the emerging realization that variability is important, particularly due to our increasingly variable climate, but we lack a framework to connect variation at multiple scales. As a result, the causes and consequences of variability across scales are relatively unknown. As technological advances allow us to better quantify and estimate variability of biological systems, we will be able to determine the role that variability has played in the evolution of life and how future changes in variability may impact the biology of all living systems on Earth.

The ability to determine the importance of variability at one scale, and what influence it has on other scales, crosses disciplinary divides and unifies biological principles and theories. Biological systems operate under basic fundamental principles, which allows scientists to make predictive models that can be applied to systems where it might not be possible to make direct measurements. For example, temperature can be used to speed up a reaction that might take place on a timescale that is not reasonable to measure, and in situ measurements of photosynthesis in a relatively small number of sites around the world are used to model global carbon flux. Further, several biological fields have successfully decomposed variability across scales, including climate science and community ecology. Climate science uses wavelet analyses to decompose climatic variability across different temporal scales (i.e., seasonality, trends, and noise). Community ecology partitions organismal trait variation from individuals to ecosystem to disentangle the contribution of trait variation and species turnover to large scale patterns of biodiversity across space. The tremendous advances made possible by embracing variability in these areas suggests that there is much to gain by a more global synthesis of variability.

One of the challenges in being able to develop meaningful models is our general inability to translate knowledge developed at one level of biology to other levels of biology. One of the roadblocks for doing this successfully is handling variability in a given system, especially as we translate across scales. A small amount of variability at a lower level may become amplified at higher levels of organization. For example, variation in animal body size may predict variation in locomotor speed both within a population, and across evolutionary lineages. Alternatively, variability at one scale may be independent of variability at other scales, such as when multiple alleles of a gene have no measurable effect on the phenotype. In other words, variability at one level of biology may be dampened at a different level due to compensatory mechanisms. Nevertheless, efforts to identify patterns of variability across biological scales cannot be improved by simply collecting additional empirical data. We do not currently have the tools or understanding to accurately predict how variability will change as a function of biological scale.

Our inability to handle variability as a function of scale is a fundamental roadblock to building predictive models, so it is important to develop the tools and procedures that allow us to understand this problem. We propose four interconnected approaches that can be used to close these gaps. First, we can expand the application of existing tools so that they can be used to measure similar biological events at multiple levels. Second, we can develop new methods that are independent of biological scale. Third, we need to significantly improve our ability to distinguish meaningful variation from measurement error. Finally, all of the approaches described above will require a cultural shift encouraging collaboration between scientists with fundamentally diverse expertise.

Applying existing tools: The data era has revolutionized biology. From personalized medicine and genetic engineering to real-time remotely sensed observations of Earth processes, the increased availability of data allows us unprecedented opportunity to measure variability across the tree of life and across hierarchical levels of biological organization. Although artificial intelligence, machine learning, image recognition, and other technological advances make it easier than ever to process large amounts of data, we argue that these tools have not been fully utilized for measuring variability. On the contrary, the use of machine learning to identify species from millions of wildlife camera traps or from billions of specimens digitized from biodiversity collections disregards within-species variability. In contrast, facial recognition systems detect subtle differences in facial features and were built on the assumption that variability matters. Applying advanced technological and modelling tools to the study of how variability scales will require genuine collaborations between biologists, statisticians, computer scientists, and data scientists (see *Collaborations* below).

Developing new methods independent of biological scale: For understanding how variability scales, there is much to learn about complex systems that are scale invariant (fractals, power laws, and temporal noise). As we understand these principles more thoroughly, we will be able to develop more methods that are independent of scale. Statistical analyses are also increasingly able to incorporate both mean and variability into their calculations. This will aid in distinguishing noise from real variation (see *Distinguishing meaningful variability* below).

It may also be possible to adapt technologies that work at one scale so that they can work at other scales. For example, CRISPR technology allows a reporter to be attached to a protein of interest to study function inside a cell without requiring overexpression of the protein. This allows *in vitro* studies to be translated into whole cell studies to validate if behaviors observed in the test tube occur *in vivo* as well.

Finally, if a unifying metric, such as information content or energy, could be identified that can be meaningfully measured across scales, there would be a broader array of techniques or technologies that may be applied to study variability in such a metric across scales.

Distinguishing meaningful variability from uncertainty: We define variability as the quantitative or qualitative difference(s) between two or more samples. This definition is purposefully broad because the scales of study and measurement methods differ widely across biology and thus the margin between variability and uncertainty (or, noise) differs. Variability refers to inherent variation in the real world, while uncertainty refers to measurement error, lack of data, or model assumptions that convert the real world into a quantitative metric. The difference between variability and uncertainty may not scale across all levels, which is why variability is challenging to measure. Further, variability may be adaptive, maladaptive, or neutral. The differing selective contexts of variation may have different implications for how variability scales and whether the signal can be distinguished from the noise. Additionally, a technological barrier may exist to accurately measure variation across scales. With improvements in measurement techniques, we hope it will become easier to differentiate between variability and noise.

Facilitating collaborations: Measuring variability and understanding how variability scales across levels of biological organization seems like a daunting task. However, some of biology's greatest achievements have been a result of unique collaborations across disciplines. The Human Genome Project leveraged international collaboration and (almost instantaneous) data sharing to read and record the entire sequence of the human genome. The Intergovernmental Panel on Climate Change similarly depends on international collaboration and dissemination of scientific information for assessing climate change. Yet biology training programs rarely integrate the science behind successful collaborations, or data management and sharing. Fostering creativity and collaboration should thus be a priority for funding agencies and institutional leaders.

Conclusion

Understanding variability is the business of Biology, and there is an array of tools that have been developed by each discipline related to the measurement of variability. However, these tools do not transcend levels, which makes it difficult to develop models that can make accurate predictions between levels. As we work to reintegrate biology, our ability to extract meaningful connections out of observed variability at different levels of biological organization will lead to new and more powerful predictive models. We will also likely find connections between levels that have been overlooked because of the lack of correlation. While most biological experiments are designed to probe or understand the variation between samples, local variations within a sample or data set can often be seen as an annoyance or a challenge to be overcome. These data points are often viewed as outliers and discarded. However, as we have tried to highlight in this vision paper, those points typically considered outliers might actually be critical to our understanding, and it is therefore valuable to develop tools to allow us to identify important signals, even when they occur at low frequency. It is also critical to develop methods that will allow us to understand the universal principles of variation that cross all biological scales. An understanding of the causes and consequences of variability in biological systems will aid in improving our ability to predict within and across scales. The ability to determine the level of importance in the variability at one scale, and what influence it has on other scales, will help set priorities in the level of data needed to be collected at each scale. A better understanding of the determinants of variability in one study or system could be leveraged to extrapolate to the potential for variability in new systems, or systems in which variability is difficult to quantify.

Key References

Dillon ME, HA Woods. 2016. Introduction to the Symposium: Beyond the mean: Biological impacts of changing patterns of temperature variation. Integr Comp Biol 56: 11-13. https://academic.oup.com/icb/article/56/1/11/2363329

Hadfield, J. D., and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23:494-508. https://doi.org/10.1111/j.1420-9101.2009.01915.x

Katz, R.W. and Brown, B.G., 1992. Extreme events in a changing climate: variability is more important than averages. *Climatic change*, *21*(3), pp.289-302.

Moran EV, E Hartig, DM Bell. 2017. Intraspecific trait variation across scales: implications for understanding global change responses. Global Change Biology 22: 137-150. https://www.ncbi.nlm.nih.gov/pubmed/26061811

Peiman KS, BW Robinson. 2017. Comparative analyses of phenotypic trait covariation within and among populations. Am Nat 190:451-468. https://www.ncbi.nlm.nih.gov/pubmed/28937814

Raffard A, F Santoul, J Cucherousset, S Blanchet. 2019. The community and ecosystem consequences of intraspecific diversity: a metaanalysis. Biol Rev Camb Philos Soc 94: 648-661. https://www.ncbi.nlm.nih.gov/pubmed/30294844

Violle, C., Enquist, B.J., McGill, B.J., Jiang, L.I.N., Albert, C.H., Hulshof, C., Jung, V. and Messier, J., 2012. The return of the variance: intraspecific variability in community ecology. *Trends in ecology & evolution*, *27*(4), pp.244-252. https://doi.org/10.1016/j.tree.2011.11.014

Williams TD. 2008. Individual variation in endocrine systems: moving beyond the “tyranny of the Golden Mean.” Philos Trans R Soc B Biol Sci 363:1687–98.