



doi:10.24908/iee.2018.11.12.e © 2018 The Author Received 4 December 2018; Accepted 5 December 2018

Editorial

A comment on computational biology and connecting the dots

Christopher J. Lortie

Christopher J. Lortie (lortie@yorku.ca), Department of Biology, York University, 4700 Keele St. Toronto ON Canada. M3J1P3 and The National Center for Ecological Analysis and Synthesis (NCEAS). UCSB. Suite 300, 735 State St., Santa Barbara, CA. 93101

Abstract

Increasingly, big data, coding, and quantitative methods contribute to contemporary ecological and evolutionary endeavours. This is not in opposition to effective ideation nor does it play to the false dichotomy of theory versus data. Computational expeditions with data, models, simulations or any other number of approaches both expand the toolkit of science and promote more structured reasoning. The implications of computational biology integrated with scientific pursuits such as experiments and theory development include the following positive outcomes: enhanced open science, better reproducibility, data literacy, author inclusivity, social good, and novel ideation opportunities. We face a climate apocalypse and unprecedented ecological challenges of collapsing ecosystem functions. Computation coupled with ideation is one mechanism to align the hearts and heads of scientists and decision makers alike.

Introduction

Humans do science, not computers. Theoretical and empirical tools do not mix. Increasingly, the lines blur between these 'almost always false dichotomies but worthwhile debates' on how to augment the many facets of scientific inquiry (Shrader-Frechette and McCoy 1990, Zeller and Carmines 1980). Computational biology is the perfect intersection—or eye of the storm—for changes in contemporary ecology and evolution. Computational biology is quantitative and at times atomic (Nussinov 2013) because it can illuminate broad patterns for

biologists from very fine scales (Leonard 2007, Rosse et al. 1998). Contemporary computation biology has become more integrated into the broader community of biological research (Nussinov 2015) through big data, models, and other computational tools that examine or support the examination of biological processes. All biology is likely computational biology to some extent "because it is the questions you address that matter, not the tools you use" (Markowetz 2017), and thus, in ecology and evolution, we are unified and overlap in our shared focus on the dynamics and interactions within natural systems. This is an exceptionally practical opportunity to do better science. Powerful ideas are needed to inform our survival through the climate apocalypse and unprecedented ecological challenges of collapsing ecosystem functions. It is germane to these crises that we include (some) computation in our ecological and evolutionary expeditions, literal and digital, because there is wisdom in statistics (Stigler 2016), logic in literate coding (Grolemund and Wickham 2016), and generalizable skills from a computational toolkit (Allesina and Wilmes 2019, Wilson et al. 2017) for all the environmental challenges we face (Cash et al. 2003). We swim in a sea of data in ecology and evolution now too (Hampton et al. 2013), and computational thinking can help use, own, manage, navigate, and reuse evidence (Di Minin, Tenkanen, and Toivonen 2015, Gandomi and Haider 2015, Zook et al. 2017). If nothing else, consideration of computation in our endeavours can force us to examine our objectivity and respective value judgements at different stages of the research process (Rudner 1953) and simplify the clutter of spurious from necessary in doing our research. A modern synthesis

including the strengths of computation in a paradigm of ecology and evolution is now viable—ideas, data, and coding map onto one another and provide a richer representation of natural systems and change. Each tool can provide a different lens for evidence and improve reasoning—i.e., the difference between novel and creative ideas is the criterion of utility (Runco and Jaeger 2012). We must connect the dots, and we can no longer afford to silo our research activities. The dots to connect are different people with various skills, diverse ideas, many tools, and fundamental scientific evidence to shape societal decision-making. I propose that there are at least six (6) critical outcomes for ecology and evolution if we embrace this perspective.

Outcomes of a connected paradigm

(1) Enhanced open science

Open science is the movement to make scientific products and outcomes transparent and accessible (Hampton et al. 2015, Lortie 2017). These products can be associated with every stage of the scientific process including open notebooks, data, code, and peer review. Open science can also promote open source tool development by the community, broad social networks, and accessible educational materials. Computational biology can contribute to these products directly for ecology and evolution through code, data storage, and peer review of tools and approaches prior to the paper. It can further engage domain-level scientists with a wider network of input from researchers exploring data reuse, efficiency, tool development, and models at both atomic and macro scales. Importantly, transparency in the process of moving from ideation and hypothesis development to analyses and hypothesis testing reduces the likelihood of hypothesizing after the results are known ("HARKing") and bias in ecology and evolution (Kerr 1998), and computation can provide mechanisms to capture this workflow.

(2) Better reproducibility

The reproducibility crisis in science is fundamentally a limited capacity to replicate experiments (Ioannidis 2005), and it has been realized by many disciplines including psychology (Makel et al. 2012), cancer research (Benjamin et al. 2017), and ecology (Kelly 2006). Admittedly, irreproducible findings do not necessarily mean false findings (Fanelli 2018); there are different forms of replication (Nakagawa and Parker 2015, Reed 2018), and many solutions can address this crisis such as publishing false findings and registered reports (Reed 2018). Nonetheless, computation also provides a crucial solution. The ability to review and

rerun code supporting a reported finding in a paper is a monumental achievement only possible through computational tools. Annotation and decisions made in handling data when completed via programming by researchers in ecology and evolution is a means to show they work even if it is not easily repeated. This accountability is even more important than replication because science is a process, and it evolves over time.

(3) Data literacy

Critical thinking is any mode of thinking that includes analysis and assessment preferably with a focus on precision, implications, and supporting assumptions (Cederbloom and Paulsen 2006, Facionie 2017). Hypotheses can be evaluated by logic and errors in likelihood given first principles and understanding of systems (Loehle 1987), but merit assessment of experiments and the critical thinking we apply to the tests of hypotheses in ecology and evolution further assumes that the data and design from empirical work are valid. Necessarily, being able to consider data critically (Hardin et al. 2015) and statistically (Good and Hardin 2003) advances better science, and computational models are heuristic because we strive for a reasonable representation of natural systems by building them from data (Patterson and Whelan 2017). Use of computational tools advances data literacy because we need to be able to read data as humans (and as machines) and comprehend significance to build models.

(4) Author inclusivity

Patterns in authorship are an important form of metascience that can help identify biases and changes in ecology and evolution (Edwards et al. 2018, Fox et al. 2018). Cognitive diversity is a relatively new form of diversity for team science composition focused on the inclusion of individuals with unique thinking approaches. Teams comprised of greater variation in this specific form of diversity are more creative (Shin et al. 2012) and more rapid in solving problems (Shin et al. 2012). Collaborations between computational and field ecologists for instance likely enjoy many of the same benefits including an extended range and scope of questions. Inclusivity is also advanced through computation in ecology and evolution because increasingly common tools in our field such as R and Python programming languages are free. GitHub is a software development platform that is also free, and it has now become a website and collaboration tool that provides version control, online capacities to share work, and project management. To some extent, this alleviates barriers to access advanced data and statistical tools because of cost but also because free computation tools and services

iee 11 (2018) 87

enable a diverse and distributed population of scientists to engage with the thinking associated with statistics (Hector 2017) and data science (Lortie 2017).

(5) Social good

Hackathons, data derbies, codefests, and kaggle challenges are common in computation. This culture of collaborative and crowd sourced problem solving is not dissimilar from citizen science in ecology and evolution but is typically very solution driven. Computation can provide a big picture, quantitatively (Markowetz 2017), and synthesis science is a critical form of advancement in ecology and evolution (Carpenter et al. 2009; Hampton and Parker 2011, Sidlauskas et al. 2009). A continued scientific focus on the benefit to the largest number of people in significant ways through primary and synthetic research engenders better decision making, and computation is a natural pathway to aggregate and summarize natural processes for social good. Purpose defines change.

(6) Novel ideation opportunities

Inspiration and insight are personal. Some ideate and embrace design thinking (Brown 2008, Dorst 2011) in our field through experiential interactions with natural systems whilst others are more abstract. Understanding, observ-ing, ideating, and testing can be expeditions in natural history, experimentation, or computation via models and simulations. Big ideas that make connections and help solve mysteries can come from computation (Nussinov 2015). Gaps in data (O'Leary et al. 2017), true versus false zeros (Martin et al. 2005), and the 'residuals' of exploratory data analyses can all ignite new ideas. At times, seeing the forest from the trees involves counting all the trees, crunching the numbers, and discovering the forest. Computation also allows us to prototype ideas and hypotheses through training models, simulation, and threshold analyses. Knowing the limits of good ideas is an important outcome for ecology and evolution.

Implications

We need to be holistic. We never could ignore the big picture, and data literacy with statistical reasoning are more critical than ever as the public struggles to reconcile evidence of global change with political discourse. There does not have to be a disconnect between evidence and decisioning or the heart and the head or the ecologist and the computational biologist. Predicted scenarios of change are computational. Ecology and evolution can inform many of these models, and we need both experiences with nature and with data to educate and inspire others and affirm what we see, what we know, and what we care about.

Acknowledgments

This synthesis was supported by NSERC DG to CJL and synthesis support by NCEAS.

References

- Allesina, S., and M. Wilmes. 2019. Computing Skills for Biologists: A Toolbox. Princeton, New Jersey: Princeton University Press.
- Benjamin, D., Mandel, D.R., and J. Kimmelman. 2017. Can cancer researchers accurately judge whether preclinical reports will reproduce? PLOS Biology 15 (6): e2002212. CrossRef
- Brown, T. 2008. Design Thinking. Harvard Business Review 86 (6): 84–92.
- Carpenter, S.R., Armbrust, V.E., Arzberger, P.W., Chapin, F.S., Elser, J.J., Hackett, E.J., et al. 2009. Accelerate Synthesis in Ecology and Environmental Sciences. Bioscience 59: 699–701. CrossRef
- Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D.H., et al. 2003. Knowledge systems for sustainable development. Proceedings of the National Academy of Sciences 100(14): 8086. CrossRef
- Cederbloom, J., and D. Paulsen. 2006. Critical Reasoning. Belmont, CA: Wadsworth.
- Di Minin, E., Tenkanen, H., and T. Toivonen. 2015. Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science 3. CrossRef
- Dorst, K. 2011. The core of 'design thinking' and its application. Design Studies 32(6): 521–532. CrossRef
- Edwards, H. A., Schroeder, J., and H.L. Dugdale. 2018. Gender differences in authorships are not associated with publication bias in an evolutionary journal. PLOS ONE 13 (8): e0201725. CrossRef
- Facionie, P. 2017. Critical thinking: What it is and why it counts. Insight Assessment: California Academic Press.
- Fanelli, D. 2018. Opinion: Is science really facing a reproducibility crisis, and do we need it to? Proceedings of the National Academy of Sciences 115 (11): 2628. CrossRef
- Fox, C.W., Ritchey, J.P., and C.E.T. Paine. 2018. Patterns of authorship in ecology and evolution: First, last, and corresponding authorship vary with gender and geography. Ecology and Evolution 1–16. CrossRef.
- Gandomi, A., and M. Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management 35 (2): 137–44. CrossRef

iee 11 (2018) 88

- Good, P.I., and J.W. Hardin. 2003. Common Errors in Statistics (and How to Avoid Them). New Jersey: John Wiley & Sons, Inc. CrossRef
- Grolemund, G., and H. Wickham. 2016. R for Data Science. Canada: O'Reilly Media.
- Hampton, S.E., and J.N. Parker. 2011. Collaboration and productivity in scientific synthesis. Bioscience 61: 900–910. CrossRef
- Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., et al. 2015. The Tao of open science for ecology. Ecosphere 6(7): 1–13. CrossRef
- Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., et al. 2013. Big data and the future of ecology. Frontiers in Ecology & the Environment 11: 156–62. CrossRef
- Hardin, J., Hoerl, R., Horton, N.J., Nolan, D., Baumer, B., Hall-Holt, O., et al. 2015. Data science in statistics curricula: Preparing students to 'think with data'. The American Statistician 69(4): 343–53. CrossRef
- Hector, A. 2017. The New Statistics with R. Book. Oxford: Oxford University Press.
- Ioannidis, J.P.A. 2005. Why most published research findings are false. PLoS Med 2. CrossRef.
- Kelly, C.D. 2006. Replicating empirical research in behavioral ecology: How and why it should be done but rarely ever is. The Quarterly Review of Biology 81 (3): 221–36. CrossRef
- Kerr, N.L. 1998. HARKing: Hypothesizing after the results are known. Personality and Social Psychology Review 2 (3): 196–217. CrossRef
- Leonard, B.P. 2007. The atomic-scale unit, entity: Key to a direct and easily understood definition of the SI base unit for amount of substance. Metrologia 44 (5): 402. CrossRef
- Loehle, C. 1987. Hypothesis testing in ecology: Psychological aspects and the importance of theory maturation. Quarterly Review of Biology 62: 397–409. CrossRef
- Lortie, C.J. 2017. Open sesame: R for Data Science is open science. Ideas in Ecology & Evolution 10: 1–5. CrossRef
- Makel, M.C., Plucker, J.A., and B. Hegarty. 2012. Replications in psychology research. Perspectives on Psychological Science 7(6): 537–42. CrossRef
- Markowetz, F. 2017. All biology is computational biology. PLOS Biology 15(3): e2002050. CrossRef
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., et al. 2005. Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. Ecology Letters 8(11): 1235–46. CrossRef
- Nakagawa, S., and T.H. Parker. 2015. Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. BMC Biology 13 (1): 88. CrossRef

- Nussinov, R. 2013. How can PLOS Computational Biology help the biological sciences? PLOS Computational Biology 9(10): e1003262. CrossRef
- Nussinov, R.. 2015. Advancements and challenges in computational biology. PLOS Computational Biology 11 (1): e1004053. CrossRef
- O'Leary, B.C., Woodcock, P., Kaiser, M.J., and A.S. Pullin. 2017. Evidence maps and evidence gaps: Evidence review mapping as a method for collating and appraising evidence reviews to inform research and policy. Environmental Evidence 6 (1): 19. CrossRef
- Patterson, E.A., and M.P. Whelan. 2017. A framework to establish credibility of computational models in biology. Progress in Biophysics and Molecular Biology 129: 13–19. <u>CrossRef</u>
- Reed, W.R. 2018. A primer on the 'Reproducibility Crisis' and ways to fix it. Australian Economic Review 51 (2): 286–300. CrossRef
- Rosse, C., Mejino, J.L., Modayur, B., Jakobovits, R., Hinshaw, K.P., and J.F. Brinkley. 1998. Motivation and organizational principles for anatomical knowledge representation: The digital anatomist symbolic knowledge base. Journal of the American Medical Informatics Association 5: 17–40. CrossRef
- Rudner, R. 1953. The scientist qua scientist makes value judgments. Philosophy of Science 20 (1): 1–6. CrossRef
- Runco, M.A., and G.J. Jaeger. 2012. The standard definition of creativity. Creativity Research Journal 24: 92–96. CrossRef
- Shin, S.J., Kim, T-Y., Lee, J-Y., and L.I.N. Bian. 2012. Cognitive team diversity and individual team member creativity: A cross-level interaction. Academy of Management Journal 55 (1): 197–212. CrossRef
- Shrader-Frechette, K.S., and E. McCoy. 1990. Theory Reduction and Explanation in Ecology. Oikos 58: 109–14. CrossRef
- Sidlauskas, B., Ganapathy, G., Jenkins, K.P., Lapp, H., McCall, L.W., Price, S., et al. 2009. Linking big: The continuing promise of evolutionary synthesis. Evolution 64: 871–80. CrossRef
- Stigler, S.M. 2016. The Seven Pillars of Statistical Wisdom. Book. Cambridge, Massachusetts: Harvard University Press. CrossRef
- Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and T.K. Teal. 2017. Good enough practices in scientific computing. PLOS Computational Biology 13(6): e1005510. <u>CrossRef</u>
- Zeller, R.A., and E.G. Carmines. 1980. Measurement in the Social Sciences. the Link Between Theory and Data. New York: Cambridge University Press.
- Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S.P., et al. 2017. Ten simple rules for responsible big data research. Plos Computational Biology 13 (3). CrossRef

iee 11 (2018) 89