FutureArctic - beyond Computational Ecology

Steven Latre, Dimitri Papadimitriou * Ivan Janssens, Eric Struyf, Erik Verbruggen †
Ivika Ostonen ‡ Josep Penuelas § Boris Rewald ¶ Andreas Richter ‖ Michael Bahn **

Abstract: This paper presents the Future Arctic initiative, a multi-disciplinary training network where machine learning researchers and ecologists cooperatively study both long- and short-term responses to future climate in Iceland.

1. Introduction

Computational ecology recognizes ecological systems as complex and adaptive. It places a large emphasis on mathematical methods and tools that can handle, or even require, a certain degree of stochasticity. Second, it understands that data are the final arbiter of any simulation or model; this favors the use of data-driven approaches and analysis. Finally, it accepts that some ecological systems are too complex yet to be formulated in mathematical or programmatic terms. Depicting ecology as a field in which mathematical and data-driven modeling do not interact much, and instead the research work develops separately, Levin (1) suggested that ecology should move towards combining them. Subsequently, the development of statistical models (7) and numerical approaches (3) based on multivariate statistics have shown some ability to explain data. However, this combination yields models that very rarely do give new predictions. One of the reasons is that current empirical and experimental research is often lacking in integration of multiple ecosystem compartment interactions: most project-based analysis is restricted to well-established system boundaries, current paradigms and hypotheses based thereon. Even if this approach has moved forward the scientific knowledge on the functioning of individual ecosystem compartments, it has strong limitations as shown by the current in-ability of models to predict whole ecosystem carbon balances.

Consequently, despite decades of research, scientists still struggle to accurately determine the scale of future carbon export (2) essential information to understand how climate change will affect Arctic and Subarctic ecosystems. Key questions such as "How much carbon will Arctic emit under future climate conditions?" and "How do the multitude of ecosystem processes, driven by plant activities and growth, microbial activities and soil characteristics, interact to determine soil carbon storage capacity?" remain unanswered. Colin Prentice, leading climate researcher of the Imperial College of London, worded it in a provocative but also a confronting way: *Our current ecosystem-climate interaction models are better at predicting the past than the future.*

In this paper, we present the Future Arctic initiative (`http://www.futurearctic.be`), a multi-disciplinary training network where machine learning researchers and ecologists study jointly the effects of climate change in Iceland.

2. Experimental Evidences: the ForHot site

The ForHot site (`www.forhot.is`) in Iceland offers a geothermally-controlled soil temperature warming gradient, where Subarctic ecosystem processes are affected by temperature increases as expected through climate change (6). Due to sub-soil lava streams, the soil on the ForHot site is

* IDLab - Department of Mathematics and Computer Science, University of Antwerp - imec, Sint-Pietersvliet 7, 2000 Antwerp, Belgium `{steven.latre,dimitri.papadimitriou}@uantwerpen.be`
† Department of Biology, University of Antwerp, Antwerp, Belgium
‡ Physical Geography and Landscape Ecology, University of Tartu, Tartu, Estonia
§ CSIC, Global Ecology Unit CREAF-CSIC-UAB, Bellaterra, Spain
¶ University of Natural Resources and Life Sciences, Vienna, Austria
‖ Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria
** Department of Ecology, University of Innsbruck, Innsbruck, Austria

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

heated from 0 up to +50°C. The site consists of two different zones: a long-term zone, which is known to be exposed to such heats for centuries and a short-term zone, which has only been heated up to 8 years, due to a recent earthquake. This availability of a large range of temperature gradients, at multiple time scales, is unique in the world, as they enable studying both long- and short-term responses to future climate change now.

To exploit this unique site, a common network of permanent sampling plots has been set up, along a soil temperature increase from +1°C to +10°C, both in the long- and short-term zone. Observations collected over the last 5 years from the ForHot site diverge from the results expected in the most recent ecosystem models. These observations point to the existence of small-scale ecosystem feedbacks between ecosystem factors (either biotic or abiotic) and climate that are currently unidentified by projections based on Earth System models (4). Thus, although current models are usually well-enabled to explain the past, intricate climate-feedbacks at the ecosystem scale significantly complicate the forecasting of the future.

It is now recognized that machine learning techniques provide massive potential to transform ecological understanding (5). However, machine learning has so far not been adopted to identify complex interactions that influence ecosystem scale carbon-fluxes. The challenge is nevertheless significant: data acquisition has to be achieved across a gradient of expected soil temperature changes in the highest possible frequency. Observation data from interaction processes do not necessarily verify the working assumptions of standard machine learning techniques. For instance, empirical data samples do not necessarily verify the independent and identically distributed (i.i.d.) assumption when training/calibration datasets and testing/running datasets do not follow the same distribution. Related measurement (variables) are affected by random errors that violate key assumptions of ordinary least squares fitting since often non-Gaussian and heteroskedastic (their variance is not constant instead of being unrelated to any of the explanatory/independent variables). Also, applicability of nonlinear dimensionality reduction methods to obtain a low dimensional representation for the data, i.e., uncover a set of points parameterizing the data, gets challenged. In addition to the high-dimensionality of the data (of the order of 1000 for some bio-geochemical sets), nonlinear dimensionality reduction techniques have to cope with the nonconvex nature of the parameter space due to the complex topology of the data and/or their incompleteness. By enabling the automatic discovery of complex/hidden patterns and extraction of spatio-temporal features from observational data, these advances in machine learning would in turn steer significant qualitative and quantitative progress in characterizing the interactions that influence ecosystem scale carbon-fluxes.

3. Methodology - The FutureArctic training initiative

Starting from the ForHot research area, the FutureArtic initiative is articulated around a set of activities that combine research on key unknowns in both current ecosystem science and machine learning for a mutual reinforcement to produce both explanatory and predictive models:

a) Soil and root system functioning: with studies focusing on i) the assessment of plant roots growth (e.g., production, turnover) and changes in the timing of seasonal events (phenology), and ii) soil microbial community physiology, composition and functioning, with multiple interactions to assess the root rhizobiome-microbiome. Specific synergetic interactions will be realized by developing new imaging technology for root growth assessment and for the identification of root taxa by hyperspectral imaging.

b) Plant functioning: with a specific focus on i) plant and vegetation traits, community composition and interaction with environmental controlling variables, ii) plant phenology and plant stress adaptation/evolution to climate change, and iii) the development of novel unmanned aerial vehicles (UAV) hyperspectral assessments.

c) Ecosystem carbon balance and the effects of warming on ecosystem-level CO_2 fluxes by means of i) detailed CO_2 exchange measurements, ii) permanently coupled gas-flux chamber-lysimeter analyzers, and iii) focus on the ecosystem source of soil CO_2 emissions and seasonal variability through isotope analysis.

Databases of i) root-rhizobiome-microbiome interactions, ii) ecosystem carbon balance and associated carbon source, and iii) ecosystem productivity, plant stress and soil-plant metabolome and elementome, will be used as input to identify complex interactions between multiple compartments that influence ecosystem scale carbon-fluxes and predict their evolution.

References

[1] Levin SA, Towards the marriage of theory and data, Interface Focus, 2(2):141-143, 2012.

[2] Euskirchen et al., Ecosystems, 20(5):960-974, 2017.

[3] Legendre and Legendre, Numerical Ecology, 24, 1998.

[4] Reichstein et al., Deep learning and process understanding for data-driven Earth system science, Nature, 566, 2019.

[5] Ryo and Rillig, Ecosphere, 8(11):e01976, 2017.

[6] Sigurdsson et al., Icelandic Agricultural Science, 29:53-71, 2016.

[7] Warton, Model-based thinking for community ecology, 216(5):669-682, 2015.