Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-321-RC1, 2020

Interactive comment on "Multi-scale assessment of a grassland productivity model" *by* Shawn D. Taylor and Dawn M. Browning

Anonymous Referee #1

Received and published: 13 October 2020

The manuscript "Multi-scale assessment of a grassland productivity model" uses the PhenoGrass model to evaluate vegetation cover predictions among sites grouped by ecoregions and vegetation types in North America (focusing on the lower 48 US). The main insights include that the model performs poorly when applied for across ecoregions and across vegetation types, but performs well for non-desert ecoregion x grassland combinations. Phenological studies make important contribution to climate change impact science and also to climate change science by providing long-term observed records. Understanding and modelling what drives phenological patterns in different regions and ecosystems and how these are responding to climate change are important scientific question with direct applications for land management, farming, and forestry. I found the discussion on why the model may have performed poorly for shrubs or in agricultural settings (8.4-9.15; all numbers refer to page.line) and how the model may be improved particularly interesting and a great addition to the manuscript.

General comments

- Data and code are made available online, but I have not tested whether I can re-create the analysis.

- Terminology of forecasts

* page 1.line 1 ("forecasting . . . in the coming decades"), 1.10, 1.20, and throughout: I would welcome a more careful representation of what exactly is implied with forecast claims, e.g., "this work allows us to perform long-term forecasts" (1.10). While I agree that the term "forecast" is used quite generally (e.g., White et al. 2019), it can also be interpreted more specifically for quantitative predictions (e.g., Clark et al. 2001), e.g., weather forecasts. I interpret that the claims made here include time-frames over which a considerable amount of climate change is continuing to occur and thus entail "projections" under specific climate scenarios which then "provide an indication of possibility" instead of "definitive probabilities" (Clark et al. 2001). To make this distinction clear, the climate change community including IPCC do not talk about "predictions"– instead, they are "projections", e.g., quote from the glossary of the AR5 (IPCC 2014 p. 1451): "A climate projection is the simulated response of the climate system to a scenario of future emission or concentration of greenhouse gases and aerosols, generally derived using climate models. Climate projections are distinguished from climate predictions by their dependence on the emission/concentration/radiative forcing scenario used, which is in turn based on assumptions concerning, for example, future socioeconomic and technological developments that may or may not be realized. See also Climate scenario."

* This study does not appear to evaluate/discuss the capability of the model to transfer in time or under climate change type conditions. Thus, the conclusion that "this work allows us to perform long-term forecasts" (1.10) appears to be not based on results

and insights generated by this study.

- Confusing terminology: spatial scale, iteration, spatial extent and spatial resolution (grain) appear to be used as interchangeable, but see, e.g., Wiens 1989. I interpret that the study didn't explicitly assess spatial scale or extent or resolution, but rather differences among ecoregions and ecoregion x vegetation type combinations (e.g., Fig. 2, Table 1). This has implications for the framing, discussion, and conclusions (inclusive title).

- Gaps in method section: no definition or details provided on "parametrization" (e.g., p3.3, p3.14, and throughout)

* Does "parametrization" refer to the "estimation of model parameters" or is it rather how other branches of science use it, e.g., climate scientists as defined in the Glossary to AR5 (IPCC 2014): "technique of representing processes that cannot be explicitly resolved at the spatial or temporal resolution of the model (sub-grid scale processes) by relationships between model-resolved larger-scale variables and the area- or time-averaged effect of such subgrid scale processes."

* I would like to know what was parametrized, e.g., number and types of parameters or processes respectively.

* And it would be valuable to know how this was achieved (estimation method or representing structure respectively). For instance, the result section writes in 5.26 "The fitted model, which minimized the mean CVME among the 5 sites"; I interpret that ecoregion x vegetation-type wide optimization pooled data from 5 sites whereby a metric "CVME" (that is not mentioned or defined elsewhere) was minimized.

- Model evaluation (section 2.4)

* What is the impact of unequal sample size (both number of sites and number of years) among ecoregions (Fig. 2) on model evaluation, particularly based on means across sites? It seems possible that a model may appear to perform worse/better

just by chance in ecoregions with fewer data points. Maybe assess sensitivity with rarefaction?

* How exactly was R2 calculated? It seems that there are a number of different possibilities (4.3).

* I don't believe that RMSE (4.3) is a very useful metric to compare different datasets (e.g., comparisons among ecoregions) because mean cover values differ among regions (see Fig. 3). For instance, a RMSE of 0.09 for shrublands cannot be directly compared to the RMSE of 0.16 for grasslands if grasslands have on average higher cover than shrublands. Unfortunately, mean cover values are not reported (e.g., they could be added to Table 1). A normalized form of RMSE (e.g., the coefficient of variation of RMSE = RMSE / mean) would make among-region comparisons easier to interpret.

* Why was the estimate of the "scaling coefficient" not part of the cross-evaluation (4.12)? Depending on the sensitivity of the model to this parameter, the "out-of-sample" estimate may be considerably influenced (as in augmented) by this estimate based on all data.

* I agree that mean R2 and mean RMSE among sites may provide a reasonable estimate of average performance across sites (4.14); however, use cases at specific sites would likely need to also consider expected worst case performance. For instance, Fig. 3 hints at (not possible to know for sure because point density is not shown) that a large number of data points are predicted at 0 cover irrespective of the observed cover value (all but "All Shrubland" model) – this could be driven by to one or a few poorly performing site or by some years, etc.

* A blue line is shown and labelled "Correlation" in Fig. 3. This looks rather like a simple linear regression line than a "correlation" (point estimate)? Was this regression the basis to estimate R2 (see comment to 4.3)?

* Result section (5.8ff): The entire result section focuses on R2 only and does not report on RMSE results – despite method section and figures.

Specific comments

- It is not clear to me why the manuscript argues that the theoretical expectation is that process-based transfer in general worse to new conditions than others, e.g., statistical models (1.21f.). It seems that at least several authors have argued for exactly the opposite expectation, e.g., Grimm and Berger 2016, Radchuk et al. 2019, while others have pointed out real-world limitations while maintaining the theoretical expectation, e.g., that process-based models are often limited because they require large amounts of data and a complete understanding of relevant processes, e.g., Pennekamp et al. 2017, Yates et al. 2018, Bouchet et al. 2019).

- 1.23-2.4 appear to refer to the problematic of overfitting versus building a generalizable model without referring to relevant literature and without defining what exactly is meant by "most accurate". It seems that approaches to minimize overfitting have been frequently discussed in the literature including parsimony (e.g., likelihood-based or information theoretic approaches).

- 2.18: Maybe clarify whether "fractional vegetation cover" includes all vegetation cover combined (whether grasses or not) or estimates are produced separate for grass, shrub, and "agricultural" vegetation types (as hinted at in 3.1)

- 3.7: It seems that the original Daymet product has a 1-km resolution. Is this a typo or was it aggregated here to 4-km. If the latter, explain why and how.

- 3.7: "Climate time series": daily meteorological time series data rather sound like weather data to me

- 3.11: Why use a 20+ year old global soil dataset instead of using one of the many updated and improved ones, e.g., WISE-based HWSD (Batjes 2016), SoilGrids (Hengl et al. 2017) etc., and why not use a regional dataset, e.g., gNATSGO (NRCS 2020),

Printer-friendly version

Discussion paper

POLARIS (Chaney et al. 2019), etc.? How sensitive is the PhenoGrass model to differences in soil variables that occur typically between Global Soil Data Task Group (2000) and others?

- 5.15: Why are the different model fits/parametrizations now suddenly called "iterations"? This is confusing because "iteration" can have a specific and different meaning in parameter estimation/model fitting than what appears to be implied here.

- Fig. 3: Explain what the individual data points are, days pooled from all sites and years?

References - Batjes, N. H. 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma 269:61–68. https://doi.org/10.1016/j.geoderma.2016.01.034.

- Bouchet, P. J., A. T. Peterson, D. Zurell, C. F. Dormann, D. Schoeman, R. E. Ross, P. Snelgrove, A. M. M. Sequeira, M. J. Whittingham, L. Wang, G. Rapacciuolo, S. Oppel, C. Mellin, V. Lauria, P. K. Krishnakumar, A. R. Jones, S. Heinänen, R. K. Heikkinen, E. J. Gregr, A. H. Fielding, M. J. Caley, A. M. Barbosa, A. J. Bamford, H. Lozano-Montes, S. Parnell, S. Wenger, and K. L. Yates. 2019. Better Model Transfers Require Knowledge of Mechanisms. Trends in Ecology & Evolution 34:489–490. https://doi.org/10.1016/j.tree.2019.04.006.

- Chaney, N. W., B. Minasny, J. D. Herman, T. W. Nauman, C. Brungard, C. L. S. Morgan, A. B. McBratney, E. F. Wood, and Y. T. Yimam. 2019. POLARIS soil properties: 30-meter probabilistic maps of soil properties over the contiguous United States. Water Resources Research 55:2916–2938. https://doi.org/10.1029/2018WR022797.

- Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological Forecasts: An Emerging Imperative. Science 293:657–660. https://doi.org/10.1126/science.293.5530.657.

- Grimm, V., and U. Berger. 2016. Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue. Ecological Modelling 326:177–187. https://doi.org/10.1016/j.ecolmodel.2016.01.001.

- Hengl, T., J. M. de Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen. 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12:e0169748. https://doi.org/10.1371/journal.pone.0169748.

- IPCC. 2014. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change. 1535 pages. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

- Pennekamp, F., M. W. Adamson, O. L. Petchey, J.-C. Poggiale, M. Aguiar, B. W. Kooi, D. B. Botkin, and D. L. DeAngelis. 2017. The practice of prediction: What can ecologists learn from applied, ecology-related fields? Ecological Complexity 32:156–167. https://doi.org/10.1016/j.ecocom.2016.12.005.

- Radchuk, V., S. Kramer-Schadt, and V. Grimm. 2019. Transferability of Mechanistic Ecological Models Is About Emergence. Trends in Ecology & Evolution 34:487–488. https://doi.org/10.1016/j.tree.2019.01.010.

- Soil Survey Staff. 2020. Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States. Available online at https://nrcs.app.box.com/v/soils. (FY2020 official release). United States Department of Agriculture, Natural Resources Conservation Service.

- White, E. P., G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, and S. K. M. Ernest. 2019. Developing an automated iterative near-term fore-

Printer-friendly version

Discussion paper

casting system for an ecological study. Methods in Ecology and Evolution 10:332–344. https://doi.org/10.1111/2041-210X.13104.

- Wiens, J. A. 1989. Spatial Scaling in Ecology. Functional Ecology 3:385–397.

- Yates, K. L., P. J. Bouchet, M. J. Caley, K. Mengersen, C. F. Randin, S. Parnell, A. H. Fielding, A. J. Bamford, S. Ban, A. M. Barbosa, C. F. Dormann, J. Elith, C. B. Embling, G. N. Ervin, R. Fisher, S. Gould, R. F. Graf, E. J. Gregr, P. N. Halpin, R. K. Heikkinen, S. Heinänen, A. R. Jones, P. K. Krishnakumar, V. Lauria, H. Lozano-Montes, L. Mannocci, C. Mellin, M. B. Mesgaran, E. Moreno-Amat, S. Mormede, E. Novaczek, S. Oppel, G. Ortuño Crespo, A. T. Peterson, G. Rapacciuolo, J. J. Roberts, R. E. Ross, K. L. Scales, D. Schoeman, P. Snelgrove, G. Sundblad, W. Thuiller, L. G. Torres, H. Verbruggen, L. Wang, S. Wenger, M. J. Whittingham, Y. Zharikov, D. Zurell, and A. M. M. Sequeira. 2018. Outstanding Challenges in the Transferability of Ecological Models. Trends in Ecology & Evolution 33:790–802. https://doi.org/10.1016/j.tree.2018.08.001.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-321, 2020.