

April 5, 2022 Theory Working Group Call

Attendees: Abby Lewis, Elyssa Collins, Amanda Gallinat, Andrew Allyn, Cole Brookson, Steph Brodie, Jaime Ashander, Glenda Wardle, Christy Rollinson, Gerbrand Koren

Regrets: Noel Juvigny-Khenafou, Mike Dietze

Agenda:

1. Announcement - May 23-25 EFI is hosting a Virtual Conference. We are looking for people to present 10 minute talks or posters and/or 30 or 60 minute workshops about a specific task, skill, walk through a forecasting workflow, etc. Details and registration at: <https://ecoforecast.org/efi-2022-conference/>
2. Icebreaker for the group at the start of the call: Spitball ideas for what the group is interested in doing next
 - a. Test some of the hypotheses put forth in the manuscript
 - b. Several versions of figure 1 that lays out forecasting - produce something in a non-manuscript format, but through EFI that lays out underlying theory of forecasting and expand and provide through EFI (not necessarily through traditional manuscripts). Graphical abstract of what forecasting is about
 - c. Test transferability
 - d. Work on/test the toy model that Abby and Elyssa started a while ago
 - e. Use existing experiments to reanalyze with a forecasting approach. Start to understand how to do science differently. Put into practice the idea of how forecasting can change what we are currently doing. Especially some of these experiments that have been done across multiple sites
 - f. Journal club - dive into and discuss papers that cover existing theory. (student group has had fun and productive discussions on complexity and simplicity - could possibly have cross-over call)
 - g. Opportunity for synthesis and to learn from each other. Thinking about different scales of ecological organization and where predictability scales across that. Synthesis across domains (marine, terrestrial, freshwater). Think about different end users - resource managers, ag management, water management. Sounds like some of the ideas in the very first manuscript outline-- comparing predictability across levels of biological organization, geography, life history, etc
 - h. Model validation/model selection using large datasets or past datasets. Transferability
 - i. Use manuscript as starting point to put things into practice to show benefit of forecasts. Like having the manuscript to motivate to read new papers and organize thoughts more than just discussing a topic.
 - j. Working on the uncertainty component of the manuscript - is uncertainty the key part of forecasting or required? Can work to explain the different sources and show how to do it and visualize.

3. Final Items to Discuss

a. Introduction: unresolved comments about *future* predictions

i. 2nd paragraph of Intro

1. Beginning of paragraph follows what other papers have said. So at the end Abby is trying to convey what is different. Forecasting is predicting the future and humans are altering conditions, so we have to do more than predict from historical
 - a. Don't want to leave out forecasting efforts that aren't predicting the future explicitly (hindcasting without predicting future states)
 - b. Retrospective forecast to validate forecast? Or inherent difference in how we define forecasts.
 - c. Example - creating forecast now in 2022 from data from 6 years ago up through Oct 2019 and making prediction for Nov 2019
 - d. Look at ocean warming from past decade and look at warming in the future decade. Pretend we go back to 2010 and pretend we are forecasting forward, but we have all the data and know what happened
 - e. Maybe it is the term: historical modeling, retrospective modeling
 - f. Whether using all the data available and what the goal is - model 1990 from 1980 data. It lays the foundation for the forecasting because you can validate and understand aspects of the model. Could be viewed as calibration. Are we using all available data or are we withholding some information to do assimilation or bias assessment
 - g. From Steph: Two typical terms used to describe this historical prediction/forecast: 'out-of-sample' prediction and 'retrospective forecast'.
 - h. Difference where you have truly withheld the data and you are blind to the decade you are forecasting (e.g., 2010-2020) and you build a model up to that point to use the first 10 years to get a sense of how it works and then get ready to forecast forward (e.g. 2020-2030) vs. reconstructing posthoc when we do exploratory testing. What would it look like to do a strong test with data that is withheld (could be a form of data that we couldn't get before)
 - i. Could resolve in the text by using terms that are common - out of sample prediction (only fitting model with subsample of data and testing with withheld data) use forecast field in retrospective
 - j. Perhaps its getting technical, but I'm thinking of the historical modeling that is state-space based rather than

something that's more regression-based or descriptive in nature where you do it all in one go

- k. Retrospective forecasting - often used interchangeably with hindcast, but hindcast means something different in the climate field. Generally the idea if you have a forecasted system, instead of predicting next week and waiting until next week to validate it, go back to the past and make prediction two weeks ago for one week ago and see how the prediction did since we have the data
- l. Benefits of forecasting that we describe are that it is forced to be out of sample. Limits to predictability could be tested with retrospective forecasting. So may be helpful to clarify that. Seems like there is a subset of things we talk about that is tied to true forecasting. Could add clarification in Sec 2.2 Increasing reproducibility. Fits with the not overfitting argument of that section.
- m. If anyone has suggestions for this section now or after meeting feel free to add to this section

b. Box 2

- i. Box 2 - is it useful? Are there ways to adjust it?
 1. Think there are significant edits based on Jono's comments.
 2. The discussion around empirical and mechanistic models can be its own paper and we treat it pretty lightly even though it is in a box.
 3. Add statements that we assume process based models predict outside of the range of observed conditions (but not necessarily true in all cases)

c. Figure 2

- i. Figure 2 (pg 14) - give examples for each scale listed on the x-axis. Is this figure useful/needed? The hypothesis it is trying to show is that predictability increases with ecological aggregation.
 1. We've seen from lit from other disciplines and from recent forecasting papers that predictability increases with increasing aggregation, but we don't know what the different aggregation there is.
 2. Add curve shapes to the figure? What if it logistic or exponential? That is what Abby was getting with in the current figure but the lines look like CI right now. Can change the color, weight, etc of the lines
 3. We don't know what affects the shapes. Can we add what we can get

- a. Have a hypothesis that gets at the pattern that we see in this figure, but it doesn't get at the "why". Not sure if there is lit to support the different mechanisms. One example is life history traits - longer lived species have increased predictability. Could be reflected by age, size, and something else. But doesn't get us
- b. Steph looking at figures from papers - linear relationships for species length for trophic level, but not for age.
- c. This is good for generating ideas. E.g., saturating line puts an upper limit on predictability - why
- d. If we introduce the paper with the tentativeness and as a suggestion for how to think about it. We are not coming with a fully formed answer for theory. We should think about it and here are some processes for doing this.
- e. Could we use the plots that Steph has referenced for what is on the x-axis for what is in the lit and then have dashed lines for other shapes? Then could show examples of shapes of the curves that are published and then what other curves are
- f. Could also add other x-axis variables that just haven't been studied yet.
- g. Abby will take a shot at that and Steph will send the papers she is thinking about

d. Wording questions:

- 1. Forecast Skill vs something more general such as Forecast Performance. Some people talk about skill as the performance of your forecast vs performance of a null model.
- 2. Do we keep it as forecast skill or use a more general term?
- 3. Steph prefers skill. But doesn't feel strongly. If you just say performance then could be model performance. If using forecast performance have to use both words. Whereas if you say skill that reflect forecast skill. Think skill is more widely accepted in forecasting community
- 4. Andrew - doesn't have strong feeling as long as it is defined. So don't have skill, proficiency, performance, and predictability
- 5. Abby had comment from collaborator about this who sent a paper about it.
 - a. Maybe I missed it, but do we define forecast skill vs. accuracy and are we using skill properly throughout? My understanding is that skill is how well a forecast compares to a null model, so I'm not sure RMSE/R² would count as skill, but definitely defer to Quinn and others if I'm misunderstanding! E.g., see WaPo article ;-P

- b. https://www.washingtonpost.com/blogs/capital-weather-gang/post/weather-forecast-accuracy-versus-skill-skill-is-what-matters/2012/04/05/gIQA42BixS_blog.html
 - 6. What do people feel about accuracy?
 - a. Think it is the way to define skill. So perhaps interchangeable.
 - b. Gerbrand - think skill is preferred term. Think performance is a technical term that relates to how the model runs.
 - 7. Next step: Abby will take look at papers on this and is planning to go with skill or accuracy (no immediate objections from the group)
 - ii. Definition of ecological forecasting and including uncertainty- think we are coming to consensus.
 - 1. Are we could to say “preferably including uncertainty” - group was on board with this
 - iii. Section 4.2. Hypothesis 2.
 - 1. Aggregation vs scale
 - 2. Mike thinks scale is more sexier
 - 3. Christy - aggregation is more specific and descriptive than scale and gets more at
 - 4. Jaime - scale is sexier, btu more ambiguous
 - 5. Could go with aggregation across scales
 - 6. Glenda - picks aggregation if we had to pick one
 - 7. Amanda/Steph - aggregation
 - e. Is everyone ready to submit? Do we need another version?
 - i. Abby will continue to work on making edits and will check in with people who wants to see the next version
 - f. Next tasks: 2-3 volunteers for a final read through
 - i. Cole, Amanda, Jody will help with final copy editing. Jaime also offered to help with copy editing
 - ii. Road map section - Abby needs to take time think about reframing and if anyone wants to have another look, let her know.
- 4. Timeline (DUE April 15th)
 - a. **April 5 Meeting**
 - b. April 5–15 ASL resolve comments, ping people as necessary
 - i. **2–3 volunteers to do a final read through of MS and cover letter**
 - c. April 15 ASL submit