December 7, 2022 Theory Working Group Call

Attendees: Christy Rollinson, Noel Juvigny-Khenafou, Abby Lewis, Cole Brookson, Jody Peters, Shubhi Sharma, Carl Boettiger, Mike Dietze, Gerbrand Koren

Agenda:

BIG PICTURE AGENDA:
15 minutes of discussion on the overall idea (Abby push to keep things moving at 2:30)
10 min of discussion on big-picture tasks that need to be accomplished to answer the research questions we have
10 min on nitty gritty first steps
10 min dividing up tasks, scheduling intermediate meetings for small groups if relevant

1. Notes from Today
    a. Dig into the NEON Challenge. Generating set of hypotheses and use the Challenge to test the hypotheses
    b. On previous calls we had talked about generating our own forecasts. But started to move away from this to 2 other ideas
        i. Use a theoretical framework from Mike's 2017 Ecological Applications paper to dig into Hypotheses 1 to compare predictability across the NEON themes. See why one system may be more predictably than another
            1. Use simulations to do an uncertainty analysis
            2. This is something we can do right now while we wait to get the forecasts up and running for point 2 below
        ii. Create forecasts that we can analyze. Abby interested in this and is setting up ARIMA models for the 5 forecasting challenge themes
    c. What framework to use to answer hypothesis 1 - why is one system may or may not be more predictable than another
        i. Should the model be processed based or not. Leaning towards moving away from process based models.
        ii. What does it mean to be good at making predictions? Multiplicity of ways to look at the effectiveness of predictions. Good at short horizons, but not at 30 days out. Phenology good at make predictions except for at bud burst. Carbon models predictions can look good at small scales, but when added up, then they do poorly.
        iii. What is the right way to score
        iv. People use ML and tune them without looking at probabilistic scores.
        v. How to look at what is better - in which way. Have that rooted in theory.
        vi. Need framework to parse these questions. Want to get started and get simulations running to look at uncertainty at an arbitrary horizon.
        vii. Could test different models or analyze data submitted to the Challenge

viii. Are forecasts also available for download? Yes, they are easy to download. Can't Google for the data, but can get 2 lines of code (which is in the phenology repo and the theory repo) to get the data.

ix. There is an R package neon4cast which has code for combining scores method or can name theme and get scores for the theme

x. Are there papers, metrics, frameworks that people like for thinking about predictability

1. Shubhi has been reading papers about information theory that she thinks may be useful
2. How have people quantified predictability in the past?
3. Permutation entropy as a way to measure predictability of a time series. Could be worth exploring this more
4. Interesting to compare permutation entropy (or other metric) that claims to estimate predictability of a time series with different sources of uncertainty when running the simulations
5. Based on the idea that if your new timestep brings in new information that is not redundant then it will be less predictable. Could put this in simulations and see how it works when forecasting a highly stochastic vs cyclical process
6. But ignores different complexity - seasons have different patterns, different lengths of horizons
7. Is the metric applied to the time series of the data itself or is it applied to the model or the forecast? It is applied to the time series.
    a. It is based on the likelihood. So whatever stats model you are using, it will have an effect on it.
8. Mike's lab has done analyses - think about predictability in that it tells about how skill changes as function of lead time. For phenology broke it down by day of year. Breaking down uncertainties as function of lead time and day of year. Percentage uncertainty attributed to each of 5 bins. So can see predictability of phenology in the winter is very easy and is influenced by different types of uncertainty compared to time of green-up
9. Christy - trying to ignore the models and thinking about predictability. Think about the uncertainty in the observations in the data assimilation or what is forecast and use that variability. Wonder if how much changes from observation to observation can be a way to frame questions about uncertainty. Ex. there are no leaves on the tree, so the uncertainty will be small. So thinking about predictability in terms of change
10. Cole's question - What characteristics could be applied to a set of simulations - you aren't just interested in phenology, but interested in a yearly cycle. Where along the gradient is helpful to focus?

a. Don't think it is just about the next time step, but multiple time steps. Forecasts over a certain time horizon
b. How to make it a comparative study, given that each system behaves differently
c. Think about what are the properties of the data. Those are bugs or carbon vs fast changing things or slow changing things. How to translate everything into widgets (fast or slow, big or small) instead of bugs and carbon

xi. Papers shared in chat
1. https://esajournals.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/ecm.1359
   a. the paper looks interesting. They compare a measure of uncertainty in the time-series itself (permutation entropy) to RMSE, a skill score for point-forecasts which themselves don't have uncertainty.
2. https://doi.org/10.1016/j.ecolind.2020.106113

d. Coming back to the big idea of the analysis: Want to do comparative analysis of predictability using an uncertainty framework to decompose how predictable something is and why or why not
   i. Nuances - what do we mean by uncertainty and at what level (spatial, temporal, aggregation)
   ii. Start with temporal analysis and do a comparative analysis
   iii. Would not use forecast submissions. Use simulations
   iv. Simulations aren't the goal. The analyses are. What would those analyses be?
   v. Once we come up with the framework/protocol, then apply it to the forecasts submitted to the Challenge
e. Inputs/Outputs of the protocol/framework
   i. Purpose - have some measure (or multiple measures) of predictability that come out of forecasts. Output is some sort of scoring metric of how well the model does
   ii. Mike: Uncertainty partitioning to not just assessing skill, but to analyze the uncertainty partitioning - what are the dominant uncertainties across lead time and day of year and how does it map on to how skill maps on to lead time and day of year. Once you have those dimensions in one system, then do it for k different systems (at the same site or across sites). Then ask what is consistent what is idiosyncratic?
   iii. If you were to extrapolate to the k+1 systems, what could you say about how well the forecast will do. If we can answer this than we can answer what is predictable or not
f. How are we defining predictability?
   i. Is it predictable given the current suite of models or is it something inherent?

        ii. Is it fair to evaluate the predictability of a system based on the model used?
            1. Also tied to the question is if the framework/score is model specific or model agnostic?
        iii. We're interested in the predictability of variables in general.
            1. Want to say that dissolved oxygen is more predictable than beetles. But it is hard to determine/measure. Have different proxies to measure this and start to get at this question. All of the potential ways to approximate predictability is useful
            2. Can we disentangle it from the type of model used?
    g. How do we build something/think about predictability when there is limited data (one beetle collection) vs when there are tons of data?
        i. Long amount of data collection? Lots of data collected?
        ii. Interesting to take hyperconceived data sets and remove data so the data availability is comparable and look at models at that scale - this is the reverse of what we have been talking about
        iii. Cole's suggestion: Can we use the simulation approach to compare models by taking multiple approaches. Is it that we are missing data or is it that the model is a poor reflection of what to expect?
    h. Does anyone want to get involved with the coding and setting up simulations - will have smaller group meetings to work on this
2. Abby shared a figure that shows team science and collaborations through time. There is the groan zone, which is what she thinks the group is experiencing. As you develop collaboration have initial phase with lots of ideas, then hit the groan zone where there are new ideas that can be helpful, then eventually through processes become more convergent

3. RCN June 21-23, 2023 Unconference at NEON HQ Save the Date
    a. Bringing People Together to Do Forecasting: Training, Technology, Theory, and Translation
    b. The goal of the Unconference is for participants to work together to produce a product such as getting a forecast up and running, developing teaching materials, finalizing tutorials, refining or creating tools, analyzing forecasts for a manuscript, and/or developing visualizations. The event will also include a poster session for attendees to present their research.
    c. Space is limited to 50 people. Short application will be available Jan 1 and due Feb 1. Applicants will be notified Feb 15.
    d. Travel funds will be available for >30, <50

4. What hypotheses from the manuscript could be explored within one or across the NEON Forecasting Challenge themes or with forecasts listed on the EFI forecasting profiles webpage or from the EFI community
    a. Hypothesis 1: The rate of decline in predictability over increasing forecast horizons differs across variables and scales

b. Hypothesis 2: Predictability increases with biological and ecological aggregation