September 11, 2023 Theory Working Group Call

Attendees: Cole Brookson, Caleb Robbins, Abby Lewis, Kathryn Wheeler, Alyssa Willson, Jody Peters, Shubhi Sharma, Bilgecan Sen, Gerbrand Koren
Regrets: Freya Olsson

Agenda:

1. EFI 2024 Conference Announcement
   a. Use this short feedback form to submit proposals for workshops, panels, short courses, socials before/after/during this conference.
   b. Deadline for proposals : 01 Nov 2023
   c. Registration and Abstract submissions open: 01 Dec 2023
   d. Accepted contributions are announced: 01 Mar 2024
   e. Final program announced: 01 Apr 2024

2. Original questions that led to the creation/submission of multiple forecasts across all the NEON Forecasting Challenge themes and the exploration of intrinsic predictability (Abby, Shubhi)
   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
   c. Figures developed with expectations from these hypotheses

3. Preliminary results related to these questions

4. Manuscript co-leads - Shubhi, Caleb, others?
   a. Could be one or two manuscript - could do comparative predictability of models across challenges and intrinsic predictability of models and why some did better than other
   b. Forecast Challegne submissions were looking at which forecasts are most forecastable and how does the decline in predictability over time differ
   c. Intrinsic predictability
      i. Novelty is combining intrinsic predictability work which has been recently been published in ecology with the uncertainty decomposition work. How predictable is your forecast and how much can you gain with model development?
      ii. Hoping to put together a manuscript looking at the 2 frameworks in combination with each other with ARIMA or time series models to look at what the metrics.
         1. Hard to look at intrinsic predictability with the ML models
      iii. Is the simulation approach explain results or motivating a question?
         1. Could do both.

2. Challenge for intrinsic predictability to explain why patterns jump out in empirical work will then have hard time to explain patterns.
3. Is the goal to use the simulations to support the empirical work?
   a. Depends on what the group wants to do
4. Cole finds it compelling to use the intrinsic predictability simulations to remove the messiness found in the NEON Challenge themes.
5. Shubhi, Cole have been working with clean, close form process model to get parameters back from (as compared to the NEON Challenge themes).
6. Simulations gives freedom that focusing on the Challenge does not.
   a. But think having an empirical example makes the simulation in a paper stronger.
   b. Would like to use a simple model for this
   c. Being able to constrain the uncertainty decomposition using ARIMA is appealing

iv. Papers looking at intrinsic predictability and uncertainty decomposition would be a novel contribution
v. Bilgecan has been thinking about the hypotheses from a population ecology perspective. Thinking about scale
vi. Get realized predictability from the forecasting model, this doesn't tell you much. R2 and RSME tells you about your model, but not predictability. Intrinsic predictability would contribute to understand the models predictive performance. It will also tell you potential background processes that you are trying to model. Munch et al. have used it to identify chaos (intrinsic predictability was in the top 3 of the methods to identify chaos in the Munch paper).
vii. Pennekamp 2019 paper
viii. Bilgecan is currently writing a paper about is your time series more predictable than white noise.
ix. Could look at it across NEON Challenge themes - see what is dominated by stochasticity. The goal is to find the scale in time/space where it is predictable
x. If you have daily or monthly time series that spans multiple years and multiple seasons. You'll see seasonality in the patterns, but there will be lots of variability if looking overall. But there are specific time periods where you can make predictions, so get the WPE or moving window average for specific time periods in the forecast.
xi. Use WPE this to guide you in building the model.
xii. Calculating WPE is easy, compared to building the model.
xiii. Uncertainty framework reveals which part of ecological variable is the noisiest. This reveals whether collecting more data will help improve understanding of the system or if your process is stochastic.

1. Uncertainty decomposition gets at the why

xiv. Thinking about what time step to forecast to maximize predictability is really interesting. Haven't thought as much about this.

xv. Thought about questions of scale originally. Didn't come to any conclusions to that during the initial brainstorm. But would be interesting to apply to the simulations easier than applying it to the Challenge because the scale in that is already in place.

d. Who is leading what, what capacity, what timeline?

i. Abby is dissertating. Excited about submitting models and has been working with Freya on this. Excited to help out with projects, but not able to lead at this point.

ii. What part of forecast submissions is Caleb most excited about and what capability to lead things?

1. Really interested in the relation between realized and intrinsic predictability - asking what is predictable, but underlying question is what does predictable mean

2. Driver data for ML no initial conditions, nor process error

3. What do we need to change to get the models that ask the questions we want to - can do the hindcasting

4. Caleb interested in leading a manuscript. Most interested in aquatics, but pushing on the general aspect is

a. Not committed to one vs two manuscripts

5. Like the idea of one manuscript and linking realized and intrinsic predictability. But there is the aspect of what is low hanging fruit. For example, getting the uncertainty included in ML will take longer to do

iii. Depending on people's interest and time could see two topics. 2 approaches to answer a similar but different question that have overlapping components.

1. Could focus on the realized predictability with reference to calculations of intrinsic predictability which is a small portion of the work that Shubhi and Cole have done. Realized predictability confronted with some intrinsic predictability

2. Uncertainty analysis that decomposes different uncertainties and tie that to intrinsic predictability which would have some analyses from the Challenge forecasts. Simulations confronted with some data.

iv. Shubhi's capacity - tied up until end of October. So can't contribute for the next couple of months

v. Cole's capacity - tied up until October

vi. Time management of manuscript from Abby's experience - want meetings make significant progress while continuing to be open to new people joining projects

        1. Nice thing about realized predictability - can look at the data as it comes in. Anyone can jump in and contribute
        2. Harder when it gets to the writing and editing portion of the project
    vii. If we wanted to focus an analysis on the year of 2023, then get as many models up and running as possible and the fill in with hindcasts as needed.
    viii. Next steps - Caleb draft up specific questions to analyze, methods, workflows. Put it in a Google doc for the group to give feedback on.
    ix. Shubhi and Cole have an outline that we can look at at the next meeting
    x. Next meeting - write down proposed outline
        1. Have 2 proposed outlines to see concrete goals
    xi. Bilgecan is very interested in helping out

5. Should we schedule a hackathon or collaborative working time as part of the monthly calls or separately to push ideas/projects forward?
    a. Potential working group funding opportunity from ESIIL (Environmental Data Science Innovation and Inclusion Lab)
        i. **Working Groups**
            1. In line with our mission to foster collaboration across a broad range of disciplines related to environmental data science, ESIIL is accepting applications for **working groups**. These groups will play a pivotal role in promoting the integrative approach ESIIL champions, pooling knowledge and expertise from various disciplines to tackle environmental issues with a data-driven perspective.
        ii. **What are working groups?**
            1. Working groups are self-organized research teams focused on well-defined scientific questions that advance environmental data science and require insights from a diverse group of researchers and other stakeholders. A single working group may have up to 15 participants and a quorum (50% or more) shall meet in person up to 2 times over a 2 year period, with each meeting lasting between 3 and 5 days.
        iii. **Important Dates**
            1. Proposals for working groups must be submitted in PDF format to esiil@colorado.edu by November 1, 2023. Funding decisions will be announced by early 2024, with anticipated start dates of Working Groups as early as February 2024.
            2. Please read the attached Request for Proposals or visit ESIIL's website for more information.

6. GitHub repos
   a. GitHub repo: eco4cast/predictability - comparative analysis of predictability
   b. GitHub repo: Forecast_submissions - forecasts submitted to the Challenge

7. Model Development for the NEON Challenge
   a. GitHub repo: eco4cast/Forecast_submissions
   b. New model descriptions document