November 13, 2023 Theory Working Group Call

Attendees: Freya Olsson, Caleb Robbins, Abby Lewis, Marcus Lapeyrolerie, Alyssa Willson, Bilgecan Sen, Jody Peters, Hassan Moustahfid, Kathryn Wheeler, Saeed Shafiei Sabet
Regrets: Shubhi Sharma, Cole Brookson

Agenda:

1. Code Review Update (Abby, Jody)
    a. This topic came up on the last call. It isn't directly related to the main discussions the group has had, but was something that was of interest
    b. A subgroup of individuals met together since the last call to talk about developing a formal review process
    c. Guidelines were developed from: Implementing code review in the scientific workflow: Insights from ecology and evolutionary biology. https://onlinelibrary.wiley.com/doi/full/10.1111/jeb.14230
    d. Abby has developed a rubric for an Advanced R course that she is leading
    e. https://github.com/abbylewis/Advanced_R/tree/master/class_exercises
        i. 05a, b, and c is the overview
        ii. Checklist for the class
        iii. Self assess and then another person also does the assessment
        iv. The class assignment did not have a chance to think about potential updates for the rubric
    f. Notre Dame also updated the rubric with the view that it would be used mainly by grad students
        i. Also talked about co-authorship and the potential to be co-author on a paper that was reviewed and the opportunity for helping other grad students with the knowledge that other grad students will help review code
        ii. Caz will be testing Alyssa's code for a paper in the next few weeks, so the rubric may get updated
    g. Another approach to co-authorship. If you are already a co-author, then you should be helping to review code.
        i. This works well in a lab where it is common for people in the lab to be co-authors
    h. Other resources:
        i. Cracking the code review process- https://www.nature.com/articles/s43588-022-00261-w
        ii. https://uwescience.github.io/neuroinformatics/2017/10/08/code-review.html

2. Manuscript Outline - using the NEON Forecasting Challenge to explore predictability across variables and scales (Caleb)
    a. Last time talked about methodology
    b. Re-ran a bunch of forecasts using the random forest models
    c. Worked on developing a workflow

i. Trying to estimate how forecasts decline over time. Aggregate by site to see how forecast change for different months
ii. Have plots of the R2 for the forecast horizon for the different months for the different variables
    1. Rolling window issue - it takes an average of the forecast horizon for an entire month.
    2. The forecasts look linear - not too surprising GAMs collapse back to a linear pattern
    3. There is a lot of site variability. If we look at R2 at the sites, they are spanning a large range across the sites.
iii. There is a convergence of chlorophyll a, water temp, and redness which is weird
    1. Need to check with Gavin Simspon or Nick Clark about the GAM modeling - think there is something going on there
iv. Also the increase in R2 over 30 days is hard to think about why that would be happening
v. For each forecast horizon used to calculate the R2 is there difference in sample size?
    1. Equal sample size across forecast horizon
    2. If forecast is made on Jun 30, the 30 day horizon is in Jul, but it ends up in the Jun plot
vi. Can you account for the number of sites - there are different number of sites for the different variables
    1. It should be accounting for that
    2. Uncertainty in some of the forecast is interesting. Some have a rapid increase in uncertainty
        a. Separate question we may want to go down
    3. Have you looked at the skill of the forecast
        a. R2 is the observed vs expected
        b. RMSE normalized against a normal. Caleb started to try to do that, but then decided to start with the R2 which seemed more interpretable for a first shot
    4. Marcus has been looking at percent improvement of CRPS over the historical/climatology model
        a. Have that in the code to compare across the persistence model
    5. R2 is a great starting point.
        a. R2 as a metric is nice to be able to compare across variables and doesn't depend on a null model persistence vs climatology
        b. A summary plot not broken down by month would be useful.
        c. Certain variables may be more forecastable at t+1 time point and then may change over the time horizon

d. Month comparison is cool , but separate question.
e. Baseline predictability vs seasonal question that shorter seasons are harder to predict (spring/fall vs winter/spring). For seasons may be useful to use RSME or CRPS
f. Keeping in R2 for the reviewers is good because in Kathryn's experience reviewers don't understand CRPS

6. Weird that R2 increasing for greenness - for example October
    a. When is it easier to forecast throughout the year may depend on the variation
7. Beetle and ticks are not included because of the delay of data processing
    a. But the beetle and tick data go back to 2014 so could make forecasts (that are actually hindcasts)
    b. Would be cool to include in this paper, but if not could look at the differential availability of data will influence the predictability - may not be included in this paper, but could be included in Shubhi and Cole's paper
8. Is there a reason to do the GAM specification instead of plotting the raw R2?
    a. It is a massive amount of data that is really messy and hard to interpret
    b. Cool to see both the raw R2 and the GAM
9. Summary plot of R2 for all horizons for each variable may be interpretable if we don't do the months.
10. There should only be one point per horizon per variable
11. Are there a handful of sites that have all the themes?
    a. They don't have the same site ID but they are paired
    b. Could you bin by the NEON domains/ecoregions

vii. Repo where Caleb is working on this: https://github.com/robbinscalebj/NeonPredictability

d. Next meeting - To do: walk through the dashboard to see how the forecasts are doing to see the raw data and forecasts

3. Save for December: Manuscript Outline - Uncertainty analysis that decomposes different uncertainties and ties that to intrinsic predictability which would have some analyses from the Challenge forecasts. Simulations confronted with some data. (Shubhi and Cole)

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

5. Model Development for the NEON Challenge
    a. GitHub repo: eco4cast/Forecast_submissions

b. New model descriptions document

Previous Notes and Links for Reference

6. 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

7. Resources the group pulled together to test hypotheses
    a. Google sheet with a summary of drivers, data availability, number of sites, etc for the Challenge themes
    b. Lit review of models typically used for the NEON Forecasting Challenge themes
        i. Here is a google doc to compile the models
    c. Figures of hypotheses that can be examined using the forecast challenge output
        i. Google slides with images
    d. GitHub repo with code that lets people drop in models to create forecasts for the challenge: https://github.com/abbylewis/EFI_Theory