August 14, 2023 Theory Working Group Call

Attendees: Mark Lowell, Jody Peters, Abby Lewis, Cole Brookson, Bilgecan Sen, Caleb Robbins, Jonathan Borelli, Shubhi Sharma, Mike Dietze
Regrets: Freya Olsson

Agenda:

1. Check in with Caleb about ML work from the Unconference
    a. Goal was to extend the current ML to include python to do deep learning of residuals of the forecasts
    b. Caleb worked with Marcus Lapeyrolerie (student of Carl Boettiger)
    c. Python is good at something that R isn't so workflow includes outputs from both
    d. The upcoming work to continue to work on is working the kinks out in getting Python to work in a container
    e. On last call, talked with Freya about taking submitted forecasts and combine them into ensemble of forecasts. Abby hasn't got to this yet

2. Check in with Mike about uncertainty work from the Unconference
    a. https://github.com/eco4cast/UncertaintyTutorial
    b. Is this work with the models set up for the Theory group or separate? Yes to both.
    c. Thinking about how to apply uncertainty partitioning to different models and how to teach people
    d. Started with one of the theory's example repo and tried to come up with 2 case studies for doing uncertainty partitioning - ARIMA time series model and another that was black box ML, random forest
    e. Didn't finish writing the tutorial, but thinking abotu how to include parameters, etc for a wider range of modeling approaches
    f. Need to schedule a call with this group. They agreed to meet every 3-4 weeks to continue to work on that project.
    g. Once it is done, then go over it with this group to think about how to apply it to a broader subset of models that are being run by the Theory group
    h. Abby willing to join those calls and apply them to the Theory models
    i. If we got models running through different iterations of different parameters, iterations and have ensembles, then could use the tutorial to partition out the forecasts using the SOBOL that is part of the tutorial
    j. The group was working with SOBOL because it is more global
    k. Caleb uploaded a forecast that loops through different types of model fits to get at different estimations of parameter uncertainty. So far it seems to be working, but needs to take another step and will want to run ideas by this group.
    l. Mike's group found that distinguishing parameter uncertainty didn't work well with ML models. But could still distinguish it from the other types of uncertainty

      m. For the ARIMA, Mike coded it manually and used some assumptions. But want to go back and set it up more black box, ML format.

      n. How does SOBOL separate process error and parameter uncertainty?

            i. Think of SOBOL as the classic sensitivity distinction between 1 at a time vs global

            ii. Your answer is sensitive to the order you add uncertainty in

            iii. With SOBOL - all uncertainty is on at all times. You run A ensemble and B ensemble and can borrow an swap things out and by understanding what is going on, then you can look at uncertainty more jointly and globally. Remove the artifacts of order and how things interact when you do it as 1 at a time

            iv. With random forest, take different realizations of fitting. Leverages that ML methods are ensemble based

            v. There were still conceptual issues that the group worked was trying to work through at the Unconference

3. Are there needs for the forecasting challenge that the Theory group can contribute to?

    a. The uncertainty work will be helpful

    b. Getting ensemble over the models that are submitting forecasts up will also be helpful - either doing adaptive weighting

    c. Freya has it turned on for aquatics so far. But seems like it would be low hanging fruit to apply it to all the targets

    d. Either convince Freya to do this or use her repo to set this up

    e. Do ensemble of ML models, time series models, and basic time series models and then do something that combines across all

          i. Could also do this for all process based models from other groups, but getting the metadata will be the hard part of this

          ii. Could do this to contribute to the Challenge, but not sure how well it would contribute to the Theory group questions and analyses

          iii. There are process based models that are being submitted for Terrestrial. Not sure about for phenology or ticks.

    f. Looking at the full suite of models the Theory group is contributing - what would be useful for grouping?

          i. Think about what groupings would be most useful.

          ii. With Freya's grouping - the simple stats models are doing quite well, but adding the process models hedge your bets a bit.

              1. Abby thinks you might see more of these affects when combining across different types of models.

              2. The effect of having ensemble of time series may be less than by combining time series and ML

              3. The more correlate the models are the less effective the ensemble is - this is what Freya presented at ESA

                  a. Freya is working on the aquatics challenge and did an analysis of the forecasts that have been submitted. Then she did analysis of ensemble that combines

      b. Has an analysis that looks at the proportion of time that each model does best
          i. The ensembles that combine across different types of models are not the best, but they are never the worst
          ii. Simple stats model like persistence does well
          iii. Process models generally do a bit worse
          iv. The three mechanistic models created by 3 different teams (e.g., 3 different models) had similar assumptions and had correlation coefficient of 0.9
          v. Think would see something similar for terrestrial models since there is a lot of shared assumptions
          vi. The ensemble averages out the extremes

4. Think about how similar the models are/parameters that are included. Make predictions before creating the ensembles. If the model structures are very similar then the ensembles may not give more information
5. In land modeling have examples where in hindcast mode the ensemble across many process models beat all of them
      a. Think this is that everyone makes mistakes and by ensembling you cancel out the mistakes
6. Freya has ensemble forecasts - if there are 10 ensemble members for a forecast then she grabs 3 and combines it into a new ensemble.
      a. There isn't a weighting scheme. But in other research have found that not weighting works just as well or better than weighting (Mike gave example from previous EFI meetings)
7. Would like to get ensemble predictions for the Theory forecasts
      a. Abby isn't able to commit to do this, but will talk to Freya to see if she can set it up or at least talk through how to do that.
      b. It would be helpful to hear Freya's process on the next call.
      c. Abby will invite Freya for the next call to hear more about this.

4. Comparative analysis of predictability using an uncertainty framework to decompose how predictable something is and why or why not (Shubhi, Cole, Noel, etc)
      a. [GitHub repo: eco4cast/predictability](#)
      b. From the July call, Shubhi and Cole were planning to work on the aquatics theme and will dig into the intrinsic predictability
      c. Cole and Shubhi have been swamped with other projects the past month, hope to have updates for the next call

5. Long term plans
    a. See 1-2 manuscripts coming from the work that is being done by the group
    b. Abby is eager for the manuscripts to be written, but not sure if she can commit
    c. Shubhi happy to co-lead or lead in a couple of months
    d. Intrinsic predictability results would have to be really cool to have it be a separate manuscript. Not sure if it is novel enough to merit its own publication.
    e. Caleb is in a similar boat. Interested in seeing where it will go, but don't think it will go anywhere without a leader. Caleb is happy to help, but doesn't want to step on toes.
    f. **Go back to the questions that was generating the ideas. Could revisit this on the next call given what has taken place. Put this on the Agenda for the next call.**
        i. **If there are preliminary results that apply to the results will also make the questions more tangible**
    g. Plan for now - Shubhi/Caleb (check with Cole as well) to be co-leads
    h. Look at the results and decide how many papers to write.
    i. Check to see if Freya is interested in being in the mix. Abby's impression is that Freya is excited to be involved in co-author role and push it forward.
    j. Are there things that EFI can facilitate?
        i. Find a collective date for a sprint in a location in person
        ii. Sponsoring a hackathon
        iii. Finding time before/after a meeting
        iv. Theory group workshop
        v. Do a sprint to do analysis, get words on page. Have a co-working time to work together.
        vi. Could set aside time at AGU.
        vii. Could have 1-2 small sprints before then and then use AGU to wrap things up.
        viii. Could see about time the week before the next call
        ix. Do people have time for a sprint before the next call? Or use the next call to prioritize for the sprint. Or use the working group as time for the sprint.

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

7. Discussion questions (Jody left this in here from the March call):
    a. Which variables do we expect to be most predictable, why
    b. How do we expect the relative performance of persistence and climatology to differ across variables/themes
    c. Are there certain times of year that we expect to be less predictable? Are these consistent across variables?
    d. How does the level of biological aggregation change predictability?

    e. Do we want to tackle spatial predictability at all? Our conversations so far have been focused on temporal predictability

Previous Notes and Links for Reference

8. 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

9. 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