February 27, 2023 Theory Working Group Call

Attendees: Glenda Wardle, Christy Rollinson, Shubhi Sharma, Abby Lewis, Kathryn Wheeler, Caleb Robbins, Cole Brookson, Noel Juvigny-Khenafou, Jody Peters
Regrets: Jono Tonkin

Agenda:

1. NEON Challenge Forecast Submissions Updates (Abby)
    a. Reminder that the motivation is to use the same model across multiple themes and see if there are differences in the forecasts and why
    b. Updates and Next steps
        i. Submitted 9 models so far (we can add more), they are automatically running for every day for the past month.
        ii. You can see the results for the "tg-" models so far here (click through the tabs to see all the themes): https://projects.ecoforecast.org/neon4cast-dashboard/phenology
        iii. Goal is to compare across models across themes.
        iv. Analyze the realized forecast performance for the different models
        v. Patterns seen so far - which model does best is different across the different themes and whether or not they do better than other models submitted varies widely as well. So there is a difference in forecast performance so far. But still need more forecasts to do have a more robust analysis. The forecasts in Jan-Feb are pretty stable in the US. But the period of change in the spring in the US will be interesting.
        vi. Anyone can create forecasts to run across themes. If you need help plugging in your own model Abby is happy to help.
        vii. Here is the GitHub repo: https://github.com/abbylewis/EFI_Theory

2. Comparative analysis of predictability using an uncertainty framework to decompose how predictable something is and why or why not (Shubhi)
    a. Thinking about this using simulations and time series dynamics
    b. Shubhi/Cole did a lit review about what people in climatology/meteorology have done. Presentation from them:
        i. Information Criterion as a useful form of predictive quantification
        ii. This idea came from the MEE paper where we made hypotheses about prediction and how forecast performance and predictability are related to the forecast horizon.
        iii. Quantifying predictability literature rabbit hole
        iv. How to quantify when asking about forecastability or predictability
        v. How predictable is system X in time and/or space
        vi. Realized vs intrinsic predictability
            1. Intrinsic - the greatest forecastability of a system

2. Realized - predictability that is actually achieved with your model/parameters
3. The difference between the two is how much your model can be improved
4. An event is unpredictable if the forecast distribution is identical to the climatological distribution

vii. Is the system intrinsically unpredictable (has low predictability) or do we have a lot of space to move from our realized prediction to the intrinsic predictability
viii. Information theory
ix. Making comparison across models for 1 system - can use RMSE. But this only works to compare multiple models for 1 system.
x. But what about comparing across systems - the same model for multiple themes or different models to forecast multiple themes
xi. Relative entropy (realized predictability) - estimates when your model is doing no better than a null model
xii. Attempt to measure intrinsic predictability - permutation entropy. How much more predictability can you get with a perfect forecast (how good can I do in theory?)
xiii. Permutation entropy is sensitive to time series quality
xiv. Application to the Forecasting Challenge?
    1. Could apply this approach to simulated data and to the Forecast Challenge
xv. Discussion
    1. Carbon vs beetles example
        a. Thinking about beetles - if trying to predict the true population of beetles. It is a noisy data set. It has to do with the type of sampling. Maybe some measurements/sampling are closer to the true physical process/estimate
        b. Permutation entropy does account for the observation error. So can make an estimate (a prior on your prior) of how good your observations are. The less confident you are about the observations will influence the permutation entropy
        c. The distribution is also useful because you can account for uncertainty
        d. Also thinking about decomposing the different uncertainty estimates - so if it is a data issue, then can get at that from the different sources of uncertainty
        e. When we choose to sample a process poorly, is that intrinsic or realized?

            i. If we sample at a low rate so it is highly noisy then that might be intrinsic. What is external and what is internal
    c. Simulation overview from Shubhi - thinking about it from the information theoretic framework
        i. Climatology is the probability of y, whereas the forecast is the probability of y based on conditions of the past
        ii. Shubhi looked at relative entropy and mutual information (for the call we just focused on relative entropy today)
            1. Relative entropy: How different are 2 distributions from each other
        iii. Simulations - looking at different AR processes
            1. AR1
                a. When relative entropy = 0 we are not learning anything new
                b. At timestep 12 relative entropy is 0
            2. AR3
                a. Less noise and less autocorrelation in te times series.
                b. Relative entropy goes to 0 more slowly - at timestep 33
            3. Whitenoise process
                a. Unpredictable system. Relative entropy is at timestep 4
            4. Data deficiency simulations
                a. Cut training data to ¼ (had 1000, but cut to 250 data points) and randomly sampled data points
                b. When randomly sampled, relative entropy drops to 0 immediately
                    i. Think this might be like what is happening with the beetles
                c. Some NEON time series have chunks of missing data and other times will be more frequent random observations
                d. Beetle abundance is dependent on random walk vs the previous time step - relative entropy does not change between these two.
                e. If your model changes slightly, relative entropy will stay constant. If you add a variable into your model, relative entropy doesn't change.
                f. Would you expect for a more lagged AR process (AR10) would you expect relative entropy to change? Is the amount of data that you take. Is the only thing governing it the lag of the AR process.
                    i. If you have AR10 process but only have the last 6 points and you don't fully capture the dependence, then think you will land in the middle, but you won't get as good of an estimate as you would with the whole structure of the data.

5. Relative entropy = realized predictability
6. The stronger the dependence on past states, the longer the forecast horizon
7. Sensitive to data quality!
8. Next steps
    a. For Forecasting Challenge - use relative entropy to see if we can understand across the challenges, how good is our realized prediction?
    b. Using variance decomposition (this is coming from Mike's 2017 paper) - can we understand why/why not we are able to predict well?
        i. Do you need to explicitly include all the sources of uncertainty? Yes
            1. Currently, Abby is not including all sources of uncertainty for the Challenge models so far. Will add that to the To Do list for the models
            2. Shubhi is hung up on the parameter drivers uncertainty - doesn't see those results anywhere
                a. Abby is doing this in some models because using the NOAA ensembles
            3. If you don't account for the different uncertainties, then it gets lumped in the process error
            4. Does drier uncertainty get pushed to observation error?
                a. Think so
    c. Using permutation entropy - can we theoretically understand how close we could get to intrinsic predictability
9. Discussion
    a. If you know that you have a long time series, but you can use the last 250 data points.
    b. How to balance sampling? Rely on old data? Or need to use the last sampling
    c. What if NEON stops? What would we then be able to do if we pick it up a decade later? Would beetles still behave the same way or would we say we have no understanding of the beetles?
    d. Taken as a given that the climatology is known. But with beetles, a lot of it is capturing the sampling. We can get more from the physical processes, but from the biological information, we don't have an informed prior so have to

work hard to get that first. So any data makes it more predictable the no data at all.

e. Someone should work on this for biological systems - what is a good null model and how do you choose it? The closer the null model is to the actual process the harder it is to be able to make predictions that are better than the null model
    i. If you are challenging your forecast with a tough null model, you could say your forecast isn't good, but your null model could just be really good.
    ii. Caveat - you are only doing better relative to whatever standard you pick.

10. What can the group do that would be helpful for the work Cole/Shubhi are working on?
    a. Plan to apply the framework to the data we have for the NEON Challenge
    b. In the Theory GitHub, they will set up code that anyone can work on. Create models that people can go in an tinker with.
    c. Shubhi will put up the simulation code up. Will also add a couple more metrics
    d. Then apply it to the forecasting challenge

3. Unconference ideas - didn't get to this during the Feb call. Save it for the March call.
    a. Definitely worth going to the Unconference repo and suggesting this project so people who have not heard about what the Theory group is working on can know about this
    b. https://github.com/eco4cast/unconf-2023