March 20, 2023 Theory Working Group Call

Attendees: Abby Lewis, Jonathan Borrelli, Glenda Wardle, Cole Brookson, Caleb Robbins, Noel Juvigny-Khenafou, Shubhi Sharma, Cayelan Carey, Christy Rollinson, Glenda Wardle, Mike Dietze
Regrets: Jono Tonkin

Agenda:

1. NEON Challenge Forecast Submissions Updates (Caleb, Abby)
    a. 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
    b. Goal is to compare across models across themes.
    c. GitHub repo with code to submit forecasts to the Challenge: https://github.com/EFI-Theory/Forecast_submissions (new shared EFI Theory org)
    d. Trying to tune models can be intensive and hard to get GitHub to do it automatically.
        i. For Caleb's machine learning model - Split into 2 steps
            1. Variables are all the predictors of meteorological variables from NOAA
            2. Have the training data.
            3. Using the tidymodels framework. Inputs are treated the same as outputs.
            4. Take input data (site target). Then do something to it. Then update it and say the target variable is the output variable - what it wants to predict.
            5. Searches through grid of lasso parameters.
            6. Uses k-fold cross validation. Trying to balance minimizing variance in predictions with potential overfitting.
            7. For each site selects tuning parameters based on lowest RMSE
            8. Takes the trained model and makes a forecast, looks for a fit based on the site/target variable, makes the predictions, puts it into the forecast format, and spit it all out
        ii. Abby asks how uncertainty is handled in this machine model
            1. Not sure right now, but Caleb wants to look into conformal (?) prediction
            2. Mike wonders if you could save multiple models instead of one model and treat it analogously to a bootstrap
            3. Right now, Caleb says it's just multiple ensemble models that's creating the uncertainty
        iii. Mike curious what the spread of the prediction looks like, suspicion there is an overconfidence in early time

iv. The lasso model is currently doing the best out of any of the Theory Group models
v. Mike says that for skill to be improving over time, you're probably getting penalized for overconfidence in the short term
vi. Chlorophyll - hold off for now. NEON is checking the data and will be updating them with a data release. So they haven't been updated in the Challenge targets data. Freya (from VT) has been talking with Bobby from NEON so they are working on it. But don't expect anything for 1-2 weeks
vii. Next steps for Caleb's models - xgboost has a tidy model framework. Could do something with xgboost. Could also look at variables. Currently using all met data, but could think about other variables.
viii. Forecasts are currently being run at each site. But could do something where it is learning across sites.
ix. Thinking about the computation issues - Is Caleb refitting the model every day or calibrating the model and refitting every day as met comes available? It is the same trained model every day. No new training and fitting. The same model fit to the new forecast. So fit model offline, save the object in GithUb and GitHub actions makes forecast everyday.
    1. Forecasts run fast. There are no problems on the automation side with the computation.
x. Want to get as many models up as possible with the tidymodels framework.
    1. Can anyone help with this? How do we make it happen?
    2. Can we adapt Caleb's code and insert other models into it?
    3. xgboost is a next step. It will be more computationally difficult.
    4. If anyone is interested in propagating uncertainty or uncertainty in ML that is something Caleb would like help with.
    5. Abby wants to get a better representation of uncertainty for her models as well. It meshes nicely with the simulations
    6. Like that Caleb has a record of the model, even though it is static, that is helpful. Abby's models don't have parameters that are saved when the forecasts are made on a given day.
xi. Are there other simple models on our wishlist that haven't been implemented yet?
    1. GAMS - Caleb likes this
    2. Something where we can decompose variance into process, parameter, etc uncertainty
    3. Does Mike have code that lays out how to do variance decomposition for time series?
        a. Yes. Has code online for a one at a time approach.
        b. But moving towards a more global sobol approach
        c. As Mike's student is getting the sobol approach put together then can share this with the group.

d. Does anyone know if Nick Clark been writing dynamic GAMS for the Challenge?
    i. Nick gave a presentation at Glenda's department. When the recording from that presentation is available, she will make it available to the group

xii. Would it make sense to set aside time on one of these meetings to do a short tutorial on how to use the tidymodels framework?
    1. Caleb can do this on the next call.
    2. Highlight where it would be useful for others to make modifications. Caleb will share how to swap out models (after looking at the available tutorials).
    3. https://www.tidymodels.org/find/parsnip/

2. Comparative analysis of predictability using an uncertainty framework to decompose how predictable something is and why or why not (Shubhi, Cole)
    a. Relative entropy - we discussed this on the Feb call. Anything else to follow up on?
    b. Permutation Entropy - we didn't get to this on the last call.
    c. What are the steps for applying the uncertainty frameworks to the Challenge? Are we ready to discuss this or wait until we have developed the framework further?
    d. Noel, Cole, Shubhi met a few weeks ago. Shubhi's next steps - apply the metrics to the forecasts from the Challenges. But wanted to continue to explore the simulations further.
    e. Using population dynamic models as time series and using metrics to quantify predictability.
    f. Reminder that there are 2 types of predictability - intrinsic predictability and realized predictability
    g. Shubhi worked on 2 examples with a stable time series and an unpredictable time series
    h. Logistic map - x = growth rate. Y = population at equilibrium. When growth rate is >2 then in chaotic
    i. Looked at 2 metrics
    j. 1) permutation entropy - measure of intrinsic predictability. Not looking at fittedmodel and how well it is doing. Just looking at time series. 0 = very predictable. 1= chaos. Measure how repetitive a time series it is. The more redundant it is the more predictable it is. When there is a lot of new info at every point then it becomes less predictable.
    k. 2) relative entropy - measure of realized predictability. Calculated relative to anothermeasurure - a flat prior distribution. If your distribution is doing as well as your forecast, then your observations are not informing your next timesetp so system is not that predictable. Used an ARIMA model. When very stable, very predictable. When chaos, less predictable.

l. Overall goal - identify places in a given time series where there is some intrinsic predictability and where we are differing from taht without realized predictability. So it tells us how good the forecast can be and where we have room for improvement.
m. These were deterministic models so would be interesting to add process variance
n. How do these handle lead time? Calculate them by one time ahead?
    i. Permutation entropy is on the observed time series. 1 value for the whole time series. Relative entropy is by time step.
o. Could make another version of the image where the color is lead time. How many time steps into the future are you better than chance? Expect the chaotic domain is a small number of timesteps, but not necessarily 0.
p. Shubhi has only seen a few papers using these metrics in ecology. One from 2017 that said we should pair the permutation entropy with variance decomposition ot get at the question about what is predictable vs not predictable but also why.
q. The y-axes are on unique scales from each other with the 2 examples Shubhi showed today.
r. Is there a way to ask when relative entropy hits permutation entropy? When do they converge?
    i. This would be interesting. Your model is doing the best it can, but no you can't do any better. And if you can decompose the variability then you can know why.
s. This is the sort of stuff that will be great for the Gordon Conference - would be good to have a discussion of this in the breakouts. There are a number of people in the group that are going, so that will be cool.
t. Next steps?
u. Anyone interested in joining discussions about the simulations and relative and permutation entropy should reach out to Shubhi
v. Move the code to the Eco4cast repo: https://github.com/eco4cast
w. An interesting next step on the entropy side would be to calculate the intrinsic predictability of the NEON challenges (which you could do from the targets files). How predictable are the ticks vs fluxes, vs aquatics, etc? Then from there think about calculating the relative predictability for the forecasts that have been submitted. If we prototype it for a couple of models, then could probably get it added to the scoring container. So then it can be another score that is calculated for any forecast submitted.
x. Last thoughts by Shubhi (that Jody didn't quite catch fully) Measure distance with distribution on all possible outcomes vs forecasts. As long as it is a function of the observation (linear combination). Also invariante to transformations, so if you add or take away an extra term you can't game the system for it to do better.
y.

3. Discussion questions:

   a. https://www.neonscience.org/impact/observatory-blog/water-quality-chlorophyll-measurements-assigned-incorrect-units - unit issues for NEON data. How is this being treated for the challenge?
   b. Which variables do we expect to be most predictable, why
   c. How do we expect the relative performance of persistence and climatology to differ across variables/themes
   d. Are there certain times of year that we expect to be less predictable? Are these consistent across variables?
   e. How does the level of biological aggregation change predictability?
   f. Do we want to tackle spatial predictability at all? Our conversations so far have been focused on temporal predictability

4. Unconference ideas - didn't get to this during the Feb 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. Add ideas or comment on current ideas at: https://github.com/eco4cast/unconf-2023
   c. The list of Unconference ideas is meant to 1) provide opportunities for Unconference to participants to work on creating concrete products (could be something simple like recommendations for next steps to achieve a certain goal, or could be the development of code to complete a task, etc) and 2) provide the start to a general wishlist of tasks that would lower the barrier to ecological forecasting creation, analysis, or use.