## September 14, 2022 Theory Working Group Call

Attendees: Christy Rollinson, Mike Dietze, Caleb Robbins, Jody Peters, Cole Brookson, Shubhi Sharma, Hassan Moustahfid
Regrets: Abby Lewis, Noel Juvigny-Khenafou, Jaime Ashander

Agenda:

1. 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
2. Designing experiments testing model predictability and transferability
3. Tasks from the August Call
    a. Contribute to a “lit review” of models that are commonly used (in the broader, non-forecasting world) for each of the NEON challenge variables
        i. Tick populations
            1. Cole added a couple of example from the lit that he found
            2. Is John Foster’s thesis available that could contribute? John is in the process of incorporating edits from his committee. Could email him to get copy to mine for citations. Cole can reach out to John
        ii. Beetle communities
            1. Shubhi added lit
        iii. Aquatic temperature, dissolved oxygen, chlorophylla
            1. Hassan may add example for HAB from Great Lakes
            2. Aquatics is focused on lakes and streams since this is related to the NEON Forecasting Challenge
        iv. Terrestrial carbon and water fluxes
        v. Plant phenology
        vi. Here is a google doc to compile the models
        vii. Predictability and transferability is a rich area of study in terrestrial systems. Shubhi added papers for this under the Beetles section on the Google doc (particularly enjoyed the Yates 2018 paper)
    b. Sketch a figure or two that you would like to be able to make using results from the forecast challenge, then take a look at the forecast standards to see if it would be possible to make that figure using existing standards/metadata.
        i. Figures and discussion can be added to this google slideshow.
            1. Slide 2 - forecasts in time (or space) looking
                a. Predictability is looking at model accuracy in the place where it is trained to. Transferability is looking at accuracy outside the place it is trained.

   b. Starting at time 0 looking at predictability. Then through time it is looking at transferabilty
   c. Show different combination of species that are predictable or transferable or neither
   d. If thinking about making forecasts at a given site out into the future - Mike thinks of this as predictabiltyi. Thinks transferabilyt is predicting in a new place or to a new species.
   e. In Shubhi's diagram, think that climate change is built in so the species see new conditions
   f. The question is at what point at a given location are you talking about transferability rather than predictability. Would need some environmental space to know the model was built on.
   g. Time scale of forecasts make it interesting. If we are talking about climate, talking about long term. If forecast is very short, then won't have chance to inject climate
   h. If talking about climate, will be thinking at least on decadal time scale.
   i. Thinking of testing this - have a data that goes back 30 years, if we started at 30 years ago and hold out all the data collected after that and train model on data from that year. Could be one approach to see how well your forecast is doing and whether it is predictability or transferability
   j. Could look at species on a spatial gradient
   k. Shubhi and Cole talked at ESA about using simulations to test hypotheses
2. Slide 3 - Cole's figure
   a. Thinking about how forecast horizon might change depending on the modeling approach
   b. May see a decline in forecast skill along the forecast horizon by different types of models
   c. Cole hasn't thought about the shape/location of these curves for different models
   d. Describing process models vs other model types - If we as a group are going to pick models and run forecast for them, then we would set up the models. So it is an easy time to for the group to decide how to define a process model
   e. This could be done for any Challenge - it is not Challenge specific
   f. Do we want to look for places/themes we haven't created forecasts for
3. Slide 4 - Mike's figure

a. Conceptual figure Mike has been using. Gets at the same thing Cole is getting at. But axis is changed.
b. Added this figure as a reminder that it is probably the same thing we are talking about
c. Mike has long standing questions about components in the function (initial conditions, drivers & covariates, parameters, random effects, process error)
d. Have a table about predictions from 2017 Eco App paper for these. But hasn't made predictions about what modeling approach would be better
e. Open question how the forecast limit will change for different systems and what uncertainty will dominate in different systems
f. This decomposed uncertainty in one model. Then do it over a bunch of models to see if there are patterns.
g. Assumes that everyone is propagating uncertainty, which they aren't.
h. A lot of the metadata asked for the Challenge is designed around understanding the uncertainty
i. If you choose a different function it will change. Example from John's dissertation - was initially calculating a 3 stage model and all forecasts were dominated by process error, after he added an additional stage the process error dropped dramatically and the other uncertainties came into play. Good example of where the model's error was dominated by process error because had the wrong model. Once there was a better structured model, then more interesting questions could be answerable because the model was no longer dominated by process error.
j. Another example from HAB forecast - has a model that is consistently dominated by process error no matter what is changed.
k. One thing that is hard with process error is distinguishing hard cases, are we dominated by process error because 1) we have the wrong model or 2) is there a lot of inherent stochasticity in the process?
    i. When you add white noise to a model is that one way to disentangle between the two?
    ii. You put stochasticity in the model when you believe it is key to the process. On the back end, if fit a model that has a process error term and it is large, how do I know it is because stochasticity is key part of the process or I have the wrong model.

            iii. If you have competing models then you can tell what structures are better

4. Slide 5 - from Mike
    a. One of the ancillary variable asked for in the metadata is that if you have included one of the five types of uncertainty, how many do they include
    b. Can be # of parameters, # of initial conditions, # of drivers
    c. Classic model selection parsimony hypothesis
    d. Haven't thought through for the 5 NEON themes where will the most parsimonious model lie? What types of problems need a complex model to be predictable? What can rely on a simple model?
    e. Would this involve looking at where the models made their errors
    f. If you had enough models in a challenge theme you can make this plot for all the models used. Look to see if there is an optimal complexity and then compare that across the 5 systems.
        i. Or ask a priori what the hypotheses would be.
        ii. Could make good phenology model with a simple model, Carbon models tend to be complex. Expect Ticks and Aquatics to be somewhere in the middle and beetles would probably be simple.
        iii. Ticks could probably get by with simpler model then aquatic
        iv. Could you relate that to the features of the data?
            1. Think Aquatic/Terrestrial rely on a certain type of model (Jody didn't catch it)
            2. They have both seasonal and diurnal cycles. Don't care about diurnal cycles in any of the other three themes
        v. Hypothesis - the two systems with 2 cycles need more complexity than the three systems that don't
        vi. Phenology has 1 cycle (goes up in the spring and down in the fall)
        vii. How consistent is the error within site?
            1. Simple models may have high error, but may consistently have high error
            2. Or all terrestrial process models have low error at HF because they have been trained there, but elsewhere they have higher error because it is a new system
    g. Think perhaps simple models may have higher error when transferred

h. Do we have models that were specifically trained at a location?
    i. Not sure if people are documenting which sites they are using for calibration vs which ones are completely out of sample
    ii. Ideal thing would be a sub-challenge that asks people to perform a transferability analyses
    iii. E.g., take HF parameter and apply them at Oak Ridge, take Oak Ridge parameters and apply them at UNDERC, etc
    iv. Come up with a protocol that looks at transferability without requiring everyone to do n-factorial experiments

i. Challenge Update
    i. The Target files for all 5 themes have data from all relevant sites (relevant because for ticks, the only sites that are being processed are where the species are within their range).
    ii. Met drivers have been processed from all the sites

j. Subchallenge idea - train your model on this data and see how well it does in these other sites where you haven't trained your model

ii. Mike provided an overview of the standards in [this YouTube recording]() - see the first ~14 minutes. The remaining material on the recording is Carl Boettiger giving an overview of the cloud cyberinfrastructure used for the NEON Challenge

iii.

c. Help develop a modular framework that we can use to try many different models for several challenge themes. Abby is going to work on this based on code Mike has.
    i. Abby has been working on this, but ran into an issue with the neon4cast package and may just need to reinstall it.

4. To Dos
    a. Reminder of Hypotheses
        i. Hypothesis 1: The rate of decline in predictability over increasing forecast horizons differs across variables and scales
        ii. Hypothesis 2: Predictability increases with biological and ecological aggregation. (This is - it is easier to predict forest productivity than any individual tree. Emergent patterns)
            1. Aggregation is challenging. Is a site more predictable than the plots within a site
            2. What would be optimal for the beetles is to predict beetles at different taxonomic ranks and ask what rank is easies to predict

(order, class, family, etc). Mike has found that predictability increases at taxonomic scale
3. There would be the potential to do this for phenology, but would require using other variables than the phenocam
4. Biodiversity indicators

iii. Figures seem more pertinent to hypothesis 1, so start with that
iv. Next step - pick a challenge to get more familiar with it
   1. See details of the challenge themes here: https://projects.ecoforecast.org/neon4cast-docs/
v. Shubhi to check with Abby to see about pulling out the submissions and get familiar with the metadata and the data for the challenges to see how well they fit in for the hypotheses to test
vi. If we had a single repo that everyone has access to for everyone to play around with
vii. Here is a pointer to a lab that goes with Mike's class:
viii. https://github.com/EcoForecast/EF_Activities/blob/5e3ba17eb8bc51fa567d19b57f982d7d037c5497/Exercise_03_BigData.Rmd#L199
ix. To get forecast submissions: https://github.com/EcoForecast/EF_Activities/blob/5e3ba17eb8bc51fa567d19b57f982d7d037c5497/Exercise_03_BigData.Rmd#L199
x. To get model metadata: https://github.com/eco4cast/neon4cast-phenology/blob/master/Challenge2021_manuscript.Rmd
xi. These are fairly large scripts and Abby/Shubhi would need to consolidate so folks didn't have to look through scripts for the code required