April 22, 2024 Theory Working Group Call

Attendees: Abby Lewis, Freya Olsson, Jody Peters, Cole Brookson, Alyssa Willson, Caleb Robbins, Marcus Lapeyrolerie

Agenda:

1. #phylogenetics-predictions Slack channel with the aim to build theory & methods for asking whether phylogenetic or functional relationships among species can be leveraged to give better predictions of nonlinear biodiversity trends.
    a. Blog post: https://ecogambler.netlify.app/blog/phylogenetic-smooths-mgcv/
    b. GitHub repo: https://bit.ly/3PY8DuC

2. Check in with Cole and Shubhi - simulations, weighted permutation entropy and handling data gaps.
    a. Using simulations and NEON data - can we pull info from time series using weighted permutation entropy and realized permutation entropy - use it to look at how predictable the time series are
    b. 2 things -
        i. Simulate from models and permutation entropy
        ii. Looking at the NEON data
    c. How does data gaps affect understanding of predictability
        i. NEON Aquatics data
        ii. Plot of ARIK site dissolved oxygen - how do rolling averages fil in data gaps. As you decrease the rolling average more gaps pop up.
        iii. There will be times of year with larger gaps
        iv. Want to figure out gap length vs gap distribution have more effect WPE
        v. For every NEON site and variable combination - pulled every consecutive 30 data points data sets and calculating WPE for each data set
            1. This is running on the cluster
    d. During the call was waiting for a simulation to finish but it didn't complete so checked in with the group about what might be going on with the aquatic data.
        i. For aquatics - for DO, temp, and chla - will there be difference in when the gaps will be generated? Is there a single sensor that measures all 3? Or can one variable that will be measured while the other 2 won't?
            1. Yes.
            2. Chla is very tough. Sensors get fouled. Even though it is collecting data, the data may not be reliable because the sensors are fouled.
            3. Less confidence in chla. Wonder if NEON is filtering out data based on QA/QC rules
            4. Expect temp to be most reliable.
                a. It is turning out to be the most reliable for Cole. But it has the most gaps. Wondering why that is.

            b. The temp generates more 1-day observation gaps then the other 2 variables
         5. Rolling windows are helpful with 1 day gaps. But not helpful when something goes offline for 3 weeks.
            a. Temp has the most gaps that get filled in
            b. The other variables don't have the 1-day gaps
         6. Is it surprising that temp generates more 1-day gaps?
            a. Yes
         7. This seems to be happening across sites
         8. It isn't just that temp has shorter gaps, there are more gaps overall
            a. There are more gaps overall due to the fact that there are more 1-day gaps
            b. Not more gaps that are multi-week gaps
         9. Are there weird QA/QC on temp that aren't applied to other variables?
         10. What Abby has done for course QA/QC on global dataset has been to set the temp threshold of 0-40 degrees. Anything outside of that is removed.
         11. Bobby Hensley and Kaelin Cawley are the NEON contacts that work on aquatics measurements
      ii. Repo to go from NEON raw download data to processed data for Forecasting Challenge
         1. https://github.com/eco4cast/neon4cast-ci/tree/main/targets
         2. You can't run it since you need credentials, but you can see the script
         3. Cole will dig into this more. Trying to get cross site measurements

3. Marcus - machine learning forecasts and NEON Forecasting Challenge. Any other things to check in about?
   a. Has made a number of changes since the last time he presented to the group 2 months ago
   b. After making changes, the initial results (all the different neural networks were performing the same), now one is performing better than others
   c. Added a naive ensemble model. The one neural network model works and then the naive model works second best
   d. Best working model is the temporal fusion transformer
   e. Transformer has become popular for natural language processing. Haven't typically been used for time series.
   f. Transformer is different from recurring neural networks. There is a different mechanism to make predictions.
   g. Background of the project is comparing neural network predictions.
   h. Use the Python library Darts to get the neural network models
   i. Looking at 8 models applied to the aquatics challenge

   j. Marcus has found that the way the data is pre-processed affects the output/performance of the models
   k. Comparing the suite of neural network output to a climatology model and the naive neural network model

4. Check in with Caleb - Using the NEON Forecasting Challenge to explore predictability across variables and scales
   a. Updates from the April 15 call
      i. Also saw there was chla issues where chla is at 0 and then pops very high. Freya is looking into this
      ii. Goal of the call was to get the across forecast analysis
         1. Plan is to pare down the question
         2. Will focus on measurements from sensors (so not include analyses from ticks and beetles) - so ecosystem measurements
      iii. Caleb is working on cleaning up the Repo and wants to set up every other weekly meeting
      iv. Abby - considering which models would be helpful to add. Could pull in Carl's prophet model (check with Carl on this). Could include ensemble models
         1. Want to think about what broad classes of models we are missing
         2. Have 3 takes on empirical time series models, have basic linear regressions, have typical ML approaches
            a. Ensemble across or within model class would be interesting
      v. How to treat climatology and persistence models for this - could think of them as baseline models for comparison or think of them as models themselves
      vi. Freya has code written for an ensemble model that can be used and can help Abby with this
      vii. How is ensemble different from combining predicting from model for every time point and calculating a score based on that?
         1. This would be true if all individual models were normally distributed. There is potential for skew in prediction, so just taking the main, you won't get the full distribution of the model
         2. The ensemble takes into account the spread from each of the individual models. Mean can be the same, even though distribution is different
         3. Cool idea or wish list item - think of an adaptive weighing scheme so models that perform better get weighted more strongly
         4. When you choose a weighting scheme you are estimating new parameters and you need to balance if that offsets the gains in accuracy

a. There is a lot of lit about if weighting is good or not (and it is split on the usefulness or not). The weights have uncertainty
b. You could not treat parameters as a distribution, but there is uncertainty that won't get incorporated

viii. Poll for biweekly meetings

b. Feel free to add to existing or create additional discussion or methodological points as issues in the Github Repo https://github.com/robbinscalebj/NeonPredictability/issues
c. Google Drive folder set up where any drafting/rationale can happen as this moves forward: https://drive.google.com/drive/folders/1JaJ7d5-C7Zr9n4NNnvLPTQXeDj9OEVsW?usp=drive_link
d. Schedule every-other-weekly check-ins

5. Blog post idea for code review materials (Jody)
   a. On January call the group talked about Jody drafting the blog post and running it by the group. There is no definite timeline for this, but hopefully within the next month or two
      i. SORTEE group has a lot of energy between code review and seems like there is overlap with EFI
      ii. SORTEE: https://www.sortee.org/
      iii. Has Slack channel - here is the link to join https://join.slack.com/t/sortee/shared_invite/zt-2fnqytett-AND1mTuXBKQWYyWUXKn6YA
      iv. Library of code mistakes: https://docs.google.com/presentation/d/12QN3WUc5v1Df7OArEox2U7I_N_qnHHuwzjCYiI4idC8/edit#slide=id.p
         1. Can add anonymously issues that people have found when their code has been reviewed
         2. E.g., day/time errors, misunderstanding of function
         3. It is structured with the same headings that were in the paper on code review