July 10, 2023 Theory Working Group Call

Attendees: Christy Rollinson, Freya Olsson, Abby Lewis, Shubhi Sharma, Glenda Wardle, Mark Lowell, Jody Peters
Regrets: Mike Dietze

Agenda:

1. Update from Caleb about Unconference ML work
   a. Focused on the uncertainty and creating ensembles with multiple sources of uncertainty
   b. We'll check in with Caleb on the next call
   c. Note from Mike who also couldn't make the call - he has updates from the partitioning uncertainty Unconference group

2. Update from Freya about the work at the Unconference on making updates to the Challenge forecast visualizations. This may also feed into the conversation from the last call about comparing across forecasts.
   a. Initially started with 2 projects - aquatic synthesis and redesigning the Challenge dashboard
   b. Dashboard is oriented towards looking at how well each model is doing and scores
      i. Want to reorient it towards community learning and theory
      ii. What can we learn from the Challenge rather than it being a competition/leaderboard - what can we take from the submissions and see how forecasts are performing at particular sites
   c. Used the aquatics as a first example for the dashboard
   d. Had social scientists in the group so had their perspective on creating visualization - what do we want to get from the visualizations (top down approach) vs bottom up building from the code
   e. Designed a couple of maps for multiple locations across the US
      i. Austin is working to implement the code on the Dashboard
   f. Thinking about from aquatics side - compare rivers vs lakes
   g. Thinking about how to synthesize across different themes. For some themes the site type (e.g, river vs lake) is not
   h. The group came away with a list of synthesis questions to ask
   i. Other idea is to have a build your own dashboard template - so people can build the dashboard with your model and the null model so you can assess your own model. Whereas the main dashboard is about synthesis and theory
   j. Want the new dashboard to expand so it is not just for people submitting forecasts. Want it to be for people who haven't not participated in the Challenge, but may want to still use the forecasts
   k. Follow the #dashboard channel on the Slack if you are interested in the synthesis and what would be good to go on the dashboard

l. Would be fun to have an axis that is performance in lakes and another axis that is performance in rivers/streams and is there a tradeoff
m. Currently all the best lake forecasting temp are specific lake models - they are not modeling rivers. 7 out of the top 10 models were hydrodynamic models. So do you need something specific to do well or is a general approach appropriate?
    i. Not so surprising since the other models are simple models that aren't as well calibrated as the lake models
    ii. There must be river people doing process based modeling of rivers that we haven't engaged for participation yet

3. Beetle group at the Unconference updates from Glenda
    a. Start with the motivation of all the themes to in the Challenge, beetles has the least engagement - the group focused on
    b. Forecast is about abundance of beetles with all species lumped or species richness - there was nothing to combine them
    c. By looking at the data - realized if we are doing comparisons across forecasts, then some are not suitable for the comparisons because of the data structure
    d. beetle s is collected in pitfalls over 14 days and then there is a gap so there are days with no observations. So some of the off the shelf models have a whole lot of 0s.
    e. Ex. ARIMA has a flat line because of so many 0s
    f. Group took the aquatics tutorial that Freya has used for workshops and modified it for the beetles. Tried to create the tutorial to be accessible for undergrads to create models
    g. Beetle data - there are 0s sometimes because there are no beetles and NA because the traps are not out, but it is implied that it is 0 instead of NA
    h. Group decided that the target dataset did not do a good job for what people wanted to model
        i. The group made an effort to make a second target dataset to make it available to the broader community
    i. Meghan reached out to Abby about putting together a tutorial to try out with a friendly group and then be able to share it at conferences
    j. Would be nice to have this type of tutorial developed for all data themes
    k. The group also found as they were exploring the NEON data that there was a lot of good information about the data themes across the NEON website, but it was not easy to find it all, so that was another resource the group worked on making it easier for everyone
        i. Freya found this to also be the case with the aquatics data
    l. From Shubhi - sound like the beetles and ticks are set up well for an occupancy model. 0 could mean one of 2 things, either they are not there or they were not trapped. The group didn't get a chance to do this, but were more focused on getting the pieces together for others to do that

m. Shubhi took the summer course and worked with someone that curated the beetle data. There was talk about that there are years when it is safe to impute the 0 data.
    i. Glenda's student Vihanga poster at the Unconference and her work is on imputing 0 data
        1. Most off the shelf stuff puts in a mean, which is not appropriate for the time series
    ii. The group found that by working on the dataset it was really easy to come up with new questions
n. The Unconference group felt the beetles tutorial would also be pretty easy to modify for the Ticks data stream

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. Update from the Gordon Conference - Shubhi had a chance to talk to Mike about the projects about the nitty gritty details about the computation
    c. Shubhi's poster was next to another person's (Bilgecan Sen) who works on weighted permutation entropy - discussed a number of ideas, but Shubhi and Cole haven't had a chance to work on updates yet
    d. How does data availability affect weight permutation entropy - this is something Bilgecan is working on so Shubhi was able to connect about that
    e. Whether or not the time series should be cleaned and then calculate the intrinsic predictability. Or use the raw data but then the intrinsic predictability is very low because of the stochasticity
    f. For the sensor data (e.g., aquatics) there intrinsic predictability is very low because there is observation error. Think moving towards using a random walk to predict means assuming normal distribution of error
        i. This is on a daily time step. The daily time step for aquatics is very noisy.
        ii. Abby would expect the aquatic to fluctuate alot, but not sure it is sensor error
        iii. Maybe intrinsic predictability on daily time step is very low, but if you subset to weekly or monthly. Shubhi has found that when you aggregate from daily to weekly/monhly then it goes from chaos to almost perfect.
        iv. Interesting because some forecasts are doing quite well over short time periods based on RMSE and CRPS
        v. The weighted entropy values are positively correlated with predictability. Bilgecan found a linear trend between RMSE and weighted entropy.
        vi. Think raw air temp would have low intrinsic predictability so it appears it is chaotic, but we can capture it with weather forecasts. Abby guesses that for aquatic temp targets where it appears to be chaotic time series, but that it is predictable on a daily time scale. Not sure how to break down

knowledge gained from met forecasts to partition that from ecological dynamics

vii. Assumption of weighted permutation entropy is that if you don't have redundant information - if your pattern isn't repeating, then it is stochastic and stochasticity is inherently unpredictable.

viii. If you have 7 days of data and it randomly fluctuates around a line and your prediction is a line. If tiny fluctuations around mean and your model predicts the mean, is that a good prediction?
   1. Does Freya's work looking at where models beat persistence
   2. Think this is a forecast evaluation issue - is RMSE a good metric to use in this case?
   3. A flat line is not attempting to forecast the fluctuations.
   4. Comes back to what we are trying to forecast? If we don't care about the up adn down fluctuations then the flat line isn't useful.

ix. Another metric Shubhi/Cole is working on - comparing the climatological prediction to recent data. For a predictable system it gets better.

x. Climatology may reflect the long term variation

xi. Persistence is a random walk from the last observation. Climatology is on this day, historically, these were the values.
   1. For the Forecasting Challenge, there are not many models that are doing better than persistence
   2. ARIMA could be the same as persistence

xii. Maybe want ot compare null model to the intrinsic predictability

xiii. Nex steps - Shubhi and Cole are going to work on the aquatics theme and will dig into the intrinsic predictability for next month!

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

6. How to compare forecast performance for models that are not submitting for every single date. E.g., New forecasts submitted now, won't have forecasts for January

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