

January 23, 2023 Theory Working Group Call

Attendees: Abby Lewis, Kathryn Wheeler, Mike Dietze, Christa Torrens, Jody Peters, Caleb Robbins, Cole Brookson, Jono Tonkin
 Regrets: Shubhi Sharma

Agenda:

1. Announcement
 - a. [Applications](#) are open June 21-23, 2023 [EFI Unconference](#) at NEON HQ in Boulder, CO Save the Date. **Applications are due February 1**
 - i. Goal: Bringing People Together to Do Forecasting: Training, Technology, Theory, and Translation
 1. Participants will work together to produce products they want to work on. Examples are (but not limited to): getting a forecast up and running, developing teaching materials, finalizing tutorials, refining or creating tools, analyzing forecasts for a manuscript, and/or developing visualizations.
 2. The event will also include a poster session for attendees to present their research.
 - ii. Travel funds within the US are available
 - iii. Space is limited to 50 people.
 - iv. The application to attend is due on Feb 1. But the nominations of project ideas are not due at that time. So this group will have more time to think about specific project ideas to propose.
2. Workflow to submit models for multiple Challenge themes (Abby)
 - a. Motivation
 - i. Use the same model across multiple themes. See if there are differences and why
 - b. Demo current workflow
 - i. Linear regression between air temp and variable of interest by theme
 - ii. Done for terrestrial daily, aquatics, phenology for 30 days
 - iii. Think it could be extended to ticks and beetles. Could look at historical air temps at time of year to extend forecasts. Not sure what would work best
 - iv. Goal is to make it easy for people to adapt and add new models that can be run in the same framework to submit to all the Challenges
 - v. Step 2 - Getting NOAA driver data
 1. Having issue with arrows and M1 Mac so Abby has a workaround for this, but takes a long time to download
 - vi. Step 3 - the model
 - vii. This would need to be adapted for the auto-regressive models. Don't think it would be hard, but needs to be set up

c. Next steps

- i. Get it to run automatically for the 2 models set up so far
- ii. Think simple ARIMA models would be good to set up
- iii. Could extend covariates beyond temp and extend to more flexible models. Could take lm and replace with GAM
- iv. Better representation of uncertainty would also be helpful. This will connect nicely with what Shubhi is thinking about as well.
- v. Think the models should spit out predictive intervals easily.
 1. Have the NOAA ensemble and that is all that has been incorporated so far
 2. Could do the met by parameters uncertainty. This might require manual Monte Carlo simulations
- vi. Goal is to compare the model performance and where the horizon is met
- vii. Could add any driver we want (e.g., wind speed). Wind speed may not matter for something like ticks, but it documents a difference across systems which can be informative
- viii. For a single predictor variable vs using a single variable and see how well each process can be predicted by a variable across the 3 response variables. If you do a simple regression and you restrict your model structure - can any of the themes be predictable?
 1. What would be good predictors other than temp?
 - a. Anything that would exist at the same scale
 - b. Precip, humidity, solar radiation, day length
- ix. Setting up forecasts with precip, humidity, etc is easy for Abby to set up pretty quickly
- x. Are running the models by site and with all sites combined
- xi. All Challenges have historical mean forecasts that are running
- xii. In the future will be able to compare between historical means and persistence models since they are already up and running
- xiii. ARIMA model - who used the Prophet model? cb_prophet = Carl Boettiger
 1. Ethan White has also set up a bunch of models for the Portal Project - all the code for those models are available - just need to look them up
 2. portal project is here: <https://github.com/weecology>
- xiv. Think it is logit to run the beetles/ticks for 35 days. Portal Project also has code for downloading MME which is a seasonal/sub-seasonal data
- xv. Ticks can forecast 4 weeks. Same for beetles
- xvi. Anyone want to contribute to the model development and walk through the code to discuss how/where to modify. Either develop your own forecast or working on models that can be applied across all themes
 1. Caleb, Christa, Kathryn

2. Repo:
 - https://github.com/abbylewis/EFI_Theory/tree/main/Generate%20forecasts
 3. Kathryn is getting into ML so think it could be good to do across NEON data streams
- d. A couple of logistical qs:
- i. generate_metadata() function not working?
 1. neon4cast:generate_metadata - doesn't seem to be working or doing anything
 - a. This definitely used to exist. Mike ran it in the past
 - b. Check with Quinn
 - ii. Help with automation?
 1. GitHub actions - how to set it up in an automated way? Seems easy if you create a new repository for that specific forecast. But want it to be in the EFI repository so everyone can access it and it is consolidated in one spot
 - a. Mike has found the example yaml for the Challenge workflow example on GitHub is pretty good
 - b. This is Mike's explanation for how to set up GitHub Actions and how it works
 - i. https://github.com/EcoForecast/EF_Activities/blob/0e8d93daeb519d2a30ae3fffb7287794cfc0feeb/Exercise_04_PairCoding_NEON.Rmd#L482
- e. Unconference ideas
- i. 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
 - ii. <https://github.com/eco4cast/unconf-2023>
 - iii. Looking at how sites that are close in space can be used to create forecasts
- f. Next step - Abby will copy and paste to run the models for other NOAA data
- g. For the group in the next couple of weeks - think about other models that could be run
- h. Nick Clark has some nice papers on dynamic GAMS. Leah Johnson had a lot of luck in the CDC mosquito forecasting challenge using Gaussian process models
- i. At one point we talked about comparing "state of the art" models across all the themes
- i. So this would be a different model for each theme
 1. Think we need to rely on what has been submitted by other teams
 2. May be some of the themes don't have anything that is considered "state of the art"
 3. Had previously brainstormed what would be considered state of the art models and could see if those have been submitted

3. 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. (saving for next month)