

March 9, 2026

Attendees: Jon Borrelli, Pamela Rueda Cediel, John Foster, Andrew Allyn, Bilgecan Sen, Abby Lewis, Lenny Smith, Freya Olsson, Cole Brookson, Hassan Moustahfid, Jody Peters, Freya Olsson

Agenda:

1. Predictability of Nature synthesis updates, reminders, next steps (Bilgecan)
 - a. Paper outline
 - b. Bilgecan has reached out to separate groups of people working on different sections of the paper to work out details for the direction of the section. Writing will start taking place in April with the goal of finishing in May (with flexibility on that)
 - c. Want people who are writing sections. To work in the tab for your section.
 - i. People could edit the outlines, but think it will be better to move to working in the tabs
 - d. Then once the sections are drafted, Bilgecan will combine into one doc with everything resolved
 - e. Bilgecan will write the workflow and introduction - will wait to see what the other groups decide on
 - f. Expect to have a bigger update with the first draft on the June call
2. Prep for the EFI2026 Conference Theory working group (Cole)
 - a. Cole will have more time to think about this and we can discuss further on the May call
3. Forecast evaluation/scoring follow-up and learning from other forecasting disciplines (Lenny)
 - a. Lenny's thoughts (see this doc), following up on [Simonis et al. 2021](#), which provides a review for ecologists about probabilistic forecast evaluation tools from climatology, economics, and epidemiology disciplines
 - b. Lenny ran a center thinking about time series data so brought
 - c. Would like to get comments from the group - do we need a discussion about why autocorrelation, root mean square error
 - i. Mark up the document for what is known (green) and what needs to be clarified or updated (red)
 - d. Autocorrelation is a measure of predictability with a linear model that is perfect
 - e. RMSE only works with Gaussian distributions
 - f. How to get the same base point
 - g. Expectation that people will respond that there may be a better score, but ecological understanding won't change
 - h. Difference between optimal and suboptimal score in theory vs that the conclusion or outcome would change
 - i. Is my application good enough, do you want to bet on the outcome
 - j. Lenny needs ecological forecast examples. What are we worried about. Is giving a mathematical example enough
 - k. Previous theory working group work was working on the NEON Forecast Challenge - seems like there is a building block there. If we are trying to come up with a general conclusion about modeling systems, then could use the models that were forecasted.
 - i. Could think about how to rescore those forecasts. We already know what people use to model the different challenge themes

- l. For the predictability synthesis paper: Simple example - if we use rmse for a range of species. Higher rmse with lower latitudes. What does it tell us about predictability? Is it wrong - there is no predictability, but we found something. Or rmse is not suitable. Or is it a noisy way to detect predictability. If we measure it wrong, then shooting ourselves in the foot.
- m. When getting to the point of measuring model performance, the data is messy, the model is messy, there are many assumptions
- n. To be convincing would be good to have examples. Think about decision making from applied/conservation stand point and decision making from science stand point.
- o. NEON Forecast data could be great for this. But there are also population time series available to make predictions with RMSE and weighted permutation entropies.
 - i. Would be good to make generalizations - insects are nonlinear, birds are linear, insects are hard to forecast, birds are easier to forecast
- p. All models are wrong, some are dangerous. Want to avoid the dangerous ones.
- q. If system is best described by linear model. If you use linear model, you are making assumptions about the distribution. Need to check internal consistency with whatever is used.
- r. Rms makes sense for linear models. Need internal consistency checks
- s. Briar score - would be wrong for some probabilistic forecasts. Some scores were adopted because of the technology of the day, but now we don't need to do that
- t. Distinguish philosophical questions
- u. Some scores are not measures of predictability
- v. Decay of predictability. Sometimes there is not the info you want in the dataset you have. There is a big divide between using the best available model vs the model fit for purpose. Need to figure out if the available model can answer the question
- w. Species distribution modeling - predictability in space and time. There are some metrics better for time series vs spatial piece. There is an inevitable question about scale.
 - i. There isn't consistency in distribution modeling
- x. Presence and pseudo-absence data. If the model is logistic and it predicts 0.5 then you move to presence/pseudo-absence you don't know if that holds
- y. **Predictability of the Decay of the 1982/83 El Niño-**
https://journals.ametsoc.org/view/journals/mwre/114/6/1520-0493_1986_114_0967_potdot_2_0_co_2.xml
- z. It comes down to what is the question you are trying to answer
 - i. What to do with 0s, what to do with missing data
- aa. Sometimes we confuse prediction and forecasting (and projection). Need to use the same jargon - don't make up different terminology.
- bb. Lenny will pick out some examples to share

4. April call: Shubhi Sharma to share a brief recap from her dissertation defense