October 14, 2020 Theory Working Group Call

Attendees: Amanda Gallinat, Christy Rollinson, Abby Lewis, Alex Young, John Foster, Jody Peters, Ann Raiho

Agenda:

1. Go through the Questions in the RCN NEON Forecasting Challenge Phenology example to continue to clarify what the questions are asking (see further info about using the Forecasting Challenge topics to explore ecological forecasting theory)
    a. Consider the definitions from the forecasting vocab we've discussed. Think about how to apply those definitions to that specific case and how we would assess them. As a Theory Group what would we want people to report on (even conceptually)? What would we want to test in the phenology example?
    b. Background about why we want to think about the Phenology (and other NEON Forecasting Topics)
        i. The motivation for the Phenology example was thinking through questions about scale and uncertainties that had been bouncing around in earlier discussions, things that Peter brought up in his talk, and the conceptual slides.
        ii. Can we apply these ideas a priori to any of the NEON Forecast Challenge areas
        iii. Can we make a prediction for what we expect to see for these scale and uncertainty questions from the conceptual figures and do a community of forecasters agree on any of it
        iv. Our plan is to go through this exercise for the 5 NEON Forecasting Challenge Topics and then go back to the Conceptual figures to see if we are still on the same page.
        v. This exercise will also be useful for making predictions for the NEON Forecasting Challenges
    c. Notes for the Discussion during the Call
        i. The questions in the Example came from the Theory group, not the RCN steering committee/design teams
        ii. Goal is to use the Forecasting Challenge themes as a way to think through the theory of forecasting
        iii. On the September call it had been challenging to separate the predictability of systems vs judgment calls of how to model that process
            1. They are not the same thing
            2. What is the timescale of the process is different from what is the frequency that we can measure the processes to put into the forecast
            3. The relevant process of the timescale is also important to consider. We can make a case that everything is driven by sub-

minute processes. But the process of what things change practically is what matters

4. Temporal scale and spatial scale are connected. The relevant temporal scale will depend on the spatial scale you are studying

iv. Going back to Row 8 in the Phenology Example - the timescale relevant to the process
    1. There is a big spread in everyone's responses
    2. Clarify - how we talk about the processes or the timescale to complete itself (thinking of phenology and bud burst and senescence - how long does it take for the whole season or part of the season to occur)
        a. You would have different answers to different parts of the process
            i. Spring split into bud burst and leaf growth
            ii. Fall - triggering of onset of leaf change from senescence vs abscission
            iii. These are different but are coupled. In fall they are integrating over non-trivial time scales (days, weeks, possibly months - probably not years - but if you include multiple years and thinking about plant health the year scale may have some influence)
            iv. Leaf growth responds on much shorter time scales (hot days vs cold days)
            v. Plant health impacts
                1. Spring leaf growth uses last years C storage so there is longer memory
                2. This connects with the concentric rings of the ecological theory.
    3. When thinking about time scales - do we want to separate it into season? If there are different drivers, do we want to model the different seasons separately.
        a. Separate by what season and what process we are forecasting
    4. Start with spring
        a. How do the timescales differ?
            i. Short process for bud burst, longer process for leaf growth
        b. Or reverse it to think about the cues that signal bud burst. The process and the change in the process happens faster, but if thinking of presence/absence of bud burst. Can have warm day in Feb to get bud burst. But then the leaf growth is constrained. Then the change can be more predictable in leaf growth than in bud burst.

c. A single warm day may influence bud burst, but wouldn't expect it to happen immediately, expect that it is building up to it.
d. If using warming degree day and chilling degree day to think about a way to estimate bud burst. Warming is probably 1-2 months before. Chilling is happening the year before if it is accumulation of chilling components
e. Warming accumulation - at what point would your decision to set your start date not change your answer. The threshold will be different, but could calibrate to get the same answer. For example, if you set the threshold for spring warmth now/today for next spring will probably get the same answer if you set the spring warmth to next January.
f. Knowing the uncertainty contribution of the initial conditions
    i. But with phenology initial condition is the date you start the sum on. How long your spin up is. This is an artifact of the models used
g. -omics data could come into play into this. What is turning the genes on? Are the things used in the models arbitrary options for the

5. The relevant timescale for the process does not mean the relevant times scale we are expecting to observe.
6. What are the initial conditions or what is the date that we aren't getting extra information or change?. Would want to think about this for all the RCN Challenge Examples
7. Think not only the process but the drivers of the process
8. Phenology is hard compared to the other Challenge themes because we need the genetic information.
    a. Are there models for genes in trees? Not sure. There are things ecologists think about and what people who studied Arabadopsis think about but there is a big gulf between
    b. Perhaps the weighted ensemble models would be a way to incorporate the genetic models and see what works best

v. One person's process is another person's correlation. For those where the process is if the leaf comes out or not, then the multi-model approach is helpful.

vi. Going back to the different scales of organization and Amanda's honeysuckle example
    1. Generally all the buds burst at once. But then have variation within individuals

vii. Another example - viburnums. Amanda has and image of viburnum with green, yellow, and red leaves on the same individual. So there is variation within individuals
viii. Ecosystem level forecasts - have site effect that averages over each site
ix. Spatial heterogeneity driven by species, microclimate, and soils and vertical canopy effect
x. Would expect differences within species and individuals - the variation will increase with the level of organization

d. Relevant time scales - think about the timescale of the process. Split it into 2 processes. Lots of uncertainty around the initial bud burst. Once you have budburst, then uncertainty collapses and there is a clear trajectory of leaf expansion.
    i. Expect that the predictability of leaf growth is more certain
        1. Think that there is higher predictability and that we have drivers on a timescale that could predict the variability in leaf growth better compared to bud burst
    ii. NOAA weather forecast just jumped from 16 day to 35 day forecast

2. Row 9 in the Phenology Example: What do you think is the forecast limit for this system at the NEON phenocam scale?
    a. A lot of entries that are dependent on weather forecasts. 1 week, 2 weeks, 1 month is the range
    b. This is a key way to provide insight into what the relevant timescale of the process
    c. In hindcast will get longer limit of predictability than we get at in the forecast because of the met uncertainty
    d. Is this a question we can answer? Or is this something we want from the forecasts once we get the output from the Phenology Challenge.
        i. Yes this is something we can get from the Phenology Challenge. But we want to get people's priors before the Challenge as well
3. Row 10 in the Phenology Example: What is the relevant spatial scale?
    a. What are we most interested in? The genetic scale that we aren't able to observe but think is important and hope to include in future models? Is it at the individual scale? Or is it a process at the canopy/community level?
        i. For space (and applies to time) - thinking of the original Stommel diagram it had a Z axis that was variance. It is clear that things are happening at all scales, but the variance will not be the same at all the scales.
        ii. Where is the variance in the big picture?

4. Theory group manuscript
    a. Before call look through the Draft Outline - provide comments
        i. This is based upon previous TG meetings and in particular the Forecasting Hypotheses document
        ii. Thoughts on this structure for a theory-oriented forecasting manuscript?

  b. Goal of the Theory group is to understand how forecasting can advance ecological theory
    i. One way to do that is to link forecasting concepts to ecological concepts (see Christy's comments on the outline)
    ii. Think about this as we go through the forecasting Phenology exercise
      1. Thinking specifically about the last question dividing the forecast time from time = 0 to forecast horizon - when do we think different things will dominate the forecast uncertainty?
    iii. Amanda pictures a conceptual figure of a forecast overlaid with ecological theory and point to what can/cannot be tested with ecological theory
  c. Something else valuable to explore is the ability to identify what are the limits of predictability and sources of uncertainty. If we vary the scale how does that change the uncertainty?
  d. Not only is it the difference between the process itself and what we are observing but also what that means for the scale of process vs the scale of observation. This was a big aha moment for Amanda
    i. Forecasting is set apart from other statistical ways of analyzing data by the iterative nature of the forecast. Can have some things mismatched in time. Don't need 1 driver point for 1 observation. The iterative nature allows for understanding variability at scales better than doing it in a post-hoc manner.
    ii. In sample model fitting can mislead you about your understanding of a process
      1. For example, from Mike's experience with phenology, once they started using dynamic models, the first non-trivial model they fit was a logistic model. They took the simple logistic curve which did well getting the basic shape right, but when they tried to use that to make a forecast, it didn't work. Lesson learned was that something that fits the data well can make bad predictions. Have to think about the underlying processes.

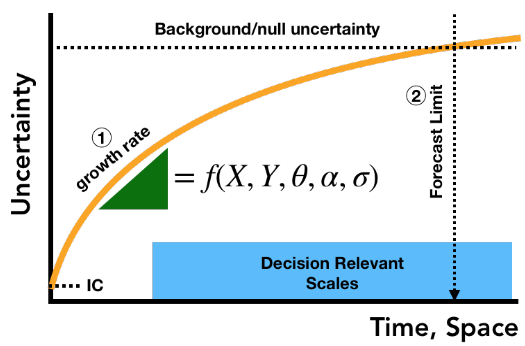
5. What are the next steps for the Forecasting Vocab Terms. We didn't get to this on the October call
  a. Summarize and put on the EFI website?
  b. Box in Forecasting Definitions manuscript led by Anna and Gretchen?

6. Background Information for Comparing Across NEON Forecasting Challenge Topics
  a. We have created Google sheets linked to each of the following forecast topics with questions 3b-3j. For question 3j the format in the Google sheet is slightly different. For each type of uncertainty, enter whether that uncertainty dominates

in the first 1/3, middle 1/3, or last 1/3 of the forecast horizon. The common frameworks slides are listed below question 3j for you to reference.

i. Leaf phenology <- We (organizing committee) decided on starting with leaf phenology (specifically forecasting PhenoCam observations) for phase 1
ii. Carbon and water fluxes (eddy flux) between land and atmosphere
iii. Aquatic - chlorophyll a. The organizing committee has decided to focus on forecasting water temp and DO
iv. Ticks - abundance. Timing of the peak or abundance through time as observations come in
v. Beetle communities

b. Questions that you will see for each Example (see Common Framework slides below for additional context):
   i. For each of the RCN forecasts, what would the units on the x axis be in the following figure?



c. What is the level of organization being forecast (organ/physiology, individual, population, community, ecosystem)?
d. What is the phylogenetic scale of the forecast (if applicable)?
e. What is the trophic scale of the forecast (if applicable)?
f. What would you say is/are the relevant timescale(s) of the process itself?
g. What do you think is the forecast horizon (time till the prediction is doing no better than chance) for this system at the NEON plot/sensor spatial scale?
h. What would you say is the relevant spatial scale of the process itself?
i. How would you describe the spatial scale of the forecast (relative to the process itself)

j. If we divide the time between t=0 and the forecast horizon into 1/3s, what input uncertainty do you think dominates the forecast uncertainty at each point in time?

uncertainty	First 1/3	Middle 1/3	Last 1/3
Initial conditions			
drivers			
parameters			
Random effects			
Process error			