Chapter 3: Optimizing forecast visualizations with incorporated uncertainty as decision support tools for water resource managers

Background

Visualizations are critical tools which efficiently enhance human understanding of complex ideas (Larkin and Simon 1987). Producing appropriate visualizations of forecasts presents a unique challenge given that forecasts are inherently uncertain and yet their output is often used to make crucial management decisions with human health implications. Additionally, uncertainty is a notoriously difficult concept for most individuals to grasp, as well as to represent graphically (Bonneau et al, in *Scientific Visualization* 2014, Potter et al 2012). There are numerous studies looking at different ways to visualize uncertainty in data (Potter et al 2012, Smith Mason et al 2017, among many others) but little consensus as to the best approach for visualization uncertainty of forecasts for decision support. Additionally, it has been well documented that different approaches to visualizing uncertainty result in differential comprehension by users (Kinkeldy et al 2017, Mckenzie et al 2016) indicating that representation of uncertainty has important implications for influencing decision-making. In this chapter, I propose to determine the best approach to visually represent uncertainty in ensemble forecasts of algal concentrations at a single location to optimize application for decision support.

Research Questions

1. How do different visual representations of uncertainty of vector forecasts (ensemble forecasts of algal concentrations (ug/L)) influence decision making of water quality managers?
2. Do managers and scientists differentially perceive visual representations of uncertainty within forecasts of algal bloom densities (ug/L)?
3. Does the comprehension of uncertainty by managers vary between forecast variables? (e.g., binary (presence/absence of bloom) or continuous (concentration of algal biomass) variables)

Proposed Methods, Q1

* Managers at the Western Virginia Water Authority will be surveyed as a baseline study. Results from this study will be used to further adjust visualizations for a broader survey of managers across geographic regions. The larger manager pool will be accessed through collaborations within the North American Lake Management Society (NALMS). A preliminary analysis to determine the number of respondents necessary in order to make a representative sample size will be conducted.
* A forecast scenario indicating a severe algal bloom (which should elicit some management decision) will be developed into several visualizations which include uncertainty of the forecast (n = 4? 6?). Managers will receive one of the visualizations.
  + Possible variables to differ between visualizations include:
    - Background primer on uncertainties?
    - Data type (point, line, polygon, network, field) (Smith Mason et al 2017)
    - Data summary (mean + 95% CI, mean + sd, all ensembles, etc.) (Wiggins et al. 2018)
    - Representation
      * Extrinsic (incorporating new geometric objects to represent uncertainty) vs intrinsic (incorporate uncertainty within the existing object) (Smith Mason et al 2017)
        + Extrinsic example: separate error bars for SD or CI
        + Intrinsic example: the entire range of the ensembles is a fuzzy polygon (Olston and Mackinlay 2002)
    - Interactivity (create a shiny app?)
    - Animation (animate time series of forecasts leading up to present? i.e., does knowing the history of forecasting at a system influence comprehension and trust?)
* After viewing the visualizations, a **qualitative survey** will be given to manager to assess 1) comprehension of uncertainty and 2) trust in the forecast output (i.e., does a better understanding of uncertainty influence a respondent’s level of trust in the forecast output as a tool) as a function of the provided forecast visualization.
  + The survey should also include questions to measure the amount of prior knowledge/experience with probability concepts and algal blooms to assess individual differences in responses to the given scenario
* Respondents will be asked as a manager of the water body the forecast visualization is based at to make a **management decision**.
  + E.g. a given decision is predicted (‘modeled’), the realized decision of the subject is observed

Chapter 4: Many potential avenues

1. Further explorations of forecast visualizations
   1. Many potential questions
      1. Do managers and scientists differentially perceive visual representations of uncertainty within forecasts of algal bloom densities (ug/L)?
      2. Does the comprehension of uncertainty by managers vary between forecast variables? (e.g., binary (presence/absence of bloom) or continuous (concentration of algal biomass) variables)
   2. Contingent upon further development of Chapter 3
   3. Motivation: learning a new body of literature for Chapter 3 that could be further applied to a fourth chapter
2. Process-based vs. empirical forecasts
   1. See Chapter 1, which has now split into two distinct projects
      1. Current Chapter 1: Successful production of a near-term iterative forecast of algal biomass in a drinking water reservoir using an empirical model approach
      2. Potential Chapter 4: Comparison of process-based vs. empirical model success at predicting near-term algal biomass
   2. Motivation: I’ve spent a lot of time thinking about and developing this question already, so there is a lot of material to build on.
3. Developing forecasts for Lake Sunapee
   1. Applying similar approach to what we have used in FCR to expand the forecasting network to a north temperate lake with different geology, weather patterns, land use, infrastructure, etc. to test the ‘robustness’ of FLARE