Response to comments on "Visualization in Bayesian Workflow"

J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman

November 2, 2018

Thank you to everyone who has participated in this discussion.

In his proposal of vote of thanks Diggle poses the following question in response to our paper: in what sense inference can be deemed Bayesian if "[...] you massage your prior until it generates realisations that are concentrated (to a greater or lesser extent) around the data [...]"? This is a very reasonable question, but we were *not* suggesting that the prior should be manipulated until realizations from the prior data generating process are concentrated around the data. Since most readers of our paper will not be experts in particulate matter air pollution, in order to make the comparison more accessible we compared the prior predictive simulations of PM 2.5 concentration to the actual measurements we have. But when a researcher is actually conducting an analysis in their area of expertise, they should have enough familiarity with the subject matter to look at prior predictive simulations on their own, without needing to make direct comparisons with the data that will be used for model fitting. For example, a researcher studying PM 2.5 levels would know that the simulated data represent concentrations that would be fatal to life on Earth. So our point is really that a reasonable prior is one that yields a reasonable prior data generating process, not that the researcher should tailor the prior to suit the particular observations in hand.

Baillie and Vandemeulebroecke say that our proposed workflow resonates with their experience as statisticians working on drug development. We are very happy to receive this confirmation that the workflow is being applied successfully in industry. Baillie and Vandemeulebroecke also point out one limitation of our paper: there are no recommendations for how to use visualization in the final step of communicating inferences to a more general audience. We used our limited space to focus on visualizations intended to help people doing statistical modeling develop better models, but we strongly agree that translating the inferences from those models effectively for other audiences is a vital part of a statistician's job. We are also thankful that Carter and other commenters have brought attention to the challenges of communicating to different audiences using visualizations. This was not something we intended to address in our paper, but it is an important topic that deserves greater attention from the statistics community.