

June 22, 2016

*To:*

Dr. Parker J. Wigington, Jr.  
Editor-in-Chief  
JAWRA, Wiley

*From:*

Dr. Marcus W. Beck  
US Environmental Protection Agency  
[beck.marcus@epa.gov](mailto:beck.marcus@epa.gov), +18509342480

Dr. Rebecca R. Murphy  
UMCES at Chesapeake Bay Program  
[rmurphy@chesapeakebay.net](mailto:rmurphy@chesapeakebay.net), +14102679837

Enclosed please find our manuscript, entitled “Numerical and qualitative contrasts of two statistical models of water quality in tidal waters”, to be reconsidered as an original research article in the Journal of the American Water Resources Association.

This resubmission is a major revision to address earlier comments provided by an associate editor (attached). We agree that our previous submission did not provide sufficient guidance or distinctions between the models to be a useful contribution. We are hopeful that the revision provides a more satisfactory approach to address the study objectives. In particular, this revision includes the following:

- More detailed description of differences in the statistical foundations of each model, in particular the section ‘Numerical contrasts of WRTDS and GAMs’ was added to the methods as a preface to the analysis.
- Detailed qualitative comparisons of the models, including a discussion of the products and use across a range of situations. In particular, table 8 compares each model to supplement the empirical comparison in the earlier draft. A discussion of specific elements of the table is also included in the section ‘Qualitative comparisons’ to assist readers with choosing an appropriate method.
- Information on k-fold cross-validation was shortened in the section ‘Selection of model parameters’. Additionally, the first paragraph of the discussion that describes use of performance metrics for model comparisons was removed. We agree with the editor that this information was too general and not helpful for choosing a technique.

Respectfully,

Marcus W. Beck

Associate Editor

Comments to the Author:

This manuscript has the potential to provide statistical guidance on selection of modeling approaches for data commonly encountered by JAWRA readers. As such, it would provide valuable information and has the potential to be well-cited. As written however, the manuscript does not meet this goal.

I recommend that the manuscript be significantly revised before formal review. The revisions would need to include the following:

1) Clear information about the statistical relationship between the two methods. A quick search in the statistical literature, for example, finds the following statement. “In the development of generalized linear models, we use the link function  $g$  to relate the conditional mean  $\mu(x)$  to the linear predictor  $\eta(x)$ . But really nothing in what we were doing required to be linear in  $x$ . In particular, it all works perfectly well if  $\eta$  is an additive function of  $x$ . We form the effective responses  $z_i$  as before, and the weights  $w_i$ , but now instead of doing a linear regression on  $x_i$  we do an additive regression, using backfitting (or whatever). This gives us a generalized additive model (GAM). Essentially everything we know about the relationship between linear models and additive models carries over. GAMs converge somewhat more slowly as  $n$  grows than do GLMs, but the former have less bias, and strictly include GLMs as special cases. The transformed (mean) response is related to the predictor variables not just through coefficients, but through whole partial response functions. If we want to test whether a GLM is well-specified, we can do so by comparing it to a GAM, and so forth.” The underlying statistical relationship suggests that it should not be surprising that the results for this dataset did not differ much by approach and yet the authors conclude “A general conclusion from our results is that both models provide similar information, both in predictive performance and trends over time in the Patuxent.”

2) There are important differences in products available with each technique - e.g., how easily can confidence intervals be calculated. If the goal of the paper is to help readers choose a method for their situation, this information is important and is only casually mentioned in the discussion.

I also note that there is quite a bit of discussion / education on what the leave-one-out approach is and on what models can be used for in a general sense (hypothesis generation etc) which is not new information and which doesn't help the reader choose a technique.

I suggest that the authors reframe the manuscript with the goal of providing explicit guidance on which modeling technique to use across a range of situations, pros and cons of each, and the underlying relationship between the two (for this last topic - a short summary will do as it is well known and well-understood in other scientific communities).