6520 Mercantile Way, Suite 5

Lansing, MI 48911

517-887-8903

http://mi.water.usgs.gov

February 8, 2015

###### MEMORANDUM

To: Wesley Brooks, Rebecca Carvin, Steven Corsi, Michael Fienen; Wisconsin Water Science Center, Madison, Wisconsin

From: Carrie Givens , USGS Michigan Water Science Center, Lansing, Michigan

Subject: Manuscript Review— Predicting recreational water quality advisories: a comparison of statistical methods

This manuscript assesses 14 different regression methods in nowcast models to determine which method most accurately predicts exceedances of fecal indicator bacteria (FIB) concentration at public beaches. This assessment is both timely and critical as more agencies are incorporating the use of nowcast models into beach management. Thus, these results would be of interest to any entity (i.e. Federal, state, and local science and health agencies) responsible for beach management, public health, or environmental science.

A few things need to be considered and revised prior to journal submission. First, the author did not provide details regarding a target journal and audience for submission. As stated previously, this information could be very useful for beach managers and local health departments. However, this manuscript is quite technical in nature and some of the specific technical details regarding regression methods and analysis of these methods may not be effectively conveyed to personnel without this technical expertise. I would advise the authors to review the manuscript and clarify some of the technical language and details for a wider audience. Additionally, the manuscript would benefit an overall review for minor issue involving grammatical errors, awkward sentence structure, and other editorial issues.

I have provided a line-by-line list of specific issues that need to be addressed prior to journal submission.

**Abstract**: The abstract currently does not contain line numbers. What does the author mean by the phrase “appearing in growing number?” Does this refer to popularity and use or the type of regression methods and models available for use.

***Response****: Changed “*methods are … appearing in growing number*” to “*methods are sometimes used*”. This reflects the fact that OLS regression is the most popular method in use, but other methods are used and there is little or no formal guidance about which method to use.*

**Line 11:** The phrase “harmful pathogens” is redundant. Does the author mean to specify pathogens harmful to human health?

***Response****: We do indeed mean to refer to pathogens harmful to human health, so the sentence was changed to use that phrase.*

**Lines 14-16:** Here and elsewhere, consider revising sentence structure for clarity.

***Response****: Changed this section from “*Accordingly, many states have established regulatory standards for water quality; Wisconsin states that a beach should be posted with a swimmer's advisory when the concentration of the FIB Escherichia coli exceeds the beach action value (BAV) of 235 colony forming units (CFU) / 100 mL*“ to “*Accordingly, the states have established regulatory standards for water quality. The state of Wisconsin says that a beach should be posted with a swimmer's advisory when the concentration of Escherichia coli (an FIB) exceeds 235 colony forming units (CFU) / 100 mL*.” and made other minor revisions for the sake of clarity throughout the manuscript.*

**Line 22:** Authors should consider providing minimal information on the traditional analysis methods further explaining comments about the labor intensive and expensive aspects of these methods.

***Response****: Changed this section from this:* “Traditional analysis methods for FIB concentration requires 18-24 hours for culturing a sample, so the decision to post an advisory is often made based on the previous day's FIB concentration, which is the so-called persistence model for beach management” *to this:* “The standard method of measuring FIB concentration is by counting CFU in a sample after 18􀀀-24 hours of incubation. Because of the lag induced by the incubation time, the most up-to-date measurement of FIB concentration is approximately one day old.”

**Lines 32-33:** Are “easily observed surrogate covariates” environmental data? Is there a reason to not simply state environmental data including the specified environmental variables? If the authors want to use “covariates” throughout the manuscript, then a definition should be provided. Please clarify “running 24 h rainfall total.”

***Response****: Yes, the surrogate variables are environmental data. We had thought that using the term “covariate” throughout the document would enhance clarity by keeping the number of terms to a minimum. However, this comment makes clear that the term “covariate” was stretched too far in some places, so we now use the terms “environmental surrogate variables”,”surrogates”, and“variables” to replace“covariates” in most cases.*

**Line 42:** Revise sentence for missing text at the end of the sentence.

***Response****: This was an error in the citation style and has been corrected.*

**Lines 46-48:** Consider moving this text earlier in the Introduction. Are there any drawbacks to any of the other regression methods besides the OLS regression?

***Response****: Moved this text so that it introduces the discussion of the current generation of predictive models. Revised it to make clear that OLS is the current standard option and that listing some drawbacks of OLS does not mean other methods have no drawbacks, but does justify looking for methods that evade the problems with OLS.*

**Line 55:** Provide specifics on what data was incorporated – FIB, epidemiological, environmental, beach closure data and statistics. Please specify which nowcast models were used for this analysis.

***Response****: An exhaustive listing of the environmental surrogate variables at each site is too lengthy to put in the main text; it appears in the appendix. The listing of the nowcast methods used in this study is provided in section 3 (“Methods”).*

**Line 57: P**rovide clarity on the phrase “may be lost” – do you mean inadvertently overlooked, not at all considered for specific reasons, or underestimated?

***Response****: Revised to clarify that we are looking for methods that consistently produce accurate predictions across a range of sites. The phrasing “may be lost” was unfortunate. We meant to say that focusing on one site doesn’t address consistency of the performance across sites.*

**Lines 59-65:** This paragraph is not needed and should be omitted from the manuscript. Consider replacing this with a concluding paragraph for the Introduction stating the overall purpose of the study and hypothesis or goal of the study.

***Response****: This paragraph was replaced with a paragraph that describes how the chosen performance measures specifically address the goals of the study.*

**Lines 67-68:** Explain the phrase “two distinct regions” – distinct based on geography, hydrology, water quality influences?

***Response****: We meant “distinct” in terms of geography, water quality influences, and land use patterns. But we think the manuscript reads better if we strike the word “distinct” and so we have.*

**Line 69:** Counties are along Lake Michigan and not “in” Lake Michigan.

***Response****: Indeed.*

**Line 71:** Provide more specifics on “automatic sensing,…and manual sampling” methods.

***Response****: Edited this paragraph to clarify that the data came from three sources: manually sampled, automatically measured by remote sensing, and automatically generated from hydrological and atmospheric models.*

**Line 73:** I did not see a listing of the covariates in the Appendix. Please add and specify Figure/Table # within the text.

***Response****: Whoops - that section of the appendix had not been included with the manuscript for review. It has been reattached.*

**Line 74:** There is a lot of detail provided in the Site Descriptions section. Please review this section and see if some of the site information is actually needed or if it can be moved to supplemental text. Providing so much information detracts from the main point regarding site-specific influences on beach water quality and land use. Also, capitalize “beach” when part of a proper name.

***Response****: Capitialized “Beach” as suggested. Struck all discussion of beach features that are not believed to affect the water quality such as structures and amenities.*

**Line 115:** Provide date range for both Memorial Day and Labor Day.

***Response****: Done.*

**Line 120:** Revise citation format.

***Response****: Done.*

**Lines 123-129:** Are all these websites available for public use? Provide specific information on where this data originates – is it collected and reported by local health departments, Federal agencies, or other groups?

***Response****: Edited the text to clarify that the data are all produced by federal agencies and are available from public web services.*

**Line 139-141:** Review this sentence and consider revising or providing further explanation regarding “scientific judgment regarding phenomena.”

***Response****: We agree that the term “scientific judgement” is hazy, but after consideration it still seems to be the best available phrasing. Clarified that the judgement applied has to do with how a surrogate variable might affect the FIB concentration. The examples that immediately follow this sentence are included to illustrate the idea of “scientific judgement”.*

**Lines 147-149:** The end of this sentence seems to be missing text.

***Response****: Revised this paragraph so that it didn’t seem to end so abruptly.*

**Line 150:** Were these observations collected by WI-WSC or some other agency?

***Response****: Added text to clarify that the manually sampled data was collected by the University of Wisconsin-Oshkosh, Northland College, and Bayfield County.*

**Methods section:** There are several embedded tables throughout this section. Is that in accordance with journal format? Also, this section introduces many terms, definitions, and abbreviations. Consider including a table for easy references that defines the various terms and abbreviations.

***Response****: Added some information to the table of regression methods. The embedded tables are in accordance with the journal’s submission format.*

**Line 174:** The BAV has already been defined so there is no need to redefine abbreviation.

***Response****: Struck the redefinition.*

**Line 197:** Fix spelling of “exceedances.”

***Response****: Fixed.*

**Line 223:** Consider moving some text from *3.2.4 Listing of model algorithms* to supplemental text.

***Response****: We considered this move but decided to keep the text in the main body at this time. The descriptions of the methods used are considered critical to assessing the study’s results and implementing its guidance.*

**Line 301 (Table 2):** Please replace the phrase “Hika here” with “Hika Beach.” Is the information in Table 2 presented elsewhere for all studied beaches? If so, I am not sure why Table 2 is needed. If not, this information needs to be included for all beaches and not just Hika Beach.

***Response****: Fixed the text. We justify including this table only for Hika and not the other beaches on the grounds that this is an example that makes the somewhat abstract discussion of section 3.4 more concrete.*

**Line 311:** Fix spacing error.

***Response****: Done.*

**Line 314:** Revise for missing text in sentence.

***Response****: There was no missing text but a period had been replaced by a comma, which gave the appearance of missing text. The comma in the equation was corrected to a period.*

**Results:** Revise the Results section so that it only includes the “results” and not additional model explanation and methodology.

***Response****: The journal allows for “Results and discussion” and Conclusions. Some revision was combined with changing the “Results” section to “Results and discussion” to clarify that it contains both results and discussion.*

**Line 386:** Consider providing minimal text on the percentage of accurate predictions, false negatives, and other relevant information.

***Response****: A new table was added, listing the specificity, sensitivity, and proportion of correct predictions by site for the AL, GBM-OOB, and GBM-CV methods. This was accompanied by text interpreting the table and discussing its significance to the choice of a modeling method and a decision threshold.*

**Line 449:** Correct the spelling of “concentration.”

***Response****: Done.*

**Line 450-452:** This paragraph belongs in Methods and should not be the concluding paragraph of this manuscript. Move this paragraph to Methods and add a summary conclusion in its place. Consider adding text summarizing the extent that the incorporation of these regression models improves current nowcast models and/or traditional analysis methods.

***Response****: As suggested, this paragraph was revised for clarity and moved to the Methods section.*

**Table 1:** Define regression method abbreviations.

***Response****: The abbreviations GBM-CV and GBM-OOB were clarified with the addition of text in the table. The other abbreviations are adequately described here and in the text.*

**Figures 2 & 3:** Increase font size so text is readable.

***Response****: Done.*

I appreciate the opportunity to review this manuscript. Please let me know if there are any questions or concerns.

Thank you,

Carrie E. Givens, Ph.D.

Microbiologist, Michigan USGS Michigan Water Science Center