

Interactive comment on “Interannual hydroclimatic variability and its influence on winter nutrients variability over the southeast United States” by J. Oh and A. Sankarasubramanian

Anonymous Referee #1

Received and published: 5 January 2012

General comments:

This paper demonstrates approaches that can be used for season-ahead nutrient predictions conditioned on climate forecasts. Specifically, the authors use precipitation forecasts to predict total nitrogen (TN) in winter for 18 watersheds in the Southeast US using two approaches: Principal Components Regression (PCR) and Canonical Correlation Analysis (CCA). The paper includes a discussion of how including an antecedent flow predictor can improve the forecast, as well as potential applications.

This is a good paper that fits within the scope of HESS and makes an important scien-

C5622

tific contribution to nutrient forecasting. However, the paper would benefit from a major revision, focusing on two main fronts: (i) strengthening the connection between the research purpose and the results/conclusions, and (ii) improving the manuscript flow and organization. The two are related, and improvement in (ii) would lead to automatic improvements in (i). Specific suggestions for (ii) are included in the Specific Comments. In terms of (i), the authors need to better articulate the purpose of the paper, and make a closer connection between the presented results (i.e., figures/tables) and interpretations/conclusions. The stated purpose is: “. . . on understanding the process controls in estimating winter nutrient loadings. . .”, but the majority of the manuscript is devoted to presenting and validating the PCR and CCA models (4 out of 6 tables and 5 out of 8 figures). From this, it seems like the purpose of the paper is to demonstrate (and compare?) two models that can incorporate precipitation forecasts to provide nutrient forecasts. This is important because despite forecast improvements, they are currently underutilized, and there aren't many examples of how they can be used for water quality forecasting. In addition, to better highlight the main points and conclusions, the results and descriptions need to be sharpened and focused (specific examples follow in Specific comments).

One main technical question arose: how are the climate forecast errors considered? You consider R2LOADEST and R2CCA/PCR, but do these consider the possibility that the precipitation forecast for JFM is “wet” but it turns out to be “dry”? (For more details and specific places in the manuscript related to this comment, see Specific Comments).

Specific comments:

Title: Consider removing second “variability” (perhaps replace with “concentrations” or “loadings”). Consider revising to better reflect purpose (see subsequent comments).

Abstract: Define abbreviations, remove “Table 2”, and revise sentence: “Stations that have very high R2(LOADEST) (>0.8) in predicting the observed WQN loadings during the winter (Table 2) exhibit significant skill in loadings.”

C5623

Introduction: This does a good job of motivating the study, but could use a revision for sentence- and word-smithing to insure that there is logical flow and natural transitions between ideas.

p. 10937 line 11. Remove sentence or edit: "Thus, it is critical to estimate the seasonal nutrient loadings conditioned on the expected runoff from nonpoint sources."

p. 10937 line 16. Define ENSO, SST in main text.

p. 10937 – In this section, consider revising to make a more explicit connection between SSTs, ENSO, and the precipitation forecasts that you will be using to make nutrient forecasts: SSTs drive ENSO, which drives precip/streamflow patterns, which drives nutrient concentrations.

p. 10938 lines 18-21. Clarification is needed to explain how the purpose of the paper is to "understand the "controls" that are required. . .". The paper does not seem to explore "controls", i.e., controlling processes or mechanisms driving nutrient concentrations. That might come into play if you were testing a suite of predictors to find the "best" combination for nutrient forecasting, but that does not seem to be the focus (you look at precip, and then briefly at flow and ENSO). It seems to be demonstrating statistical tools that one can use to incorporate seasonal forecasts to develop nutrient forecasts. Or to compare different statistical tools to see what types of models are best suited to incorporate seasonal forecasts for nutrient forecasting.

p. 10938 lines 21-24. Replace "climate forecasts" with "precipitation forecasts". Replace "land surface conditions" with "flow conditions".

Section 2. Data Sources. Consider reorganizing to clarify that there are 2-steps in this section. For instance: First, streamflow is used with the LOADEST to simulate a full-record of water quality; then it is used as a predictor for season-ahead nutrient forecasting. One idea is making Section 2. "Study Area and Data". You could start with the study area description, then introduce the data sources in light of the two steps: (i)

C5624

Water quality simulation and (ii) Season-ahead forecast predictors.

Section 2.3. Briefly, what kind of model is LOADEST? E.g., Mechanistic, empirical, other?

p. 10940 lines 24-27: Consider revising this to be less technical and more descriptive. For instance, consider removing "dtime" and instead describe why it was appropriate to exclude a time trend in the regression. Further, you indicate what's not included (i.e., the time trend), but is there a concise way to summarize what predictors are included? You could describe them instead of listing the model numbers (i.e., 1, 2, 4 and 6). Listing the predictors gives a sense of what the important parameters are for simulating the water quality data.

Table 2. Consider revising the table to only include the most necessary information, e.g., may not need station number and model number. The coefficients values may not be necessary, especially if we don't know what each predictor is. Or could you summarize the results for most of the models, but only show results from 2 locations, say the best fit and the worst fit. E.g. station 17 vs. station 18 (?).

p. 10941 line 5. Replace "in predicting" with "of". The reason being that "goodness-of-fit" only implies a good fit to the data, it doesn't test the model in a predictive mode.

p. 10941 lines 6-23. Revise paragraph to better highlight the key points and procedures. Point out the key results for the JFM LOADEST simulation from Table 2. Much later on p. 10953 you note: "stations that have very high R2 (LOADEST) (>0.8) in 20 predicting the observed WQN loadings during the winter (Table 2) exhibit significant skill in loadings." In this paragraph, you should point out the stations with low R2 values that you refer to later (e.g., station 5, 6, 18. . .).

Section 2.4. Technical comment: How accurate/reliable are the 3-month ahead precipitation forecasts that you use? That is, let's say the retrospective forecast for JFM of 1989 was wet, is it possible that it actually ended up being dry? This would con-

C5625

tribute to errors in your results. Consider revising the technical description of how the forecasts are constructed (e.g., “To force the ECHAM4.5 with SST forecasts, retrospective monthly SST forecasts were developed based on the observed SST conditions in that month based on the constructed analogue approach”) to better highlight how that method affected the confidence/correctness of the forecasts that you use. (E.g., consider: “By forcing the ECHAM4.5 with an SST approach based on xyz, it insured that the precip forecasts were lmnop...”)

Figure 2. Good figure.

Table 3. It's very interesting that there is little to no difference between the variance explained by PC1 of Q or TN. This helps to justify what you are doing (i.e., going directly from precip to TN, by-passing concurrent streamflow). I think this could be highlighted in the text more. (See 10943 line 10-12).

Figure 1 caption: Add “precipitation forecast” before “grid points”.

Section 3. Technical comment: From the paper you indicate: “. . . we first identify relevant grid points (Table 3) of JFM precipitation forecasts that have statistically significant correlation with JFM observed precipitation for each watershed. Nearest grid points that are significantly correlated to each watershed (Fig. 1) are selected.” Is this how you insure that the forecasts are accurate/reliable? It might help to add information on why you did this here. In the Discussion you later say: “By selecting grid points of precipitation forecasts that are statistically significant with the observed precipitation in the basin, we ensure that the skill in predicting nutrient loadings is related to the basin process as well.” I'm unsure of what you mean by “related to the basin process”. If this step is not to insure that the forecasts are accurate/reliable, then you would have to add the forecast error too (i.e., Independent errors would be from (i) LOADEST simulation, (ii) low-dimensional model, (iii) forecast model).

Section 3.1. It would be helpful if you could point out the difference/advantage of developing these low-dimensional models as compared to just using the regressions

C5626

already developed from LOADEST. This might become clear when you add information about the LOADEST predictors. The low-dimensional models are suited to using the GCM forecasts, but do any of the LOADEST models contain precipitation, or other large-scale information, as predictors?

p. 10943, line 21. Consider revising this sentence, not sure what is meant by “marginal bias”.

Section 3.1. It might help here to clarify the paper purpose here. (See earlier comments).

Table 5, Figure 3, Figure 4. This table and two figures show the validation for PCR models in terms of RMSE and R2 from LCV and SSV. I'm wondering if there is a way to distill these results to show the key conclusions. For instance, with PCR the stated conclusion is: “Thus, based on two different validation methods, we understand that eleven stations (2–4, 7–11 and 13–15) exhibit statistically significant skill in predicting the observed WQN loadings using the PCR model developed separately for each site.” These are conclusions based mainly on the validation method that had the most conservative result – namely Figure 4 (it found less sites significant than did Figure 3). It might be interesting to only show the results from Figure 3 that are different than what you found in Figure 4 (e.g., site 17 is significant in Figure 3 but not in Figure 4, why?). What are the most important values in Table 5 RMSE, i.e., which parts bolster your conclusions?

Table 5, Figure 5, Figure 6. This table and two figures show the validation for CCA models. Similar to above comments, can these be distilled? Would it make sense to only show the more conservative results – i.e., where less sites are significant (Figure 6), and then only the stations that show different results from Figure 5? Also, why do you show the result from the LCV for CCA as a map (Figure 5) versus the LCV for PCR with box plots (Figure 3)? You might want to comment on why you think the SSV results in less significant sites than the LCV.

C5627

p. 10947 line 7. Replace “However” with “Here”.

p. 10949 line 25. Replace “CCA model in explaining” with “the CCA model to explain” Section 4.3. This section has interesting information, but it comes as a bit of a surprise in the manuscript. Although it is mentioned in the Abstract, it is not mentioned in the Introduction, Section 3, or Section 4. Up to this point, the majority of the paper indicates that you are only going to look at precipitation forecasts as a predictor of water quality. Was this work motivated from the results of the low-dimensional models? Consider removing or revise so that this fits more logically. (Perhaps move this to the discussion section?) Or introduce it earlier and indicate that Section 4 includes results from two parts: (i) using ECHAM4.5 precipitation forecasts alone and (ii) adding antecedent streamflow.

Section 4.4. Similar to section 4.3 – this section is interesting but does not fit where it is in the manuscript. Consider removing or revise so that this fits more logically. (One idea: This is more of a motivation – it could almost go after Figure 2. Fig 2 establishes that there is a strong precip-TN correlation, but this would be going one step further by saying there is also a strong correlation between ENSO and TN. Further, it would be good because precip forecasts tend to have more skill during ENSO years).

Section 5. Consider removing some information in the discussion that is more of a summary than discussion (or move to a summary or methods section): e.g., “Since obtaining long continuous records of daily observations of nutrients is difficult particularly over a large region, we employed simulated nutrient loadings from the LOADEST model to understand the role of climate variability in modulating the interannual variability in nutrients over the SEUS. However, to account for the errors in the LOADEST model in predicting the observed WQN database, the reported skill measures (Eqs. 2 and 3), R2 and RMSE, are adjusted for both LOADEST model error as well as the error of the low-dimensional models.”

I’m uncertain about the statement made here in the Discussion: “Thus, the intent of

C5628

this study is to understand how well climate and basin storage conditions control the seasonal TN loadings rather than developing a skillful nutrient forecasts using low-dimensional models.” The majority of the paper is devoted to evaluating the skill of nutrient forecasts using low-dimensional models, so I’m uncertain what point this is trying to get across. This goes back to the question of the purpose of the paper. I interpret the intent of the study to demonstrate the potential of climate forecasts for water quality prediction. The results imply that while these models may not be ready for operational forecasting, there is great potential to further develop water quality forecasting tools.

The paragraph that begins: “Perhaps the most important utility...” makes a good case for the utility of the season-ahead forecasts of nutrient loadings. This might be a good place to discuss the comparison between the results of the CCA and the PCR. For instance, if a manager wanted to use one of your models for water quality trading, which model – the CCA or the PCR or a hybrid – and for what cases would you recommend each? What are the advantages/disadvantages of each, which had better results for what? It would strengthen your paper if you could link these applications more closely to the analysis and results of this paper – even if it is only hypothetical. This might also be a good opportunity to motivate the results from section 4.3. You could indicate that to implement in practice, you need to improve the forecasts, which led to including the antecedent conditions.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 10935, 2011.

C5629