

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

**Interannual  
hydroclimatic  
variability**

J. Oh and  
A. Sankarasubramanian

# Interannual hydroclimatic variability and its influence on winter nutrients variability over the southeast United States

**J. Oh and A. Sankarasubramanian**

Department of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC 27695-5908, USA

Received: 17 November 2011 – Accepted: 5 December 2011 – Published: 12 December 2011

Correspondence to: A. Sankarasubramanian (sankar\_arumugam@ncsu.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Abstract

It is well established in the hydroclimatic literature that the interannual variability in seasonal streamflow could be partially explained using climatic precursors such as tropical Sea Surface Temperature (SST) conditions. Similarly, it is widely known that streamflow is the most important predictor in estimating nutrient loadings and the associated concentration. The intent of this study is to bridge these two findings so that nutrient loadings could be predicted using season-ahead climate forecasts forced with forecasted SSTs. By selecting 18 relatively undeveloped basins in the Southeast US (SEUS), we relate winter (January-February-March, JFM) precipitation forecasts that influence the JFM streamflow over the basin to develop winter forecasts of nutrient loadings. For this purpose, we consider two different types of low-dimensional statistical models to predict 3-month ahead nutrient loadings based on retrospective climate forecasts. Split sample validation of the predictive models shows that 18–45% of interannual variability in observed winter nutrient loadings could be predicted even before the beginning of the season for at least 8 stations. Stations that have very high  $R^2_{(\text{LOADEST})}$  ( $>0.8$ ) in predicting the observed WQN loadings during the winter (Table 2) exhibit significant skill in loadings. Incorporating antecedent flow conditions (December flow) as an additional predictor did not increase the explained variance in these stations, but substantially reduced the RMSE in the predicted loadings. Relating the dominant mode of winter nutrient loadings over 18 stations clearly illustrates the association with El Niño Southern Oscillation (ENSO) conditions. Potential utility of these season-ahead nutrient predictions in developing proactive and adaptive nutrient management strategies is also discussed.

## 1 Introduction

Concerted efforts to improve national water quality conditions resulted in the enactment of the 1972 Clean Water Act with section (303)d requiring the states and territories to

**HESSD**

8, 10935–10971, 2011

### Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



list impaired water bodies and develop Total Maximum Daily Loads (TMDLs) for these waters. Despite these efforts and frequent updates to TMDLs, the US Environmental Protection Agency's recent update reveals that nutrients affect 20 % of impaired and 12 % of the assessed river miles (EPA, 2006). The increase in aquatic nutrients might result from population growth as well as from increased fertilizer application (Meybeck 1982; Vitousek et al., 1997). However, natural variability associated with weather (e.g. hurricanes) and climatic events (e.g. El Niño) could also induce significant increase in nutrient concentrations beyond critical levels (Chen et al., 2007) even if the basin is not experiencing any pressure from urban development or changes in agricultural practice. Thus, it is critical to estimate the seasonal nutrient loadings conditioned on the expected runoff from nonpoint sources.

The National Research Council (NRC, 2001, 2002) has emphasized that a detailed understanding of various sources of uncertainties, including the role of climate change and climate variability, is required for improving water quality prediction in natural systems. One of the dominant and well understood modes of global climatic variability is ENSO that has a periodicity of 3–7 yr and exhibits anomalous warm/cold SST conditions in the equatorial Pacific, thereby modulating the climate particularly in the tropics and sub-tropics (Ropelewski and Halpert, 1987). Considerable research now exists on the recurrence and regime structure of ENSO and its teleconnections to rainfall/streamflow, and their potential predictability of interannual hydroclimatic variability over the United States (Ropelewski and Halpert, 1987; Dettinger and Diaz, 2000; Devineni and Sankarasubramanian, 2010). It is also well known that instream nutrient concentration and loadings primarily depend on streamflow variability (Borsuk et al., 2004; Paerl et al., 2006; Lin et al., 2007) and antecedent flow conditions (Vecchia, 2003; Alexander and Smith, 2006). Recent studies on the relationship between coastal water quality conditions and SST conditions also show that there is a strong association between climatic modes and concentrations of phosphorous (Childers et al., 2006), aquatic vegetation (Cho and Poirrier, 2005), and chlorophyll and phytoplankton levels (Arhonditsis et al., 2004). However, systematic research in associating climatic

**Interannual  
hydroclimatic  
variability**J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



variability to instream nutrient variability and utilizing that linkage to estimate season-ahead nutrient loadings is very limited.

Most of the studies on estimating instream nutrient concentrations have focused primarily on predicting the average annual concentrations using runoff and various basin attributes (Smith et al., 1998, 2003; Mueller and Spahr, 2006; Mueller et al., 1997). Studies have also recommended approaches to predict daily and seasonal loadings and concentrations nutrients using streamflow and their time of observation (Cohn et al., 1992; Runkel et al., 2004). However, these nutrient models rely on the observed information (e.g. streamflow) during that season, which has limited utility in developing season-ahead estimates of nutrients. Findings from the hydroclimatic literature clearly show that interannual variability in streamflow can be predicted by developing low dimensional models contingent on SST conditions (Devineni et al., 2008) as well as using precipitation forecasts from General Circulation Models (GCMs) (Sankarasubramanian et al., 2008). Similarly, water quality literature emphasizes that streamflow is the most important descriptor in explaining nutrient variability (Cohn et al., 1992; Runkel et al., 2004; Cohn, 2005). To our knowledge, this is the first effort that associates the interannual variability in all of the above noted three variables – climate, streamflow and total nitrogen (TN) – to develop TN forecasts over a region. The purpose is to understand the “controls” that are required for developing a skillful seasonal nutrient forecasts and also to assess how the skill in hydroclimatic predictions translate to skill in nutrient forecasts over the regional scale. For this purpose, we consider two low-dimensional models that consider season-ahead climate forecasts and land surface conditions as predictors for developing season-ahead estimates of winter nitrogen loadings over the SEUS.

The manuscript is organized as follows: a brief description of climate forecasts, streamflow and water quality databases employed in the study is first provided in Sect. 2. Following that, Sect. 3 provides the details of the low-dimensional statistical models and skill measures utilized in developing and evaluating the season-ahead nitrogen loadings forecasts. In Sect. 4, we present results from the winter nutrient

**Interannual hydroclimatic variability**

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



forecasts developed using the low-dimensional models. Next, we discuss about the potential implications of the findings in the context of developing adaptive water quality management plans. Finally, in Sect. 6, we summarize the findings and conclusions from the study.

## 2 Data sources

In this section, we discuss various hydroclimatic and water quality databases employed for associating climate forecasts with the nutrient loadings over 18 watersheds from the Southeast US.

### 2.1 HCDN streamflow database

Given the intent of the study is to associate interannual variability in winter nutrient loadings to climatic variability, we focus our analysis on 18 undeveloped basins over the Southeast United States (SEUS) from the Hydro-Climatic Data Network (HCDN) database (Slack et al., 1993). Daily streamflow records in the HCDN basins is purported to be relatively free of anthropogenic influences such as upstream storage and groundwater pumping and the accuracy ratings of these records are at least “good” according to United States Geological Survey (USGS) standards. The HCDN database contains the mean daily discharge for about 1600 sites across the continental United States with an average length of 48 yr. Figure 1 shows the location of 18 HCDN stations and Table 1 provides the list of the 18 stations considered in this study along with their drainage areas. Since the streamflow data ( $Q$ ) in the HCDN database is available only up to 1988, we have extended it up to 2009 based on the USGS historical daily streamflow database.

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2.2 WQN Water Quality Network database

USGS provides national and regional descriptions of stream water quality conditions in Water Quality monitoring Network (WQN) across the nation (Alexander et al., 1998). The WQN database comprises water quality data from USGS monitoring networks from both large watersheds (National Stream Quality Accounting Network, NASQAN) and minimally developed watersheds (Hydrologic Benchmark Network, HBN). We employ the observed daily concentrations of Total Nitrogen (TN) available for these 18 stations from the NASQAN. Observed streamflow during the time of sampling is also available as part of the WQN database. The available water quality data varies from 10–30 yr depending on the measured water quality variable and station. By ensuring the selected watersheds are from HCDN basins, we basically ensure that both the streamflow and water quality data are minimally affected by anthropogenic influences. For additional details about WQN, see Alexander et al. (1998). The selected 18 HCDN stations have observed TN concentrations for 12–22 yr (Table 1). However, the number of samples for each station ranges from 54–152 daily observations with an average of five-seven observations per year.

## 2.3 Simulated nutrients database

Though nutrient data in the WQN database is available for 12–23 yr over 18 watersheds (see Table 2), their samplings are intermittent. Using the daily observation over this period, we first obtain continuous daily nutrients for the observed period using the LOADING ESTimation (LOADEST) program developed by USGS (Runkel et al., 2004). The LOADEST model allows the user to select the best-fitting regression model from eleven predefined regression models using the Akaike Information Criterion (AIC) (Akaike, 1974). Five regression models that include “dtime” term are not appropriate to use for extrapolation since those models incorporate a linear time trend. Therefore, the simulated nutrient loadings based on the remaining regression models (i.e. model forms: 1, 2, 4 and 6) in the LOADEST program do not have any time trend. For details

### Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



on model forms, see Runkel et al. (2004). Table 2 shows the “goodness of fit” statistics (coefficient of determination ( $R^2$ ) and AIC) in predicting the observed daily loadings in the WQN database (Table 1) and the coefficients of the best-fitting regression model for TN for the selected 18 stations. From Table 2, we infer that  $R^2$  ranges from 0.83–0.97 indicating good fit in predicting the observed daily loadings over 18 stations.

To relate retrospective climate forecasts (discussed in the next section), we estimated the daily TN loadings from 1957–2009 using the observed streamflow data available from the extended HCDN database and the best fitting regression model (in Table 2). The simulated daily loadings obtained from the LOADEST model over the period 1957–2009 are aggregated during JFM to develop winter loadings ( $L_t$ ) of TN. Given that the estimated loadings are based on Adjusted Maximum Likelihood Estimation (AMLE) procedure in the LOADEST model, the simulated daily and the aggregated winter loadings are statistically unbiased (Cohn, 2005). We also computed  $R^2_{(\text{LOADEST})}$  and the root-mean square error ( $\text{RMSE}_{(\text{LOADEST})}$ ) for the simulated winter TN loadings obtained from the LOADEST model. For additional details on computing errors in seasonal predictions, see Cohn (2005). Further, to ensure that there is no trend in the winter loadings, we performed Mann-Kendall test. At 1 % significance level, null-hypothesis with Kendall’s tau being not equal to zero was rejected in all of the sites for TN. We also performed regional Mann-Kendall test to account for spatial correlation among the 18 stations (Douglas et al., 2000). The p-value for TN is 4 % indicating no-trend at the regional level. Our study will consider the simulated winter TN loadings ( $L_t$ ) available during 1957–2009 for relating the interannual hydroclimatic variability to nutrient variability over 18 stations in the SEUS.

## 2.4 Climate forecasts database

For developing season-ahead TN forecasts, we utilize the retrospective winter precipitation forecasts from ECHAM4.5 General circulation model forced with constructed analogue SSTs ([http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/.ECHAM4p5/.Forecast/ca\\_sst/.ensemble24/.MONTHLY/.prec/](http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/.ECHAM4p5/.Forecast/ca_sst/.ensemble24/.MONTHLY/.prec/), International Research

Institute of Climate and Society (IRI) data library) (Li and Goddard, 2005). Retrospective precipitation forecasts from ECHAM4.5 are available for 5 months in advance for every month beginning January 1957. 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. For additional details on forcing ECHAM4.5 using constructed analogue SST forecasts, see Li and Goddard (2005). Figure 1 also shows the locations of 56 grid points of precipitation forecasts from ECHAM4.5 along with their latitude and longitude over SEUS. For this study, we utilize only the forecasted mean (which is obtained by computing the average of 24 ensembles) of winter retrospective precipitation forecasts issued in the beginning of January for developing 3-month ahead retrospective nutrient forecasts over the period 1957–2007.

### 3 Low-dimensional models development and performance validation metrics

Given that winter streamflow over the SEUS is predominantly rainfall driven with limited snow accumulation, we hypothesize that precipitation is the primary driver in controlling the JFM loadings. To verify this, we correlate simulated JFM loadings with both observed precipitation (Fig. 2) and principal components of the forecasted precipitation from ECHAM4.5 (Table 3). In this study, we only employ Spearman rank correlation for performing all correlation analyses. Similarly, the computed rank correlation was checked for statistical significance (i.e.  $1.96/(n - 3)^{0.5}$  at 95 % confidence interval, where “ $n$ ” denotes the number of data points used in calculating the correlation). Thus, the computed correlation in Fig. 2 needs to be greater than 0.29 ( $n = 50$ ) to indicate statistically significant relationship between the observed precipitation and simulated loadings.

From Fig. 2, we infer that the correlation between observed precipitation and simulated loadings is statistically significant and greater than 0.55 for all the basins. Given this dependency, we first identify relevant grid points (Table 3) of JFM precipitation

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





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. The variance explained by the first principal component (PC1) of the precipitation forecasts from these grid points is around 74–95 % indicating the strong spatial correlation among the gridded forecasts. Further, from Table 3, we also infer that rank correlations between PC1 of precipitation forecasts with streamflow, and seasonal loadings of TN are statistically significant for all stations ( $>0.29$ ) with the only exception being station 18. The primary reason for such low correlation for station 18 is due to the poor coefficient of determination from the LOADEST model (Table 2) in predicting observed WQN data. To summarize, Fig. 2 and Table 3 provide the scope for using the low-dimensional components of precipitation forecasts for developing season-ahead forecasts of TN loadings over 18 selected stations.

### 3.1 Low-dimensional models

Given that our interest is primarily in understanding how large-scale hydroclimatic information could be utilized for seasonal nutrient predictions over the SEUS, we consider two low-dimensional models: principal components regression (PCR) and canonical correlation analysis (CCA). Low-dimensional models reduce the correlated predictors and predictands so that a subspace of uncorrelated predictors and predictands could be used for regression model development (Tippet et al., 2003; Sankarasubramanian et al., 2008). Further, these low-dimensional models also recalibrate the GCM forecasts so that any marginal bias in predicting the observed precipitation could be adjusted based on the regression model (Landman and Goddard, 2003). Brief description of the low-dimensional models is provided next.

#### 3.1.1 Principal Components Regression (PCR)

PCR, which is otherwise known as Model Output Statistics (MOS) (Wilks, 1995), eliminates systematic errors and biases in GCM fields and also recalibrates the principal

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



components (PCs) of GCM fields to predict the hydroclimatic variable of interest using regression analyses. The predictand could be streamflow ( $Q_t$ ) or loadings ( $L_t$ ) over a watershed. Since the gridded precipitation forecasts over a given region are spatially correlated, employing precipitation forecasts from multiple grid points as predictors would raise multicollinearity issues in developing the regression. To avoid this, we employ PCR based on Eq. (1):

$$\ln(L_t) = \hat{\beta}_0 + \sum_{k=1}^K \hat{\beta}_j \cdot \text{PC}_t^k + \hat{\varepsilon}_t \quad (1)$$

where  $L_t$  denotes the estimate of daily average TN loadings during the JFM season in year “ $t$ ”,  $\text{PC}_t^k$  denotes the “ $k$ ”th PCs from the retained “ $K$ ” PCs of precipitation forecasts and  $\hat{\beta}$ s denote the regression coefficients whose estimates are obtained by minimizing the sum of squares of error. We employ step-wise regression to select “ $K$ ” PCs out of the rotated grid points of precipitation (given in Table 3) for developing the PCR model. From Table 5, we infer that most of the stations (except stations 8, 10 and 11) require only up to the first four principal components for developing the PCR model.

### 3.1.2 Canonical Correlation Analysis (CCA)

In PCR, we develop separate regression models for each site. Given that the predictands, the winter loadings, across the basins are also spatially correlated, one could utilize that information to develop a reduced set of regression models. This could help in utilizing the inter-site correlations to develop many (multiple predictands)-to-many (multiple predictors) regression relationships. Consider winter loadings available from “ $m$ ” sites represented by  $L^T = (L_1, L_2, \dots, L_m)$  (dimension:  $n \times m$ ) whose corresponding “ $p$ ” grid points of precipitation forecasts ( $p > m$ ) are represented as  $X^T = (X_1, X_2, \dots, X_p)$  (dimension:  $n \times p$ ), then canonical correlation analysis finds a linear combination of the “ $p$ ” predictors,  $Y^* = b^T Y$ , that maximally correlates with the linear combination of “ $m$ ” predictands ( $X^* = a^T X$ ). Mathematically, the canonical

**Interannual hydroclimatic variability**

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



correlation is obtained by choosing the vectors  $\mathbf{a}$  and  $\mathbf{b}$  that maximizes the relationship  $\left(\mathbf{a}^T \sum_{xy} \mathbf{b}\right) / \left\{ \left(\mathbf{a}^T \sum_{xx} \mathbf{a}\right) \left(\mathbf{b}^T \sum_{yy} \mathbf{b}\right) \right\}^{1/2}$  where  $\sum$  denotes the variance-covariance matrix between the two variables in the subscript. For a detailed mathematical treatment of CCA, see Wilks (1995). Number of components from “ $m$ ” predictands and “ $p$ ” predictors to be retained for the regression is decided based on step-wise regression. Squared values of canonical correlation represent the percentage of variance explained in each predictand by the predictors under that dimension. Thus, the skill in predicting the loadings for each site could be obtained based on the precipitation forecasts by developing a reduced set of models.

Before performing CCA, we first group the basins based on k-means clustering (Hartigan and Wong, 1979) so that CCA could be performed on each cluster. Based on clustering, four groups were identified (Table 4) with the sites having the highest loadings placed under group 1 and the lowest average loadings placed under group 4. Separate CCA was performed for each group. For instance, CCA on group 1 is performed on loadings from two sites ( $m = 2$ ) and the corresponding grid points of precipitation forecasts for the two sites (#13 and #18) from Table 3 are combined ( $p = 20$ ) as predictors. The skill in predicting the winter loadings for each station is evaluated based on two different skill scores, which are discussed next.

### 3.2 Skill scores for nutrients forecasts

To evaluate the skill in predicting the interannual variability in winter TN loadings using climate forecasts, we consider two error metrics – coefficient of determination ( $R^2$ ) (Eq. 2) and root mean square error (RMSE, defined in Eq. 3) per unit area of the watershed ( $A$ ). These metrics need to account both sources of errors: error in predicting the observed JFM nutrient loadings from the WQN database using the LOADEST program ( $R^2_{(\text{LOADEST})}$ ,  $\text{RMSE}_{(\text{LOADEST})}$ ) (see Table 2) and the error in predicting simulated JFM nutrients from LODEST based on the low dimensional model ( $R^2_{(\text{PCR/CCA})}$ ,

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

$RMSE_{(PCR/CCA)}$ ). Since these two models are developed independently,  $R^2$  and  $RMSE$  in predicting winter nutrient loadings using climate information could be expressed as follows:

$$R^2 = R^2_{(LOADEST)} \cdot R^2_{(PCR/CCA)} \quad (2)$$

5  $RMSE = (RMSE_{(LOADEST)} + RMSE_{(PCR/CCA)})/A \quad (3)$

For each station,  $R^2_{(PCR/CCA)}$ ,  $RMSE_{(PCR/CCA)}$ , were computed based on the estimated TN loadings from the low-dimensional models and the simulated winter TN loadings for the period 1957–2006. Thus, we compute skill measures using Eqs. (2) and (3) to quantify our ability to explain the interannual variability in winter TN loadings using precipitation forecasts from ECHAM4.5.

## 4 Results and analyses

To ensure that the skill in forecasting winter nutrients is reliable, we evaluate the low dimensional models based on two different types of validation namely, leave- $X$  out cross-validation (LCV) and split-sample validation (SSV). Both these methods are commonly adapted in forecasting literature for validating the model (Wilks, 1995).

### 4.1 TN loadings forecasts based on PCR models

For validating the PCR models under LCV, the methodology suggested by Towler et al. (2009) is modified to evaluate the skill of the model over 51 yr (1957–2008). The LCV steps for PCR models are described as follows: (i) 10 % of the data (5 yr) are randomly removed along with the year for which the prediction is desired, (ii) a PCR model is developed using the remaining 45 yr of loadings ( $L_t$ ) and retained PCs (iii) the developed model is then used to predict the left-out year, and (iv) steps (i) to (iii) are repeated to develop prediction for each year and skill measures ( $R^2$  and R-RMSE)

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



were computed based on the 51 yr of predicted data. This entire procedure (i)–(iv) is repeated 100 times and a box-plot of  $R^2$  (Fig. 3) and the median of RMSE (in Table 5) are presented.

Figure 3 shows the box-plot of  $R^2$  under LCV for 18 stations. Under LCV, we compute  $R^2$  based on the predicted loadings for 51 yr. Hence,  $R^2$  needs to be higher than 0.08 (correlation  $> 0.29$ ) to demonstrate statistically significant skill in predicting season-ahead nutrient loadings. However, more than 12 stations exhibit  $R^2$  greater than 0.16 over 100 trials of LCV. From Fig. 3, sixteen stations show statistically significant skill for TN. The developed PCR model under LCV explains more than 10% of interannual variability in TN loadings in all the 100 different fittings (Fig. 3) except stations 6 and 18. For the rest of the 16 sites, the correlation between the predicted nutrient loadings obtained using climate forecasts and the loadings simulated from LOADEST using the WQN database is greater than 0.29, which is statistically significant for the 51 yr of data considered. The forecasted TN loadings show statistically insignificant relationship with observed TN loadings that the correlation coefficients are 0.28 and 0.16 for stations 6 and 18 respectively. Poor goodness of fit (see Table 2,  $R^2_{(\text{LOADEST})}$ ) from the LOADEST model is the primarily reason behind the poor performance of these two stations during the winter season. Further, station 18 shows poor correlation between the principal components of precipitation forecasts and JFM loadings (Table 3). Another possible reason for such poor prediction by LOADEST model in those two stations is the limited number of years of data availability (see Table 1) with station 6 (18) WQN observations spanning 14 (12) years having a total of 56 (57) daily samples. The median RMSE (Table 5) computed under LCV also shows that error in predicting the observed WQN loadings during the winter season is lesser than  $1 \text{ kg day}^{-1} \text{ km}^{-2}$  for most of the stations.

Under split-sample validation (SSV), PCR models are developed using  $L_t$  and PCs available over the calibration period (PCR: 1957–1986) and skill measures are computed in predicting  $L_t$  during the validation period (1987–2007). Hence,  $R^2$  needs to be higher than 0.21 (correlation  $> 0.46$  for 21 yr of data) to demonstrate statistically

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



significant skill in predicting season-ahead nutrient loadings. Based on this, Fig. 4 indicates that eleven stations (2–4, 7–11 and 13–15) show significant skill in predicting TN loadings. Stations 6 and 18 perform poorly because of the limited number of years of WQN data which results in very low  $R^2$  of the LOADEST model. Apart from these two stations, stations 1, 5, 12, 16 and 17 also show insignificant skill ( $R^2 < 0.21$ ). Stations 5 and 16 perform poorly due to the poor skill of the LOADEST model during JFM (see Table 2,  $R^2_{(LOADEST)}$ ). Station 17 also has limited number of years of WQN observations (54 daily samplings, see Table 1). For stations 1 and 12 perform well under LCV, but the skill is statistically insignificant under SSV even though the simulated TN (Fig. 2) loadings exhibit significantly correlation with both observed precipitation (Table 2) and forecasted precipitation (Table 3). 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. Next, we evaluate the ability to predict the loadings in these stations under a different low-dimensional model – Canonical Correlation Analyses – that utilizes the spatial correlation in the TN loadings to develop a predictive model.

## 4.2 TN loadings forecasts based on Canonical Correlation Analyses

Four different CCA was performed on each group (listed in Table 4) and the developed models were evaluated under LCV and SSV. For LCV, we simply perform leave-5 out cross-validation instead of repeated fitting of the model (as described in Sect. 4.1 for the PCR model). Under leave-5 out cross-validation, we randomly leave five predictands and predictors along with the year for which the prediction is desired for each station under a given group (in Table 4) and CCA model is developed using the rest of the 46 years of data. The developed CCA model was employed to predict the year for which the prediction is desired. This procedure was repeated for all the years of observation under a given group to develop the CCA model estimated loadings. The  $R^2_{CCA}$  and  $RMSE_{CCA}$  was estimated between the simulated winter loadings from the LOADEST model and predicted loadings from the CCA model for each station under all the

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



four groups.  $R_{CCA}^2$  and  $RMSE_{CCA}$  were further adjusted according to Eqs. (2) (Fig. 5) and (3) (Table 5) to account for the errors in the LOADEST model in predicting the WQN database. From Fig. 5, stations 5, 6 and 18 do not exhibit statistically significant correlation in predicting the loadings from the WQN loadings. As discussed under PCR model, stations 5, 6 and 18 did not perform well because of the limited number of years of WQN data and low  $R^2$  of the LOADEST model during the winter season. Rest of the sites exhibited statistically significant relationship in explaining the observed variability in the WQN loadings. For stations 8 and 9, CCA model explain 30–40 % (correlation 0.55–0.63) of the observed winter variability in TN within the WQN database. With regard to RMSE, the performance of CCA model and PCR model is almost similar to PCR with most of the stations having an error  $1 \text{ kg day}^{-1} \text{ km}^{-2}$ .

Under SSV (Fig. 6 and Table 5), we compute  $R^2$  based on the predicted loadings during 1987–2007 using the CCA model developed over the period 1957–1986. From Fig. 6, CCA model did not exhibit any skill in predicting the winter loadings in stations 5, 6, 11–13, 16–18. Comparing this with the PCR model performance, CCA model performs similar with the exception being very low  $R^2$  in stations 11. For station 11, PCR ( $R^2 = 0.47$ ) performs significantly better than the CCA model ( $R^2 = 0.14$ ). One possible reason for such poor performance of CCA model is that station 11 has low correlation with the rest of the sites under group #4 which has 8 stations. Under SSV,  $R^2$  of the CCA model for the rest of the stations is almost similar to that of  $R^2$  of the PCR model. However, the  $RMSE_{CCA}$  is consistently higher than the  $RMSE_{PCR}$  (Table 5). This implies the conditional bias (over prediction and under prediction) of the CCA model is much higher. One possible reason for such increased conditional bias under CCA is due to increased heteroscedasticity in the observed loadings under a given group. However, the ability of CCA model in explaining the observed variance in loadings is almost comparable to that of the PCR model indicating the source of interannual variability in winter nutrients being the same across the region (discussed further in Sect. 4.4). To summarize, using ECHAM4.5 precipitation forecasts alone, we infer both low-dimensional models demonstrate significant ability in predicting the

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



observed winter TN loadings in nine coastal stations (#2–4, 7–10 and 14–15) based on two different validation methods.

### 4.3 Role of antecedent flow conditions in improving season-ahead TN forecasts

Though instream loadings primarily depend on streamflow and precipitation variability during the season, antecedent moisture/flow conditions also play a critical role in influencing the nutrient loadings from the watershed (Vecchia, 2003; Alexander and Smith, 2006). At seasonal time scales, antecedent flow conditions could be considered as the surrogate for basin storage or initial conditions in influencing the streamflow variability. To understand the role of antecedent storage conditions, we consider the observed December streamflow at each station as an additional predictor along with the gridded precipitation forecasts (Table 3) to develop nutrient forecasts for each station. Forecasts of TN loadings were developed using both PCR model (Fig. 7a) and CCA model (Fig. 7b) and the modified  $R^2$  and RMSE (Table 6) are computed (Eqs. 2 and 3) based on SSV by evaluating the model over the period 1987–2007. Comparing Fig. 7a and b with Figs. 4 (PCR model) and 6 (CCA model), we infer that only station #13 (under CCA model) has resulted in statistically significant skill by adding streamflow as an additional predictor. However, by using streamflow as an additional predictor,  $R^2$  of the CCA model substantially improved over all stations which indicate the importance of incorporating local information in spatial-dimension reduction. Further, RMSE of both PCR and CCA models are substantially reduced by adding the observed December streamflow as an additional predictor. This implies that antecedent storage/flow conditions are very critical in reducing the conditional bias in developing season-ahead TN forecasts resulting in reduced over/under prediction compared to the models developed using the precipitation forecasts alone. Thus, from a process control perspective, given the good skill in the reconstructed seasonal nutrient loadings, the interannual variability in nutrient loadings could be partially explained based on climatic variability. But, to obtain improved prediction (i.e. RMSE), it is important to incorporate both climatic variability and antecedent storage conditions in developing season-ahead nutrient forecasts.

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





#### 4.4 Source of climatic information influencing the winter TN variability

To understand the source of climate information that modulates the TN variability over the SEUS, we performed principal component analysis on the simulated loadings ( $L_t$ ) of TN over 18 stations. The first component approximately explains 59 % of total variability in TN loadings over 18 stations. It is well-known in the hydroclimatic literature that ENSO is one of the important climatic conditions that influence the winter precipitation, temperature and streamflow over the SEUS (Ropelewski and Halpert, 1987). Figure 8 shows the correlation between the first component of JFM TN loadings over 18 stations and JFM Nino3.4 – an index used to denote ENSO conditions by averaging the SST's (Kaplan et al., 1998) over the tropical Pacific (170° W–120° W; 5° S–5° N). From Fig. 8, we infer that roughly 36 % of the variability in the first principal component of nutrient loadings over SEUS could be explained purely based on ENSO conditions. ENSO plays an important role on the winter climate of the US since its peak activity typically coincides during December–February. In fact, the precipitation forecasts from ECHAM4.5 incorporate the forecasts of tropical SST conditions (i.e. Nino3.4 region), which are obtained from constructed analogue SST forecasts for forcing the ECHAM4.5. Thus, ENSO is one of the sources of climatic variability that primarily influence both JFM hydroclimatic and nutrient variability over the SEUS.

#### 5 Discussion

Analyses presented in Figs. 2–8 show that interannual variability in nutrient loadings could be predicted well before the beginning of the season contingent on the climate forecasts. 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. 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

HESSD

8, 10935–10971, 2011

### Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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),  $R^2$  and RMSE, are adjusted for both LOADEST model error as well as the error of the low-dimensional models.

Thus, the intent of 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. In principle, the analyses provided here could also be extended with other sophisticated statistical models including nonparametric and Bayesian hierarchical models to estimate the entire conditional distribution of loadings. Similarly, one could also develop nutrient forecasts by forcing the mechanistic water quality model with forecasted streamflow and water temperature which in turn could be obtained based on dynamical downscaling (Leung et al., 1999) or statistical downscaling (Devineni et al., 2008) based on the climate forecasts.

Perhaps the most important utility of the season-ahead forecasts of nutrient loadings is in promoting water quality trading. Some of the successful water quality trading programs in the country (e.g. Tar-Pamlico River basin and Neuse River basin in NC) typically allow trading nutrient loadings across different point sources as well as with nonpoint sources (e.g. farmers participating in the voluntary nutrient reduction program) through the basin-level trading association so that the seasonal/annual load caps are always met from the basin. Research on climate forecasts and water allocation clearly show that probabilistic streamflow forecasts could be effectively utilized to specify the failure probability (1-reliability) of reservoir releases as well as in ensuring the end of season target storage conditions being met with high probability (Sankarasubramanian et al., 2009). Similarly, in the context of seasonal water quality management, the developed forecasts of loadings could be used to estimate the probability of violation of target loadings for the upcoming season. One could also develop an optimal nutrient loading model such that the probability of violating the total loadings from multiple sources is within the acceptable level. Thus, utilizing season-ahead forecasts

**Interannual hydroclimatic variability**

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



of nutrient loadings and updating them throughout the season provide an opportunity to develop adaptive nutrient control strategies that ensure target nutrient loadings and desired concentration.

## 6 Summary and conclusions

The study primarily focused on understanding the process controls in estimating winter nutrient loadings by considering 18 HCDN watersheds over the SEUS. Given the discontinuous observed daily TN loadings, the study reconstructed simulated TN loadings using the LOADEST model for the winter season. The ability to predict these simulated loadings were validated with two low-dimensional models that utilize winter nutrient forecasts and pre-season flow conditions. However, the reported skill in predicting the TN loadings account for both error from the LOADEST model as well as the error from the low-dimensional models.

Out of 18 stations, totally nine stations (#2–4, 7–10 and 14–15) exhibited statistically significant skill in predicting the observed winter nutrient loadings under both low-dimensional models based on two different validation methods. Given that these stations exhibit skill under two different validation methods (LCV and SSV), the reported skill is also significant over the entire validation period (1957–2006). Findings from the study could be summarized as the following “controls” that influence the skill in predicting seasonal TN loadings: stations that have very high  $R^2_{(LOADEST)}$  ( $>0.8$ ) in predicting the observed WQN loadings during the winter (Table 2) exhibit significant skill in loadings. Incorporating antecedent flow conditions (December flow) as an additional predictor did not increase the explained variance in these stations, but substantially reduced the RMSE in the predicted loadings. Understanding the source of climatic variability that control the TN variability revealed that Nino3.4, an index denoting ENSO conditions over the tropical Pacific, accounted for 36 % of the observed spatial variability in the TN loadings over the SEUS. Given that using climate forecasts has been very beneficial in improving reservoir management over seasonal time scale

### Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(Sankarasubramanian et al., 2009), we argue the need to develop nutrient loadings forecasts conditioned on climate forecasts. Our future work will utilize these seasonal nutrient forecasts in developing adaptive water management plans over the SEUS.

*Acknowledgements.* The authors of this work were partially supported by the US NSF under CAREER grant CBET-0954405. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not reflect the views of the NSF. Authors also wish to thank Rob Runkel of USGS for his support in setting up the LOADEST model.

## References

- Alexander, R. B. and Smith, R. A.: Trends in the nutrient enrichment of U.S. rivers during the late 20th century and their relation to changes in probable stream trophic conditions, *Limnol. Oceanogr.*, 51, 639–654, 2006.
- Alexander, R. B., Slack, J. R., Ludtke, A. S., Fitzgerald, K. K., and Schertz, T. L.: Data from selected US Geological Survey national stream water quality monitoring networks, *Water Resour. Res.*, 34, 2401–2405, 1998.
- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Arhonditsis, G. B., Winder, M., Brett, M. T., and Schindler, D. E.: Patterns and mechanisms of phytoplankton variability in Lake Washington (USA), *Water Res.*, 38, 4013–4027, 2004.
- Borsuk, M. E., Stow, C. A., and Reckhow, K. H.: Confounding effect of flow on estuarine response to nitrogen loading, *J. Environ. Eng.*, 130, 605–614, 2004.
- Chen, C. F., Ma, H. W., and Reckhow, K. H.: Assessment of water quality management with a systematic qualitative uncertainty analysis, *Sci. Total Environ.*, 374, 13–25, 2007.
- Childers, D. L., Boyer, J. N., Davis, S. E., Madden, C. J., Rudnick, D. T., and Sklar, F. H.: Relating precipitation and water management to nutrient concentrations in the oligotrophic “upside-down” estuaries of the Florida Everglades, *Limnol. Oceanogr.*, 51, 602–616, 2006.
- Cho, H. J. and Poirrier, M. A.: Response of submersed aquatic vegetation (SAV) in Lake Pontchartrain, Louisiana to the 1997–2001 El Nino Southern Oscillation shifts, *Estuaries*, 28, 215–225, 2005.
- Cohn, T. A.: Estimating contaminant loads in rivers: An application of adjusted

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

maximum likelihood to type 1 censored data, *Water Resour. Res.*, 41, W07003, doi:10.1029/2004WR003833, 2005.

Cohn, T. A., Caulder, D. L., Gilroy, E. J., Zynjuk, L. D., and Summers, R. M.: The Validity of a Simple Statistical-Model for Estimating Fluvial Constituent Loads – an Empirical-Study Involving Nutrient Loads Entering Chesapeake Bay, *Water Resour. Res.*, 28, 2353–2363, 1992.

Dettinger, M. D. and Diaz, H. F.: Global characteristics of stream flow seasonality and variability, *J. Hydrometeorol.*, 1, 289–310, 2000.

Devineni, N. and Sankarasubramanian, A.: Improved categorical winter precipitation forecasts through multimodel combinations of coupled GCMs, *Geophys. Res. Lett.*, 37, L24704, doi:10.1029/2010GL044989, 2010.

Devineni, N., Sankarasubramanian, A., and Ghosh, S.: Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, 44, W09404, doi:10.1029/2006WR005855, 2008.

Douglas, E. M., Vogel, R. M., and Kroll, C. N.: Trends in Flood and Low Flows in the United States, *J. Hydrol.*, 240, 90–105, 2000.

EPA Total Maximum Daily Load Report: available at: <http://www.epa.gov/owow/tmdl/>, 2006.

Hartigan, J. A. and Wong M. A.: Algorithm AS 136: A k-means clustering algorithm, *Applied Statistics*, 28.1, 100–108, 1979.

Kaplan, A., Cane, M., Kushnir, Y., Clement, A., Blumenthal, M., and Rajagopalan, B.: Analyses of global sea surface temperature 1856–1991, *J. Geophys. Res.*, 103, 18567–18589, 1998.

Landman, W. A. and Goddard, L.: Model Output Statistics Applied to Multi-Model Ensemble Forecasts for Southern Africa, *Proceedings of the Seventh International Conference on Southern Hemisphere Meteorology and Oceanography*, 249–250, 2003.

Leung, L. R., Hamlet, A. F., Lettenmaier, D. P., and Kumar, A.: Simulations of the ENSO Hydroclimate Signals in the Pacific Northwest Columbia River Basin, *B. Am. Meteorol. Soc.*, 80, 2313–2329, 1999.

Li, S. and Goddard, L.: Retrospective Forecasts with the ECHAM4.5 AGCM, *IRI Technical Report*, 05–02, 2005.

Lin, J., Xie, L., Pietrafesa, L. J., Ramus, J. S., and Paerl, H. W.: Water quality gradients across Albemarle-Pamlico estuarine system: Seasonal variations and model applications, *J. Coastal Res.*, 23, 213–229, 2007.

Maybeck, M.: Carbon, Nitrogen, and Phosphorus Transport by World Rivers, *Am. J. Sci.*, 282,

---

**Interannual  
hydroclimatic  
variability**

J. Oh and  
A. Sankarasubramanian

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



401–450, 1982.

Mueller, D. K. and Spahr, N. E.: Nutrients in Streams and Rivers across the nation – 1992–2001, USGS Scientific Investigations Report, 2006.

Mueller, D. K., Ruddy, B. C., and Battaglin, W. A.: Logistic model of nitrate in streams of the upper-midwestern United States, *J. Environ. Qual.*, 26, 1223–1230, 1997.

National Research Council: Assessing the TMDL approach to water quality management, 2001.

National Research Council: Florida Bay Research Programs and Their Relation to the Comprehensive Everglades Restoration Plan, 2002.

Paerl, H. W., Valdes, L. M., Peierls, B. L., Adolf, J. E., and Harding, L. W.: Anthropogenic and climatic influences on the eutrophication of large estuarine ecosystems, *Limnol. Oceanogr.*, 51, 448–462, 2006.

Ropelewski, C. F. and Halpert, M. S.: Global and Regional Scale Precipitation Patterns Associated with the El-Nino Southern Oscillation, *Mon. Weather Rev.*, 115, 1606–1626, 1987.

Runkel, R. L., Crawford, C. G., and Cohn, T. A.: Load Estimator (LOADEST): A FORTRAN Program for Estimating Constituent Loads in Streams and Rivers, US Geological Survey Report, 2004.

Sankarasubramanian, A., Lall, U., Souza Filho, F. D., and Sharma, A.: Improved Water Allocation utilizing Probabilistic Climate Forecasts: Short Term Water Contracts in a Risk Management Framework, *Water Resour. Res.*, 45, W11409, doi:10.1029/2009WR007821, 2009.

Sankarasubramanian, A., Sharma, A., Lall, U., and Espinueva, S.: Role of retrospective forecasts of GCMs forced with persisted SST anomalies in operational streamflow forecasts development, *J. Hydrometeorol.*, 9, 212–227, 2008.

Slack, J. R., Lumb, A., and Landwehr, J. M.: Hydro-Climatic Data Network (HCDN) Streamflow Data Set, 1874–1988, US Geological Survey Report, USGS open-file report: 92–129, 1993.

Smith, R. A., Alexander, R. B., and Schwarz, G. E.: Natural background concentrations of nutrients in streams and rivers of the conterminous United States, *Environ. Sci. Technol.*, 37, 3039–3047, 2003.

Smith, R. A., Schwarz, G. E., and Alexander, R. B.: Regional interpretation of water-quality monitoring data, *Water Resour. Res.*, 33, 2781–2798, 1998.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S.: Ensemble square-root filters, *Mon. Weather Rev.*, 131, 1485–1490, 2003.

Towler, E., Rajagopalan, B., and Summers, R. S.: Using Parametric and Nonparametric Meth-

- ods to Model Total Organic Carbon, Alkalinity, and pH after Conventional Surface Water Treatment, *Environ. Eng. Sci.*, 26, 1299–1307, 2009.
- Vecchia, A. V.: Relation Between Climate Variability and Stream Water Quality in the Continental United States, *Hydrological Science and Technology*, 19, 77–98, 2003.
- 5 Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., and Tilman, G. D.: Human alteration of the global nitrogen cycle: Sources and consequences, *Ecol. Appl.*, 7, 737–750, 1997.
- Wilks, D. S.: *Statistical Methods in the Atmospheric Science*, Academic Press, 1–630, 1995.

---

**Interannual  
hydroclimatic  
variability**

J. Oh and  
A. Sankarasubramanian

---

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 1.** Baseline information for 18 selected stations showing the number of years of observed daily records of TN available in the WQN database. Values in the parentheses under number of years column show the total number of daily observations available for each station.

Station Index	Station Number	Station Name	Drainage Area (km <sup>2</sup> )	Number of years (# of daily Obs.)
1	2047000	Nottoway river near Sebrell, VA	3732.17	17 (95)
2	2083500	Tar river at Tarboro, NC	5653.94	22 (152)
3	2126000	Rocky river near Norwood, NC	3553.46	14 (65)
4	2176500	Coosawhatchie river near Hampton, SC	525.77	13 (100)
5	2202500	Ogeechee river near Eden, GA	6863.47	20 (141)
6	2212600	Falling creek near Juliette, GA	187.00	14 (56)
7	2228000	Satilla river at Atkinson, GA	7226.07	20 (123)
8	2231000	St. Marys river near Macclenny, FL	1812.99	14 (108)
9	2321500	Santa Fe river at Worthington Springs, FL	1489.24	21 (82)
10	2324000	Steinhatchee river near Cross city, FL	906.50	19 (92)
11	2327100	Sopchoppy river near Sopchoppy, FL	264.18	22 (125)
12	2329000	Ochlockonee river near Havana, FL	2952.59	22 (133)
13	2358000	Apalachicola river at Chattahoochee, FL	44 547.79	23 (152)
14	2366500	Choctawhatchee river near Bruce, FL	11 354.51	21 (119)
15	2368000	Yellow river at Milligan, FL	1616.15	21 (123)
16	2375500	Escambia river near Century, FL	9885.98	22 (145)
17	2479155	Cypress creek near Janice, MS	136.23	16 (54)
18	2489500	Pearl river near Bogalusa, LA	17 023.99	12 (57)



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

**Table 2.** Performance of LOADEST model in predicting the observed TN loadings from the WQN database. Models with linear time components (Model No.: 3, 5, 7–9) are not considered. The skill in predicting JFM WQN loadings are separately given in the last two columns.

Station Index	Station No.	$R^2$ (Daily)	AIC (Daily)	Model No.	Coefficients of selected LOADEST model					Skill for JFM	
					a0	a1	a2	a3	a4	$R^2_{(\text{LOADEST})}$	RMSE $_{(\text{LOADEST})}$ [kg day $^{-1}$ km $^{-2}$ ]
1	2047000	0.948	0.892	4	6.768	1.114	-0.283	-0.069		0.897	0.432
2	2083500	0.966	-0.131	4	8.122	0.980	0.108	-0.018		0.948	0.403
3	2126000	0.966	0.496	4	8.863	1.066	-0.195	0.090		0.929	0.896
4	2176500	0.956	0.905	6	4.446	1.013	0.026	0.238	-0.036	0.867	0.702
5	2202500	0.916	0.837	4	7.721	1.069	-0.084	-0.317		0.014	0.756
6	2212600	0.853	2.094	1	2.647	1.095				0.004	0.855
7	2228000	0.968	0.518	6	7.521	1.005	-0.025	-0.083	0.103	0.887	0.701
8	2231000	0.963	0.250	6	6.428	1.088	-0.075	-0.027	0.187	0.925	0.242
9	2321500	0.986	-0.219	6	5.690	1.086	-0.037	-0.078	0.059	0.977	0.211
10	2324000	0.979	0.279	6	5.549	1.241	-0.069	-0.096	0.071	0.826	0.608
11	2327100	0.979	0.516	6	4.351	1.139	-0.043	0.187	0.007	0.959	0.344
12	2329000	0.923	0.585	1	7.341	0.846				0.839	0.415
13	2358000	0.902	0.193	4	10.563	0.981	0.074	0.165		0.728	0.625
14	2366500	0.835	0.423	4	9.077	0.931	-0.145	-0.042		0.884	0.375
15	2368000	0.834	1.085	6	7.238	1.123	-0.131	-0.004	0.176	0.835	0.595
16	2375500	0.873	0.758	4	8.868	1.039	0.147	0.032		0.424	0.798
17	2479155	0.912	1.233	4	4.555	1.188	0.206	0.328		0.999	1.531
18	2489500	0.899	0.853	1	10.193	1.047				0.075	6.544

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

[⏪](#) [⏩](#)

[◀](#) [▶](#)

[Back](#) [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

**Table 3.** Rank correlation between observed winter streamflow, TN loadings with the first principal component of the winter precipitation forecasts for the 18 selected stations. Locations of grid points indicated in the Table are shown in Fig. 1.

Station Index	Station Number	Grid points	% Variance Explained	Correlation between PC1 and $L_t$ (TN)	
				$Q$	$L_t$ (TN)
1	2047000	37–40, 45–48, 53–56	73.6	0.453	0.453
2	2083500	29–32, 37–40, 45–48	84.9	0.479	0.476
3	2126000	36–39, 44–47	79.8	0.456	0.445
4	2176500	27–31, 35–39, 45–47	86.2	0.710	0.611
5	2202500	19–23, 27–31, 35–39, 45–47	85.6	0.598	0.480
6	2212600	37–40, 45–48	86.5	0.418	0.405
7	2228000	19–23, 27–31, 35–39	91.3	0.596	0.601
8	2231000	18–22, 26–30, 34–38	90.8	0.556	0.561
9	2321500	19–22, 27–30, 35–38	90.8	0.561	0.558
10	2324000	18–22, 26–30, 34–38	90.8	0.575	0.572
11	2327100	10–14, 18–22, 26–30	96.0	0.526	0.561
12	2329000	18–22, 26–30, 34–38	90.8	0.469	0.490
13	2358000	26–30, 34–38, 45–46	86.8	0.554	0.555
14	2366500	18–21, 26–29, 34–37	88.0	0.491	0.500
15	2368000	18–21, 26–29, 34–37	90.8	0.499	0.531
16	2375500	26–31, 34–39, 45–47	86.9	0.404	0.404
17	2479155	18–21, 26–29, 34–37	90.1	0.453	0.432
18	2489500	26–27, 34–35	94.7	0.287	0.284

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

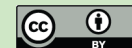
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Table 4.** Grouping of 18 selected stations based on K-Means clustering.

Group	Stations	Averaged TN loading (kg day <sup>-1</sup> )
1	13, 18	59 608.83
2	2, 3, 7, 14, 16	11 394.76
3	1, 5, 12	4043.69
4	4, 6, 8, 9, 10, 11, 15, 17	895.95

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

[Back](#) [Close](#)

[Full Screen / Esc](#)

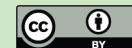
[Printer-friendly Version](#)

[Interactive Discussion](#)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

**Table 5.** Skill, expressed as RMSE (based on Eq. 3), in predicting winter TN loadings using climate forecasts. Table also gives the number of principal components considered and the percentage variance explained by them for the total grid points selected (given in Table 3) for each station.

Station ID	# of PCs	% variance explained	PCR		CCA	
			RMSE (kg day <sup>-1</sup> km <sup>-2</sup> )		RMSE (kg day <sup>-1</sup> km <sup>-2</sup> )	
			LCV	SSV	LCV	SSV
1	1	73.6	0.585	0.924	0.583	0.884
2	2	96.6	0.652	0.672	0.679	1.246
3	3	99.2	1.955	1.809	1.968	3.461
4	1	86.2	0.938	2.190	0.891	1.654
5	2	97.4	0.801	0.833	0.795	0.968
6	1	86.5	1.038	1.284	1.039	1.472
7	1	91.3	1.034	1.547	0.938	1.285
8	6	99.7	0.678	0.652	0.524	2.319
9	1	90.8	0.967	1.458	0.674	3.619
10	8	99.9	1.235	1.830	1.005	6.453
11	6	99.8	0.865	1.518	0.757	3.691
12	1	90.8	0.906	0.923	0.757	2.014
13	1	86.8	0.790	0.818	0.785	1.582
14	2	98.1	0.510	0.555	0.523	1.013
15	2	98.5	0.690	0.817	0.682	1.199
16	2	96.9	0.965	1.106	0.960	1.418
17	1	90.1	1.656	2.051	1.645	2.132
18	1	94.7	6.703	6.722	6.694	7.673



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Table 6.** RMSE (Eq. 3) of forecasted TN loadings based on PCR and CCA models that consider ECHAM4.5 precipitation forecasts and December streamflow as predictors under SSV.

Station Index	Station Number	RMSE (kg area <sup>-1</sup> )	
		PCR	CCA
1	2047000	0.653	0.866
2	2083500	0.671	1.233
3	2126000	1.748	3.420
4	2176500	2.136	1.487
5	2202500	0.845	0.930
6	2212600	1.247	1.438
7	2228000	1.312	1.283
8	2231000	0.631	2.359
9	2321500	1.294	3.190
10	2324000	1.266	6.061
11	2327100	1.510	3.417
12	2329000	0.846	1.419
13	2358000	0.750	1.370
14	2366500	0.508	1.008
15	2368000	0.810	1.193
16	2375500	1.058	1.411
17	2479155	2.022	2.038
18	2489500	6.713	7.032

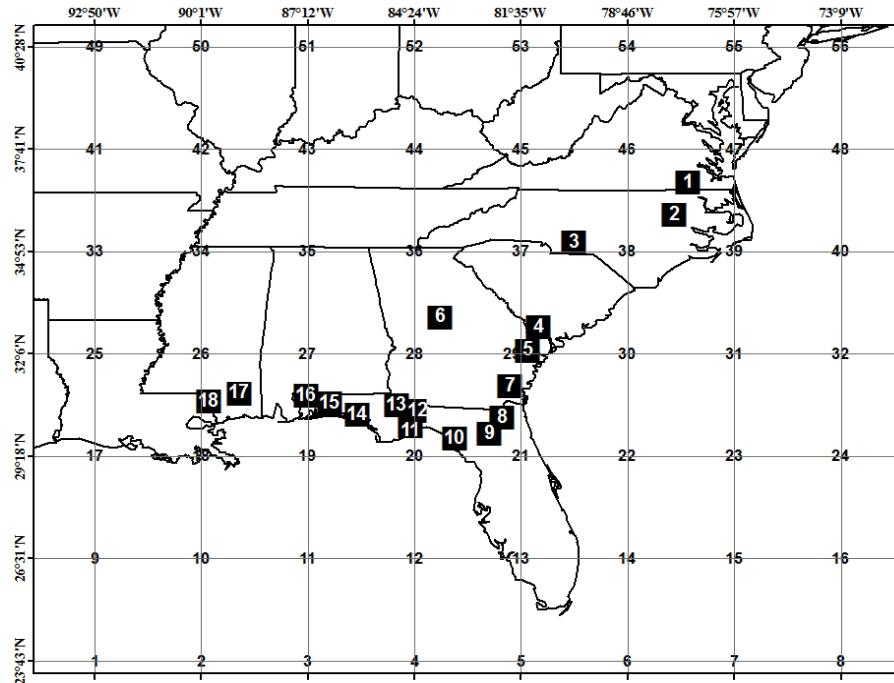
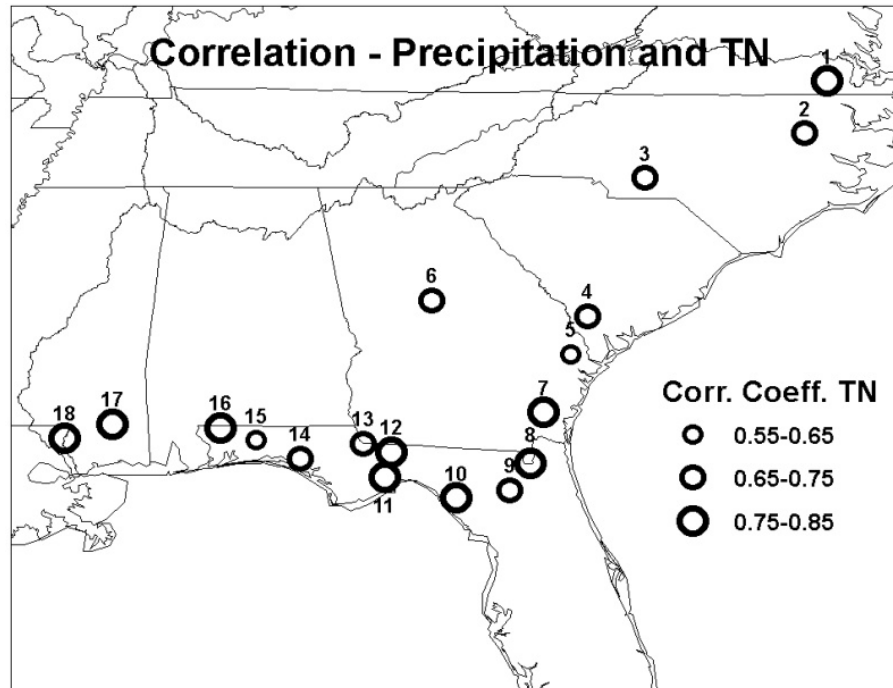
Interannual  
hydroclimatic  
variabilityJ. Oh and  
A. Sankarasubramanian

Fig. 1. Location of the 18 HCDN stations along with the considered grid points over the SEUS.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

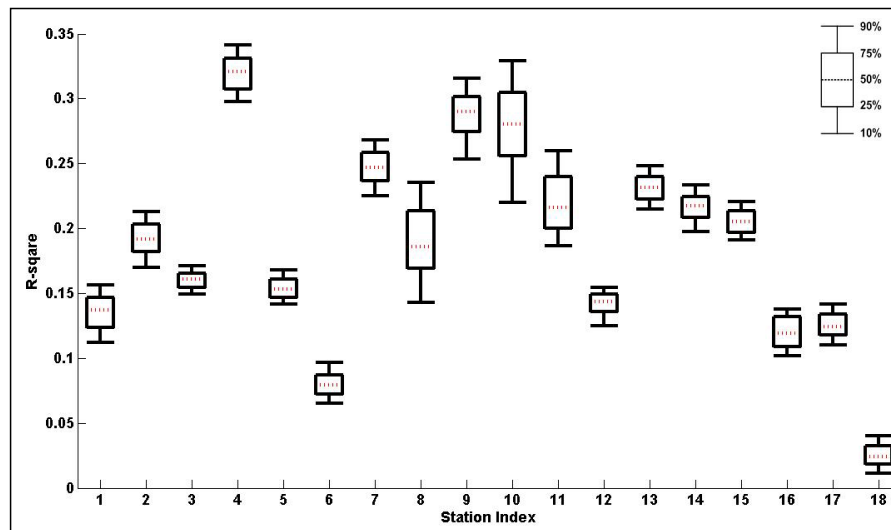
Interannual  
hydroclimatic  
variabilityJ. Oh and  
A. Sankarasubramanian

**Fig. 2.** Rank correlation between the simulated TN loadings from the LOADEST model and observed precipitation over the selected 18 stations.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian



**Fig. 3.** Box-plot of  $R^2$  (based on Eq. 2) of PCR model predicted TN loadings obtained using PC's of forecasted precipitation under LCV.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[⏪](#)
[⏩](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)



## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

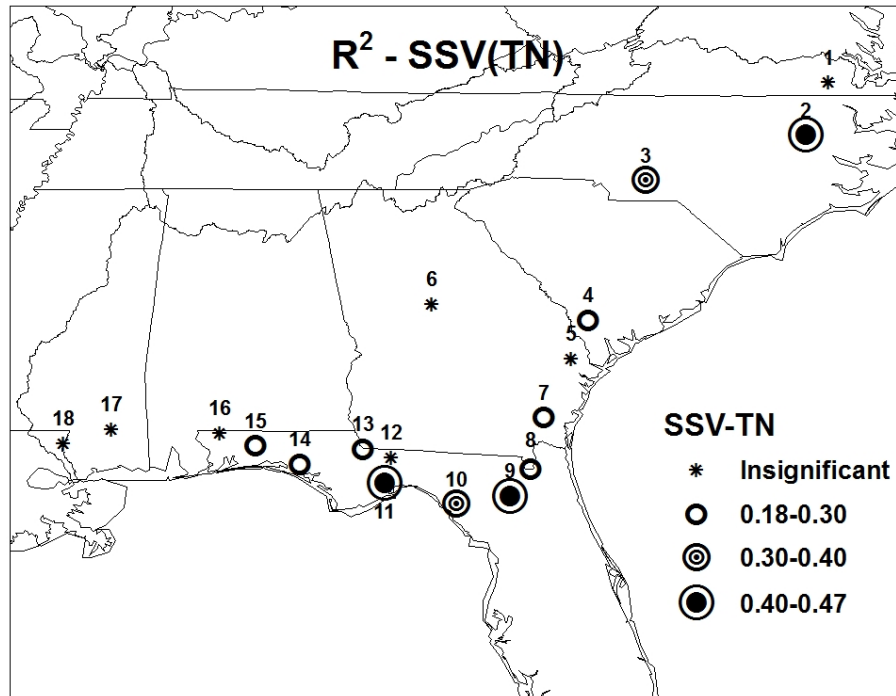
Back

Close

Full Screen / Esc

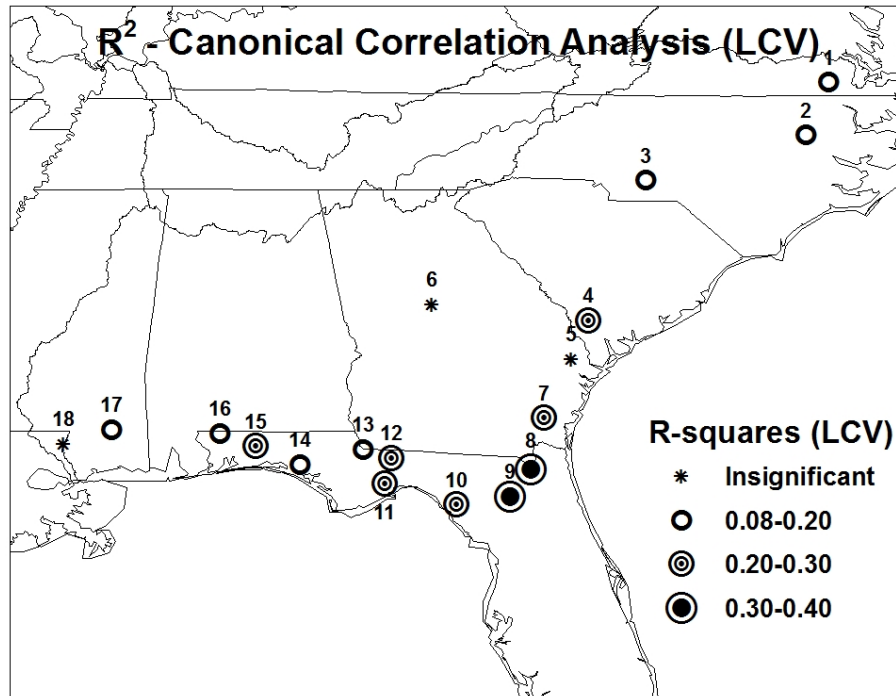
Printer-friendly Version

Interactive Discussion



**Fig. 4.** Modified  $R^2$  (based on Eqs. 2 and 3) of PCR model predicted TN loadings obtained using PC's of forecasted precipitation under SSV.

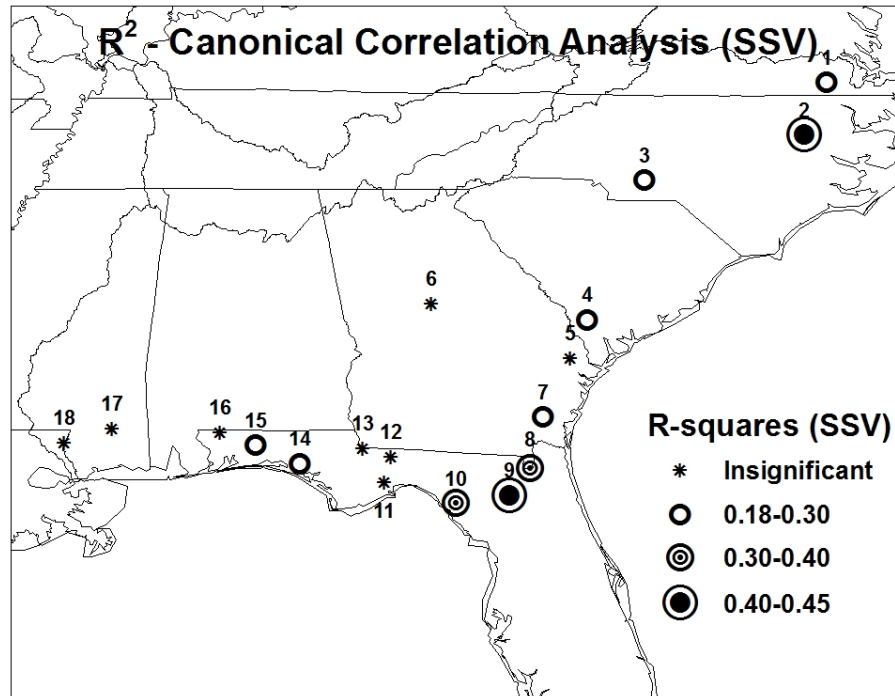
## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian

**Fig. 5.** Modified  $R^2$  (based on Eq. 2) of CCA model predicted TN loadings obtained using PC's of forecasted precipitation under LCV.

## Interannual hydroclimatic variability

J. Oh and  
A. Sankarasubramanian



**Fig. 6.** Modified  $R^2$  (based on Eq. 2) of CCA model predicted TN loadings obtained using PC's of forecasted precipitation under SSV.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

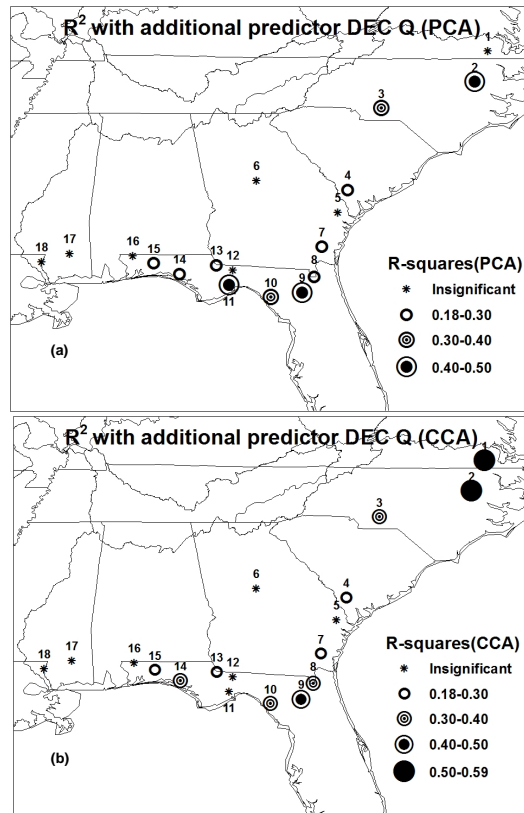
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Interannual hydroclimatic variability

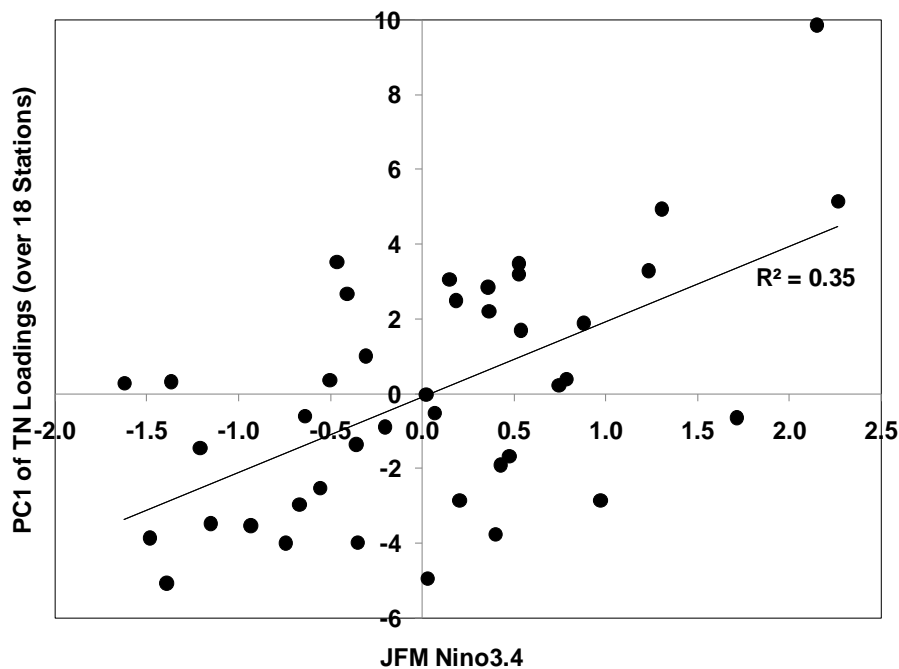
J. Oh and  
A. Sankarasubramanian



**Fig. 7.**  $R^2$  (based on Eq. 2) of PCR (a) and CCA (b) model predicted TN loadings obtained using both PC's of forecasted precipitation and December streamflow under SSV.

[Title Page](#)  
[Abstract](#)   [Introduction](#)  
[Conclusions](#)   [References](#)  
[Tables](#)   [Figures](#)  
⏪   ⏩  
◀   ▶  
[Back](#)   [Close](#)  
[Full Screen / Esc](#)  
[Printer-friendly Version](#)  
[Interactive Discussion](#)



**Interannual hydroclimatic variability**J. Oh and  
A. Sankarasubramanian

**Fig. 8.** Relationship between the first principal component of TN loadings over SEUS and ENSO conditions, which is indicated by Nino3.4.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)