WILLIAM & MARY
CHARTERED 1693

W&M ScholarWorks

VIMS Articles

Virginia Institute of Marine Science

4-24-2019

A data-driven modeling approach for simulating algal blooms in the tidal freshwater of James River in response to riverine nutrient loading

Jian Shen
Virginia Institute of Marine Science

Qubin Qin
Virginia Institute of Marine Science

Ya Wang

Mac Sisson
Virginia Institute of Marine Science

Follow this and additional works at: https://scholarworks.wm.edu/vimsarticles

Part of the Environmental Sciences Commons, and the Marine Biology Commons

Recommended Citation

A data-driven modeling approach for simulating algal blooms in the tidal freshwater of James River in response to riverine nutrient loading

Jian Shen[1]*, Qubin Qin[1], Ya Wang[2], and Mac Sisson[1]

[1]Virginia Institute of Marine Science, College of William & Mary,

Gloucester Point, VA 23062, USA

[2]The Third Institution of Oceanography, Ministry of Natural Resources, Xiamen, China

*Corresponding Author:

Jian Shen

Virginia Institute of Marine Science

College of William & Mary

Gloucester Point, VA 23062

Email: shen@vims.edu

Phone: (804) 684-7359

Highlights:

1. A successful application of the Support vector machine (SVM) for algal bloom simulation provides new approach for predicting harmful algal bloom.

2. Combining Empirical Orthogonal Function and SVM enables simulations of algal blooms for the entire tidal freshwater region in one model.

3. Applying variable transformation is crucial for improving model predictive skill.

4. The data-driven model is capable of assessing algal blooms responding to changes of nutrients if it is trained appropriately.

1

2 **Abstract**

3   Algal blooms often occur in the tidal freshwater (TF) of the James River estuary, a tributary

4 of the Chesapeake Bay. The timing of algal blooms correlates highly to a summer low-flow

5 period when residence time is long and nutrients are available. Because of complex interactions

6 between physical transport and algal dynamics, it is challenging to predict interannual variations

7 of bloom correctly using a complex eutrophication model without having a high-resolution

8 model grid to resolve complex geometry and an accurate estimate of nutrient loading to drive the

9 model. In this study, an approach using long-term observational data (from 1990-2013) and the

10 Support vector machine (LS-SVM) for simulating algal blooms was applied. The Empirical

11 Orthogonal Function was used to reduce the data dimension that enables the algal bloom

12 dynamics for the entire TF to be modeled by one model. The model results indicate that the data-

13 driven model is capable of simulating interannual algal blooms with good predictive skills and is

14 capable of forecasting algal blooms responding to the change of nutrient loadings and

15 environmental conditions. This study provides a link between a conceptual model and a dynamic

16 model, and demonstrates that the data-driven model is a good approach for simulating algal

17 blooms in this complex environment of the James River. The method is very efficient and can be

18 applied to other estuaries as well.

19

22

23

1. Introduction

26   The tidal freshwater (TF) region is located in the upstream of an estuary where tidal

27   forcings extend inland but beyond the limits of salinity intrusion. The TF ecosystem is highly

28   influenced by freshwater discharge and the net transport is downstream, even as it experiences

29   tidal fluctuations. The TF is often associated with complex geometry involving a meandering

30   channel with irregular channel depth and cross-section. The interactions of complex geometry

31   and the large fluctuation of freshwater discharge result in strong seasonal variations of dynamic

32   conditions, which can alter pollutant transport and algal growth (Bukaveckas et al., 2017; Shen et

33   al., 2016). Although the TF only accounts for a small portion of an estuary, it interfaces with the

34   drainage basin and is sensitive to any perturbations occurring in the drainage basin. The seasonal

35   and interannual changes of the retention time of pollutants can have a profound impact on the

36   downstream estuary (Paerl, 2009). As it is located in a transition zone between river and estuary,

37   the change of ecosystem in the TF is indicative of the changes in land use applications in the

38   drainage area.

39   Algal blooms, including harmful algal blooms (HABs), often occur in the TF region

40   (Seitzinger, 1991; Paerl et al., 2001; Bukaveckas et al., 2011). Algal community metrics have

41   been used for a long time in the assessment of water quality conditions as indicators of biotic

42   responses to environmental stressors such as eutrophication and acidification (Buchanan et al.,

43   2005; Marshall et al., 2006). The United States Environmental Protection Agency (EPA) and

44   Chesapeake Bay Program have developed specific Chlorophyll a (Chl-a) concentration criteria

45   for the TF regions in the Chesapeake Bay (USEPA, 2010) and use numerical models to assess

46   the impact of algal bloom on water quality (Cerco and Noel, 2004). As the algal growth is highly

47   controlled by the nutrient inputs from the non-point sources, the Chl-a concentration criteria are

48   often used to evaluate the efficiency of nutrient reduction in the drainage basins and estuary

49   restoration.

50   Complex water quality models have been widely used to understand algal blooms in

51   response to the flow and nutrient discharges and to determine nutrient loading reduction

52   (Thomann and Mueller, 1987; Cerco and Noel, 2004; Shen, 2006). There are many applications

53   for using 2D and 3D water quality models to simulate algae blooms (Wu and Xu, 2011; James,

54   2016; Kim et al., 2017; Jiang and Xia, 2017). However, it is always a challenge to calibrate a

55   complex model to properly simulate an algal bloom because both physical transport and

56   biological processes modulating algal biomass dynamics are highly variable under different

57   residence times and biological timescales (Lucas et al., 2009; Qin and Shen, 2017). To simulate

58   hydrodynamics well, a spatially fine resolution of the model grid is often required due to the

59   complex geometry in TF portions of estuaries (e.g., Shen et al., 2016). Moreover, the accuracy of

60   model simulation depends highly on model kinetic processes, while large uncertainties are

61   always associated with the selection of model kinetic parameter values due to high correlations

62   among these parameters and non-uniqueness of the parameter values (van Straten, 1983; Shen,

63   2006; Jiang et al., 2018). On the other hand, the accuracy of eutrophication models depends

64   highly on the nutrient loadings from both point and nonpoint sources, which are often simulated

65   by the watershed model (Shen et al., 2005; Riverson et al., 2013). As large uncertainties are

66   associated with the watershed model as well, the linked watershed-receiving water quality

67   modeling approach is often associated with a high level of uncertainty (Wu et al., 2006), which

68   increases the difficulty for accurate simulation of an algal bloom.

69   In addition to the use of complex numerical models, statistical and empirical modeling

70   approaches based on observational data have been applied to simulate algae, dissolved oxygen

71 concentration (DO), and other pollutants in aquatic systems (Recknagel et al., 1997; Shen et al.,

72 2008; Zhang et al., 2009; Rounds, 2002; Shen and Zhao, 2010; Xie et al., 2012; Kong et al.,

73 2017). The empirical approaches have the advantage of providing a relationship between

74 independent and dependent variables and estimating dependent variables according to the

75 changes of a set of independent variables. Empirically based algorithms have been playing an

76 increasingly important role in HAB modeling, providing an important link between conceptual

77 and dynamical modeling approaches (McGillicuddy, 2010; Blauw et al., 2010; Anderson et al.,

78 2010; Wang and Tang, 2010; Kong et al., 2017). As more and more observational data become

79 available, many effective methods can be used to build empirical models, such as multiple

80 variable regression and neural network. Recently, more sophisticated methods based on the field

81 of machine learning have also been applied for water quality modeling in multiple ways (e.g.

82 Recknagel, 2001; Muttil and Chau, 2006; Volf et al., 2011; Liang et al., 2015; Kong et al., 2017).

83 Lui et al. (2007) used a vector autoregressive model to simulate algal blooms. Ribeiro and Torgo

84 (2008) compared different methods for algal simulations, which showed that the support vector

85 machine (SVM) has a good modeling skill. Xie et al. (2012) demonstrated the effective use of

86 SVM for simulating freshwater algal bloom in a reservoir. Crisci et al. (2012) reviewed

87 supervised machine learning used for ecological data. Recently, Park et al. (2015) developed an

88 early warning protocol of algae using SVM in freshwater and reservoirs. Moe et al. (2016)

89 applied Bayesian network technology to study cyanobacteria bloom in lakes. Kong et al. (2017)

90 applied SVM to evaluate eutrophication statuses in coastal seas successfully. These studies

91 indicate that the machine-learning approach is an effective tool for algal bloom simulations.

92 However, most temporal variations of algal bloom simulations in the literature are limited to the

93 simulation of algae at individual observation stations. When applying a statistical model or a

94   machine-learning model to an entire region of an estuary with multiple observation stations,

95   different models need to be created for the prediction at different stations, which could cause

96   inconsistency in response to change of environmental conditions at different stations. Although

97   the machine-learning has been applied to predict algal blooms, it is not well-studied if the model

98   also is capable of responding to the change of watershed condition due to change of nutrient

99   loadings as many model simulations are trained using input variables observed at the same

100   station to be predicted (Xie et al., 2012; Park et al., 2015).

101   In this study, we investigated a data-driven modeling approach to simulate algal blooms in

102   the James River. The James River estuary is a western tributary of the Chesapeake Bay. In the

103   TF region of the James River, a bloom of cyanobacteria, a freshwater HAB species, often occurs

104   in summer, and microcystin is often observed when the Chl-a concentration is high (Bukaveckas

105   et al., 2018). The Chl-a distribution is strongly influenced by hydrodynamic conditions when the

106   geometry changes from a narrow stream to a wider cross-section because of the limited mobility

107   of algae. Bukaveckas et al. (2011) found that the location of the maximum of Chl-a

108   concentration in the TF James River is determined in part by the natural geomorphic features of

109   the channel. The transition from a riverine-type (narrow and deep) cross-sectional morphology to

110   a broad channel with shallow lateral areas provides favorable light conditions for the algal

111   growth. The residence time increases during the low-flow period, which coincides with the

112   summer period of algal bloom (Shen et al., 2016; Qin and Shen, 2017). Consequently, the algal

113   bloom occurs frequently during summer in this region. This is an ideal area for investigating the

114   data-driven modeling approach.

115   The purpose of this study is to apply the learning machine technique to the TF portion of the

116   James River to simulate the algal blooms using long-term monitoring data at multiple stations to

117   provide new capability for harmful algal bloom prediction. The difference of the current

118   modeling approach is that we applied Empirical Orthogonal Function (EOF) to separate Chl-a

119   concentrations at multiple stations in the TF region to both spatial and temporal components. We

120   simulated the principle temporal vectors based on the variation of environmental variables.

121   Therefore, the entire TF region can be simulated by one model. We also used nutrient loadings

122   and flow data at Full-Line as dependent variables in order to ensure the model to response to

123   change of environmental variables. In addition, we conducted a transformation of variables and

124   introduced combined new variables to improve the model prediction skill and ensure that the

125   model would respond to the changes of nutrients discharged from the watershed. As a result, the

126   model shows an improved predictive skill. The model sensitivity tests indicate that the model is

127   suitable for investigating the responses in algal growth to changes in nutrient loadings when the

128   model is trained appropriately. The approach is efficient in terms of model effort and can be

129   applied to other estuaries.

130   **2. Methods**

131

132   *2.1 Data collection*

133   The tidal freshwater (TF) segment of the James River (salinity < 0.5 ppt) extends from

134   the Fall Line (at Richmond, VA) to the downstream with a total length of approximately 115 km.

135   The drainage area is 26,165 km^2, which is predominantly forested (about 71%) (Bukaveckas et

136   al., 2017). The Virginia Department of Environmental Quality conducts monthly monitoring

137   surveys in the James River. Observations of Chl-a concentration near the surface, together with

138   temperature, DO, total suspended solid (TSS), total nitrogen (TN) and total phosphorus (TP),

139   dissolved inorganic nitrogen (DIN), and phosphate (PO$_4$), are available from 1990-2013. The

140   station locations are shown in Fig. 1. The stations located in the TF region include TF5-2, TF5-

141   2A, TF5-3, TF5-4, TF5-5, TF5-5A, and TF5-6 from upstream to downstream, and Station TF5-2

142   is located at the Fall Line near Richmond. The distribution of the Chl-a concentration along the

143   James River mainstem is shown in Fig. 2. High Chl-a concentration is observed in the region

144   with relatively wide cross-sections and the concentration decreases in both the upstream and

145   downstream directions. The daily freshwater discharge is available at USGS freshwater gauge

146   station (US02037500) near Richmond. The water quality station near the Fall Line is mainly

147   controlled by the freshwater and nonpoint source discharges of nutrients and TSS. There is a

148   statistically significant relationship between Chl-a concentration and discharge (Bukaveckas et

149   al., 2017). Loadings of TN, TP, and TSS were estimated by multiplying daily flow and

150   interpolated daily TN, TP, and TSS concentrations at the Fall Line. Hourly solar radiation was

151   obtained from the Richmond International Airport. In this study, the data set of Chl-a

152   concentration for all stations excluding TF5-2 from 1990 to 2013 was used, leading to the data

153   size matrix of 274×6 in total. We excluded Station TF5-2 because it is located at the Fall Line

154   and the measurements were used to estimate loadings for algae, TN, TP, and TSS for the

155   nonpoint sources.

156   *2.2 Variable transformation*
157

158   Observations of Chl-a data and environmental variables used for the model were

159   transferred first to improve the accuracy of the model prediction. The logarithmic transformation

160   was applied to Chl-a data. Initial analysis showed that transforming discharge Q to $Q^{1/3}$ has a

161   high correlation between flow and Chl-a concentration. A 5-day backward moving average flow

162   prior to the date of observation was applied to the flow to account for the accumulating effect as

163   the USGS flow station is located upstream of the study area, which provides better correlation

164   between flow and log (Chl-a). An example of correlations of independent variables and log (Chl-

165   a) at Station TF5-5 is listed in Table 1.

166   Both the observations of TN and TP at Fall Line were linearly interpolated and multiplied by

167   flow to obtain daily loadings. Both TN and TP loadings are high in the spring and low in the

168   summer, which are negatively correlated to high concentrations of Chl-a (Table 1). If we directly

169   use it for the model, the Chl-a concentration will not respond correctly to the nutrient level. On

170   the other hand, the summer high algal bloom depends not only on the total spring runoff of TN

171   and TP, but also on the summer bottom fluxes of DIN and DIP from the bottom sediment due to

172   the later winter and spring (February-May) deposition of organics and subsequent

173   remineralization, which will be on the order of 100 days (Thomann and Mueller, 1987). To better

174   reflect the real signal of TN and TP in summer when an algal bloom occurs, we first backward-

175   average the loading (moving average) for a 120-day period prior to the date of Chl-a observation

176   to obtain the accumulative effect of spring runoff. The 120-day moving average was determined

177   based on the spring runoff period, time required for remineralization (Park et al., 1995), and the

178   model performance. Both TN and TP daily loading were transferred to the new variables as

179   follows:

180      $$TN_{new} = \frac{TN}{H_{TN}+TN} \theta^{T-20}$$

181      $$TP_{new} = \frac{TP}{H_{TP}+TP} \theta^{T-20} \hspace{3cm} (1)$$

182   The approach is similar to the Monod-type nutrient limiting function applied in the water quality

183   model (Thomann and Mueller, 1987; Park et al., 1995). By doing this transformation, the signal

184   for high nutrients during spring was reduced. The correction of temperature, θ^{T-20}, was used to

185   amplify the release of recycled nutrients in summer ($\theta = 1.03$). Note that it is not a good

186   approach to use temperature directly as an independent variable as it has the same annual cycle

10

187   as algal blooms, which will be discussed more in the Discussion Section. We used the 75th

188   percentile of loading values as the half-saturation coefficients for both TN and TP based on

189   model test runs. With these changes, the respective correlations of Chl-a and TN and TP were

190   improved (Table 1). In addition, as Chl-a concentration values were obtained on different dates

191   for each month, a 15-day average of light was used for the model. Although the moving average

192   of light did not show an improvement of the correlation, it did improve the model simulations. A

193   detailed description of environmental variables used for model input and transforming are listed

194   in Table 2.

195   *2.3 Empirical orthogonal function analysis*

196   There are many observation stations located in the estuary. A traditional approach for

197   developing an empirical model is to develop a model for each observation station, which is not

198   efficient and may not be consistent with changes of environmental conditions for each station.

199   We applied the EOF method to reduce the data dimension and to be able to simulate the entire

200   system based on the principle components of Chl-a data. The EOF method has often been

201   applied to analyze complex data sets to understand the spatial and temporal patterns and

202   distributions of state variables (Bergamino et al., 2007; Wang and Tang, 2010; Du et al., 2018).

203   The purpose of using the EOF method in this study is to separate spatial variations and temporal

204   variations of Chl-a based on principal components. Therefore, we can focus on the prediction of

205   a few temporal vectors for all stations rather than develop a model for each station. The EOF

206   analysis is based on the singular value decomposition method, which decomposes the data matrix

207   *F* (log(Chl-a)) into the form:

208   *F= SVD*   (2)

209   where *S* is the temporal vector of the matrix (274×6), *D* is an orthonormal matrix (6×6) of spatial

210   vectors, and V is a diagonal matrix (6×6) storing the eigenvalues. Once we obtain the temporal

211   variations for the principal components, the spatial variations at each station can be obtained

212   based on Eq. (2).

213   *2.4 Support vector machine LS-SVM*

214   We used support vector machines (SVM) (Vapnik, 1999) for this study. SVM is a powerful

215   learning machine for classification, and it can be applied to time-varying simulations. SVM has

216   been first introduced within the context of statistical learning theory and structural risk

217   minimization. The idea of SVM is to map the training data nonlinearly into a higher-dimensional

218   feature space and then to construct a separating hyperplane with maximum margin there. LS-

219   SVM, proposed by Suykens and Vandewalle (1999) and Suykens et al. (2002) is an extended

220   version of the standard SVM. Different from the standard SVM, LS-SVM takes a squared loss

221   function for the error variable and uses equality constraints instead of inequality constraints.

222   LS-SVM has been widely applied in fields of pattern recognition, classification and

223   function estimation (Zhang et al., 2011). Recently, it was also combined with a water quality

224   model to estimate model kinetic parameters (Liang et al., 2015; Park et al., 2015; Kong et al.,

225   2017). Park et al. (2015) applied it successfully for predicting the eutrophication status in a

226   coastal water.

227   The method is to estimate a function $f: R^N \rightarrow \{\pm 1\}$ using training data of N-dimension

228   patterns x_i and class labels y_i, $(x_1, y_1), \ldots (x_l, y_l) \in R^N \times \{\pm 1\}$. Data can be mapped into the

229   higher dimensional space via a nonlinear function φ (x) and:

230    $y(x) = w^T \varphi(x) + b$ (3)

231   where $w \in R^N$ and $b \in R$ are regression parameters to be determined. The following

232   optimization is formed:

233    $\min J(w, e) = \frac{1}{2} w^t w + \frac{1}{2} \gamma \sum_{k=1}^{N} e_k^2$ (4)

234   Subject to:

235    $y(x) = w^T \varphi(x) + b + e_k, \qquad k = 1, 2, \dots, N$

236   The problem can be solved using non-linear optimization (Lagrangian method), and the LS-SVM

237   model can be expressed as:

238    $y(x) = \sum_{k=1}^{N} \alpha_k K(x_k, x) + b$ (5)

239   where $\alpha = [a_1 a_2, \dots, a_N]^T$ are the Lagrangian multipliers, and K(x_k, x_l)= $\varphi(x)' \varphi(x)$ is the kernel

240   function. The linear SVM kernel is $K(x_k, x) = (x_k^T x + 1)^d$ and the RBF kernel is $K(x_k, x_l) = $

241    $\exp\{-||x - x_k||_2^2 / \sigma^2\}$, where σ is kernel parameters. Different kernels were tested and the RBF

242   kernel was used for this study, which provides satisfactory performance.

243   We used flow, TSS, TN, TP, and Chl-a loadings at the Fall Line together with light and

244   temperature as independent variables for the model. We first used the LS-SVM learning machine

245   to conduct training for six temporal mode of eigenvectors obtained from Eq. 2 using the same

246   independent variables. Although the 1st eigenvector has the highest contribution, the contribution

247   of this vector to Chl-a concentration at each station is different. Therefore, LS-SVM was trained

248   for each eigenvector. We used data from 1992-2005 for the model training because the algal

249   concentration is much higher during the period from 1990-2002 and it decreases after 2002, the

250   selection of data for training spanning both of these two periods was important. We compared

251   the results using either date set of the first 14-year (1990-2003) or the data set of the last 14-year

13

252 (2000-2013) for model training to that of using 1990-2002 data set for training, the model has the

253 best predictive skill using 1992-2005 data set. Adding more data for the training did not improve

254 model performance much, which may cause over-fitting of the model.

255   Once the model was trained, the data from 1990, 1991, and 2006-2013 were used for

256 verification. After having completed training and verification processes, the Chl-a concentration

257 can be computed by combining three principal temporal and spatial eigenvectors at each station

258 as follows:

259

260     $\ln(Chl\ a(x_{t,i})) = \sum_{k=1}^{3} S(t,k)V(k,k)D(k,i), \quad i = 1,\ldots, 6$ (6)

261 We also compared model predictions and observations of Chl-a at each station as verification.

262 The sensitivity tests were also conducted to evaluate the model response to change of riverine

263 loading. All components including data transformation, EOF analysis, and LS-SVM simulation

264 were implemented in the Matlab. A detailed flow-chart of the procedure for machine learning is

265 shown in Fig. 3.

266 **3. Results**
267

268 *3.1 EOF analysis*

269   The EOF results are listed in Table 3. The first 4 eigenvalues have a total contribution of

270 91%. Fig. 4 shows the spatial pattern of these stations based on the 1st and 2nd dominant modes.

271 It can been seen that Stations TF5-2A and TF5-3 are close to each other in the lower right corner,

272 while Stations TF5-4, TF5-5, and TF5-5A concentrate in the upper left corner, and Station TF5-6

273 is between these two groups. The upper tidal freshwater region, where Stations TF5-2A and TF5-

274 3 are located, has the negative spatial value of the second mode, indicating that the change of

275 eigenvector value is in the opposite direction as the downstream TF region. The correlations

14

276   among parameters are similar to the distance between different stations with respect to the 1st

277   mode (Fig. 4). High correlations exist between stations close to each other. The pattern of the

278   distribution appears to be determined by similarities in geomorphology within and between TF

279   segments. Stations TF5-2A and TF5-3 are located in a narrow upper TF (Fig. 1), where water

280   moves fast and residence time is less than 5 days under the mean flow condition (Shen and Lin,

281   2006) and less algae can accumulate. Stations TF5-4, TF5-5, and TF5-5A, in contrast, are

282   located in a wide segment with a relatively long residence time, which can create a favorable

283   condition for algae to grow (Bukaveckas et al., 2011). Station TF5-6 is located downstream

284   where the channel becomes narrow again and it can be influenced by nutrient loadings from the

285   Fall Line and upstream transport of nutrients from the saline-water region due to estuarine

286   circulation.

287   *3.2 Simulation using LS-SVM*

288   The model results for training and verification to fit eigenvectors for the first four modes

289   are shown in Fig. 5. It can be seen that the model has the best skill for the first three modes with

290    $r^2 = 0.77, 0.60$, and 0.36 ($p < 0.0001$), respectively. The performance decreases and varies for

291   different modes. There is no predictive skill for the last three modes as they contribute minor

292   contributions (Table 3) and are distributed randomly.

293   Using the first three modes of the EOF prediction, the Chl-a concentration can be

294   computed using Eq. 6. The prediction results are shown in Fig. 6. The correlation (r^2) and the

295   root-mean-square error (RMSR), mean error (ME) ($\sum(M - O)/n$), absolute error (AE)

296   ($\sum |M - O|/n$), and model skill $SS = 1 - \frac{\sum(M-O)^2}{\sum(O-\bar{O})^2}$ are used to quantify the model performance,

297   where M is model output, O is observations, and n is the number of observations. These statistics

298   are commonly used for model skill assessment (Cerco and Noel, 1993; Allen et al., 2007;

15

299 Maréchal, 2004; Willmott, 1981). Statistical results for data used for training and prediction are

300 listed in Table 4. It can be seen that the model prediction skill at each station is different. The

301 skill for the model prediction is lower than the skill for the model training period. Performance

302 levels are often categorized by SS as: > 0.65 excellent; 0.65–0.5 very good; 0.5–0.2 good; < 0.2

303 poor (e.g., Maréchal, 2004; Allen et al., 2007). The very good predictions are found at Stations

304 TF5-4, TF5-5, and TF5-5a ($r^2 > 0.56$, SS > 0.5). The prediction skill decreases at Stations TF5-

305 2A, TF5-3 and TF5-6. The worst station is TF5-2A in term of SS ($r^2 = 0.53$ and SS $= 0.11$)

306 though the correlation coefficient is still high, suggesting that it is difficult to simulate high

307 variations of algal blooms at this station. Both bias and the absolute difference between model

308 training and prediction are on the same order. Based on model skill assessment statistics, overall,

309 the model prediction skill is satisfactory based on the model skill assessment statistical measures

310 (Maréchal, 2004; Allen et al., 2007). Compared to previously published applications of Chl-a

311 simulations, the model skill is lower than that of Xie et al. (2012) based on correlations and mean

312 errors. One of the reasons is that we only used seven environmental variables, while more

313 independent variables at model station were used for training by Xie et al. (2012). The

314 performance is comparable to Park et al. (2015) at most stations. Predictive skills at many

315 stations are comparable to complex water quality models as well (e.g., Wu and Xu, 2011). The

316 model simulation period is from 1990 to 2013, which covers both wet and dry seasons. Qin and

317 Shen (2017) compared the interaction of biological and physical transport processes under

318 different timescales and found that there is a good correlation between algal biomass and

319 residence time under seasonal to annual scales in the TF portion of the James River. The inverse

320 relationship between algal biomass and the flushing effect of physical transport in this area was

321 successfully reflected by the model results, which show that the Chl-a concentration is lower

322 during the high-flow period from 2003-2006 than during the 2000-2003 low-flow period.

323   The model results show a discrepancy in observations during summer when HABs exist

324 (very high Chl-a concentration). It may be due to some factors (or variables) controlling HABs

325 that are not exclusively included in the current model (e.g. competition of nutrients and light

326 between species), as it is still not well-known why microcystin is often observed when Chl-a

327 concentration is high (Bukaveckas et al., 2018). Chl-a observations are conducted monthly,

328 which may be not insufficient for simulating microcystin. It appears that a high-frequency

329 observation of Chl-a is needed to improve the model skill.

330 *3.3 Response to nutrient reduction*

331   The LS-SVM learning machine maps the training data nonlinearly into a higher-

332 dimensional feature space and constructs a separating hyperplane with a maximum margin there.

333 It then classifies new data based on the distance from the training data and separates these data

334 into different classes. However, though the model prediction skill is satisfactory, the application

335 of the model other than the prediction of Chl-a concentration may be limited as the model

336 depends on training data. For example, it may be questionable if the model will respond to the

337 changed nutrient reduction because the model may not be trained based on the underlying

338 biological processes. However, with effective transformation of nutrient data (e.g. making model

339 sensitive to low nutrients) and sufficient training data, the response of model to nutrients is

340 feasible. To evaluate the reliability of the model application for nutrient reduction, it is useful to

341 examine if the model responses to the changes of nutrients are reasonable. In this study, a model

342 simulation was conducted by simultaneously reducing the loadings of TN and TP by 50%. After

343 TN and TP are reduced from the watershed, the Chl-a concentration at the Fall Line will

344   decrease proportionally by 50% as well. The model results compared with the baseline condition

345   is shown in Fig. 7. The Chl-a concentration decreased correspondingly with reductions of TN

346   and TP loadings. In the upper TF region, the reduction of Chl-a concentration is about 45%,

347   lower than 50%. In the middle to lower TF, the reduction ranges from 36-41%. The reduction is

348   about 36% at the downstream Station TF5-6. This comparison shows that the model response to

349   the loading reduction is reasonable, which varies at different stations. More discussion of the

350   model response to loading reduction will be presented in the Discussion Section.

351   **4. Discussion**
352

353   *4.1 Contribution of EOF mode*

354   The purpose of applying EOF analysis to the entire TF region is to use a single model to

355   simulate algal blooms at different locations in the TF region rather than building a series of

356   models at each station. The approach of applying EOF analysis has the potential to be applied to

357   the entire estuary.

358   As shown in Table 3, the 1st mode accounts for about 62% of the variance using the matrix of

359   data at the 7 stations. However, the contribution of the 1st mode to the variations in Chl-a

360   concentration at each station is different. Fig. 8 shows examples of model simulations with

361   respect to using different numbers of EOF modes at Stations TF5-3 and TF5-4, respectively. It

362   can be seen that the 2nd and 3rd modes are important to improve the model prediction skill as well

363   as the 1st mode at Station TF5-3, where the r^2 value improved from 0.54 to 0.76 and the RMSE

364   value reduced from 7.24 to 5.75 µg/L. In contrast, the 1st mode has the dominant contribution to

365   predict Chl-a concentration at Station TF5-4, while adding the 2nd and 3rd modes have much

366   smaller contributions. The correlation r^2 value increases from 0.66 to 0.70 and the RMSE value

367   decreases from 14.13 to 13.39 µg/L, suggesting that the contribution of each mode to different

18

368   stations varies and the LS-SVM learning machine is able to fit Chl-a concentrations at different

369   stations with the use of the same independent variables.

370   *4.2 Variable transformation and model response to load reduction*

371   For this study, we conducted variable transformations for TN and TP concentrations. If we

372   directly use TN and TP loadings for the model, the Chl-a concentration will decrease because

373   both TN and TP loadings are high in spring and low in summer. The model will also not respond

374   to the nutrient reduction correctly as the Chl-a concentration will increase rather than decrease in

375   summer.

376   Because algal blooms are highly temperature-dependent, including the temperature effect

377   implicitly rather than using temperature itself as an independent variable is also important for the

378   model. If we use temperature as an independent variable directly without transformation (Eq. 1),

379   the model can simulate Chl-a concentration well with the same or improved skill for the model

380   training (Table 5). However, the model response to nutrient reduction will be incorrect (Fig. 9).

381   Compared to Fig. 7, the Chl-a concentrations increase at Station TF5-2A and the maximum

382   reduction is less than 19%. Our approach, instead, is to apply a temperature correction to the

383   nutrients. The approach is similar to the approach for nutrient limitation by using the Monod

384   function for algal growth (Eq. 1), while the temperature modification is to amplify the effects of

385   nutrient limitation and the benthic fluxes of nutrients from the bottom sediment in summer. The

386   model is sensitive to the selection of the half-saturation nutrient value. We used 75th percentile

387   values of TN and TP concentrations and the selection of the values are based on model

388   performance.

389   4.3 Model limitation

390 The current model is built based on nonpoint source loadings of TN and TP and not explicitly

391 expressed as DIN and DIP, and we did not include the point source loadings as independent

392 variables as we assume they are close to constant based on the designed discharge flow without

393 much seasonal variations, which the discharge maybe not always be constant. It is expected that

394 the total reduction is lower than the reduction of DIN and DIP loadings, especially from point

395 sources during the summer period. When time-varying point source data become available,

396 especially including time-varying DIN and DIP loadings at downstream of Fall Line, the model

397 response to nutrient loading reduction will be more accurate. It can be seen that the model can

398 simulate interannual variations of algal blooms, but frequently under-estimate high bloom

399 concentrations. As the cause of the HAB does not depends solely on hydrodynamic conditions

400 and nutrients, the competition of nutrients between different algal species can also contribute to

401 the variations. The model has no prediction skill for the last three modes of EOF indicating that

402 some variations can be due to nonlinear and random effects. Occasionally, we can see that the

403 Chl-a concentration increases while the nutrient concentration decreases. This is partially due to

404 the non-linear behavior of algae. For example, as algal growth decreases, the light condition can

405 be improved and nutrient may become available at the downstream in reality, and the Chl-a

406 concentration can increase at some stations if it is light-limited. Therefore, a detailed evaluation

407 is needed when applying the model to realistic simulations. Nevertheless, the model can be used

408 to evaluate the impact of the non-point source of nutrients on algal blooms in the TF area.

409 **5. Conclusions**

410 An approach using long-term observational data and the LS-SVM learning machine for

411 simulating algal bloom in the TF region of the James River estuary was conducted. The

412 simulation period spanned from 1990-2013, which included both wet years and dry years. The

413   EOF method was introduced to reduce the data dimension that enables us to model the algal

414   bloom in the entire TF region using only one model. The model simulated well seasonal and

415   interannual variations of an algal bloom during the summer low-flow periods and the low Chl-a

416   concentrations during a high-flow years. The model performance has a good modeling skill ($r^2 >$

417   0.5 and SS > 0.5) for most stations based on statistical measures. The results show that the bloom

418   is highly modulated by the hydrodynamic condition. The model experiments with changes in

419   nutrient loadings indicate that it has a correct response to nutrient loading reduction. Our

420   modeling exercise indicates that an adequate data transformation is needed in order to use LS-

421   SVM to adequately simulate an algal bloom and its response to loading changes.

422   As only nonpoint source nutrient loadings were included in the model, the algal bloom

423   simulated can be considered as the response to the upstream nutrient loading. The model

424   simulation results can be further improved if DIN, DIP, and additional parameters are included.

425   This study demonstrates that the use of the LS-SVM learning machine is a good approach for

426   simulating algal blooms in the complex environment of the TF portion of the James River with

427   high efficiency, which can be applied to many other estuaries.

433

434

435 **7. References**
436

437 Allen, J.I., Somerfield, P.J., Gilbert, F.J., 2007. Quantifying uncertainty in high-resolution
438 coupled hydrodynamic-ecosystem models, *J. Mar. Syst.*, 64, 3–14.
439
440 Anderson, C.R., Sapiano, M., Prasad, M., Long, W., Tango, P.J., Brown, C., Murtugudde, R.,
441 2010. Predicting potentially toxigenic Pseudo-nitzschia blooms in the Chesapeake Bay. *J.*
442 *Marine Syst.* 83(3-4), 127-140.
443
444 Bergamino, N., Loiselle. S.A., C´ozar, A., Arduino M. Dattilo, A.M., Bracchini, L., Claudio
445 Rossi, C. 2007. Examining the dynamics of phytoplankton biomass in Lake Tanganyika using
446 Empirical Orthogonal Functions. *Ecological Modeling*, 204,156-162.
447
448 Blauw, A., Los, H., Huisman, J., Peperzak, L., 2010. Nuisance foam events and Phaeocystis
449 globosa blooms in Dutch Coastal waters analyzed with fuzzy logic. *J Marine Syst.* 3(3):115-126.
450
451 Buchanan, C., Lacouture, R., Marshall, H.G., Olson, M., Johnson, J.M., 2005. Phytoplankton
452 reference communities for Chesapeake Bay and its tidal tributaries. *Estuaries* 28: 138-159.
453
454 Bukaveckas, P.A., Barry, L.E., Beckwith, M.J., David, V., Lederer, B., 2011. Factors
455 determining the location of the chlorophyll maximum and the fate of algal production within the
456 tidal freshwater James River. *Estuarine, Coasts*, *34*, 569–582.
457
458 Bukaveckas, P., Beck, R., Devore, D., Lee, W.M., 2017. Climatic variability and its role in
459 regulating C, N and P retention in the James River Estuary. *Estuarine, Coastal and Shelf*
460 *Science*, doi.org/10.1016/j.ecss.2017.10.004.
461
462 Bukaveckas, P.A., Franklin, R., Tassone, S., Trache, B, Egerton, T. 2018. Cyanobacteria and
463 cyanotoxins at the river-estuarine transition. *Harmful Algae,* 76, 11–21
464
465 Cerco, C., Noel, M.R., 1993. Three-dimensional eutrophiciton model of Chesapeake Bay.
466 *Journal of Environmental Engineering*, 1906-1025.
467
468 Cerco, C., Noel, M.R., 2004. Process-based primary production modeling of in Chesapeake Bay.
469 *Marine Ecology Progress Series*, 282:45-58.
470
471 Crisci, C., Ghattas, B., Perera, Ghattas, 201. A review of supervised machine learning algorithms
472 and their applications to ecological data. *Ecological Modeling* 240,113-122.
473
474 Du, J., Shen, J., Park, K., Wang, Y.P., Yu, X., 2018. Worsened physical condition due to climate
475 change contributes to the increasing hypoxia in Chesapeake Bay. *Science of the Total*
476 *Environment* 630, 707–717.
477
478 James, R. T., 2016. Recalibration of the Lake Okeechobee water quality model (LOWQM) to
479 extreme hydro-meteorological events. *Ecological Modeling*, 325,71-83.

480
481 Jiang, L., Lia,Y., Zhaoa,X. Tillotsonc,M.R., Wangb, W., Zhang, S., Sarpong, L., Asmaab, Q.,
482 Pane,B 2018. Parameter uncertainty and sensitivity analysis of water quality model in Lake
483 Taihu, China, *Ecological Modeling*, 375, 1-2.

484
485 Jiang, L., Xia, M. 2017. Wind effects on the spring phytoplankton dynamics in the middle reach
486 of the Chesapeake Bay. *Ecological Modeling*, 363(10),68-80.

487
488 Kim, J. Lee, T., Seo, D. 2017. Algal bloom prediction of the lower Han River, Korea using the
489 EFDC hydrodynamic and water quality model. *Ecological Modeling*, 266(24), 27-36.

490
491 Kong, X., Sun, Y., Su, R., Shi, X., 2017. Real-time eutrophication status evaluation of coastal
492 waters using support vector machine with grid search algorithm. *Marine Pollution*
493 *Bulletin* 119, 307-319.

494
495 Liang, S. Han, S., Sun, Z., 2015. Parameter optimization method for the water quality dynamic
496 model based on data-driven theory. *Marine Pollution Bulletin* (98), 137-147.

497
498 Lucas, L.V., Thompson, J.K., Brown, L.R., 2009. Why are diverse relationship observed
499 between phytoplankton biomass and transport time?, *Limnol. Oceanogr.*, 54(1), 381–390.

500
501 Lui, G.C.S.,Li, W.L.,Leungb,K.M.Y., Lee, J.H.W., Jayawardenac,A.W. 2007. Modelling algal
502 blooms using vector autoregressive modelwith exogenous variables and long memory filter.
503 *Ecological Modeling*, 200, 130-138.

504
505 McGillicuddy Jr, D.J., 2010. Models of harmful algal blooms: conceptual, empirical, and
506 numerical approaches. *J Mar Syst*. 83(3-4): 105–107. doi:10.1016/j.jmarsys.2010.06.008.

507
508 Maréchal, D., 2004, A soil-based approach to rainfall-runoff modeling in ungauged catchments
509 for England and Wales, Ph.D. thesis, Cranfield Univ., Cranfield, U. K.

510
511 Marshall, H.G., Lacouture, R., Buchanan, C., Johnson, J.M., 2006. Phytoplankton assemblages
512 associated with water quality and salinity regions in Chesapeake Bay, USA. *Estuarine, Coastal*
513 *and Shelf Science* 69: 10-18.

514
515 Moe, S. J., Haande, S., Couture, R. 2016. Climate change, cyanobacteria blooms and ecological
516 status of lakes: A Bayesian network approach. *Ecological Modeling*, 337, 330-347.

517
518 Muttil, N., Chau, K-W., 2006. Neural networks and genetic programming for modeling coastal
519 algal blooms. *International Journal of Environment and Pollution*, 28 (3-4), 223-238.

520
521 Paerl, H.W., Fulton III, R.S., Moisander, P.H., 2001. Harmful freshwater algal bloos, with an
522 emphasis on cyanobacteria. *The Scientific World*, 1, 76-113. DOI 10.1100/tsw.2001.16.

523
524 Paerl, H.W., 2009. Controlling eutrophication along the freshwater–marine continuum: dual
525 nutrient (N and P) reductions are essential. *Estuaries and Coasts* 32,592-601.

526
527 Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning
528 protocol for predicting chlorophyll-a concentration using machine learning models in freshwater
529 and estuarine reservoirs, Korea. *Science of the Total Environmental*, 202, 31-41.
530
531 Park, K., Kuo, A.Y., Shen, J., Hamrick, J.M., 1995. A Three-Dimensional Hydrodynamic
532 Eutrophication Model (HEM-3D): Description of Water Quality and Sediment Process
533 Submodels; Special Report in Applied Marine Science and Ocean Engineering No. 327; Virginia
534 Institute of Marine Science: Gloucester Point, VA, USA, p. 102.
535
536 Qin, Q., Shen, J., 2017. The contribution of local and transport processes to phytoplankton
537 biomass variability over different timescales in the Upper James River, Virginia. *Estuarine,*
538 *Coastal and Shelf Science* 196, 123-133.
539
540 Recknagel, F. 2001. Applications of machine learning to ecological modelling. *Ecological*
541 *Modelling* 146, 303–310
542
543 Recknagel, F., French, M., Harkonen, P., Yabunaka, K-I., 1997. Artificial neural network
544 approach for modelling and prediction of algal blooms. *Ecological Modelling* 96 (1997) 11-28.
545
546 Ribeiro, R., Torgo, L., 2008. A comparative study on predicting algae blooms in Douro
547 River, Portugal. *Ecological Modelling* 212, 86–91.
548
549 Riverson, J., Coats, R., Costa-Cabral, M., Dettinger, M., Reuter, J., Sahoo, G., Schladow, G.,
550 2013. Modeling the transport of nutrients and sediment loads into Lake Tahoe under projected
551 climatic changes. *Climatic Change*. 116:35-50.Rounds, S.A., 2002. Development of a neural
552 networks model for dissolved oxygen in the Tualatin Rier, Oregon. In Proceedings of the Second
553 Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada.
554
555 Seitzinger, S.P., 1991. The effect of pH on the release of phosphorus from Potomac estuary
556 sediments: implications for blue-green algal blooms. *Estuary, Coastal and Shelf Science* 33(4),
557 409-418.
558
559 Shen, J., Wang, Y., Sisson, M., 2016. Development of the hydrodynamic model for long-term
560 Simulation of water quality processes of the tidal James River, Virginia. Journal of Marine
561 Science and Engineering. *J. Mar. Sci. Eng.* 2016, *4*(4), 82.
562
563 Shen, J., Zhao, Y., 2010. Combined Bayesian statistics and load duration curve method for
564 bacteria nonpoint source loading estimation. *Water Research*, 44, 77-84.
565
566 Shen, J., Wang, T., Herman, J., Mason, P., Arnold, G.L., 2008. Hypoxia in a Coastal Embayment
567 of the Chesapeake Bay: A Model Diagnostic Study of Oxygen Dynamics. *Estuaries and Coasts*,
568 31,652-663.
569
570 Shen, J., Lin, J., 2006. Modeling Study of the Influences of tide and stratification on age of water
571 in the tidal James River. *Estuarine*, *Coastal and Shelf Science,* 68 (1-2): 101-112.

Shen, J., 2006. Optimal estimation of parameters for an estuarine eutrophication model. Ecological Modeling 191 (3–4), 521–537.

Shen, J., Parker, A., Riverson, J., 2005. A new approach for a Windows-based watershed modeling system based on a database-supporting architecture, *Environmental Modelling & Software* (20), 1127–1138.

Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural Process. Lett*. 9 (3), 293–300.

Suykens, J.A., Van Gestel, T., De Brabanter, J., et al., 2002. Least Squares Support Vector Machines. World Scientific (ISBN 981-238-151-1).

Thomann, R.V., Mueller, J., 1987. Principles of Surface Water Quality Modeling and Control. Harper Collins Publishers, New York, 644 pp.

USEPA, 2010. Ambient Water Quality Criteria for Dissolved Oxygen, Water Clarity and Chlorophyll a for the Chesapeake Bay and Its Tidal Tributaries 2008 Technical Support for Criteria Assessment Protocols Addendum, EPA 903-R-10-002.

van Straten, G., 1983. Maximum likelihood estimation of parameters and uncertainty in phytoplankton models. In: Beck, M.B., van Straten, G. (Eds.), Uncertainty and Forecasting of Water Quality. Springer–Verlag, pp. 157–171.

Vapnik, V.N., 1999. An overview of statistical learning theory. *Neural Networks, IEEE Trans*. 10 (5), 988–999.

Volf, G., Atanasova, N., Kompare, B., Precali, R., Oani, N., 2011. Descriptive and prediction models of phytoplankton in the northern adriatic. *Ecological Modelling* 222, 2502–2511.

Wang, J., Tang, D., 2010. Winter phytoplankton bloom induced by subsurface upwelling and mixed layer entrainment southwest of Luzon Strait. *Journal of Marine Systems*, 83(3–4), 141-149.

Willmott, C.J., 1981. On the validation of models. *Physical Geography* 2, 184-194.

Wu, J., Zou, R., Yu., S.L., 2006. Uncertainty analysis for coupled watershed and water quality modeling systems. Journal of Water Resources Planning and Management 132 (5), 351–361.
Wilmott, C.J., 1981. On the validation of models. *Physical Geography* 2, 184-194.

Wu, G., Xu, Z. 2011. Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake. *Ecological Modeling*, 222, 1245-1252.

Xie, Z., Lou, I., Ung, W.K., Mok, K.M., 2012. Freshwater Algal Bloom Prediction by Support Vector Machine in Macau Storage Reservoirs. Hindawi Publishing Corporation. *Mathematical Problems in Engineering*, Volume 2012, Article 397473, 12 pages,doi:10.1155/2012/397473.

618
619   Zhang, C., Zhang, T., Yuan, C., 2011. Oil holdup prediction of oil-water two phase
620   flow using thermal method based on multiwavelet transform and least squares
621   support vector machine. *Expert Syst. Appl.* 38 (3), 1602–1610.
622
623   Zhang, X., Srinivasan, R., Liew, M.V., 2009. Approximating SWAT Model Using
624   Artificial Neural Network and Support Vector Machine. *Journal of the American Water*
625   *Resources Association (JAWRA)* 45(2):460-474. DOI: 10.1111/j.1752-1688.2009.00302.

626

627 *Table 1. Correlation of selected independent variables and Chl-a concentration.*

628

	Flow	TSS	TN	TP	light	$(light)^{1/2}$
Original data	-0.43	-0.26	-0.03	-0.01	0.44	0.42
Transformed	-0.65	-0.26	0.46	0.35	0.44	0.42

632

633

634

635

636

637

Table 2. List of Variables and transformation used for model input

Name	Variable	Transformation	Parameter values
Chlorophyll a (Chl-a)	Observations at each station (state variable)	logarithmic transformation for Chl-a at each station	
Chlorophyll a (Chl-a)	Observation at Full Line	Convert to loading (concentration ×flow×86400) (ug d^{-1})	
Flow (Q)	Daily observation at USGS flow station	Convert to Q$^{1/3}$, backward 5-day running average	
Temperature (T)	Observation at full line	θ^{T-20}	θ = 1.03
Suspended solid	Observation at full line	Convert to loading (concentration ×flow×86400) (g d^{-1})	
Total nitrogen (TN)	Observation at full line	Convert to loading (concentration ×flow×86400) (g d^{-1}), backward 120 moving average, and introduce new independent variable[1] $$TN_{new} = \frac{TN}{H_{TN}+TN}\theta^{T-20}$$	H_{TN} = 75th percentile of loading
Total phosphorus (TP)	Observation at full line	Convert to loading (concentration ×flow×86400) (g d^{-1}), backward 120 moving average, and introduce new independent variable[1] $$TP_{new} = \frac{TP}{H_{TP}+TP}\theta^{T-20}$$	H_{TP} = 75th percentile of loading
Solar radiation	Observation at full line	15-day average	

639

640

641

28

642

643 *Table 3. Contribution of EOF modes.*

mode	1	2	3	4	5	6
Eigenvalue	90.79	20.43	12.40	8.36	8.06	5.72
Contribution	62%	14%	9%	6%	6%	4%
Accumulative contribution	62%	76%	85%	91%	96%	100%

644

645

648

Table 4. A summary of model skill.

	RMSE		r²		ME		AE		SS	
Station	Train.	Pred.	Train.	Pred.	Train.	Pred.	Train.	Pred.	Train.	Pred.
TF5-2A	3.00	6.79	0.67	0.53	-0.75	-0.79	1.62	2.73	0.76	0.11
TF5-3	4.34	6.93	0.76	0.67	-1.13	-0.59	2.31	3.17	0.69	0.29
TF5-4	13.88	12.87	0.70	0.58	-3.55	-3.77	7.56	8.79	0.58	0.54
TF5-5	14.61	14.21	0.72	0.71	-4.61	-3.66	9.06	9.29	0.62	0.50
TF5-5A	15.47	13.10	0.63	0.56	-4.53	-2.98	9.63	9.75	0.56	0.52
TF5-6	7.92	8.78	0.51	0.29	-2.34	-0.72	4.96	6.28	0.50	0.27

649

650

651

Table 5. A summary of model skill using temperature as an independent variable.

	RMSE		r²		ME		AE		SS	
Station	Train.	Pred.	Train.	Pred.	Train.	Pred.	Train.	Pred.	Train.	Pred.
TF5-2A	2.22	7.24	0.66	0.45	-0.41	-0.16	1.15	3.21	0.87	-0.01
TF5-3	3.14	7.37	0.78	0.64	-0.68	0.24	1.75	3.58	0.84	0.20
TF5-4	12.79	13.86	0.68	0.50	-2.53	-3.79	6.70	9.67	0.65	0.47
TF5-5	12.65	14.29	0.74	0.65	-3.41	-3.99	7.85	9.54	0.72	0.50
TF5-5A	12.90	14.70	0.67	0.50	-3.38	-3.05	8.38	10.78	0.69	0.40
TF5-6	7.19	9.16	0.55	0.30	-1.90	-0.45	4.34	6.45	0.58	0.20

652

653

654

655

Figure 1. Map of the tidal freshwater James River Estuary and the monthly monitoring locations in the mainstem.

Figure 2. The distribution of the Chl-a concentration in the log scale along the James River mainstem.

Figure 3. A flow-chart for simulation procedure.

Figure 4. Spatial pattern of EOF for each observation station.

Figure 5. Comparison of model simulation of temporal vectors for each of the first four EOF modes (data with red circles are used for training).
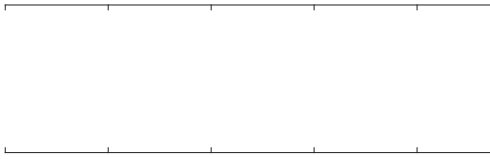
Figure 6. Comparison of model simulation and observations of Chl-a concentration (Black circles are observations, blue lines are training, and red lines are model predictions. Numbers show root-mean-square-error and r^2 for training data and model prediction inside brackets).
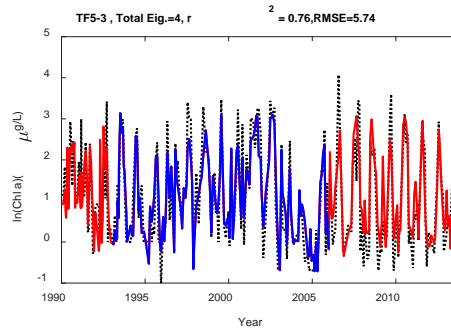
Figure 7. Comparison of model simulation with reduction of TN, TP, and Chl-a loadings by 50% simultaneously to the baseline condition (Black lines are baseline simulation and red lines are simulation with load reduction).

Figure 8. Comparison of contribution of each modes to the accurate prediction of Chl-a concentrations at Stations TF5-3 and TF5.4 (Black lines are observations, red lines are model simulations, and r^2 values are for training).
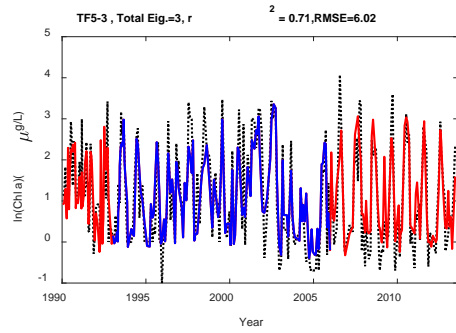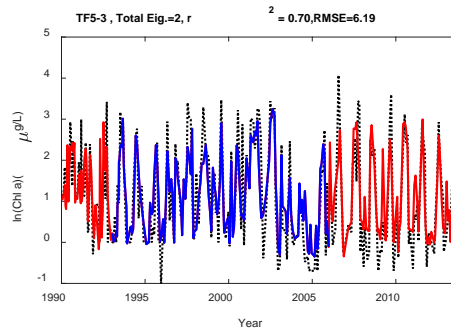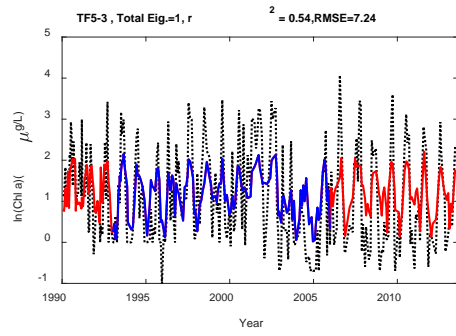
Figure 9. Comparison of model simulation with reduction of TN, TP, and Chl-a loadings by 50% simultaneously to the baseline condition using temperate as an independent variable (Black lines are baseline simulation and red lines are simulation with load reduction).
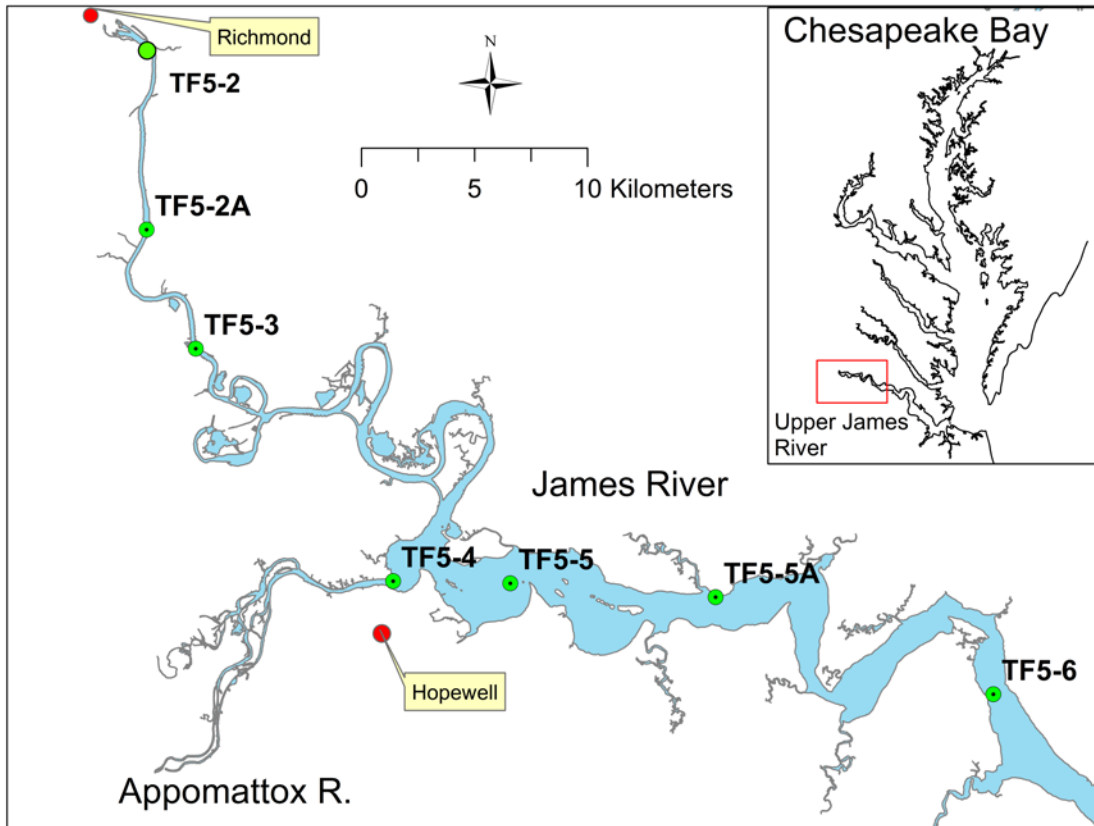
688

689

690 Figure 1. Map of the tidal freshwater James River Estuary and the monthly monitoring locations
691 in the mainstem.

692

693
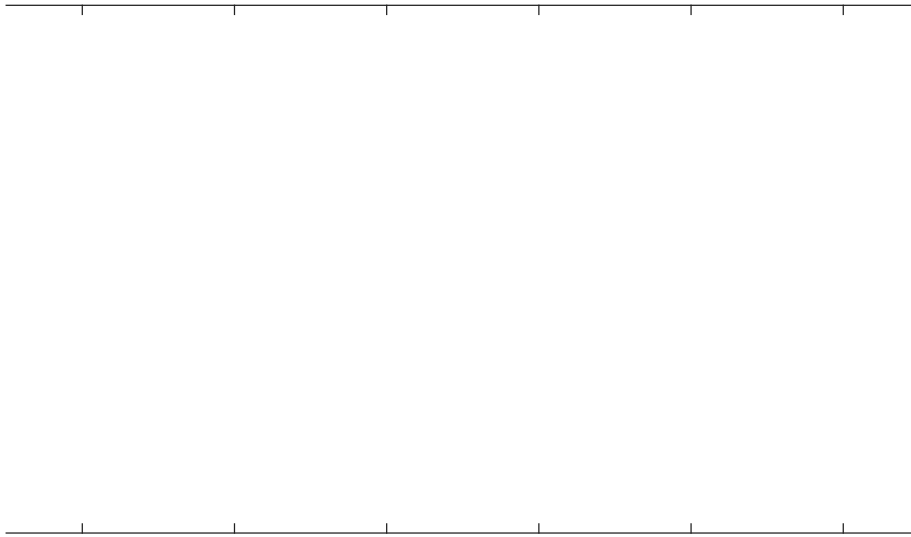
694

695

696　Figure 2. The distribution of the Chl-a concentration in the log scale along the James River
697　mainstem.

698

699

700

701

702

703

Figure 3. A flow-chart for simulation procedure.
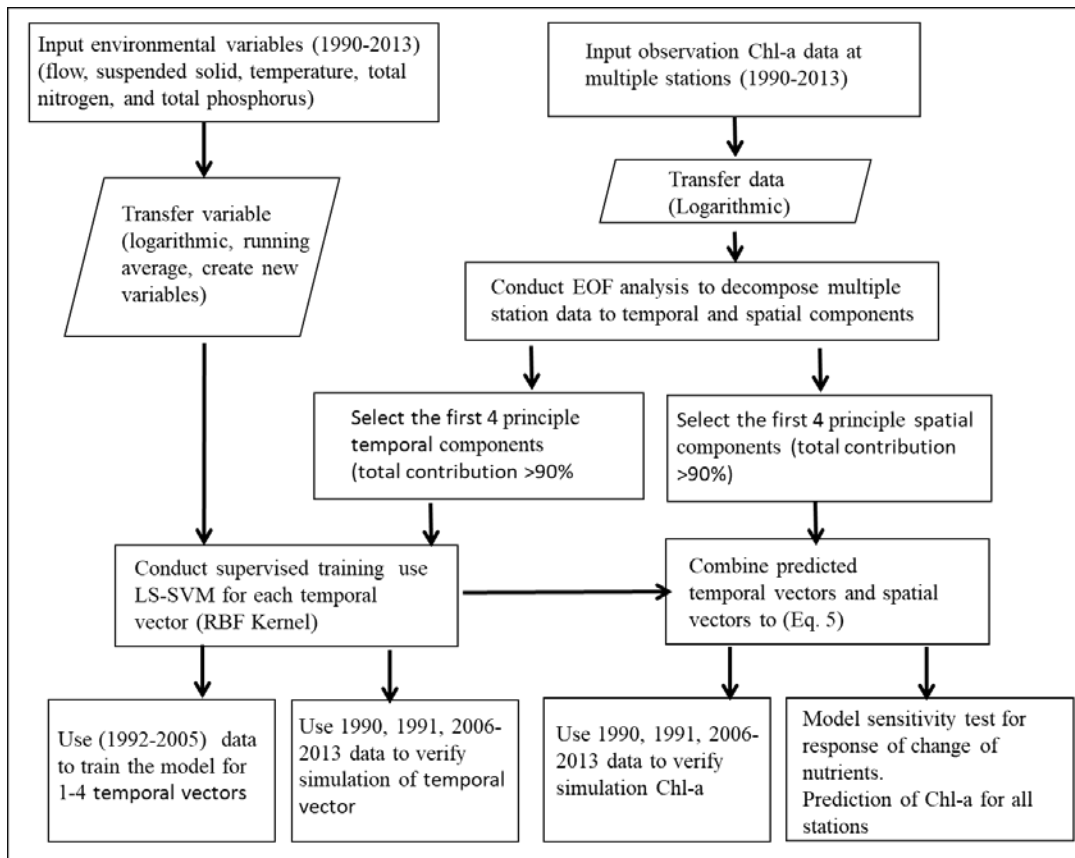
704
705
706
707
708
709
710

711
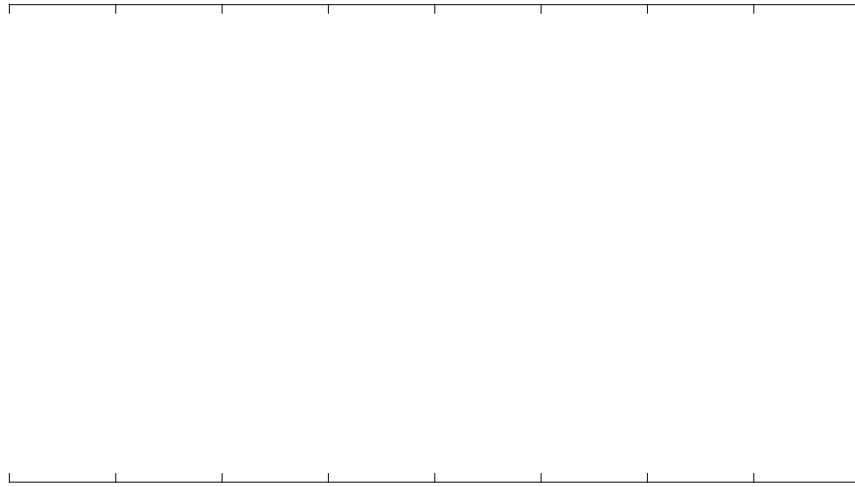
712

713   Figure 4. Spatial pattern of EOF for each observation station.

714

715
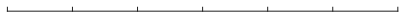
716

717

718   Figure 5. Comparison of model simulation of temporal vectors for each of the first four EOF
719   modes (data with red circles are used for training).
720

721

722
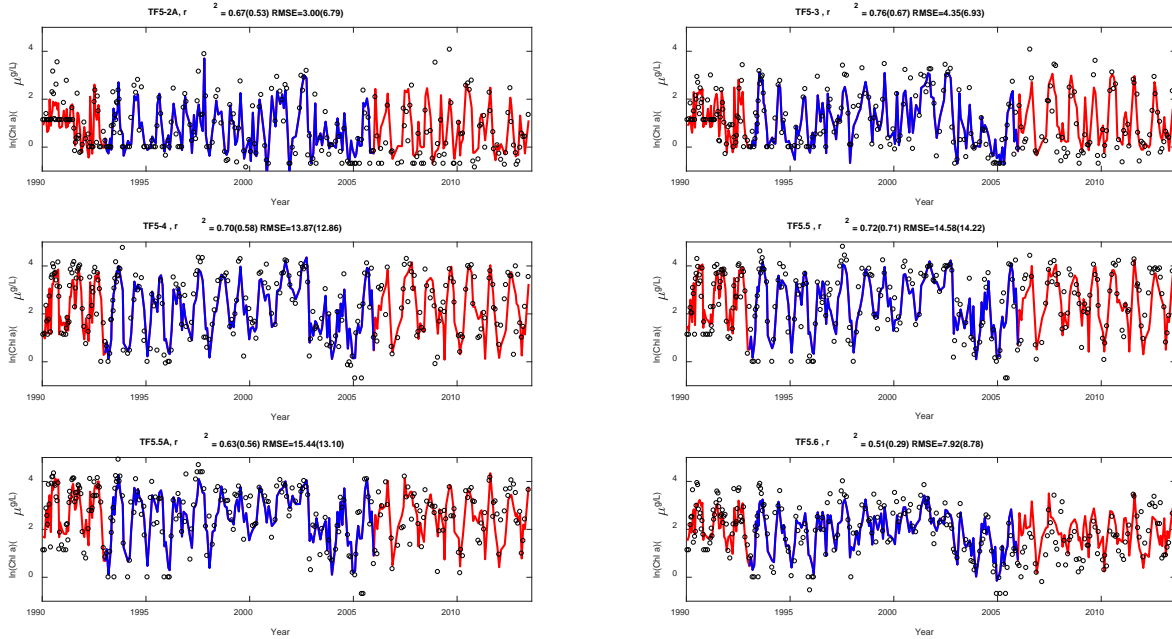
723

724 Figure 6. Comparison of model simulation and observations of Chl-a concentration (Black
725 circles are observations, blue lines are training, and red lines are model predictions. Numbers
726 show root-mean-square-error and r^2 is for training data and model prediction inside brackets).

727

728

729

730

731 Figure 7. Comparison of model simulation with reduction of TN, TP, and Chl-a loadings by 50%
732 simultaneously to the baseline condition (Black lines are baseline simulation and red lines are
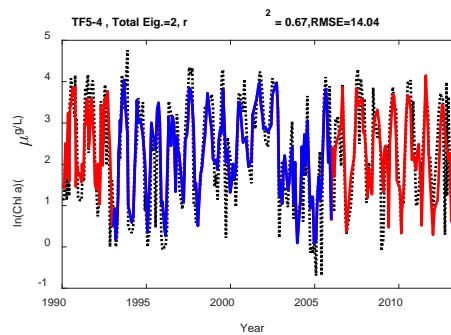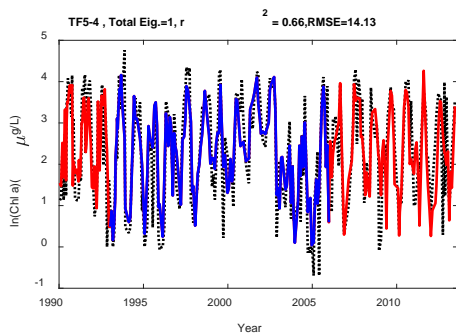733 simulation with load reduction).
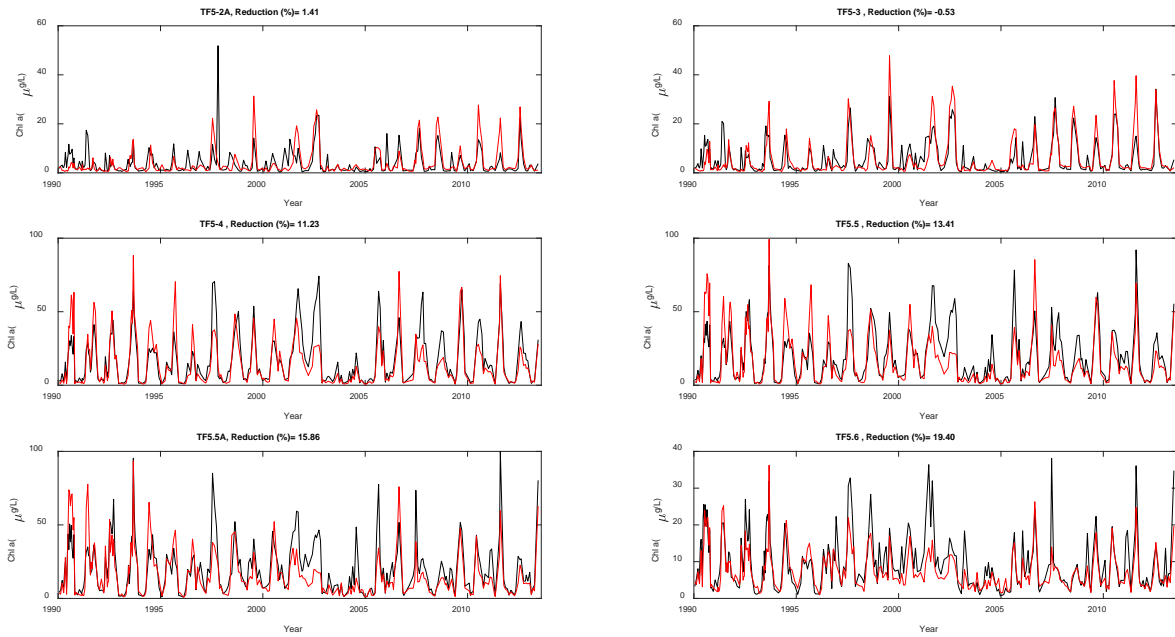734

735

736

737
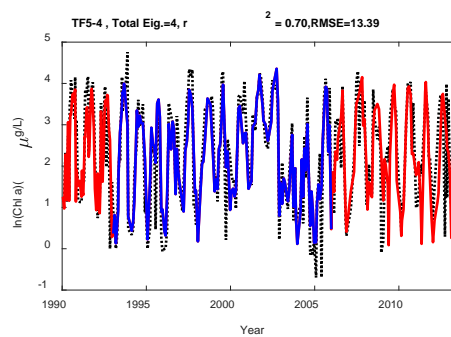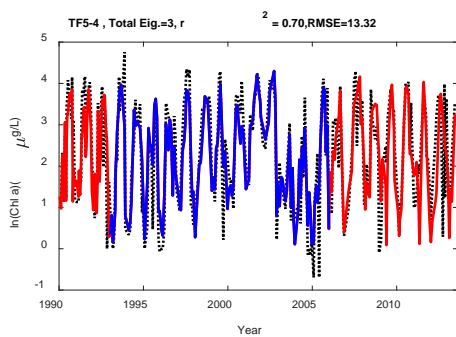
738

739   Figure 8. Comparison of contribution of each model to the accurate prediction of Chl-a
740   concentrations at Stations TF5-3 and TF5-4 (Black lines are observations and red lines, model
741   simulations, and r^2 values are for training).

742

743

744

745

746

Figure 9. Comparison of model simulation with reduction of TN, TP, and Chl-a loadings by 50%
simultaneously to the baseline condition using temperate as an independent variable (Black lines
are baseline simulation and red lines are simulation with load reduction).

750