Theor Appl Climatol
DOI 10.1007/s00704-016-1906-7

ORIGINAL PAPER

The development of a non-linear autoregressive model with exogenous input (NARX) to model climate-water clarity relationships: reconstructing a historical water clarity index for the coastal waters of the southeastern USA

Cameron C. Lee[1] · **Scott C. Sheridan**[1] · **Brian B. Barnes**[2] · **Chuanmin Hu**[2] · **Douglas E. Pirhalla**[3] · **Varis Ransibrahmanakul**[3] · **Karsten Shein**[4]

Received: 23 December 2015 / Accepted: 12 August 2016

Abstract The coastal waters of the southeastern USA contain important protected habitats and natural resources that are vulnerable to climate variability and singular weather events. Water clarity, strongly affected by atmospheric events, is linked to substantial environmental impacts throughout the region. To assess this relationship over the long-term, this study uses an artificial neural network-based time series modeling technique known as non-linear autoregressive models with exogenous input (NARX models) to explore the relationship between climate and a water clarity index (KDI) in this area and to reconstruct this index over a 66-year period. Results show that synoptic-scale circulation patterns, weather types, and precipitation all play roles in impacting water clarity to varying degrees in each region of the larger domain. In particular, turbid water is associated with transitional weather and cyclonic circulation in much of the study region. Overall, NARX model performance also varies—regionally, seasonally and interannually—with wintertime estimates of KDI along the West Florida Shelf correlating to the actual KDI at $r > 0.70$. Periods of extreme (high) KDI in this area coincide with notable El Niño events. An upward trend in extreme KDI events from 1948 to 2013 is also present across much of the Florida Gulf coast.

Electronic supplementary material The online version of this article (doi:10.1007/s00704-016-1906-7) contains supplementary material, which is available to authorized users.

✉ Cameron C. Lee
cclee@kent.edu

[1] Department of Geography, Kent State University, Kent, OH 44242, USA

[2] College of Marine Science, University of South Florida, St. Petersburg, FL, USA

[3] National Oceanic and Atmospheric Administration, National Ocean Service, Silver Spring, MD, USA

[4] NOAA National Centers for Environmental Information, Asheville, NC, USA

1 Introduction and background

The coastal waters of the southeastern USA encompass important habitats and resources for marine animals and plants as well as for human beings. Within this system are many marine and estuarine resources, including coral reefs, beaches, seagrasses, wetlands, and fisheries that provide critical functions and value to tourism and the economy (Causey 2002). However, data from the NOAA National Centers for Environmental Information (2015) indicate that this region has led the nation in negative economic impacts of climate/weather disasters between 1980 and 2014, with much of the damage resulting from tropical cyclones. One of the immediate impacts of these natural disasters is on coastal water clarity, as substantial sediment resuspension and coastal runoff often lead to significant degradation in water clarity, thus affecting all marine animals and plants that rely on ambient light (see Short and Wyllie-Escheverria 1996). Indeed, one of the long-term goals of the Tampa Bay Estuary Program has been improving water clarity to foster seagrass growth through regulating nutrient releases, which has resulted in tremendous progress (Greening et al. 2011, 2014). Monthly water clarity data have been used as one of the inputs to form a decision matrix to help make management decisions (Janicki et al. 2000). Likewise, water clarity also has a direct impact on coral reef health as it determines the

Published online: 25 August 2016

amount of available light in both the visible and ultraviolet wavelengths to the benthos (Barnes et al. 2015). Due to its importance, assessment of the long-term trend of water clarity in these environments is a priority.

The inherent role of climate on coastal water conditions and processes is often best understood in the context of singular weather events and/or extremes in regional climate. For example, changes in distribution patterns of chlorophyll, particulate and dissolved matter, and other bio-optical properties across the southeastern and Gulf coasts are often the result of regional events: tropical cyclones, extreme heat and cold events, and winter storms (Lohrenz et al. 1999; Liu and Weisberg 2005; Ault 2006; Hu and Muller-Karger 2007; Conmy 2008; Lirman et al. 2011; Sheridan et al. 2013; Pirhalla et al. 2014). Sheridan et al. (2013) used a synoptic climatological classification to connect variability in atmospheric circulation pattern frequencies across the southeastern USA to anomalies of chlorophyll off the west coast of Florida using Sea-viewing Wide Field of View Sensor (SeaWiFS) data. Particularly evident were the statistically significant associations between cyclonic/anticyclonic circulation pattern frequencies and chlorophyll levels across part of the study region. While these correlations suggest short-term variability in water parameters can be related to changes in atmospheric circulation, yet unexplored are the impacts that climate change trends have upon water parameters in the region.

Such analysis, however, is difficult due to lack of long-term synoptic datasets of coastal water clarity. The use of satellite observations is perhaps the only way to routinely provide systematic assessment of coastal water changes; however, reliable, uniform data on water clarity is only available since 1997 with the launch of the SeaWiFS satellite. Even within the last 20 years, satellite-derived water clarity data are only available under clear skies, making the data set spatially and temporally incomplete, rendering the evaluation of water clarity trends particularly difficult.

To assess more properly the trends and spatiotemporal variability in water clarity, in this paper, we utilize a novel approach to reconstruct a 66-year (1948–2013), daily-scale estimate of water clarity for the coastal waters of the southeastern USA. Specifically, we use the available historical observations of water clarity, along with their statistical relationship to atmospheric circulation over the region to develop a non-linear, neural network-based, time series model to recreate a complete historical time series of water clarity.

2 Data and methodology

Non-linear autoregressive models with exogenous input (NARX models) were chosen to define the relationship between climate and water clarity in each region. While considerable detail is provided below, outlining the specific methodology used and decisions made to optimize the NARX models used herein, a detailed discussion of these models is beyond the scope of this paper (please see: Maier and Dandy (2000) for a thorough discussion of artificial neural network models in general, Diaconescu (2008) for a discussion on NARX modeling, and Beale et al. (2014) for a description of implementing NARX models in Matlab specifically). While not commonly used in applied climatological or oceanographic research, the structure of NARX models still contain a predictand variable that we aim to model (the time series of a water clarity index in each region) and multiple predictor variables (time-lagged climate-related variables) that are used in the model itself. In the sections that follow, the data used as predictand and predictors are discussed, followed by the set-up of the NARX models. Due to the subdomain-scale variability in the relationship between meteorological factors and water clarity (Table 1), a separate model is produced for each of the regions, as defined below.

2.1 Kd-index and regionalization

Water clarity in the study area (southeastern USA) was defined through the calculation of a Kd-Index (KDI), derived using SeaWiFS and Moderate Resolution Imaging Spectroradiometer (MODIS, onboard the satellite Aqua) data. K_d (λ) (m^{-1}) is a measure of the attenuation of downwelling light at a certain wavelength λ (nm), which defines how fast light disappears when it propagates through the water. Light availability at a certain depth z (m), referenced against the surface light, is proportional to exp. ($-K_d \cdot z$). Higher K_d values indicate lower water clarity. Level-2 remote sensing reflectance (R_{rs}) data from SeaWiFS and MODIS/A within the bounds of 24° to 31° N, 78° to 98° W were downloaded from NASA Goddard Space Flight Center. Pixels with negative R_{rs} at any wavelength were excluded from analysis, as were pixels identified by standard Level-2 processing flags (Patt et al. 2003, see Barnes and Hu 2015 for specific flags used) and targets deeper than 50 m. The standard K_d_lee algorithm (Lee et al. 2005, 2009) and the K_d_lee modification optimized for optically shallow water (Barnes et al. 2013) were applied to the mapped R_{rs} data to derive $K_d(488)$ (i.e., K_d at 488 nm) for optically deep and optically shallow targets, respectively. A 5 × 5 median filter was applied to all $K_d(488)$ scenes, whereby pixels were also excluded if any previously masked pixels (either due to negative R_{RS} or Level-2 processing flags) were within the 5 × 5 box with the pixel of interest in the center. Using both SeaWiFS and MODIS data, daily mean images were created for each day from 1997 to 2013. From these mean images, mean and standard deviation climatologies (spanning all years) were calculated for each ordinal day (i.e., day-of-year). In the creation of these climatologies, data from the preceding

Table 1 Pearson bivariate correlation statistics between anomalous frequencies of each SSC weather type (top) and circulation pattern (bottom) and KDI for each of the nine regions. Correlations are for winter months (December–February) only and are aggregated to the weekly level

	DM	DP	DT	MM	MP	MT	MT+	MT++	TR
R1	0.07	0.11	−0.04	−0.02	−0.06	−0.13	−0.10	0.01	**0.17**
R2	**0.04**	0.21	**0.01**	**−0.04**	−0.11	−0.14	−0.14	**−0.02**	0.29
R3	**−0.03**	**0.05**	0.07	**−0.02**	−0.13	−0.06	**−0.04**	**0.00**	0.29
R4	0.08	0.04	0.03	0.03	0.07	**−0.14**	**−0.16**	−0.01	0.08
R5	0.01	**0.14**	−0.03	**0.19**	**0.19**	**−0.27**	**−0.16**	**−0.19**	**0.29**
R6	0.12	**0.16**	−0.01	**0.15**	**0.21**	**−0.26**	**−0.22**	**−0.23**	**0.22**
R7	**0.30**	0.02	−0.05	0.08	0.11	**−0.29**	**−0.24**	0.09	**0.40**
R8	**0.24**	−0.02	−0.07	0.03	0.04	**−0.20**	**−0.16**	0.05	**0.29**
R9	**0.19**	−0.07	−0.06	0.08	**0.17**	**−0.24**	−0.04	−0.05	0.11

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
R1	−0.02	0.00	**0.22**	−0.06	−0.09	0.06	−0.04	−0.08	−0.05	0.09
R2	**0.16**	−0.07	**0.16**	−0.09	−0.03	**0.19**	−0.07	−0.10	**−0.18**	−0.05
R3	0.09	**−0.21**	**0.16**	−0.13	−0.02	**0.21**	0.01	0.00	**−0.15**	−0.07
R4	0.09	0.04	0.06	−0.13	0.00	**0.21**	−0.07	−0.12	−0.05	−0.07
R5	**0.21**	0.13	**0.15**	**−0.25**	−0.03	0.12	−0.12	−0.14	−0.11	0.00
R6	**0.24**	**0.20**	0.08	−0.13	0.01	0.00	**−0.22**	−0.10	−0.13	0.03
R7	**0.22**	**0.24**	**0.17**	−0.13	0.06	−0.06	**−0.21**	**−0.17**	**−0.15**	0.00
R8	0.13	**0.25**	0.11	−0.10	0.01	−0.09	−0.13	−0.09	−0.05	−0.01
R9	0.00	**0.35**	**0.21**	−0.07	0.07	−0.13	**−0.14**	−0.14	−0.07	0.00

Bold and underlined values indicate statistically significant ($p < 0.05$) correlations. Increasingly positive (negative) correlations are darker shades of red (green)

and following 17 ordinal days were used. KDI were calculated for each pixel on each day in the time series as [(mean-mean_climatology)/standard_deviation_climatology].

In addition to the near-daily KDI used in the modeling analysis below, monthly averaged KDI values for each grid point in the domain were calculated for the purpose of stratifying climate-KDI models and results from over 260,000 pixels into nine more-manageable geographical regions. Following a two-part clustering procedure common in synoptic climatology (Jiang 2011; Lee and Sheridan 2011; Sheridan et al. 2013; Sheridan and Lee 2014), we used a t-mode decomposition of the data (grid points as rows and months as columns, with each cell, therefore representing a monthly KDI value for a grid point) to perform an initial principal components analysis (PCA) on the MODIS satellite data (2002–2013), replacing any missing values with the mean. The 31 principal components (PCs) with eigenvalues greater than one (accounting for 67.3 % of the variance in the data set) were retained for use in a subsequent cluster analysis, using the Two-Step Clustering component in SPSS Statistical Software (SPSS, 2001). Multiple realizations of the cluster analysis were undertaken with varying numbers of clusters/regions, with nine regions ultimately being chosen, as it represented the most spatially cohesive and intuitive realization, with the regions broadly corresponding to adjacent drainage basin domains. These spatially coherent regions, based upon observed patterns in KDI variance within the domain, formed the basis for the regional boundaries demarcated in Fig. 1. Note that while all optically shallow waters (those processed using the modified Kd_lee; Barnes et al. 2013) clustered within region 8, the boundary used for switching between the standard and modified K_d algorithms is not the same as the boundary of region 8.

Daily regional KDI were then calculated for each region based on the mean value of all available pixels, generally well below 100 % due to clouds and satellite coverage. Through statistical trials, it was determined that if at least 5 % of pixels in a region were available on a given day, the mean KDI of that subset was not statistically different from the overall mean. Thus, prior to further analysis, all days on which less than 5 % of the pixels were available in a region (eliminating about 13 % of the days with some pixels), were eliminated (marked as

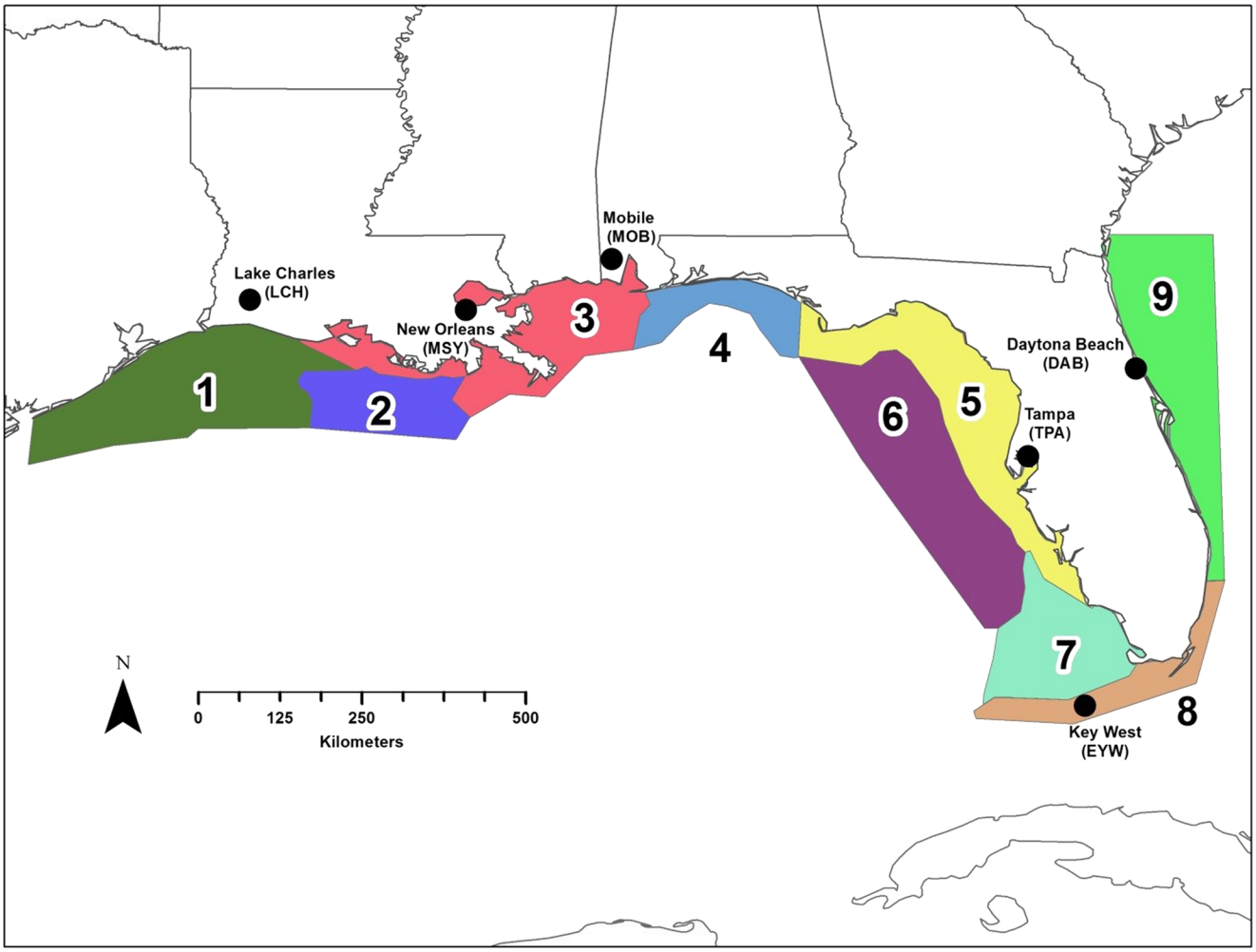

Fig. 1 The nine KDI regions and corresponding SSC weather stations

missing), leaving total temporal coverage of the KDI at about 67 % of days (the other 20 % are days which had no satellite data). As cloud cover reduces satellite data availability and is affected by atmospheric conditions, the circulation patterns and weather types (discussed below) associated with clearer weather conditions have a greater relative sample size of KDI.

2.2 Circulation pattern data and classification methodology

In order to classify atmospheric circulation patterns (CPs), daily gridded mean sea-level pressure (SLP) data were obtained from the NCEP/NCAR Reanalysis 1 project (Kalnay et al. 1996) for the spatial domain bounded by 20° to 40° N and 100° to 60° W, at a resolution of 2.5°× 2.5°, for all days from 1 January 1948 through 23 August 2013. These SLP data were then converted into spatial anomalies—subtracting the daily, domain-wide mean SLP from the SLP for each individual grid point on that day—thus, creating fields of SLP gradients that depict overall flow direction and magnitude within the domain.

Following the methodology detailed by Sheridan et al. (2013) and similar to the regionalization process described in the section above, these spatial anomalies of SLP (more simply referred to as SLP hereafter) were then subjected to a two-part clustering procedure whereby an initial s-mode-based (days as rows, grid points as columns) PCA was undertaken to reduce data dimensionality and yield orthogonal principal component scores (PCs) necessary for clustering. The 10 PCs with eigenvalues greater than one (accounting for 96.1 % of the variance in the data set) were retained for the second part of the classification, whereby the Two-Step Clustering component was employed, effectively classifying each day in the data set into one of 10 circulation patterns based upon the similarity of the shape of daily SLP-gradient patterns in the domain (Fig. 2).

2.3 Other climate data

In addition to CPs, to incorporate the smaller-scale weather conditions interacting with the ocean surface, two other sources of climate data were used to help model water clarity

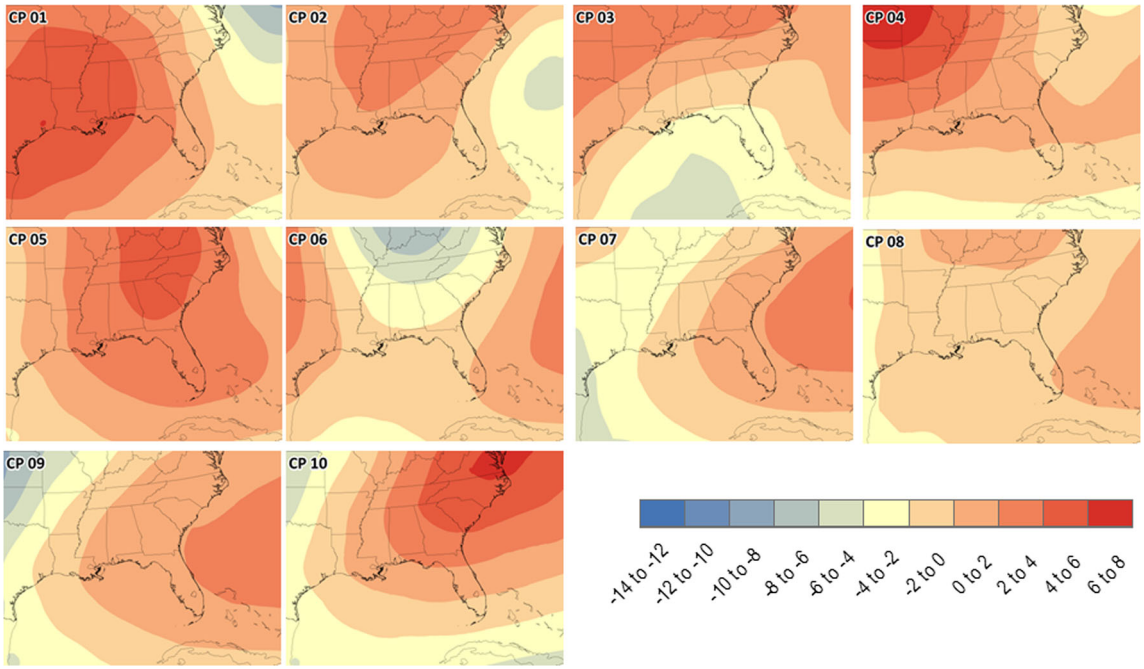

Fig. 2 Composite averaged anomalous SLP for all days classified into each of the 10 circulation patterns (*CPs*)

in the region: surface weather types and precipitation. Daily calendars of surface weather type data were obtained from the Spatial Synoptic Classification (SSC; Sheridan 2002) homepage (http://sheridan.geog.kent.edu/ssc.html) for the six weather stations labeled in Fig. 1. The SSC uses four-times daily values of six different near-surface meteorological variables in order to classify each day at a location into one of several different multivariate weather types (WTs): DM (dry moderate), DP (dry polar), DT (dry tropical), MM (moist moderate), MP (moist polar), MT (+, ++) (moist tropical, (plus, double plus)); TR (transition) signals a change between two weather types. Since precipitation is not directly accounted for in either CP or WT data, daily precipitation amounts (PRCP) were obtained from the NOAA National Climatic Data Center (now National Centers for Environmental Information) for each of the same six weather stations from which SSC data were obtained. In addition to raw daily PRCP, a binary variable was created to demarcate days on which PRCP ≥ 20 mm as heavy precipitation events (HPEs) for each location. All WT, PRCP, and HPE data are obtained for the same 1948–2013 period as outlined above.

2.4 Potential predictors, missing data treatment, and input variable selection

Potential predictors in the NARX model for each region included dummy variables for each of the 10 CPs and the 10 retained PCs used in the classification of CPs. In addition, each region's potential predictors included PRCP, HPEs, and dummy variables for each of the SSC WTs for the station corresponding to each region (Fig. 1). Missing precipitation data (and HPEs) for Key West Airport (EYW) for 1953–1957 were substituted with precipitation from the nearby Key West City Bureau data record and for 3 days in September 1998 with data from Duck Key. Missing SSC data for EYW from 1948 to 1960 was filled with SSC data from Miami (MIA). Missing precipitation data for New Orleans (MSY) on 2 May 1998 was set to 0 mm based upon inspection of nearby station data.

Input variable selection (IVS) for each region was conducted using partial Spearman correlations (a variant on one of many possible IVS procedures suggested by May et al. 2011) of the 5-day-prior moving averages between each of the potential predictor variables and a "filled KDI." Filled KDI was computed to replace any missing actual KDI values using simple linear interpolation between days available in the actual KDI record for each region. Potential predictor variables for each region that showed significant ($p < 0.001$) partial correlation with the filled KDI were selected as the final predictors in the model for each region (see supplementary material Table S1). Dummy variables for month are also included in each region's model to incorporate any important seasonal interactions between predictors and KDI. As is standard when using NARX models, each potential predictor variable is then normalized throughout the 1948–2013 period to fit within the range of −1 to +1 (using the "mapminmax" function in Matlab).

2.5 NARX model set-up and decisions

NARX models are time series models that incorporate the autoregressive nature of KDI and that of the climate

variables in order to model the non-linear relationship between predictors and predictand. NARX models use an iterative training process whereby weights assigned to input variables and bias terms are iteratively adjusted to improve model performance at each step. The entire NARX model development process used herein was completed using functions in the Matlab Neural Network Toolbox (version R2013b) and is outlined in Fig. 3. The Levenberg-Marquardt backpropagation algorithm was used for optimization during training, as it is among the fastest backpropagation algorithms for feedforward networks (Mathworks 2014; Hagan and Menhaj 1994; Beale et al. 2014), especially in Matlab (Mathworks 2014). All training of the NARX models in each region was done on time-blocks of KDI and climate data from the satellite record only (i.e., since September 1997). Once fully trained, these models were then *run* (using the finalized weight and bias terms learned during training) for the entire 1948–2013 time series.

Two user-defined parameters must be determined prior to training each NARX model: the number of neurons in the hidden layer (h) and the number of delays included in the model (d). The former parameter indicates the degree of complexity to be incorporated into the modeling framework, while the latter simply indicates the amount of lag (in days) that is to be incorporated into the time series model for predictors and predictand. However, the ideal settings for these parameters are not evident without experimentation. Therefore, a model with each of the 16 possible combinations of $h = 1$ to $h = 4$ and $d = 1$ to $d = 4$ was trained. Setting the maximum of both parameters (h and d) to 4 was one of several measures taken to prevent overfitting, as each additional parameter (i.e., delays, neurons, input variables) adds complexity, and the closer that the total number of effective parameters is to the sample size of the training data set, the more likely the model will be overfitted (Maier and Dandy 2000, 2001). Since training proceeds using a random initialization (leading to slightly different outcomes with each model), each of these 16 models were trained using multiple (10) permutations (p), which will be used later for ensemble averaging.

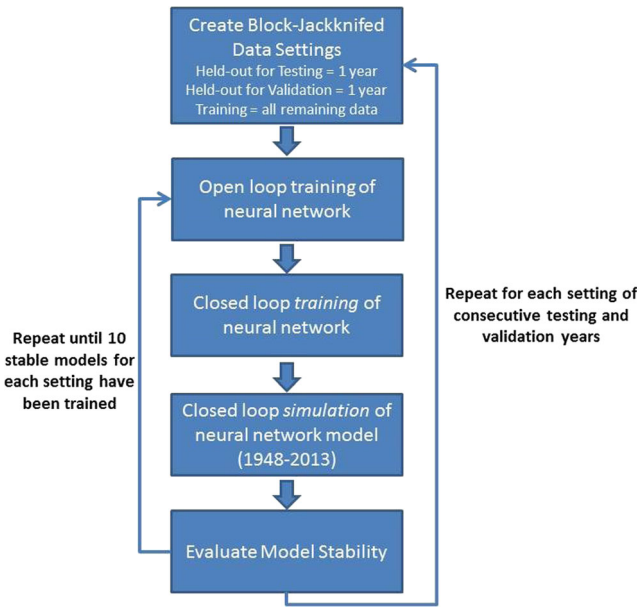

Fig. 3 Flow chart depicting the NARX methodology that is repeated for each combination of number of delays, number of neurons, and each region

Prior to training each model, all time series were divided into three separate "time-blocks" of data, representing the training-, validation-, and testing-blocks of the time series. Training-block data were used for the "learning" part of neural network modeling where weights and biases of the neural network were incrementally adjusted during each epoch to improve (decrease) the mean squared error (MSE) of the training portion of the data. After each adjustment of the model's weights and biases, the updated model was then run on the validation-block of the data set. If, after 20 epochs of training and validation, the MSE of the model run on the validation-block of the data failed to improve (decrease), then the model with the lowest MSE (i.e., the model trained 20 epochs ago) was selected as the best model and training stopped. If the MSE of the validation data set did improve, then training continued, and this iterative process continued in this manner until this qualification was met (i.e., there was no better model in 20). Incorporating an "internal validation" step in this manner prevents overfitting of the model onto the training-block of the data set (Beale et al. 2014), since it was also the best model in 20 with the internal validation-block of the data set. Further, the testing-block of the data was held out from training (and internal validation), and thus can be considered completely independent from the training process, and model performance (e.g., Table 2 and section 2.6) is quantified only using these testing-blocks of the data set.

Since interannual variability impacted model performance, each model was also run using 32 different settings for the partitioning of the data into the time blocks of training, validation, and testing data sets, whereby training was completed using all but two different year-long blocks of the data set, with 1 year held out each for validation and testing purposes. Thus, for each region, 5120 different models were trained, one for each combination of neurons, delays, time-block division settings, and permutations, equating to over 46,000 models across the entire nine-region domain.

Each of these models was first trained using an "open-loop" format on data from 1997 to 2013 (covering the availability of SeaWiFS and MODIS data); meaning, actual KDI values at previous lags (delays) are used during the training process. While this would be ideal for applications

Table 2 Model evaluation statistics based upon ensemble testing data time series (1998–2012) for each region

Month	R1	R2	R3	R4	R5	R6	R7	R8	R9
Jan	0.578	0.512	0.456	0.304	0.758	0.451	0.673	0.472	0.432
Feb	0.531	0.483	0.401	0.353	0.733	0.451	0.647	0.441	0.396
Mar	0.515	0.491	0.373	0.370	0.695	0.454	0.646	0.361	0.407
Apr	0.472	0.401	0.294	0.329	0.556	0.396	0.567	0.340	0.389
May	0.358	0.295	0.226	0.323	0.413	0.351	0.431	0.334	0.395
Jun	0.257	0.157	0.157	0.312	0.227	0.288	0.265	0.296	0.291
Jul	0.281	0.133	0.166	0.195	0.247	0.280	0.289	0.351	0.281
Aug	0.460	0.266	0.259	0.099	0.276	0.184	0.348	0.386	0.399
Sep	0.524	0.386	0.356	0.134	0.438	0.252	0.467	0.386	0.402
Oct	0.596	0.498	0.454	0.329	0.616	0.417	0.588	0.424	0.398
Nov	0.629	0.540	0.492	0.371	0.721	0.490	0.680	0.470	0.381
Dec	0.624	0.526	0.492	0.340	0.745	0.500	0.727	0.471	0.403
All	0.470	0.414	0.333	0.403	0.563	0.341	0.531	0.377	0.353
Best year	0.591	0.621	0.490	0.505	0.723	0.633	0.659	0.528	0.600
Worst year	0.381	0.190	0.166	0.137	0.427	0.155	0.326	0.272	0.168
Hit rate	45%	44%	38%	46%	50%	41%	50%	37%	38%
MdAE	0.326	0.381	0.348	0.391	0.292	0.375	0.262	0.264	0.345
MAE	0.426	0.479	0.416	0.494	0.404	0.488	0.408	0.385	0.455

Hit rates are the percent of extreme KDI events (days on which observed KDI and modeled KDI > 80th percentile of their respective time series) predicted by the model. Note that monthly and best/worst year correlations are based upon 89-day centered moving correlations averaged for each month/year. Overall correlations are traditional. Cells with increasingly darker gray coloring signify increasingly stronger correlations

MdAE median absolute error, *MAE* mean absolute error

where previous values of KDI are available, the goal of the research herein was to extend the KDI back to 1948 using climate as the driver. However, reliable estimates of KDI were only available back to the beginning of the SeaWiFS time series (September 1997), and thus a key component of the model—the previous days' values of KDI—would be missing from 1948 to 1997. Thus, after training was completed on open-loop, each model was then trained using a "closed-loop" format initialized from the open-loop model's final state. A closed-loop format of the model feeds back the previous days' values of modeled KDI into the model rather than the actual KDI during the training process in order to incorporate the autoregressive nature of KDI. While the closed-loop training could have been completed without first using the open-loop format, the selected method resulted in a reduction of computational time by a factor of about 60 in preliminary analyses, yielding a possible savings of over 68 weeks of modeling time in the final analysis. After training on closed-loop over the 1997–2013 portion of the time series, each fully trained model was then run on closed-loop for the entire 1948–2013 time period. Since closed-loop simulation would fail if there were any missing predictor variables, any missing SSC-WT days (only 0.6 % of all days/locations) were set to 0 for each WT's dummy variable on that day.

Preliminary analyses revealed that a small percentage of individual models' KDI output varied around an implausibly high or low KDI value over the course of years to decades. Thus, once closed-loop simulation was complete, any model with an average annual KDI > 1 or KDI < −1 for any calendar year between 1948 and 2013 was discarded and another was fully trained through all of the aforementioned steps.

Once all 5120 models were fully trained, the ideal number of neurons and delays for modeling KDI in each region were selected (i.e., a winning model architecture; Table S2 in supplementary material). The winning model architecture was determined by averaging the MSE of the different testing-block portions of the 320 (10 permutations × 32 settings) closed-loop trained models for each of the 16 possible model architectures (h = 1:4 and d = 1:4), with the lowest averaged MSE of the 16 determining the winner. This winning set of 320 models was then averaged on a day-by-day basis (ensemble mean) and used as the final model-ensemble time series and the reconstructed KDI from 1948 to 2013 for each region.

2.6 Independent model validation using a block-jackknifing technique

In addition to providing a temporally robust determination of the winning model architecture, the block division of

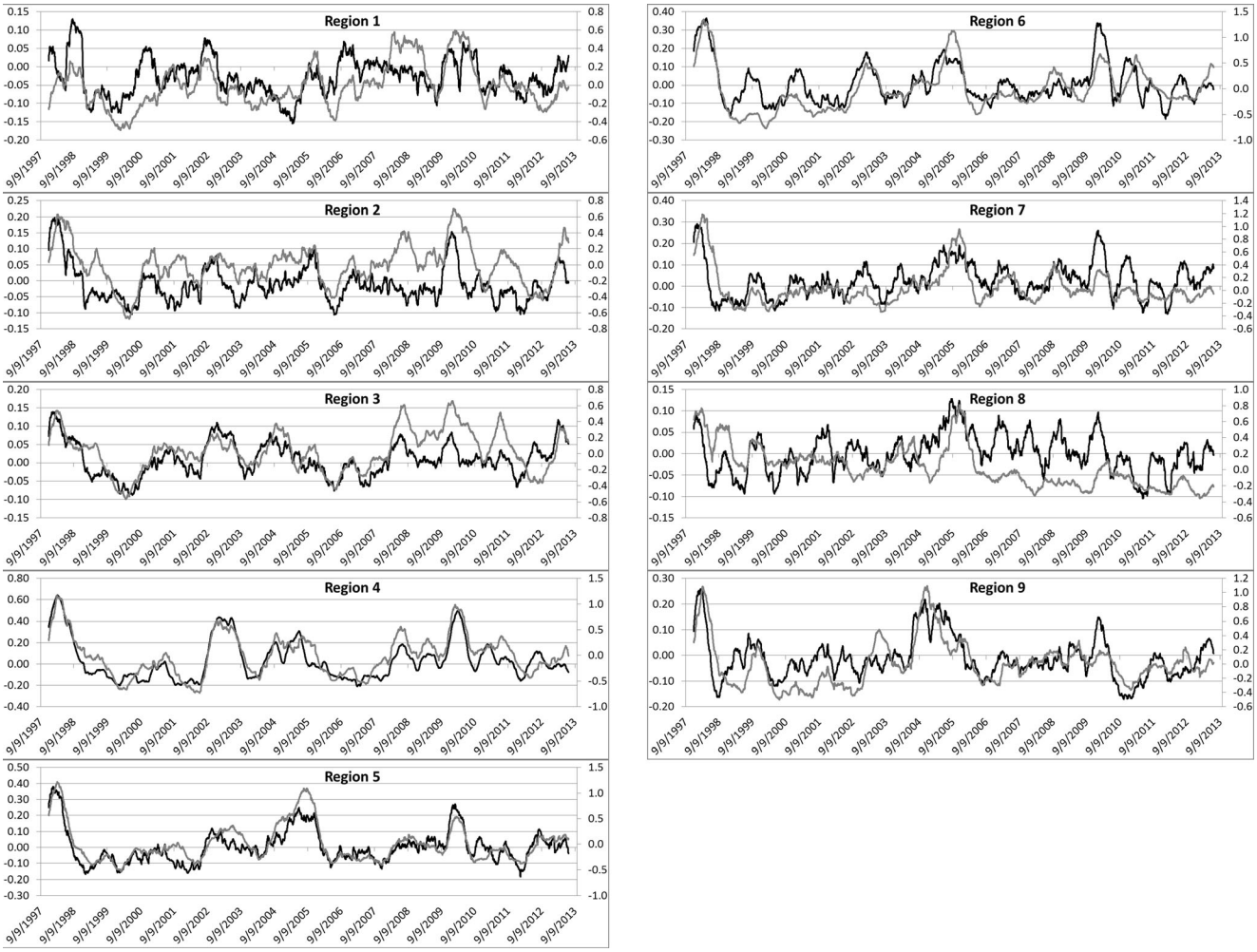

Fig. 4 Final model ensemble time series of the 181-day centered moving average actual (*gray line*, *right axis*) and modeled (*black line*, *left axis*) KDI for each region, September 1997–September 2013

the data into 32 different time-block division settings also allowed for an independent evaluation of ensemble model performance across the entire 1997–2013 time series by using a "block-jackknifing technique." Since each of the 32 settings used one of 16 different year-long periods as a testing-block, each year-long period was used exactly twice specifically as the independent testing-block. Additionally, each setting of the winning model underwent 10 permutations. Together, this yielded 20 separate testing-block representations of each day from 1997 to 2013 that can be considered independent of the model-training process. We chronologically reconstructed the output of these testing-block time periods of each of the 320 closed-loop simulation models into a set of 20 time series that span the entire 1997–2013 period and calculated day-by-day (ensemble) means of these time series (i.e., an independent testing-block time series) and report these results below as well. Important to note is

Table 3 Estimated relative importance of predictor variables in the final ensemble model for each region

	R1 (%)	R2 (%)	R3 (%)	R4 (%)	R5 (%)	R6 (%)	R7 (%)	R8 (%)	R9 (%)	AVG (%)
CPs	0	13	20	5	24	4	3	4	7	10
WTs	7	29	37	60	52	77	71	48	14	46
PRCP	85	39	–	–	–	–	23	36	78	32
HPE	–	–	16	2	–	–	–	–	–	0
PCs	8	20	27	33	24	19	3	13	2	12

Blank cells indicate that the variable was not selected for use in that region's models

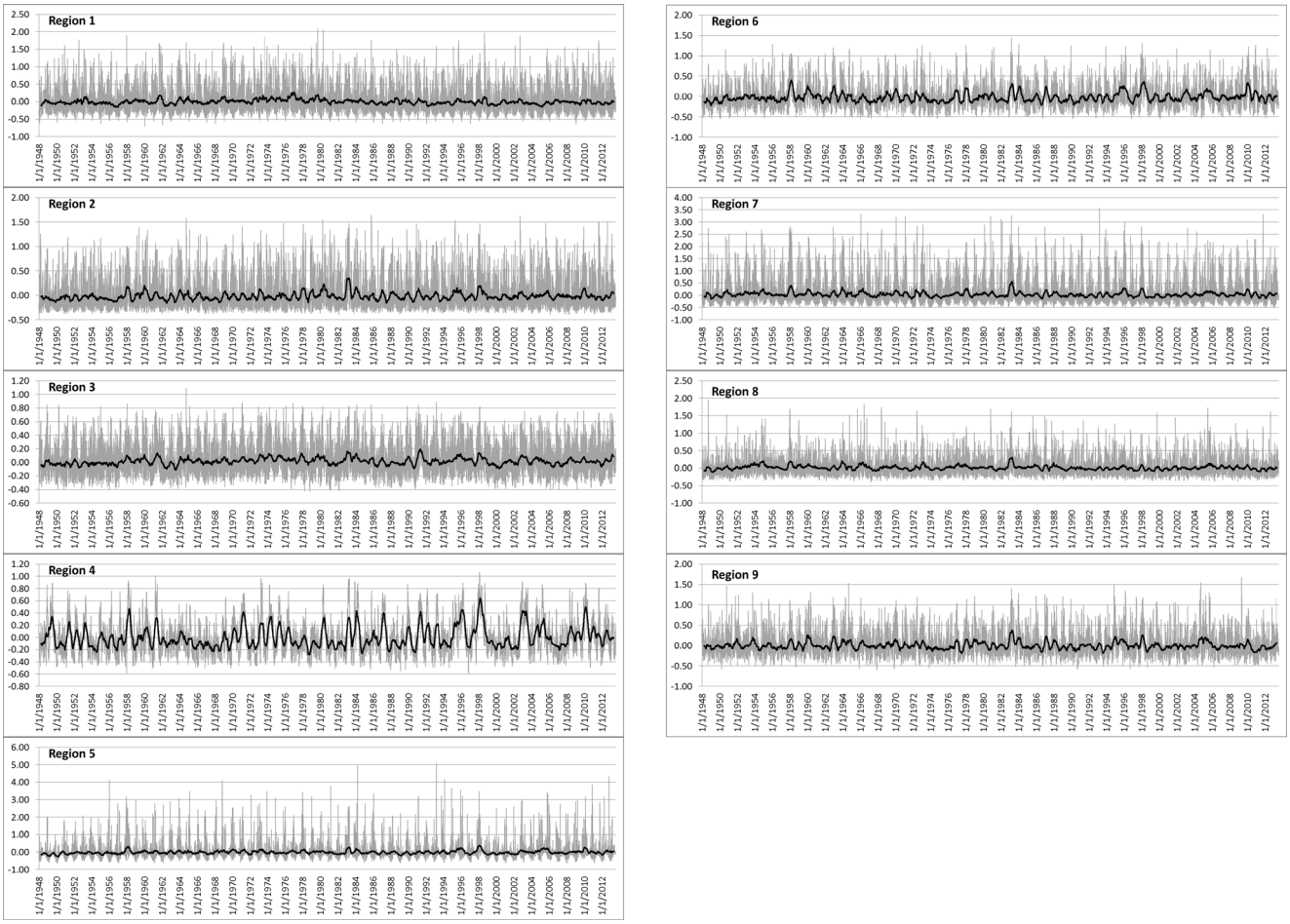

Fig. 5 Reconstructed 66-year time series of daily (*gray line*) and 181-day centered moving averages (*black line*) of the ensemble NARX-modeled KDI for each region, 1948–2013. Note the different scales of the *y*-axis in each graph

that this testing-block time series is different from the final model ensemble time series, as the former was reconstructed from the separate testing-block portions of the data sets, while the latter was made up of the entire time series (i.e., the training-, validation- and testing-block portions of the time series) for each of the 320 model ensemble members. Correlation analyses with the chronologically reconstructed testing-block ("block-jackknifed") time series, therefore, give an estimate of the confidence with which the 66-year historical KDI time series can be modeled.

3 Results and discussion

3.1 Circulation patterns, SSC types, and correlations with KDI

Ten circulation patterns (CPs; Fig. 2) were identified, spanning the range of weather conditions observed across the domain. These include several strong cyclonic patterns: centered in the Ohio River Valley (CP 6), the Gulf of Mexico (CP 3), the Southeastern Coast (CP 2), and the Mid-Atlantic (CP1). Strong continental anticyclones can be observed in CPs 4, 5, and 10; CPs 7, 8, and 9 are all manifestations of the Bermuda High. Correlations between weekly KDI and the circulation patterns show a broad tendency for increased KDI (i.e., poorer water clarity) associated with the cyclonic patterns and decreased KDI for the anticyclonic patterns (Table 1), though considerable spatiotemporal variability exists, with the strongest correlations observed in the winter months. High KDI values could be the result of high chlorophyll-*a* (or colored dissolved organic matter, CDOM) concentrations in association with cyclonic conditions and windy weather that churn up nutrients or increase nutrient load via runoff, a result noted in Sheridan et al. (2013).

Correlations between SSC frequencies and weekly-aggregated KDI also yield statistically significant results across multiple categories. Generally, the WTs associated with more stagnant conditions (MT, MT+, MT++) are

associated with clearer water (Table 1), while rainier patterns (such as MM or MP) are associated with higher KDI values. The TR type has the greatest positive correlations and the most significant results, particularly along the west coast of Florida, unsurprising considering the inherently unstable and windy conditions of the TR weather type would likely churn up benthic sediments in the coastal zone, decreasing water clarity.

3.2 Development and validation of the KDI time series

In comparing correlations between observed KDI and modeled KDI (with the testing-block time series), the model performed considerably better in all seasons except summer (Table 2), with peak ability from November to March. Overall, model performance is best off of the west coast of Florida (Regions 5 and 7) and worst in the waters downstream of the Mississippi River (Region 3), whose length and catchment area results in longer lags and much greater upstream influence, though considerable variability exists spatially as well as interannually. Of particular note is the performance of the NARX models in regions 5 and 7 in winter, where correlations between the model and observed KDI exceed $r = 0.70$ in some months. Time series (e.g., Fig. 4) show the broad agreement between the observed and modeled data; the model tends to render the general patterns well, though underestimates daily variability, particularly in extreme events, as evidenced by the difference between the mean absolute error (MAE) and the median absolute error (MdAE; Table 2). Hit rates (defined as the match percentage between days that were above the 80th percentile of observed KDI and above the 80th percentile of NARX-modeled KDI) were near 50 % for Regions 5 and 7 and between 37 and 46 % for other regions.

Due to the complex and interactive nature of NARX modeling, evaluating the importance of an individual predictor variable in a model is not straightforward (see Olden and Jackson 2002). However, in order to estimate this importance, each of the final 320 ensemble members in each region were re-run leaving a single predictor variable set to a constant, and the change in the performance (MSE) of the model was noted. This process was repeated for each of the predictors in each model, the change in MSEs were averaged across the 320 members in a region and made into a percentage relative to the summed MSE differences of all other predictors. In short, this yields a relative importance of each predictor among all other predictors in a region (Table 3). Among the predictors, SSC WTs play the most prominent role in the regions spanning the Louisiana coast to the Florida Keys. Raw precipitation is most important in the outer regions of the project domain (Texas and East Florida especially), but HPEs figure prominently only in region 3 (the Mississippi River area). The CP classification plays at least a moderate role in model performance in every region but region 1, and its estimated importance is high in region 5, one of the better-modeled areas. It is important to note that in NARX modeling, previous-day (s) values of KDI are also used to model KDI. While the relative importance of lagged KDI cannot be determined using this method, it likely plays a substantial role as well.

3.3 Reconstructing the historical KDI

As noted in the methodology above, all NARX models were run for the full period for which historical atmospheric data are available (since 1948; Fig. 5). Long-term patterns vary across the regions in terms of extreme KDI events (i.e., >80th percentile; Fig. 6). Across the western Gulf, in Regions 1, 2, and 3, extreme KDI events peaked in the 1970s and 1980s. Contrastingly, across the regions east and south of the Florida peninsula (Regions, 7, 8, and 9), there appears to be a slight decreasing trend in the long-term frequency of extreme KDI

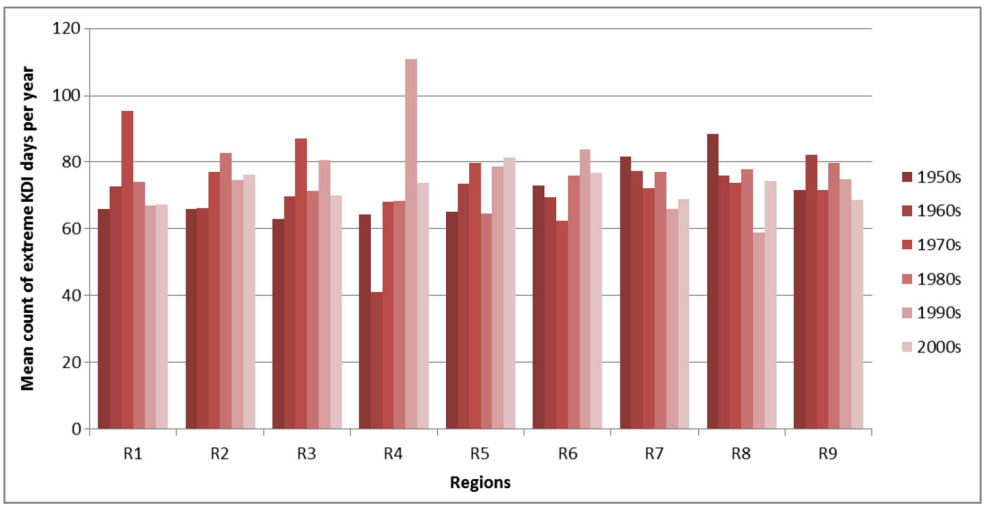

Fig. 6 Decade-by-decade counts of extreme KDI events (>80th percentile) in the NARX-modeled reconstruction time series for each region

Table 4 Region-by-region association between ENSO and wintertime NARX-modeled KDI, 1950–2012

Rank	Year	R1	R2	R3	R4	R5	R6	R7	R8	R9
1	1983	9	1	1	9	4	4	1	1	1
2	1998	14	2	2	1	1	3	3	11	2
3	1992	56	35	37	10	18	28	42	51	30
4	1987	54	9	17	7	11	5	5	7	3
5	1958	26	3	6	6	2	1	2	2	7
Spearman rho		0.170	0.647	0.467	0.335	0.561	0.576	0.578	0.483	0.673

The left column indicates the rank of the top 5 years of the bimonthly Multivariate ENSO Index (MEI; NOAA 2015) as averaged from December–January through March–April of the year indicated in column two. The nine rightmost columns have the corresponding rank of the monthly averaged modeled KDI for each region for the December–April period of the corresponding years (December of the prior year), with a rank of 1 being the highest averaged KDI of the 63 years ranked. Bottom row shows Spearman's rank correlations of the MEI and the monthly modeled KDI values for the months described above, throughout the entire 1950–2012 period

events. Only in Regions 4 and 5, across the West Florida Shelf, is a general upward trend in extreme KDI events observed. This trend may be associated with the increased wintertime frequency of cyclonic CPs (CPs 1, 3, and 6) and TR weather types (at the Tampa weather station), and the secular decline in winter MT weather type occurrences in these locations over the 1948–2012 period, though trends in other CPs and WTs confound these relationships. Notably strong El Niño Southern Oscillation (ENSO) winters (e.g., 1957–58, 1982–83, 1997–98) as defined by the Multivariate ENSO Index (MEI; NOAA 2015) correspond with the greatest sustained, short-term increases in modeled KDI in most regions east of Texas (Table 4). This is especially interesting considering the models' relative strength in winter compared to other seasons, and the fact that no ENSO index was directly incorporated into modeling.

4 Summary and conclusions

This research examines the impact of meteorological variability on a day-to-day water clarity index (KDI) in the coastal waters off the southeastern USA. Analyses of individual variable relationships to the KDI show that cyclonic circulation and transitional atmospheric situations coincide with turbid water, while stagnant, stable conditions lead to greater water clarity. A novel modeling methodology using an ensemble of non-linear autoregressive models with external input (NARX models) allowed for a 66-year reconstruction of the KDI in each region. Overall, surface weather types showed the greatest influence on model performance especially in the regions off the west coast of Florida. While total precipitation figured prominently into the NARX models in over half of the regions, specific heavy precipitation events did not (Table 3). Model performance during periods of actual satellite observation (1997–2013) varies by season, region, and interannually, but generally, models perform best in the winter and for the West Florida Shelf, lending more confidence in the historic reconstructions in these regions and seasons. Analyses of the historical reconstruction revealed higher KDI values during winters with noted ENSO events. While an original premise of the research was to investigate the applicability of water clarity as an indicator of climate change, noticeable upward trends in extreme KDI events were only present in two regions (regions 4 and 5).

Non-climatic factors (e.g., land use and land cover changes) undoubtedly account for a significant portion of KDI variability and thus explain the low correlations for some regions. Of particular note are changes in water management strategies over the latter half of the twentieth century that impact river discharge into the Gulf, especially diluting the effect of precipitation on KDI variability in some locations, for example, the managed diversion of Lake Okeechobee runoff into Florida Bay starting in 1980 (see Fourqurean and Robblee 1999; Barnes et al. 2014). Another potential limitation of the research in finding a consistent climate change signal is the relatively short period-of-record of KDI data, especially relative to the longer-term trends indicative of climate change. The intermittent nature of the data may also limit the ability of the NARX models to incorporate enough lag into the model in each region. This likely leads to precipitation (and HPEs) having a less prominent role in the final models than may be expected, as the lag between precipitation event and KDI response might not be within the lag periods incorporated into the models. Combined, these last two factors (period of record and the intermittent nature of KDI) limited the sample size, and thus the complexity (via the maximum number of neurons) that could be safely incorporated into any region's ensemble-member models. With a greater sample size of KDI, more complex models could be constructed, perhaps yielding better results. Further, cloudy conditions are most often the cause of missing KDI data, and thus, the models trained herein may be slightly biased towards days with clear skies, as would any water clarity study using such data. However, by incorporating thousands of individual pixels into regions as the unit of study (including the clear pixels in an

otherwise cloudy image), the effect of this bias is thought to be fairly limited.

Despite these limitations, the novel NARX modeling methodology and the use of synoptic weather types and circulation patterns showed promising results, especially in vital ecosystem regions along the West Florida Shelf. Future research will aim to further refine these models based upon the availability of new KDI and climate data, incorporate anthropogenic factors into the models where such data are available, and transport this methodology to investigate water clarity in other locations. Further, NARX modeling is easily adapted to make multi-step-ahead predictions using the closed-loop framework; hence, mid- to long-range forecasts of water clarity using forecast weather data is another likely future avenue, as this could provide valuable information to ecosystem managers in these regions.

Acknowledgments This research was supported by the National Aeronautics and Space Administration's (NASA's) Research Opportunities in Space and Environmental Sciences (ROSES) funding opportunity, Development and Testing of Potential Indicators for the National Climate Assessment, Award NNX13AN31G. The authors would like to thank Dr. Michael J. Allen from the Department of Political Science and Geography at Old Dominion University for his contributions to the early portions of this project.

References

Ault, D. N. (2006) Temporal and spatial distribution of chlorophyll on the West Florida Shelf. University of South Florida Theses and Dissertations. Paper 3864

Barnes BB, Hu C (2015) Cross-sensor continuity of satellite-derived water clarity in the Gulf of Mexico : insights into temporal aliasing and implications for long-term water clarity assessment. IEEE Trans Geosci Remote Sens 53:1761–1772

Barnes BB, Hu C, Schaeffer BA, Lee Z, Palandro DA, Lehrter JC (2013) MODIS-derived spatiotemporal water clarity patterns in optically shallow Florida keys waters: a new approach to remove bottom contamination. Remote Sens Environ 134:377–391

Barnes BB, Hu C, Holekamp KL, Blonski S, Spiering BA, Palandro D, Lapointe B (2014) Use of Landsat data to track historical water quality changes in Florida keys marine environments. Remote Sens Environ 140:485–496

Barnes BB, Hallock P, Hu C, Muller-Karger F, Palandro DA, Walter C, Zepp R (2015) Prediction of coral bleaching in the Florida keys using remotely sensed data. Coral Reefs 34:491–503

Beale MH, Hagan MT, Demuth HB (2014) Neural Network ToolboxTM User's Guide, R2014a. White Paper. The Mathworks Inc.

Causey BD (2002) The role of the Florida keys National Marine Sanctuary in the South Florida ecosystem restoration initiative, pp. 883–894. In: Porter JW, Porter KG (eds) The Everglades, FloridaBay, and coral reefs of the Florida keys, an ecosystem source book. CRC Press, Boca Raton

Conmy, R.N. (2008) Temporal and spatial patterns in optical properties of colored dissolved organic matter on Florida's gulf coast: shelf to stream to aquifer. University of South Florida Theses and Dissertations, Paper 187

Diaconescu E (2008) The use of NARX neural networks to predict chaotic time series. WSEAS transactions on computer. Research 3(3):182–191

Fourqurean JW, Robblee MB (1999) Florida Bay: a history of recent ecological changes. Estuaries 22:345–357

Greening HS, Cross LM, Sherwood ET (2011) A multiscale approach to seagrass recovery in Tampa Bay, Florida. Ecol Restor 29:82–93. doi:10.3368/er.29.1-2.82

Greening H, Janicki A, Sherwood ET, Pribble R, Johansson JOR (2014) Ecosystem responses to long-term nutrient management in an urban estuary: Tampa Bay, Florida, USA. Estuar Coast Shelf Sci 151:A1–A16. doi:10.1016/j.ecss.2014.10.003

Hagan MT, Menhaj M (1994) Training feed-forward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5(6): 989–993

Hu C, Muller-Karger FE (2007) Response of sea surface properties to hurricane Dennis in the eastern Gulf of Mexico. Geophys Res Lett 34:L07606. doi:10.1029/2006GL028935

Janicki, A., Wade, D., & Pribble, R. J. (2000) Developing & establishing a process to track the status of chlorophyll—a concentrations and light attenuation to support seagrass restoration goals in Tampa Bay. Tampa Bay Estuary Program Technical Report # 04–00

Jiang N (2011) A new objective procedure for classifying New Zealand synoptic weather types during 1958–2008. Int J Climatol 31(6): 863–879

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–471

Lee CC, Sheridan SC (2011) A six-step approach to developing future synoptic classifications based on GCM output. Int J Climatol 32(12):1792–1802

Lee Z, K-P D, Arnone RA (2005) A model for the diffuse attenuation coefficient of downwelling irradiance. J Geophys Res 110:C02016

Lee Z, Lubac B, Werdell PJ, Arnone RA (2009) An update of the quasi-analytical algorithm (QAA_v5). International Ocean Color Group Software Report. Available from: http://www.ioccg.org/groups/Software_OCA/QAA_v5.pdf

Lirman D, Schopmeyer S, Manzello D, et al. (2011) Severe 2010 cold-water event caused unprecedented mortality to corals of the Florida reef tract and reversed previous survivorship patterns. PLoS One 6: e23047. doi:10.1371/journal.pone.0023047

Liu Y, Weisberg RH (2005) Patterns of ocean current variability on the West Florida shelf using the self-organizing map. J Geophys Res 110:C06003. doi:10.1029/2004JC002786

Lohrenz SE, Fahnenstiel GL, Redalje DG, Lang GA, Dagg MJ, Whitledge TE, Dortch Q (1999) Nutrients, irradiance, and mixing as factors regulating primary production in waters impacted by the Mississippi River plume. Cont Shelf Res 19:1113–1141

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ Model Softw 15:101–124

Maier HR, Dandy GC (2001) Neural network based modelling of environmental variables: a systematic approach. Math Comput Model 33:669–682

Mathworks (2014) www.mathworks.com/help/nnet/ref/trainlm.html?searchHighlight=trainlm as accessed on 4 December 2014

May R, Dandy G, Maier H (2011) Review of input variable selection methods for artificial neural networks, Artificial Neural Networks-Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978–953–307-243-2, InTech. Available from: http://www.intechopen.com/books/artificialneural-networks-methodological-advances-and-biomedical-applications/review-of-input-variable-selectionmethods-for-artificial-neural-networks

NOAA (2015) NOAA's Earth Systems Research Laboratory's Bimonthly MEI Index. http://www.esrl.noaa.gov/psd/enso/mei/table.html as retrieved on 7 December 2015

NOAA National Centers for Environmental Information (2015) Billion Dollar Weather and Climate Disasters. https://www.ncdc.noaa.gov/billions/mapping

Olden JD, Jackson DA (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol Model 154:135–150

Patt FS, Barnes RA, Eplee RE, Franz BA, Robinson WD, Feldman GC, Bailey SW, Gales J, Werdell PJ, Wang M, Frouin R, Stumpf RP, Arnone RA, Gould, R. W. J, Martinolich PM, Ransibrahmanakul V, O'Reilly JE, Yoder JA (2003) Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, NASA Tech Memo 2003–206892, Volume 22. In: Hooker SB, Firestone ER (eds) SeaWiFS Postlaunch Technical Report Series

Pirhalla, D.E., Sheridan, S.C., Ransibrahmanakul, V., Lee, C.C. (2014) Assessing cold-snap and mortality events in South Florida coastal ecosystems: development of a biological cold stress index using satellite SST and weather pattern forcing. Estuaries and Coasts (Accepted, in press). DOI: 10.1007/s12237–014-9918-y

Sheridan SC (2002) The redevelopment of a weather-type classification scheme for North America. Int J Climatol 22:51–68

Sheridan SC, Lee CC (2014) Synoptic climatology. In: Warf B (ed) Oxford bibliographies in geography. Oxford University Press, New York City, USA. doi:10.1093/OBO/9780199874002-0088

Sheridan SC, Pirhalla DE, Lee CC, Ransibrahmanakul V (2013) Evaluating linkages of weather patterns and water quality responses in South Florida using a synoptic climatological approach. J Appl Meteorol Climatol 52(2):425–438. doi:10.1175/JAMC-D-12-0126.1

Short FT, Wyllie-Escheverria S (1996) Natural and human-induced disturbance of seagrasses. Environ Conserv 23(1):17–27

SPSS (2001) The SPSS two-step cluster component: a scalable component to segment your customers more effectively. White paper, technical report, SPSS, Chicago, Illinois, USA