Estimation of Hourly Salinity Concentrations Using an Artificial Neural Network

Vladimir J. Alarcon[1(✉)], Anna C. Linhoss[2], Christopher R. Kelble[3], Paul F. Mickle[4], Joseph Bishop[3], and Emily Milton[3]

[1] Civil Engineering Department, Universidad Diego Portales, 441 Ejercito Av., Santiago, Chile
vladimir.alarcon@udp.cl

[2] Department of Biosystems Engineering, Auburn University, 3101 Shelby Center, Auburn, AL 36849, USA
alinhoss@auburn.edu

[3] NOAA Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, USA
{chris.kelble,joe.bishop,emily.milton}@noaa.gov

[4] Northern Gulf Institute, Mississippi State University, 2 Research Blvd., Starkville, MS 39759, USA

Abstract. Estimating salinity concentrations in coastal waters allows characterization of the spatial and temporal dynamics of the freshwater/saltwater interface. In Southeast Florida (USA) the saltwater interface is monitored and evaluated for potential impacts to public supply wellfields and biological communities. In this research, a closed-loop autoregressive neural network with exogenous inputs was developed to estimate salinity concentrations at a coastal water quality station (BISCC4) in Biscayne Bay, Florida. The neural network (ANN) is shown to successfully simulate hourly salinity concentrations for years 2015 through 2019. A statistical comparison of simulated concentrations versus observed data demonstrates that the ANN simulates salinity concentration values and trends within acceptable margin of errors ($R^2 = 0.59$, K-G = 0.64, NSE = 0.33, d = 0.86, PBIAS = 1.5%, RSR = 0.82). In its current form, the ANN model performs better in simulating salinity concentrations and trends, than an existing hydrodynamic model. These results have the potential to be applied to other coastal locations in Biscayne Bay where freshwater inputs from inland streams and canals are affecting salinity concentrations.

Keywords: Biscayne Bay · Salinity · Neural network · Southeast Florida · Prediction

1 Introduction

Salinity is a critical response variable that integrates hydrologic and coastal dynamics including sea level, tides, winds, precipitation, and streamflow [1]. The position and temporal dynamics of the freshwater–saltwater interface for larger coastal rivers results from the interaction of three principal forces: streamflow, mean tidal water levels, and

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12954, pp. 629–640, 2021.
https://doi.org/10.1007/978-3-030-86979-3_44

tidal range [2]. The position of the freshwater interface determines the composition of freshwater and saltwater aquatic communities as well as the freshwater availability for water intakes [1]. Estimating salinity concentrations in coastal waters allows characterizing the spatial location of the freshwater/saltwater interface which has implications for biological communities and for urban settlements. For example, in Southeast Florida (USA) the saltwater interface is mapped as part of the water use permitting review process and to monitor for potential impacts to public supply wellfields [3].

Salinity in estuarine systems is influenced by climate and weather. The climate in Southeast Florida is characterized by hot and humid rainy summer, and mild winters. The annual average minimum, mean, and maximum temperatures in the area are 15 °C, 26 °C, and 32 °C, respectively, and the total annual precipitation average is 1507 mm. The total annual average potential evapotranspiration (ETP) loss in the area ranges from 1220 mm to 1320 mm per year [4]. Stalker et al. [5] estimated a freshwater input ratio canal/precipitation/groundwater of 37%:53%:10% in the wet season, and 40%:55%:5% in the dry season, with an error of ±25%. Approximately 75% of the total precipitation falls in the wet season, generating substantial run-off. Furthermore, discharge from regional groundwater circulation and controlled release from upstream canals provide a considerable volume of freshwater input to the water balance in the area [4]. Moreover, Biscayne Bay drains the Everglades (an extensive wetland located in Southeast Florida). This wetland is critical for water storage and recharge of the Biscayne aquifer, and is the principal source of freshwater for South Florida [6].

A recent modeling study on the hydrodynamic regime and salinity spatial distribution in Biscayne Bay (Southeast Florida, USA) was able to replicate salinity concentration values and seasonal trends occurring within the bay [7]. In this study, salinity estimations were poorest for locations close to the coast, especially during the wet season. A non-linear effect produced by freshwater inputs, which cannot be quantitatively described by mechanistic approaches, seems to be substantially affecting salinity concentrations.

The artificial neural network (ANN) method has been successfully applied to model non-linear processes. Special configurations of ANN have been successfully used to predict flow and sediment concentration time-series for rivers around the world [8, 9]. The most common models are artificial neural networks with dynamic architecture (DANN), autoregressive neural networks (ARNN), non-linear autoregressive exogeneous neural networks (NARX), and hybrid configurations that combine ANN with other modeling approaches. There are various ANN configurations that reported using artificial neural networks to predict salinity. Shahid et al. [10] used feed-forward back propagation ANN models to estimate salinity in the Grant Line Canal (California) using electrical conductivity, temperature, and pH as data inputs. Lin et al. [11] used an Elman neural network (ENN) to predict salinity intrusion at the Pearl River Delta. Zhou et al. [12] developed a hybrid model coupling the techniques of wavelet transform (WT) and artificial neural network (ANN) for forecasting estuarine salinity in the Pearl River Estuary, China. The WT-ANN was able to forecast 1 to 3 days of salinity concentrations. Rath et al. [13] developed an empirical model of salinity in the San Francisco Bay-Delta estuary using a Bayesian artificial neural network (ANN) model.

However, in most of these previous ANN applications the validation period of the model output corresponded to within-year or seasonal events. There are no ANN model

applications that simulate continuous hourly salinity estimations at multiyear temporal scales.

This research presents an autoregressive neural network (with exogenous inputs) developed to forecast salinity concentrations in the coastal waters at Biscayne Bay (Florida, USA). Details are presented on input data selection, model calibration/training (optimization of model parameters), and model validation. Moreover, the developed ANN is assessed to determine if it can be used for forecasting hourly salinity concentrations for 5 continuous years.

2 Methods

2.1 Study Area

Biscayne Bay is a coastal water body located in southeast Florida (adjacent to the Miami metropolitan area) covering an area of approximately 700 km^2 (Fig. 1). The water depth in the bay ranges between 1.8 m and 4 m, except in dredged areas where depths can exceed 12 m [14]. The Biscayne Bay area includes the largest marine national park in the U.S. national park system: Biscayne National Park, an important part of the recreational, social, economic, and cultural life of Southeast Florida [15].

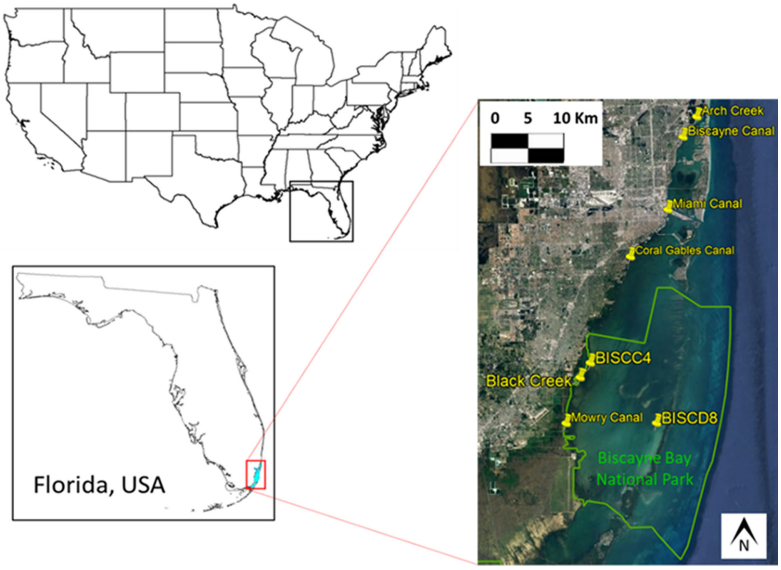

Fig. 1. Study area. Biscayne Bay and location of main streams draining to the bay.

Figure 1 shows the main canals that drain waters into Biscayne Bay including: Arch Creek, Biscayne Canal, Little River, Miami Canal, Coral Gables Canal, Snapper Creek, Black Creek, Princeton Canal, and Mowry Canal. Over a century of water management had led to no natural rivers or creeks draining into Biscayne Bay. Two water quality stations that collect salinity concentrations (BISCD8 and BISCC4) are also shown. Hourly salinity concentration data collected at those stations were used in this research.

2.2 Neural Network Development

The Nonlinear Autoregressive Exogenous Neural Network Model is a dynamical neural architecture commonly used for nonlinear dynamical systems. When applied to time-series prediction, this network configuration is designed as a feedforward time-delayed neural network, provided with feedback connections which enclose at least two layers of the network, and can use past values of predicted or true (measured) time-series [16]. In this research, salinity is forecasted using a nonlinear autoregressive exogenous neural network. Details on model development are presented in the following sections.

Input Data

The neural network was developed using hourly salinity from BISCD8 and BISCC4 water quality stations, and hourly flow from a gauge station located near the outfall of Black Creek to Biscayne Bay. Figure 2 illustrates the relationship among the observed data at those water quality stations.

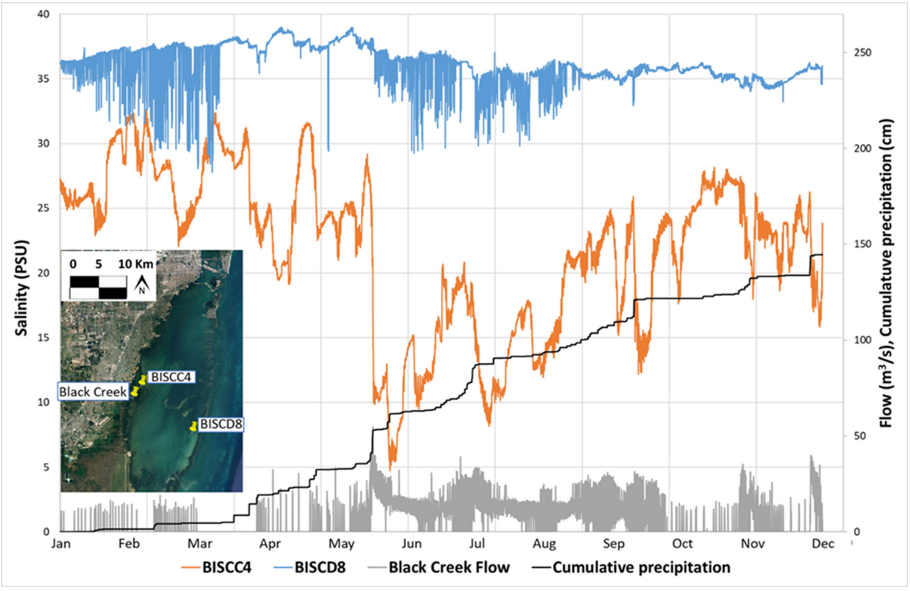

Fig. 2. Hourly salinity, flow, and precipitation during year 2013.

Figure 2 shows hourly salinity (BISCC4 and BISCD8 stations) and flow data (Black Creek) for year 2013. Cumulative precipitation in the Biscayne Area is also shown. The difference between salinity concentrations at BISCC4 and BISCD8 is striking. While BISCD8, which is closer to the ocean, recorded salinity values between 30 and 38 PSU, observed salinity at the coastal station (BISCC4) ranges between 5 and 35 PSU. The influence of freshwater inputs from rivers and canals (e.g., Black Creek, the closest stream to BISCC4), and precipitation, is evident because some of the lowest salinity values at BISCC4 occur after storm events and high flows observed at Black Creek.

However, there are decreases in salinity that do not co-occur with precipitation or flow events. At some degree, those minimum concentrations guard a relationship to salinities observed at BISCD8. This lack of flow and precipitation during periods when salinity decreases significantly suggests groundwater inputs may also play an important role in the salinity trends observed at BISCC4.

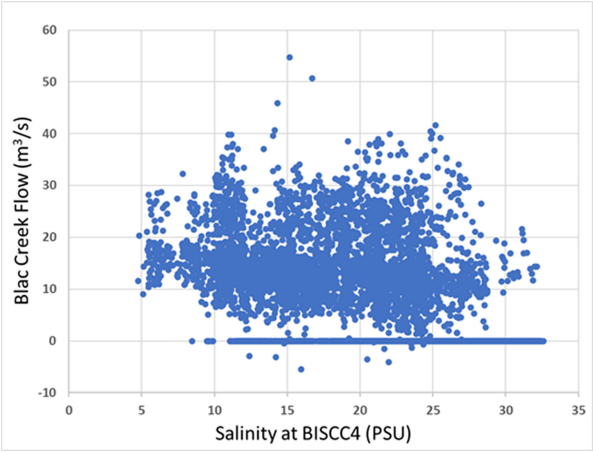

Fig. 3. Non-linear relationship between Black Creek flow and Salinity at BISCC4. Low salinity concentrations at BISCC4 are observed either for low or high flows at Black Creek.

Having identified the non-linear trend in salinity concentrations at BISCC4 (Fig. 3), an artificial neural network (ANN) was developed to capture this non-linear trend.

Neural Network Training

Observed hourly data (salinity at BISCC4 and BISCD8, and flow data at Black Creek), for years 2012 through 2019 were used for the development of the ANN. In order to explore the effect of training data size to achieve a successful prediction of salinity at station BISCC4, the number of years for training and prediction was varied. Table 1 details the process.

Table 1. Training data size

Model Rrn	Number of years used for training		
	Salinity BISCD8	Salinity BISCC4	Black Creek Flow
1	2012–2014	2012–2014	2012–2014
2	2012–2013	2012–2013	2012–2013
Final	2013–2014	2013–2014	2013–2014

Initially, observed hourly salinity data (collected at BISCC4 and BISCD8 stations), and flow data collected at Black Creek, from years 2012 through 2014, were used for training the neural net. Several combinations of yearly datasets were tested until optimal statistical indicators of fit between simulated and observed daily SST concentrations data were achieved. The final training data set corresponded to years 2013 through 2014 (Table 1) with respective input delay, k, and feedback delay, r, which are discussed below.

Input and feedback delays partitioned the input time-series data into subsets, during the computation of the forecasted SST time-series. Physically they represented the set of past BISCD8 salinity and Black Creek flow data that determine the forecasted salinity concentrations at BISCC4. By exploring the behavior of the observed data through time (Fig. 2), initial numerical values of input and feedback delays were determined. From the initial estimation of the delay values, a trial-and-error approach was followed until obtaining a calibrated neural network. In this research, input delays (k) were varied from 1:15 to 1:80 and feedback delays (r) from 1:1 to 1:6. All computations were performed using MATLAB R2017. Table 2 shows the parameter values that produced a trained network.

2.3 Forecasting Salinity at BISCC4 Using the Trained Neural Network

For the prediction phase, exogenous data consisted of a combination of observed data and predicted data. To forecast BISCC4 hourly salinity corresponding to year 2015, the ANN used observed hourly salinity and flow at stations BISCD8, BISCC4, and Black Creek, for years 2013 and 2014. The BISCC4 salinity forecasted for year 2016 was estimated using BISCD8 observed salinity and observed Black Creek flow data. However, the 2015-predicted BISCC4 salinity data was used as exogenous input. The forecast of BISCC4 salinity for years 2017–2019 was performed using BISCC4 ANN-predicted salinities. Table 2 summarizes the strategy.

Table 2. Forecasting strategy.

Run	Exogenous input data			Predicted output
	Observed Salinity BISCD8	Observed Black Creek Flow	BISCC4	Salinity BISCC4
1	2013–2014	2013–2014	Observed 2013–2014	2015
2	2014–2015	2014–2015	Observed 2014–Predicted 2015	2016
3	2015–2016	2015–2016	Predicted 2015–Predicted 2016	2017
5	2016–2017	2016–2017	Predicted 2016–Predicted 2017	2018
6	2017–2018	2017–2018	Predicted 2017–Predicted 2018	2019

As explained in Table 2, the neural network model developed in this research is able to perform predictions of BISCC4 salinity making full use of previously predicted salinity concentrations for that station.

The measurement of how well predicted salinities matched observations was performed using several statistical indicators (Table 3). The goodness of fit was statistically assessed by the computation of the correlation coefficient (R), coefficient of determination (R^2), Nash Sutcliffe coefficient (NSE), Kling-Gupta efficiency (K-G), and Willmott's index of agreement (d). The statistical error was evaluated by the root-mean-squared-error to standard deviation ratio (RSR), and the bias of predicted salinity concentrations with respect to observed data was determined by the percent bias coefficient (PBIAS). These indicators are used for assessing ANN output [8] and are also widely used to assess hydrological and water quality modeled output [17–19]. Table 3 also shows the acceptability ranges recommended by the cited literature.

Table 3. Statistical indicators of fit and acceptability ranges.

Indicators of Fit	Formulae	Range
Root-mean-squared-error to standard deviation ratio, $RSR = \frac{RMSE}{STDEV_{Obs}}$	$\frac{\sqrt{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Sim}).^2}}{\sqrt{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Mean}).^2}}$	<0.79
Percent bias, *PBIAS*	$\frac{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Sim})*100}{\sum_{i=1}^{n}(Y_i^{Obs}).}$	<±18%
Correlation coefficient, *R*	$\sqrt{\frac{\sum_{i=1}^{n}(Y_i^{Sim}-Y_i^{Mean}).^2}{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Mean}).^2}}$	>0.71
Coefficient of determination	R^2	>0.50
Nash–Sutcliffe efficiency, *NS*	$1-\frac{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Sim}).^2}{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Mean}).^2}$	>0.50
Kling-Gupta efficiency, *K-G*	$1-\sqrt{\left(\frac{Y_{Sim}^{Mean}}{Y_{Obs}^{Mean}}-1\right)^2+\left(\frac{STDEV_{Sim}}{STDEV_{Obs}}-1\right)^2+(R-1)^2}$	>0.50
Willmott's index of agreement, *d*	$1-\frac{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{Sim}).^2}{\sum_{i=1}^{n}(\lvert Y_i^{Sim}-Y_{Obs}^{Mean}\rvert+\lvert Y_i^{Obs}-Y_{Obs}^{Mean}\rvert).^2}$	>0.65
Y_i^{Obs} = Observed SST concentration Y_i^{Sim} = Simulated SST concentration Y_{Obs}^{Mean} = Mean of observed SST concentration Y_{Sim}^{Mean} = Mean of simulated SST concentration n = Total number of daily SST concentrations		

3 Results

3.1 Training Phase

Table 4 summarizes the main neural network parameters calculated from the training phase. The parameter values shown in the table were selected from sensitivity analysis, where each parameter was varied until the network produced predictions of salinity concentrations for BISCC4 that matched observed concentrations.

Table 4. Artificial neural network parameters and features.

Feature	Adopted value
Number of layers	3 (input, hidden, output)
Number of nodes in hidden layer	1
Architecture	Closed loop
Number of iterations	100
Input delays	1:40
Feedback delays	1:1
Train ratio	70/100
Value ratio	15/100
Test ratio	15/100
Training algorithm	Bayesian regularization

As shown in Table 4, a closed-loop network architecture was adopted. This type of artificial neural network feeds back the actual output data, therefore it is a real-time configuration [20]. However, the network was trained under an open loop configuration in which it uses *a priori* target information to simulate output feedback.

Figure 4 shows the results for the training phase. As detailed above, observed data for years 2013 and 2014 were used. The training targets were salinity concentrations for BISCC4. The simulated concentrations for BISCC4 (training outputs) were calculated using observed Black Creek flow data and observed salinity data at BISCD8.

As illustrated in Fig. 4, the trained network captures observed BISCC4 salinity concentration values and trends very well for years 2013 and 2014 ($R > 0.99$, $R^2 > 0.99$). Nevertheless, the actual test of a trained ANN is in the prediction phase, where it is used to estimate future targets based on past information.

3.2 Forecasting/Prediction Phase

The trained network was used to predict hourly salinity concentrations at BISCC4, for years 2015 through 2019. Figure 5 shows the forecasting results achieved using the trained neural network. In addition to showing the observed and ANN-predicted salinity concentrations, the figure shows the predictions performed by the hydrodynamic model.

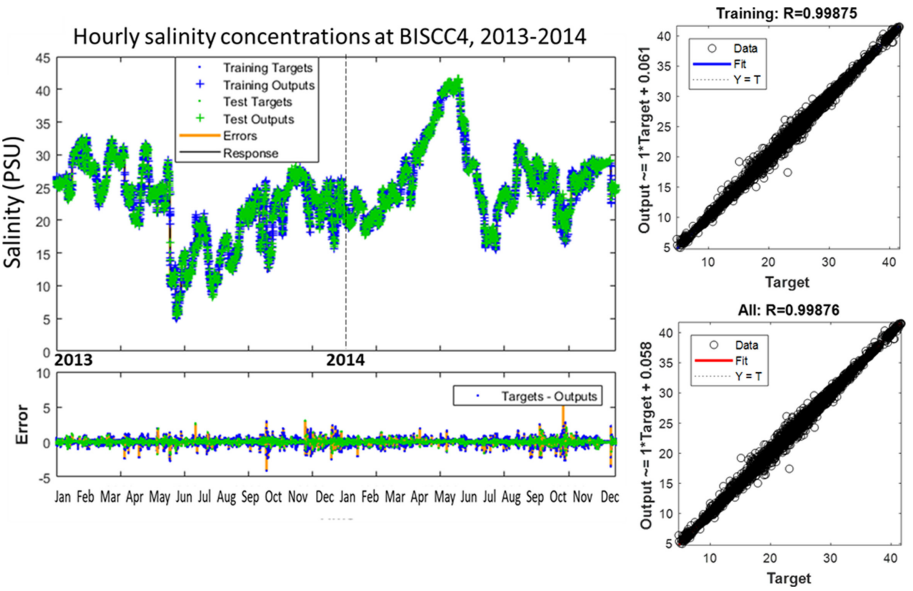

Fig. 4. Training phase results for years 2013 and 2014. Simulated and observed hourly salinity concentrations at BISCC4 are compared.

Clearly, the ANN captures the salinity temporal dynamics better than the hydrodynamic model. The goodness of fit of predicted concentrations to observed concentrations was measured by the indicators described in Table 3.

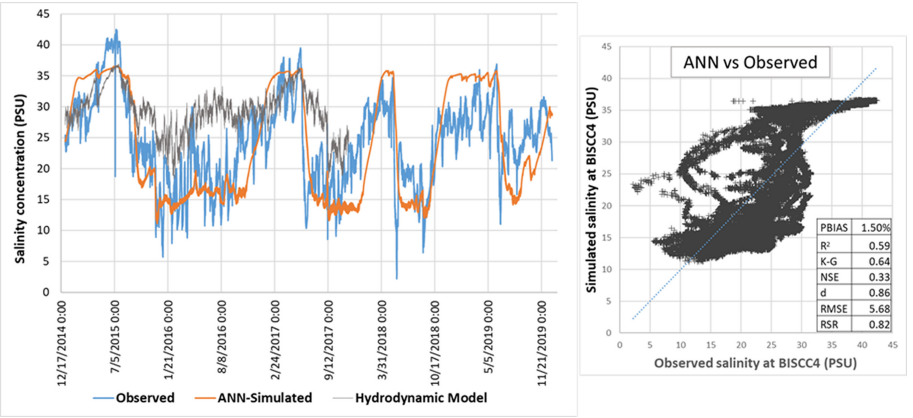

Fig. 5. Predicted hourly salinity concentrations at BISCC4 for years 2015–2019.

As evidenced in Fig. 5 and Table 5, the ANN developed in this research simulates hourly salinity concentration values and trends (for years 2015 through 2019) well. All statistical indicators are within their ranges of acceptability with the exception of

NSE and RSR. ANN simulations are an improvement from the salinity concentrations calculated by the hydrodynamic model.

Table 5. Comparison of statistical indicators.

Indicator	R	R^2	K-G	NSE	d	RSR	PBIAS
Acceptability range	>0.71	>0.5	>0.5	>0.5	>0.65	<0.79	<±18%
ANN model	0.77	0.59	0.64	0.33	0.86	0.82	1.5%
Hydrodynamic model	0.65	0.43	0.33	−0.08	0.54	1.04	−22.2%

While the statistics corresponding to the hydrodynamic model are all outside ranges of acceptability (Table 5), the indicators corresponding to the ANN show acceptable goodness of fit ($R^2 = 0.59 > 0.5$; K-G $= 0.64 > 0.5$; d $= 0.86 > 0.65$) and statistical error (PBIAS $= 1.5\% < \pm 18\%$; RSR $= 0.82 < 1$), except for NSE and RSR.

4 Conclusions

In this research, a closed-loop autoregressive neural network (with exogenous inputs) successfully simulated hourly salinity concentrations for 5-years, 2015 through 2019, at a coastal water quality station (BISCC4) in Biscayne Bay, Florida, USA. Training of the network was performed using observed salinity concentrations for BISCC4, observed Black Creek flow data (a nearby stream), and observed salinity data at BISCD8 (a water quality station located 10 km away from BISCC4).

Statistical indicators of goodness of fit show that the ANN developed in this research correctly simulates salinity concentration values and trends ($R^2 = 0.59$, K-G $= 0.64$, NSE $= 0.33$, d $= 0.86$, PBIAS $= 1.5\%$, RSR $= 0.82$). The ANN could be improved in capturing salinity peaks and minimums to achieve better NSE and RSR indicator values. However, in its current form the ANN model performs better in simulating salinity concentrations and trends, than an existing hydrodynamic model.

The results presented in this paper have the potential to be applied to other coastal locations in Biscayne Bay and elsewhere in which important freshwater inputs from inland streams are likely affecting salinity concentrations. Moreover, the developed neural network could be used to estimate groundwater flows contributions, in combination with simulations from an existing hydrodynamic model.

References

1. Conrads, P.A., Darby, L.S.: Development of a coastal drought index using salinity data. Bull. Am. Meteor. Soc. **98**, 753–766 (2017). https://doi.org/10.1175/BAMS-D-15-00171.1
2. Conrads, P.A., Roehl Jr., E.A.: Analysis of salinity intrusion in the Waccamaw River and the Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002. USGS Scientific Investigations Rep. 2007–5110, 41 pp (2007)

3. Shaw, J E., Zamorano, M.: Saltwater Interface Monitoring and Mapping Program. Water Resources Division, South Florida Water Management District (2020). https://www.sfwmd.gov/sites/default/files/documents/ws-58_swi_mapping_report_final.pdf
4. Abiy, A.Z., Melesse, A.M., Abtew, W., Whitman, D.: Rainfall trend and variability in Southeast Florida: Implications for freshwater availability in the Everglades. PLoS ONE **14**(2), e0212008 (2019). https://doi.org/10.1371/journal.pone.0212008
5. Stalker, J., Price, R., Swart, P.: Determining spatial and temporal inputs of freshwater, including submarine groundwater discharge, to a subtropical estuary using geochemical tracers, Biscayne Bay. South Florida. Estuaries and Coasts **32**, 694–708 (2009). https://doi.org/10.1007/s12237-009-9155-y
6. Lorenz, J.J.: A review of the effects of altered hydrology and salinity on vertebrate fauna and their habitats in Northeastern Florida Bay. Wetlands **34**(1), 189–200 (2013). https://doi.org/10.1007/s13157-013-0377-1
7. Alarcon, V.J., Linhoss, A., Kelble, C., Sanchez, G., Mardonez, F., et al.: Tidally forced saltwater intrusion into the Coral Gables Canal. Florida, USA (2021). (In process)
8. Alarcon, V.J.: Hindcasting and forecasting total suspended sediment concentrations using a NARX neural network. Sustainability **13**, 363 (2021). https://doi.org/10.3390/su13010363
9. Afan, H.A., El-Shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M.: Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. J. Hydrol. **541**, 902–913 (2016). https://doi.org/10.1016/j.jhydrol.2016.07.048
10. Shahid, E.S., Salari, M., Rastegar, M., Sheibani, S.N., Ehteshami, M.: Artificial neural network and mathematical approach for estimation of surface water quality parameters (Case study: California, USA). Desalin. Water Treat. **213**, 75–83 (2021). https://doi.org/10.5004/dwt.2021.26709
11. Lin, K., Lu, P., Xu, C.-Y., Yu, X., Lan, T., Chen, X.: Modeling saltwater intrusion using an integrated Bayesian model averaging method in the Pearl River Delta. J. Hydroinf. **21**(6), 1147–1162 (2021). https://doi.org/10.2166/hydro.2019.073
12. Zhou, F., Liu, B., Duan, K.: Coupling wavelet transform and artificial neural network for forecasting estuarine salinity. J. Hydrol. **588**, art. no. 125127 (2020). https://doi.org/10.1016/j.jhydrol.2020.125127
13. Rath, J.S., Hutton, P.H., Chen, L., Roy, S.B.: A hybrid empirical-Bayesian artificial neural network model of salinity in the San Francisco Bay-Delta estuary. Environ. Model. Softw. **93**, 193–208 (2017). https://doi.org/10.1016/j.envsoft.2017.03.022
14. Caccia, V., Boyer, J.: Spatial patterning of water quality in Biscayne Bay, Florida as a function of land use and water management. Mar. Pollut. Bull. **50**, 1416–1429 (2005). https://doi.org/10.1016/j.marpolbul.2005.08.002
15. USGS.: Changing Salinity Patterns in Biscayne Bay, Florida. Prepared in cooperation with South Florida Water Management District and Biscayne National Park (2004). https://doi.org/10.3133/fs20043108
16. Boussaada, Z., Curea, O., Remaci, A., Camblong, H., Mrabet, N.B.: A Nonlinear Autoregressive Exogenous (NARX) Neural network model for the prediction of the daily direct solar radiation. Energies **11**, 620 (2018). https://doi.org/10.3390/en11030620
17. Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE **50**, 885–900 (2007). https://doi.org/10.13031/2013.23153
18. Ang, R., Oeurng, C.: Simulating streamflow in an ungauged catchment of Tonlesap Lake Basin in Cambodia using Soil and Water Assessment Tool (SWAT) model. Water Sci. **32**, 89–101 (2018). https://doi.org/10.1016/j.wsj.2017.12.002

19. Knoben, W.J.M., Freer, J.E., Woods, R.A.: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol. Earth Syst. Sci. **23**, 4323–4331 (2019). https://doi.org/10.5194/hess-23-4323-2019
20. Hussein, A.A.: Derivation and comparison of open-loop and closed-loop neural network battery state-of-charge estimators. Energy Procedia **75**, 1856–1861 (2015). https://doi.org/10.1016/j.egypro.2015.07.163