National Park Service
U.S. Department of the Interior

South Florida Natural Resources Center
Everglades National Park

HYDROLOGIC
MODEL
REPORT

SFNRC Technical Series
2016:2

Hypersalinity in Florida Bay:

A Low-Dimensional Nonlinear Model

Hypersalinity in Florida Bay: A low-dimensional Nonlinear Model

HYDROLOGIC MODEL REPORT
SFNRC Technical Series 2016:2

South Florida Natural Resources Center
Everglades National Park
Homestead, Florida

National Park Service
U.S. Department of the Interior

Hypersalinity in Florida Bay: A low-dimensional nonlinear model

HYDROLOGIC MODEL REPORT

SFNRC Technical Series 2016:2

EXECUTIVE SUMMARY

Hypersalinity events in the coastal basins of Florida Bay are an annual occurrence driven by a combination of meteorologic, hydrologic, and oceanographic influences. Episodically, climatic conditions prevail that produce extreme hypersalinity events (salinity greater than 50 g/kg) associated with large scale seagrass die-offs triggering a cascade of ecological impacts and regional collapse of an entire ecosystem. Statistical regression models that estimate salinity based on linear predictors in a high–dimensional phase space are found to be less robust than nonlinear predictors in a low–dimensional phase space. A composite logistic-Gaussian function is used to model the nonlinear relation between basin runoff and salinity, and this nonlinear predictor performs better than linear models in the estimation of hypersalinity events in coastal basins of Florida Bay.

TABLE OF CONTENTS

CONTRIBUTING AUTHORS

Joseph Park, Erik Stabenau, Jed Redwine, and Kevin Kotun
South Florida Natural Resources Center, Everglades National Park
950 N. Krome Avenue, Homestead, FL 33030-4443

Comments and Questions: Joseph_Park@nps.gov, Erik_Stabenau@nps.gov, Jed_Redwine@nps.gov, Kevin_Kotun@nps.gov

Please reference this report as follows:

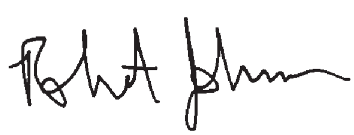
Park, J., Stabenau, E., Redwine J., and Kotun K., (2016). Hypersalinity in Florida Bay: A low-dimensional nonlinear model. U. S. Department of Interior, Everglades National Park, South Florida Natural Resources Center, SFNRC Technical Series 2016:2. 14 pp.

FOREWORD

Everglades National park is globally recognized as a beautiful and delicately balanced ecosystem. This recognition stems back to the early 20th century as famously expressed in Marjorie Stoneman Douglas' 1947 book *The Everglades: River of Grass*. What may be less well-known is that nearly one-third of Everglades National park consists of Florida Bay, an estuarine and shallow marine wilderness spanning the southern terminus of the river of grass. Florida Bay is home to vast areas of mangroves and seagrass beds forming a vital nursery for many species of fish and invertebrates, sustaining both a pristine wilderness and an economically vibrant recreational fishing industry.

While Florida Bay is geologically young, less than 5,000 years, instrumented records of its' physical characteristics span only decades. Hypersalinity events in the coastal basins of Florida Bay are an annual occurrence, but we have observed only two instances of extreme hypersalinity events. The first occurred in 1987 and a second in 2015, leading to cascading collapses of the marine and estuarine ecosystems. As scientists, our role is to try and understand how environmental conditions lead to these hypersalinity events with the goal of informing regional water management decisions, as well as to develop ecosystem indicators for the Comprehensive Everglades Restoration Plan (CERP).

One way to investigate and understand relationships between the meteorologic and hydrologic conditions that lead to hypersalinity events is with statistical and functional analysis of observed salinities and environmental conditions. In this report, the authors demonstrate synthesis of a computer model representing freshwater flow from the Everglades river of grass into Florida Bay, with nonlinear functions describing relationships between freshwater flow, salinities in the Gulf of Mexico, and salinities in the northern basins of Florida Bay. Their model relies on just two physically important input variables, hence the characterization low-dimensional, providing an improvement over existing statistical models in the attribution and prediction of hypersalinity events in Florida Bay. Analysis and models as demonstrated in this report mark progress in our collective efforts to protect and restore the Everglades, while underscoring the need for continued environmental monitoring and diligence.



Robert Johnson
Director
South Florida Natural Resources Center
Everglades National Park

October 2016

1 Introduction

Florida Bay is a shallow marine and estuarine wilderness at the southern end of the Florida peninsula situated between the Everglades, Gulf of Mexico, and Atlantic Ocean (figure 1). The bay is largely contained within the boundaries of Everglades National Park and supports diverse ecological habitats including freshwater marshes, mangroves, and seagrasses. Collectively, Florida Bay spans an area of approximately 2,200 km^2, however, it is not a contiguous open water environment but a tessellation of interconnected shallow basins separated by carbonate mud banks and mangroves.

Hydraulic connectivity between basins varies greatly with relatively large exchanges in relation to basin volume in the marine areas of the southern and western bay, and very little exchange among the estuarine coastal basins along the peninsula. These shallow, isolated coastal basins experience phenomenal excursions in salinity ranging from near zero during times of heavy rainfall and high water levels in the Everglades, to above 60 g/kg during hypersalinity events associated with drought and high evaporation conditions. The relative isolation of the coastal basins coupled with the potential for large amplitude, localized precipitation, runoff, and evaporation allows both hypersaline and hyposaline estuarine conditions to exist simultaneously in different basins across the bay.

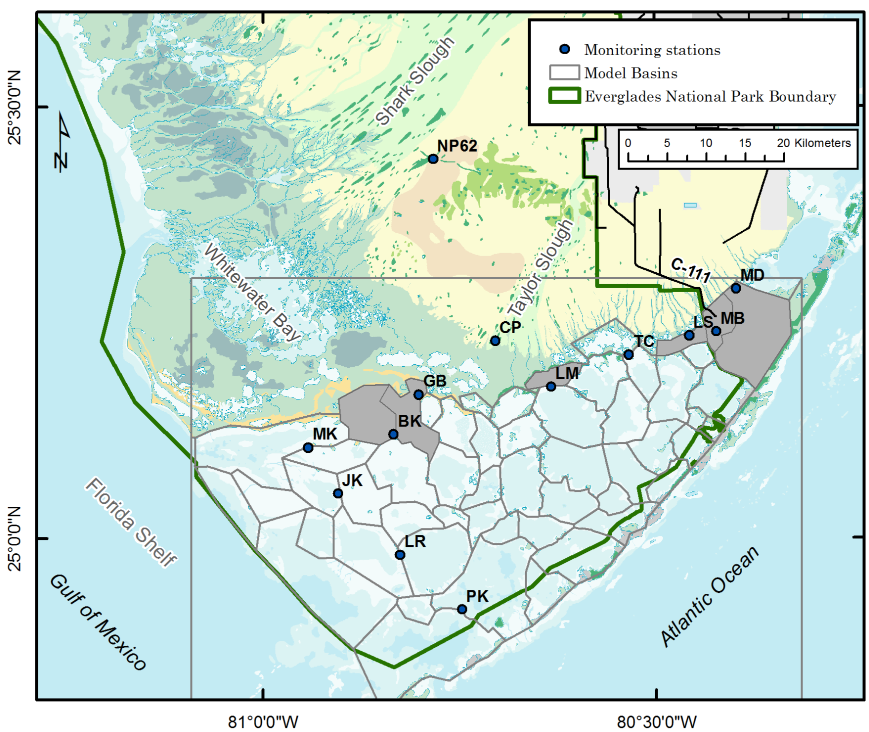


Figure 1.

Map of Florida Bay including physiographic features of the terrestrial Everglades. Salinity and water level measurement stations are denoted with the two and four letter abbreviations (e.g., LM, NP62). Basin boundaries of the BAM model are shown in Florida Bay. Coastal basins where the models are applied are shaded and correspond to observation stations as follows: Snake Bight : BK, Rankin Lake : GB, Little Madeira : LM, Long Sound : LS, Manatee Bay : MB, Barnes Sound : MD.

Seagrasses are widely present in Florida Bay and form the base of the ecological web that thrives there. They release dissolved carbon consumed by microorganisms at the bottom of the food chain, and seagrass leaves provide food for snails, urchins, sea turtles, fish, waterfowl, and manatees. Seagrass detritus is another primary food source for protozoans and nematodes, which in turn provide sustenance for shrimp, crabs and fish, which are then consumed by larger fish. It is thought that historically freshwater runoff from the Everglades, and dilution of Florida Shelf salinities in the Gulf of Mexico from Shark River flows, mitigated the development of hypersaline conditions in the northern Bay (*Marshall and Wingard*, 2014). Since the turn of the 20^{th} century, an ambitious series of water control and drainage features have fundamentally altered the historic flow of freshwater into the bay and homogenized the seagrass diversity to be dominated by a single species, *Thalassia testudinum*, commonly referred to as turtle grass (*Forqurean and Robblee*, 1999).

In 1987, a lingering drought combined with high temperatures resulted in hypersaline conditions and high water temperatures reducing the dissolved oxygen capacity of the water. This event triggered widespread turtle grass mortality followed by decomposition of the detritus resulting in high oxygen consumption due to excess carbon release, thereby fueling microorganisms to further consume oxygen. This feedback to anoxic conditions is thought to accelerate sulfate reduction in sediments liberating sulfide gas which is lethal to plants (*Koch et al.*, 2007a; *Rudnick et al.*, 2005). The initial seagrass die–off covered 40 km^2, but eventually expanded to affect over 240 km^2 as a result of cascading algal blooms leading to widespread seagrass die–off and fauna mortality (*Koch et al.*, 2007b).

In 2015, a local drought resulted in record low runoff during the usually wet summer, and by June 2015 daily average salinities in the central coastal basins exceeded maximums recorded over the previous decade. By mid–July, salinity peaked at 72 g/kg, the highest value recorded in 68 years of records. Concurrent with this hypersalinity event was a large seagrass die–off which eventually covered 160 km^2. Such a collapse at the base of the ecological web cascades into a local environmental crisis where seagrass beds are severely impacted, sediments become less stable, and the bay becomes less hospitable to marine and estuarine life. Effects from such events can linger for decades.

The ability to model and forecast hypersalinity events therefore has importance to the ecological health of Florida Bay and the adjacent coastal waters, can serve to inform regional water management decisions, and can be used as an indicator for success in the Comprehensive Everglades Restoration Plan (CERP).

1.1 Salinity in Florida Bay

Salinity in Florida Bay is highly variable, both in time and space, as governed by the influences of precipitation, evaporation, runoff, and mass exchange with adjacent basins and water bodies. *Kelble et al.* (2007) classified Florida Bay as a seasonally hypersaline estuary where the net freshwater input fluctuates widely throughout the year, but is near zero on an annual basis. This is reflected in an annual cycle where hypersaline conditions prevail at the end of the dry season in early summer transitioning to estuarine conditions at the end of the wet season in early winter. Salinities and environmental variables are monitored by a network of hydrographic, oceanographic, and meteorological stations operated by Everglades National Park as shown in figure 1.

Studies of salinity in Florida Bay appear to have been initiated by *Tabb* (1967) who used linear models to relate eastern coastal basin salinities with groundwater levels in Homestead, Florida and water levels in Shark Slough to the salinity front along the western coastal basins. Tabb recognized that water in Shark Slough, which eventually enters the Gulf of Mexico from Shark River and distributaries of Whitewater Bay, does not have a substantial overland flow path into Florida Bay. Nonetheless, a hydraulic head relationship exists between water levels in the southern Everglades and freshwater input, as expressed in strong negative correlations between Everglades water level and coastal basin salinity (*Marshall et al.*, 2011).

Potential mechanisms for transport of Everglades freshwater into Florida Bay include streamflow, overland flow and submarine groundwater. Streamflow is the only component that has been quantified (*Hittle et al.*, 2001), although *Corbett et al.* (1999) reported spot measurements of groundwater with significant spatial variability. *Langevin et al.* (2004) modeled the hydrology of the coupled Everglades–Florida Bay system finding that streamflows are a primary contributor on interannual and longer timescales, overland flow can be important at daily, weekly or monthly timescales, and that groundwater is dependent upon the relative water stage relationship between the Everglades and bay.

Nuttle et al. (2000) found that increased runoff into the bay would lower salinity in the eastern bay but have little effect in the western bay. However, they only considered runoff composed of streamflow from Taylor Slough and the C-111 canal into the eastern bay. They also applied several statistical and conceptual models demonstrating the difficulty of high–fidelity salinity estimation in Florida Bay. *Kelble et al.* (2007) analyzed monthly salinity data from ship measurements over a 7–year period (1998–2005) assessing mean and regional bay salinities, finding the expected negative correlation between runoff and salinities. Interestingly, they found a negative mean annual net freshwater supply of -5.3 cm, yet no overall increase in salinity over the period. This could be a result of inadequate runoff assumptions as their runoff was determined solely from streamflow measurements of nine streams discounting numerous small streams, sheetflow, and submarine groundwater.

Marshall et al. (2011) contributed a comprehensive review and analysis of Florida Bay salinities, suggesting that multivariate linear regressions of salinity against Everglades stage, regional wind, remote sea surface elevation, and flow into Shark River Slough or Taylor Slough could provide high fidelity salinity estimates at many basins. They reported roughly the same level of accuracy as numerical simulation models, but at a significantly reduced overhead. Their work was seminal in the sense that it filled critical information gaps in the planning of Everglades restoration which develops simulated hydrologic surfaces across several decades to compare alternative water resource scenarios and their impacts on the regional ecosystem. Although the results were generally good, errors in the coastal basins could be large. It should be noted that their model coefficients were determined largely over the periods from the mid–to–late 1990s through 2002.

1.2 Linear Model Limitations

The linear models of *Marshall et al.* (2011) are attractive since their application is not resource intensive, however, a review of their independent variables suggests a lack of independence in some cases and questions of physical relevance in others. For example, their model for salinity at Garfield Bight (GB) contains water stage from two Everglades stations, CP and NP62, as well as terms for wind components at Key West and Miami. A linear regression between daily mean water stage at NP62 and CP over the period September 1, 1999 to December 31, 2015 finds a coefficient of 1.14 with $\mathrm{R}^2 = 0.75$ and p–value < 1E-5. Likewise, a linear regression of Key West and Miami wind components finds an $\mathrm{R}^2 = 0.41$ (p–value < 1E-5). This suggests that assumed independence between these variables may not hold. Application of the Durbin–Watson test for serial autocorrelation between CP water stage and GB salinity finds a value of 0.07 (p–value < 1E-5) suggesting significant autocorrelation in the stage data. These auto and cross variable correlations suggest that caution may be warranted in the interpretation of error estimates and significance tests.

Regarding physical significance, the formula for GB includes terms for wind velocity at both Key West and Miami lagged by 4 days. While local winds play an important role modulating the inter–basin mass fluxes and salinity on short time scales (*Lee et al.*, 2016; *Langevin et al.*, 2004), it seems likely that wind speeds lagged by four days some 100 km away may not have direct physical relevance. Lastly, their regressions contain from 4 to 8 assumed independent variables. With such a high–dimensional parameter space the likelihood of overfitting increases which can reduce overall robustness and prediction accuracy when the environmental parameter regimes have changed and are not contained within the phase–space that was sampled when the regression coefficients were determined. Such data overfitting is a primary weakness of automated stepwise regressions wherein blind usage of regression variable selection criteria such as Mallow's C_p may not provide the best model structure (*van der Voet et al.*, 1997).

Nonetheless, the regressions have been shown to be accurate predictors when applied close in time to the periods over which the regression coefficients were determined providing a useful and efficient tool for assessing salinity in Florida Bay.

1.3 Nonlinear Terms

Another potential difficulty with simple linear modeling is that geophysical phenomena in general, and those exhibiting threshold behavior in particular, exhibit nonlinear responses in relation to forcings. A case in point as used in the *Marshall et al.* (2011) linear regressions can be found in the NP62 stage data and salinity at Garfield Bight. Examination of daily mean stage at NP62 and daily mean salinity in Garfield Bight reveals a nonlinear relation, one that is better fit by an exponential decay $S_{\mathrm{GB}} = s_0 - a(1+r)^{\mathrm{NP62}}$ rather than a linear predictor $S_{\mathrm{GB}} = s_0 - a\,\mathrm{NP62}$, where S_{GB} is the observed salinity at Garfield Bight, NP62 the stage at NP62, s_0 a bias term, a a fit coefficient, and r the rate of exponential decay.

A primary aim of the current work is to address some of the limitations encountered by *Marshall et al.* (2011) through the use of nonlinear predictor variables and a significant dimensional reduction. Our models have only two independent variables, basin runoff and boundary domain salinity. Runoff is determined by a mass–conservative numerical model based on observed water stage in the Everglades, observed rainfall and evaporation, observed and tidally predicted ocean water level, and basin water levels estimated by the numerical model. Boundary salinity is empirically determined from gauge observations.

The relationship between runoff and salinity is nonlinear, and we use a composite logistic–Gaussian kernel designed to capture both the limiting response of salinity, and the localized peak in response to basin runoff as described below. This nonlinear predictor allows the model to capture hypersalinity events that are difficult to represent with simple linear predictors.

In the following sections, we describe the runoff and boundary salinity variables, present linear and nonlinear versions of salinity models, and compare application of these models with the model of *Marshall et al.* (2011).

2 Runoff and Boundary Salinity

Runoff in our model is an aggregate flow governed by the hydraulic potential between water levels in the Everglades and coastal basins as determined by a mass–conservative numerical model, the Bay Analysis Model (BAM *Park et al.* (2016)). BAM decomposes Florida Bay into 54 basins based on the geomorphology of the mangroves, buttonwood banks, and shoals separating individual basins. Mass–transport over the interconnecting shoals is governed by the transport velocity $v = \sqrt{2g\frac{h_u - h_d}{1+f}}$ integrated over the shoal depth and length (cross–flow dimension), where h_u and h_d are the upstream and downstream water levels, $f = 2gn^2 w\rho^{-4/3}$ a friction factor where n is the Mannings friction, w the shoal width (along–flow dimension), ρ the shoal hydraulic radius and g the vertical acceleration. We note that the BAM model convention for runoff in a basin is that positive runoff corresponds to flow leaving the basin, while negative runoff quantifies flow entering the basin.

Each basin is forced with rainfall and evaporation. Basins on the Gulf of Mexico and Atlantic Ocean boundaries are also forced across the appropriate shoals with sea levels consisting of tidal variations and interannual sea level changes. Coastal basins along the Everglades boundary are forced with water levels determined from the Everglades Depth Estimation Network (EDEN) (*Telis et al.*, 2014) with the shoal properties (length, width, depth) calibrated to match aggregate runoff from the FATHOM model (*Marshall et al.*, 2008).

Salinity on the Gulf of Mexico boundary has significant variability in comparison to open ocean seawater. Here, the Florida Shelf has a wide, flat and shallow bathymetry, with a generally weak, northerly countercurrent to the Loop current. This allows hypersaline conditions to develop during times of high evaporation and weak circulation. The shelf also receives freshwater runoff from the Shark River and Everglades distributaries, and significant subtropical rain events can also contribute to hyposaline conditions. Boundary salinity for this domain is computed from a 4–gauge average of daily mean salinity at the stations MK, JK, LR, and PK (figure 1), which has an overall standard deviation of 1.7 g/kg over the period September 1, 1999 to December 31, 2015 (N = 5943).

On the Atlantic side salinity is less variable than the Florida Shelf, but still ranges considerably in comparison to open ocean values. The Lignumvitae basin (station PK) in southern Florida Bay is a predominantly marine environment with significant exchange with the Atlantic, and we use a six–month lowpass filter applied to the PK station salinity to represent Atlantic boundary salinities.

2.1 Rain, Evaporation, and Wind

It is worth noting that rain and evaporation are not explicitly included as independent variables in the nonlinear salinity model. Rather, they implicitly influence the hydraulic gradient that determines the runoff input by influencing water levels in both the Everglades and coastal basins. The initial version of the regression model did include a term consisting of daily mean rainfall minus evaporation, however at this short timescale there was no evidence of a functional relationship between rainfall minus evaporation to basin salinities and the term was discarded. We also note that wind effects are not included in our models.

3 Regression Models

We examine three regression models for Florida Bay salinity, a simple linear model (LM) of runoff and boundary salinity, a model with nonlinear (NL) predictors for runoff and boundary salinity, and the linear regressions of *Marshall et al.* (2011). The linear model is simply:

$$S_{LM} = \gamma\, R + \eta\, S_{BC} \tag{1}$$

where R is the basin runoff and S_{BC} the boundary salinity with fit coefficients γ and η.

The model with nonlinear terms specifies the runoff to salinity relation as:

$$S_R = s_{R_0} + A/(1 + e^{-a(R-R_L)}) + B\, e^{-(R-R_G)^2/2\sigma^2} \tag{2}$$

where s_{R_0} is a constant, A and B amplitudes of the logistic and Gaussian terms respectively, a a logistic slope parameter, R the runoff, R_L and R_G location offsets for runoff in the logistic and Gaussian terms respectively, and σ a shape parameter of the Gaussian term. The nonlinear boundary salinity to observed salinity term is:

$$S_S = s_{S_0} + (1 + r_s)^{S_{BC}} \tag{3}$$

where s_{S_0} is a constant, r_s is the rate of change, and S_{BC} the boundary salinity. These two terms are superposed to form the NL model:

$$S_{NL} = \alpha\, S_R + \beta\, S_S \tag{4}$$

where α and β are fit coefficients on the nonlinear runoff and boundary salinity terms respectively.

An examination of daily mean observed salinity *versus* runoff values over the period September 1, 1999 to December 31, 2015 at four coastal basins is shown in figure 2. Each of these relationships is clearly nonlinear, and two general properties are apparent. First, as exemplified in the Snake Bight and Rankin Lake basins, there is a saturation or limiting response at high positive runoffs. This could be a reflection of wet–season dynamics where basin runoff is large throughout the area, mitigating the occurrence of hypersalinity, or an effect from annual sea level maximums where exchange between the marine waters and basin waters are maximized, again limiting the potential for hypersalinity. Regardless of the cause, such a saturation dynamic can be modeled with a logistic function whereas a linear function could be inappropriate.

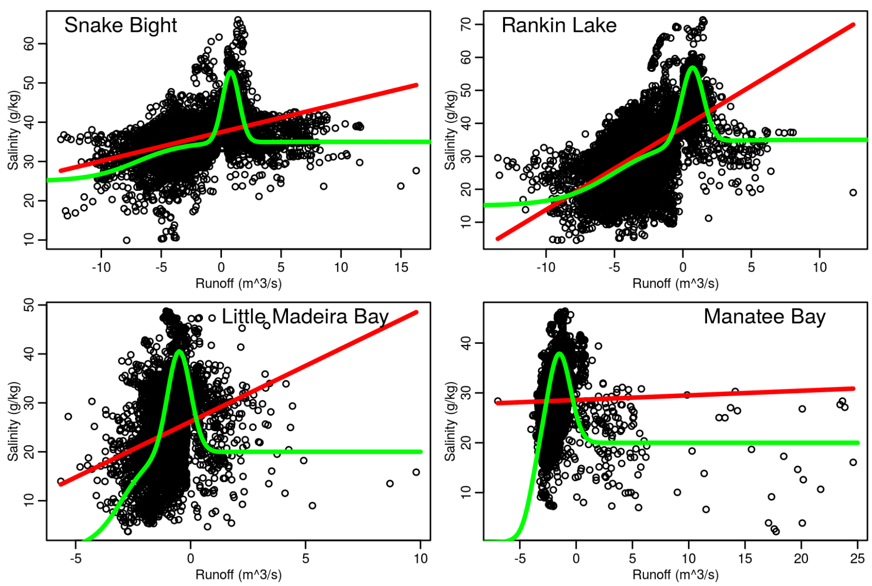


Figure 2.

Observed mean daily salinity plotted against BAM model runoff at four coastal basins over the period September 1, 1999 to December 31, 2015. Linear regressions are shown in red, with the nonlinear predictor from equation 2 in green. Negative runoff corresponds to flow entering the basin, positive runoff to flow leaving the basin.

The second property is one of localized hypersalinity at near–zero or slightly negative values of runoff, a feature inherently unsuitable for a linear function but which can be addressed with a Gaussian kernel. The presence of such localized hypersalinity features in the phase–space may be reason that simple linear models have difficulty capturing hypersalinity events. The nonlinear model parameters were determined by nonlinear optimization, and are listed in table 1.

Table 1. Nonlinear runoff and boundary salinity parameters for equations 2 and 3.

Basin	s_{R_0}	A	a	R_L	B	R_G	σ	s_{S_0}	r_s
Snake Bight	25.2	11.2	0.5	-7.2	18.2	0.8	0.7	17.5	0.084
Rankin Lake	15.7	19.1	0.5	-5.4	23.2	0.7	0.8	4.7	0.095
Little Madeira Bay	0.8	22.3	1.5	-3.1	21.8	-0.5	0.6	-4.3	0.097
Long Sound	5.4	19.3	1.2	-4.3	13.4	-2.0	0.7	-4.0	0.094
Manatee Bay	0.6	20.5	2.0	-3.8	18.3	-1.5	1.0	7.3	0.088
Barnes Sound	5.3	25.4	1.1	-4.4	8.7	-0.6	0.4	10.9	0.084

Plots of daily mean observed salinity *versus* boundary salinity values over the period September 1, 1999 to December 31, 2015 is shown in figure 3. Here, the response is approximately linear, although the nonlinear growth functions provide slightly smaller residual standard errors (Snake Bight $\varepsilon_{\mathrm{LM}} = 3.25$, $\varepsilon_{\mathrm{NL}} = 3.18$; Rankin Lake $\varepsilon_{\mathrm{LM}} = 7.71$, $\varepsilon_{\mathrm{NL}} = 7.49$; Little Madeira $\varepsilon_{\mathrm{LM}} = 5.55$, $\varepsilon_{\mathrm{NL}} = 5.31$; Manatee Bay $\varepsilon_{\mathrm{LM}} = 4.45$, $\varepsilon_{\mathrm{NL}} = 4.35$).

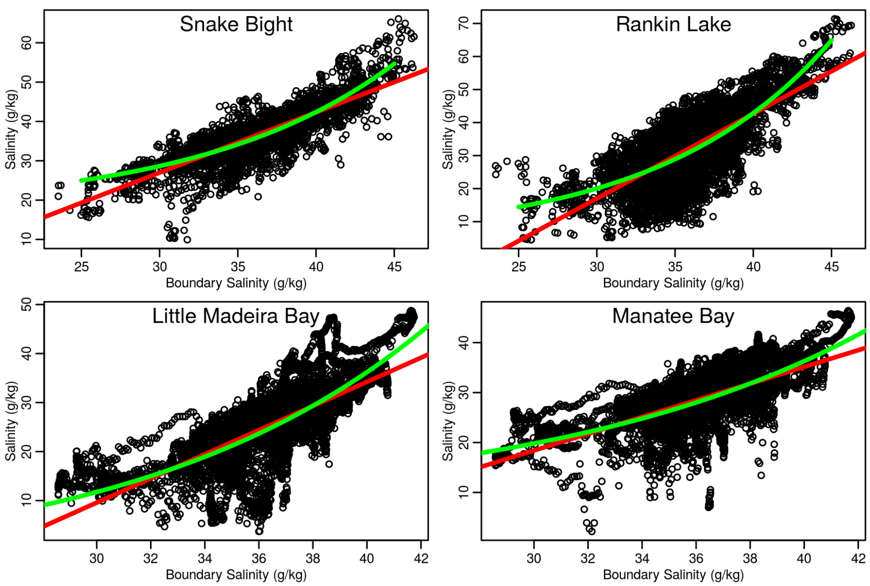


Figure 3.

Observed mean daily salinity plotted against boundary salinity at four coastal basins over the period September 1, 1999 to December 31, 2015. Linear regressions are shown in red, with the nonlinear predictor from equation 3 in green.

4 Results

The linear and nonlinear models (equations 1 and 4) were regressed against daily mean salinity at six coastal basins over the period September 1, 1999 through December 31, 2015 with the resultant fit coefficients shown in table 2. We also utilized the linear models of *Marshall et al.* (2011) and applied all three models to estimate basin salinities over this period.

Basin	γ	η	α	β
Snake Bight	0.445	1.030	0.120	0.874
Rankin Lake	2.122	1.058	0.361	0.630
Little Madeira Bay	2.013	0.729	0.107	0.852
Long Sound	2.897	0.915	0.397	0.645
Manatee Bay	0.124	0.799	0.167	0.802
Barnes Sound	2.238	0.880	0.317	0.653

Table 2.

Linear and nonlinear model fit coefficients of equations 1 and 4 over the period September 1, 1999 through December 31, 2015.

Figures 4 and 5 present the model comparison and residuals at Snake Bight, listing the RMS and maximum model errors on each plot. Since we are interested in hypersalinity events, we define the RMS error as $\varepsilon_{\mathrm{RMS}} = \sqrt{\sum_n^N (\max(0, S - \hat{S}))^2 / N}$, where S is the observed salinity, $\hat{S}$ the estimate and N the number of points in the series. Although we are assessing positive residuals, the same relative error relationships between the models hold with the canonical definition. Here we see that the *Marshall et al.* (2011) model does not capture the hypersalinity events in 2004, 2008 or 2015, and the linear model performs only slightly better in that regard. The nonlinear model does not fully reproduce the extreme hypersalinity events of 2004 and 2015, but does effectively capture these events with smaller mean and maximum errors than the linear models.

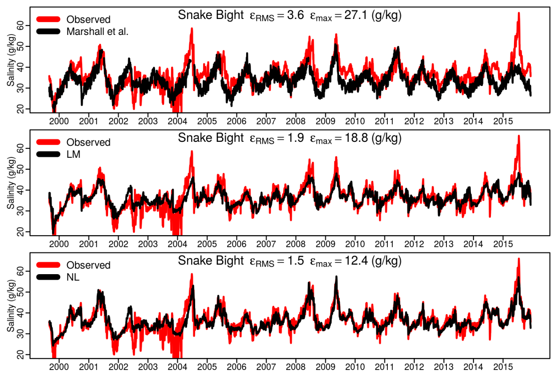


Figure 4.

Model comparison at Snake Bight. Top: Linear model of *Marshall et al.* (2011). Middle: Linear model of equation equation 1. Bottom: Nonlinear model of equation 4.

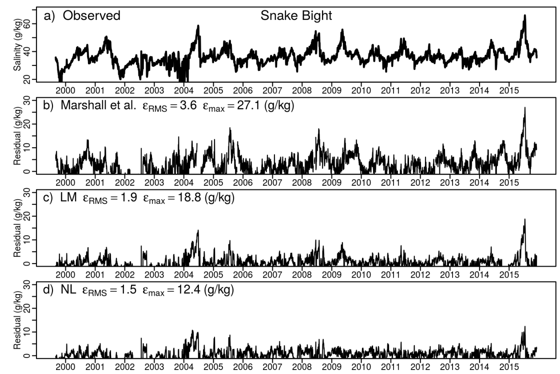


Figure 5.

Model residuals at Snake Bight. a) Observed salinity data. b) Linear model of *Marshall et al.* (2011). c) Linear model of equation 1. d) Nonlinear model of equation 4.

Results from Rankin Lake are shown in figures 6 and 7 where the *Marshall et al.* (2011) model performs well over most of the record, but fails to predict the 2015 hypersalinity event and produces higher variance over the entire record. Similar results are obtained at Little Madeira Bay in figures 8 and 9. Results at Manatee Bay are presented in figures 10 and 11 where again the 2015 hypersalinity event is not well represented in the linear models and the model of *Marshall et al.* (2011) produces estimates with a higher variance. A notable feature of the *Marshall et al.* (2011) model is that after 2004, there seems to be a general divergence from the observed salinities. This may be indicative that the finely–tuned and high–dimensional model of *Marshall et al.* (2011) has encountered a somewhat different parameter regime (phase–space), such that the model has reduced accuracy.

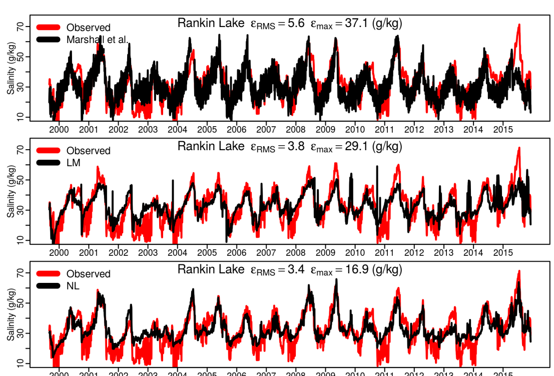


Figure 6.

Model comparison at Rankin Lake. Top: Linear model of *Marshall et al.* (2011). Middle: Linear model of equation 1. Bottom: Nonlinear model of equation 4.

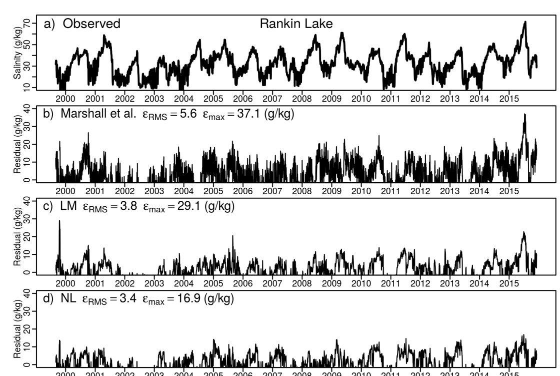


Figure 7.

Model residuals at Rankin Lake. a) Observed salinity data. b) Linear model of *Marshall et al.* (2011). c) Linear model of equation 1. d) Nonlinear model of equation 4.

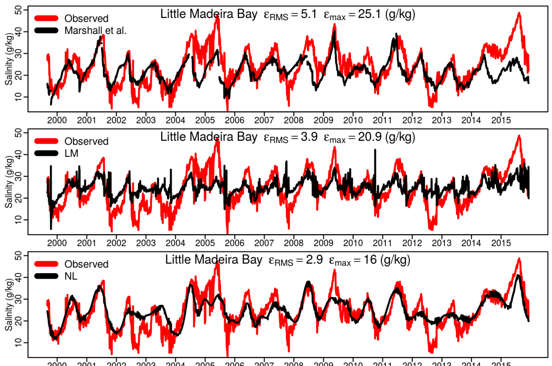


Figure 8.

Model comparison at Little Madeira Bay. Top: Linear model of *Marshall et al.* (2011). Middle: Linear model of equation 1. Bottom: Nonlinear model of equation 4.

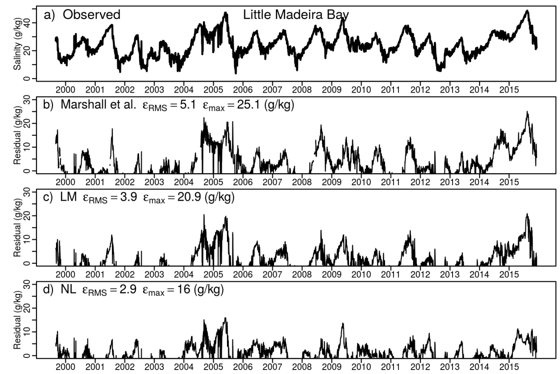


Figure 9.

Model residuals at Little Madeira Bay. a) Observed salinity data. b) Linear model of *Marshall et al.* (2011). c) Linear model of equation 1. d) Nonlinear model of equation 4.

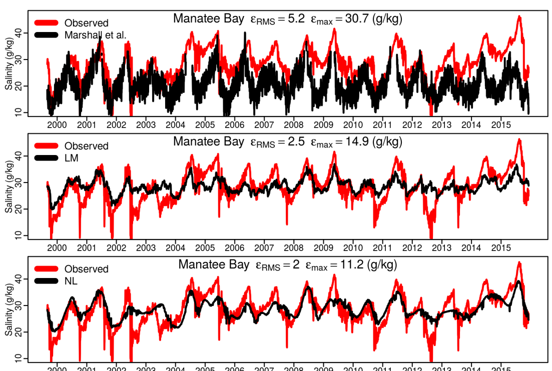


Figure 10.

Model comparison at Manatee Bay. Top: Linear model of *Marshall et al.* (2011). Middle: Linear model of equation 1. Bottom: Nonlinear model of equation 4.

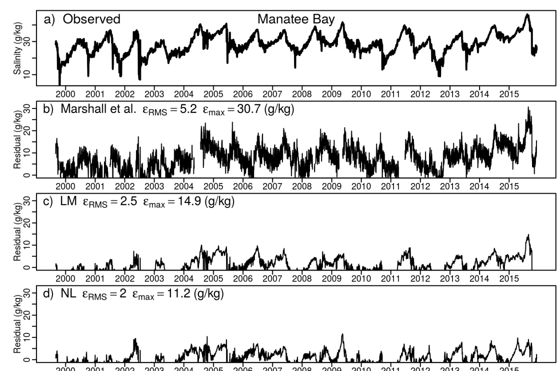


Figure 11.

Model residuals at Manatee Bay. a) Observed salinity data. b) Linear model of *Marshall et al.* (2011). c) Linear model of equation 1. d) Nonlinear model of equation 4.

A comparison of model errors at all six basins is shown in table 3. The results are consistent in that the models progress from larger to smaller errors as one considers the linear model of *Marshall et al.* (2011), a low–dimensional linear model, and the low–dimensional nonlinear model. Operationally, this suggests that the nonlinear model is better suited than the linear models for hypersalinity estimates.

Basin	Marshall $\varepsilon_{\rm RMS}$	Marshall $\varepsilon_{\rm max}$	LM $\varepsilon_{\rm RMS}$	LM $\varepsilon_{\rm max}$	NL $\varepsilon_{\rm RMS}$	NL $\varepsilon_{\rm max}$
Snake Bight	3.6	27.1	1.9	18.8	1.5	12.4
Rankin Lake	5.6	37.1	3.8	29.1	3.4	16.9
Little Madeira	5.1	25.1	3.9	20.9	2.9	16.0
Long Sound	6.8	38.2	3.9	22.6	3.3	16.4
Manatee Bay	5.2	30.7	2.5	14.9	2.0	11.2
Barnes Sound	3.0	17.5	2.3	12.7	2.1	13.1

Table 3.

Comparison of RMS and maximum model errors over the period September 1, 1999 through December 31, 2015, units are g/kg. LM is the linear model of equation 1 and NL the nonlinear model of equation 4.

However, we should also compare model errors over the common period of regression with the model of *Marshall et al.* (2011), which ended in 2002, and table 4 presents model error comparisons over the period September 1, 1999 through December 31, 2002. We find the same general model behavior with the low dimensional linear and nonlinear models performing better than the model of *Marshall et al.* (2011) except in the Barnes Sound basin, where the models perform nearly the same.

Basin	Marshall $\varepsilon_{\mathrm{RMS}}$	Marshall $\varepsilon_{\max}$	LM $\varepsilon_{\mathrm{RMS}}$	LM $\varepsilon_{\max}$	NL $\varepsilon_{\mathrm{RMS}}$	NL $\varepsilon_{\max}$
Snake Bight	2.7	13.3	1.2	7.4	1.0	7.4
Rankin Lake	4.5	26.5	3.4	29.1	2.6	13.7
Little Madeira	3.3	17.8	2.1	11.9	1.6	10.2
Long Sound	3.6	21.8	1.2	9.9	1.2	10.1
Manatee Bay	3.7	17.3	0.7	6.1	0.9	7.6
Barnes Sound	1.5	11.9	1.6	12.7	1.6	13.1

Table 4.

Comparison of RMS and maximum model errors over the period September 1, 1999 through December 31, 2002, units are g/kg. LM is the linear model of equation 1 and NL the nonlinear model of equation 4.

5 Discussion

Hypersalinity events in coastal basins of Florida Bay are associated with important ecological events such as widespread seagrass die–offs resulting in the collapse of an entire food web. While the bay is well–instrumented providing good spatial coverage and near real time observations of salinity and other physical variables, the ability to efficiently model hypersalinity events could provide actionable information in the months leading up to a hypersalinity event.

Two approaches for modeling the development of hypersalinity conditions include numerical models and statistical regressions. Numerical models entail significant resource commitments, while statistical models offer efficient estimates. *Marshall et al.* (2011) contributed an ambitious and comprehensive set of linear models for 21 basins in Florida Bay, effectively filling a void in the space of statistical models for bay salinity. While useful, these regressions may be limited by significant cross–correlations of the assumed independent variables, serial correlation in the independent variables, the nonlinear nature of the dependent variable relationships, and a highly–dimensional phase space that may allow for overfitting of the model. We have developed an alternative statistical model aimed at addressing the latter two issues: inherent nonlinearity of the dependent variables and a reduced phase space and number of assumed independent variables.

Even though our model results in a significant dimensional reduction, its data are not immune to serial or cross–correlations, although these issues are less severe than the application of *Marshall et al.* (2011). For example, while the presumed independent Everglades stage variables of CP and NP62 used by *Marshall et al.* (2011) have a linear coefficient of determination of 1.14, R^2 of 0.75, and a mutual information of 0.62 bits/measurement, the variables of runoff and boundary salinity have a linear coefficient of 0.44, R^2 of 0.13 and mutual information of 0.23 bits/measurement. A ratio of the mutual information suggests that the runoff : salinity variables have about 1/3 the interdependence of the stage variables ($(2^{0.23} - 1)/(2^{0.62} - 1) = 0.32$).

While the logistic–Gaussian function improves the estimation of hypersalinity events, its nonlinear nature, specifically the Gaussian localization of hypersalinity as a function of runoff, predisposes the model to predict hypersalinity. Figure 2 shows many runoff : salinity occurrences that are not well–represented by the function such that this model is not tailored to hyposalinity events and would not perform well in the estimation of hyposalinity. However, hyposalinity is a normal occurrence in the coastal basins of Florida Bay to which the ecosystems are well–adapted, and it is the hypersalinity events with the potential to initiate widespread ecological damage that are the focus of this work.

It should also be noted that our application is specific to coastal basins where runoff from the Everglades can be estimated. However, the use of nonlinear and dimensionally compact predictor functions is a straightforward exercise that can be applied to non–coastal basins with a model based on inter–basin flows and boundary salinity. Development of the nonlinear predictors has also raised questions regarding the nature of runoff and salinity dynamics, specifically, what physical processes govern the salinity saturation at positive runoff (outflow) values? Speculation could include large freshwater inter–basin flows during the wet–season, and large boundary inflows with seawater during seasonal cycles with elevated sea levels. It would also be informative to identify truly independent variables within limited dimensional phase spaces to improve attribution of physical significance and robustness of the estimates.

6 Conclusion

Florida Bay is an ecologically diverse marine and estuarine wilderness at the base of regional ecosystems and food webs for the Gulf of Mexico, Florida Keys, and Atlantic coastal areas of southern Florida. Hypersalinity events in the bay are part of a natural annual cycle, but climatic extremes can lead to prolonged and extreme hypersalinity events leading to a cascading collapse of the marine and estuarine ecosystems. Linear regression models offer computationally efficient means to estimate salinity, however, the use of linear predictor functions and highly dimensional interdependent variables may not be appropriate given the inherently nonlinear nature of the functional relationships, and may obscure attribution of physical relevance.

We have applied a composite logistic–Gaussian function to model the nonlinear relationship between coastal basin runoff from the Everglades and basin salinity, and a power–law growth function to model the boundary salinity and basin salinity relationship. Comparison of this model to a simple linear model based on proportional runoff and boundary salinity reveals that the nonlinear model estimates overall and hypersalinity conditions with lower mean and maximum error. Comparison of the simple linear model to a highly dimensional linear model applied by *Marshall et al.* (2011) finds that the reduced dimension model provides estimates with lower mean and maximal errors than the high–dimensional linear model.

LITERATURE CITED

Corbett D. R., Chanton J., Burnett W., Dillon K., Rutkowski, C., Fourqurean, J. W. (1999), Patterns of groundwater discharge into Florida Bay, *Limnol. Oceanogr.*, *44*(4), 1045–1055.

Fourqurean, J. W. and M. B. Robblee (1999), Florida Bay: a history of recent ecological changes, *Estuaries*, *22*, 345–357.

Hittle C., Patino E., and Zucker, M. A., (2001), Freshwater flow from estuarine creeks into northeastern Florida Bay, U.S. Geological Survey Water-Resources Investigations Report 2001-4164, 32 p. https://pubs.er.usgs.gov/publication/wri014164

Koch, M.S., Schopmeyer, S., Kyhn-Hansen, C., Madden, C.J., (2007), Synergistic effects of high temperature and sulfide on tropical seagrass, *J Exp Mar Biol Ecol*, *341*, 91–101.

Koch, M. S., Schopmeyer, S. A., Nielsen, O. I., Kyhn-Hansen C., Madden, C. J. (2007), Conceptual model of seagrass die-off in Florida Bay: Links to biogeochmical processes, *J Exp Mar Biol Ecol*, *350*, 73–88.

Kelble, C.R., Johns, E.M., Nuttle, W.K., Lee, T.N., Smith, R.H., Ortner, P.B. (2007), Salinity patterns in Florida Bay, *Estuarine, Coastal and Shelf Science*, *71*, 318–334. doi:10.1016/j.ecss.2006.08.006.

Langevin, C. D., Swain, E. D., and Wolfert, M. A. (2004), Simulation of Integrated Surface-Water/Ground-Water Flow and Salinity for a Coastal Wetland and Adjacent Estuary, U.S. Geological Survey Open-File Report 2004-1097, 30 p. http://pubs.usgs.gov/of/2004/1097/

Lee T. N., Melo N., Smith N., Johns E. M., Kelble C. R., Smith R. H., Ortner P. B., 2016, Circulation and water renewal of Florida Bay, USA, *Bull Mar Sci. 92*(2), 153–180. doi:10.5343/bms.2015.1019

Nuttle, W.K., Fourquerean, J., Cosby, B.J., Robblee, M. (1999), Influence of net freshwater supply on salinity in Florida Bay, *Water Resources Research 36*(7), 1805–1822. doi:10.1029/1999WR900352.

Marshall, F.E, Smith, D.T and Nuttle W. (2008), Simulating and forecasting salinity in Florida Bay: A review of models, Cooperative Agreement Number CA H5284-05-0006 Between The United States Department of the Interior National Park Service Everglades National Park and Cetacean Logic Foundation, Inc. http://sofia.usgs.gov/publications/reports/salinity_flbay/salinity_models.pdf

Marshall, F.E, Smith, D.T and D.M. Nickerson (2011), Empirical tools for simulating salinity in the estuaries in Everglades National Park, Florida. *Estuarine, Coastal and Shelf Science 95*, 377–387. doi:10.1016/j.ecss.2011.10.001.

Marshall, F. E., and Wingard, G. L. (2014), Florida Bay salinity and Everglades wetlands hydrology circa 1900 CE; A compilation of paleoecology-based statistical modeling analyses, U.S. Geological Survey Open-File Report 2013-1054, 32 p., http://pubs.usgs.gov/of/2012/1054

Park, J., Stabenau, E. and Kotun K. (2016). Florida Bay Assessment Model: User Manual. South Florida Natural Resources Center, Everglades National Park, Homestead, FL. Hydrologic Model Manual. SFNRC 2016:7-27. 62 pp. https://github.com/SoftwareLiteracyFoundation/BAM

Rudnick, D.T., Ortner, P.B., Browder, J.A., Davis, S.M., (2005), Florida Bay conceptual ecological model, *Wetlands*, *25* (4), 870–883.

Tabb, D.C., (1967), Predictions of Estuarine Salinities in Everglades National Park, Florida, by the Use of Ground Water Records, Ph.D. dissertation, 107 pp., University of Miami, Coral Gables, Florida.

Telis, P.A., Xie, Zhixiao, Liu, Zhongwei, Li, Yingru, and Conrads, P.A., (2015), The Everglades Depth Estimation Network (EDEN) Surface-Water Model, Version 2: U.S. Geological Survey Scientific Investigations Report 2014-5209, 42 p., doi 10.3133/sir20145209. http://sofia.usgs.gov/eden/

van der Voet H., ter Braak C. J. F., and Mallows, C., (1997), C_p and prediction with many regressors: comments on Mallows (1995), *Technometrics*, *39* (1), 119–116.

South Florida Natural Resources Center
Everglades National Park

950 N. Krome Ave, 3rd Floor
Homestead, FL 33030-4443

https://www.nps.gov/ever/learn/nature/technicalreports.htm