USING STATISTICAL MODELS TO SIMULATE SALINITY VARIATION AND OTHER PHYSICAL PARAMETERS IN NORTH FLORIDA BAY

Cooperative Agreement Number 1443CA528001020 Amendment/ Modification 0004
Between
The United States Department of the Interior National Park Service
Everglades National Park
And
Cetacean Logic Foundation, Inc.

FINAL PROJECT REPORT

Project Period October 1, 2002 through April 30, 2004

Frank E. Marshall, III
Principal Investigator

DeWitt Smith
Project Coordinator

David Nickerson
Review

April 30, 2004

USING STATISTICAL MODELS TO SIMULATE SALINITY VARIATION AND OTHER PHYSICAL PARAMETERS IN NORTH FLORIDA BAY

Table of Contents

USING STATISTICAL MODELS TO SIMULATE SALINITY VARIATION AND OTHER PHYSICAL PARAMETERS IN NORTH FLORIDA BAY

Cooperative Agreement Number 1443CA528001020 Amendment/ Modification 0004
Between
The United States Department of the Interior National Park Service
Everglades National Park
And
Cetacean Logic Foundation, Inc.

FINAL REPORT
Project Period October 1, 2002 through April 30, 2004

I. Introduction

This report describes the activities of the second year of Critical Ecosystem Studies Initiative (CESI) research into the use of statistical models to simulate salinity in Florida Bay. The activities of the first year of CESI work are presented in *Salinity Simulation Models for North Florida Bay Everglades National Park* (Marshall, et al 2003a). Because this report describes follow-on activities, reference to the first year report may be needed to understand all of the background for the second year of work. During the first year of this investigation, two types of statistical modeling procedures were found to be well-suited for use with time series data – SARIMA models (seasonal autoregressive integrated moving average models) and multivariate linear regression models (MLR models). SARIMA models were found to be useful for one-step forward predictions, but for other simulation purposes, MLR models were found to be much easier to use and almost as robust to the idiosyncrasies of time series data. Since the models are intended for use with the output from the South Florida Water Management District (SFWMD) Everglades watershed model (South Florida Water Management Model, or 2X2 model), which simulates hydrologic conditions in south Florida beginning in 1965, MLR models were selected for further development.

Tasks in the original Project Description for the Year 2 work are summarized as follows:

1. Coordinate with SFWMD to obtain 2X2 model data and evaluate the uncertainty in the 2X2 model simulations;
2. Obtain any other data that may be needed to use MLR models with 2X2 model data;
3. In conjunction with ENP staff, prepare MLR models of salinity, water level, or flow;
4. Simulate salinity, water level, or flow using the 2X2 Natural System Model and other appropriate input parameters;
5. Decode a previously prepared SARIMA model; and
6. Prepare draft and final reports.

As the project progressed considerable experience and feedback were gained using MLR salinity models. Additionally, when the uncertainty in the 2X2 model output was evaluated, the structure of the MLR salinity models was modified, as described further in the following sections of this report.

From the onset of the CESI work on statistical models, it was hoped that they would prove useful for simulating salinity in Florida Bay for the Initial CERP (Comprehensive Everglades Restoration Plan) Update (ICU) evaluations. However, when the Year 2 work began, it did not appear that the MLR models would be ready for this purpose, and the scope of the Year 2 project was made intentionally broad to investigate the use of MLR models to simulate other parameters besides salinity, such as water level or flow. As the project progressed it became clear that MLR models were capable of making acceptable simulations of salinity such that different water management schemes could be evaluated, and that the revised schedule for the ICU evaluations was going to make it possible for MLR models to be used for the evaluations. Therefore, the work being done was concentrated on the development of MLR models for that purpose (ICU evaluations).

The first tasks were completed as scheduled. Residual plots were examined and the updated dataset was assembled. In the midst of completing the project tasks, a need for the MLR models developed at Everglades National Park for use with the Interim Operational Plan (IOP) evaluation Congressional Report. ENP was tasked with analyzing the water management regimes that had been modified to lessen the impact of flow diversions on the Cape Sable seaside sparrow. MLR models were developed for use with these evaluations, and valuable experience was gained that has benefited the CESI project. In order to complete the IOP evaluations, a six-month extension of this CESI project was requested and granted. The IOP evaluation model development procedure also allowed the project dataset used for model development to be lengthened.

One committee that is charged with completing the ICU evaluations is the Southern Estuaries Sub-team of RECOVER. Beginning in the spring of 2003, the Principal Investigator has been coordinating with the Sub-team, preparing to use the models developed by this CESI project for evaluating the established salinity performance measures for Florida Bay and the southwest coast. At the time of preparation of this report, the models presented herein are intended to be used by the Sub-team in this manner.

When work re-started on this CESI project following the IOP evaluation activities, the tasks to be completed were officially modified to take into account the focus on modeling for ICU evaluation purposes. The revised Project Description included the following revised list of tasks:

1. Contact / meet with SFWMD staff to coordinate the acquisition of 2X2 model output and additional information about the modeling procedure.
2. Acquire the other data needed to create a complete input data set for running the multivariate linear regression models with 2X2 model output, including the historical record for wind at Key West and Miami weather stations.

3. Eliminate flow parameters from MLR salinity models.
4. Meet with the Southern Estuaries Sub-team to obtain their needs for MLR salinity models for ICU performance measure evaluations.
5. Prepare MLR salinity models for Joe Bay, Little Madeira Bay, Terrapin Bay, Garfield Bight and North River using 2X2 model output to calibrate the MLR salinity models. This task was completed. However, because the SFWMD updated the 2X2 model output subsequent to their development, these models were rendered obsolete. All future MLR salinity models will be developed from real (observed) data.
6. Adapt the IOP models prepared from observed data for use with the Southern Estuaries Sub-team performance measure evaluations by expanding the data used for model development, where possible for Little Madeira Bay, Terrapin Bay, Whipray Basin, Butternut Key, and Duck Key.
7. Prepare new MLR models using the longest period of record available for Taylor River, Little Blackwater Sound, Highway Creek, and Bob Allen Key.
8. Run simulations at all stations using the following 2X2 model runs: NSM 4.5, NSM 4.6, 95 Restudy, and 2000 CERP. These are the same runs being made by the Southern Estuaries Sub-team at other stations.
9. Evaluate the level of uncertainty in the models and in the simulations. Some statistical tests that may be used include the mean error, mean absolute error, root mean square error, maximum absolute error, relative mean error and relative absolute mean error.
10. Prepare draft and final Project Reports describing the activities that were completed and present the findings.

Details on the activities of these revised tasks are presented in this report.

II. Study Area and Data Set

The study area for this CESI project encompasses northeastern, north, and central Florida Bay; the extreme southwestern coast of the Florida; and the Everglades watershed within Everglades National Park. This modeling effort utilized data that have been collected at 15 to 60 minute increments and averaged to daily and monthly values. Salinity data is taken from the ENP Marine Monitoring Network (MMN) data base. The stage data are ENP Physical Monitoring Network Everglades water levels. Details about these data can be found in Everglades National Park (1997a and 1997b), and Smith (1997, 1998, 1999, and 2001). To these data other time series data were added, including wind data from the National Weather Service (Southeast Regional Climate Center), and water level data collected at Key West from the National Ocean Service. Wind data from Key West and Miami were used as these locations had the longest continuous records for wind and were considered to be representative of the regional wind patterns. Sea level data from Key West were considered to be representative of the average effect of oceanic water level influences, and, to some extent, the average water level patterns within Florida Bay.

The locations of each of the monitoring stations where water level and salinity data were collected are presented in Figure 1. The salinity monitoring stations for which MLR salinity models were prepared as part of this CESI study or the IOP evaluation (shown on Figure 1) are as follows:

1. Joe Bay
2. Little Madeira Bay
3. Terrapin Bay
4. North River
5. Whipray Basin
6. Duck Key
7. Butternut Key
8. Taylor River
9. Highway Creek
10. Little Blackwater
11. Bob Allen Key
12. Long Sound.

Continuous water level records in the Everglades begin in the 1950's in some locations, but most stage records date from the 1990's. Continuous salinity data extend back to 1988 at several locations in northeast Florida Bay. Because the shortest data record (for E146) begins on March 24, 1994, the period of data used for most of these modeling activities begins on this date. The period of record extends through October 31, 2002, which means that there are 3143 daily values in a record with no missing data. In reality, most data sets contained some missing values. Information on the parameters that were used for the modeling activities is presented in Table 1.

III. Residuals Analysis and Variable Significance Level Evaluation

The first task of the second year of this CESI project was to evaluate the residuals from the models that were developed in the first years work. Residuals (observed values minus simulated values, or deviations) were computed for all MLR salinity models including:

1. Residuals versus predicted values
2. Residuals versus time
3. Residual / probability / normal quartile.

Residual plots are presented in Appendix A. From the analysis of these diagnostic plots, it was determined that the preliminary MLR salinity models do not significantly violate any of the assumptions of linear regression model development, namely that the residuals are approximately normally distributed with a mean of 0 and a constant variance. However, the relatively large variability of the residuals indicates that there may be other significant predictor (independent) variables that are currently not in the models. The obvious example of a factor that is not currently included in the model is evaporation, and evaporation is an important process in salinity variation. However, direct measurements of evaporation on a daily basis are not available for use. Work by Nuttle (2003) has produced monthly estimates for evaporation in Florida Bay. Use of a spline-curve

Figure 1. The Everglades and Florida Bay Study Area Showing Monitoring Stations

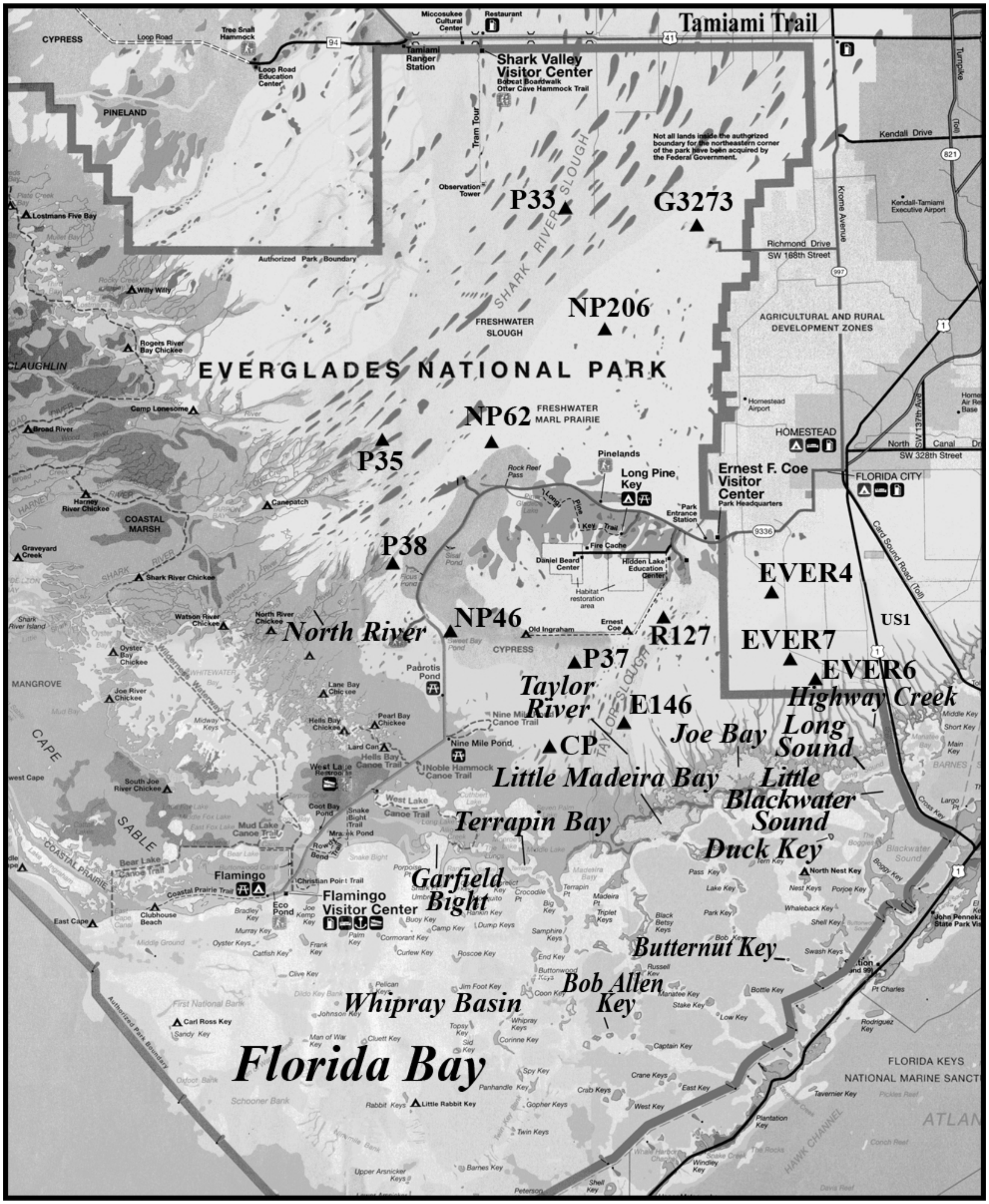


Table 1. Summary of information about the dependent and independent variables used in model development and verification, and in simulations.

Variable Name	Dependent or Independent	Variable Type	Units	Data Source	Location
Little Madeira Bay	Dependent	Salinity	PSU	ENP	Near-shore Florida Bay
Terrapin Bay	Dependent	Salinity	PSU	ENP	Near-shore Florida Bay
Long Sound	Dependent	Salinity	PSU	ENP	Near-shore Florida Bay
Joe Bay	Dependent	Salinity	PSU	ENP	Near-shore Florida Bay
Little Blackwater Sound	Dependent	Salinity	PSU	ENP	Near-shore Florida Bay
North River	Dependent	Salinity	PSU	ENP	Southwest Coast
Taylor River	Dependent	Salinity	PSU	ENP	Mangrove Zone
Highway Creek	Dependent	Salinity	PSU	ENP	Mangrove Zone
Whipray Basin	Dependent	Salinity	PSU	ENP	Open Water Florida Bay
Duck Key	Dependent	Salinity	PSU	ENP	Open Water Florida Bay
Butternut Key	Dependent	Salinity	PSU	ENP	Open Water Florida Bay
Bob Allen Key,	Dependent	Salinity	PSU	ENP	Open Water Florida Bay
Cp	Independent	Water Level	Ft, NGVD 29	ENP	Craighead Pond
E146	Independent	Water Level	Ft, NGVD 29	ENP	Taylor Slough
Ever4	Independent	Water Level	Ft, NGVD 29	ENP	So. Of FL City
Ever6	Independent	Water Level	Ft, NGVD 29	ENP	So. Of FL City
Ever7	Independent	Water Level	Ft, NGVD 29	ENP	So. Of FL City
G3273	Independent	Water Level	Ft, NGVD 29	ENP	East of S.R. Slough
NP206	Independent	Water Level	Ft, NGVD 29	ENP	East of S.R. Slough
NP46	Independent	Water Level	Ft, NGVD 29	ENP	Rocky Glades
NP62	Independent	Water Level	Ft, NGVD 29	ENP	East of S.R. Slough
P33	Independent	Water Level	Ft, NGVD 29	ENP	Shark River Slough
P35	Independent	Water Level	Ft, NGVD 29	ENP	Shark River Slough
P37	Independent	Water Level	Ft, NGVD 29	ENP	Taylor Slough
P38	Independent	Water Level	Ft, NGVD 29	ENP	Shark River Slough
R127	Independent	Water Level	Ft, NGVD 29	ENP	Taylor Slough
uwndkw	Independent	E-W Wind	N/A	NWS	Key West
vwndkw	Independent	N-S Wind	N/A	NWS	Key West
uwndmia	Independent	E-W Wind	N/A	NWS	Miami
vwndmia	Independent	N-S Wind	N/A	NWS	Miami
Kwwatlev	Independent	Tide Elevation	Ft, NGVD 29	NOS	Key West

method of interpolation to produce daily estimates did not create a time series that was significant as a predictor variable when tested.

Other predictor variables that were not included in the preliminary MLR salinity models that were investigated include the use of some measure of the hydraulic gradient in Shark River Slough, Taylor Slough, and in the eastern panhandle area. The following gradient independent variables were defined and evaluated:

R127 - E146
R127 - P37
P33 - P37
P33 – NP206
EVER4 - EVER6
EVER7 - EVER4.

Feedback on the preliminary models prepared during the first year of the project indicated that there was concern with the high number of independent variables in some of the models. An evaluation of the models (including gradient variables) showed that the significance level threshold for keeping a parameter in a model could be raised as high as 0.999 and there would still be 5-10 independent variables in each model, and the R^2 value remained high. This means that there was not much loss in explanatory power when the lesser significant parameters were dropped from the models, many of which were expressions of cross-correlation in the data.

IV. Observed Versus Model-Produced Data for Model Development

During April, 2003 visits were made to the SFWMD to coordinate obtaining the 2X2 model output for the ICU runs when it becomes available. From these meetings, it was learned that the 2X2 model output flow data may have a higher level of uncertainty compared to the water level simulations. Additionally, some of the water management structures have not been in place for the full 36-year period of the evaluations. Because of this and the fact that the correlation analysis showed that flow data are not as highly correlated to salinity at the locations in this study as water level in the Everglades (stage), a decision was made not to include any structure flows in the updated models.

When 2X2 model output data are compared to observed data, the 2X2 data frequently show a bias, greater at some stations than at others. A decision was made to adjust the 2X2 model data before they are input to the MLR salinity models in order to obtain a "best" simulation. When this is done, a higher Pearson's correlation coefficient value is obtained for 2X2 stage output and observed data. Initially, the bias was computed from the overlap period of 1995. When 2X2 model version 5.0 became available, this period of comparison was increased to 1996-2000. The bias between the two series' is then added or subtracted to/from the 2X2 model data.

The Southern Estuaries Sub-team is charged with the development of tools for Interim CERP Update (ICU) evaluations. They have developed performance measures for salinity in the embayments of Florida Bay. MLR salinity models were considered for use with the ICU performance measure evaluations at their July 2003 meeting. A recommendation was made by the Sub-team to use the CESI MLR salinity models for their performance measure evaluations for Florida Bay salinity. When the choice was made between models developed from observed stage data and models developed from 2X2 model stage data, the decision was made to develop the models to be used for ICU evaluations from the 2X2 model calibration/verification stage values, assuming that the 2X2 model would not be updated again for ICU evaluations.

However, the 2X2 model was subsequently re-calibrated, leading to the finding that observed data are the appropriate data for model development are observed data. Additionally, there is a strong aversion within the scientific community to using models that were developed from other modeled data, despite the fact that they have the ability to provide more accurate predictions, and are statistically sound.

Nonetheless, the models that were developed from 2X2 model output are presented below. These models should only be used with 2X2 model version 5.0.19 stage data, and historical wind and sea level data.

JOE BAY = 68.2 - 6.6 (P33 –P35) + 3.2 (EVER4 – EVER6) - 6.7 E146[lag2]
- 6.3 EVER6[lag6] - 5.7 P35[lag7] - 0.094 uwndkw + 0.074 uwndkw[lag2]
- 0.155 uwndmia[lag1] - 0.161 vwndmia[lag1]+ 7.0 kwwatlev[lag2]

LITTLE MADEIRA BAY = 34.6+ 2.2 (P33-P35) - 1.44 CP - 4.4 CP[lag21]
+ 1.9 NP46[lag17] - 2.4 R127[lag8] - 2.9 P33 - 0.15 vwndmia[lag1]
+ 3.8kwwatlev

TERRAPIN BAY = 32.5 - 4.0 (EVER4 – EVER6) - 8.7 CP[lag1] - 4.5 E146
+ 4.4 G3273[lag2] + 2.2 NP206[lag1] - 5.1 P33[lag2] - 4.4 P35[lag6]
- 0.31 uwndkw[lag1] - 0.24 vwndkw[lag2] + 2.8 kwwatlev + 5.2 kwwatlev[lag2]

GARFIELD BIGHT = 4.5 + 5.9 (P33-P35)[lag1] - 4.3 CP - 8.0 E146[lag1]
- 7.8 EVER4 - 6.0 NP46[lag1] + 4.4 P37[lag1] + 10.2 R127 - 0.19 uwndkw
- 0.14 uwndkw[lag2] + 0.09 vwndmia + 3.8 kwwatlev[lag1]

NORTH RIVER = 18.3 + 4.6 (P33-P35)[lag3] + 2.9 (EVER4-EVER6) - 4.3 CP
+ 4.8 E146[lag2] - 4.9 NP206[lag3] - 2.4 NP46[lag2] - 2.8 P37[lag2]
+ 1.8 kwwatlev [lag2]

In all of the models that are presented in this report, the following naming conventions have been adopted: kwwatlev is the water level measured at Key West; uwndmia and vwndmia are the *U* and *V* vectors of wind as measured at the Miami weather station; uwndkw and vwndkw are the *U* and *V* vectors of wind measured at Key West. These components are computed as follows:

U = (Resultant wind speed) * Cosine (Resultant direction)

V = (Resultant wind speed) * Sine (Resultant direction).

Resultant wind speed and direction are the daily average values as reported in the National Weather Service data archives. “Lag” refers to the value of the independent variable at the day in the past to be used in the model with the present day values of the other parameters.

The adjusted-R^2 values for these models prepared from 2X2 model output are as follows:

Joe Bay Salinity – 0.87
Little Madeira Bay Salinity – 0.79
Terrapin Bay Salinity – 0.81
Garfield Bight Salinity – 0.74
North River Salinity – 0.88.

V. MLR Salinity Models for the IOP Evaluation

At a CESI project progress meeting in early August 2003, it was decided that MLR salinity models would be used for the ENP IOP evaluations. The IOP evaluations were deemed a priority by ENP. To complete them, a 6-month extension to this CESI project was requested and granted. In the end, the IOP evaluation project was instrumental in showing that MLR salinity models could be used to compare various operational alternatives. It was also instrumental in determining the final activities for the second year CESI project. After the IOP evaluation project started, it became clear that some of the IOP models, prepared using observed data, could also be used for the ICU evaluations.

Updated MLR salinity models were prepared for the IOP evaluation. These models are physically defensible (see the Discussion section below) with terms in each model that are reasonable. Examination of each model shows that the most important Everglades water level station is Craighead Pond (CP), which appeared in all of the near shore models. Some combination of wind vectors also appeared in all models except the North River model (including all open water locations), which is as expected. Sea level (tide) appeared in most models, but not all. Because the significance level was set at a very high level for inclusion of a parameter in a model (0.999), it is expected that there are other parameters that would have been significant had the significance level been specified at a lower level more typically seen in other statistical evaluations (say 0.95 or 0.90). However, the fact that the significance level is so high means that there is little doubt as to the importance of the parameters in the models in explaining the variation in salinity when all of the other parameters are also being included.

Comparisons of water management operational scenarios were made using salinity estimated by the IOP models. Stage simulations for 31 years of data from the 2X2 model (ver. 4.5) for IOP, ISOP, Base 95 (same as 95 Restudy) and Natural Systems Model (NSM 4.5) operational conditions were used with historical wind and sea level data to simulate salinity with the IOP MLR salinity models. Comparative statistics prepared from the time series simulations were then evaluated, and statistically significant differences in salinity can be detected at most of the stations. This application showed that the MLR models have done their job, simulating salinity and providing consistent results that are supported by the current level of knowledge in hydrology and physiography of Florida Bay. The operational comparisons show that the MLR salinity models can adequately estimate salinity in a manner that will allow the comparisons to be made.

The conclusions of the IOP evaluation study can be summarized as follows (Marshall, 2003b):

1. Statistical models can be used for the reasonable simulation of salinity using multivariate linear regression techniques.
2. The evaluation procedure using the MLR salinity models with 2X2 model output for Everglades water levels and historical data for wind and sea level to simulate

long-term operations for Base95 and IOP / ISOP water delivery scenarios show an increase in salinity values at the following locations, primarily during the dry season, for monthly average values (80% significance level):

   Little Madeira Bay
   Terrapin Bay
   North River
   Whipray Basin
   Duck Key
   Butternut Key

3. No effect of IOP / ISOP operations compared to Base95 31-year simulations was seen in the salinity regime of Joe Bay and Long Sound.

The IOP salinity models that were developed are as follows and plots are presented in Figures 2 - 9:

JOE BAY = 37.1 - 3.1CP - 3.5EVER6[lag6] - 10.5E146[lag6] - 0.19uwndkw - 0.09uwndkw[lag2] - 0.1vwndkw - 0.16vwndmia[lag1], Adj-R^2 = 0.74

LITTLE MADEIRA BAY = 66.4 – 3.6CP[lag2] - 6.3P33[lag2] - 0.83(P33-NP206) – 0.21uwndkw + 0.15uwndmia - 0.14vwndmia[lag1] + 0.8kwwatlev[lag2], Adj-R^2 = 0.56

TERRAPIN BAY = 106.9 - 6.3CP[lag1] - 11.1P33[lag2] - 0.45uwndkw - 0.23uwndkw[lag1] - 0.2uwndkw[lag2] - 0.14vwndkw[lag2] + 0.46uwndmia + 1.9kwwatlev[lag2], Adj-R^2 = 0.76

LONG SOUND = 42.2 - 9.5CP[lag4] - 5.2EVER7[lag2] - 1.7EVER6[lag2] - 0.04vwndmia[lag1], Adj-R^2 = 0.80

NORTH RIVER = 36.7 - 4.3CP - 3.8CP[lag3] - 3.4NP206[lag3] + 0.6kwwatlev[lag2], Adj-R^2 = 0.86

WHIPRAY BASIN = 21.1 + 0.24ltmad[lag3] + 0.2terbay + 0.15terbay[lag3] - 0.04vwndkw[lag2] - 0.5kwwatlev[lag2], Adj-R^2 = 0.80

DUCK KEY = 10.2 + 0.3ltmad[lag1] + 0.4ltmad[lag3] + 0.10uwndkw[lag1] + 0.13vwndkw[lag2] + 0.5kwwatlev, Adj-R^2 = 0.70

BUTTERNUT KEY = 15.4 + 0.14ltmad[lag1] + 0.44ltmad[lag3] + 0.03terbay[lag3] - 0.08uwndkw - 0.10uwndkw[lag2] + 0.4kwwatlev, Adj-R^2 = 0.65

Figure 2. Joe Bay salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 23, 1995.

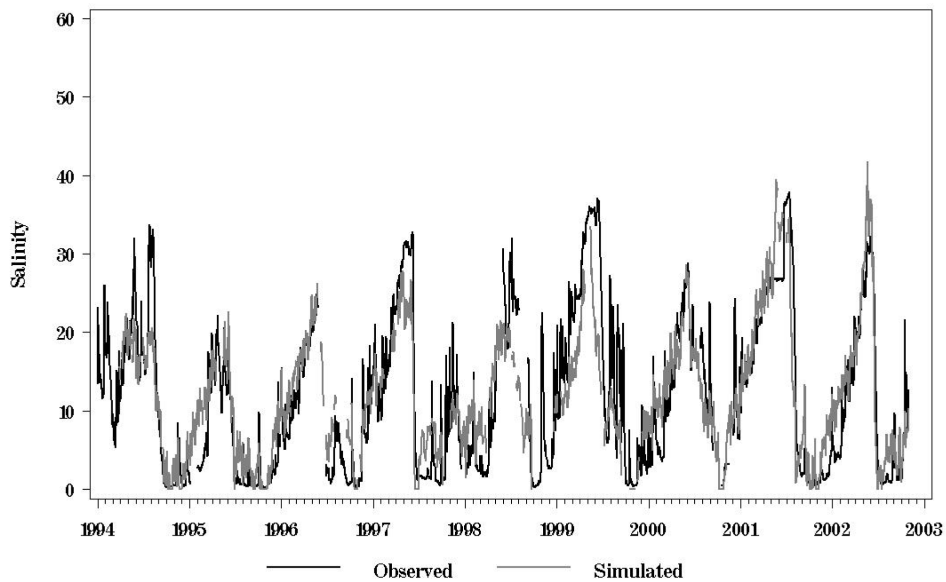


Figure 3. Little Madeira Bay salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

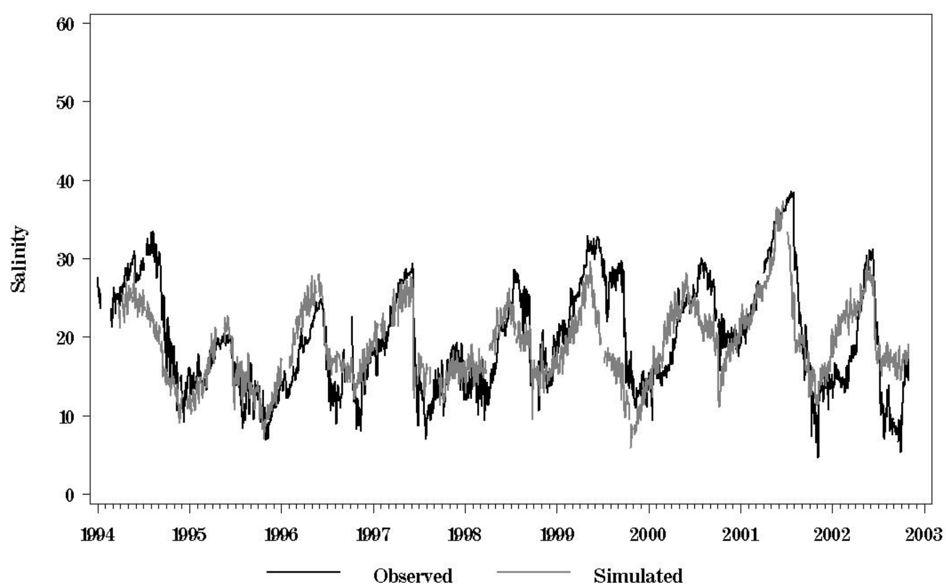


Figure 4. Terrapin Bay salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

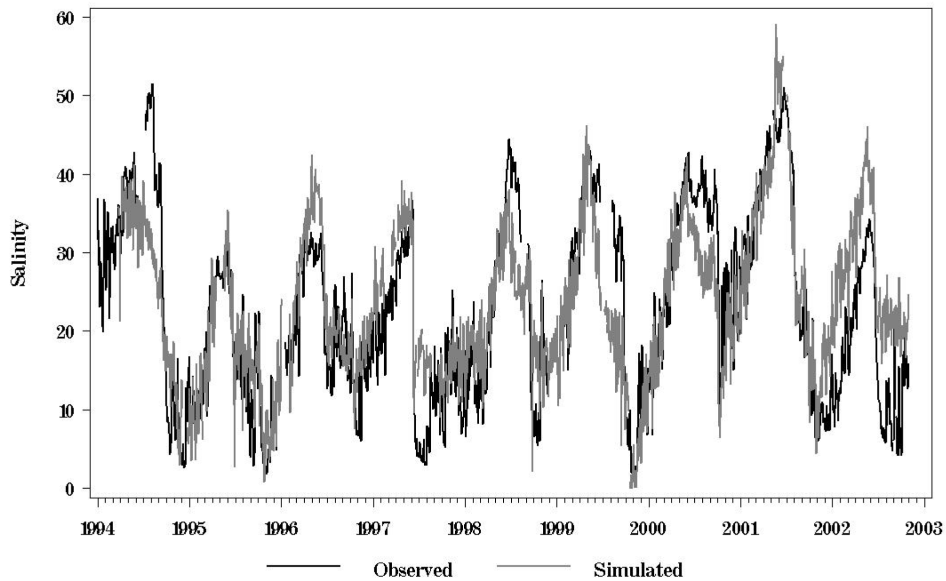


Figure 5. Long Sound salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

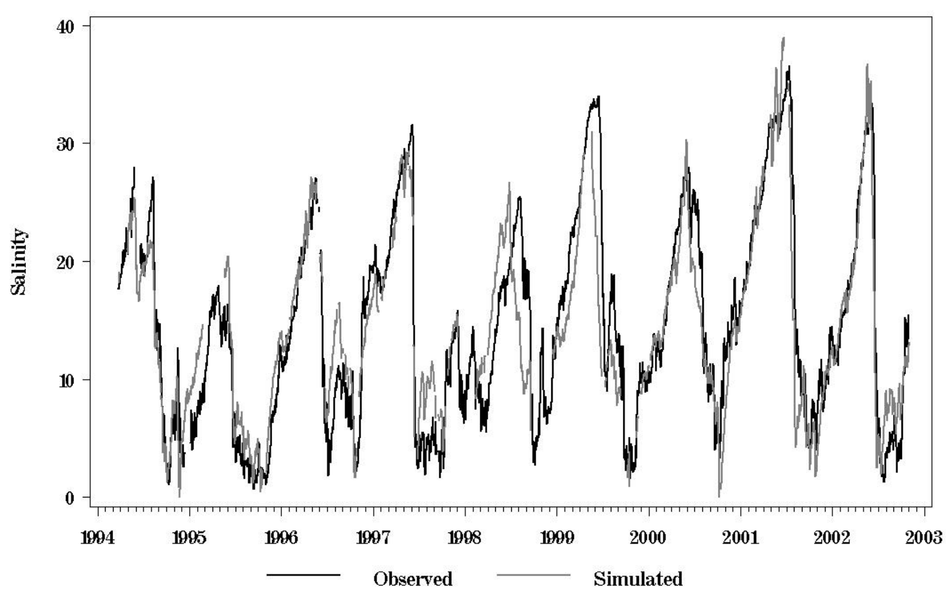


Figure 6. North River salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

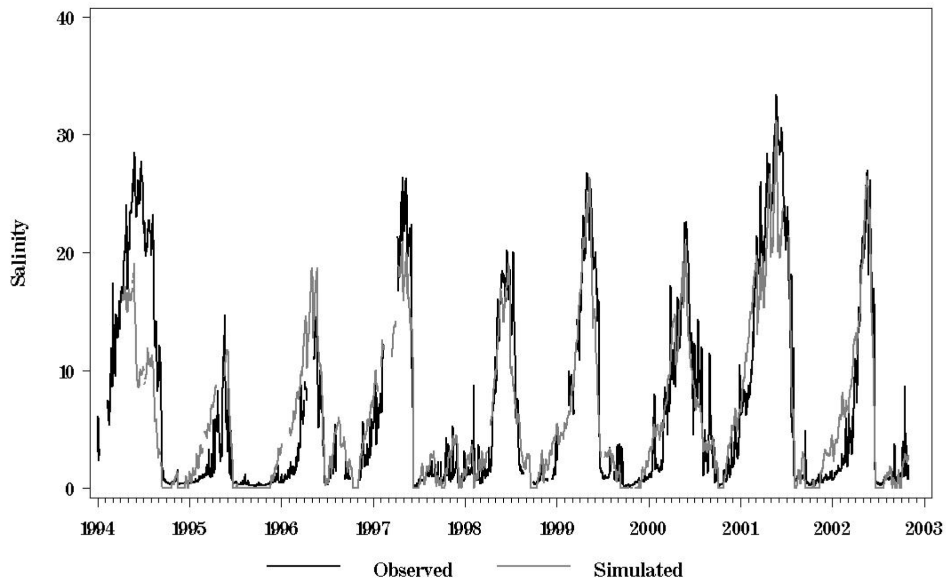


Figure 7. Whipray Basin salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

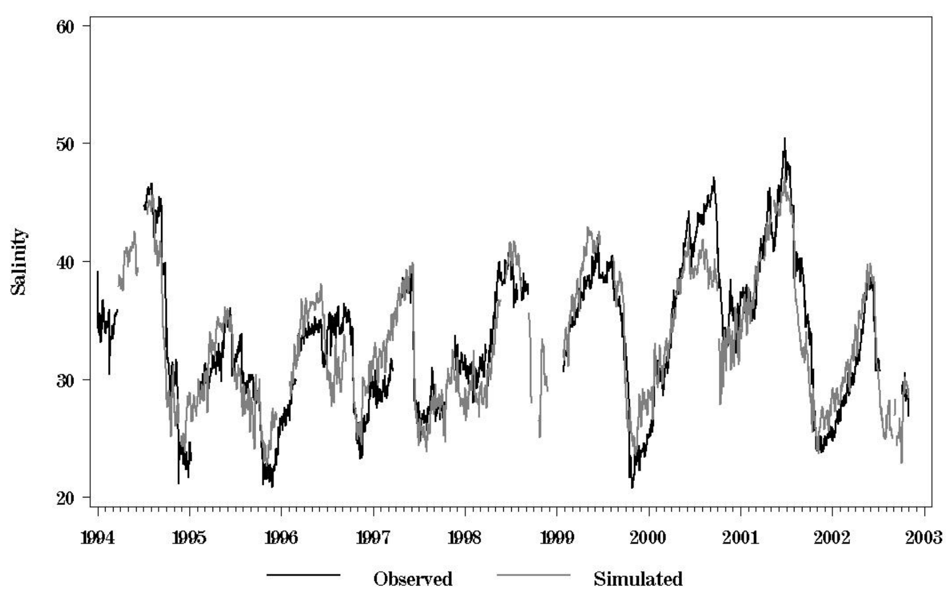


Figure 8. Duck Key salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

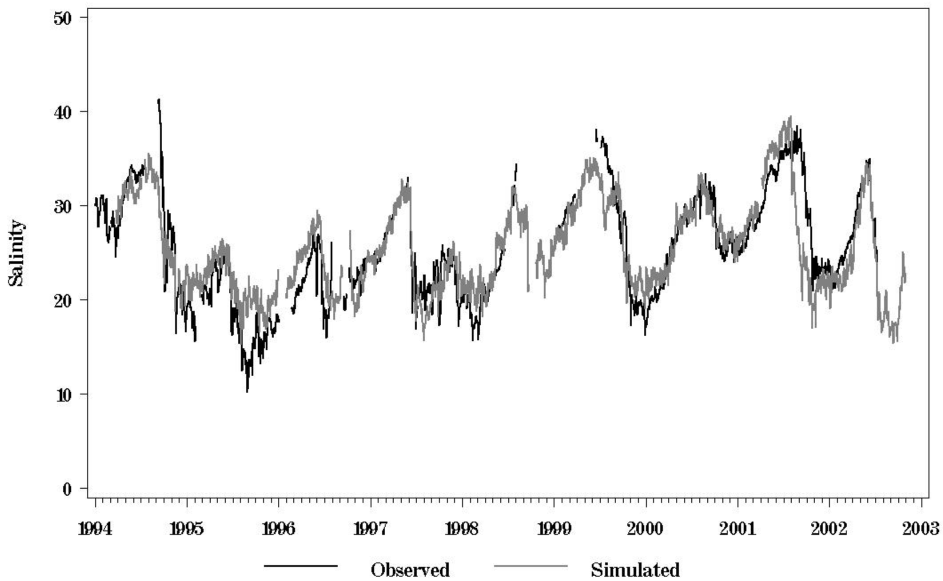
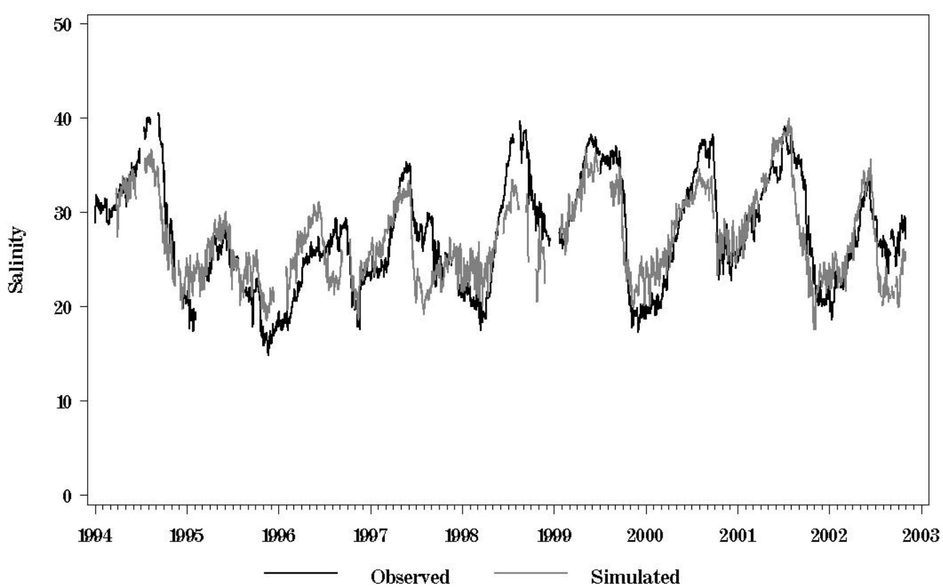


Figure 9. Butternut Key salinity model development plot (IOP) – Calibration is March 24, 1995– October 31, 2002; Verification is March 24, 1994 – March 3, 1995.



VI. Additional MLR Salinity Models

New MLR salinity models were prepared from observed data as part of this second year CESI project for Taylor River, Highway Creek, Little Blackwater Sound, and Bob Allen Key. The models are as follows:

TAYLOR RIVER = 83.2 - 15.1CP[lag4] + 0.8kwwatlev - 7.8(P33-P35)[lag1] – 4.4 (P33-P35) [lag4], Adj-R^2 = 0.78

HIGHWAY CREEK = 71.0 - 4.6E146[lag1] - 13.1EVER6[lag3] - 3.4R127[lag3] + 0.15uwndkw[lag1] + 0.1vwndkw[lag2] + 0.2uwndmia[lag3] - 4.4(P33-P37), Adj-R^2 = 0.81

LITTLE BLACKWATER SOUND = 42.5 -7.65CP[lag6] - 6.3EVER7[lag5] + 0.1vwndkw, Adj-R^2 = 0.75

BOB ALLEN KEY = 19.4 - 0.04uwndkw - 0.07uwndkw[lag2] - 0.06vwndkw[lag2] + 0.3ltmad + 0.25ltmad[lag3] + 0.08terbay[lag3], Adj-R^2 = 0.75.

Plots of the model simulations compared to the observed data for the model development and verification periods are presented as Figures 10 -13.

Figure 10. Comparison of Observed and Simulated Data for the Taylor River MLR Model – Calibration is March 24, 1995 – October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

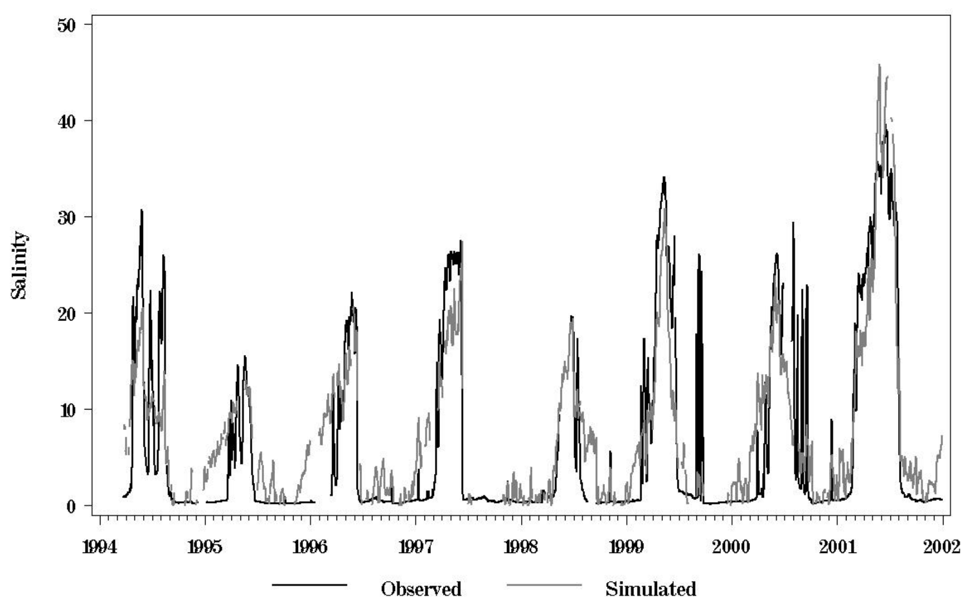


Figure 11. Comparison of Observed and Simulated Data for the Highway Creek MLR Model – Calibration is March 24, 1995 – October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

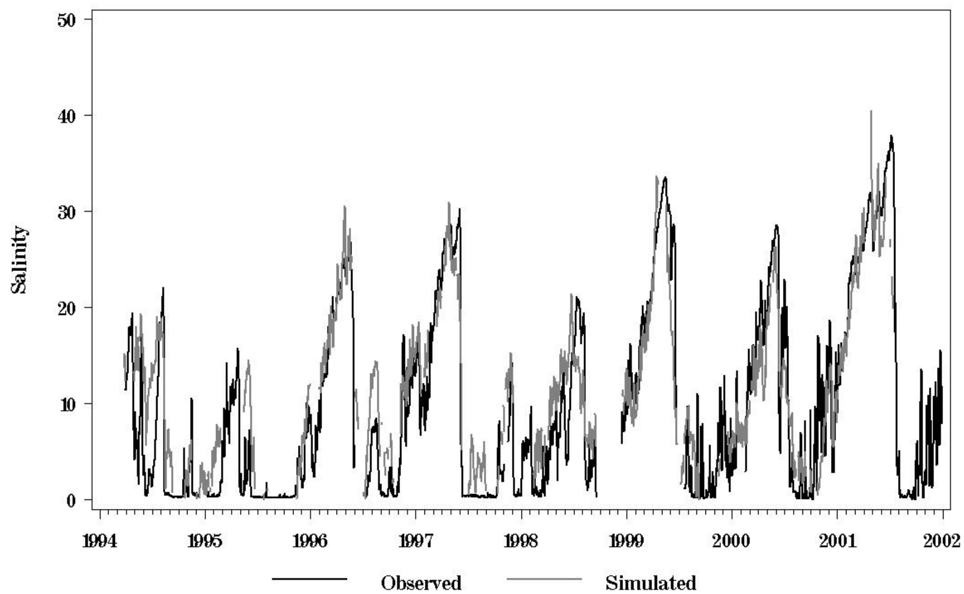


Figure 12. Comparison of Observed and Simulated Data for the Little Blackwater Sound MLR Model – Calibration is March 24, 1995 – October 31, 2002; Verification is March 24, 1994 – March 3, 1995.

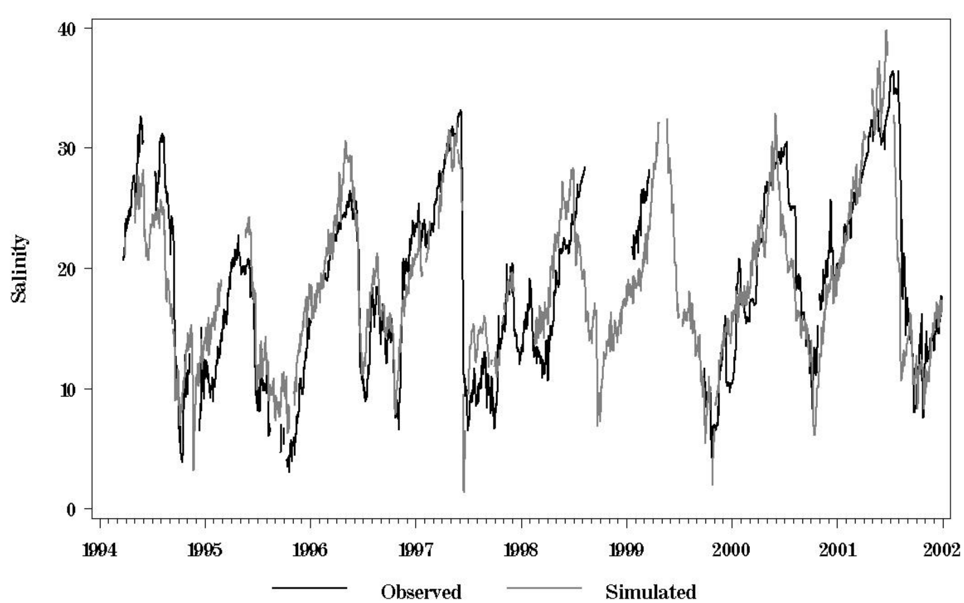
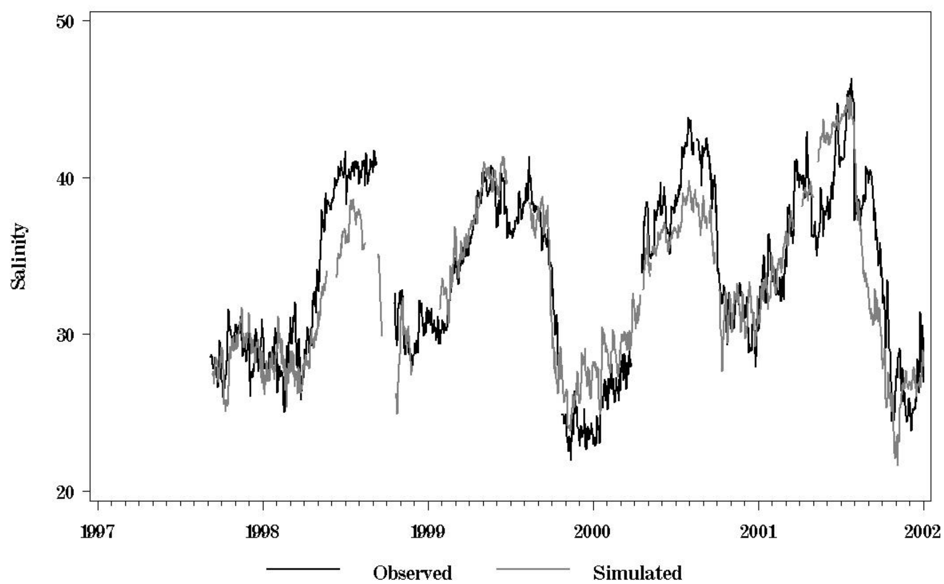


Figure 13. Comparison of Observed and Simulated Data for the Bob Allen Key MLR Model – Calibration is September 9, 1998– October 31, 2002; Verification is September 9, 1997 – September 8, 1998.



VII. Extended Period Models

The period of record for salinity and Everglades water level data at several stations extends back in time beyond March 24, 1994 (the beginning date for the this CESI / IOP dataset). For salinity, the period of record in the data available to the PI for Little Madeira Bay extends to August 25, 1988, and the period of record for Terrapin Bay extends to September 12, 1991. The period of record for the stage stations in the Little Madeira Bay and Terrapin Bay models covers the extended periods. In addition, the salinity models of Whipray Basin, Duck Key, and Butternut Key are a function of Little Madeira Bay and Terrapin Bay salinity (in addition to wind and sea level parameters), and the record at each of these open-water salinity stations extends at least as far as the Terrapin Bay salinity data. Therefore, extended period models can also be developed for these three open-water stations for the period extending to September 12, 1991. For these five extended period models, the model development data continues through September 31, 2001. The period from October 1, 2001 through September 31, 2002 was used for verification purposes in order to use the data from the start of the period for model development because the late 1980's was a period of relatively severe drought.

There were several objectives for preparing MLR salinity models with an extended period of data. Although there have been several short duration dry periods during the period of record used for development of the IOP/CESI models, there have been questions as to how well the MLR salinity models will perform during extended periods of drought, such as the severe drought experienced by south Florida in the mid-1980's. Therefore, one of the reasons for preparing the extended period models is to determine the effect on the models of including this additional data.

The MLR salinity models developed with the extended period of data are as follows:

LITTLE MADEIRA BAY = 106.1 - 0.3CP[lag2] - 12.5P33[lag2] - 1.7(P33-NP206) - 0.25uwndkw + 0.13uwndmia - 0.19vwndmia[lag1] + .95kwwatlev[lag2], Adj-R^2 = 0.65

TERRAPIN BAY = 101.2 - 7.4CP[lag1] – 10.0P33[lag2] - 0.36[uwndkw] - 0.20uwndkw[lag1] - 0.21uwndkw[lag2] - 0.19vwndkw[lag2] + 0.31uwndmia + 1.4kwwatlev[lag2], Adj-R^2 = 0.71

WHIPRAY BASIN = 21.0 + 0.0004vwndkw[lag2] + 0.21kwwatlev[lag2] + 0.2ltmad[lag3] + 0.20terbay + 0.19terbay[lag3], Adj-R^2 = 0.77

DUCK KEY = 9.6 + 0.06uwndkw[lag1] + 0.15vwndkw[lag2] + 1.1kwwatlev + 0.33ltmad[lag1] + 0.45ltmad[lag3], Adj-R^2 = 0.70

BUTTERNUT KEY = 14.6 + -0.06uwndkw - 0.09uwndkw[lag2] + 0.96kwwatlev + 0.13ltmad[lag1] + 0.47ltmad[lag3] + 0.06terbay[lag3], Adj-R^2 = 0.66

Plots of these extended periods models showing the comparison between observed and simulated data are presented in Figures 14 – 18.

Figure 14. Comparison of Observed and Simulated Data for the Little Madeira Bay Extended Period MLR Model – Calibration is August 25, 1988 – October 31, 2001; Verification is November 1, 2001 – October 31, 2002.

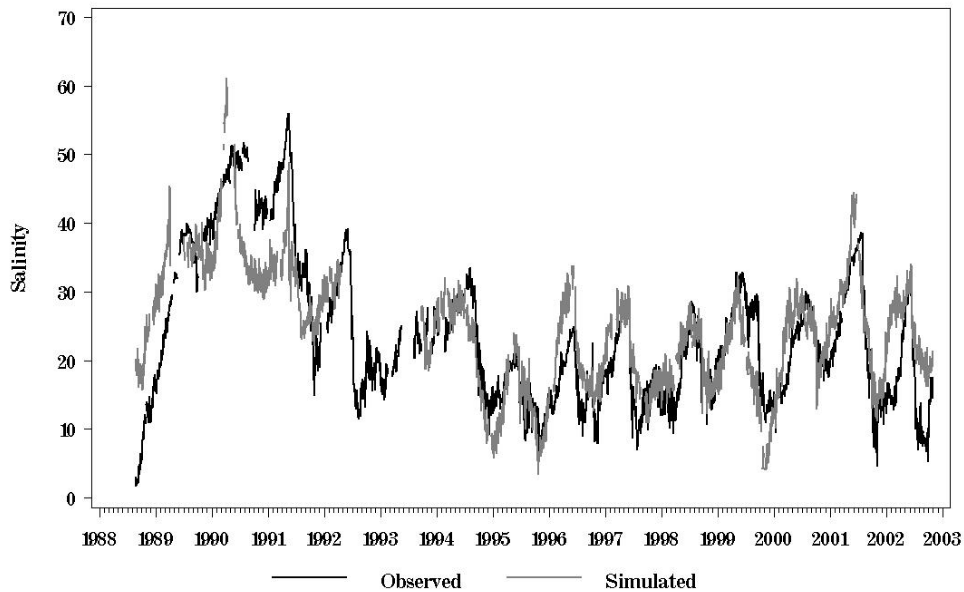


Figure 15. Comparison of Observed and Simulated Data for the Terrapin Bay Extended Period MLR Model – Calibration is September 12, 1991 – October 31, 2001; Verification is November 1, 2001 – October 31, 2002.

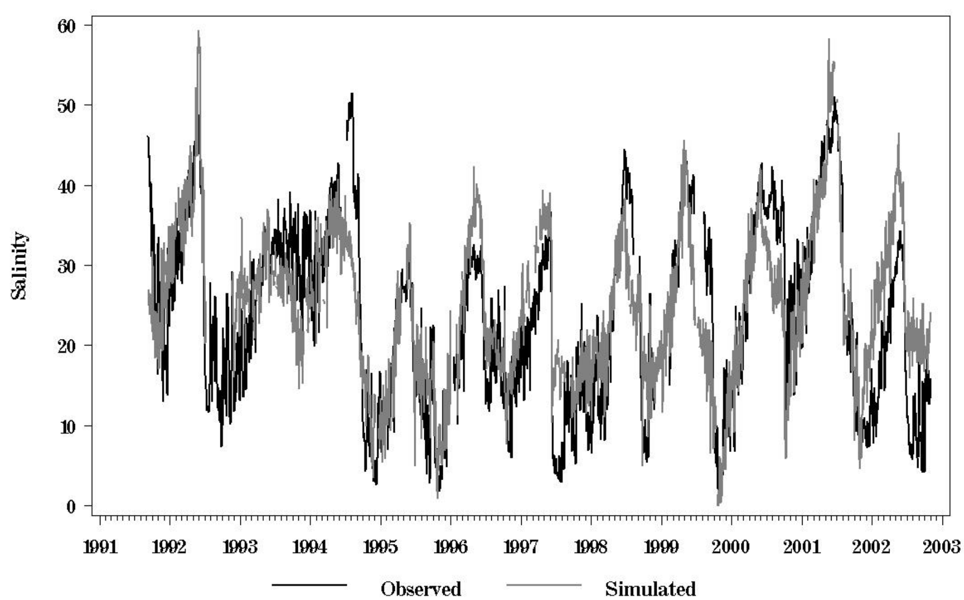


Figure 16. Comparison of Observed and Simulated Data for the Whipray Basin Extended Period MLR Model – Calibration is September 12, 1991 – October 31, 2001; Verification is November 1, 2001 – October 31, 2002.

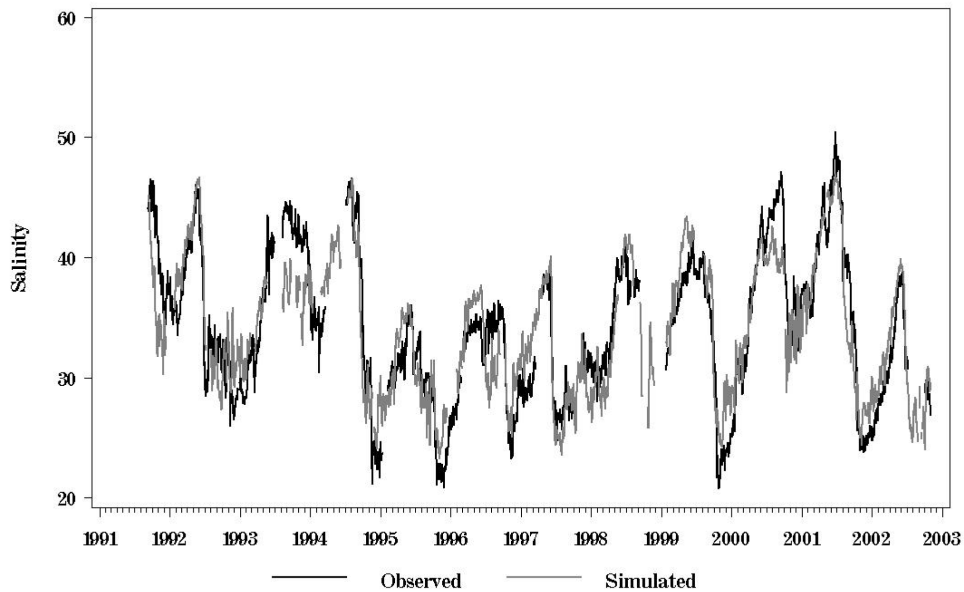


Figure 17. Comparison of Observed and Simulated Data for the Duck Key Extended Period MLR Model – Calibration is September 12, 1991 – October 31, 2001; Verification is November 1, 2001 – October 31, 2002.

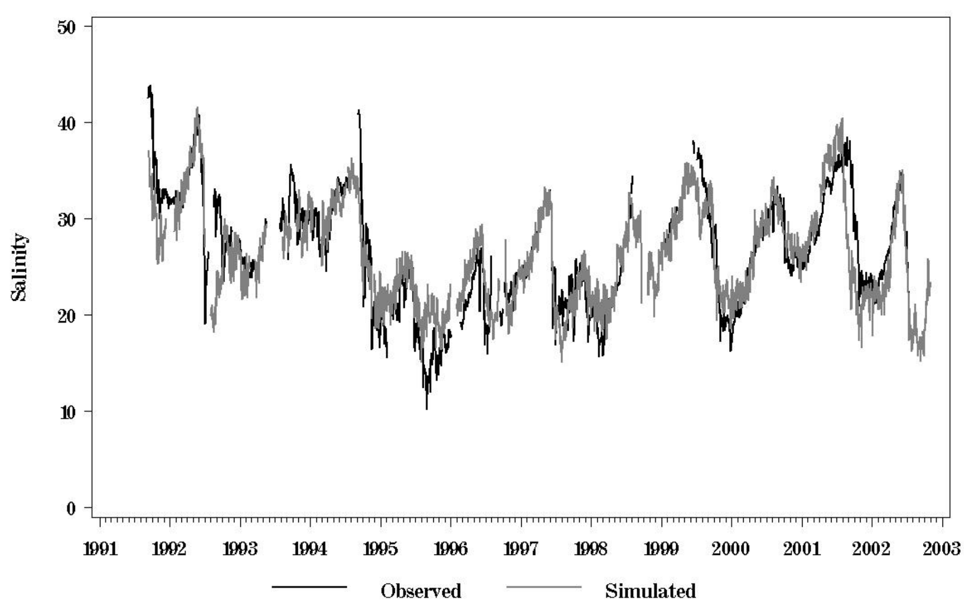
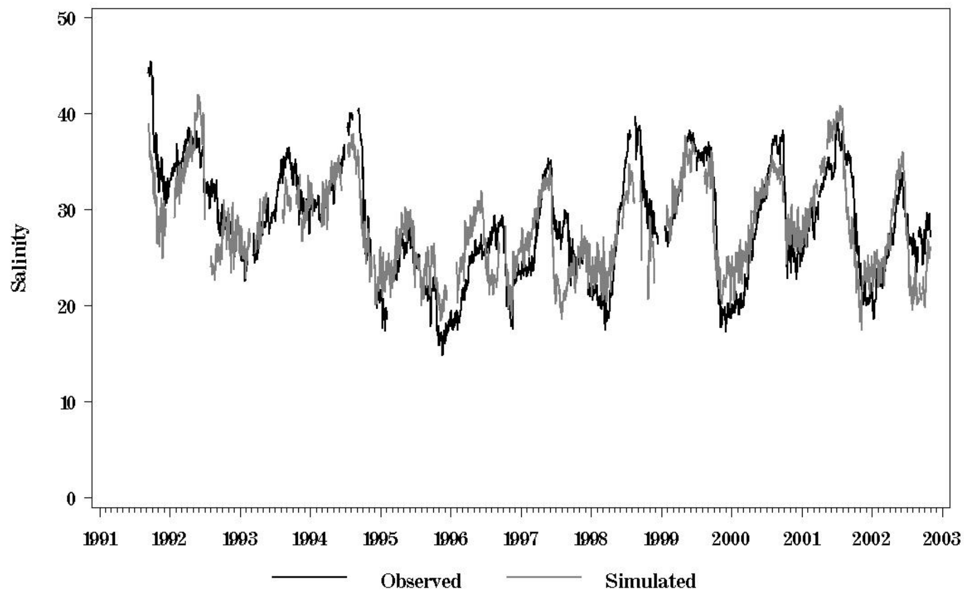


Figure 18. Comparison of Observed and Simulated Data for the Butternut Key Extended Period MLR Model – Calibration is September 12, 1991 – October 31, 2001; Verification is November 1, 2001 – October 31, 2002.



VIII. Coupling the 2X2 Model and MLR Salinity Models for Salinity Simulations

The ICU evaluations for salinity performance measures will include the analysis of simulations made using the MLR salinity models coupled with output from the 2X2 model for Everglades water level, and historic wind and sea level data. At the time of preparation of this report, the only 2X2 model ICU output available on the SFWMM (2X2) model version 5.0 website (accessed through www.evergladesplan.org) are 95 Restudy, 2000 CERP, NSM 4.5, and NSM 4.6. These 31- and 36-year stage simulations were used to simulate salinity at the following stations:

1. Joe Bay
2. Little Madeira Bay
3. Terrapin Bay
4. Whipray Basin
5. Duck Key
6. Butternut Key
7. Taylor River
8. Highway Creek
9. Little Blackwater Sound
10. Bob Allen Key
11. North River
12. Long Sound.

Before using the 2X2 model stage data for salinity simulations, the calibration / verification run was evaluated to determine how well the most recent update to the 2X2 model simulated the stage records. The average value of the 2X2 model stage estimates was compared to the average value of the observed time series for the 2X2 model verification period of January 1, 1996 through December 31, 2000 to compute the bias of the 2X2 model data. Then the 2X2 model data were adjusted by adding or subtracting the value of the bias (as appropriate) before being used for salinity simulations in the MLR salinity models. Comparison plots for each stage station are presented in Appendix B. These figures show that the 2X2 model stage data is very close to the observed data at some stations (P33 and EVER4 are examples) but not very close at others (P35 and R127 are examples). At R127 the deviation across the time series is almost perfectly constant as evidenced in a Pearson's correlation coefficient of almost 1.0 between the observed and 2X2 model data, though the 2X2 model data are offset from the observed by about 1.7 feet, which is possibly a datum problem. At most locations the deviation is similar through time, while at other locations the deviation varies over the time series. Therefore, it can be expected that the adjustment of the 2X2 model data using the bias in the 2X2 model data will improve the simulative capability for some of the 2X2 model series' more than others.

The plots of the simulations made with adjusted 2X2 model output for the 95 Restudy, 2000 CERP, NSM 4.5, and NSM 4.6, and historical wind and sea level data are presented in the Appendices. In a similar manner, the Southern Estuaries Sub-team work will produce new models for Garfield Bight, Shark River Slough, North River, Whitewater Bay, and Manatee Bay/Barnes Sound. The simulations presented herein are intended to be used as appropriate with the simulations produced by the new models, to compare the various operational scenarios, and to assist in the interpretation of performance measures.

IX. Model Error Statistics

The ability of the MLR salinity models to simulate the observed conditions can be evaluated using a number of error statistics. For this project, the statistics that were computed are described below.

1. Mean Square Error

The Mean Square Error, or MSE, is defined as the square of the mean of the squares of all the errors, as follows:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(O-P)^{2}$$

2. Root Mean Square Error

The Root Mean Square Error is defined as:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(O^{(n)} - P^{(n)}\right)^2}$$

The Root Mean Square Error is a weighted measure of the error where the largest deviations between observed and predicted values contribute most to this uncertainty statistic. This statistic has units that are the same as the observed and predicted values. It is thought to be the most rigorous tests of absolute error (Hamrick, 2003).

3. Adjusted – R^2

The Coefficient of Multiple Determination (R^2) is the most common measure of the explanatory capability of a model. It is defined as:

R^2 = Sum of Squares Regression/Sum of Squares Total, or
= 1- (Sum of Squares Error/Sum of Squares Total)

R^2 measures the percentage reduction in the total variation of the dependent variable associated with the use of the set of independent variables that comprise the model (Neter, et al; 1990). When there are many variables in the model, it is common to use the Adjusted Coefficient of Multiple Determination, which is R^2 divided by the associated degrees of freedom.

4. Mean Error

The Mean Error is another measure of model uncertainty. It is defined as:

$$ME = \frac{1}{N}\sum_{n=1}^{N}\left(O^{(n)} - P^{(n)}\right)$$

where O=observed values, P=predicted values, and N= number of observations used to develop the model. Positive values of the mean error indicate that the model tends to over-predict, and negative values indicated that the model tends to under-predict (Hamrick, 2003.)

5. Mean Absolute Error

The Mean Absolute Error is defined as:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|O^{(n)} - P^{(n)}\right|$$

Although the Mean Absolute Error tells nothing about over- or under-prediction, it is considered as another measure of the agreement between observed values and predicted values. It is preferred by some because it tends to cancel the effects of negative and positive errors, and is therefore less forgiving compared to the Mean Error (Hamrick, 2003).

6. Maximum Absolute Error

The Maximum Absolute Error is defined as:

$$MAX = \max\left|O^{(n)} - P^{(n)}\right| \quad : \quad n = 1, N$$

The Maximum Absolute Error is the largest deviation between observed and predicted values.

7. Nash-Sutcliffe Efficiency

The Nash-Sutcliffe Efficiency is a measure of model performance that is similar to R^2. It was first proposed for use with models in 1970 (Nash and Sutcliffe, 1970). It is defined as:

$$NSE = 1 - \frac{\sum_{n=1}^{N}(P - O)}{\sum_{n=1}^{N}(O - \overline{O})}$$

The value of the NSE roughly corresponds to the percentage of variation that is explained by a model.

8. Relative Mean Error

Relative measures of error are not as extreme as the absolute measures presented above. Relative error statistics provide a measure of the error relative to the observed value. The Relative Mean Error is defined as:

$$RME = \frac{\sum_{n=1}^{N} O^{(n)} - P^{(n)}}{\sum_{n=1}^{N} O^{(n)}}$$

9. Relative Mean Absolute Error

The Relative Mean Absolute Error is defined as:

$$RMA = \frac{\sum_{n=1}^{N} \left|O^{(n)} - P^{(n)}\right|}{\sum_{n=1}^{N} O^{(n)}}$$

Caution must be applied in the use of these two statistics when there can be small values of the observed and predicted variable, and when they can have both positive and negative signs (Hamrick, 2003).

10. Relative Mean Square Error

The Relative Mean Square Error is not as prone to fouling by small values and/or the presence of both positive and negative values and is defined as (Hamrick, 2003):

$$RSE = \frac{\sum_{n=1}^{N}\left(O^{(n)} - P^{(n)}\right)^2}{\sum_{n=1}^{N}\left[\left(O^{(n)} - \overline{O}\right)^2 + \left(P^{(n)} - \overline{O}\right)^2\right]}$$

The Relative Mean Square Error has values between zero and one, with a model that predicts well having a Relative Mean Square Error close to zero. According to this measure, the most reliable models are the Whipray Basin and Bob Allen Key models, but all models are considered by this measure to be very reliable.

Table 2 presents a summary of the values of these statistics for the various models.

Table 2. Comparison of Model Uncertainty Statistics for MLR Salinity Models

station	mean sq error (mse), psu^2	root mse (rmse), psu	adj R-sq	mean error, psu	mean abs error, psu	max abs error, psu	Nash-Sutcliffe Efficiency
Joe Bay	25.8	5.1	0.75	-0.14	3.7	20.6	0.76
Little Madeira Bay	20.3	4.5	0.56	0.56	3.5	15.4	0.55
Terrapin Bay	32.6	5.7	0.75	-0.99	5.4	5.4	0.67
Whipray Basin	7.2	2.7	0.8	0.11	2.2	10.1	0.77
Duck Key	9.7	3.1	0.71	-0.18	2.27	14.4	0.71
Butternut Key	10.7	3.3	0.65	0.1	2.7	11.3	0.66
Long Sound	15	3.9	0.8	0.31	2.7	18.9	0.81
Taylor River	21.4	4.62	0.78	-0.49	3.6	22.9	0.78
Highway Creek	18.2	4.3	0.81	-0.95	3.7	17.7	0.76
Little Blackwater Sound	14	3.75	0.75	-0.14	2.9	15.7	0.76
North River	8.9	3.0	0.86	0.19	2.3	18.1	0.78
Bob Allen Key	7.2	2.7	0.79	0.3	2.1	9.2	0.81

Table 2., continued.

station	rel mean error	rel mean abs error	rel mean sq error
Joe Bay	0.012	0.32	0.14
Little Madeira Bay	0.027	0.18	0.29
Terrapin Bay	0.044	0.24	0.2
Whipray Basin	0.034	0.07	0.12
Duck Key	-0.007	0.09	0.17
Butternut Key	0.003	0.1	0.21
Long Sound	0.021	0.18	0.11
Taylor River	-0.06	0.47	0.13
Highway Creek	-0.08	0.31	0.14
Little Blackwater Sound	-0.007	0.16	0.14
North River	0.03	0.35	0.11
Bob Allen Key	0.01	0.065	0.12

When taken as a whole, these error statistics show that the MLR models are good to very good at simulating salinity values.

X. Presentations

A poster presentation was made at the joint conference of the Florida Bay Science Program and the Greater Everglades Ecosystem Restoration in April 2003. In this poster, SARIMA models and MLR models were discussed, including the reasoning behind the choice of MLR salinity models for the 2X2 model evaluations. The newly developed models with gradients included were presented, along with the Whipray Basin transfer model prepared from Joe Bay, Little Madeira Bay, and Terrapin Bay salinity.

A presentation was made by the PI to the Southern Estuaries Sub-team of RECOVER at their July 11, 2003 meeting on the progress that has been made with MLR salinity modeling.

On September 16-18, a poster was presented at the 2004 Estuarine Research Federation Meeting in Seattle, Washington, detailing the progress in the development of MLR salinity models and discussing the use of the models to simulate hypersaline conditions in Florida Bay.

On October 31, 2003, a presentation was made at the Estuarine Indicators Workshop at Sanibel Island, Florida. The current status of MLR salinity models was presented.

XI. Discussion

The second year CESI activities have shown that the MLR salinity models presented herein are capable of making reasonable and reliable simulations of salinity in the near-shore embayments, the mangrove zone, and the open water of Florida Bay over a wide range of hydrologic, meteorological, and sea level conditions. During this second year of the project, the models of the near-shore embayments and mangrove zone evolved into salinity relationships that have a physical basis in the parameters of the model, which are:

1. the variation of the elevation of the freshwater in the Everglades,
2. the variation in the elevation of sea level, and
3. the effect of wind direction and speed.

The MLR salinity modeling procedure relates them using a least squares method and step-wise regression for parameter selection at a significance level of 0.999, which is a very high threshold. The result is a weighting for each independent variable when used in combination with the other independent variables in a linear combination model.

The range of salinity measured at the MMN stations varies widely. At stations in the near-shore embayments and the mangrove zone (Joe Bay, Little Madeira Bay, Terrapin Bay, Garfield Bight, North River, Long Sound, Taylor River, Highway Creek, Little Blackwater Sound) the salinity varies between 0 psu and 35-55 psu. Depending on the location, the salinity may only approach 0 psu (Little Madeira Bay and Terrapin Bay), while at other near-shore locations the salinity remains at 0 psu for longer periods (weeks in the case of Highway Creek and Taylor River). At most locations the transition from high salinity values to low salinity values is more rapid (being described as “flashy” by some) than the transition from low to high salinity values.

To be applied with confidence MLR salinity models must be developed considering the physical phenomena that affect the salinity at a particular location and time in the estuary. The Everglades and the near-shore embayments of Florida Bay are a coastal aquifer system, with the fresh water body and the salt water body competing with each other as other factors (wind, evaporation, direct rainfall) act to reinforce the effects of one or the other or to provide mixing and translocation energy. Coastal aquifers have been studied in other estuarine areas of Florida. Pandit, et al (1991) studied the coastal aquifer and the interface of the surficial aquifer with the Indian River Lagoon in east central Florida. Beneath the barrier islands / peninsulas and coastal plain mainland of the Indian River Lagoon watershed the freshwater surficial aquifer is stored in porous sandy soils at a higher elevation than the saline lagoon or ocean. The location within the soil strata of the interface between the freshwater surficial aquifer and the denser saline water body can be mapped using the Ghyben-Herzberg principle (Pandit et al, 1991). According to this principle, if the elevation of the surficial aquifer is raised sufficiently by recharge, the interface will move towards the coast. If the elevation of sea level is raised, the interface moves towards the mainland (away from the coast) as the salt water intrudes. Because of the density difference between the salt water mass and the freshwater mass, a larger volume of freshwater (compared to salt water) is needed to cause the interface to move an equal distance. In addition, the interface between the surficial aquifer and the saltwater body is known to be a zone of salinity gradient that moves in response to climatic conditions. In south Florida, the U. S. Geological Survey (USGS) has prepared a ground and surface water model that simulates the regional aquifers. Information on the USGS model can be found on the SOFIA website.

In the Everglades, the water table (surficial aquifer) emerges above the ground most wet seasons and freshwater flows as sheet flow towards Florida Bay and the southwest coast of the Gulf of Mexico creating the unique ecosystems that exist in Everglades National Park. While overland flow has less resistance than flow through the substrate, the porous nature of the substrate means that freshwater is still flowing to the coast during the dry season, as evidenced by the continued decline in stage levels at all locations in the Everglades as the dry season evolves. During the dry season, evaporation may also be contributing to the decline of water levels.

Confined to a soil matrix, the interface zone is not affected by other factors that can affect a surface water body. In the absence of wind, direct rainfall, and evaporation, the change in salinity gradient over distance is large within the interface zone, and the “width” of the zone is relatively small. In an open estuary, the conditions are somewhat different. Direct rainfall can dilute the upper layer of the salt water body. Evaporation works in the opposite manner, reducing water mass in both bodies of water. Wind works to move the fluid water bodies in translocation fashion as well as to mix horizontally and vertically the different water masses. The result in estuarine surface water bodies is a relatively wide interface with a relatively gradual salinity change over distance. Because the conductivity probe at a monitoring station is fixed both vertically and horizontally, a particular observation in the near-shore embayments and mangrove zone may be the conductivity (salinity) of the freshwater lens, the interface zone, or the saltwater body, depending on the location of the interface zone. With the exception of the most upstream stations (North River, Highway Creek, and Taylor River), the salinity record shows that the most of the monitoring stations in the near-shore embayments are usually monitoring the salinity in the transition zone. Estuaries by definition are water bodies where a transition zone can exist. In the mangrove zone, the stations are measuring the salinity of the freshwater lens for a longer time.

Therefore, based on the coastal aquifer model presented above, it can be hypothesized that the salinity at a near-shore location is correlated in some manner to elevation of freshwater in the Everglades, sea level elevation, wind, and the watershed hydraulic gradient. Salinity may also be correlated to evaporation and direct rainfall, but those two parameters were not able to be investigated in this study. The results of the correlation analysis using the SARIMA correlation coefficient plot evaluation procedure (see Marshall, 2003) shows that the hypothesis is supported, in general, at the 95% significance level. What is meant by “in general” is that salinity at all locations was correlated with the stage measured at one or more stations in the Everglades, sea level, wind parameters, and hydraulic gradient in the Everglades. At all locations, lagged values of some of the independent variables were also correlated with salinity. However, the salinity at various locations was not always correlated to the same stage observations, but salinity was always correlated with stage in some manner. This means, there are a multitude of simple (univariate) linear regression models that could be prepared from this dataset with a wide range of R^2 values. Substantially improved models can be prepared by taking advantage of cross-correlation relationships in the stage data, and by including wind and sea level.

In this study more than one correlated independent variable is used to improve the fit of the models. The decision as to which independent variable to include in the models was, at first, left to the canned step-wise regression process. The step-wise regression procedure in SAS© begins by evaluating all independent variables that were identified as candidate variables from the correlation evaluation, and then a simple linear regression model is prepared from the candidate variable that produces a model with the highest R^2 value. Then the other candidate independent variables are added to the model one by one. For an independent variable to be kept in the model it must be significant at the 0.999 level using an F-test. If not, that independent variable is dropped from further consideration. Using a lower significance level than 0.999 resulted in models with many independent variables. Setting the selection threshold this high ensured that the parameters in the final models are as highly significant as possible, and reduces the overall Type I error rate.

If this canned step-wise regression process is left to its own devices, it may choose some independent variables for the model that are not physically defensible, but were selected by the program because of statistically advantageous cross-correlation relationships. In particular, there were cases where the step-wise procedure kept a stage variable that would seemingly be increasing salinity with increasing stage, i.e. the stage variable would be in the model with a + sign instead of a – sign. Therefore, the step-wise procedure was modified by eliminating the stage terms that were positive, or sea level terms that were negative.

It was at this point in developing the procedure for building MLR salinity models that the IOP evaluation began, and models using the above described procedure from this CESI project were prepared and used. After the IOP work was completed, there were some comments received regarding the inclusion in some models of stage stations that were not thought to influence salinity at a certain location. The best example of this is the North River IOP model, which includes stage at Craighead Pond (CP). Even though there is a correlative relationship between CP stage and North River salinity, a cause-and-effect relationship is not thought to be possible, because they are many miles apart and there are physical barriers that will not allow a simple raising of the CP elevation (such as by local rainfall) to decrease the salinity at North River. The Southern Estuaries Sub-team will be modifying the IOP model of North River so that it includes primarily stage stations that are directly upstream of the station in Shark River Slough.

Additionally, many of the models include P33, which is in Shark River Slough. The presence of P33 in the models is seen as an expression of the regional nature of the hydrology of the Everglades. There is known to be connectivity between Shark River Slough and Taylor Slough, though there is thought to be less connectivity between the water in Shark River Slough and the mangrove zone of the eastern panhandle of ENP. All models also include wind parameters, and most include sea level. Therefore, the new models for Taylor River, Highway Creek, and Little Blackwater Sound were developed using the concept of a combination of variables that represent the regional hydrology, the local hydrology, wind and sea level elevation, using the step-wise selection method to choose the appropriate independent variables at the 0.999 significance level.

At the open-water stations (Whipray Basin, Duck Key, Butternut Key, and Bob Allen Key), the salinity rarely drops below 20 psu and frequently reaches above 40 psu. The relationship between salinity at the open-water stations and the hydrology of the Everglades is weaker than at the near-shore areas. However, there is a very strong relationship between the salinity in the near-shore embayments and the salinity at the open-water stations. The relationship is particularly strong for Terrapin Bay and Little Madeira Bay. Therefore, the MLR salinity models for the open-water stations were developed using the salinity at Terrapin Bay and Little Madeira Bay along with sea level elevation and wind factors. This means that simulation of open-water salinity is made using a two-step transfer function relationship (using Everglades stage to estimate Little Maderia Bay and Terrapin Bays salinities, then using these salinities to estimate open-water salinity). When Pearson correlation coefficients were compared, the two-step process was found to provide a better simulation than open-water simulations from only Everglades water levels. Therefore, the use of another transfer function to simulate open-water salinity provides improved predictions.

In terms of deliverables, this second year activity produced the following products:

1. A set of salinity models that were developed from 2X2 model output that became obsolete when the 2X2 model was re-calibrated;
2. New MLR salinity models for Taylor River, Highway Creek, Little Blackwater Sound, and Bob Allen Key;
3. New models using an extended period of data for Little Madeira Bay, Terrapin Bay, Whipray Basin, Duck Key, and Butternut Key;
4. 31- or 36-year simulations of salinity at Joe Bay, Little Madeira Bay, Terrapin Bay, Whipray Basin, Duck Key, Long Sound, Butternut Key, Taylor River, Highway Creek, Little Blackwater Sound, and Bob Allen Key for use with ICU evaluations; and
5. A detailed uncertainty analysis.

From the models that became obsolete it was learned that the improvement to fit using 2X2 model data for model development was not worth the benefit when the 2X2 model was unknowingly re-calibrated, an exercise that is likely to happen again in the future.

The new models that were developed for Taylor River, Highway Creek, Little Blackwater Sound, and Bob Allen Key have been added to the IOP models for Joe Bay, Little Madeira Bay, Terrapin Bay, Long Sound, Whipray Basin, Duck Key, and Butternut Key for the ICU evaluations that will be performed by the Southern Estuaries Sub-team of RECOVER. A revised model for North River and (perhaps) Joe Bay will be developed by the USACOE, along with new models for Barnes Sound / Manatee Bay, Garfield Bight, Whitewater Bay, Shark River Slough, and one other station not yet identified (Buckingham, per. com.). The 31- or 36-year simulations in this CESI report will be added to new simulations for four other water management scenarios (D13R Restudy, D13R CERP, 2050 CERP, 2050 Restudy) and used directly for ICU performance measure evaluations. New 31- or 36-year simulations will be generated by the USACOE contractor for the new models that they will be developing.

Therefore, this CESI project has significantly extended the spatial coverage for the ICU evaluations from five near-shore stations (Joe Bay, Little Madeira Bay, Terrapin Bay, Garfield Bight, and North River) to include four open-water stations (Whipray Basin, Duck Key, Butternut Key, and Bob Allen Key), another important near-shore station (Little Blackwater Sound), and two stations in the mangrove zone (Taylor River and Highway Creek). This will allow a more spatially comprehensive look at the operational scenarios, particularly as it relates to the potential for hypersaline events in the open-water areas of Florida Bay.

For the extended period of data the Little Madeira Bay model differed more from the IOP model than the other stations. The R^2 value for the extended period Little Madeira Bay model (0.65) is a substantial improvement over the R^2 value for the IOP model (0.56). For Terrapin Bay, the extended period model coefficients are closer to the IOP model than for Little Madeira Bay, and the R^2 value is less (0.71) than the R^2 value for the IOP model (0.76). The Whipray Basin extended period model also had a slightly improved R^2 value, but the values for the Duck Key and Butternut Key extended period models, and the model coefficients were virtually the same as the IOP models. Therefore, this evaluation of the use of the extended period to produce better models is inconclusive. However, at all stations that the range of the data used for model development has been extended, and there is greater confidence in the use of the models to estimate hypersaline conditions because the model development data included a large number of high salinity values.

The uncertainty evaluation produced a number of measures of the ability of the new models developed by this study to simulate salinity. The error statistics show that all of the MLR salinity models are good to very good in their ability to simulate salinity. The highest relative mean error for the MLR salinity models developed by this study is about 3.5%, but there are several models with absolute relative mean error higher than 25%. However, the relative absolute mean error is known to be highly affected when observed and predicted quantities can have small values or values that have both positive and negative signs, as is the case with the salinity in the near-shore waters and the mangrove zone. It is the models for the stations in these areas that that have the highest values of the relative absolute mean error.

All models also have a relative mean square error of close to 0, which means that the models are rated as highly skilled according to this statistic, except for the Little Madeira Bay model. In general, the uncertainty analysis shows that only the Little Madeira Bay model would benefit from additional model modifications using the initial data set. The extended period Little Madeira Bay model is a substantial improvement over the IOP model, and should be considered for use with the ICU evaluations.

XII. Summary and Findings

This second year CESI project has evolved as the project period passed, and the end product was a broader and more accurate set of MLR salinity models with a better handle on model uncertainty than was possible with the original model scope. The IOP evaluation for the Congressional Report provided an important opportunity to use the initial second year project tasks in an application mode, which accomplished some of the original objectives for the project. In turn this allowed additional salinity models to be developed and simulations to be prepared which can now be used to provide a wider spatial analysis in the ICU evaluations.

The following findings were produced by this project:

1. MLR salinity models that can reasonably simulate observed conditions at a daily time step can be prepared from observed daily average values of Everglades water levels, average daily sea level elevation, and average daily wind speed and direction.
2. Other factors that were identified as potentially able to explain some of the remaining error are evaporation and direct rainfall.
3. Craighead Pond is the stage station that shows up most frequently in the MLR salinity models. The next most frequent stage station in model development is P33.
4. A modification of the canned SAS© step-wise linear regression procedure for parameter selection is needed to produce models with a physically defensible basis.
5. A hydrologic model that shows promise in application to the Everglades / Florida Bay system is the coastal aquifer model with fresh water and salt water masses and an interface zone that can be affected by wind and LOCAL evaporation factors.
6. The Little Madeira Bay MLR model prepared from an extended period of data that included data from the severe drought conditions in the late 1980's and early 1990's was improved as measured by the Adjusted-R^2 value, compared to the model prepared for the IOP evaluation that began in 1995. For the other stations, there was minimal improvement using the extended period data.
7. According to a number of error statistics, the MLR salinity models prepared for this study can be considered as good to excellent for simulation purposes. However, a relatively high maximum absolute error for most of the models may mean that a measure of local evaporation and rainfall would improve this statistic and further improve the MLR salinity models.
8. The simulations produced by this project are ready to be used for the ICU evaluations in a much broader fashion than was envisioned when this project began.

This work has shown that statistical models can be used when there is not enough physical data to develop and implement detailed hydrodynamic models. In the case of the Everglades and Florida Bay, not enough is known about the distribution of freshwater flows into the Bay outside of the gauged creeks. Because of this and the lack of detailed bathymetry needed to accurately describe the banks within the bay, a “traditional” hydrodynamic model has not yet been developed that can adequately simulate salinity. Therefore, MLR salinity models can be used in place of hydrodynamic models for analysis of water management alternatives, and to serve as input to ecological models as the next step of MLR salinity model utilization. MLR salinity models are relatively easy and inexpensive to develop and utilize, and the concept of linear regression and analysis of variance are familiar to most scientists and engineers. In the future, the more sophisticated SARIMA models may be able to be used for system control using the very accurate one-step forward predictions that can be made by these models, so that changes to water delivery patterns can be evaluated in real-time.

The strength of the MLR salinity models presented herein is that they are physically based, they have shown to be adaptable to changes in hydrology by simulating salinity well for both wet and dry conditions (including severe drought), and they are capable of being used to evaluate different flow scenarios by simulating salinity conditions from a given set of input data (in this case 2X2 model data) and historical climatology. The error statistics show that the models simulate good to very good, though there is an occasional large residual.

However, there is work that still remains to be done on the MLR salinity models. For example, a surrogate for evaporation is needed to determine if a measure of evaporation will improve the simulative capability of the models. When it is considered how the MLR salinity models are being used (i.e. with 2X2 model output over 31- and 36-year periods), the surrogate must be a quantity that has been continuously measured since 1965. The only set of data that fits that requirement is the meteorological data available from the National Weather Service, and the National Ocean Service records. Fortunately, parameters that are used to estimate evaporation such air temperature and sea water temperature are amongst the data that are available. Because the evaluation of extended period models was somewhat inconclusive, additional evaluation of model performance during wet and dry periods is also needed.

Now that it has been shown that MLR salinity models are capable of adequately simulating salinity, they can be coupled with ecological models at the daily, weekly, or monthly time scales. Though not shown in this report, simulated monthly average values track observed monthly average values very well, and the output from the daily MLR salinity models are easily transformed into weekly or monthly values, with the statistical power that comes from preparing simulations at the daily level and aggregating to develop less frequent simulations. Since the extended period models showed that the MLR salinity models were capable of simulating drought conditions as well as wet periods, MLR salinity models can be used with ecological models to evaluate the effect of hypersaline conditions. To assist with those evaluations, the preparation of MLR salinity models for the remaining MMN stations should be accomplished to increase the spatial resolution of the simulations.

XIII. Literature Cited

Everglades National Park. 1997a. Everglades National Park Marine Monitoring Network 1994 Data Summary. Everglades National Park, Homestead, Florida. 67 pp.

Everglades National Park. 1997b. Everglades National Park Marine Monitoring Network 1995 Data Summary. Everglades National Park, Homestead, Florida. 67 pp.

Hamrick, J.M. 2003. Florida Bay Hydrodynamic and Salinity Model Analysis. South Florida Water Management District. West Palm Beach, Florida. 20-22.

Kashigan, S.K. 1991. Multivariate Statistical Analysis. Radius Press. New York. 117-192.

Marshall III, F.E.; D. Smith; and D. Nickerson. 2003a. Salinity Simulation Models for North Florida Bay Everglades National Park. Cetacean Logic Foundation, Inc. New Smyrna Beach, Florida. 41 pp.

Marshall III, F.E. 2003b. IOP Salinity Analysis Using Statistical Models. Cetacean Logic Foundation, Inc. New Smyrna Beach, Florida. 35 pp.

Neter, J.; W. Wasserman; M. H. Kutner. 1990. Applied Linear Statistical Models. Richard D. Irwin, Inc. Boston, Massachusetts.

Nuttle, W.K. 2004. Wetland Hydrology and Estuarine Salinity Related to SFWMM Scenarios. The Cadmus Group. Watertown, Massachusetts. 14-17.

Smith, D. 1997. Everglades National Park Marine Monitoring Network 1996 Data Summary. Everglades National Park, Homestead, Florida. 94 pp.

Smith, D. 1998. Everglades National Park Marine Monitoring Network 1997 Data Summary. Everglades National Park, Homestead, Florida. 100 pp.

Smith, D. 1999. Everglades National Park Marine Monitoring Network 1998 Data Summary. Everglades National Park, Homestead, Florida. 100 pp.

Smith, D. 2001. Everglades National Park Marine Monitoring Network 1999 Data Summary. Everglades National Park, Homestead, Florida. 65 pp.

Pandit, A.; C. El-Khazen; S.P. Sivaramapilli. 1991. Estimation of Hydraulic Conductivity Values in a Coastal Aquifer. Groundwater. Vol. 29, No.2. pp 175-180.