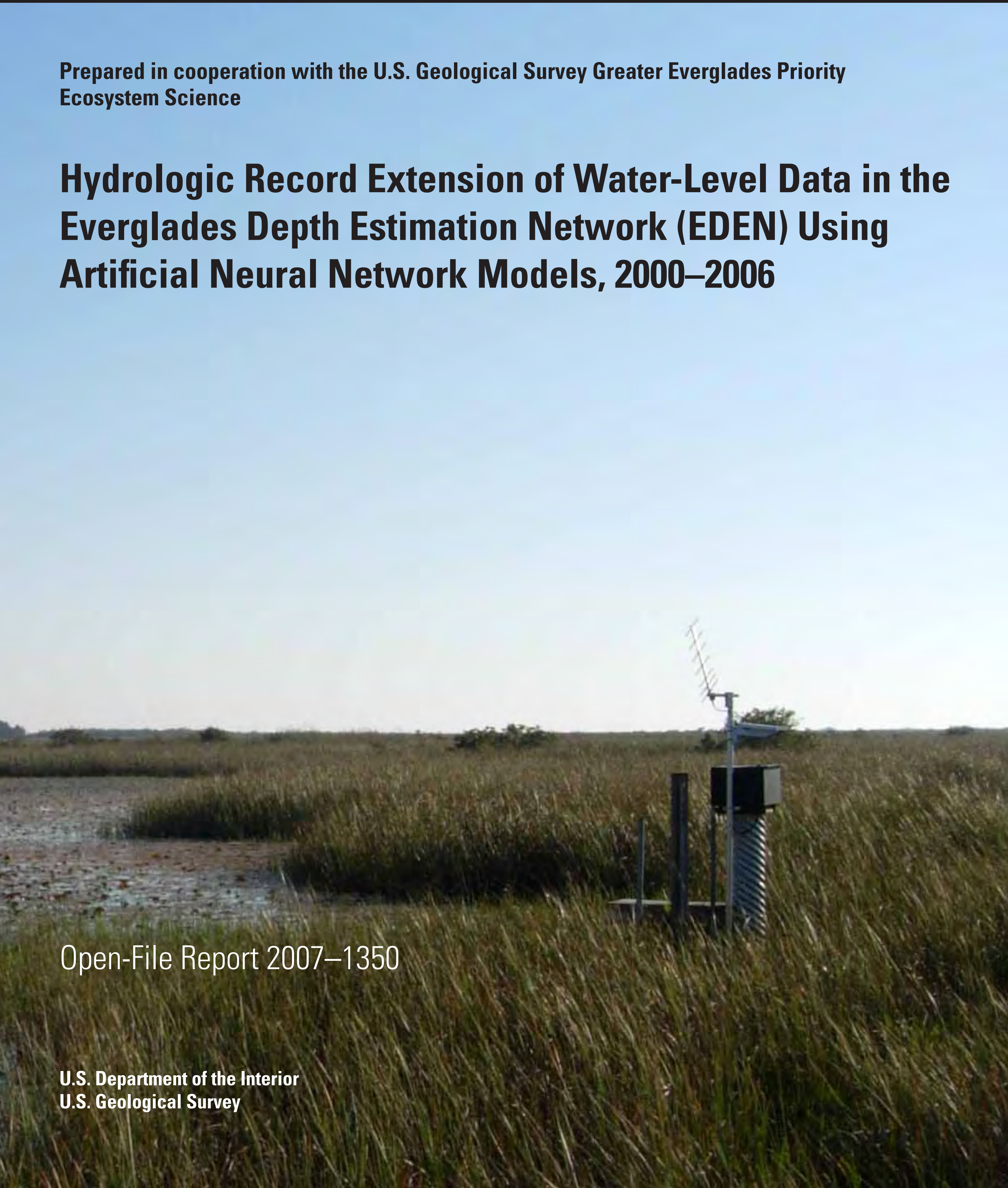
Prepared in cooperation with the U.S. Geological Survey Greater Everglades Priority Ecosystem Science

Hydrologic Record Extension of Water-Level Data in the Everglades Depth Estimation Network (EDEN) Using Artificial Neural Network Models, 2000–2006



Open-File Report 2007–1350

U.S. Department of the Interior
U.S. Geological Survey

Cover. **Gaging station W11 in Water Conservation Area 3A of the Florida Everglades** *(photograph taken by Christa Zweig, Graduate Research Assistant, Florida Cooperative Fish and Wildlife Research Unit, University of Florida)*.

Hydrologic Record Extension of Water-Level Data in the Everglades Depth Estimation Network (EDEN) Using Artificial Neural Network Models, 2000–2006

By Paul A. Conrads and Edwin A. Roehl, Jr.

Prepared in cooperation with the U.S. Geological Survey Greater Everglades Priority Ecosystems Science

Open-File Report 2007–1350

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia: 2007

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS

Suggested citation:
Conrads, P.A., and Roehl, E.A., Jr., 2007, Hydrologic record extension of water-level data in the Everglades Depth Estimation Network (EDEN) using artificial neural network models, 2000–2006: U.S. Geological Survey Open-File Report 2007-1350, 56 p. (only online at ***http://pubs.water.usgs.gov/ofr2007-1350***)

Contents

Figures

Tables

Conversion Factors and Datums

Inch/Pound to SI

Multiply	By	To obtain
	Length	
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
	Hydraulic gradient	
inch per mile (in/mi)	1.5783	centimeter per kilometer (cm/km)
foot per mile (ft/mi)	0.1894	meter per kilometer (m/km)
	Area	
square mile (mi^2)	2.590	square kilometer (km^2)

SI to Inch/Pound

Multiply	By	To obtain
	Length	
centimeter (cm)	0.3937	inch (in.)
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square meter (m^2)	10.76	square foot (ft^2)
square kilometer (km^2)	0.3861	square mile (mi^2)

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Elevation, as used in this report, refers to distance above the vertical datum.

Acronyms and abbreviations used in the report

ADM	Advanced Data Mining
ANN	artificial neural network
BCNP	Big Cypress National Preserve
BEP	back error propagation
CERP	Comprehensive Everglades Restoration Plan
CRADA	Cooperative Research and Development Agreement
EDEN	Everglades Depth Estimation Network
ENP	Everglades National Park
GIS	geographic information system
HLN	hidden layer neuron
ME	mean error
MLP	multilayer perceptron
MSE	mean square error
OLS	ordinary least squares
PME	percent model error
RBF	radial basis function
RMSE	root mean square error
R^2	coefficient of determination
SFWMD	South Florida Water Management District
SOFIA	South Florida Information Access
USGS	U.S. Geological Survey
WCA	water conservation area
WL	water level

Hydrologic Record Extension of Water-Level Data in the Everglades Depth Estimation Network (EDEN) Using Artificial Neural Network Models, 2000–2006

By Paul A. Conrads and Edwin A. Roehl, Jr.[1]

Abstract

The Everglades Depth Estimation Network (EDEN) is an integrated network of real-time water-level gaging stations, ground-elevation models, and water-surface models designed to provide scientists, engineers, and water-resource managers with current (2000–present) water-depth information for the entire freshwater portion of the greater Everglades. The U.S. Geological Survey Greater Everglades Priority Ecosystem Science provides support for EDEN and the goal of providing quality assured monitoring data for the U.S. Army Corps of Engineers Comprehensive Everglades Restoration Plan. To increase the accuracy of the water-surface models, 25 real-time water-level gaging stations were added to the network of 253 established water-level gaging stations. To incorporate the data from the newly added stations to the 7-year EDEN database in the greater Everglades, the short-term water-level records (generally less than 1 year) needed to be simulated back in time (hindcasted) to be concurrent with data from the established gaging stations in the database. A three-step modeling approach using artificial neural network models was used to estimate the water levels at the new stations. The artificial neural network models used static variables that represent the gaging station location and percent vegetation in addition to dynamic variables that represent water-level data from the established EDEN gaging stations. The final step of the modeling approach was to simulate the computed error of the initial estimate to increase the accuracy of the final water-level estimate.

The three-step modeling approach for estimating water levels at the new EDEN gaging stations produced satisfactory results. The coefficients of determination (R^2) for 21 of the 25 estimates were greater than 0.95, and all of the estimates (25 of 25) were greater than 0.82. The model estimates showed good agreement with the measured data. For some new EDEN stations with limited measured data, the record extension (hindcasts) included periods beyond the range of the data used to train the artificial neural network models. The comparison of the hindcasts with long-term water-level data proximal to the new EDEN gaging stations indicated that the water-level estimates were reasonable. The percent model error (root mean square error divided by the range of the measured data) was less than 6 percent, and for the majority of stations (20 of 25), the percent model error was less than 1 percent.

[1]Advanced Data Mining, LLC, Greenville, South Carolina.

Introduction

The Everglades Depth Estimation Network (EDEN) is an integrated network of real-time water-level gaging stations, ground-elevation models, and water-surface models designed to provide scientists, engineers, and water-resource managers with current (2000–present) water-depth information for the entire freshwater portion of the greater Everglades (Telis, 2005, 2006). EDEN is presented on a 400-square-meter (m^2) grid, and EDEN offers a consistent and documented dataset that can be used by scientists and managers to (1) guide large-scale field operations, (2) integrate hydrologic and ecological responses, and (3) support biological and ecological assessments that measure ecosystem responses to the Comprehensive Everglades Restoration Plan (CERP; U.S. Army Corps of Engineers, 1999). The target users of EDEN are biologists and ecologists who can use the information to examine trophic level responses to hydrodynamic changes in the Everglades. The EDEN database establishes a 7-year dataset of baseline conditions prior to the implementation of the CERP that offers investigators a single repository for historic hourly water-level data.

To estimate water depths in the greater Everglades, geographic information system (GIS) models have been developed to determine the ground elevation and water-surface elevation for the freshwater portion of the Everglades. The water-depth estimates are the differences between the two surfaces. Data to support the ground-elevation model include elevation measurements at over 50,000 sites (Desmond, 2003). Data to support the water-surface model include continuous water levels at 253 stations, including 25 stations that were added to the EDEN database in 2006 (fig. 1).

For the development of the ground-elevation model (Jones and Price, 2007), the EDEN domain was divided into a large number of equal-sized squares (“cells”) that in total are referred to as the “grid.” The grid includes information on the characteristics of each cell, such as centroid location, the area of the Everglades it represents, elevation, and percent vegetation type (slough, prairie, sawgrass, upland, exotic, and other). The large number of highly accurate elevation data allows for refinement of the ground-elevation model. The geostatistical technique of kriging was selected for the EDEN ground-

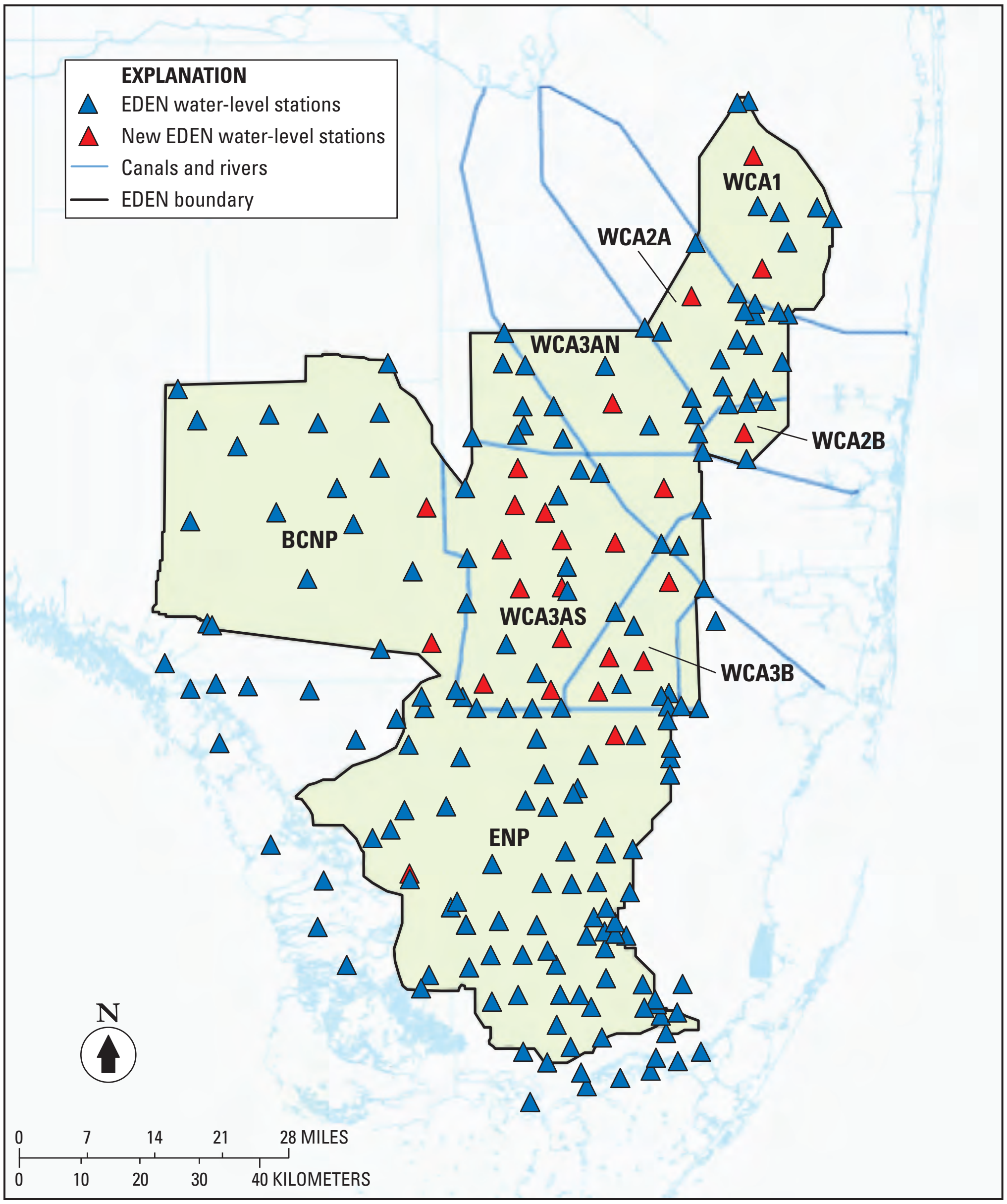


Figure 1. Locations of Everglades Depth Estimation Network (EDEN) gaging stations in southern Florida (modified from Pearlstine and others, 2007).

elevation model following extensive testing of multiple interpolation techniques. Kriging produced the lowest average error for validating elevation points and provides useful diagnostic surfaces. To account for variations within subregions of the EDEN area, individual geostatistical models were created for each water conservation area (WCA), the Everglades National Park (ENP), and portions of Big Cypress National Preserve (BCNP). These individual models were combined to create a single, 400-m^2 resolution elevation model for the entire EDEN domain (fig. 2).

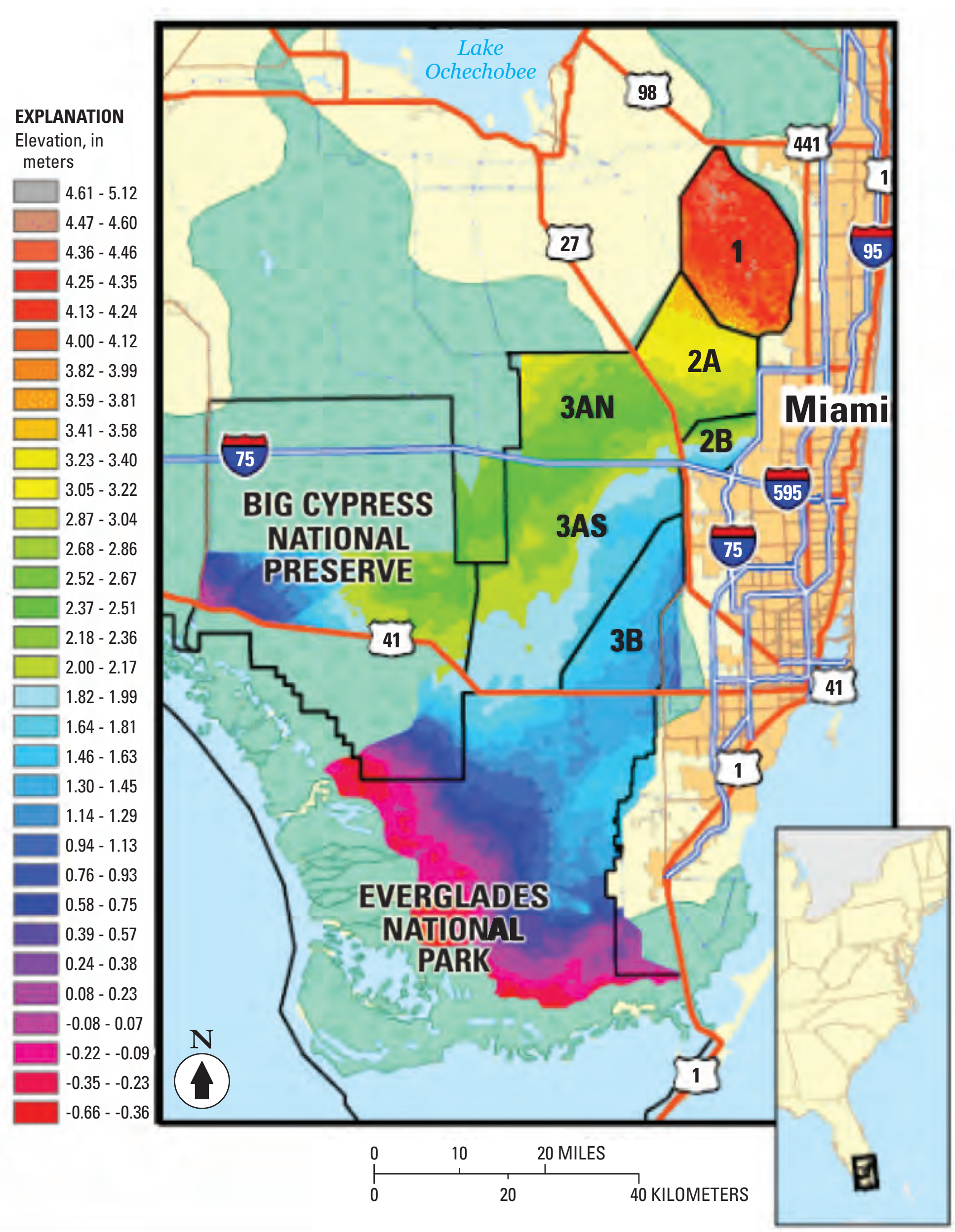


Figure 2. The Everglades Depth Estimation Network (EDEN) digital elevation model for Water Conservation Areas (1, 2A, 2B, 3AN, 3AS, 3B), Big Cypress National Preserve, and Everglades National Park (from Pearlstine and others, 2007).

A water-surface model was developed in GIS using the EDEN grid described above. The EDEN water-surface model interpolates measured water levels from the EDEN continuous monitoring network to ungaged locations using radial basis functions (RBF) with multiquadric regression (Pearlstine and others, 2007). The model produces a continuous water surface for any day within the period of record in the EDEN database. An example of the water surface for a sample day is shown in figure 3.

Twenty-five stations were added to EDEN in 2006 to address water-level data gaps identified by scientists and hydrologists in the original EDEN configuration and to improve the continuous water-surface model. To incorporate the additional data into the EDEN database, the water-level records for the new stations needed to be extended back in time (hindcasted) to be concurrent with the records in the EDEN database. The U.S. Geological Survey (USGS) South Carolina Water Science Center, as part of the EDEN project team, developed artificial neural network (ANN) models to hindcast data from the new EDEN stations. An ANN model is a flexible mathematical structure capable of describing complex nonlinear relations between input and output datasets. The architecture of ANN models is loosely based on the biological nervous system (Hinton, 1992).

Accurately hindcasting the hydrologic responses at the new locations can be challenging due to the limited number of reference gaging stations and a limited understanding of complex interactions between hydrology and topography. Techniques that often are used to hindcast hydrologic responses at ungaged locations include combinations of linear regression and interpolation; however, the dynamics between hydrology, topography, and vegetation often are nonlinear. This report presents the application of cascading ANN models to predict water levels at 23 new EDEN stations that were instrumented in 2006 and 2 EDEN stations in WCA1 with periods of record beginning in 2001. The ANN models were used to extend the 25 stations to be concurrent with the EDEN database beginning in January 2000.

To meet the objectives of this study and previous studies, the USGS entered into a Cooperative Research and Development Agreement (CRADA) with Advanced Data Mining (ADM) in 2002 to collaborate on applying data mining and ANN models to water-resources investigations. The emerging field of data mining addresses the issue of extracting information from large databases (Weiss and Indurkhya, 1998). Data mining is a powerful tool for converting large databases into knowledge for solving problems that are otherwise imponderable because of the large numbers of explanatory variables or poorly understood process physics. This knowledge encompasses understanding cause and effect relations and predicting the consequences of alternative actions. Data-mining methods come from different technical fields, such as signal processing, statistics, artificial intelligence, and advanced visualization. Data mining is used extensively in financial services, banking, advertising, manufacturing, and e-commerce to classify the behaviors of organizations and individuals and to predict future outcomes.

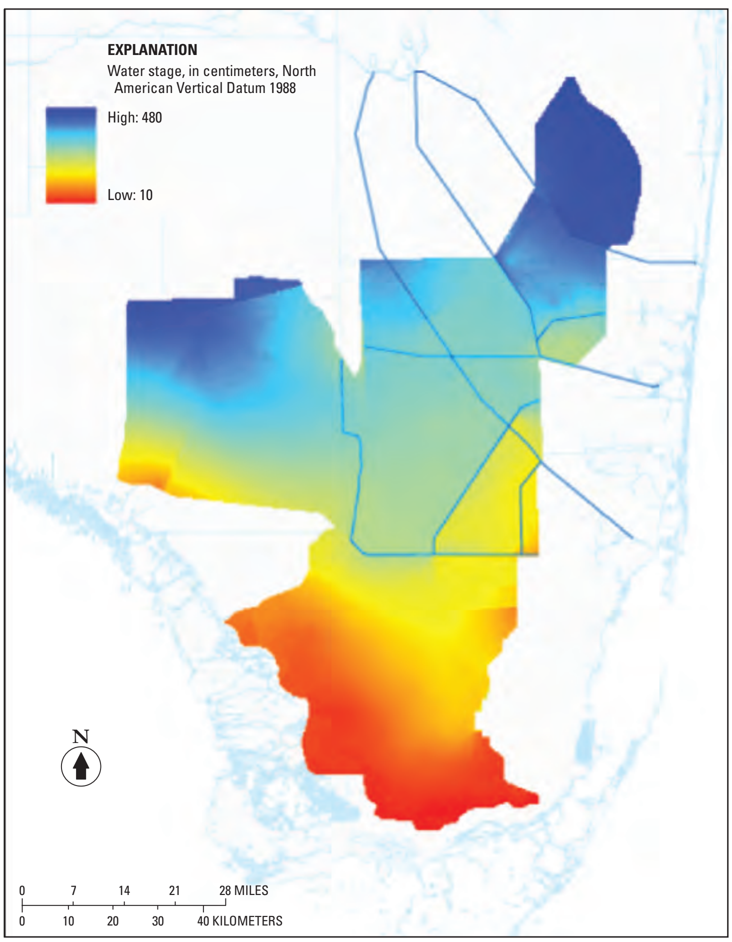


Figure 3. The Everglades Depth Estimation Network (EDEN) water surface model for September 10, 2006 (modified from Pearlstine and others, 2007).

Purpose and Scope

This report documents the water-level record extensions (hindcasts) of 25 stations in the freshwater portion of the Everglades. The geographical extent of the hindcasts includes gaging stations in WCA1, WCA2, WCA3A, WCA3B, BCNP, and the ENP. An important part of the USGS mission is to provide scientific information for the effective water-resources management of the Nation. To assess the quantity and quality of the Nation's surface water, the USGS collects hydrologic and water-quality data from rivers, lakes, estuaries, and wetlands using standardized methods, and maintains the data from these stations in a national database. The techniques presented in this report demonstrate how valuable information can be extracted from existing databases to assist local, state, and Federal agencies. The application of data-mining techniques, including ANN models, demonstrates how empirical models can be developed to hindcast time series of complex hydrologic systems and how disparate databases can be integrated.

Description of Study Area

The study area is within the greater Everglades area, which extends from south of Lake Okeechobee to the southern part of the ENP (fig. 4) This area is a wetlands system that is about 50 miles (mi) wide and about 100 mi long. The Everglades is regarded as unique in the world because it is not primarily associated with a natural river system but is itself a wide and shallow "river" that transports water by sheet flow from Lake Okeechobee to the Gulf of Mexico. The slopes with this shallow "river" are generally less than about 0.2 foot per mile (ft/mi; German, 2000).

The Everglades contains several types of environments, including freshwater marshes, tree islands, pinelands, mangrove swamps, and shallow coastal marine waters. This study is concerned with freshwater marshes, the predominant Everglades ecosystem. These marshes are characterized by sawgrass stands of varying density and height, ranging from 2 to 3 feet (ft) above land surface to 9 ft in some northern areas. Other common emergent plants in the freshwater marshes include spike rush, muhly grass, and, in some areas, cattails. Typical topographic and vegetative features include ridge and slough, tree islands, wet prairie, sawgrass, and marl prairie (German, 2000).

The annual rainfall in the Everglades generally is between 50 and 60 inches (in.), depending on location, with substantially more rainfall along the eastern edge (Lodge, 1994). The rainfall has a distinct seasonal pattern, with a wet season from May or June through September or October, that accounts for about 75 percent of the annual total. Water depths in the freshwater marshes range from 0 to 3 ft during the wet season. Minimum seasonal water levels generally occur in May before onset of the wet season. In particularly dry years, large portions of the Everglades may become dry and subject to wildfires. Heavy rainfall associated with tropical depressions, storms, and hurricanes can have a large effect on water levels. A single such event can increase water levels by a foot or more over large parts of the Everglades and because of the slow runoff rates, this can effect water levels for months (German, 2000).

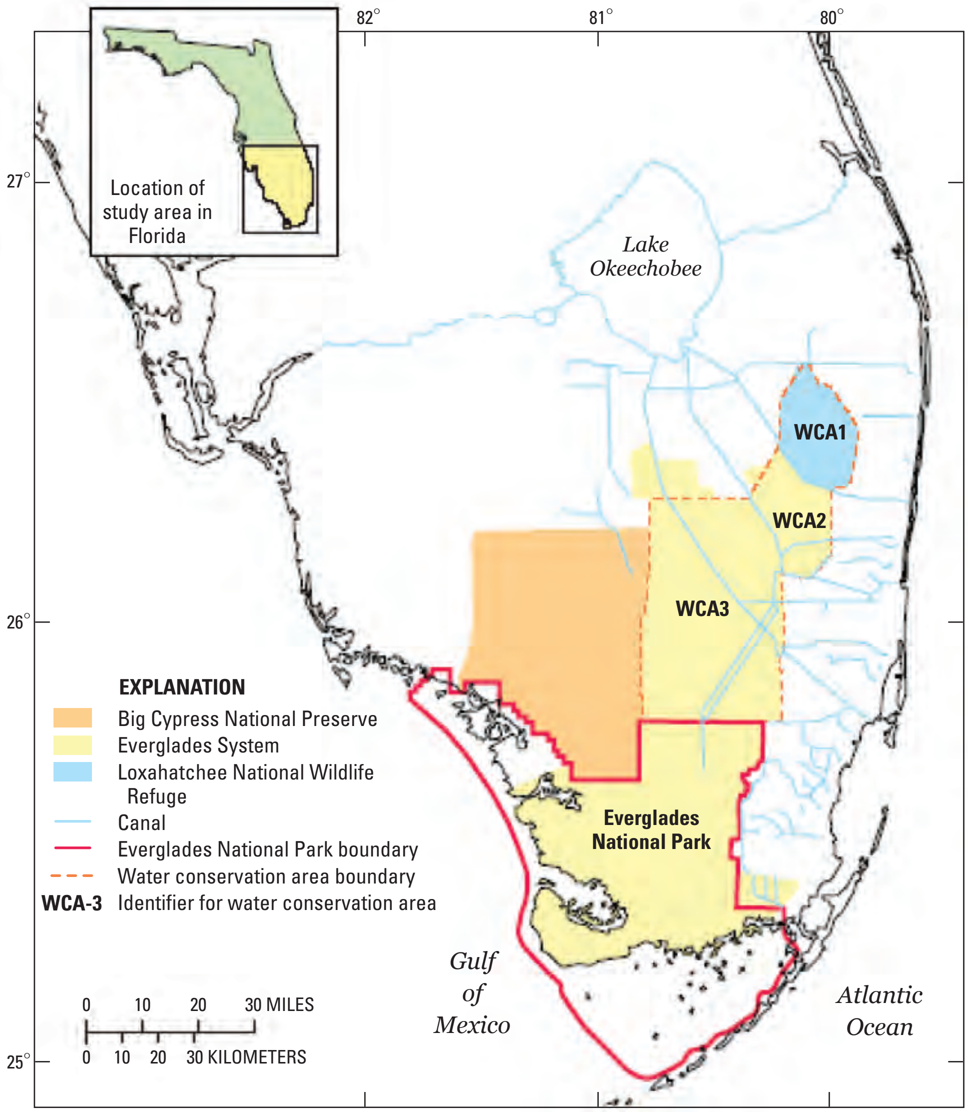


Figure 4. Greater Everglades, Florida (modified from German, 2000).

Previous Studies

Estimating hydrologic and water-quality conditions at ungaged sites by using data-mining techniques and ANN models has been used in the Everglades, western Oregon, and Wisconsin. Conrads and others (2006) used data-mining techniques in a snail kite study in WCA3A to hindcast short-term (less than 5 years) daily water-depth datasets using long-term (greater than 15 years) daily water-level data. The principal objective of the snail kite study was to separate plant community response resulting from typical seasonal and inter-annual variances in hydrologic regimes. The vegetative community structure of these sites is an expression of present and historic hydrologic conditions. A critical element of the study was to determine how the vegetative communities respond to temporal and spatial changes in hydrology. Artificial neural network models were developed to hindcast hydrologic histories at 17 transects dating back to 1991 to help ecologists in analyzing the water depth and hydroperiods over a large range of hydrologic conditions and to integrate long-term ecological data.

Artificial neural network models have been applied in western Oregon to estimate stream temperature at ungaged sites (Risley and others, 2003). In that study, a dynamic clustering technique (Roehl and others, 2006) was used to subset 142 temperature stations from first-, second-, and third-order streams into three groups of similar dynamic behaviors. Using categorical (static) variables and time-series variables, water-temperature models were developed for ungaged sites. Critical input variables included riparian shade, station elevation, and percentage of forested area of the basin. Coefficients of determination (R^2) and root mean square errors (RMSE) for the models ranged from 0.88 to 0.99 and 0.05 to 0.59 degrees Celsius (°C), respectively.

Stewart and others (2006) describe a modeling application in Wisconsin that, like the Oregon modeling, predicts stream temperatures from climate signals, such as ambient temperature and rainfall, and categorical station attributes, such as land cover and drainage area. The application's ANNs were trained on data, including summer-month daily average stream temperature time series, from 254 stations. This work was conceptually different from the Oregon application, which used concurrently measured signals. The Wisconsin stream temperatures were measured during different summers over 13 years. This required that an alternative time series clustering algorithm be developed that would still segment the signals according to their dynamic behaviors. The R^2 and RMSE for the predictions at 31 test stations not used in ANN training ranged from 0.62 to 0.75 and 1.7 °C to 2.4 °C, respectively. The ANN predictions accurately tracked the day-to-day variability at the different stations, but the primary source of error was offset station-to-station baseline (mean summer) temperatures. The predicted baselines depend largely on the ANN's categorical input variables, suggesting that the 42 variables used can only provide a partial explanation of the causes of station-to-station variability or, more likely, that the measurements provided for them are quite noisy.

An approach similar to that used in Oregon was tested to predict water depths at ungaged locations in a subdomain of EDEN (Conrads and Roehl, 2006b). Using a combination of static and dynamic variables, predictions were generated in two modeling steps. The dynamic variables were 30-month time series of daily water depths at 16 stations and water levels at 3 other stations. Static variables were obtained from the EDEN 400-m^2 grid. Values included coordinates of cell centroids and percentage vegetation types (slough, prairie, sawgrass, or upland) for approximately 2,300 cells, covering 370 square kilometers. The first ANN model simulated water depths using static (categorical) input variables to predict a constant baseline water depth (mean for the period of record). The second ANN model predicted day-to-day variability about the water-depth baseline by using a combination of static and dynamic variable inputs. A complete estimation of water depth at a given cell was computed by summing the outputs of the two models. Five of the water-depth gaging stations were withheld from model development to validate model accuracy. Using this methodology, prediction accuracy was improved, resulting in an average RMSE prediction error at validation gaging stations of only 0.1 ft (3 centimeters), or 4 percent of the dynamic range.

Approach

The majority of the hindcasts estimated for this study used a modification of the two-step modeling approach using static and dynamic data described in Conrads and Roehl (2006b). The modification was the addition of a third step error-correction model. The general approach for estimating water levels at a new station was to

1. Identify the EDEN stations that have similar hydrologic responses in a particular area,
2. Build databases for the static and dynamic variables,
3. Decorrelate time-series inputs,
4. Train the static and dynamic ANN models,
5. Compute initial water-level estimates and residual error from the measured data,
6. Train error-correction ANN models, and
7. Make final water-level estimates.

The process was then repeated for subsequent stations and areas in the Everglades. For a few stations (3 of the 25) with limited data and(or) limited data at nearby stations, a simpler single model approach was used.

Data-Collection Network

The water-level record extensions (hindcasts) use static and dynamic data from the EDEN monitoring network. The EDEN monitoring network includes ground-elevation measurements and continuous water-level data. Highly accurate ground-surface elevation data, collected by the USGS (Desmond, 2003), cover nearly the entire greater Everglades area. The elevation data were collected at over 50,000 points with an approximate spacing of 400 meters (m) to the North American Vertical Datum of 1988 (NAVD 88). The static variables were derived from the GIS cell attribute data of the 400-m EDEN grid. Attributes include location of the centroid of the cell and percent vegetation type (slough, prairie, sawgrass, upland, exotic, and other).

The EDEN database is composed of hourly water-level data from 253 gaging stations and includes freshwater marsh stations, boundary stations on canals, and coastal stations operated by the BCNP, ENP, the South Florida Water Management District (SFWMD), and the USGS. In this report, the names of the EDEN stations may follow the naming convention of the agency that maintains the stations. Stations with “site” in the name, such as Site 64 or Site 99, will be referenced with an uppercase S. All other references to a specific station will use a lowercase, such as site W2 or site North_CA1.

The dynamic variables (time series) were obtained from the marsh gaging stations of EDEN. The period of record for the EDEN water-level network is from October 1, 1999, to September 30, 2006. The periods of record for the new EDEN gaging station vary from approximately 4 months to 5 years. To extend the water-level record at the new EDEN station, a subset of 35 of the EDEN stations was used as inputs to the model for estimating water levels (referred to as “index” stations in this report). The locations and periods of record of the new EDEN stations and the index stations are listed in table 1. The stations in WCA3A were separated into two groups.

Limitation of the Datasets

As with any modeling effort, empirical or deterministic, the reliability of the model is dependent on the completeness of the datasets and on the quality of the data and range of measured conditions used for training and calibrating the model. Estimated data were not used in model development; thus, the majority of the time series used were less than 100 percent complete. The available period of record, especially the hindcasted stations, can limit the range of water level that the ANN model can accurately simulate. For the new EDEN stations, the period of record for the measured data often is a year or less (table 1). Many of the time series of the new EDEN stations provided a range of water-level behaviors corresponding to low and high water of the dry and wet seasons but did not provide a history of inter-annual variability. Some stations with limited periods of record and(or) missing data, only described a limited portion of the expected water-level range.

Table 1. Everglades Depth Estimation Network (EDEN) stations, types, periods of record, and percent complete record for hindcast and index stations used in this study.

[USGS, U.S. Geological Survey; SFWMD, South Florida Water Management District; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Station	Operating agency	Station type	Period of record: Begin date	Period of record: End date	Number of data points	Percent complete record
Water Conservation Area 1 (fig. 15)						
South_CA1	USGS	Hindcast	6/20/2001	9/30/2006	46,571	75.9
North_CA1	USGS	Hindcast	5/11/2001	9/30/2006	45,026	73.4
Site 7	USGS	Index	10/1/1999	9/30/2006	59,531	97.0
Site 8T	USGS	Index	10/1/1999	9/30/2006	60,632	98.8
Site 9	USGS	Index	10/1/1999	9/30/2006	61,055	99.5
WCA1ME	SFWMD	Index	10/1/1999	9/30/2006	60,315	98.3
Water Conservation Area 2 (fig. 17)						
EDEN11	USGS	Hindcast	6/9/2006	9/30/2006	2,723	4.4
EDEN13	USGS	Hindcast	6/8/2006	9/30/2006	2,747	4.5
WCA2F1	SFWMD	Index	10/1/1999	9/30/2006	51,177	83.4
WCA2E1	SFWMD	Index	10/1/1999	9/30/2006	54,573	88.9
Site 99	USGS	Index	10/1/1999	9/30/2006	60,645	98.8
S142H	SFWMD	Index	10/1/1999	9/30/2006	57,037	92.9
Water Conservation Area 3A (fig. 20)						
Group 1						
3A-5	USGS	Hindcast	6/6/2006	9/30/2006	2,797	4.6
EDEN4	USGS	Hindcast	6/9/2006	9/30/2006	2,728	4.4
EDEN8	USGS	Hindcast	6/7/2006	9/30/2006	2,771	4.5
EDEN9	USGS	Hindcast	6/9/2006	9/30/2006	925	1.5
EDEN12	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3
W2	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3

Table 1. Everglades Depth Estimation Network (EDEN) stations, types, periods of record, and percent complete record for hindcast and index stations used in this study. — Continued

[USGS, U.S. Geological Survey; SFWMD, South Florida Water Management District; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Station	Operating agency	Station type	Period of record		Number of data points	Percent complete record
			Begin date	End date		
			Water Conservation Area 3A (fig. 20) ***Group 1*** — Continued			
W5	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3
W11	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3
W14	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3
W15	USGS	Hindcast	1/25/2006	9/30/2006	4,991	8.1
W18	USGS	Hindcast	10/1/2005	9/30/2006	8,760	14.3
Site 64	USGS	Index	10/1/1999	9/30/2006	60,636	98.8
Site 63	USGS	Index	10/1/1999	9/30/2006	60,402	98.4
3AS3W1	SFWMD	Index	10/1/1999	9/30/2006	52,186	85.0
Site 65	USGS	Index	10/1/1999	9/30/2006	61,368	100.0
3ASW[a]	SFWMD	Index	10/1/1999	9/30/2006	60,407	98.4
			Group 2			
EDEN 5	USGS	Hindcast	6/5/2006	9/30/2006	1,657	2.7
EDEN 14	USGS	Hindcast	7/26/2006	9/30/2006	1,534	2.5
Site 62	USGS	Index	10/1/1999	9/30/2006	59,847	97.5
3A9	SFWMD	Index	10/1/1999	9/30/2006	58,027	94.5
3A12	SFWMD	Index	10/1/1999	9/30/2006	53,915	87.8
			Water Conservation Area 3B (fig. 26)			
TI-8	USGS	Hindcast	1/23/2006	9/30/2006	6,013	9.8
TI-9	USGS	Hindcast	1/23/2006	9/30/2006	6,014	9.8
EDEN7	USGS	Hindcast	6/8/2006	9/30/2006	2,748	4.5
EDEN10	USGS	Hindcast	6/8/2006	9/30/2006	2,748	4.5
Site 69	USGS	Index	10/1/1999	9/30/2006	61,277	99.8
Site 71	USGS	Index	10/1/1999	9/30/2006	60,457	98.5
Site 76	USGS	Index	10/1/1999	9/30/2006	61,059	99.4
SRS1[b]	USGS	Index	10/1/1999	9/30/2006	60,989	99.3
			Big Cypress National Preserve (fig. 29)			
EDEN1	BCNP	Hindcast	1/13/2006	9/30/2006	4,389	7.1
EDEN6	BCNP	Hindcast	7/10/2006	9/30/2006	1,974	3.2
BCA9	BCNP	Index	10/1/1999	9/30/2006	60,236	98.1
LOOP1T	SFWMD	Index	10/1/1999	9/30/2006	56,688	92.3
LOOP2T	SFWMD	Index	10/1/1999	9/30/2006	56,093	91.4
BCA10	BCNP	Index	10/1/1999	9/30/2006	60,084	97.8
BCA5	BCNP	Index	10/1/1999	9/30/2006	57,800	94.1
3ASW	SFWMD	Index	10/1/1999	9/30/2006	60,407	98.4
L28GAP	SFWMD	Index	10/1/1999	9/30/2006	53,515	87.1
			Everglades National Park (fig. 31)			
EDEN3	USGS	Hindcast	12/12/2005	9/30/2006	6,981	11.4
Met1	USGS	Hindcast	8/7/2006	9/30/2006	1,312	2.1
P34	ENP	Index	10/1/1999	9/30/2006	58,307	95.0
OT	ENP	Index	10/1/1999	9/30/2006	57,106	93.0
TMC	ENP	Index	10/1/1999	9/30/2006	59,178	96.4
P35	ENP	Index	10/1/1999	9/30/2006	58,954	96.0
P36	ENP	Index	10/1/1999	9/30/2006	59,223	96.4
NE1	USGS	Index	10/1/1999	9/30/2006	61,316	99.9
SRS1	USGS	Index	10/1/1999	9/30/2006	60,989	99.4
NP201	ENP	Index	10/1/1999	9/30/2006	59,252	96.6
NE2	USGS	Index	10/1/1999	9/30/2006	58,024	94.5
NESRS3	SFWMD	Index	10/1/1999	9/30/2006	52,855	86.1

[a] Station also used for hindcasts in Big Cypress National Preserve.

[b] Station also used for hindcasts in Everglades National Park.

Estimating Water Levels

The following section describes how the water levels were estimated for the new EDEN stations using ANN models. Hydrologic systems, such as the Everglades, exhibit random, chaotic, and multiple periodic behaviors that are driven by gravity flow, weather, and manmade disturbances, such as controlled flow releases. Modeling these behaviors on a large scale is challenging because of discontinuities in behaviors both spatially and temporally. Modeling requires calibration and validation data that represent the diversity of causes and effects. Subdividing a complex modeling problem into subproblems and addressing each subproblem is an effective means of achieving the best possible results. The creation of the various water conservation areas (WCA) in the Everglades transformed the system from a continuous "shallow" river from Lake Okeechobee to the Gulf of Mexico into a series of discontinuous compartments. The WCAs are logical boundaries to subdivide the EDEN into a subnetwork of water-level gaging stations to model the new EDEN water-level stations separately for WCA1, WCA2, WCA3A, WCA3B, BCNP, and ENP.

Within the compartments, water levels respond to large-scale, hydraulic gradients of the "shallow" river and small-scale changes in topography and vegetation. As water levels change, the restrictions (impedances) to flow change as more vegetation is inundated during rising and high water levels or as more water is channelized during falling and low water levels. The three-step modeling approach presented in this report uses information about vegetation and topography in addition to time-series data to generate accurate estimates of water levels. This section describes ANN models, statistical measures of prediction accuracy, the decorrelation of input variables, and an example that describes how to estimate water levels at a new EDEN station.

Artificial Neural Networks

Models generally fall into one of two categories—deterministic (mechanistic) or empirical. Deterministic models are created from first-principle equations, and empirical models adapt generalized mathematical functions to fit a line or surface through data from two or more variables. The most common empirical approach is ordinary least squares (OLS), which relates variables using straight lines, planes, or hyper-planes, whether the actual relations are linear or not. Techniques such as OLS and physics-based finite-difference models prescribe the functional form of the model's fit of the calibration data. Machine-learning techniques, like ANNs, synthesize a best fit of the calibration data.

The calibration of deterministic and empirical models attempts to optimally synthesize a line or surface through the measured data. Calibrating models is difficult when data have substantial measurement error or are incomplete, and the variables for which data are available may only be able to provide a partial explanation of the causes of variability. The principal advantages that empirical models have over deterministic models are they can be developed much faster and are more accurate when the modeled systems are well characterized by data. Empirical models, however, are prone to problems when poorly applied. Overfitting and multicollinearity caused by correlated input variables can lead to invalid mappings between input and output variables (Roehl and others, 2003).

Although there are numerous types of ANNs, the most commonly used type of ANN is the multilayer perceptron (MLP; Rosenblatt, 1958). As shown in figure 5, MLP ANNs are constructed from layers of interconnected processing elements called neurons, each executing a simple "transfer function." All input layer neurons are connected to every hidden layer neuron (HLN), and every HLN is connected to every output neuron. There can be multiple hidden layers, but a single hidden layer is sufficient for most problems.

Typically, linear transfer functions are used to scale input values to fall within the range that corresponds to the most

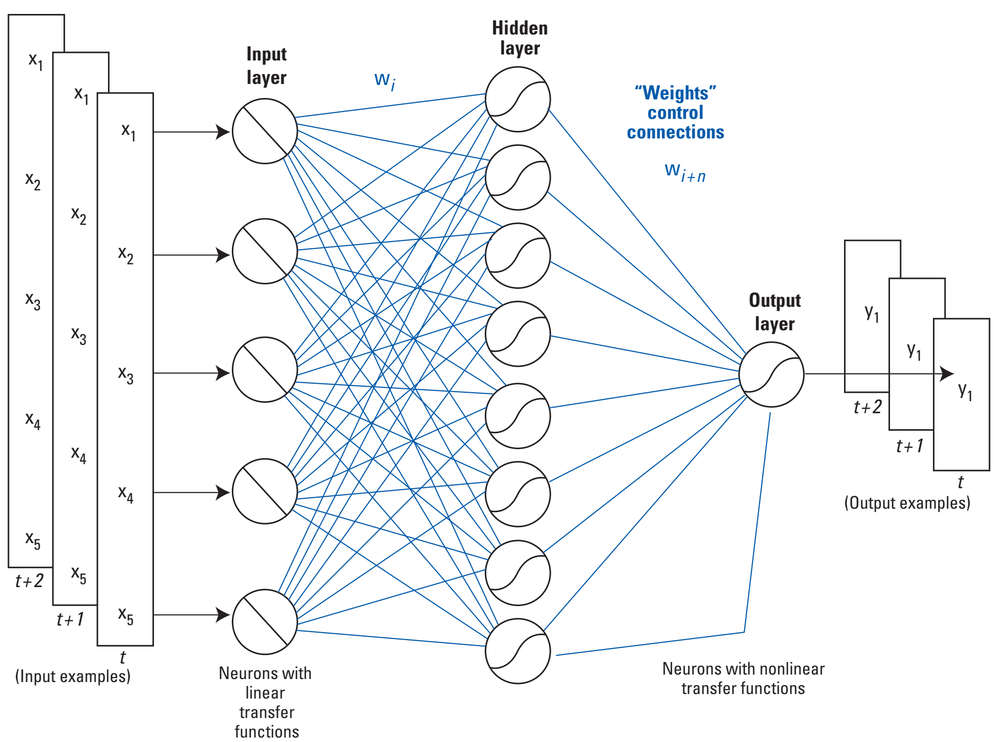


Figure 5. Multiperceptron artificial neural network architecture.

linear part of the s-shaped sigmoid transfer functions used in the hidden layers. Each connection has a "weight," w_i, associated with it, which scales the output received by a neuron from a neuron in an antecedent layer. The output of a neuron is a simple combination of the values it receives through its input connections and their weights, and the neuron's transfer function.

An ANN is "trained" by iteratively adjusting its weights to minimize the error by which it maps inputs to outputs for a dataset composed of "input/output vector pairs." Prediction accuracy during and after training can be measured by a number of metrics, including R^2 and RMSE. An algorithm that is commonly used to train MLP ANNs is the back error propagation (BEP) training algorithm (Rumelhart and others, 1986). Jensen (1994) describes the details of the MLP ANN, the type of ANN used in this study. Multilayer perceptron ANNs can synthesize functions to fit high-dimension, nonlinear multivariate data. Devine and Roehl (2003) and Conrads and Roehl (2005) describe their use of ANNs in multiple applications to model and control combined manmade and natural systems, including disinfection by-product formation, industrial air emissions monitoring, and surface-water systems, affected by point and nonpoint-source pollution.

Experimentation with a number of ANN architectural and training parameters is a normal part of the modeling process. For correlation analysis or predictive modeling applications, a number of candidate ANNs are trained and evaluated for their statistical accuracy and their representation of process physics. Interactions between combinations of variables also are considered. Input variables to the models are selected to minimize correlation between variables (typically Pearson coefficient R of less than 0.5). Finally, a satisfactory model can be configured for end-user deployment. In general, a high-quality, predictive model can be obtained when

- The data are well distributed throughout the behavioral range of interest,
- The input variables selected by the modeler share "mutual information" about the output variables, and
- The functional form "prescribed" or "synthesized" for the model used to "map" (correlate) input variables to output variables is a good one. Techniques, such as OLS and physics-based finite-difference models, prescribe the functional form of the model's fit of the calibration data. Machine-learning techniques like ANNs synthesize a best fit to the data.

Decorrelation of Variables

Often, explanatory variables share information about the behavior of a response variable. It is difficult, if not impossible, to understand the individual effects of these variables (sometime known as confounded or correlated variables), on a response variable. Using correlated inputs to the models also can spuriously increase the model accuracy statistics, such as R^2. Empirical models have no notion of process physics, nor the nature of interrelations between input variables. To clearly analyze the effects of confounded variables, the unique informational content of each variable must be determined by "decorrelation." Decorrelation is accomplished by ordering confounded variables according to a criterion. For example, the relative independence of the correlated variables is determined and then empirical functions (ANN models) of the less independent variables are developed using the more independent variables as inputs. The empirical function's residual error is computed by subtracting its predicted values from the actual measurements. The residual error manifests the "unshared" information between the model's more independent input variables and its less independent output variable. The residual error is the decorrelated version of the output variable (decorrelated variable) and can be used in water-level models. For example, in WCA3A, there are five existing EDEN stations that could be used to predict the water levels at the new stations—Site 63, Site 64, Site 65, 3AS3W1, and 3ASW. Four of the stations, Site 63, Site 65, 3AS3W1, and 3ASW, are systematically decorrelated from Site 64, using cascading models (fig. 6). Note that the residual error (decorrelated variable) from each anteceding model becomes an input to a subsequent model. The input variables and statistical summaries for the decorrelation models are listed in Appendixes 1 and 2 at the back of the report.

In this study, the decorrelation order was determined by minimizing the incomplete record of the decorrelated variable. A model cannot compute an output value when an input has

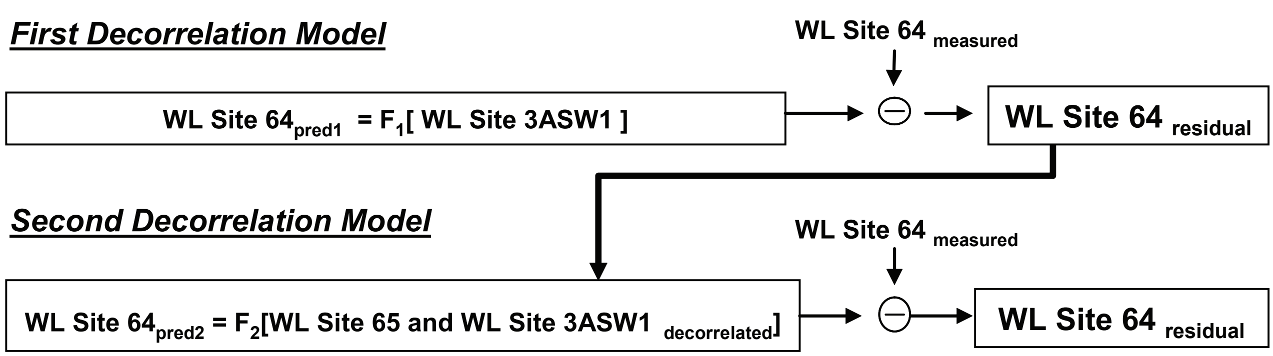


Figure 6. Decorrelation models.

a missing value. For example, if three time series needing to be decorrelated and the time series are 90, 75, and 50 percent complete records, respectively, and the decorrelation order was the third, second, and first datasets, all the decorrelated variables would be 50 percent complete or less (missing data may not be concurrent between datasets). Using the third dataset in the first decorrelation model (the 50 and 75 percent complete datasets) would result in a decorrelated variable only 37.5 percent complete. Subsequent decorrelation models would produce a decorrelated variable that is 34 percent complete (90 percent x 37.5 percent = 34 percent). The better decorrelation order would be the 90-percent, 75-percent, and 50-percent datasets. The decorrelation model (the 90- and 75- percent datasets) would result in a decorrelated variable 68 percent complete. The subsequent decorrelation model would, like the first decorrelation order, result in a decorrelated variable that is 34 percent complete (68 percent x 50 percent = 34 percent).

Statistical Measures of Prediction Accuracy

Statistical measures of prediction accuracy were computed for the final water-level estimates and for the decorrelation, static, dynamic, and error-correction models. The statistics for the final water-level estimates capture the ability of the three-step modeling approach to accurately estimate the water levels at the station. The statistics for the decorrelation models and individual step models (static, dynamic, and error-correction models) document these intermediate models. Because several models are used, the statistics for the individual models may not be an indication of the quality of the final water-level estimates. For example, the static models generally have very low R^2, especially in the test dataset, as would be expected when static variables are used to predict a dynamic time series. Ultimately, the hindcasts should be evaluated by the statistics for the final estimates. The decorrelation models typically have high R^2 values, but the results from the model used in the dynamic models are the residuals which have a low R^2.

The R^2, mean error (ME), RMSE, and percent model error (PME) were computed for the training and testing datasets for each model and are listed in Appendix 2. Model accuracy commonly is reported in terms of R^2 and is interpreted as the “goodness of the fit” of a model. A second interpretation may answer the question, “How much information does one variable or a group of variables provide about the behavior of another variable?” In the first context, an $R^2 = 0.6$ might be disappointing, whereas in the latter, it is merely an accounting of how much information is shared by the variables being used.

The ME and RMSE statistics provide a measure of the prediction accuracy of the ANN models. The ME is a measure of the bias of model predictions—whether the model over or under predicts the measured data. The ME is presented as the adjustment of the simulated values to equal the measured values; therefore, positive and negative MEs indicate an over or under prediction bias by the model, respectively. Mean errors near zero may be misleading because negative and positive discrepancies in the simulations can cancel each other. Root mean square errors address the limitations of ME by computing the magnitude, rather than the direction (sign) of the discrepancies. The units of the ME and RMSE statistics are the same as the variable simulated by the model.

The minimum and maximum values of the measured output are listed in Appendix 2. The accuracy of the models, as given by RMSE, should be evaluated with respect to the range of the output variable. A model may have a low RMSE, but if the range of the output variable is small, the model may only be accurate for a limited range of conditions and the model error may be a relatively large percentage of the model response. Likewise, a model may have a large RMSE, but if the range of the output variable is large, the model error may be a relatively small percentage of the total model response. The PME is computed by dividing the RMSE by the range of the measured data.

Example of Estimating Water Level Using Three-Step Modeling Approach

Generally, the same modeling approach for estimating water levels at a new EDEN site was used in all six areas. The three-step modeling approach is described in this section along with a detailed example of estimating the water level at W2 in WCA3A. Variable inputs for the models used in the W2 example are listed in table 2, and summary statistics for the models are listed in table 3. The input variables, summary statistics, and variable descriptions for all the models are listed in Appendixes 1, 2, and 3, respectively, at the back of the report (tables 2 and 3 are excerpts of Appendixes 1 and 2, respectively). (The summary statistics for 25 hindcasts are listed in table 4, p. 19.) The ANN models and plots discussed in this report were developed using the iQuest™ data-mining software[1] (Version 2.03C DM Rev31).

The three-step modeling approach was developed to estimate water levels at the new EDEN stations. The approach uses output from the first model as input to the second and output from the second model as input to the third model. A schematic of the three-step modeling approach is shown in figure 7. The first model (F_1) predicts the water level (WL-Site_{pred1}) using inputs for only the static variables of cell location (cell_x and cell_y) of the gaging station and the percent vegetation types (pctslough, pctprairie, pctsawgrass, pctupland, pctexotic, and pctother). Although, this model (also called the “static” model) is not able to predict the dynamic variability of the water level, it is able to discriminate general station-to-station differences based on location and percent vegetation.

---

[1] The iQuest™ software is exclusively distributed by Advanced Data Mining, LLC, 3620 Pelham Road, PMB 351, Greenville, SC 29615-5044 Phone: (864) 201-8679, email: info@advdatamining.com, *http://www.advdatamining.com*.

Table 2. Model type, model name, input variables, model prediction variables, and variables used to estimate water levels at site W2.

Model type	Model name	Input variables[a]	Model prediction variable	Variable used from model	Comment
static	cluster1static	CELL_Y PCTEXOTICS PCTSLOUGH PCTPRAIRIE PCTOTHER	WLSITE	RES_WLSITE	Static model for Group 1 stations in WCA3a—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for model for dynamic model.
dynamic	cluster1dynamic	SITE_63DEC SITE_64 3AS3W1DEC CELL_Y PCTSLOUGH PCTEXOTICS PCTOTHER	RES_WLSITE	RES_WLSITE	Dynamic model for Group 1 stations in WCA3a—RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
decorrelation	3as3w1dec	3AS3W1	SITE_64	3as3w1dec	Model to decorrelate 3AS3W1 from SITE_64.
decorrelation	65dec	SITE_65 3AS3W1DEC	SITE_64	65dec	Model to decorrelate Site_65 from SITE_64 and 3AS3W.
decorrelation	63decr1	3AS3W1DEC SITE_65DEC SITE63	SITE_64	63decr	Model to decorrelate Site_63 from SITE_64, 3AS3W1, and Site_65.
decorrelation	3aswdec	3AS3W1DEC SITE_65DEC SITE_63DEC 3ASW	SITE_64	3aswdec	Model to decorrelate 3ASW from SITE_64, 3AS3W1, Site_65, and Site_63.
error correction	w2res	SITE_63DEC SITE_64 3AS3W1DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W2.

[a] Descriptions of variables are provided in Appendix 3.

Table 3. Summary statistics for the models used to estimate water levels for site W2.

[HLN, hidden layer neurons; min, minimum; max, maximum; WL, water level; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model type	Model name	Number of HLN	Training									
			Range of output variable		n	R^2	ME, WL	SSE, WL	MSE, WL	RMSE, WL	PME
			Min, WL	Max, WL							
static	cluster1static	2	8.26	10.28	5	1.000	-0.0002	0.0001	0.000	0.004	0.2
decorrelation	3as3w1dec	1	8.51	11.74	11	0.991	0.005	0.11	0.010	0.109	3.4
decorrelation	65dec	1	8.30	11.95	90	0.995	-0.005	0.41	0.005	0.068	1.9
decorrelation	63decr1	1	8.29	11.95	358	0.941	-0.014	14.50	0.041	0.202	5.5
decorrelation	3aswdec	1	8.73	11.95	962	0.951	-0.047	28.67	0.030	0.173	5.4
dynamic	cluster1dynamic	3	-2.39	2.11	1,622	0.969	0.012	37.84	0.023	0.153	3.4
error correction	w2res	2	0.76	2.06	128	0.976	-0.002	0.40	0.003	0.056	4.3

Model type	Model name	Number of HLN	Testing								
			Range of output variable		n	R^2	ME, WL	SSE, WL	MSE, WL	RMSE, WL	PME
			Min, WL	Max, WL							
static	cluster1static	2	6.55	13.39	97,891	0.446	-0.0002	74,110.44	0.757	0.870	12.7
decorrelation	3as3w1dec	1	8.28	11.95	51,742	0.970	-0.021	1,047.08	0.020	0.142	3.9
decorrelation	65dec	1	8.28	11.95	51,742	0.993	-0.029	260.47	0.005	0.071	1.9
decorrelation	63decr1	1	8.28	11.95	50,776	0.939	0.044	2,098.85	0.041	0.203	5.5
decorrelation	3aswdec	1	8.69	11.95	49,815	0.959	-0.015	1,337.04	0.027	0.164	5.0
dynamic	cluster1dynamic	3	-10.26	2.17	83,968	0.960	0.012	2,312.33	0.028	0.166	1.3
error correction	w2res	2	0.73	2.11	7,648	0.974	0.008	22.42	0.003	0.054	3.9

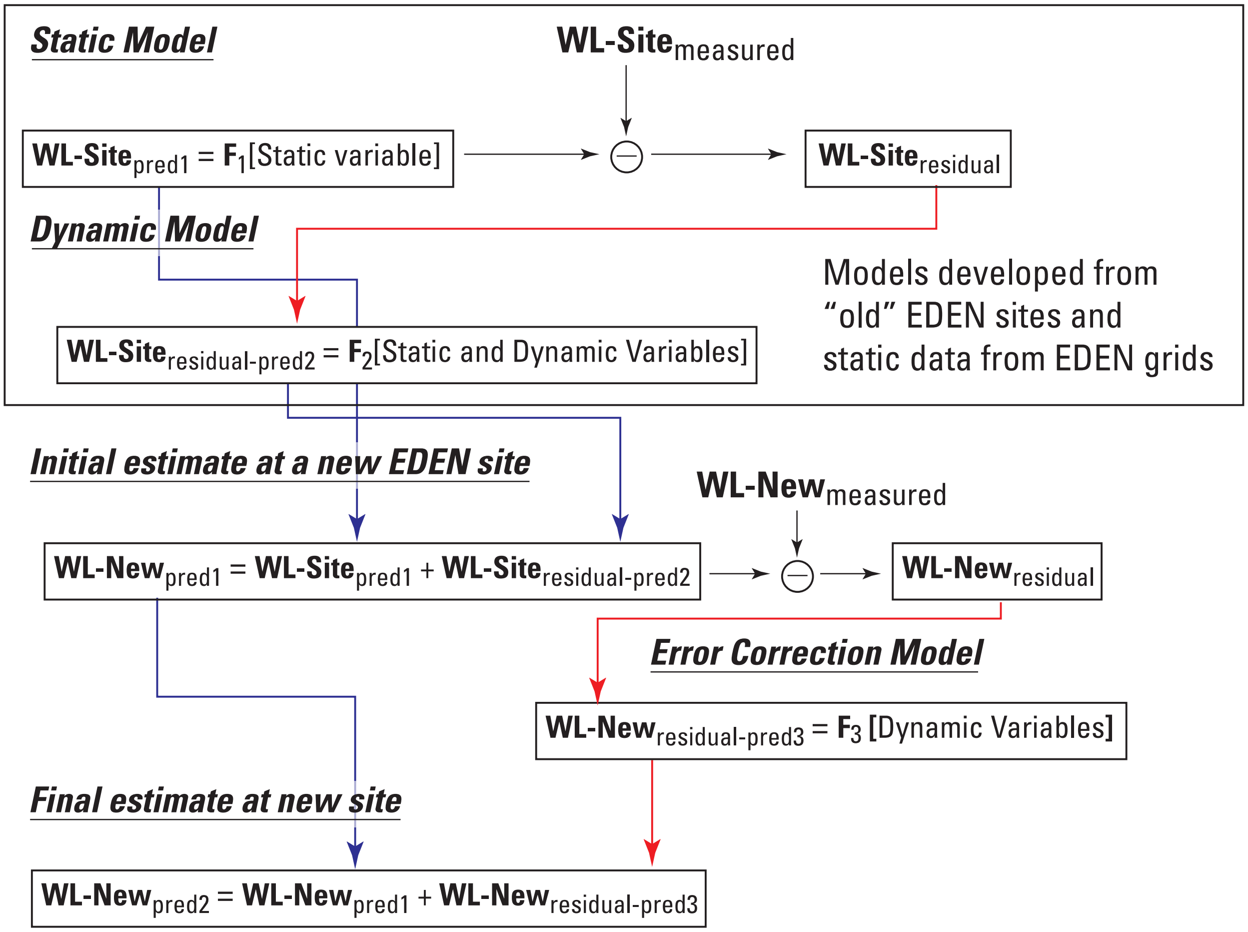


Figure 7. Three-step modeling approach to make final water-level estimates at a station.

To train and test the static model, a "stacked" dataset (Roehl and others, 2006) was generated that included the time series and static variables from the index EDEN gaging stations (fig. 8). For example, in WCA3A, five index EDEN stations were used as inputs (explanatory variables) to develop the models to hindcast water-level data at the new stations. The static model used in WCA3A uses 5 index stations for predicting water levels at 11 of the new EDEN stations. The index stations are Site 63, Site 64, Site 65, 3AS3W, and 3ASW. The static model uses five static variables: CELL_Y, PCTEXOTICS, PCTSLOUGH, PCTPRAIRIE, and PCTOTHER (table 2). The static model uses two hidden layer neurons (HLN). The R^2 of the training and testing datasets are 1.0 and 0.44, respectively (cluster1static, table 3). The water-level predictions from the static model for five stations in WCA3A are shown in figure 9. The apparent baseline shifts ("steps") in the simulated time series represent a different station used in the static model. The model is able to discriminate relative differences in water levels between the stations using only the static variables of location and percent vegetation.

Using the water-level predictions and the measured data at the five index stations, the residual error (the difference between the predicted and measured water level), WL-Site$_{residual}$, is computed for each station. The residual error manifests the dynamic variability at the station that was not simulated by the static model and is simulated by a second model (F_2, fig. 7). The second model (also called the "dynamic" model) uses time series of water level in addition to static variables to predict WL-Site$_{residual}$. For the WCA3A example, the dynamic model uses three dynamic input variables (Site_64, Site_63dec, 3AS3W1dec; table 2) and four static input variables (CELL_Y, PCTSLOUGH, PCTEXOTICS, and PCTOTHER; table 2). The input variables and summary statistics for the decorrelation models are listed in tables 2 and 3, respectively. The dynamic model uses three HLNs, and the R^2 of the training and testing datasets are 0.969 and 0.960, respectively (cluster1dynamic, table 3). The measured and predicted residual water levels are shown in figure 10. The dynamic model uses the same "stacked" dataset as the static model.

To compute the initial estimate of water level at a new EDEN station, the static variables are used in the static (F_1) and dynamic (F_2) models. The initial prediction is the sum of the water-level predictions from the static model and the

		RT	DATETIME	WLSITE	SITE_63DEC	SITE_64	SITE_65DEC	3AS3W1DEC	3ASWDEC	CELL_X	CELL_Y	PCTSLOUGH	PCTPRAIRIE	PCTSAWGRASS	PCTEXOTICS
	1775		36655.7500	9.46	-0.27	9.19	-0.01	-0.03	-0.16	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1776		36655.8750	9.46	-0.27	9.19	-0.01	-0.02	-0.17	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1777		36656.0000	9.46	-0.28	9.18	-0.02	-0.03	-0.17	547000.0	2896600.0	0.0	0.0	96.0	3.0
Site 63	1778		36656.1250	9.46	-0.28	9.18	-0.02	-0.03	-0.17	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1779		36656.2500	9.46	-0.28	9.18	-0.02	-0.03	-0.17	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1780		36656.3750	9.46	-0.28	9.18	-0.01	-0.03	-0.17	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1781		36656.5000	9.46	-0.28	9.18	-0.01	-0.03	-0.16	547000.0	2896600.0	0.0	0.0	96.0	3.0
	1782		36656.6250	9.46	-0.28	9.18	-0.02	-0.02	-0.16	547000.0	2896600.0	0.0	0.0	96.0	3.0
	22302		36655.5000	9.20	-0.28	9.20	-0.01	-0.03	-0.17	533000.0	2873000.0	1.0	3.0	92.0	4.0
	22303		36655.6250	9.20	-0.28	9.20	-0.01	-0.02	-0.16	533000.0	2873000.0	1.0	3.0	92.0	4.0
	22304		36655.7500	9.19	-0.27	9.19	-0.01	-0.03	-0.16	533000.0	2873000.0	1.0	3.0	92.0	4.0
Site 64	22305		36655.8750	9.19	-0.27	9.19	-0.01	-0.02	-0.17	533000.0	2873000.0	1.0	3.0	92.0	4.0
	22306		36656.0000	9.18	-0.28	9.18	-0.02	-0.03	-0.17	533000.0	2873000.0	1.0	3.0	92.0	4.0
	22307		36656.1250	9.18	-0.28	9.18	-0.02	-0.03	-0.17	533000.0	2873000.0	1.0	3.0	92.0	4.0
	22308		36656.2500	9.18	-0.28	9.18	-0.02	-0.03	-0.17	533000.0	2873000.0	1.0	3.0	92.0	4.0
	42831		36655.5000	8.73	-0.28	9.20	-0.01	-0.03	-0.17	528200.0	2855400.0	11.0	2.0	86.0	2.0
	42832		36655.6250	8.72	-0.28	9.20	-0.01	-0.02	-0.16	528200.0	2855400.0	11.0	2.0	86.0	2.0
	42833		36655.7500	8.72	-0.27	9.19	-0.01	-0.03	-0.16	528200.0	2855400.0	11.0	2.0	86.0	2.0
Site 65	42834		36655.8750	8.71	-0.27	9.19	-0.01	-0.02	-0.17	528200.0	2855400.0	11.0	2.0	86.0	2.0
	42835		36656.0000	8.71	-0.28	9.18	-0.02	-0.03	-0.17	528200.0	2855400.0	11.0	2.0	86.0	2.0
	42836		36656.1250	8.71	-0.28	9.18	-0.02	-0.03	-0.17	528200.0	2855400.0	11.0	2.0	86.0	2.0
	42837		36656.2500	8.71	-0.28	9.18	-0.02	-0.03	-0.17	528200.0	2855400.0	11.0	2.0	86.0	2.0
	63360		36655.5000	7.43	-0.28	9.20	-0.01	-0.03	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63361		36655.6250	7.42	-0.28	9.20	-0.01	-0.02	-0.16	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63362		36655.7500	7.42	-0.27	9.19	-0.01	-0.03	-0.16	523000.0	2859800.0	1.0	1.0	83.0	15.0
3AS3W	63363		36655.8750	7.41	-0.27	9.19	-0.01	-0.02	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63364		36656.0000	7.41	-0.28	9.18	-0.02	-0.03	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63365		36656.1250	7.41	-0.28	9.18	-0.02	-0.03	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63366		36656.2500	7.41	-0.28	9.18	-0.02	-0.03	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	63367		36656.3750	7.41	-0.28	9.18	-0.01	-0.03	-0.17	523000.0	2859800.0	1.0	1.0	83.0	15.0
	83889		36655.5000	8.10	-0.28	9.20	-0.01	-0.03	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83890		36655.6250	8.09	-0.28	9.20	-0.01	-0.02	-0.16	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83891		36655.7500	8.08	-0.27	9.19	-0.01	-0.03	-0.16	516600.0	2874600.0	17.0	1.0	61.0	18.0
3ASW	83892		36655.8750	8.08	-0.27	9.19	-0.01	-0.02	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83893		36656.0000	8.08	-0.28	9.18	-0.02	-0.03	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83894		36656.1250	8.08	-0.28	9.18	-0.02	-0.03	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83895		36656.2500	8.08	-0.28	9.18	-0.02	-0.03	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0
	83896		36656.3750	8.08	-0.28	9.18	-0.01	-0.03	-0.17	516600.0	2874600.0	17.0	1.0	61.0	18.0

Figure 8. Example of a stacked dataset used to train the static and dynamic models. Each block represents a limited number of rows for each index site in the dataset. Note the dynamic variables are in columns 3–8 (WLSITE to 3ASWDEC) and static variables begin in column 9 (CELL_X). Not all variables in the dataset are shown.

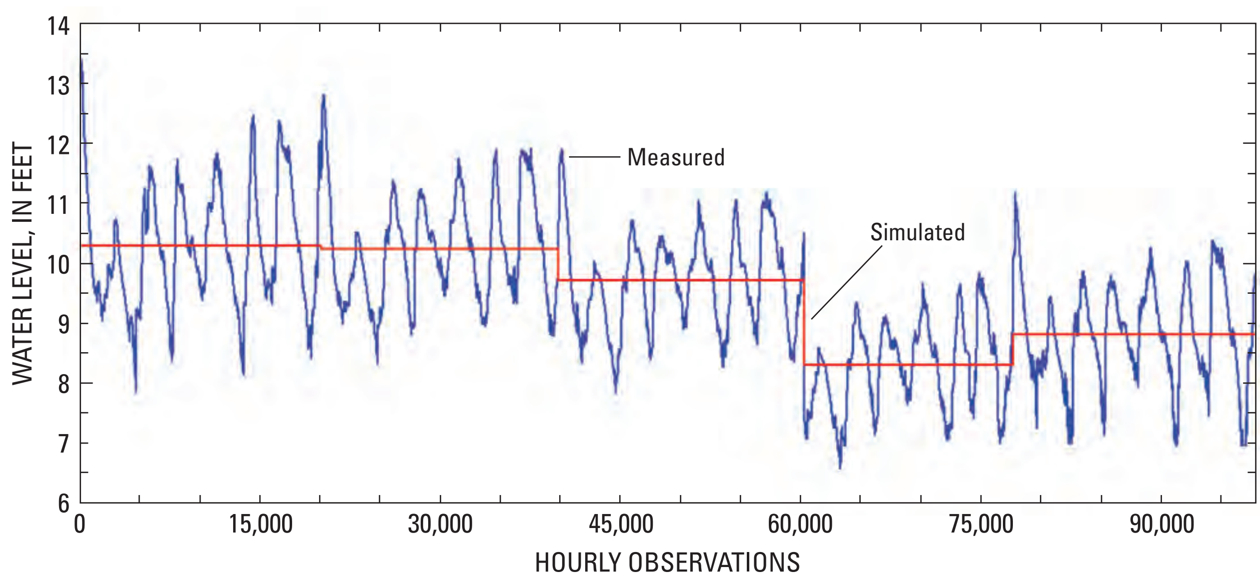


Figure 9. Measured (blue trace) and simulated (red trace) water levels from the static model for Water Conservation Area 3A. The “steps” in the simulated water level indicate a different site.

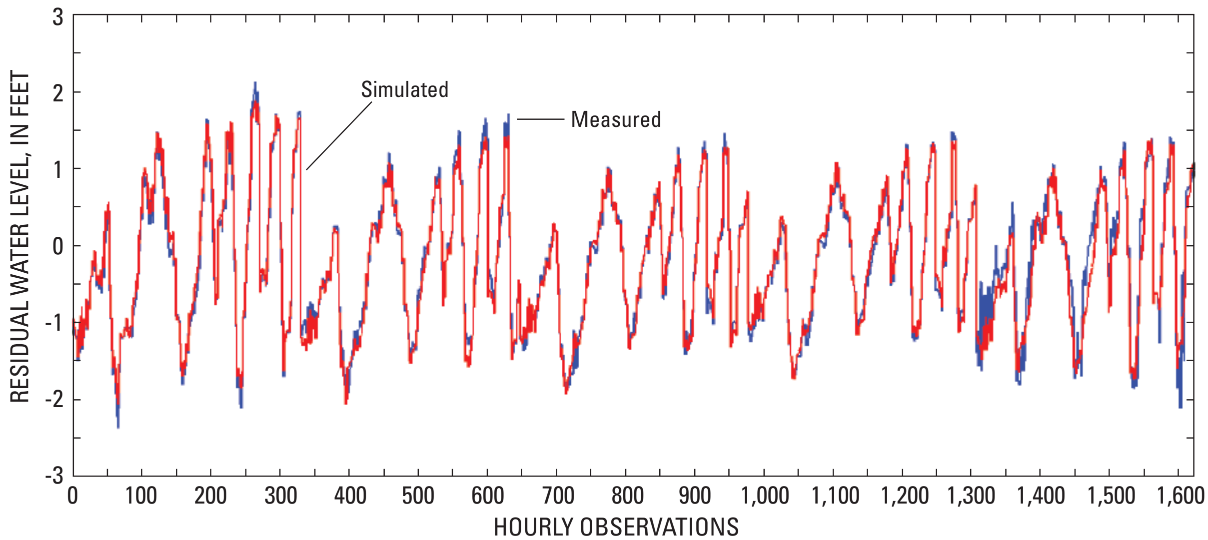


Figure 10. Measured (blue trace) and simulated (red trace) residual water levels from the dynamic model for Water Conservation Area 3A. The residual water levels were computed from the water-level prediction from the static model. The training dataset is shown. There are approximately 325 observations from each of the five index sites.

residual from the dynamic model. For example, to compute the initial water-level prediction for site W2 in WCA3A (WL-New$_{pred1}$, fig. 7), the static variables for the new station are used in the static and dynamic models (CELL-Y, PCTEXOTICS, PCTSLOUGH, PCTPRAIRIE, AND PCTOTHER) to compute WL-Site$_{pred1}$ and WL-Site$_{residual\text{-}pred2}$ (fig. 7). The results from the static and dynamic models, and the initial water-level predictions and measured data for site W2 are shown in figure 11. Although the initial water-level estimates capture the dynamic variability of the measured data, the absolute predictions are in error by approximately 1 to 2 ft.

To improve the accuracy of the water-level predictions, a third-step model (F_3, error correction model, fig. 7) is used to estimate the residual error between the initial water-level estimate and the measured data at the new EDEN station. The residual error is computed by subtracting the initial water-level

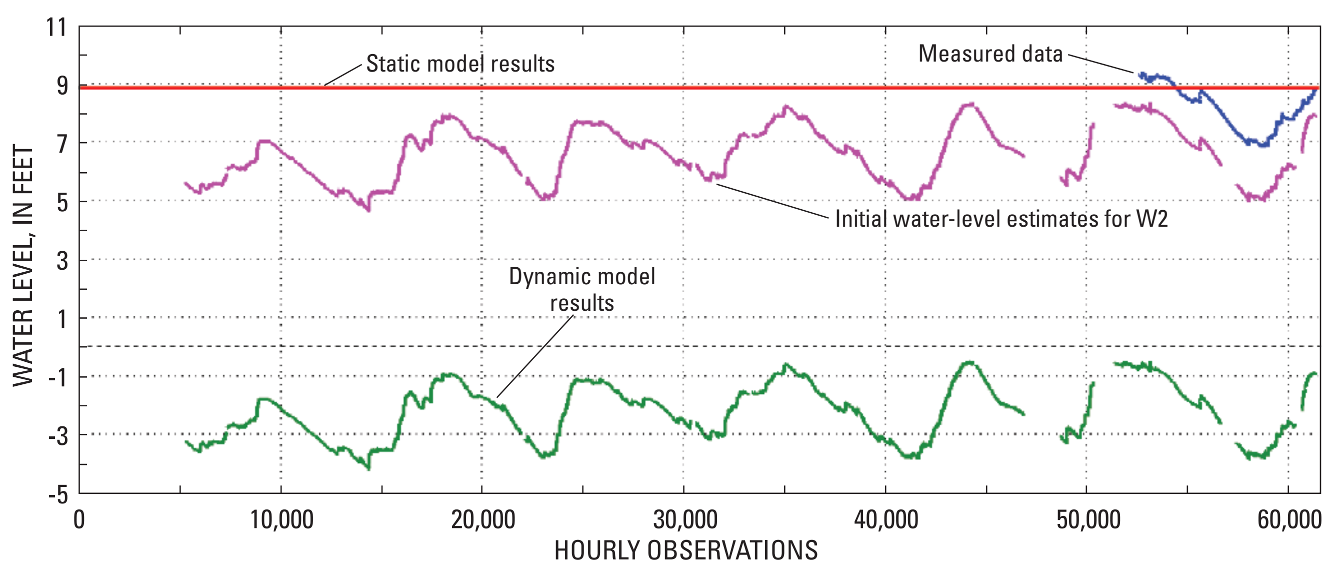


Figure 11. Initial water-level estimates for site W2. Results from the static and dynamic models are shown with the measured data. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

estimates from the measured data (fig. 12). The error-correction model uses a set of dynamic variables to predict the residual error from the initial water-level estimate. The time series of computed and predicted error is shown in figure 13. The error-correction model for site W2 (W2res) uses three dynamic variables: Site_63DEC, Site_64, and 3AS3W1DEC (table 2) and two HLNs. The R^2 for the training and testing datasets for the error-correction model is 0.976 and 0.974, respectively (table 3).

The final water-level estimate at a new EDEN station is the sum of the initial water-level estimate and the predicted residual from the error-correction model (WL-New$_{residual\text{-}pred3}$, fig. 7). The results of the error-correction model and final water-level estimates for site W2 are shown in figure 14. The

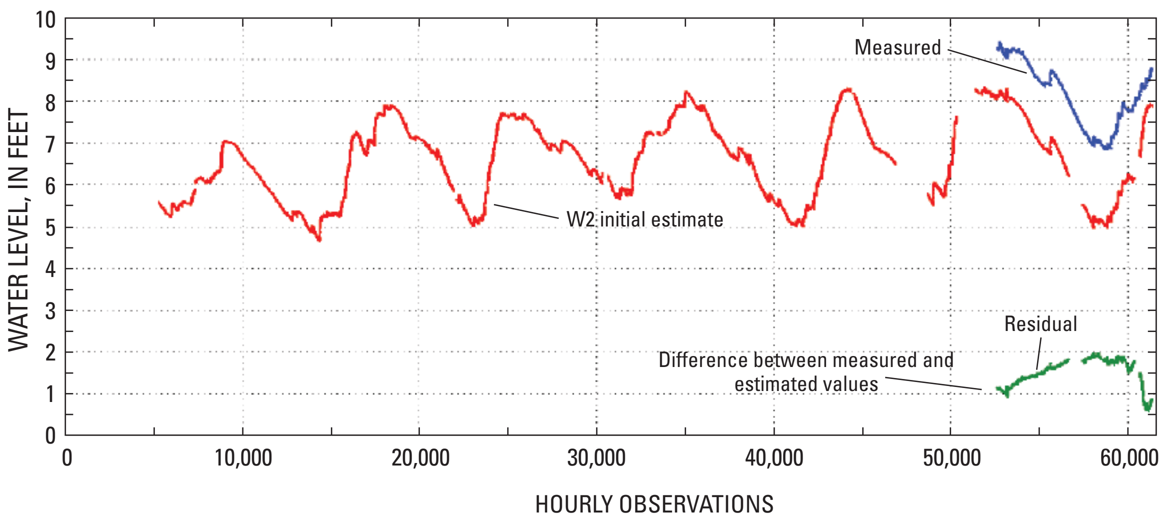


Figure 12. Measured water levels, initial water-level estimates, and residual time series (difference between the measured and initial water-level estimates) for site W2. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

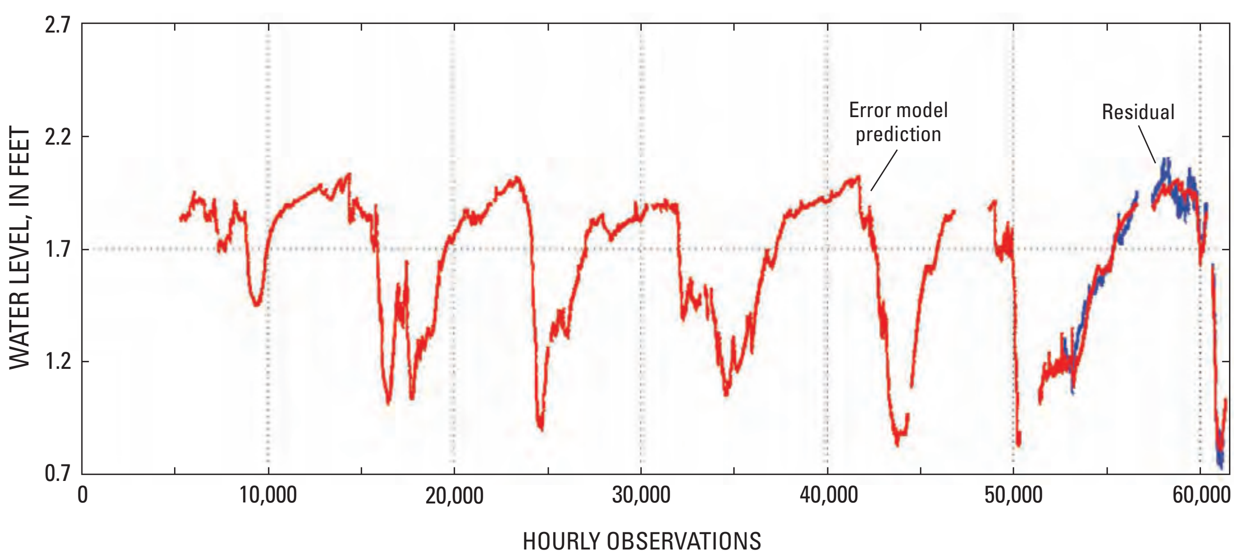


Figure 13. Residual time series between the measured and initial estimates of water levels for site W2. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

R^2 and PME for the final water-level estimate are 0.995 and 0.1 percent, respectively (table 4). Water levels for Site 64 also are plotted to show the general hydrologic response concurrent with the record extension for W2.

Empirical and mechanistic models perform best when interpolating within the range of data used for training or calibration. Although the statistics of the final water-level estimate are good, the hindcasts should be evaluated for periods when the model extrapolates to estimate the hydrologic response. The range of interpolation of water-level estimation model for site W2 (fig. 14) shows that the model estimates were extrapolated only for a brief period of low-water conditions. The breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

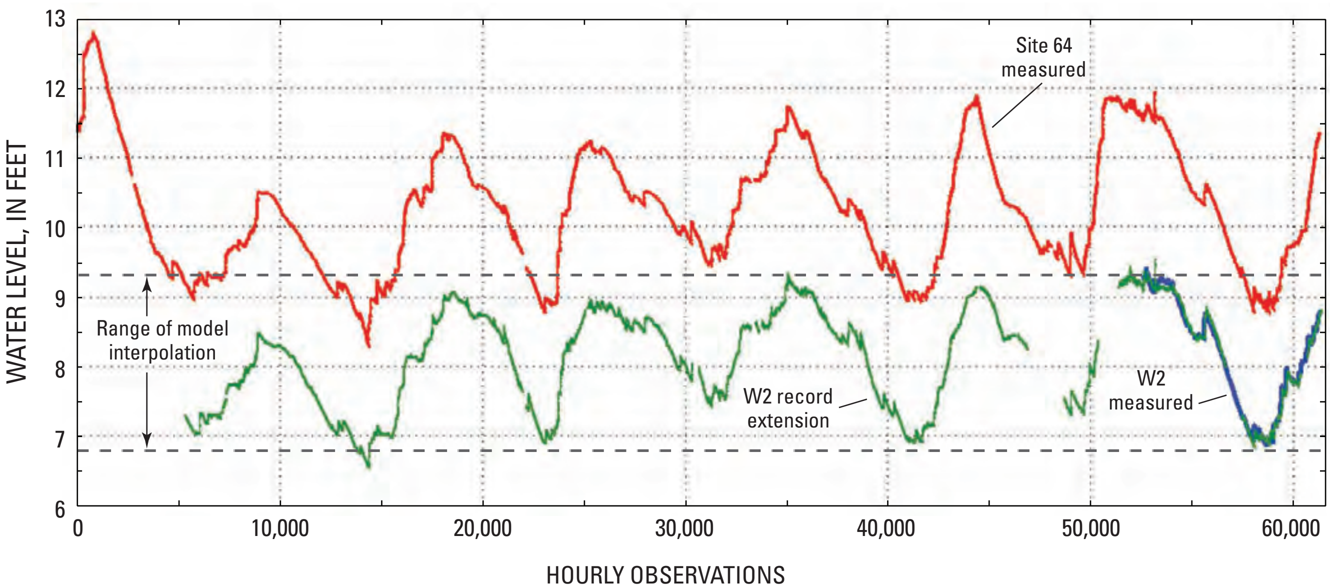


Figure 14. Measured and record extension (predicted) water levels at site W2 and water levels for Site 64. The range of measured water levels at W2 is shown with two dashed lines and used as an indication of when the W2 model is interpolating within conditions the model was trained. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Table 4. Summary statistics for water-level estimates for new Everglades Depth Estimation Network (EDEN) stations.

[n, number of data values; R, Pearson coefficient; R^2, coefficient of determination; Min, minimum; ft, feet; Max, maximum; ME, mean error; RMSE, root mean square error; PME, percent model error]

Site	n	R	R^2	Data range Min, in ft	Data range Max, in ft	ME	RMSE	PME
Water Conservation Area 1 (fig. 15)								
North_CA1	41,721	0.943	0.889	13.82	16.38	-0.046	0.134	5.2
South_CA1	43,409	0.991	0.983	12.85	15.90	0.008	0.088	2.9
Water Conservation Area 2 (fig. 17)								
EDEN11	2,637	0.950	0.902	10.97	13.19	-0.054	0.044	2.0
EDEN13	1,968	0.955	0.912	7.07	7.69	-0.018	0.006	1.0
Water Conservation Area 3A (fig. 20)								
3A-5	5,684	0.998	0.995	8.15	10.10	0.002	0.001	0.1
EDEN4	2,399	0.999	0.998	6.96	10.30	0.001	0.002	0.1
EDEN5	1,653	0.999	0.999	8.01	10.20	0.004	0.001	0.1
EDEN8	2,442	0.999	0.999	6.79	9.19	0.001	0.001	0.0
EDEN9	925	0.999	0.998	7.69	10.55	-0.006	0.004	0.1
EDEN12	7,648	0.999	0.998	6.84	9.59	-0.002	0.001	0.0
EDEN14	969	0.906	0.821	8.66	9.59	0.011	0.010	1.1
W2	7,648	0.998	0.995	6.86	9.42	0.008	0.003	0.1
W5	7,648	0.999	0.998	6.84	9.59	-0.001	0.001	0.0
W11	7,628	0.999	0.998	7.08	10.08	-0.012	0.002	0.1
W14	7,628	0.998	0.997	7.00	10.00	-0.013	0.003	0.1
W15	3,859	0.999	0.998	7.47	9.67	-0.001	0.001	0.0
W18	7,628	0.998	0.996	7.92	10.21	-0.000	0.002	0.1
Water Conservation Area 3B (fig. 26)								
TI-8	5,684	0.996	0.993	4.59	6.16	0.002	0.001	0.1
TI-9	5,684	0.996	0.992	5.29	6.42	0.009	0.001	0.1
EDEN7	2,419	0.998	0.996	5.24	6.96	0.008	0.001	0.1
EDEN10	2,419	0.995	0.990	5.06	6.32	0.007	0.002	0.1
Big Cypress National Preserve (fig. 29)								
EDEN1	3,864	0.960	0.921	7.13	7.82	0.018	0.003	0.4
EDEN6	1,591	0.984	0.968	8.76	10.73	0.030	0.007	0.4
Everglades National Park (fig. 31)								
EDEN3	5,294	0.989	0.978	0.07	1.75	-0.042	0.006	0.4
Met1	1,238	0.994	0.989	5.23	5.76	0.000	0.000	0.0

Extension of Water-Level Records for the Everglades Depth Estimation Network (EDEN)

The water-level record extensions for six areas of the Everglades are summarized in the following sections of the report. The hindcast and index stations for each area are shown in a figure followed by the measured and estimated time series. Water levels for one of the index stations in each area are also plotted to show the general hydrologic response of the area for the 7 years concurrent with the record extension. The datasets with the record extensions are available on the EDEN Web page on the South Florida Information Access (SOFIA) Web site (*http://sofia.usgs.gov/eden/index.php*).

Water Conservation Area 1 (WCA1)

For WCA1, water-level records were extended for two stations—North_CA1 and South_CA1—with four stations being used as index stations—Site 7, Site 8T, Site 9, and WCA1ME (fig. 15; table 1). The static and dynamic models used three static-input variables (CELL_Y, PCTPRAIRIE, and PCTEXOTICS). The dynamic models also used four water-level input variables (SITE_7, SITE8DEC2, WCA1MEDEC, and SITE9DEC). The measured and estimated water levels for the two stations are shown in figure 16. Site 7 also is plotted to show the general water-level behavior through the record extension. The water-level estimates for the South_CA1 gaging station are more accurate than those for the North_CA1 gaging station with R^2 values of 0.98 and 0.89, respectively, and PME values of 2.9 and 5.2, respectively (table 4). Both sets of water-level estimates capture the full range of the measured data.

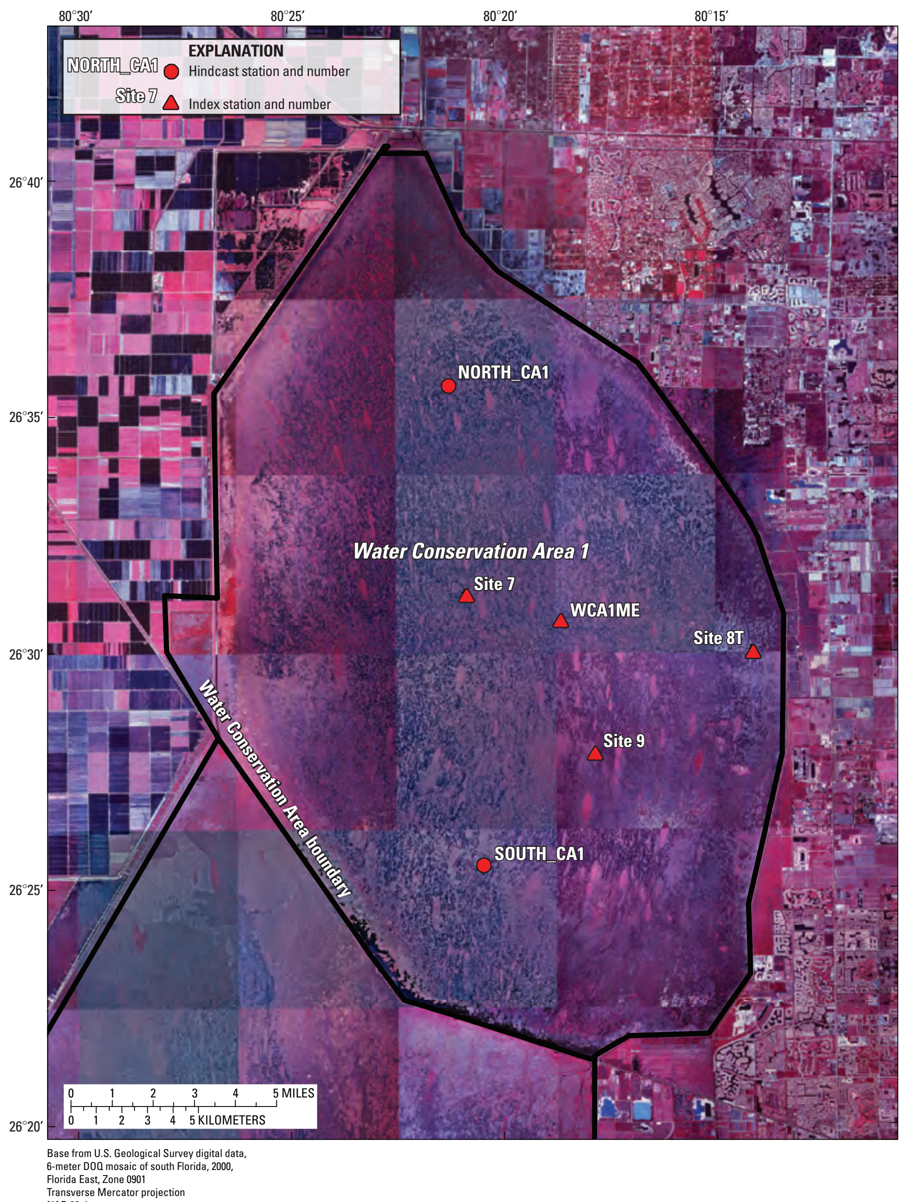


Figure 15. Index and hindcast stations for Water Conservation Area 1.

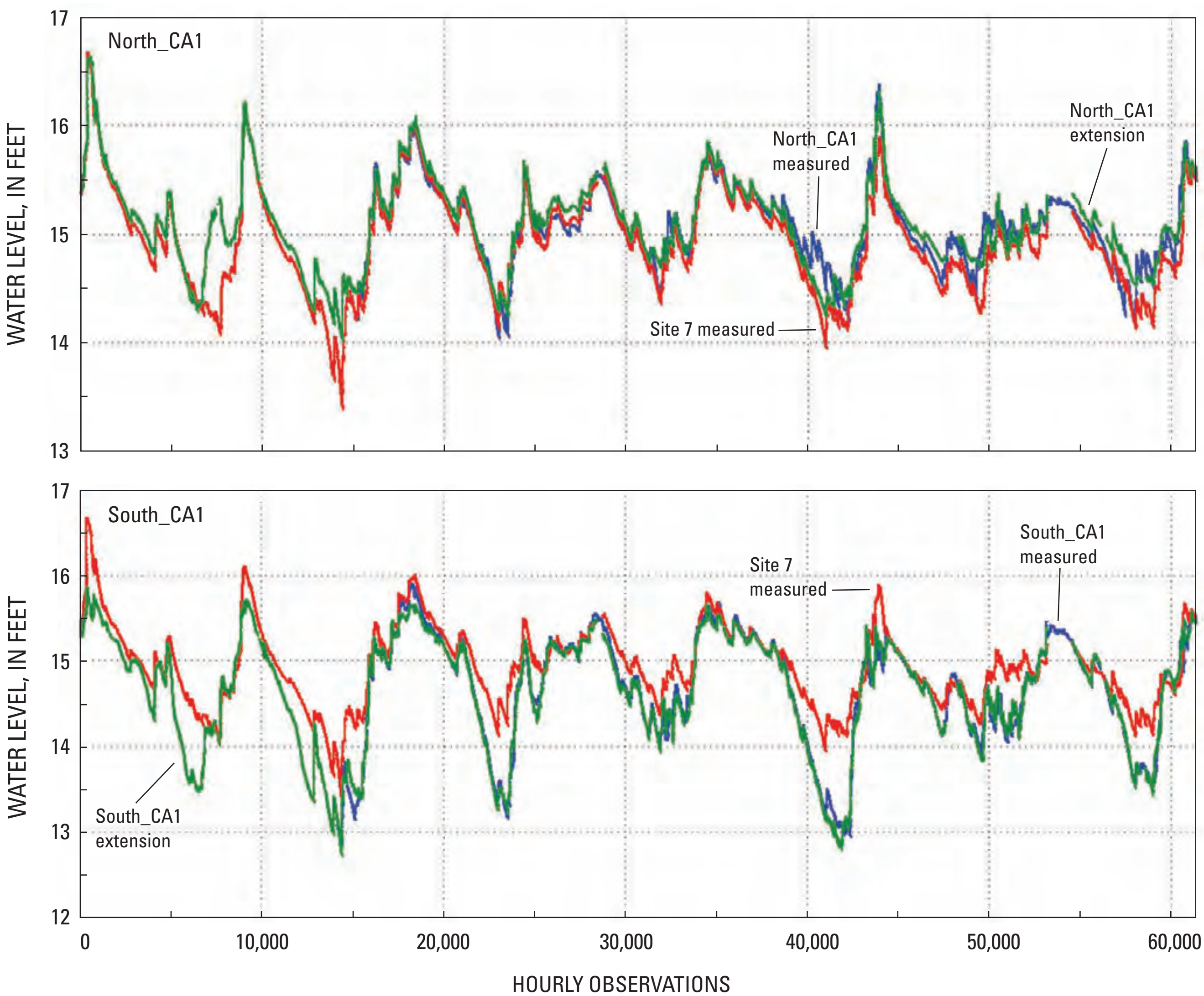
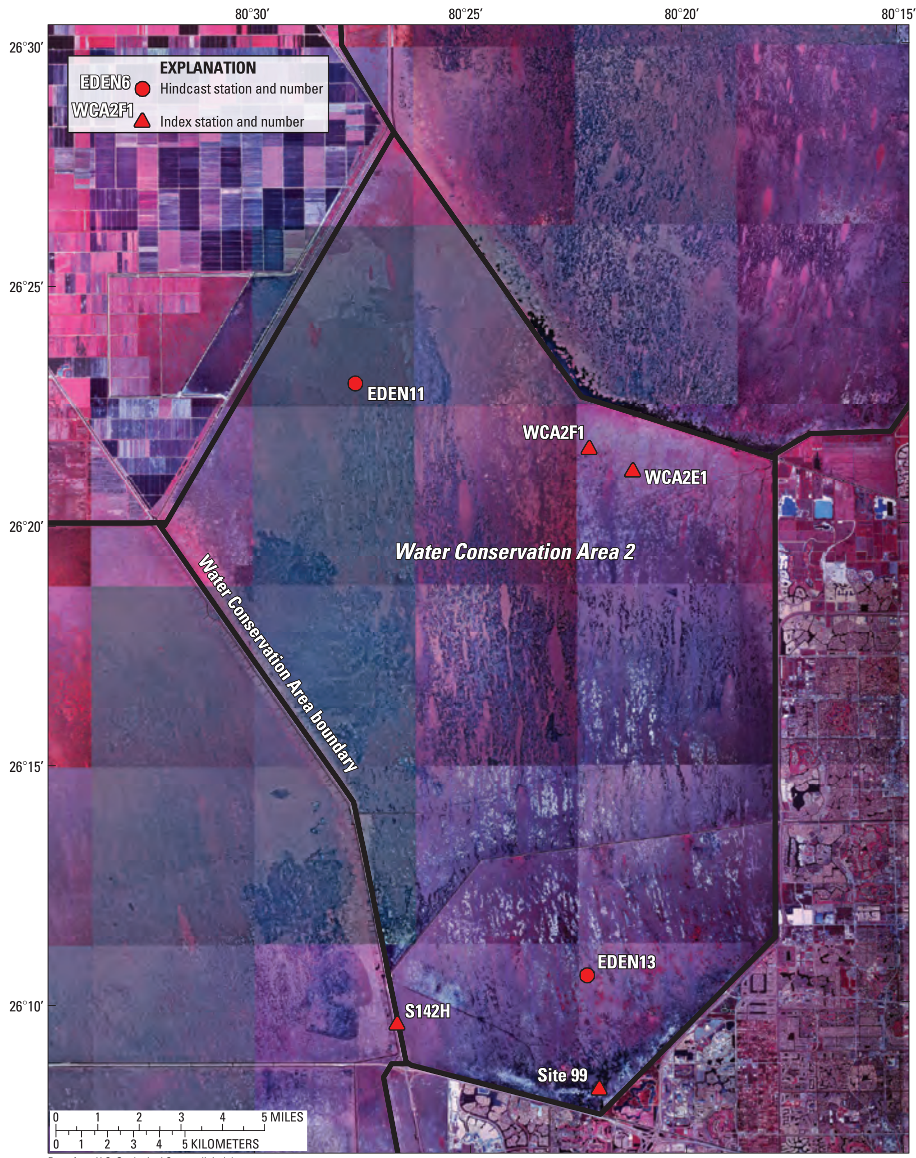


Figure 16. Water-level record extensions for sites North_CA1 and South_CA1 in Water Conservation Area 1 for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Water Conservation Area 2 (WCA2)

For WCA2, water-level records were extended for two stations—EDEN11 and EDEN13. The two new EDEN stations had very limited periods of record of approximately 4 months (table 1). For EDEN11, the water-level estimates were made using the three-step model described previously. The static and dynamic models used four static-input variables (CELL_X, PCTPRAIRIE, PCTEXOTICS, and PCTUPLAND). Two index stations were used in the models—WCA2E1 and WCA2F1 (fig. 17; table 1). The measured and estimated water levels for the EDEN11 station are shown in figure 18. WCA2E1 is also plotted with the hindcast to show the general water-level behavior through the record extension. The R^2 for the estimates is 0.90 with a PME of 2.0 percent (table 4). The measured data cover a large range of water levels, and there are only a few periods where the models extrapolated to estimate low and high water-level periods.

For EDEN13, the water-level estimates using the three-step modeling approach produced an unsatisfactory result, probably due to the limited number and spatial extent of index stations. An approach used in the snail kite study (Conrads and others, 2006) was used to extend the water levels for this station. Highly correlated data, by definition, are dynamically similar. The modeling challenge is not how the time series are similar but how they are dynamically different. The difference between EDEN13 and Site 99 (the index station closest to the station, fig. 17) was computed (fig.19A) to emphasize the



Base from U.S. Geological Survey digital data,
6-meter DOQ mosaic of south Florida, 2000,
Florida East, Zone 0901
Transverse Mercator projection
NAD 83 datum

Figure 17. Index and hindcast stations for Water Conservation Area 2.

dynamic difference between the two stations. The difference between Site 99 and EDEN13 was modeled using Site 99 and S142H as inputs. The final prediction is the sum of the EDEN13 difference and the Site 99 water level, which is shown in figure 19B with the Site 99 measured data. The R^2 for the estimates is 0.91, and the PME is 1.0 percent (table 4). Although these statistics indicate a good model, it should be noted that because of missing data at index station S142H, the range of EDEN13 water-level estimates was limited to approximately half of the hindcasted range. Water levels beyond this range require an extrapolation of the model.

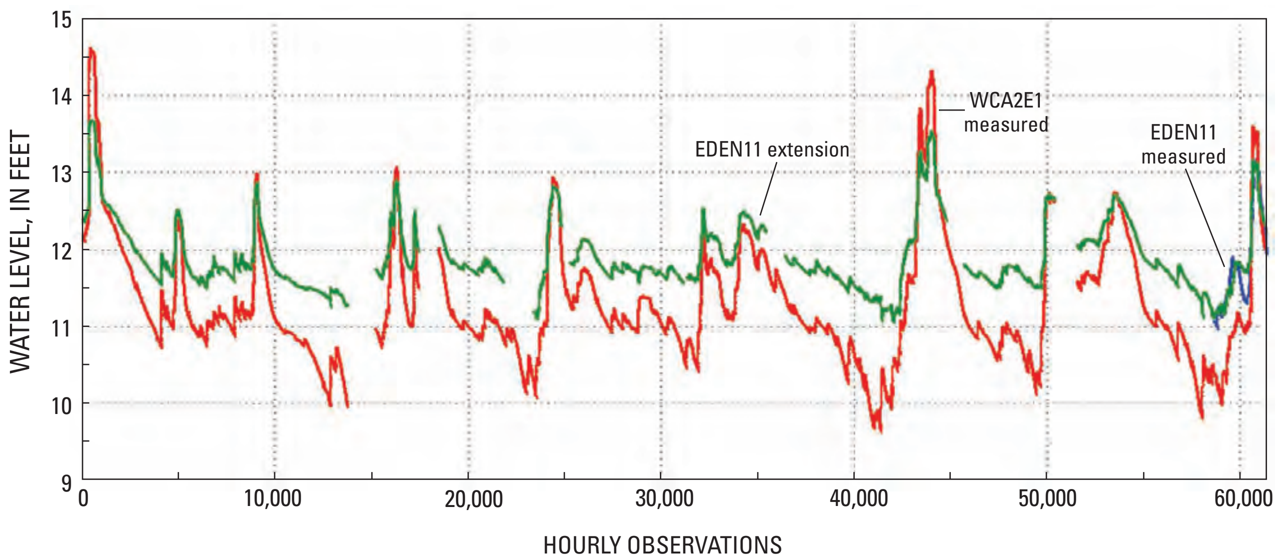


Figure 18. Water-level record extensions for site EDEN11 in Water Conservation Area 2 for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

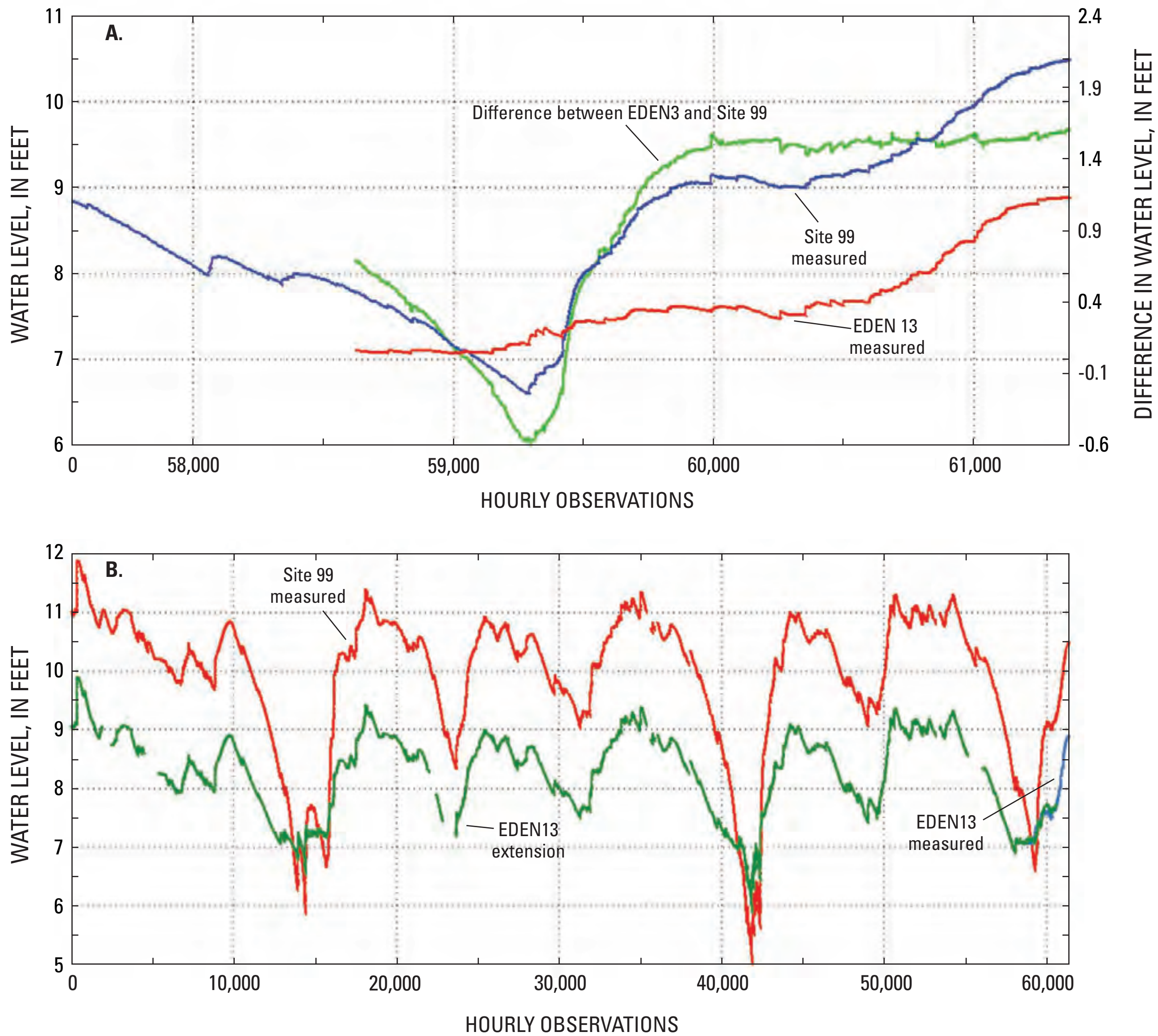


Figure 19. Difference in (A) water levels between Site 99 and site EDEN13 and (B) water-level record extensions for site EDEN13 in Water Conservation Area 2 for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Water Conservation Area 3A (WCA3A)

For WCA3A, water-level records were extended for 13 stations. The large number of index stations was separated into groups of similar hydrologic behaviors. Previous cluster analysis of the EDEN stations in the area indicated two groups of stations that had similar water-level behaviors (Conrads and Roehl, 2006a). Water-level data for EDEN, including WCA3A, manifests short-term variability and long-term hydrologic behavior. A dynamic clustering technique (Roehl and others, 2006) was used to cluster stations with similar long-term behavior. The hourly data were converted to daily values of 7-day moving-window averages, and a cross-correlation matrix of the Pearson coefficient of selected water-level time series was generated. The Pearson coefficient provides a numerical value of the amount of information that is shared between two time series, or signals. The cluster analysis of the cross-correlation matrix grouped together time series of similar behaviors. The rows were clustered using a k-means algorithm. The number of classes was determined by the sensitivity of the mean square error to k. The analysis showed two classes (groups) of stations of similar behavior in WCA3A (fig. 20).

Of the two groups of index and hindcast stations in WCA3A, 11 of the new EDEN stations are in group 1, and 2 are in group 2 (table 1; fig. 20). For all of the new EDEN stations in WCA3A, the three-step modeling approach was used for making the water-level estimates. For the WCA3A group 1 stations, five stations were used as index stations: Site 63, Site 64, Site 65, 3ASW, and 3AS3W1. The static and dynamic models used five static-input variables (CELL_X, PCTEXOTICS, PCTSLOUTH, PCTPRAIRIE, and PCTOTHER). The R^2 for the water-level estimates for the 11 new EDEN stations in group 1 were all above 0.99, and the PME were all 0.1 percent or less (table 4). The measured and estimated water levels for the 11 stations in group 1 are shown in figures 21–24. Site 64 also is plotted with the hindcasts for group 1 to show the general water-level behavior through the record extension. For the majority of the stations, the water-level estimation models cover a large range of the data and do not have to extrapolate over large portions of the range of the hindcasted data. The exception is EDEN9 (fig. 22) for which there was less than 40 days of measured data (tables 1 and 4).

For group 2 in WCA3A, there are two hindcast stations (EDEN5 and EDEN14) and three index stations (Site 62, 3A9, and 3A12) (fig. 20). The static model used three static-input variables (CELL_X, CELL_Y, and PCTPRAIRIE) and the dynamic model used two static-input variables (CELL_Y and

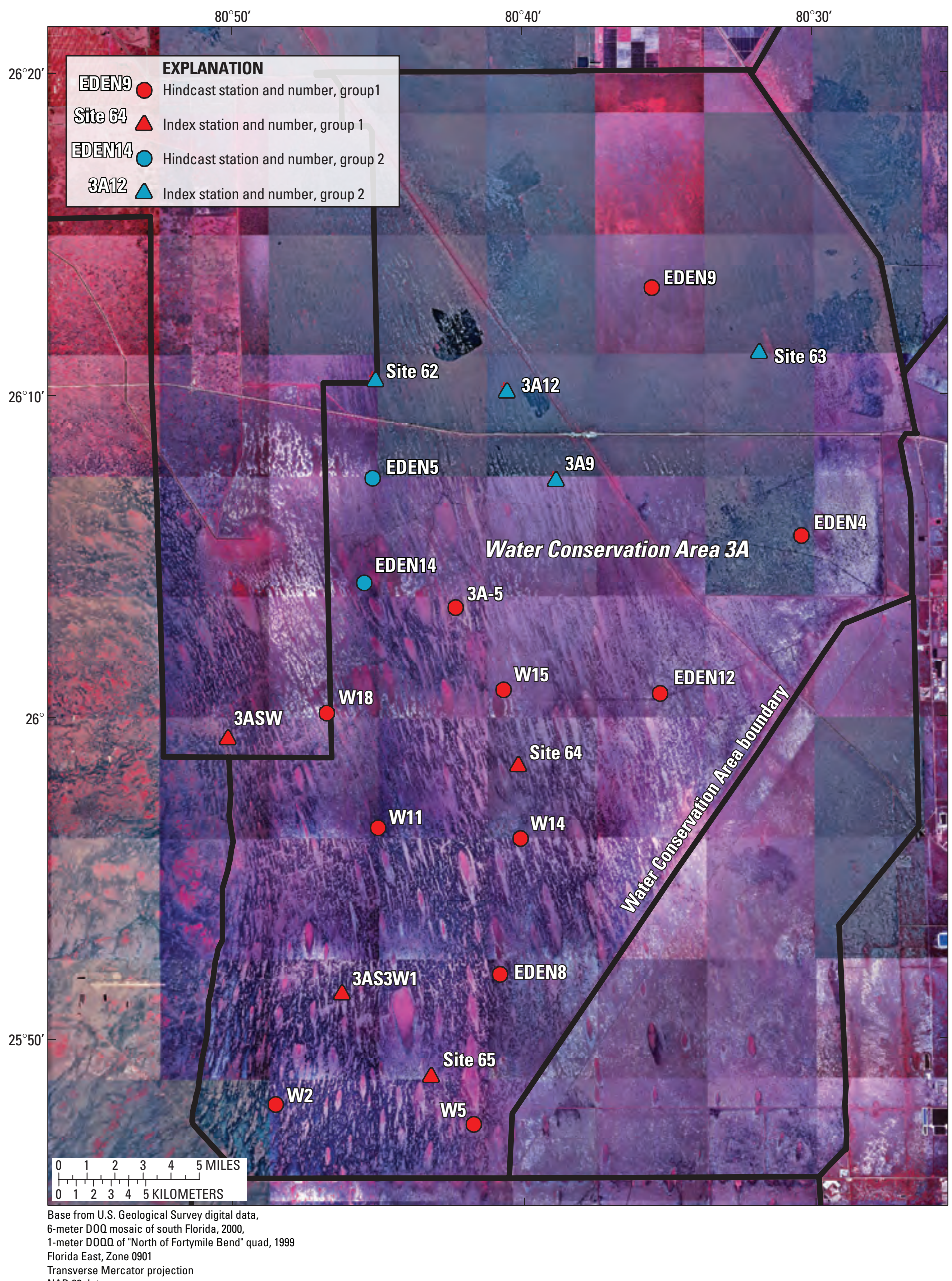


Figure 20. Index and hindcast stations for Water Conservation Area 3A.

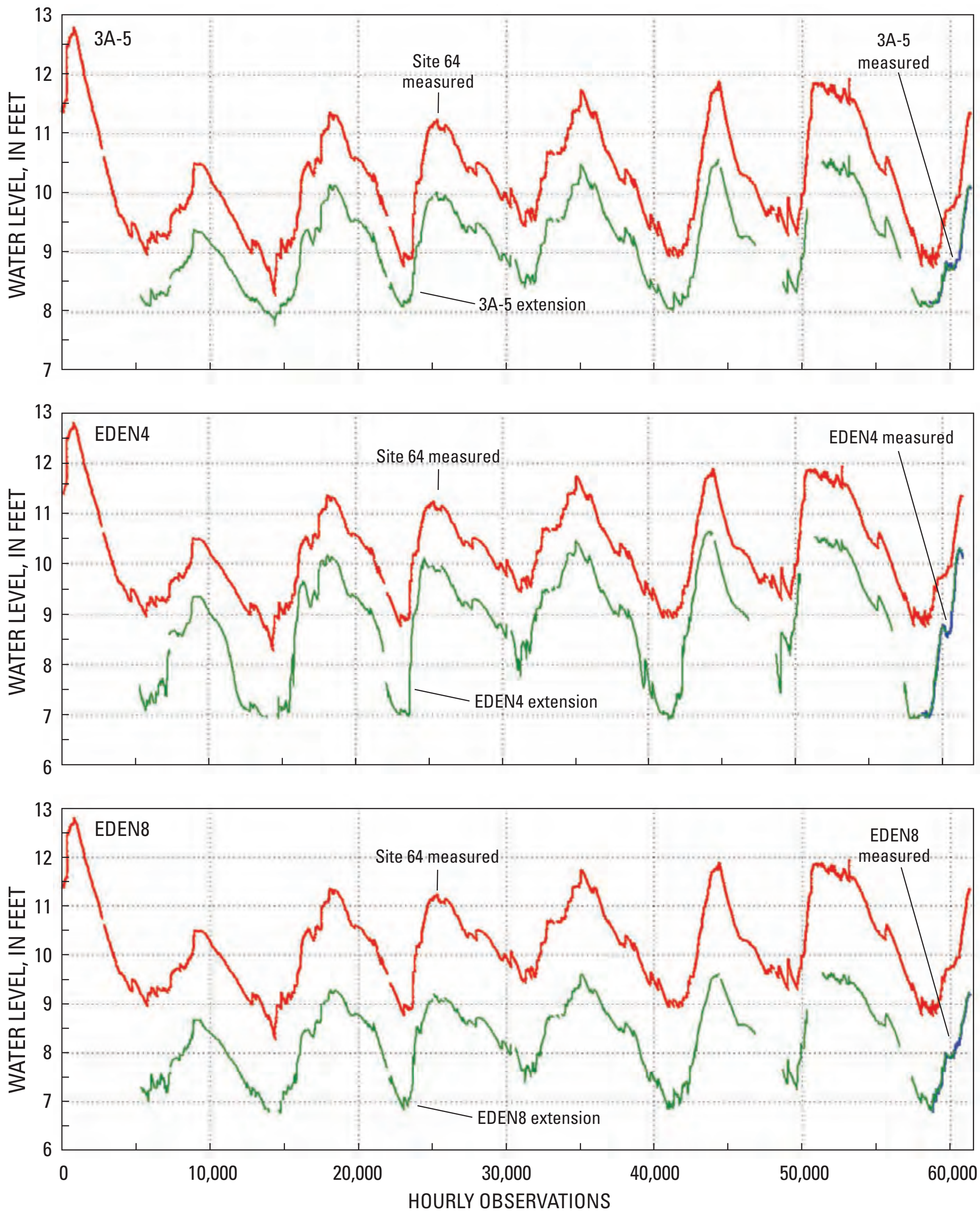


Figure 21. Water-level record extensions for sites 3A-5, EDEN4, and EDEN8 in Water Conservation Area 3A for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

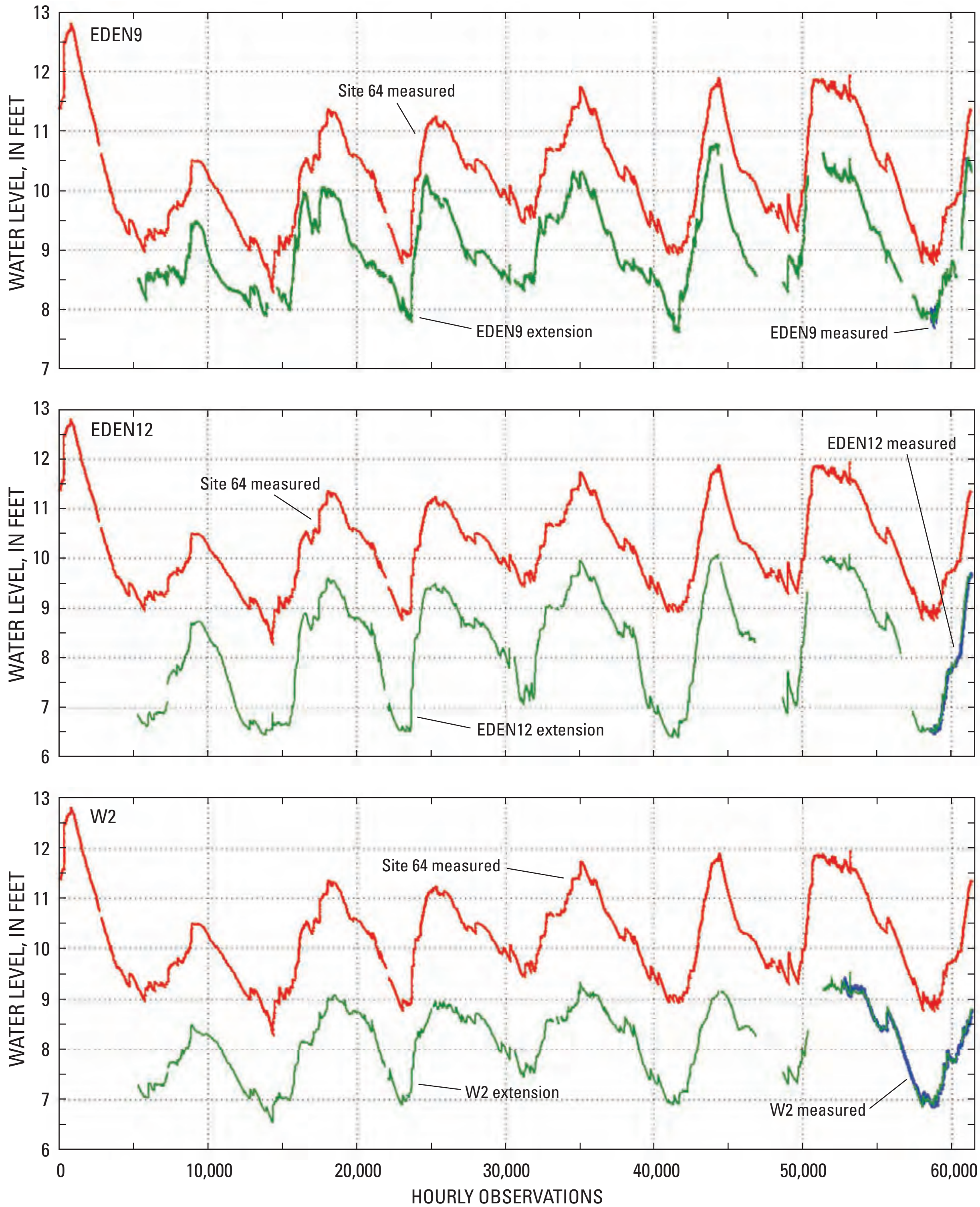


Figure 22. Water-level record extensions for sites EDEN9, EDEN12, and W2 in Water Conservation Area 3A for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

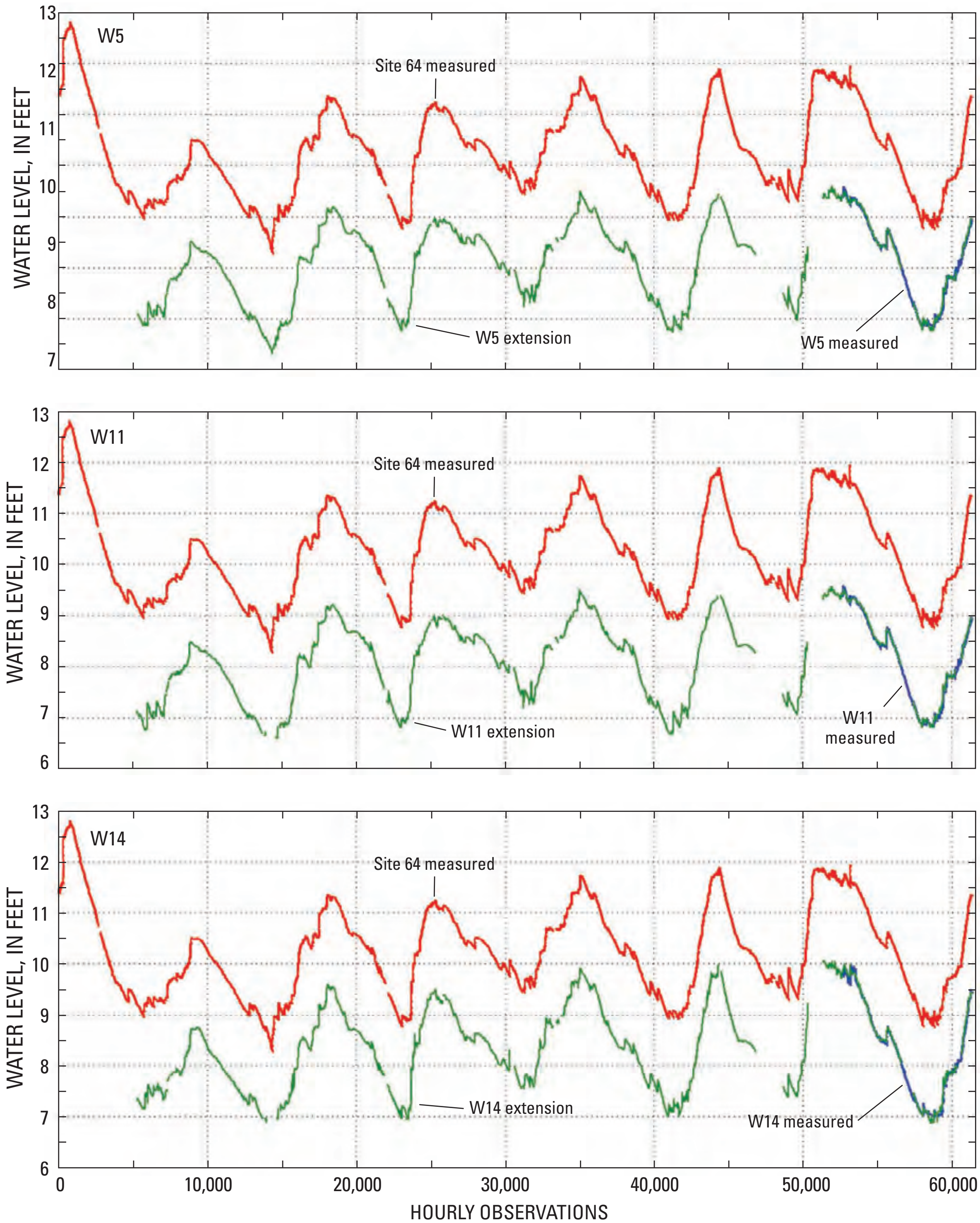


Figure 23. Water-level record extensions for sites W5, W11, and W14 in Water Conservation Area 3A for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

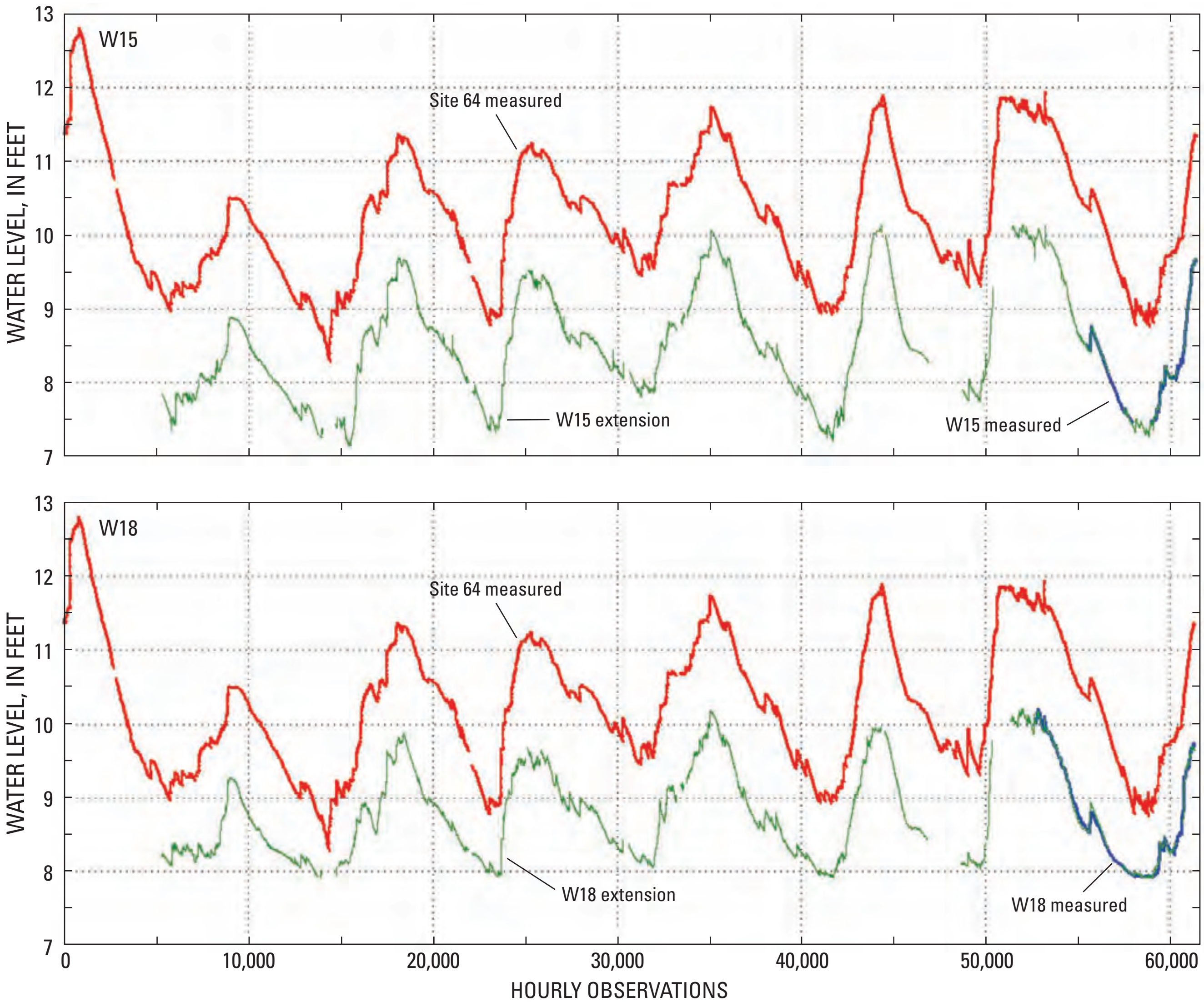


Figure 24. Water-level record extensions for sites W15 and W18 in Water Conservation Area 3A for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

PCTPRAIRIE) in addition to the dynamic input variables from the three index stations. The measured and estimated water levels for the group 2 hindcasts are shown in figure 25. Similar to EDEN9 in the group 1 stations, EDEN5 and EDEN14 have short periods of record. EDEN5 has less than 4 months of data, and EDEN14 has 2 months of data. The measured data for EDEN5 is at the extremes of the measured range of data with a large data gap in the mid-range of the data. The water-level estimates capture the overall trend of the high and low water levels as evident in the R^2 of 0.99 (table 4). Because the measured water levels are during periods of the low and high water levels, the model only has to extrapolate for a few high and low water periods. The measured data for EDEN14 are limited to the mid and lower water levels. The R^2 for the water-level estimate for EDEN14 is 0.82, the lowest of all the water-level estimates. The model has to extrapolate to estimate all of the high-water conditions.

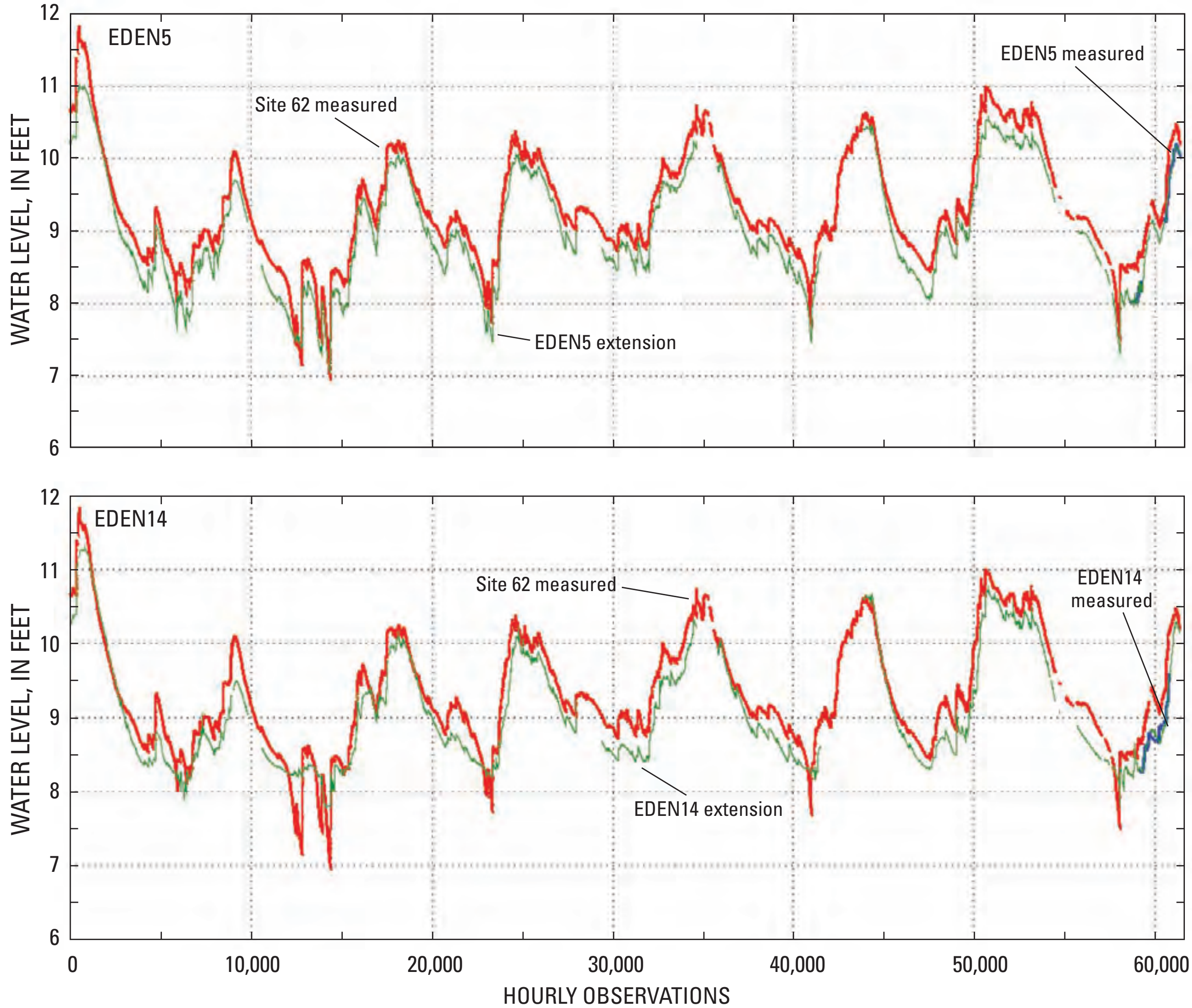


Figure 25. Water-level record extensions for sites EDEN5 and EDEN14 in Water Conservation Area 3A for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Water Conservation Area 3B (WCA3B)

For WCA3B, water-level records were extended for four stations—TI-8, TI-9, EDEN7, and EDEN10 (fig. 26; table 1). The static and dynamic models used three static-input variables (pctprairie, pctsawgrass, and pctupland). The dynamic models used dynamic-input variables from four index stations. The measured and estimated water levels for the WCA3B hindcasts are shown in figures 27 and 28 in addition to water levels for index station Site 69. All of the final estimates were able to accurately simulate the measured data (R^2 above 0.99), but only one station, EDEN7 (fig. 28), had measured data with a range similar to the range of the hindcasts. The other three stations did not have data in the higher water-level range, and the model had to extrapolate to estimate high-water conditions at these stations.

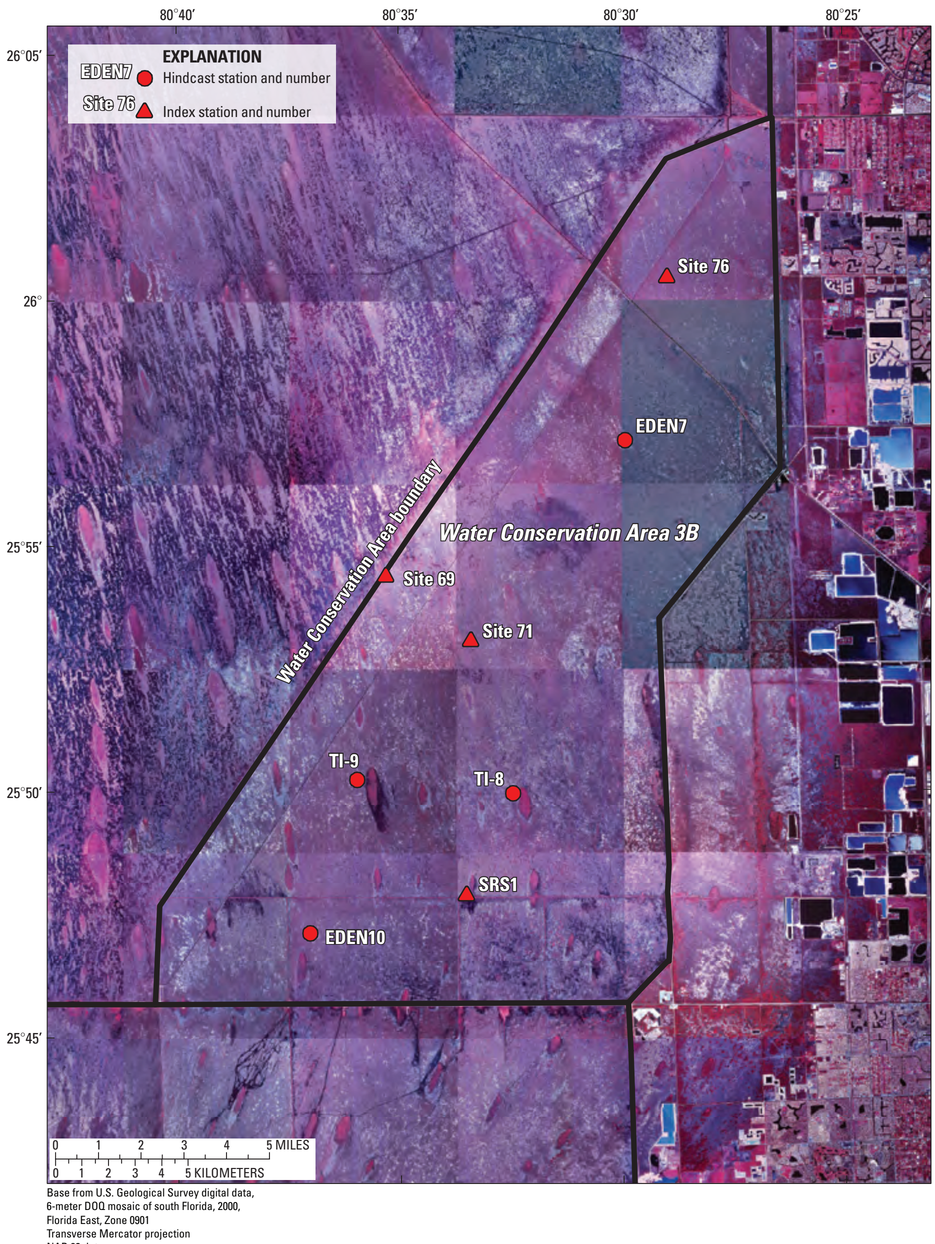


Figure 26. Index and hindcast stations for Water Conservation Area 3B.

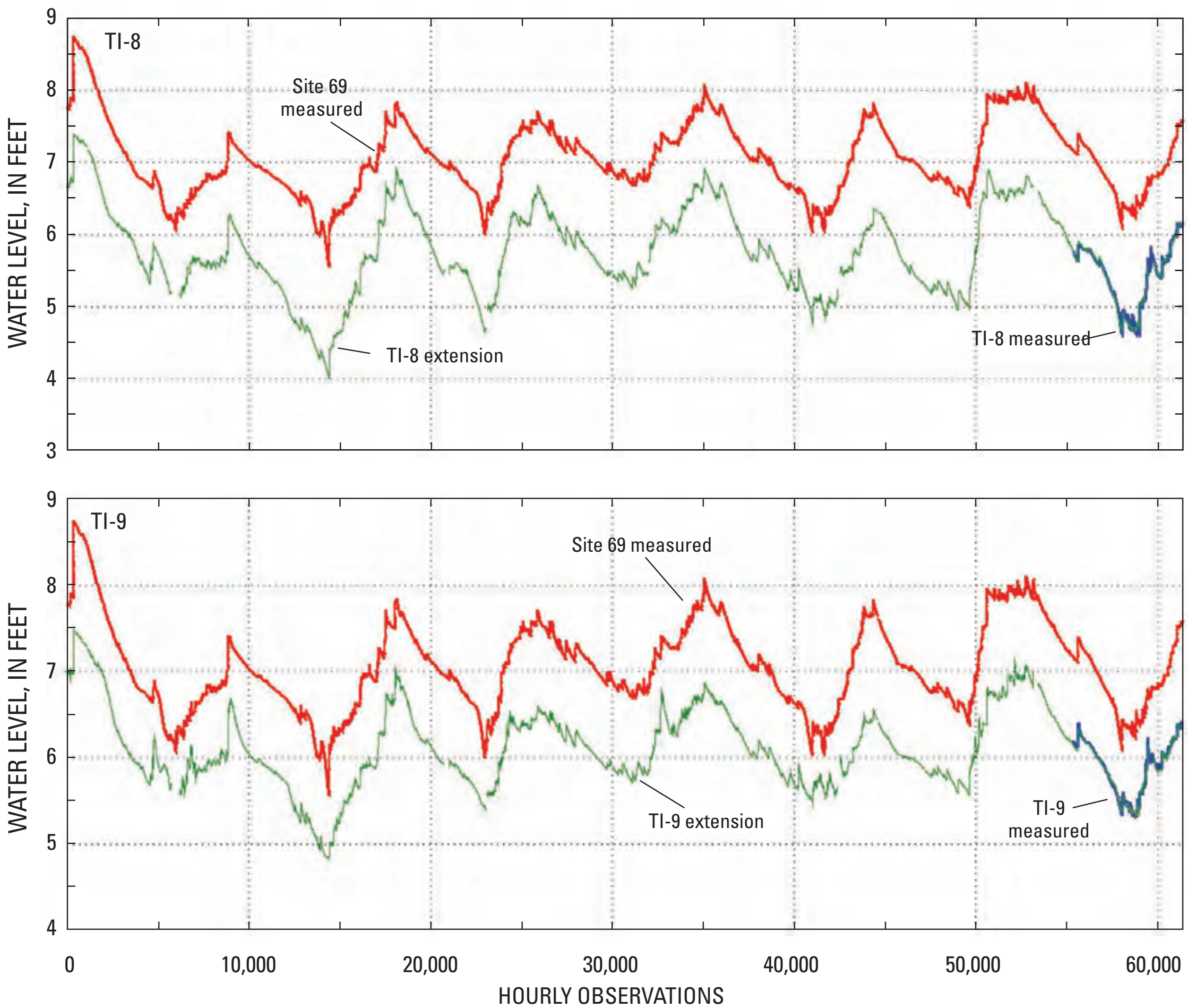


Figure 27. Water-level record extensions for sites TI-8 and TI-9 in Water Conservation Area 3B for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

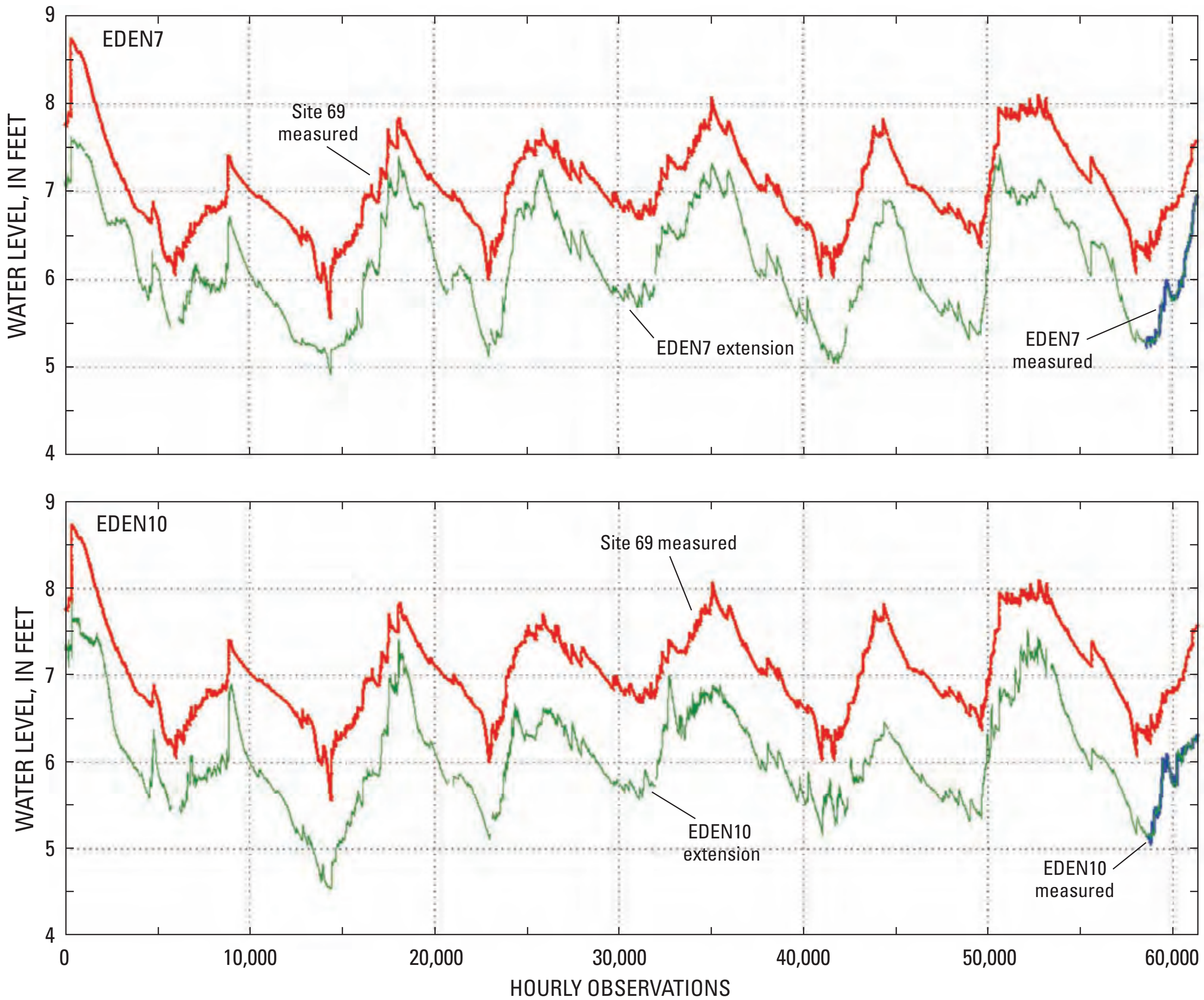


Figure 28. Water-level record extensions for sites EDEN7 and EDEN10 in Water Conservation Area 3B for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Big Cypress National Preserve (BCNP)

For BCNP, water-level records were extended for two stations—EDEN1 and EDEN6 (fig. 29; table 1). The initial three-step model produced satisfactory water-level estimates for EDEN1 but not for EDEN6. To improve the water-level estimates for EDEN6, the same three-step modeling approach was used but with a different combination of static- and dynamic-input variables. The dynamic model for EDEN6 also used water-level data from 3ASW in WCA3A (fig. 20). As with the other areas, the measured data from the new EDEN station were not used until computing the residual error of the initial water-level estimates. The static model for EDEN1 used five static-input variables, whereas the static model for EDEN6 used two static-input variables. The dynamic model for EDEN1 used four static-input variables and four dynamic-input variables, whereas the dynamic model for EDEN6 used two static variables and three dynamic variables.

The measured and estimated water levels for the two stations are shown in figure 30. The R^2 for the water-level estimates for EDEN1 and EDEN6 are 0.92 and 0.97, respectively (table 4). Water levels for BCA9 also are shown with the hindcasts for EDEN1, and water levels for BCA5 are shown with the hindcasts for EDEN6 (fig. 30). The measured data for the two new EDEN stations in BCNP are limited to the mid- and high-water ranges. The measured data range for EDEN1 is less than a foot, whereas the range of the hindcasts is more than 3 ft. For EDEN6, the range of the measured data is about 2 ft, and the range of the hindcasts is about 5 ft.

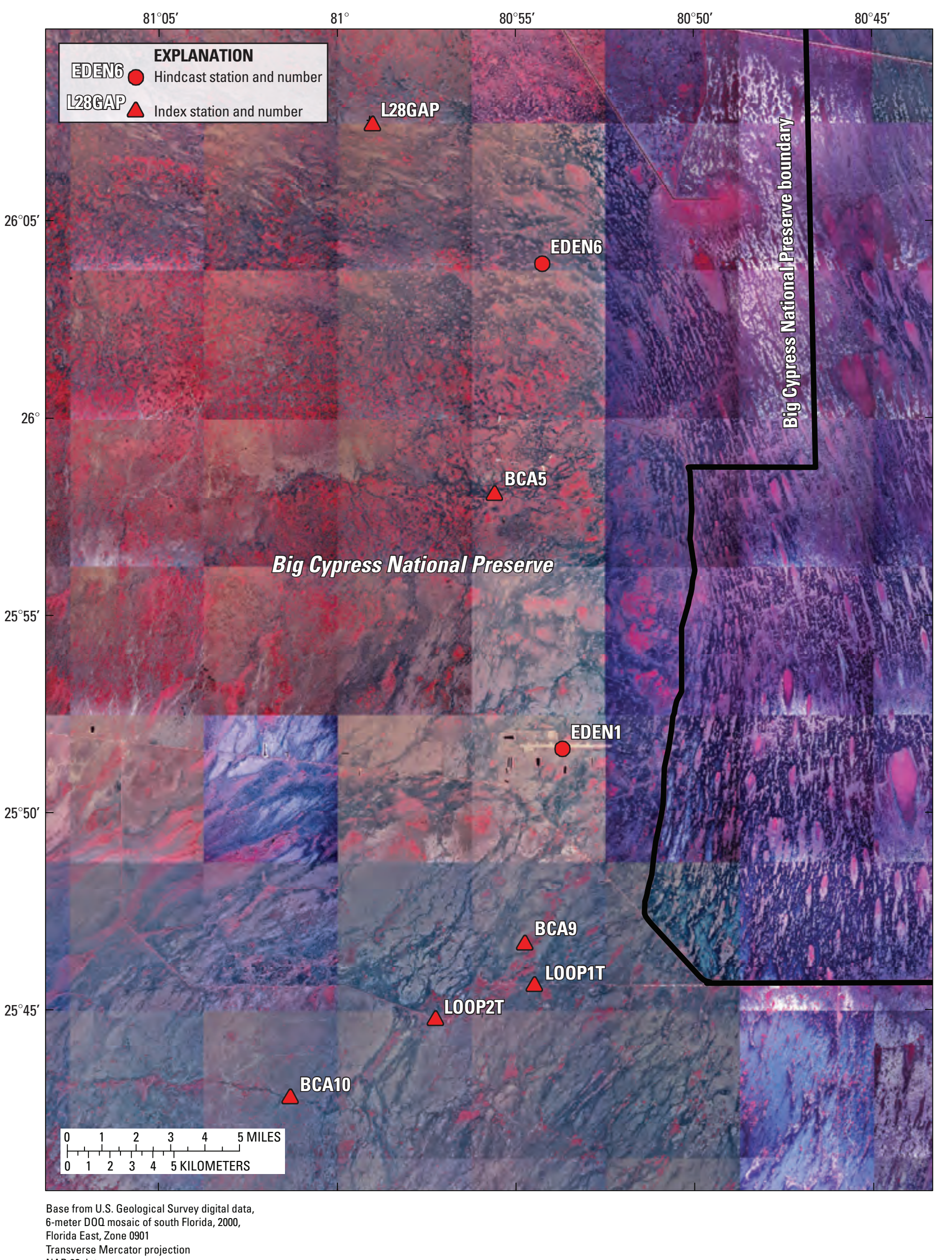


Figure 29. Index and hindcast stations for Big Cypress National Preserve.

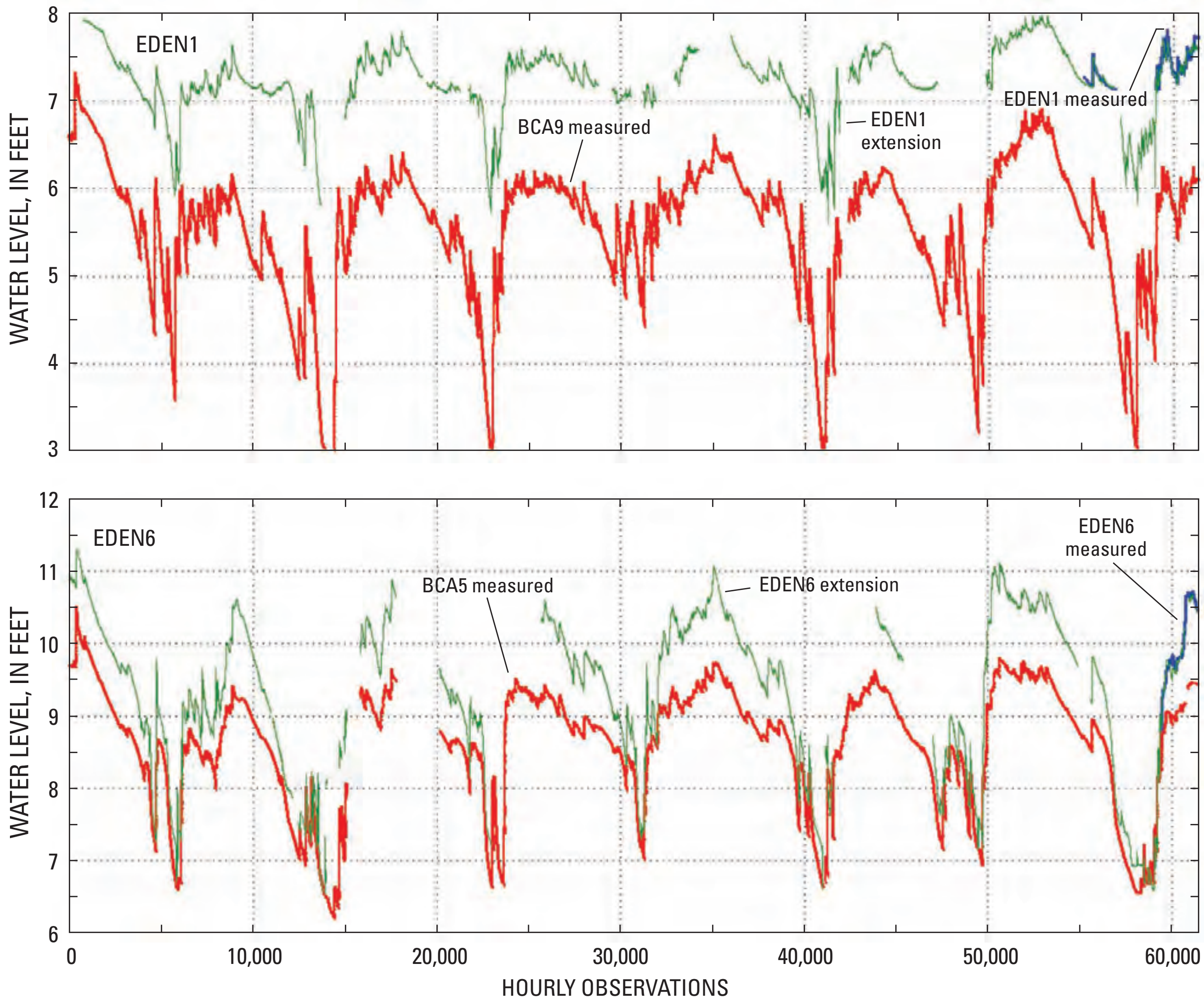
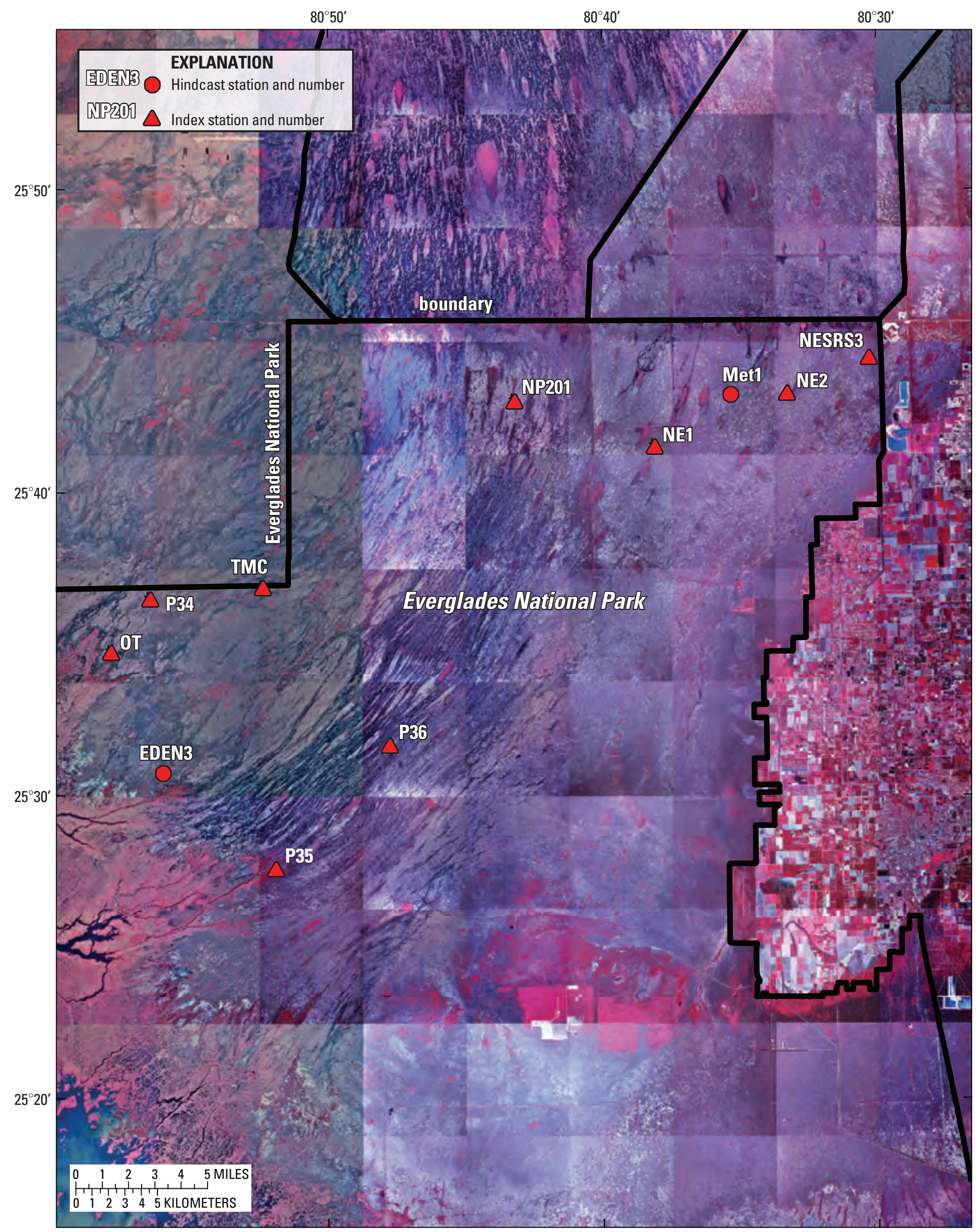


Figure 30. Water-level record extensions for sites EDEN1 and EDEN6 in Big Cypress National Preserve for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Everglades National Park (ENP)

For the ENP, water-level records were extended for two stations—EDEN3 and Met1 (fig. 31; table 1). Similar to the two new EDEN stations in BCNP, the stations were modeled separately with the three-step modeling approach. Both static models used four static-input variables. Three of the static-input variables were the same (CELL_Y, pctsawgrass, and pctSLOUGH) for both models. The static model for Met1 included percent upland (PCTUPLAND), and the static model for EDEN3 included percent prairie (PCTPRAIRIE). Both dynamic models used four static variables but used two different combinations of five dynamic variables. The measured and estimated water levels for the two stations are shown in figure 32. The R^2 for the water-level estimates for EDEN3 and Met1 is 0.98 and 0.99, respectively (table 4). Water levels for P34 also are shown with the hindcasts for EDEN3, and water levels for NE1 are shown with the hindcasts for Met1. The range of the measured data for EDEN3 covers the majority of the range of the hindcasts (fig. 32). The range of measured data for Met1 is less than 0.5 ft, and the model must extrapolate to estimate annual high and low water.



Base from U.S. Geological Survey digital data,
6-meter DOQ mosaic of south Florida, 2000,
Florida East, Zone 0901
Transverse Mercator projection
NAD 83 datum

Figure 31. Index and hindcast stations for Everglades National Park.

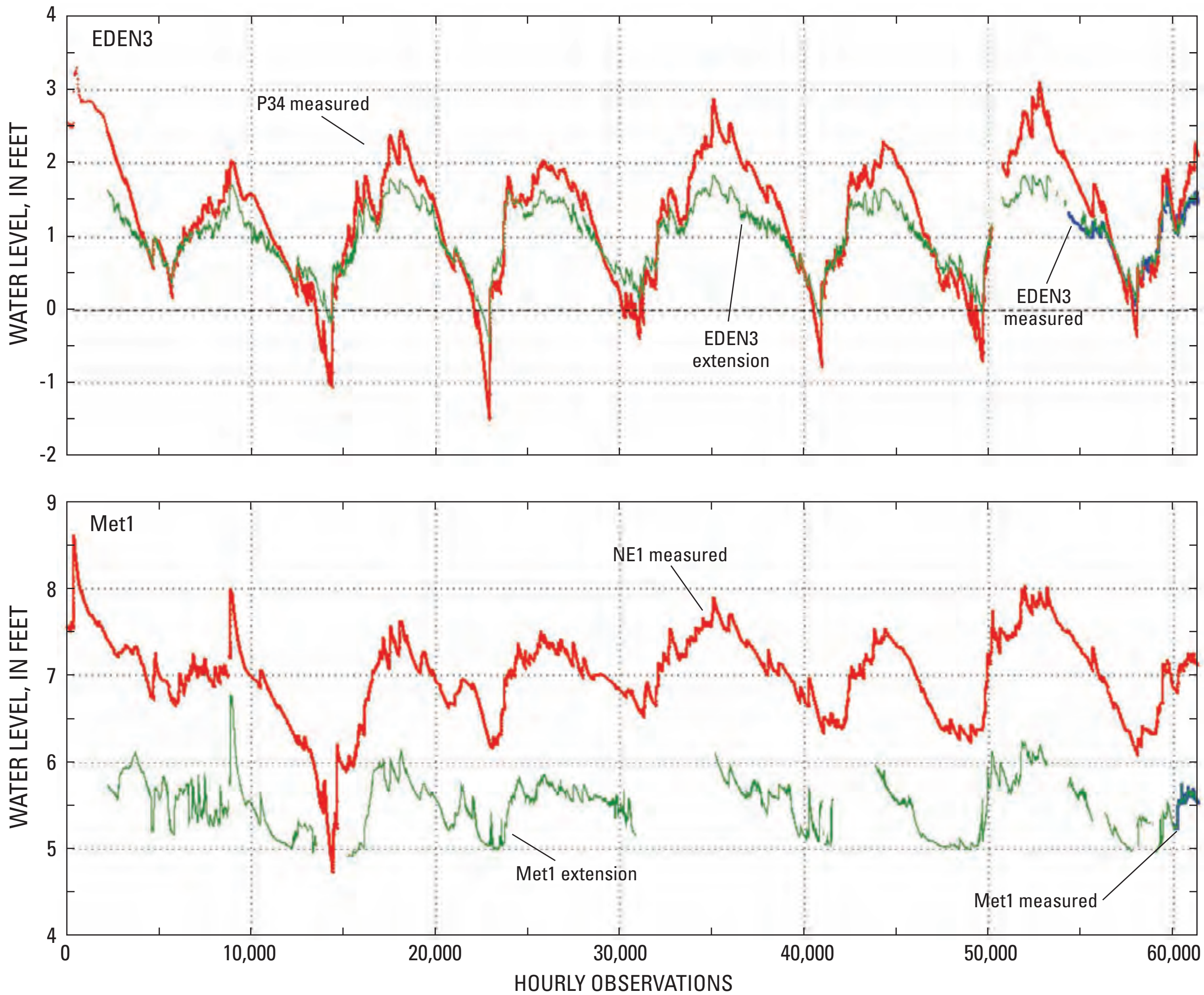


Figure 32. Water-level record extensions for sites EDEN3 and Met1 in the Everglades National Park for the period October 1, 1999, to September 30, 2006. Breaks in the water-level estimates are caused by incomplete time series for one or more of the model inputs.

Summary and Discussion

The Everglades Depth Estimation Network (EDEN) is an integrated network of real-time water-level gaging stations, ground-elevation models, and water-surface models. The network provides scientists, engineers, and water-resource managers with current (2000–present) water-depth information for the entire freshwater portion of the greater Everglades. To increase the accuracy of the water-surface models, 25 real-time gaging stations were added to EDEN. To incorporate the newly added stations to the 7-year EDEN database of 253 water-level gaging stations in the greater Everglades, the short-term water-level records (generally less than 1 year) needed to be simulated back in time (hindcasted) to be concurrent with data from the established gaging stations. A three-step modeling approach using artificial neural networks (ANN) models was used to estimate the water levels at the new stations. The ANN models used static variables of gaging station location and percent vegetation in addition to dynamic variables of water-level data from the established EDEN gaging stations. The final step of the modeling approach was to simulate the computed error of the initial estimate to increase the accuracy of the final water-level estimate.

The three-step modeling approach for estimating water levels produced satisfactory results, with coefficients of determination for 21 of the 25 for the new EDEN stations greater than 0.95 and all of the estimates greater than 0.82. For some new EDEN stations having limited data, the record extension (hindcasts) included periods beyond the range of data used to train the water-level estimate models. A comparison of the hindcasts from these models with long-term water levels proximal to the new EDEN station indicated that the water-level estimates showed a similar hydrologic response to the long-term water levels.

There are opportunities to improve the completeness and accuracy of the hindcasts of the new EDEN stations. The completeness of the hindcasts could be improved by filling in the missing data for the established EDEN stations using similar techniques as presented in this report, which in turn would provide more complete input time series to the water-level estimation models for the new stations.

For the majority of the new EDEN stations, there was 1 year or less of data to train and test the water-level estimation models. Over time, a greater record of hydrologic behaviors at the new EDEN stations will be collected, including inter-annual variability. Retraining the water-level estimation ANN models will improve the ability of the models to predict the behaviors at the new EDEN stations. Additional cell attribute data for the EDEN grid and enhancements to the existing attribute data will also improve the models.

Acknowledgments

The complexity of this study required interagency cooperation in addition to individual contributions. The authors thank the members of the EDEN project team: Ikuko Fujisaki, Aaron Higer, Monica Palaseanu, Leonard Pearlstine, and Darcy Thomas from the University of Florida and Heather Henkel, John W. Jones, Eduardo Patino, Carolyn Price, Pamela Telis, and Arturo Torres of the USGS and Roy Sonenshein of the National Park Service (formerly with the USGS). The authors also thank the funding agencies that support the EDEN project team—U.S. Army Corps, Jacksonville District, and the USGS Greater Everglades Priority Ecosystems Science.

References Cited

Conrads, P.A., and Roehl, E.A., Jr., 2005, Integration of data mining techniques with mechanistic models to determine the impacts of non-point source loading on dissolved oxygen in tidal waters: Proceedings of the South Carolina Environmental Conference, March 2005, Myrtle Beach, SC, 8 p.

Conrads, P.A., and Roehl, E.A., Jr., 2006a, Application of dynamic clustering algorithm to the water-level hydrographs of the EDEN hydrologic network [abs. and poster]: Greater-Everglades Ecosystem Restoration Conference—Planning, Policy, and Science, June 2006, Lake Buena Vista, FL, University of Florida, Institute of Food and Agricultural Gainesville, FL, p. 39.

Conrads, P.A., and Roehl, E.A., Jr., 2006b, Estimating water depths using artificial neural networks, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics, August 2006, Nice, France, v. 3, p. 1643–1650.

Conrads, P.A., Roehl, E.A., Daamen, R.C., and Kitchens, W.M., 2006, Using artificial neural network models to integrate hydrologic and ecological studies of the snail kite in the Everglades, USA, *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics, August 2006, Nice, France, v. 3, p. 1651–1658.

Desmond, G.D., 2003, Measuring and mapping the topography of the Florida Everglades for Ecosystem Restoration: U.S. Geological Survey Fact Sheet 021-03, 4 p.

Devine, T.W., and Roehl, E.A., Jr., 2003, Virtual sensors—Cost effective monitoring: Proceedings from the Air and Waste Management Association Annual Conference, June 2003, San Diego, CA, 10 p.

German, E.R., 2000, Regional evaluation of evapotranspiration in the Everglades: U.S. Geological Survey Water-Resources Investigations Report 00-4217, 48 p.

Hinton, G.E., 1992, How neural networks learn from experience: Scientific American, v. 267, no. 3, p. 145–151.

Jensen, B.A., 1994, Expert systems—neural networks, Instrument engineers' handbook (3d ed.): Chilton Book Company, Radnor, PA, p. 48–54.

Jones, J.W., and Price, S.D., 2007, Everglades Depth Estimation Network (EDEN) digital elevation model research and development: U.S. Geological Survey Open-File Report 2007-1034, 29 p.

Lodge, T.E., 1994, The Everglades handbook, Understanding the ecosystem: Delray Beach, FL, St. Lucie Press, 228 p.

Pearlstine, L., Higer, A., Palaseanu, M., Fujisaki, I., and Mazzotti, F., 2007, Spatially continuous interpolation of water stage and water depths using the Everglades Depth Estimation Network (EDEN): Gainesville, FL, Institute of Food and Agricultural, University of Florida, CIR1521, 18 p., 2 apps.

Risley, J.C., Roehl, E.A., Jr., and Conrads, P.A., 2003, Estimating water temperatures in small streams in western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02-4218, 60 p.

Roehl, E.A., Jr., Conrads, P.A., and Cook, J.B., 2003, (Discussion of) Using complex permittivity and artificial neural networks for contaminant prediction: Journal of Environmental Engineering, v. 129, p. 1069–1071.

Roehl, E., Risley, J., Stewart, J., and Mitro, M., 2006, Numerically optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems—part 1, *in* Voinov, Alexey, Jakeman, A.J., and Rizzoli, A.E., eds., Proceedings of the International Environmental Modelling and Software Society (iEMSs) Third Biennial Meeting, Summit on Environmental Modelling and Software, July 2006, Burlington VT, 6 p.; accessed November 5, 2007, at *http://www.iemss.org/iemss2006/sessions/all.html/*

Rosenblatt, F., 1958, The Perceptron, a probabilistic model for information storage and organization in the brain: Psychological Review, v. 65, p. 386–408.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986, Learning internal representations by error propagation, in Rumelhart, D.E., McClelland, J.L., and the PDP Research Group, Parallel distributed processing—Explorations in the microstructure of cognition, v. 1: Cambridge MA, MIT Press, p. 318–362.

Stewart, J., Mitro, M., Roehl, E., and Risley, J., 2006, Optimized empirical modeling of highly dynamic, spatially expansive, and behaviorally heterogeneous hydrologic systems—part 2: *in* Gourbesville, Philippe, Cunge, Jean, Guinot, Vincent, and Liong, Shie-Yui, eds., Proceedings of the 7th International Conference on Hydroinformatics, August 2006, Nice, France, v. 2, p. 831–838.

Telis, P.A., 2005, Project description, South Florida Surface Water Monitoring Network for the Support of MAP Projects (known as EDEN, Everglades Depth Estimation Network); accessed October 2, 2007, at *http://sofia.usgs.gov/projects/eden/*

Telis, P.A., 2006, The Everglades depth estimation network (EDEN) for support of ecological and biological assessments: U.S. Geological Survey Fact Sheet 2006-3087, 4 p.

U.S. Army Corps of Engineers, 1999, Central and Southern Florida Project comprehensive review study: Jacksonville, FL, Final Integrated Feasibility Report and Programmatic Environmental Impact Statement, variously paged, 4 annexes, 15 apps.

Weiss, S.M., and Indurkhya, Nitin, 1998, Predictive data mining—A practical guide: San Francisco, Morgan Kaufmann Publishers, Inc., p. 1.

(Blank)

Appendixes

1. Summary of artificial neural network models used in the study.
2. Summary statistics for the models used in the study.
3. Variables used in the artificial neural network models.

Appendix 1. Summary of artificial neural network models used in the study.

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
WCA1 Models					
wca1-static	static	CELL_Y PCTPRAIRIE PCTEXOTICS	WLSITE	WLSITE RES_WLSITE	Static model for WCA1- Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
wca1-dynamic	dynamic	CELL_Y PCTPRAIRIE PCTEXOTICS SITE_7 SITE8TDEC2 WCA1MEDEC SITE9DEC	RES_WLSITE	RES_WLSITE	Dynamic model for WCA1- RES_WLSITE prediction from dynamic model and WLSITE prediction from static model are used for initial water-level estimate.
site8tdecor	decorrelation	SITE_8T	SITE_7	SITE8TDEC2	Model to decorrelate SITE8t from SITE_7.
wca1medecor2	decorrelation	WCA1ME SITE8TDEC2	SITE_7	WCA1MEDEC	Model to decorrelate WCA1ME from SITE_8 and SITE_7.
northres	error correction	SITE_7 SITE8TDEC WCA1MEDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site NORTH.
southres	error correction	SITE_7 SITE8TDEC WCA1MEDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site SOUTH.
WCA2 Models					
wca2a-static	static	CELL_X PCTPRAIRIE PCTEXOTICS PCTUPLAND	WLSITE	WLSITE RES_WLSITE	Static model for EDEN11 in WCA2 - Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
wca2a-dynamic2	dynamic	CELL_X PCTPRAIRIE PCTEXOTICS PCTUPLAND WCA2E1 WCA2F1DEC	RES_WLSITE	RES_WLSITE	Dynamic model for EDEN11 in WCA2 RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
wca2f1decor	decorrelation	WCA2F1	WCA2E1	WCA2F1DEC	Model to decorrelate WCA2F1 from WCA2E1.
eden11res2	error correction	WCA2E1 WCA2F1DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN11.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
eden13	error correction	SITE_99 S142H	EDEN13DIF		Model to estimate water-level difference with SITE_99. Sum of difference and water-level estimate for SITE_99 are the water-level estimates for site EDEN13.
			WCA3A Models		
			Group 1 Models		
cluster1static	static	CELL_Y PCTEXOTICS PCTSLOUGH PCTPRAIRIE PCTOTHER	WLSITE	WLSITE RES_WLSITE	Static model for Group 1 stations in WCA3A—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
cluster1dynamic	dynamic	SITE_63DEC SITE_64 3AS3W1DEC CELL_Y PCTSLOUGH PCTEXOTICS PCTOTHER	RES_WLSITE	RES_WLSITE	Dynamic model for Group 1 stations in WCA3A—RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
3as3w1dec	decorrelation	3AS3W1	SITE_64	3AS3W1DEC	Model to decorrelate 3AS3W1 from SITE_64.
65dec	decorrelation	SITE_65 3AS3W1DEC	SITE_64	SITE_65DEC	Model to decorrelate Site_65 from SITE_64 and 3AS3W1.
63decr1	decorrelation	3AS3W1DEC SITE_65DEC SITE63	SITE_64	SITE_63DEC	Model to decorrelate Site_63 from SITE_64, 3AS3W1, and Site_65
3aswdec	decorrelation	3AS3W1DEC SITE_65DEC SITE_63DEC 3ASW	SITE_64	3ASWDEC	Model to decorrelate 3ASW from SITE_64, 3AS3W1, Site_65, and Site_63.
3a-5res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site 3A-5.
eden4res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN4.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
eden8res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN8.
eden9res2	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN9.
eden12res2	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN12.
w2res	error correction	SITE_63DEC SITE_64 3AS3W1DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W2.
w5res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W5.
w11res2	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W11.
w14res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W14.
w15res	error correction	SITE_63DEC SITE_64 3AS3W1DEC SITE_65DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W15.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
w18res	error correction	SITE_63DEC SITE_64 3AS3W1DEC 3ASWDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site W18.
				Group 2 Models	
cluster3_static	static	CELLX CELLY PCTPRAIRIE	WLSITE	WLSITE RES_WLSITE	Static model for Group 2 stations in WCA3A—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
cluster3_dynamic	dynamic	SITE_62 3A9DEC 3A12DEC CELLY PCTPRAIRIE	RES_WLSITE	RES_WLSITE	Dynamic model for Group 2 stations in WCA3A—RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
3a9decor	decorrelation	3A9	SITE_62	3A9DEC	Model to decorrelate 3A9 from SITE_62.
3a12decor	decorrelation	3A9DEC 3A12	SITE_62	3A12DEC	Model to decorrelate 3A12 from SITE_63 and 3A9.
eden5res	error correction	SITE_62 3A9DEC 3A12DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN5.
eden14res		SITE_62 3A9DEC 3A12DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN14.
				WCA3B Models	
cluster1-static	static	PCTPRAIRIE PCTSAWGRASS PCTUPLAND	WLSITE	WLSITE RES_WLSITE	Static model for Group 2 stations in WCA3A—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
cluster1-dynamic	dynamic	PCTPRAIRIE PCTSAWGRASS PCTUPLAND SITE_69C SITE71DEC SRS1DEC SITE76DEC	RES_WLSITE	RES_WLSITE	Dynamic model for WCA1- RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
site71dec	decorrelation	SITE_71C	SITE_69C	SITE71DEC	Model to decorrelate SITE_71 from SITE_69.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
srs1dec	decorrelation	SITE71DEC SRS1C	SITE_69C	SRS1DEC	Model to decorrelate SRS1 from SITE_69 and SITE_71.
site76dec	decorrelation	SITE71DEC SRS1DEC SITE_76	SITE_69C	SITE76DEC	Model to decorrelate SITE_76 from SITE_69, SITE_71, and SRS1.
TI-8res	error correction	SITE_69C SITE71DEC SRS1DEC SITE76DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site TI-8.
TI-9res	error correction	SITE_69C SITE71DEC SRS1DEC SITE76DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site TI-9.
eden7res	error correction	SITE_69C SITE71DEC SRS1DEC SITE76DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN7.
eden10res	error correction	SITE_69C SITE71DEC SRS1DEC SITE76DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN10.
BCNP Models					
bcnp-static	static	CELL_X CELL_Y PCTPRAIRIE PCTSAWGRASS PCTEXOTICS	WLSITE	WLSITE RES_WLSITE	Static model for EDEN1 in BCNP - Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
bcnp-dynamic	dynamic	CELL_X CELL_Y PCTPRAIRIE PCTEXOTICS BCA9 LOOP1TDEC LOOP2TDEC BCA10DEC	RES_WLSITE	RES_WLSITE	Dynamic model for EDEN1 in BCNP- RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
loop1tdecor	decorrelation	LOOP1T	BCA9	LOOP1TDEC	Model to decorrelate LOOP1T from BCA10.
loop2tdecor	decorrelation	LOOP1TDEC LOOP2T	BCA9	LOOP2TDEC	Model to decorrelate LOOP2T from BCA9 and LOOP1T.
bac10decor	decorrelation	LOOP1TDEC LOOP2TDEC BCA10	BCA9	BCA10DEC	Model to decorrelate BCA10 from BCA9, LOOP1T, and LOOP2T.
eden1res		BCA9 LOOP1TDEC LOOP2TDEC BCA10DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN1.
eden6-static	static	PCTUPLAND CELL_Y	WLSITE	WLSITE RES_WLSITE	Static model for station EDEN6 in BCNP—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for model for dynamic model.
eden6-dynamic	dynamic	PCTUPLAND CELL_Y BCA5 3ASWDEC L28GAPDEC	RES_WLSITE	RES_WLSITE	Dynamic model for station EDEN6 in BCNP—RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
l28gapdecor2	decorrelation	L28GAP 3ASWDEC	BCA5	L28GAPDEC	Model to decorrelate L28GAP from BCA5.
eden6res2	error correction	BCA5 3ASWDEC L28GAPDEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN6.
ENP Models					
eden3stack-static	static	CELL_Y PCTSLOUGH PCTPRAIRIE PCTSAWGRASS	WLSITE	WLSITE RES_WLSITE	Static model for station EDEN3 in ENP - Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
eden3stack-dynamic	dynamic	CELL_Y PCTSLOUGH PCTPRAIRIE PCTSAWGRASS P34 OTDEC TMCDEC P36DEC P35DEC	RES_WLSITE	RES_WLSITE	Dynamic model for station EDEN3 in ENP - RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
otdecor	decorrelation	OT	P34	OTDEC	Model to decorrelate OT from P34.
tmcdecor	decorrelation	TMC OTDEC	P34		Model to decorrelate TMC from P34 and OT.
p36decor	decorrelation	OTDEC P34 TMCDEC	P34	P36DEC	Model to decorrelate P36 from P34, OT, and TMC.
p35decor	decorrelation	OTDEC TMCDEC P34 P36DEC	P34	P35DEC	Model to decorrelate P35 from P34, OT, TMC, and P36.
eden3res	error correction	P34 OTDEC TMCDEC P36DEC P35DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site EDEN3.
met1-static	static	CELL_Y PCTSLOUGH PCTSAWGRASS PCTUPLAND	WLSITE	WLSITE RES_WLSITE	Static model for station MET1 in ENP—Water-level prediction (WLSITE) used for initial water-level estimate and residual (RES_WLSITE) used for dynamic model.
met1-dynamic	dynamic	NE1 SRS1DEC NP201DEC NE2DEC NESRS3DEC CELL_Y PCTSLOUGH PCTSAWGRASS PCTUPLAND	RES_WLSITE	RES_WLSITE	Dynamic model for WCA—RES_WLSITE prediction from dynamic model and WLSITE prediction from static model used for initial water-level estimate.
srs1decor	decorrelation	SRS1	NE1	SRS1DEC	Model to decorrelate SRS1 from NE1.
np201decor	decorrelation	SRS1DEC NP201	NE1	NP201DEC	Model to decorrelate NP201 from NE1 and SRS1.
ne2decor	decorrelation	SRS1DEC NP201DEC NE2	NE1	NE2DEC	Model to decorrelate NE2 from NE1, SRS1, and NP201.

Appendix 1. Summary of artificial neural network models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park]

Model	Model type	Input variables[a]	Model prediction variable	Variable used from the model	Comment
nesrs3decor	decorrelation	SRS1DEC NP201DEC NESRS3 NE2DEC	NE1	NESRS3	Model to decorrelate NESRS3 from NE1, SRS1, NP201, and NE2.
met1res	error correction	NE1 SRS1DEC NP201DEC NE2DEC NESRS3DEC	WLESTERR	WLESTERR	Model to estimate error between initial water-level estimate and measured data for site Met1.

[a] Descriptions of variables are provided in Appendix 3.

Appendix 2. Summary statistics for the models used in the study.

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Training: Range of output variable: Min, Water level, ft	Max, Water level, ft	n	R^2	ME, Water level, ft	SSE, Water level, ft	MSE, Water level, ft	RMSE, Water level, ft	PME
WCA1 Models											
wca1-static	static	2	14.70	14.94	4	0.948	0.000	0.03	0.007	0.12	52.1%
site8tdecor2	decorrelation	1	13.86	16.58	11	0.939	-0.279	0.43	0.039	0.22	8.1%
wca1medecor2	decorrelation	1	13.53	16.31	60	0.983	0.007	0.56	0.009	0.10	3.5%
wca1-dynamic	dynamic	3	-2.27	1.84	2,108	0.984	0.058	16.64	0.008	0.09	2.2%
northres	error correction	2	-1.12	0.43	252	0.861	-0.026	4.24	0.017	0.13	8.4%
southres	error correction	2	-2.04	-0.62	234	0.877	0.004	2.82	0.012	0.11	7.8%
WCA2 Models											
wca2a-static	static	1	9.77	11.37	5	0.987	0.000	0.02	0.005	0.09	5.4%
wca2f1decor	decorrelation	1	10.17	14.37	11	0.998	0.003	0.03	0.003	0.06	1.4%
wca2a-dynamic2	dynamic	3	-1.61	3.35	317	0.961	0.007	17.41	0.055	0.24	4.7%
eden11res2	error correction	1	-0.73	1.10	40	0.863	-0.008	1.55	0.039	0.20	11.0%
eden13	difference model	1	-0.54	1.51	35	0.987	0.000	0.17	0.005	0.07	3.6%
WCA3A Models											
cluster1static	static	2	8.26	10.28	5	1.000	-0.000	0.000	0.000	0.004	0.2%
3as3w1dec	decorrelation	1	8.51	11.74	11	0.991	0.005	0.11	0.010	0.109	3.4%
65dec	decorrelation	1	8.30	11.95	90	0.995	-0.005	0.41	0.005	0.068	1.9%
63decr1	decorrelation	1	8.29	11.95	358	0.941	-0.014	14.50	0.041	0.202	5.5%
3aswdec	decorrelation	1	8.73	11.95	962	0.951	-0.047	28.67	0.030	0.173	5.4%
cluster1dynamic	dynamic	3	-2.39	2.11	1,622	0.969	0.012	37.84	0.023	0.153	3.4%
3a-5res	error correction	2	-0.88	0.26	188	0.978	0.001	0.38	0.002	0.045	3.9%
eden12res2	error correction	3	-3.21	-2.43	188	0.953	-0.001	0.39	0.002	0.046	5.9%
eden4res	error correction	2	-3.30	-2.35	261	0.938	-0.006	1.27	0.005	0.070	7.3%
eden8res	error correction	3	-2.65	-2.27	266	0.903	-0.001	0.19	0.001	0.027	6.9%
eden9res2	error correction	3	1.59	2.43	132	0.991	-0.001	0.10	0.001	0.028	3.4%

Appendix 2. Summary statistics for the models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Training: Range of output variable: Min, Water level, ft	Training: Range of output variable: Max, Water level, ft	Training: n	Training: R^2	Training: ME, Water level, ft	Training: SSE, Water level, ft	Training: MSE, Water level, ft	Training: RMSE, Water level, ft	Training: PME
w11res2	error correction	3	-0.23	0.30	383	0.803	-0.011	1.03	0.003	0.052	9.8%
w14res	error correction	3	-2.16	-1.46	383	0.788	-0.014	1.36	0.004	0.060	8.6%
w15res	error correction	3	-2.49	-1.80	275	0.967	-0.003	0.32	0.001	0.034	4.9%
w18res	error correction	3	0.64	1.85	239	0.972	-0.005	0.62	0.003	0.051	4.2%
w2res	error correction	2	0.76	2.06	128	0.976	-0.002	0.40	0.003	0.056	4.3%
w5res	error correction	3	-2.85	-2.15	259	0.927	-0.001	0.38	0.001	0.038	5.5%
cluster3_static	static	1	8.91	9.41	3	1.000	0.000	0.00	0.000	0.000	0.0%
cluster3_dynamic	dynamic	2	-2.44	2.64	624	0.980	0.093	14.69	0.024	0.154	3.0%
3a9decor	decorrelation	1	7.27	11.52	11	0.996	0.002	0.07	0.006	0.088	2.1%
3a12decor	decorrelation	1	6.95	11.77	78	0.983	0.042	2.34	0.030	0.175	3.6%
eden14res	error correction	2	-1.17	-0.67	71	0.761	-0.000	0.26	0.004	0.061	12.0%
eden5res	error correction	3	-0.87	0.05	61	0.981	-0.001	0.06	0.001	0.032	3.5%
WCA3B Models											
cluster1-static	static	1	7.05	7.86	4	0.994	-0.000	0.002	0.000	0.032	3.9%
cluster1-dynamic	dynamic	3	-2.29	2.16	3,152	0.982	0.014	24.74	0.008	0.089	2.0%
site71dec	decorrelation	1	5.75	8.63	11	0.995	0.003	0.05	0.004	0.071	2.5%
srs1dec	decorrelation	1	5.57	8.74	70	0.979	0.018	0.98	0.014	0.120	3.8%
site76dec	decorrelation	1	5.56	8.74	263	0.972	-0.013	3.22	0.012	0.111	3.5%
ti-8res	error correction	1	-4.56	-3.23	227	0.983	-0.001	0.44	0.002	0.044	3.4%
ti-9res	error correction	3	-2.33	-1.59	227	0.961	0.004	0.20	0.001	0.030	4.0%
eden7res	error correction	2	-3.51	-2.08	137	0.993	-0.001	0.13	0.001	0.032	2.2%
eden10res	error correction	2	-2.07	-1.09	137	0.968	0.001	0.18	0.001	0.036	3.7%
BCNP Models											
bcnp-static	static	1	2.52	8.60	5	0.997	-0.005	0.05	0.010	0.131	2.2%
bcnp-dynamic	dynamic	4	-3.25	1.41	2,125	0.911	0.116	143.35	0.067	0.260	5.6%
loop1tdecor	decorrelation	1	3.06	7.15	11	0.993	0.019	0.15	0.014	0.130	3.2%

Appendix 2. Summary statistics for the models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Training: Range of output variable: Min, Water level, ft	Max, Water level, ft	n	R^2	ME, Water level, ft	SSE, Water level, ft	MSE, Water level, ft	RMSE, Water level, ft	PME
loop2tdecor	decorrelation	1	3.06	6.85	69	0.882	-0.043	5.60	0.081	0.289	7.6%
bac10decor	decorrelation	1	3.03	6.87	278	0.895	0.152	23.20	0.083	0.290	7.5%
eden1res	error correction	3	0.55	1.92	181	0.973	0.009	0.36	0.002	0.045	3.3%
eden6-static	static	1	8.60	10.42	3	0.790	-0.018	0.42	0.139	0.646	35.5%
eden6-dynamic	dynamic	3	-2.42	2.08	594	0.944	0.007	37.87	0.064	0.253	5.6%
l28gapdecor2	decorrelation	1	6.70	10.28	68	0.908	0.019	5.75	0.085	0.295	8.3%
eden6res2	error correction	2	-1.36	-0.55	53	0.862	0.003	0.24	0.004	0.068	8.4%
ENP Models											
eden3stack-static	static	1	0.29	2.72	5	0.999	-0.002	0.01	0.002	0.050	2.1%
eden3stack-dynamic	dynamic	3	-3.40	2.19	7,630	0.937	0.072	299.37	0.039	0.198	3.5%
otdecor	decorrelation	1	-1.06	3.17	11	0.990	0.000	0.17	0.016	0.139	3.3%
tmcdecor	decorrelation	1	-1.33	3.04	86	0.953	-0.003	5.03	0.058	0.245	5.6%
p36decor	decorrelation	1	-1.48	3.30	311	0.880	-0.001	40.39	0.130	0.362	7.6%
p35decor	decorrelation	1	-1.48	3.30	957	0.876	-0.011	94.54	0.099	0.315	6.6%
eden3res	error correction	1	-0.98	-0.34	521	0.752	-0.047	2.27	0.004	0.066	10.3%
met1-static	static	1	5.30	7.52	5	0.996	-0.000	0.01	0.002	0.062	2.8%
met1-dynamic	dynamic	1	-2.68	2.37	5,420	0.961	0.008	110.27	0.020	0.143	2.8%
srs1decor	decorrelation	1	5.16	8.09	11	0.986	-0.003	0.13	0.012	0.119	4.1%
np201decor	decorrelation	1	4.78	8.50	80	0.914	0.016	5.01	0.063	0.253	6.8%
ne2decor	decorrelation	1	5.68	8.54	263	0.980	-0.036	1.95	0.007	0.087	3.0%
nesrs3decor	decorrelation	1	5.84	8.54	764	0.959	-0.026	8.28	0.011	0.104	3.9%
met1res	error correction	3	-0.68	-0.32	215	0.950	0.001	0.06	0.000	0.016	4.5%

Appendix 2. Summary statistics for the models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Testing: Range of output variable: Min, Water level, ft	Range of output variable: Max, Water level, ft	n	R^2	ME, Water level, ft	SSE, Water level, ft	MSE, Water level, ft	RMSE, Water level, ft	PME
WCA1 Models											
wca1-static	static	2	12.44	16.68	80,510	0.023	0.000	24997.05	0.000	0.56	13.1%
site8tdecor2	decorrelation	1	13.39	16.68	59,142	0.819	-0.301	2327.94	-0.129	0.20	6.0%
wca1medecor2	decorrelation	1	13.26	16.54	57,804	0.970	-0.005	458.57	-0.005	0.09	2.7%
wca1-dynamic	dynamic	3	-2.27	1.87	77,068	0.982	0.061	435.26	0.061	0.08	1.8%
northres	error correction	2	-1.15	0.45	41,721	0.813	-0.046	659.76	-0.046	0.13	7.9%
southres	error correction	2	-2.06	-0.60	43,409	0.850	0.016	323.44	0.016	0.09	5.9%
WCA2 Models											
wca2a-static	static	1	8.41	14.67	89,154	0.340	0.004	56750.07	0.004	0.80	12.7%
wca2f1decor	decorrelation	1	9.63	14.63	53,469	0.986	0.012	480.12	0.012	0.09	1.9%
wca2a-dynamic2	dynamic	3	-1.61	3.40	80,897	0.889	-0.017	5745.87	-0.017	0.27	5.3%
eden11res2	error correction	1	-0.87	1.17	2,637	0.828	-0.054	108.68	-0.054	0.20	9.9%
eden13	difference model	1	-0.59	1.57	1,968	0.988	0.038	12.51	0.038	0.08	3.7%
WCA3A Models											
cluster1static	static	2	6.55	13.39	97,891	0.446	-0.000	74110.43	0.757	0.870	12.7%
3as3w1dec	decorrelation	1	8.28	11.95	51,742	0.970	-0.021	1047.08	0.020	0.142	3.9%
65dec	decorrelation	1	8.28	11.95	51,742	0.993	-0.029	260.47	0.005	0.071	1.9%
63decr1	decorrelation	1	8.28	11.95	50,776	0.939	0.044	2098.85	0.041	0.203	5.5%
3aswdec	decorrelation	1	8.69	11.95	49,815	0.959	-0.015	1337.04	0.027	0.164	5.0%
cluster1dynamic	dynamic	3	-10.26	2.17	83,968	0.960	0.012	2312.33	0.028	0.166	1.3%
3a-5res	error correction	2	-0.89	0.27	2,486	0.979	0.009	4.81	0.002	0.044	3.8%
eden12res2	error correction	3	-3.23	-2.42	2,488	0.960	0.003	4.87	0.002	0.044	5.5%
eden4res	error correction	2	-3.30	-2.33	2,399	0.941	0.008	11.34	0.005	0.069	7.1%
eden8res	error correction	3	-2.65	-2.26	2,442	0.894	0.001	1.76	0.001	0.027	6.8%
eden9res2	error correction	3	1.59	2.43	925	0.989	-0.000	0.83	0.001	0.030	3.5%

Appendix 2. Summary statistics for the models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Testing								
			Range of output variable								
			Min, Water level, ft	Max, Water level, ft	n	R^2	ME, Water level, ft	SSE, Water level, ft	MSE, Water level, ft	RMSE, Water level, ft	PME
w11res2	error correction	3	-0.27	0.36	7,628	0.860	-0.012	14.15	0.002	0.043	6.9%
w14res	error correction	3	-2.16	-1.44	7,628	0.828	-0.013	21.58	0.003	0.053	7.4%
w15res	error correction	3	-2.51	-1.79	3,859	0.978	-0.001	3.11	0.001	0.028	3.9%
w18res	error correction	3	0.64	1.87	7,628	0.978	0.000	16.27	0.002	0.046	3.7%
w2res	error correction	2	0.73	2.11	7,648	0.974	0.008	22.42	0.003	0.054	3.9%
w5res	error correction	3	-2.87	-2.15	7,648	0.935	-0.001	9.45	0.001	0.035	4.9%
cluster3_static	static	1	6.74	12.01	58,701	0.072	-0.000	35385.10	0.603	0.776	14.7%
cluster3_dynamic	dynamic	2	-2.45	2.70	53,841	0.982	0.125	593.08	0.011	0.105	2.0%
3a9decor	decorrelation	1	6.95	11.84	58,016	0.953	-0.008	1654.52	0.029	0.169	3.5%
3a12decor	decorrelation	1	6.95	11.84	53,837	0.983	0.075	541.43	0.010	0.100	2.1%
eden14res	error correction	2	-1.17	-0.66	1,416	0.580	0.007	9.92	0.007	0.084	16.3%
eden5res	error correction	3	-0.93	0.06	1,654	0.975	0.004	1.82	0.001	0.033	3.3%
WCA3B Models											
cluster1-static	static	1	5.21	9.80	81,262	0.218	-0.000	25801.41	0.318	0.563	12.3%
cluster1-dynamic	dynamic	3	-2.33	2.22	79,576	0.983	0.025	437.13	0.005	0.074	1.6%
site71dec	decorrelation	1	5.56	8.75	60,366	0.958	-0.023	662.95	0.011	0.105	3.3%
srs1dec	decorrelation	1	-0.36	0.50	59,987	0.944	-0.000	878.27	0.015	0.121	14.0%
site76dec	decorrelation	1	5.56	8.75	59,678	0.938	0.014	957.11	0.016	0.127	4.0%
ti-8res	error correction	1	-4.56	-3.20	5,684	0.983	0.003	8.26	0.001	0.038	2.8%
ti-9res	error correction	3	-2.34	-1.58	5,685	0.962	0.009	4.65	0.001	0.029	3.8%
eden7res	error correction	2	-3.51	-2.07	2,419	0.991	0.008	2.56	0.001	0.033	2.3%
eden10res	error correction	2	-2.08	-1.09	2,419	0.969	0.007	3.62	0.001	0.039	3.9%
BCNP Models											
bcnp-static	static	1	-1.43	10.51	115,719	0.837	-0.032	73866.48	0.638	0.799	6.7%
bcnp-dynamic	dynamic	4	-3.25	1.58	99,626	0.844	0.091	8532.91	0.086	0.293	6.1%
loop1tdecor	decorrelation	1	3.01	7.33	55,719	0.993	0.019	0.15	0.000	0.002	0.0%

Appendix 2. Summary statistics for the models used in the study. — Continued

[WCA1, Water Conservation Area 1; WCA2, Water Conservation Area 2; WCA3A, Water Conservation Area 3A; WCA3B, Water Conservation Area 3B; BCNP, Big Cypress National Preserve; ENP, Everglades National Park; HLN, hidden layer neurons; min, minimum; max, maximum; feet, ft; n, number of vectors; R^2, coefficient of determination; ME, mean error; SSE, sum of square error; MSE, mean square error; RMSE, root mean square error; PME, percent model error]

Model name	Model type	Number of HLN	Testing: Range of output variable: Min, Water level, ft	Testing: Range of output variable: Max, Water level, ft	Testing: n	Testing: R^2	Testing: ME, Water level, ft	Testing: SSE, Water level, ft	Testing: MSE, Water level, ft	Testing: RMSE, Water level, ft	Testing: PME
loop2tdecor	decorrelation	1	3.02	6.92	51,354	0.886	0.038	3145.49	0.061	0.247	6.3%
bac10decor	decorrelation	1	3.02	6.92	50,807	0.928	0.272	1964.69	0.039	0.197	5.0%
eden1res	error correction	3	0.50	1.92	3,864	0.830	0.018	9.70	0.003	0.050	3.5%
eden6-static	static	1	6.21	12.17	57,244	0.366	-0.035	49317.07	0.862	0.928	15.6%
eden6-dynamic	dynamic	3	-2.42	2.11	49,053	0.955	-0.026	1880.37	0.038	0.196	4.3%
l28gapdecor2	decorrelation	1	6.48	10.52	49,047	0.831	0.086	5630.72	0.115	0.339	8.4%
eden6res2	error correction	2	-1.36	-0.53	1,591	0.706	0.030	9.59	0.006	0.078	9.4%
ENP Models											
eden3stack-static	static	1	-1.49	3.59	93,051	0.673	-0.005	44887.56	0.482	0.695	13.7%
eden3stack-dynamic	dynamic	3	-3.41	2.20	93,270	0.941	0.058	2943.13	0.032	0.178	3.2%
otdecor	decorrelation	1	-1.49	3.32	56,161	0.919	-0.015	2870.91	0.051	0.226	4.7%
tmcdecor	decorrelation	1	-1.49	3.32	56,087	0.924	0.061	2689.96	0.048	0.219	4.6%
p36decor	decorrelation	1	-1.49	3.32	56,080	0.892	0.007	3826.45	0.068	0.261	5.4%
p35decor	decorrelation	1	-1.49	3.32	55,804	0.783	0.154	7665.94	0.137	0.371	7.7%
eden3res	error correction	1	-1.02	-0.34	5,294	0.784	-0.042	21.80	0.004	0.064	9.5%
met1-static	static	1	2.90	9.71	94,949	0.566	0.001	42924.03	0.452	0.672	9.9%
met1-dynamic	dynamic	1	-2.73	2.37	80,735	0.963	0.011	1182.90	0.015	0.121	2.4%
srs1decor	decorrelation	1	4.74	8.62	60,937	0.903	0.011	1462.12	0.024	0.155	4.0%
np201decor	decorrelation	1	4.74	8.54	58,821	0.832	0.048	2294.77	0.039	0.198	5.2%
ne2decor	decorrelation	1	5.60	8.54	55,743	0.952	-0.035	479.66	0.009	0.093	3.2%
nesrs3decor	decorrelation	1	5.83	8.54	48,255	0.965	-0.020	293.35	0.006	0.078	2.9%
met1res	error correction	3	-0.68	-0.32	1,238	0.966	0.000	0.24	0.000	0.014	3.8%

Appendix 3. Variables used in the artificial neural network models.

[NAVD88, North American Vertical Datum of 1988]

Variables	Description
3A12	water level at site 3A12
3A12DEC	decorrelated water level at site 3A12
3A9	water level at site 3A9
3A9DEC	decorrelated water level at site 3AS3W1
3AS3W1DEC	decorrelated water level at site 3A10
3ASW	water level at site 3ASW
3ASWDEC	decorrelated water level at site 3ASW
BCA10	water level at site BCA10
BCA10DEC	decorrelated water level at site BCA11
BCA5	water level at site BCA5
BCA9	water level at site BCA9
CELL_X	UTM Easting of cell center
CELL_Y	UTM Northing of cell center
L28GAP	water level at site L28GAP
L28GAPDEC	decorrelated water level at site L28GAP
LOOP1T	water level at site LOOP1T
LOOP1TDEC	decorrelated water level at site LOOP1T
LOOP2T	water level at site LOOP2T
NE1	water level at site NE1
NE2	water level at site NE2
NE2DEC	decorrelated water level at site NE2
NESRS3	water level at site NESRS3
NESRS3DEC	decorrelated water level at site NESRS3
NP201	water level at site NP201
NP201DEC	decorrelated water level at site NP201
OT	water level at site OT
OTDEC	decorrelated water level at site OT
P34	water level at site P34
P35	water level at site P35
P35DEC	decorrelated water level at site P35
P36	water level at site P36
P36DEC	decorrelated water level at site P36
PCTEXOTICS	percent exotics
PCTOTHER	percent other
PCTPRAIRIE	percent prairie
PCTSAWGRASS	percent sawgrass
PCTSLOUGH	percent slough
PCTUPLAND	percent upland
S142H	water level at site S142H
SITE_62	water level at Site 62
SITE_63	water level at Site 63
SITE_63DEC	decorrelated water level at Site 63
SITE_64	water level at Site 64
SITE_65	water level at Site 65
SITE_65DEC	decorrelated water level at Site 65
SITE_69C	water level at Site 69 adjusted to NAVD88
SITE_7	water level at Site 7
SITE_71C	water level at Site 71 adjusted to NAVD88
SITE_76	water level at Site 76
SITE_8T	water level at Site 8T
SITE_9	water level at Site 9
SITE_99	water level at Site 99
SITE71DEC	decorrelated water level at Site 71
SITE76DEC	decorrelated water level at Site 76
SITE8TDEC	decorrelated water level at Site 8T
SITE8TDEC2	decorrelated water level at Site 8T-second version
SITE9DEC	decorrelated water level at Site 9
SRS1	water level at site SRS1
SRS1C	water level at site SRS1adjusted to NAVD88
SRS1DEC	decorrelated water level at site SRS1
TMC	water level at site TMC
TMCDEC	decorrelated water level at site TMC
WCA1ME	water level at site WCA1ME
WCA1MEDEC	decorrelated water level at site WCA1ME
WCA1MEDEC2	decorrelated water level at site WCA1ME-second version
WCA2E1	water level at site WCA2E1
WCA2F1	water level at site WCA2F1
WCA2F1DEC	decorrelated water level at site WCA2F1

Prepared by:
USGS Enterprise Publishing Network
Raleigh Publishing Service Center
3916 Sunset Ridge Road
Raleigh, NC 27607

For additional information regarding this publication, contact:
Director
USGS South Carolina Water Science Center
720 Gracern Road, Suite 129
Columbia, SC 29210
phone: 1-803-750-6100
email: dc_sc@usgs.gov

Or visit the South Carolina Water Science Center website at:
http://sc.water.usgs.gov

Conrads and Roehl—**Hydrologic Record Extension of Water-Level Data in the EDEN Using Artificial Neural Network Models, 2000–2006—Open-File Report 2007–1350**