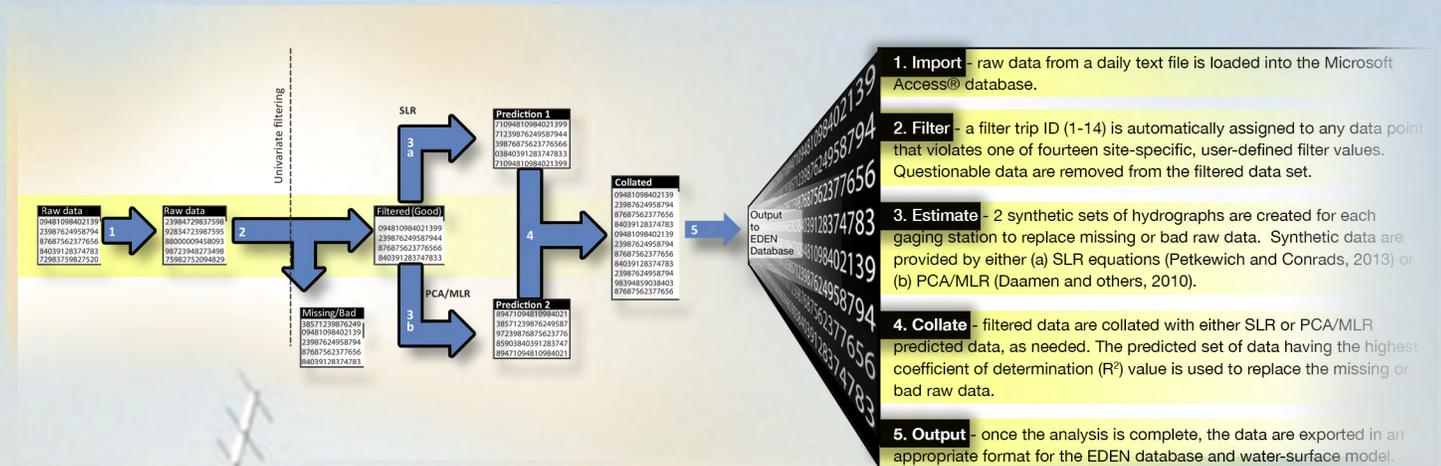


Greater Everglades Priority Ecosystems Science

Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data



Scientific Investigations Report 2016–5094



Cover. Gaging station W11 in Water Conservation Area 3A of the Florida Everglades, Broward County, Florida. Photograph by Christa Zweig, South Florida Water Management District, West Palm Beach, Florida.

Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

By Matthew D. Petkewich, Ruby C. Daamen, Edwin A. Roehl, and Paul A. Conrads

Greater Everglades Priority Ecosystems Science

Scientific Investigations Report 2016–5094

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Petkewich, M.D., Daamen, R.C., Roehl, E.A., and Conrads, P.A., 2016, Using inferential sensors for quality control of Everglades Depth Estimation Network water-level data: U.S. Geological Survey Scientific Investigations Report 2016–5094, 25 p., <http://dx.doi.org/10.3133/sir20165094>.

ISSN 2328-0328 (online)

Contents

Abstract.....	1
Introduction.....	1
Data Collection Network.....	4
Automated Data Assurance and Management (ADAM) Software.....	4
Univariate Filters.....	6
Linear Regression Models.....	8
Simple Linear Regression Models.....	8
Multivariate Linear Regression Models.....	10
ADAM Software Architecture.....	13
Using ADAM Software to Process Everglades Depth Estimation	
Network Data.....	16
Automated Processing.....	16
Manual Processing.....	16
Summary.....	24
References Cited.....	24

Figures

1. Map showing location of the water-level stations within the Everglades Depth Estimation Network.....	2
2. Maps showing the April 30, 2011, Everglades Depth Estimation Network water surface and the April 30, 2011, EDEN water surface with an erroneous data value.....	3
3. Schematic diagram showing flow of Everglades Depth Estimation Network data from agency databases to the Web.....	5
4. Schematic diagram showing Automated Data Assurance and Management Process.....	6
5. Graph showing measured and estimated water level for station S146_H for the period October 1, 2010, through December 31, 2010.....	8
6. Graph showing exceedance percentage for coefficient of determination for first, second, third predictor, and all 667 predictor water-level estimation equations.....	9
7. Graph showing measured and predicted water levels using simple linear regression and multivariate linear regression for station NP205 for October 1, 2011, to October 1, 2012.....	11
8. Graph showing coefficient of determination for simple linear regression and multivariate linear regression models for station NP205 for October 1, 2011, to October 1, 2012.....	12
9. Schematic diagram showing the general architectural elements for the Automated Data Assurance and Management software.....	14
10. Screenshot of Automated Data Assurance and Management tool in “Review” mode showing part of Control worksheet.....	17
11. Screenshot of Automated Data Assurance and Management tool in “Review” mode showing Review_Chart and Review_Tab worksheets.....	18

12. Screenshot of Automated Data Assurance and Management tool in “Review” mode showing the chart controls and graph of the Review_Chart worksheet20
13. Screenshots of Automated Data Assurance and Management tool in “Review” mode showing parts of the Review_Tab and Review_Chart worksheets and, example of preliminary results, prediction detail information, manual review procedure using the “Review Selection” dropdown menu, manual review procedure using “Data Correction” dropdown menu, manual review procedure using “Manual Entry” column and “Review Selection” dropdown menu and graphical results of a manual review.....21

Tables

1. Automated Data Assurance and Management filter descriptions7
2. Minimum, median, and maximum values for the summary statistics for 667 water-level estimation equations9
3. Measured and predicted water levels and statistics for station NP205 for Oct. 1, 2011, to Oct. 1, 201211
4. Range of dates and coefficient of determination for simple linear regression and multivariate linear regression models for station NP205 for Oct. 1, 2011, to Oct. 1, 2012.....12
5. Input, workbook, databases, and output files used in the ADAM software program.....13
6. Worksheets in the ADAM.xlsm workbook for automated review of the data15
7. Worksheets in the ADAM_Review.xlsm workbook for manual review of the data.....15

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
	Length	
foot (ft)	0.3048	meter (m)

Datum

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.

Abbreviations

ADAM	Automated Data Assurance and Management
BCNP	Big Cypress National Preserve
CERP	Comprehensive Everglades Restoration Plan
EDEN	Everglades Depth Estimation Network
eFTP	Enterprise file transfer protocol
ENP	Everglades National Park
EVE	Explore and View EDEN
FTP	file transfer protocol
GIS	geographic information system
GUI	graphical user interface
ME	mean error
MLR	multivariate linear regression
NPS	National Park Service
NWIS	National Water Information System
PC	principal component (eigenvector)
PCA	principal component analysis
PME	percent model error
R^2	coefficient of determination
RMSE	root mean square error
SFWMD	South Florida Water Management District
SLR	simple linear regression
SOFIA	South Florida Information Access
SPC	statistical process control
USGS	U.S. Geological Survey
VBA	visual basic for applications
WCA	water conservation area

Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

By Matthew D. Petkewich,¹ Ruby C. Daamen,² Edwin A. Roehl,² and Paul A. Conrads¹

Abstract

The Everglades Depth Estimation Network (EDEN), with over 240 real-time gaging stations, provides hydrologic data for freshwater and tidal areas of the Everglades. These data are used to generate daily water-level and water-depth maps of the Everglades that are used to assess biotic responses to hydrologic change resulting from the U.S. Army Corps of Engineers Comprehensive Everglades Restoration Plan. The generation of EDEN daily water-level and water-depth maps is dependent on high quality real-time data from water-level stations. Real-time data are automatically checked for outliers by assigning minimum and maximum thresholds for each station. Small errors in the real-time data, such as gradual drift of malfunctioning pressure transducers, are more difficult to immediately identify with visual inspection of time-series plots and may only be identified during on-site inspections of the stations. Correcting these small errors in the data often is time consuming and water-level data may not be finalized for several months. To provide daily water-level and water-depth maps on a near real-time basis, EDEN needed an automated process to identify errors in water-level data and to provide estimates for missing or erroneous water-level data.

The Automated Data Assurance and Management (ADAM) software uses inferential sensor technology often used in industrial applications. Rather than installing a redundant sensor to measure a process, such as an additional water-level station, inferential sensors, or virtual sensors, were developed for each station that make accurate estimates of the process measured by the hard sensor (water-level gaging station). The inferential sensors in the ADAM software are empirical models that use inputs from one or more proximal stations. The advantage of ADAM is that it provides a redundant signal to the sensor in the field without the environmental threats associated with field conditions at stations (floods or hurricane, for example). In the event that a station does malfunction, ADAM provides an accurate estimate for the period of missing data. The ADAM software also is used

in the quality assurance and quality control of the data. The virtual signals are compared to the real-time data, and if the difference between the two signals exceeds a certain tolerance, corrective action to the data and (or) the gaging station can be taken. The ADAM software is automated so that, each morning, the real-time EDEN data are compared to the inferential sensor signals and digital reports highlighting potential erroneous real-time data are generated for appropriate support personnel. The development and application of inferential sensors is easily transferable to other real-time hydrologic monitoring networks.

Introduction

The Everglades Depth Estimation Network (EDEN) project was initiated to provide scientists working on Everglades restoration with spatially continuous, quality-assured and quality-controlled hydrologic data at any location within the freshwater part of the Everglades. EDEN consists of an integrated network of real-time water-level gaging stations (fig. 1), a ground-elevation model, and a water-surface elevation model designed to provide scientists, engineers, and water-resource managers with current (1991–2016) water-level and water-depth information for the entire freshwater portion of the Everglades (Telis, 2005, 2006; Telis and others, 2015). The EDEN domain is presented on a grid consisting of more than fifty-thousand 400-square-meter (4,300-square-foot) cells and offers a consistent and documented dataset that can be used by scientists and water-resource managers to (1) guide large-scale field operations, (2) integrate hydrologic and ecological data and their analysis, 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). In addition, EDEN, with its integration of real-time data and models, provides opportunities for real-time evaluation of water-level conditions and water-resource management operation. The EDEN database is a 26-year dataset of baseline conditions (1991–2016) prior to the full implementation of the CERP and offers investigators a single repository for historical daily and hourly water-level data.

¹U.S. Geological Survey.

²Advanced Data Mining International.

2 Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

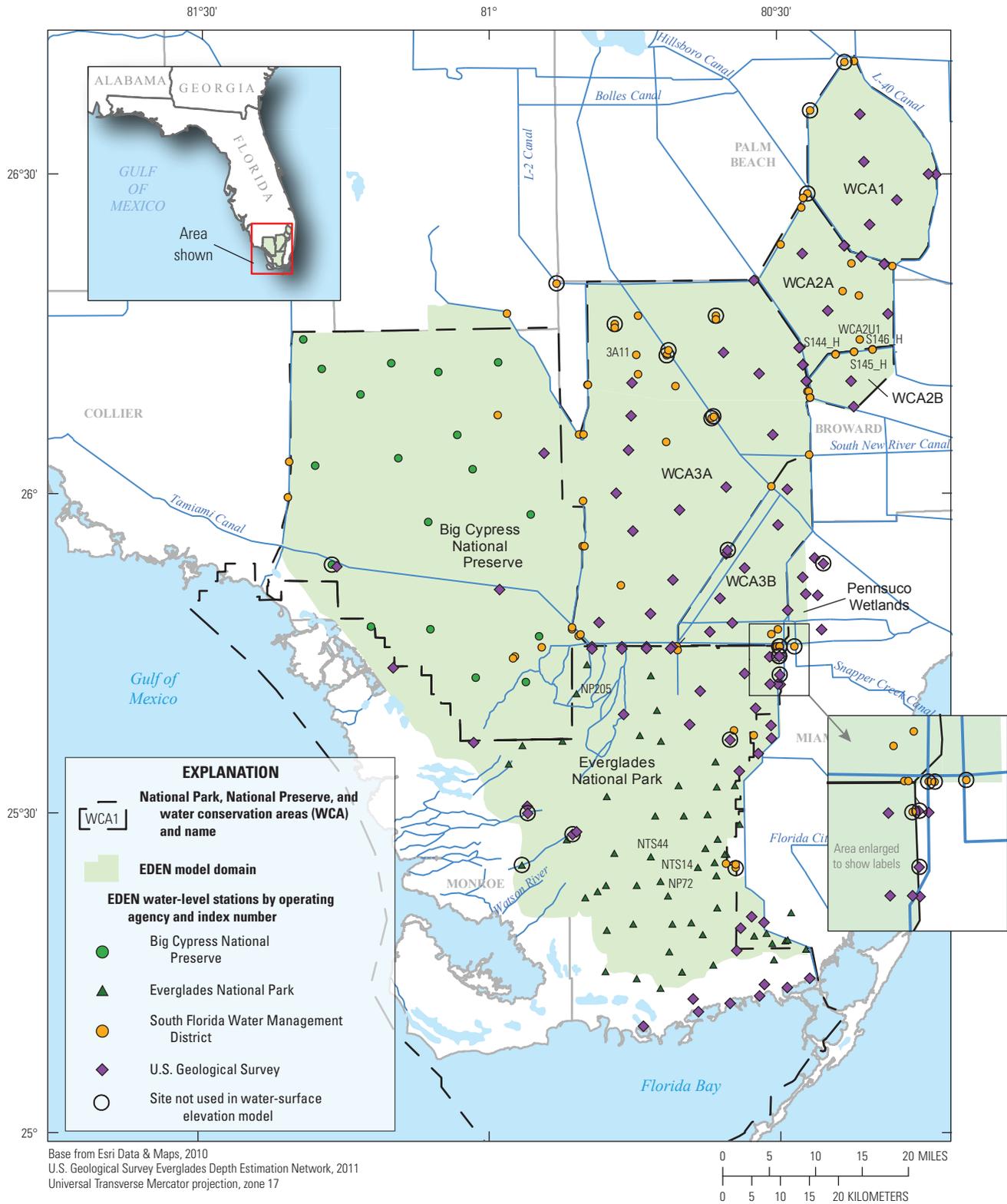


Figure 1. Location of the water-level stations within the Everglades Depth Estimation Network (EDEN).

A water-surface elevation model for the freshwater portion of the EDEN domain was developed using a geographic information system (GIS) and the EDEN grid just described for the ground-elevation model (Pearlstine and others, 2007; Palaseanu and Pearlstine, 2008; Telis and others, 2015). The model produces an areally continuous water surface within the EDEN domain for any day within the period of record for the EDEN database (1991–2016). Data to support the water-surface model include continuous water levels collected at 247 stations. Currently (2016), the EDEN water-surface model interpolates measured daily water levels from 223 of these 247 stations in the EDEN continuous monitoring network to the EDEN grid (Telis and others, 2015). The remaining 24 stations are available for quality-assurance evaluations and estimation of missing or erroneous data. The water-surface map for April 30, 2011, is shown in figure 2A.

The generation of daily water-level surfaces is dependent on high quality, real-time data because missing or erroneous data can compromise the quality of the modeled water-surface

maps. Figure 2B shows the April 30, 2011, water-surface map with an example of erroneous data. In a number of instances, data for the 223-station network may be missing because of instrumentation failure or data telemetry transmission problems. In addition, nine stations currently do not have telemetry for real-time transmission of data. When data from a particular station are missing, the water-surface model does not use that station for generating the water-surface map for that day. The quality of the water-surface maps therefore can be diminished, depending on the number of stations with missing data and the locations of those stations.

Real-time data can easily be checked for outliers using minimum and maximum thresholds for each station. Less obvious errors in the real-time data, such as a gradual drift caused by malfunctioning pressure transducers, are more difficult to immediately identify by visual inspection of time-series plots. These types of errors are typically identified during station visits by comparing manual field measurements to data logger results. Correcting these types of errors often

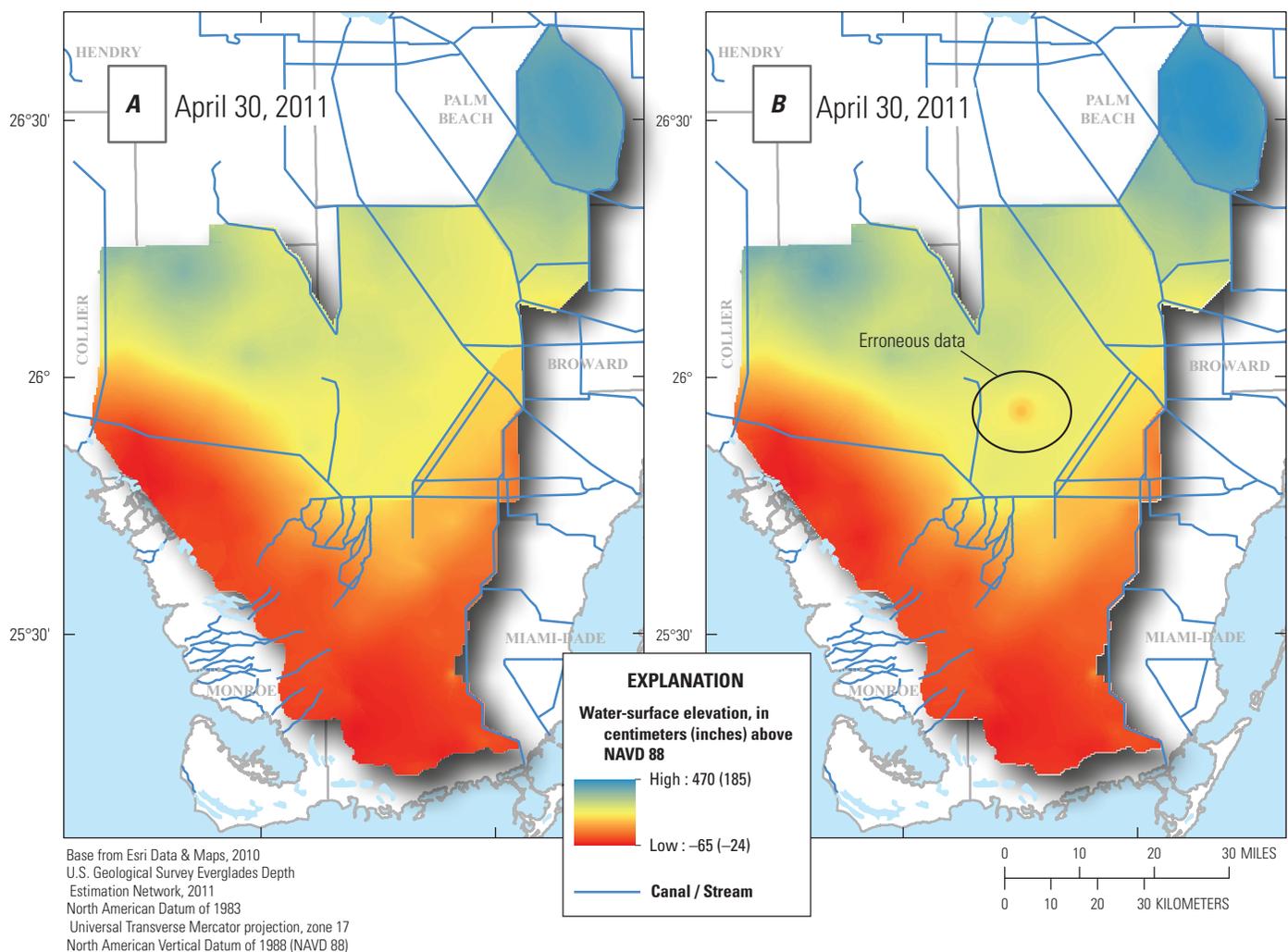


Figure 2. A, The April 30, 2011, Everglades Depth Estimation Network (EDEN) water surface and, B, the April 30, 2011, EDEN water surface with an erroneous data value.

4 Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

is time consuming, and in such cases, water-level data may not be finalized for several months. In order to provide daily modeled water surfaces on a near real-time basis, EDEN needed an automated process to identify errors in water-level data and to provide estimates for the missing or erroneous data. To increase the accuracy of the daily water-surface maps, the Automated Data Assurance and Management (ADAM) tool was created by the U.S. Geological Survey as part of Greater Everglades Priority Ecosystems Science to facilitate an accurate quality-assurance review of the real-time data from the EDEN network and allow estimation or replacement of missing or erroneous data. Incorporation of ADAM in the review process of all EDEN datasets has improved the consistency and utility of the EDEN data. The development and application of inferential sensors is easily transferable to other real-time hydrologic monitoring networks.

The purpose of this report is to describe how inferential sensors are used to quality assure the water-level data used in the generation of near-real time EDEN water-surface maps. This report documents the development of the ADAM software for EDEN and includes examples of the methods used to identify errors in real-time water-level data and provide estimates for those data. The user's manual for the installation and operation of the ADAM software is presented in Petkewich and others (2016).

Data Collection Network

The 247 stations that compose EDEN provide hydrologic data for freshwater and tidal areas of the Everglades (fig. 1). Water-level data from these surface-water and shallow-groundwater stations are used to create the EDEN water-surface maps. The stations are managed by the South Florida Water Management District (SFWMD), National Park Service (NPS), and U.S. Geological Survey (USGS). The NPS includes personnel in Big Cypress National Preserve (BCNP) and Everglades National Park (ENP). Data for the stations are managed by each agency and stored in the individual agency databases: DBHYDRO (http://my.sfwmd.gov/dbhydroplsqli/show_dbkey_info.main_menu); DataForEver; and National Water Information System (NWIS, <http://waterdata.usgs.gov/fl/nwis/>) for SFWMD, NPS, and USGS, respectively. One set of exceptions is stations BCA1 through BCA20, which are NPS stations managed by BCNP, but stored in DBHYDRO.

In this report, the names of the EDEN stations follow the naming convention used by EDEN (<http://sofia.usgs.gov/eden/explanation.php#stationname>) and are generally similar

to the names used by the respective agencies that maintain the stations. Station datasets, along with estimates for missing periods, are available from the EDEN Web page on the South Florida Information Access (SOFIA) Web site, <http://sofia.usgs.gov/eden/index.php>. Station information, including location, vegetation type, land-surface elevation, real-time and historical water-level elevation, and other site-specific data also can be accessed there.

The general flow of the data from the agency databases to the EDEN Web server is shown in figure 3. Data for the real-time stations are transmitted to the agency databases by satellite or radio telemetry. Every morning, hourly data for the previous day are exported from NPS and SFWMD databases and transmitted by file transfer protocol (FTP) to the USGS Enterprise (e)FTP server. Automated scripts merge the three data files from the USGS, NPS, and SFWMD into a single data file that is transferred to a file server to be analyzed by the ADAM program. Once the data have been processed by ADAM, the data are transferred back to the USGS eFTP server to be retrieved and loaded into the EDEN database using scripts and the USGS eFTP. Automated scripts then extract water-level data from the EDEN database, process the data using the daily median program and EDEN surface-water model, and publish the data to the EDEN Web page as daily median water-surface maps. Telis and others (2015) describe the daily median program and EDEN V2 surface-water model. In addition, water-level data from the EDEN database can be interactively viewed on graphs or retrieved as data tables through the Explore and View EDEN (EVE; <http://sofia.usgs.gov/eden/eve/>) Web tool. All water-level data used in the generation of the EDEN water-surface maps are stored in the EDEN database to produce the water-surface maps and allow a common platform for viewing hydrographs and retrieving data for all of the EDEN stations.

Automated Data Assurance and Management (ADAM) Software

The ADAM software is a Microsoft Excel and Access database tool created for fast and accurate quality-assurance review of the real-time water-level data from the EDEN network for estimation or replacement of missing or erroneous data and for digital archiving of original and processed data. The ADAM software is conceptually based on inferential sensor technology, which is often used in industrial applications. Rather than installing and maintaining a sensor in a

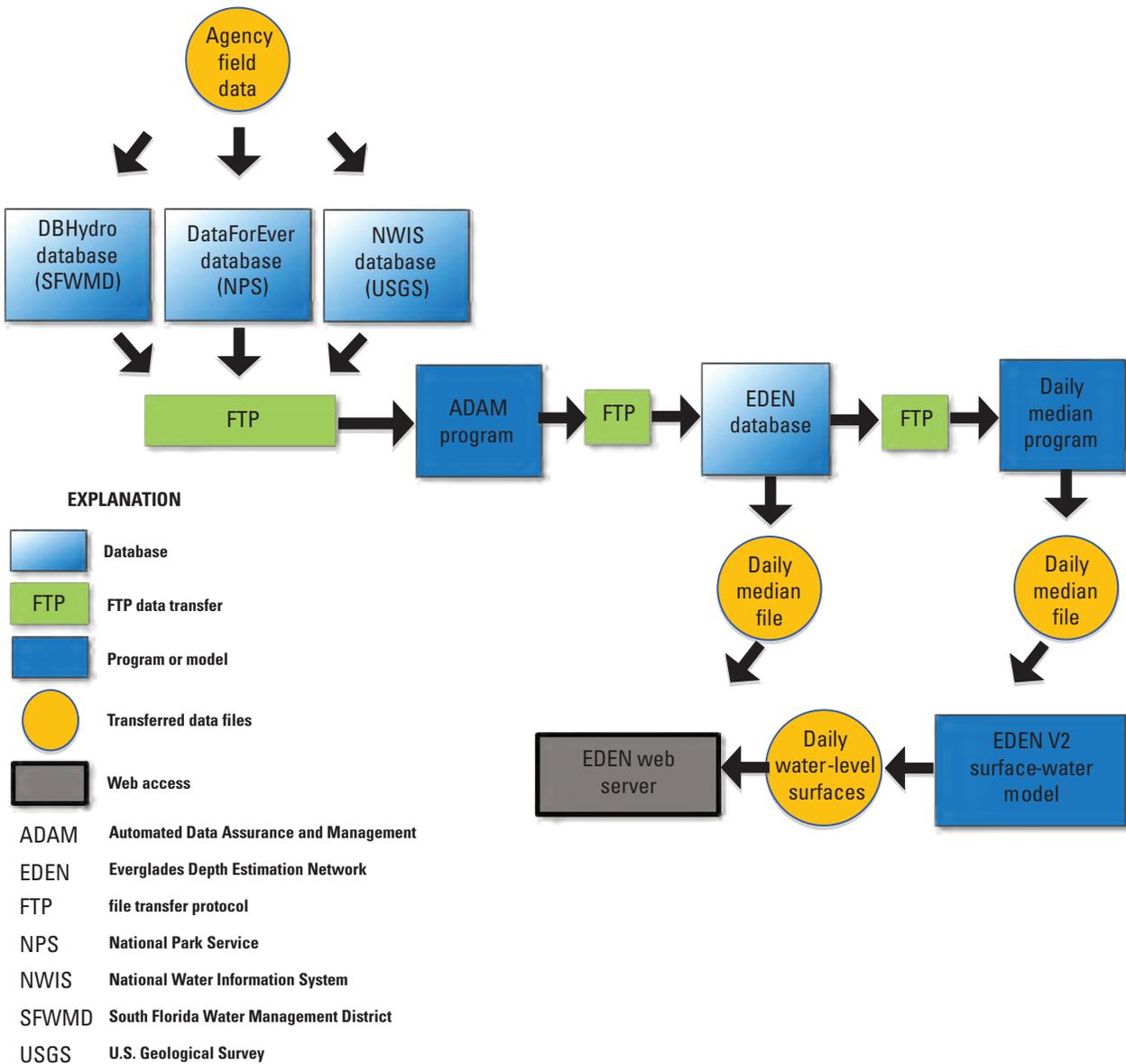


Figure 3. Schematic diagram showing flow of Everglades Depth Estimation Network data from agency databases to the Web.

harsh environment (for example, a high temperature exhaust port) to measure a process, an inferential sensor (or virtual sensor) is developed to make accurate estimates of the process measured by the hard sensor. The inferential sensor becomes a redundant sensor. The inferential sensors in the ADAM software are empirical models for each EDEN station that provide redundant signals to the water-level stations in the field without the risk of damage caused by the environmental setting. In the event that the sensor at a station malfunctions, ADAM provides accurate estimates for the period of missing or erroneous data.

The ADAM software uses a sequence of steps for quality-assuring data. The first step provides information about the

quality of the data. This is accomplished using a set of 14 univariate filters that identifies missing data and data that violate user-defined thresholds related to maximum, minimum, and rate-of-change values. The output from the univariate filters comprises a filtered dataset that is deemed of good quality to use for estimating missing data. The second ADAM step creates the inferential sensors (empirical models) for each station in the network using the subset of good filtered data as model input. The inferential sensors are used to replace missing or erroneous data and to quality-assure the measured data. The architecture of the ADAM software is shown in figure 4, and the major elements of the software are described in the following sections.

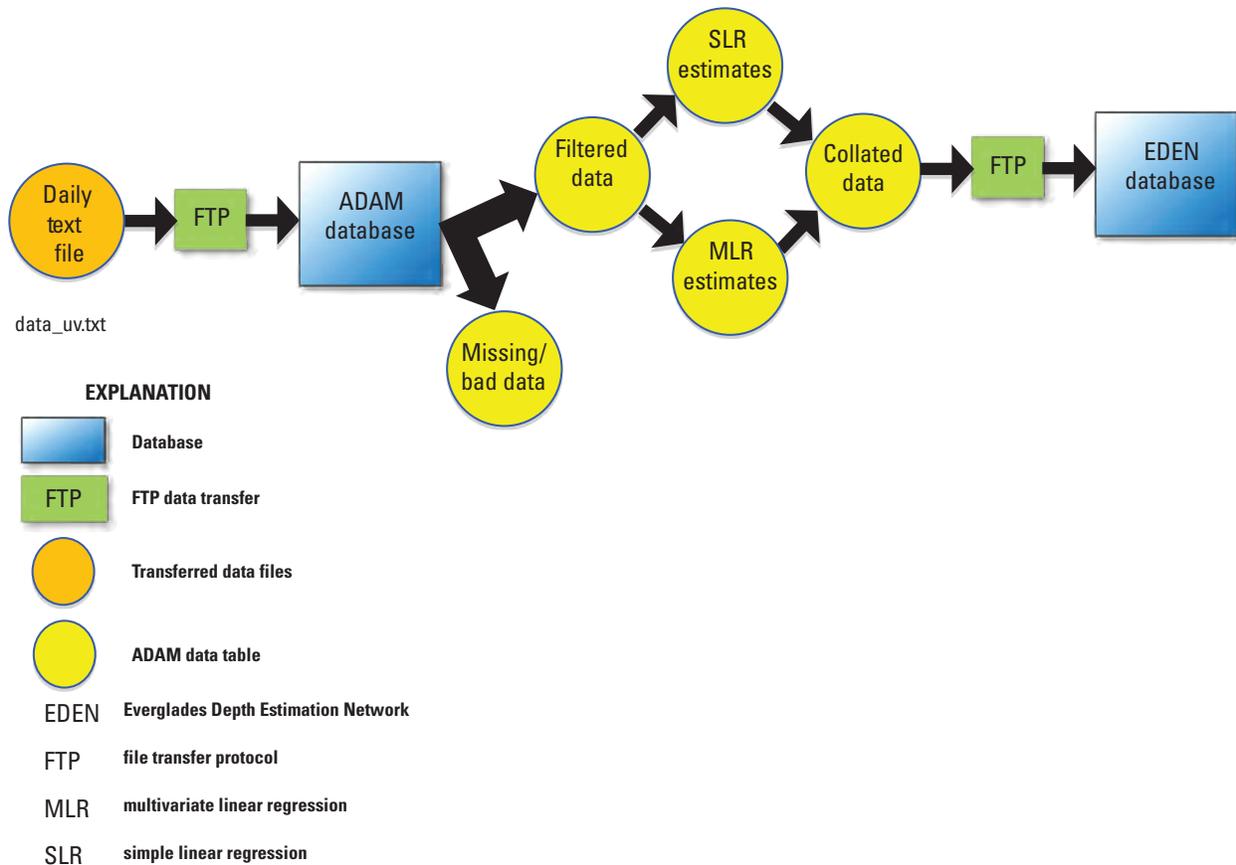


Figure 4. Schematic diagram showing Automated Data Assurance and Management (ADAM) Process.

Univariate Filters

The first ADAM step implements a Statistical Process Control (SPC) series of 14 univariate filters (Cook and others, 2008) to determine the subset of data considered to be of good quality (called “filtered data”; fig. 4). The univariate filters (1–14 in table 1) identify data that violate user-defined thresholds related to missing, maximum, minimum, and rate-of-change values. Thresholds for EDEN were established on the basis of a review of historical data for each EDEN station. Data flagged by filters 1–14 are removed from the filtered dataset so that they are not used for quality-assurance procedures employed by ADAM.

Two types of filters are used for initial evaluation: acceptable filters (1, 2, 3, 7, and 8; table 1) and warning filters (filters 4, 5, 6, 9, 10, 11, 12, 13, and 14; table 1). The following types of record violate acceptable filter thresholds, and thus are considered erroneous and not used in the preliminary dataset: missing data (filter 1), near-exceedance of maximum recorded water level (filter 2), minimum recorded water level (filter 3), very fast and very large increase (rate of change) in water level (filter 7), and very fast and very large decrease in water level (filter 8). Filters 2 and 3 were set to values representing 95 and 5 percent of the historical range of the recorded water level for an individual EDEN station, respectively. Filters 7 and 8, similarly, were set to 95 percent of the historical maximum hourly

rate-of-change increase or decrease at a station, respectively. For EDEN, these filters can vary greatly depending on the location of the station in the Everglades and its proximity to water-control structures. For example, stations located in the northern part of the Everglades, such as Water Conservation Area 1 (WCA1, fig. 1), have recorded water levels as high as 17 feet (ft) above the North American Vertical Datum of 1988 (NAVD 88), whereas coastal stations have recorded water levels as low as 3 ft below NAVD 88. Stations located in canals or near water-control structures are subject to fast and large changes in water levels caused by the opening and closing of water-control structures. The maximum rate of change at stations located in marsh environments away from water-control structures is much lower than that observed at stations near canals or structures.

The thresholds used for warning filters are less extreme variants of those used for acceptable filters 2, 3, 7, and 8. Although values exceeding the warning thresholds are not considered erroneous data, use of these values is limited in the quality-assurance procedure of ADAM. Values exceeding the warning thresholds are still removed from the filtered dataset so as not to affect the quality of the water-level estimation models and their predictions. Filters 4 and 5 are upper and lower control limit filters and represent the 90th and 10th percentiles of the historical range of data, respectively.

Table 1. Automated Data Assurance and Management (ADAM) filter descriptions.

[MLR, multivariate linear regression]

Filter	Description	Filter name	Filter type	Where in ADAM process
1	No data.	LOST_SIGNAL	Acceptable	Initial evaluation
2	Data value is greater than a value equal to 95 percent of the historical range of data.	GT_RNG_UL	Acceptable	Initial evaluation
3	Data value is less than a value equal to 5 percent of the historical range of data.	LT_RNG_LL	Acceptable	Initial evaluation
4	Data value is greater than upper control value (90 percent of historical range).	GT_UCL	Warning	Initial evaluation
5	Data value is less than lower control value (10 percent of historical range).	LT_LCL	Warning	Initial evaluation
6	Data has remained unchanged (flat) for a period greater than acceptable.	FLATLINE	Warning	Initial evaluation
7	Positive rate of change between two measurements exceeds a tolerable rate of change (95 percent of the maximum historical rise in water level).	VFVL_INC	Acceptable	Initial evaluation
8	Negative rate of change between two measurements exceeds a tolerable rate of change (95 percent of the maximum historical drop in water level).	VFVL_DEC	Acceptable	Initial evaluation
9	Positive rate of change over three measurements exceeds a tolerable rate of change (95 percent of the maximum historical rise in water level).	FVL_INC	Warning	Initial evaluation
10	Negative rate of change over three measurements exceeds a tolerable rate of change (95 percent of the maximum historical drop in water level).	FVL_DEC	Warning	Initial evaluation
11	Positive rate of change between two measurements exceeds a rate of change “warning” threshold (90 percent of the maximum historical rise in water level).	VFL_INC	Warning	Initial evaluation
12	Negative rate of change between two measurements exceeds a rate of change “warning” threshold (90 percent of the maximum historical drop in water level).	VFL_DEC	Warning	Initial evaluation
13	Positive rate of change over three measurements exceeds a rate of change “warning” threshold (90 percent of the maximum historical rise in water level).	FL_INC	Warning	Initial evaluation
14	Negative rate of change over three measurements exceeds a rate of change “warning” threshold (90 percent of the maximum historical drop in water level).	FL_DEC	Warning	Initial evaluation
15	Measured value is below a user-defined value that represents dry conditions at the station.	DRY_PROTOCOL	Warning	Signal evaluation
16	Difference between measured value and MLR value exceeds a user-defined value.	PRED_V_ACT	Acceptable	Signal evaluation

Filter 6 is a special case of a rate-of-change threshold that identifies any data values that have not changed over a specified amount of time. For example, if a sensor malfunctions and constantly records a single value, the ADAM software will warn the user that the data may need to be corrected. The time increment for each rate-of-change filter is set by the user and will represent changes over the user-defined time periods (for example: 1 hour, 3 hours, 96 hours, etc.). Filters 11 and 12 are

set to 90 percent of the historical maximum hourly rate-of-change increase or decrease at a station, respectively. Filters 9, 10, 13, and 14 represent rate-of-change thresholds based on a timespan of 3 hours. Filters 9 and 13 represent 95 and 90 percent of the maximum rate-of-change increase in water levels over a 3-hour span, respectively. Filters 10 and 14 represent 95 and 90 percent of the maximum rate-of-change decrease in water levels over a 3-hour timespan, respectively.

Linear Regression Models

In the second step for quality assuring the data, ADAM generates water-level estimates to augment actual measurements to be used when data are (1) missing, (2) removed because they violate acceptable filter thresholds, or (3) determined to be erroneous during a manual review (fig. 4). This step uses the subset of good filtered data either as inputs for existing simple linear regression (SLR) models or to create new multivariate linear regression (MLR) models to use as inferential sensors for each station in the network.

Simple Linear Regression Models

The SLR models are static equations entered into ADAM by the user for each station. One to three SLR models were developed for each EDEN station using daily water-level data from March 1, 2006, to September 30, 2011, as described in Petkewich and Conrads (2013). The resulting 667 water-level estimation equations (called “Gapfill” equations in the ADAM application) are in the form of $y = mx + b$, where y is the estimated value, m is the slope, x is the value from the input (predictor) station, and b is the y -intercept. To minimize the possibility of missing data from the input station, up to three linear regression equations (designated as P1, P2, and P3) were developed for each EDEN station by using the most highly correlated EDEN input stations. An example of the measured and estimated water levels for station S146_H is shown in figure 5.

Statistical measures of prediction accuracy were computed for the 667 water-level models and are provided in appendix 1 of Petkewich and Conrads (2013). Each statistic measures a different aspect of the accuracy of the prediction equations. Estimation accuracy commonly is reported in terms

of the coefficient of determination (R^2) and is interpreted as the goodness-of-fit of an equation or model. The standard error is the measure of the scatter of the actual observations about the regression line and is the standard deviation of the error of the predicted values in the regression. The mean error (ME) and root mean square error (RMSE) statistics provide a measure of the prediction accuracy of the estimation equations. The ME is a measure of the bias of model predictions—whether the model over- or underpredicts the measured data. The ME is the overall adjustment of the estimated values required to equal the measured values; therefore, positive and negative MEs indicate over- and underprediction bias by the model, respectively. MEs near zero may be misleading, however, because negative and positive discrepancies in the simulations can cancel each other. The RMSEs 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 those of the variable simulated by the model. The percent model error (PME) is computed by dividing the RMSE by the range of the measured data.

The majority of the SLR water-level estimation equations provide good estimates for missing values. Figure 6 shows the exceedance frequencies of the R^2 values for the P1, P2, and P3 estimation equations and the cumulative frequency for all 667 predictor equations. More than 99 percent of the 223 P1 equations have R^2 values greater than 0.70 and over 83 percent have R^2 values greater than 0.90. Fifty percent of all 667 equations have R^2 values greater than 0.95, more than 72 percent (484 equations) have R^2 values greater than 0.90, and more than 97 percent (647 equations) have R^2 values greater than 0.70. Summary statistics describing the range of R^2 , RMSE, standard error, and percent model error for all 667 equations are listed in table 2.

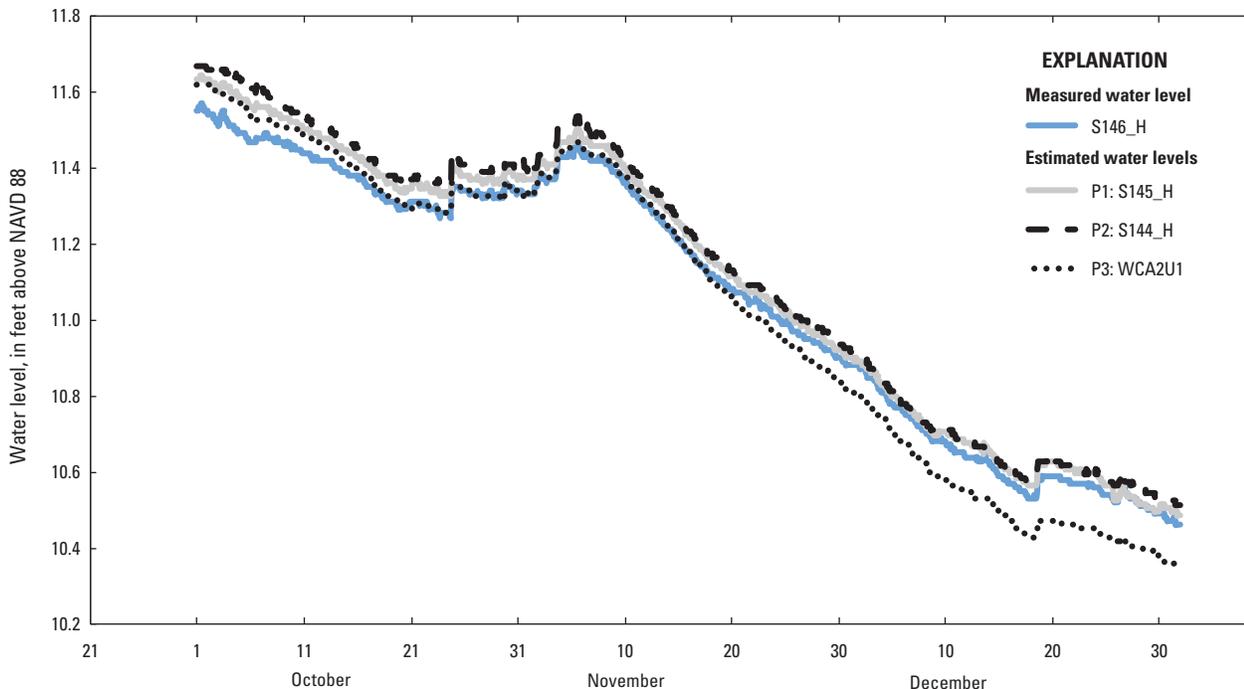


Figure 5. Measured and estimated water level for station S146_H for the period October 1, 2010, through December 31, 2010 (modified from Petkewich and Conrads, 2013).

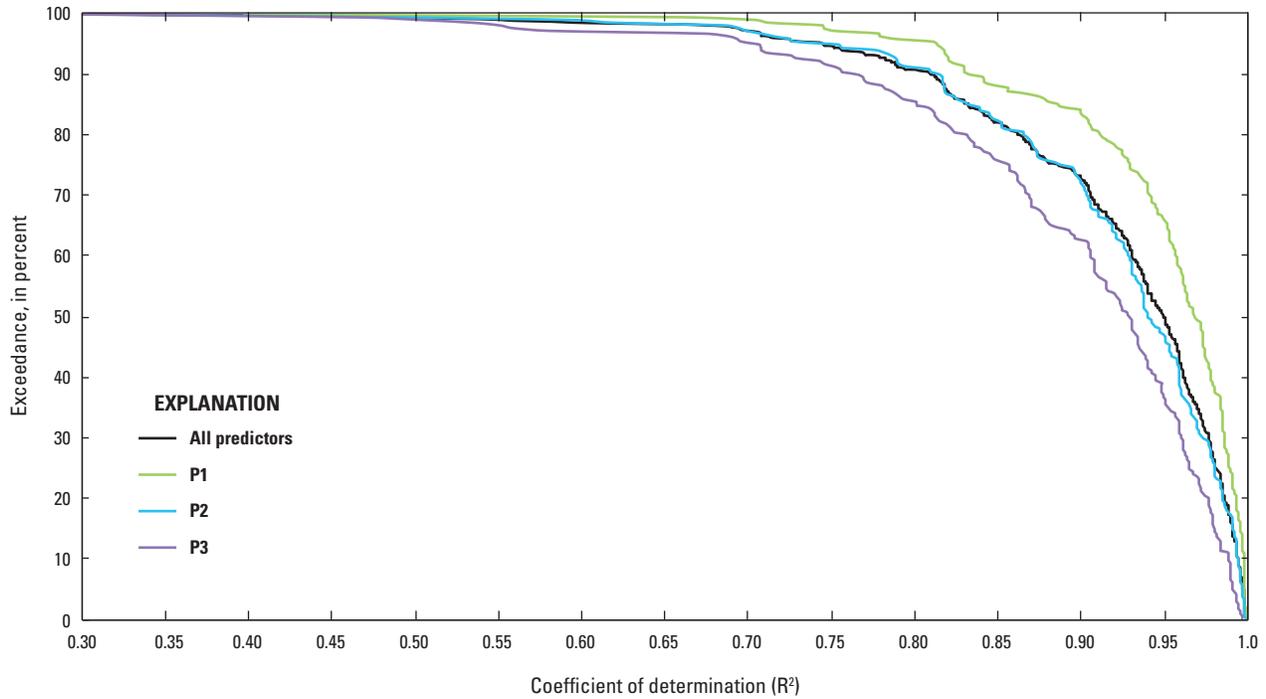


Figure 6. Exceedance percentage for coefficient of determination (R^2) for first, second, third predictor, and all 667 predictor water-level estimation equations (from Petkewich and Conrads, 2013).

Table 2. Minimum, median, and maximum values for the summary statistics for 667 water-level estimation equations (from Petkewich and Conrads, 2013).

[R^2 , coefficient of determination; RMSE, root mean square error; %, percent]

Statistic	Minimum	Median	Maximum
R^2	0.292	0.949	1
RMSE	0.013	0.165	1.036
Standard error	0.01	0.16	0.68
Percent model error	0.25%	4.30%	20.30%

Multivariate Linear Regression Models

A limitation of using the Gapfill equations is that there could be times when none of the three predictor input stations are available for model input or some of the input station data are erroneous. The Gapfill equations also are limited by their use of a single input variable, and there may be periods for which better water-level estimates could be generated using multiple model inputs. An alternate approach is to create regression models “on the fly” by using the available subset of good data from the first step in the ADAM process. The challenge of using this approach is to develop the computer algorithms to select the most correlated input stations and to decorrelate selected station data if multivariate regression equations are used. To create multivariate regression equations on the fly, a cross-correlation matrix of Pearson coefficients is calculated using the most recent period of filtered data. The matrix consists of values indicative of how well correlated data from potential input stations are to data from the output station. Candidate stations selected from the matrix as input to a multivariate regression model for a given station are selected based on degree of correlation. The selected signals must then be automatically decorrelated from each other. A statistical technique known as principal component analysis (PCA; Jolliffe, 2002), which has been widely used in data analysis and compression, was selected to compute the best set of input stations, decorrelate the data, and compute the regression coefficients.

The PCA statistical technique is utilized to reduce the dimensionality of a dataset consisting of a large number of correlated variables (Jolliffe, 2002; Rencher, 1998). In this process, a few uncorrelated principal components (PC) are used to explain a large part of the total sample variance of the original variables. The PCs are ordered so that the first components explain most of the variation of the original variables. Each subsequent component explains less variation than the previous component. In simple terms, PCA is performed by calculating the eigenvalues and eigenvectors of the covariance matrix of the assembled dataset (each eigenvector has an associated eigenvalue). When sorted by eigenvalue, highest to lowest, the first eigenvector (which is the PC) explains most of the variance associated with the original variables. Because all eigenvectors of a symmetric matrix, in this case the covariance matrix, are orthogonal to each other, each PC is decorrelated from any other. The PCs often are used as input for other analyses, such as MLR (Rencher, 1998).

The MLR models are dynamic estimates for individual stations that are calculated during the ADAM analysis using

recent filtered data. The implementation of the MLR model development process needs to be constrained by the number of stations and the time period to use for the MLR. In ADAM, user settings specify the number of stations (1 to 5) and the number of days (30 to 366) to be used in the model development. Once the input stations are determined and decorrelated from one another, MLR is used to model the relationship between the decorrelated inputs and the output station, and the predicted values are calculated. The MLR setup was as follows: include up to three stations; use between 30 and 366 days of historical data, with at least 50 percent of the time period with overlapping/concurrent data overlap; and recalculate MLR every 60 days by using the most recent data or reuse current MLR if the best match had no data available. Stations used for the MLR calculations are restricted to the same WCA. Model statistics are not presented for the MLR models because the models are developed automatically by the ADAM program and are dependent on the period of record selected. Model statistics can be computed, however, from the ADAM output data (measured and predicted values) along with the stations used for inputs to the MLR model.

An example of the SLR and MLR estimates for station NP205 located in ENP are shown in figure 7 for the period October 1, 2011, to October 1, 2012. During certain periods of the year, one model clearly outperforms the other. Summary statistics of R^2 , RMSE, and PME for the full period are listed in table 3. Overall, the SLR estimates are better than the MLR estimates, as indicated by the lower RMSE and PME values for SLR. For both the SLR and MLR estimates, the model performance changes with time, depending on the available inputs for the models. Figure 8 shows time series of R^2 values for the estimates during the simulation period. The R^2 for the SLR model increased from 0.59 to 0.90 in November 2011 when the SLR model switched from using station LOOP2_H for an input to using station SPARO. The SPARO station instrumentation was installed and activated on November 2, 2011; prior to that date the SPARO station was not available. The R^2 for the MLR model is much more variable than for the SLR model, with the R^2 values ranging from 0.56 to 0.99. The variability of the MLR model can be attributed to (1) MLR stations restricted to the same conservation area as the station being estimated, (2) the cycling of the best correlated input stations into and out of the pool of available stations, and (3) the recalculation of MLR equations every 60 days. Three input stations were used in the MLR model per estimation period for this example and are listed in table 4.

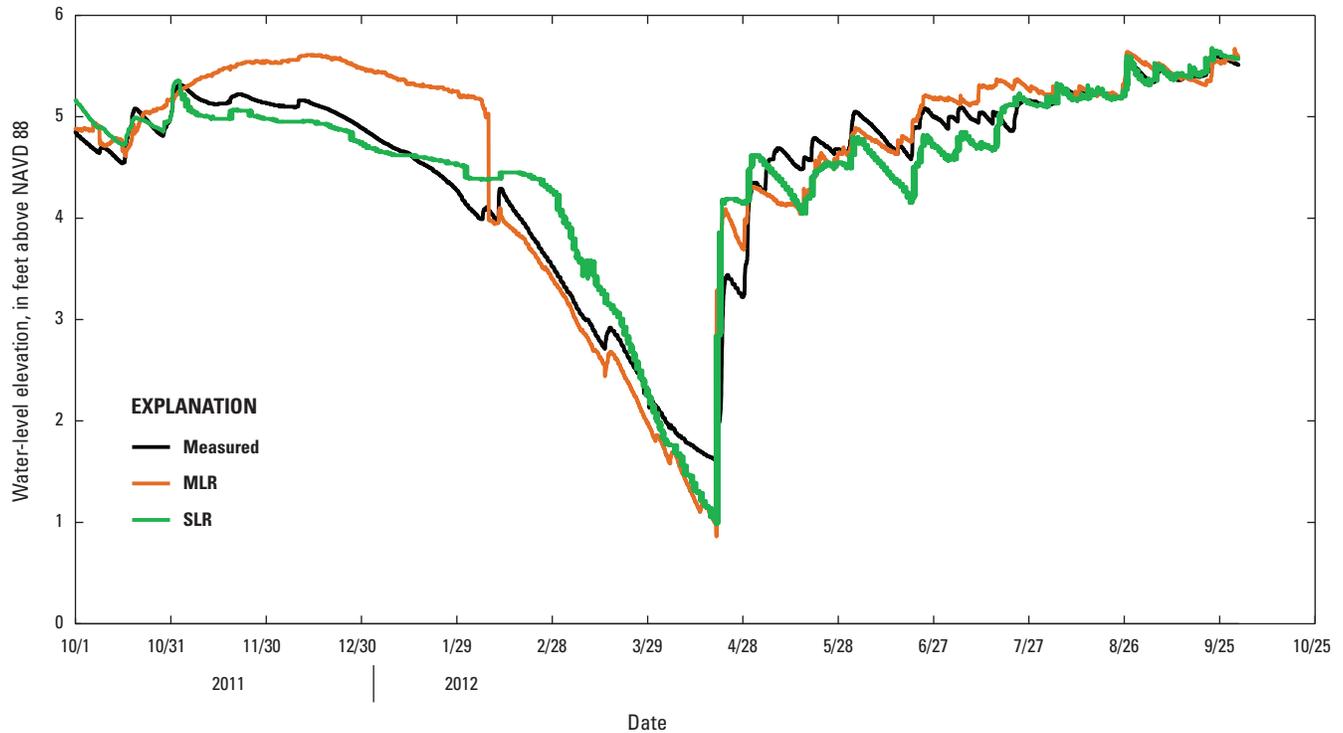


Figure 7. Measured and predicted water levels using simple linear regression and multivariate linear regression (SLR and MLR) for station NP205 for October 1, 2011, to October 1, 2012.

Table 3. Measured and predicted (MLR and SLR) water levels and statistics for station NP205 for Oct. 1, 2011, to Oct. 1, 2012.

[MLR, multivariate linear regression; SLR, simple linear regression; R^2 , coefficient of determination; RMSE, root mean square error; PME, percent model error]

Measured			MLR			SLR		
Minimum	Maximum	Range	R^2	RMSE	PME	R^2	RMSE	PME
1.62	5.60	3.98	0.906	0.389	0.098	0.905	0.301	0.076

12 Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

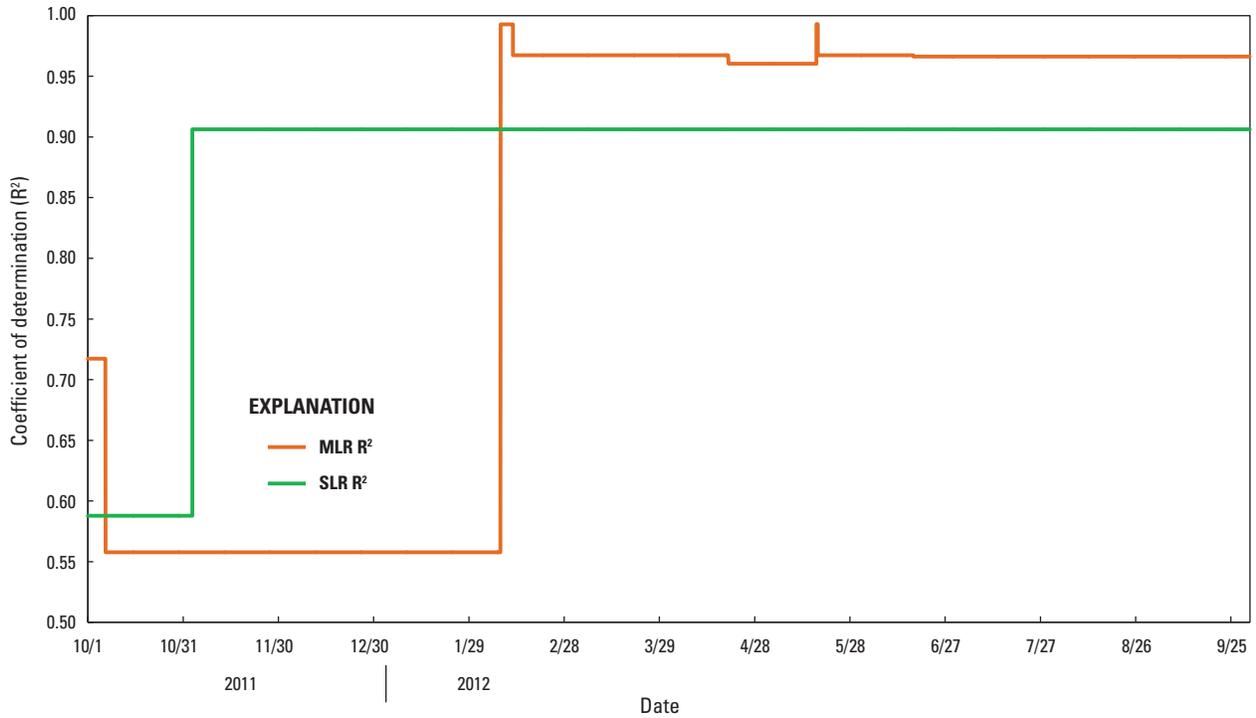


Figure 8. Coefficient of determination (R²) for simple linear regression and multivariate linear regression (SLR and MLR) models for station NP205 for October 1, 2011, to October 1, 2012.

Table 4. Range of dates and coefficient of determination (R²) for simple linear regression and multivariate linear regression models for station NP205 for Oct. 1, 2011, to Oct. 1, 2012.

[MLR, multivariate linear regression; SLR, simple linear regression. Dates shown as month, day, year; –, station not used for this range of dates]

Method	Station	Range of dates							
MLR	NP203	10/1/2011-10/6/2011	10/7/2011-2/7/2012	–	–	–	–	–	–
	EP1R	10/1/2011-10/6/2011	10/7/2011-2/7/2012	–	–	–	–	–	–
	NESRS1	10/1/2011-10/6/2011	–	–	–	–	–	–	–
	EPSW	–	10/7/2011-2/7/2012	–	–	–	–	–	–
	SPARO	–	–	2/8/2012-2/11/2012	–	–	5/17/2012	–	–
	NP206	–	–	2/8/2012-2/11/2012	2/12/2012-4/19/2012	–	5/17/2012	5/18/2012-6/16/2012	6/17/2012-9/30/2012
	G-620	–	–	2/8/2012-2/11/2012	2/12/2012-4/19/2012	4/20/2012-5/16/2012	–	5/18/2012-6/16/2012	–
	RG1	–	–	–	2/12/2012-4/19/2012	4/20/2012-5/16/2012	5/17/2012	5/18/2012-6/16/2012	6/17/2012-9/30/2012
	G-1502	–	–	–	–	4/20/2012-5/16/2012	–	–	–
	NP62	–	–	–	–	–	–	–	6/17/2012-9/30/2012
MLR R ²	–	0.72	0.56	0.99	0.97	0.96	0.99	0.97	0.97
SLR	LOOP2_H	10/1/2011-11/2/2011		–					
	SPARO	–		11/3/2011-9/30/2012					
SLR R ²	–	0.59		0.90					

ADAM Software Architecture

The general architectural elements for the ADAM software are graphical user interfaces (GUIs), a series of databases, and a set of input and output files (fig. 9 and table 5). There are two methods for running the ADAM software—automated reviews and manual reviews, the latter of which are interactive and allow user modification. For automated reviews, the process of data input, review, modification, and

export is programmed using a series of computer scripts. For the manual reviews, a user can evaluate the automated reviews, make modifications to estimated data, and export modified data to the ADAM databases. Overall, the ADAM software consists of 2 Microsoft Excel workbooks that are the GUIs for setting up the automated and manual review processes, 6 Microsoft Access databases, and 2 output files. A general description of elements in the ADAM software is presented next. Additional details about the software are provided in the ADAM user manual (Petkewich and others, 2016).

Table 5. Input, workbook, databases, and output files used in the ADAM software program.

[ADAM, Automated Data Assurance and Management]

File name	File type	Description
ADAM_Review.xlsm	Workbook	Used to set up and execute a manual review.
ADAM.xlsm	Workbook	Used to set up and execute an automated review.
ADAM_daily.mdb	ADAM program database	Archival database for daily ADAM runs.
ADAM_qtly.mdb	ADAM program database	Archival database for quarterly ADAM runs.
ADAM_annual.mdb	ADAM program database	Archival database for annual ADAM runs.
ADAM_Setup.mdb	ADAM program database	Used for all reviews. Contains station information, filter settings, prediction equations and status notes for the predictions, preliminary values, reviewed values, and output run identification.
ADAM_Import.mdb	ADAM program database	Used for automated review. Temporarily stores data imported from text file to facilitate data analysis.
ADAM_Query.mdb	ADAM program database	Used for automated review. Temporarily stores current and historical data and multivariate linear regression results to facilitate data analysis.
data_uv.txt	Input file	Used for all reviews. This file is a composite of all agency data for the given ADAM run.
schema.ini	Input file	Used for all reviews. The schema information file provides information regarding the format of the data_uv.txt file.
schema_reference.ini	Input file	Used for all reviews. The schema_reference information file provides information regarding the format of the data_uv.txt file.
ADAM_Run nm .mdb	Output file	Stores all data for a given run and includes the setup tables which contain details regarding the filter settings and predictions made. The nm part of the name is a unique numeric variable indicating the ADAM Run identifier.
ISOutput_Run.txt	Output file	Text output file.

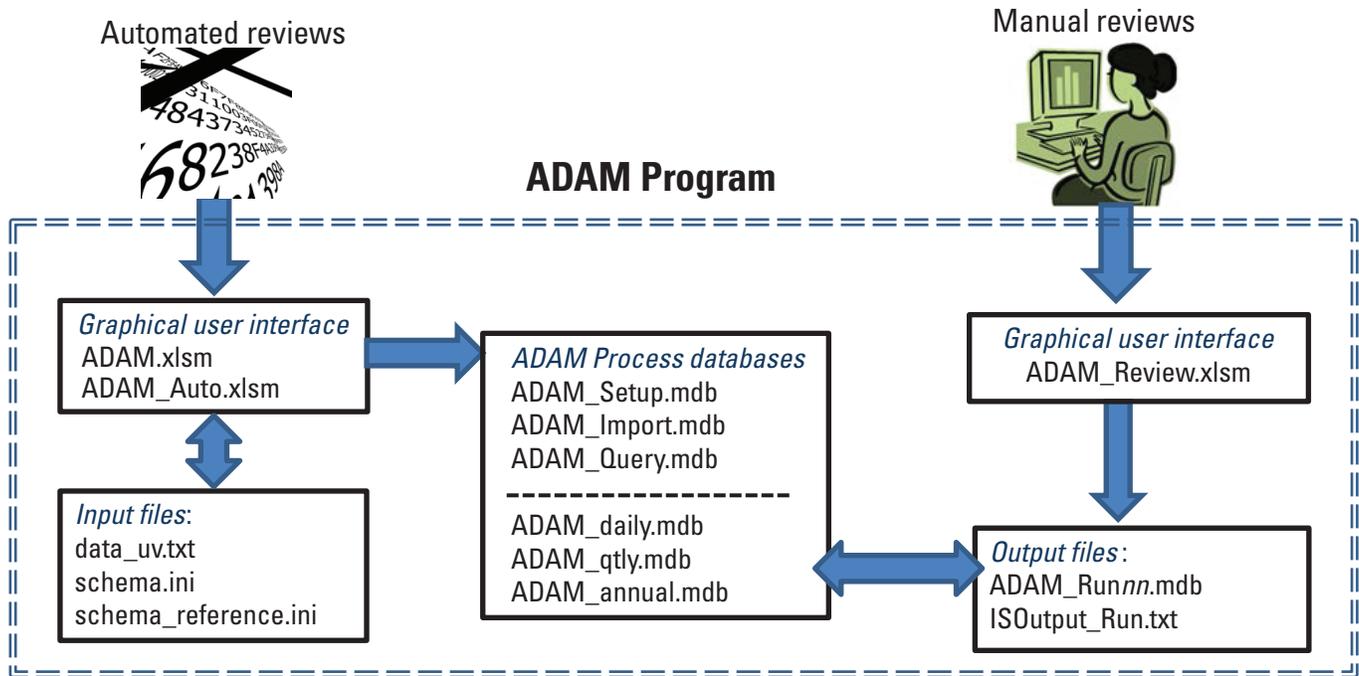


Figure 9. General architectural elements for the Automated Data Assurance and Management (ADAM) software.

The ADAM.xlsm workbook is the GUI for the automated review of the data and consists of 15 worksheets that contain the control settings and Visual Basic for Application (VBA) computer code used to run ADAM (table 6). The Control worksheet allows the setup and execution of ADAM, including selecting the directory where the analysis will be stored, adjusting the number of stations and period of time used in the MLR, and database management such as adding new stations to the review process or archiving the current database. The Filter_Adjustments worksheet allows adjustments to the filter settings for each station through the use of control buttons and a graph showing the historical data and current filter settings. The ADAMSiteList worksheet allows management of stations that are included in an ADAM analysis. The PrelimLogic worksheet allows adjustment of the ADAM settings that control when to override measured data and when to select MLR or SLR as the estimated data. The Gapfill worksheet allows the management of SLR equations, including the ability to add, delete, or modify specific equations. Other worksheets included in the ADAM.xlsm workbook generally are used for documentation purposes only. The design of the ADAM.xlsm workbook is discussed in more detail in Petkewich and others (2016).

Running the ADAM program accesses and (or) updates 6 Microsoft Access databases: ADAM_Setup.mdb, ADAM_Import.mdb, ADAM_Query.mdb, ADAM_daily.mdb, ADAM_qtly.mdb, and ADAM_annual.mdb (fig. 9 and table 5). The ADAM_Setup.mdb stores the list of stations to be analyzed by ADAM, filter settings used for each station, ADAM run identifiers, MLR and SLR prediction equations, and static status notations for the predictions, preliminary values, and reviewed values that will be determined during

the processing. The ADAM_Import.mdb stores all the data analyzed by ADAM, including data not used in the analysis such as data not recorded on an hourly basis (15-minute data). The ADAM_Query.mdb stores tables and queries used in managing data for the ADAM program. All data are deleted from the ADAM_Query.mdb after each ADAM program run conclusion. The ADAM_daily.mdb, ADAM_qtly.mdb, and ADAM_annual.mdb are the archival databases that store finalized results of the ADAM analyses. In addition to these databases that are updated during ADAM processing, a unique ADAM output database is created after each ADAM program run and is called “ADAM_Runnn.” The *nn* part of the name is a unique numeric variable indicating the ADAM Run identifier (fig. 9).

The workbook ADAM_Review.xlsm (fig. 9 and table 7) is the GUI for the manual review of data and allows for adjustment of the automated results with user-input controls on worksheets that display site-specific graphical and tabular data. The workbook consists of seven worksheets and VBA computer code. The Control worksheet allows the user to select the working ADAM directory, select the output file of interest for reviewing or exporting results, and select a specific date range for evaluating the filter trips within a file. When reviewing a specific output file, VBA code simultaneously opens the Review_Chart and Review_Tab worksheets. The Review_Chart worksheet is used to scroll through and view graphs of the various stations of interest in the output database. This worksheet allows the user to scroll through all stations in the database, only stations that have any filter trip for a specified date range, or only stations that have missing data for a specified date range. Items included on the Review_Chart can include preliminary values, reviewed values, estimated values

(MLR and SLR), similar stations, change in predictor indicators (indicating when a MLR predictor changes to a different predictor), and ground-surface elevation so that questionable data can be evaluated. The Review_Tab worksheet opens below the Review_Chart worksheet and lists time-specific data for all the stations, but most importantly the measured data, filtered data, estimated data (MLR or SLR), preliminary values, filter trips, and reviewed values. The preliminary values represent the results selected by the automated process. During the manual review process, a user can select preliminary, actual,

filtered, MLR, or SLR values as the reviewed value for each time stamp. The user also can apply complete or partial shifts to the most appropriate data source(s). When the manual review of a specific station is complete and any appropriate adjustment to the data or estimates of the data are made, the user saves the data to the ADAM output database and moves on to the next station. When all reviews are complete, the user saves the data to the archival database (ADAM_daily.mdb, ADAM_qtly.mdb, or ADAM_annual.mdb) and a output text file is created.

Table 6. Worksheets in the ADAM.xlsm workbook for automated review of the data.

[ADAM, Automated Data Assurance and Management; MLR, multivariate linear regression; PCA, principal component analysis; <, less than]

Worksheet name	Description	User modifiable
About	Lists current ADAM version and contact information.	No
Control	Used to setup and run ADAM.	Yes
Filter_Adjustments	Used to set/modify filter settings.	Yes
ADAMSiteList	Used for making changes to stations analyzed.	Yes
UserSiteList	Used for making changes to stations analyzed.	Yes
PrelimLogic	Used to set logic for selecting preliminary ADAM output (non-reviewed).	Yes
Gapfill	Used to manage Gapfill (simple linear regression) equations.	Yes
FilterGraph_Data	Data used to provide visual aid to setting filter setpoints.	No
RevisionHistory	Lists the history of modifications/updates to ADAM.	No
RegSelectionData	Contains all source data for MLR analysis. Data is extracted from the Access archival database at the start of multivariate linear regression analysis.	No
PCAPREP	Contains all data to be used in PCA Analysis for a particular station. Incomplete rows are removed and stations may be culled if there is not enough data overlap (i.e., row count < minimum amount of data for PCA analysis per user setting).	No
PCAData	Contains culled PCA data and results of PCA analysis (eigenvectors and eigenvalues).	No
RegressionData	Contains calculated principal components and results of multivariate linear regression.	No
SetupAPP	Hidden—reference sheet only.	No
pcacheck	Hidden—used for debugging only.	No

Table 7. Worksheets in the ADAM_Review.xlsm workbook for manual review of the data.

[ADAM, Automated Data Assurance and Management]

Worksheet name	Description	User modifiable
About	Lists current ADAM version and contact information.	No
Control	Used to setup and run ADAM.	Yes
Review_Tab	Tabular review of a selected run. Used to select reviewed data and save station review to output run database. Upon review completion, reviewed data can be saved to archival database.	Yes
Review_Chart	Used to select station to be reviewed. Graphical display of station data and predictions.	Yes
Current_TroubleList	List of all filter trips for a selected run. Can list all or select by FilterID.	Yes
Graph_Data	Lists all data used on the Review_Chart worksheet.	No
RevisionHistory	Lists the history of modifications/updates to ADAM_Review.xlsm.	No

Using ADAM Software to Process Everglades Depth Estimation Network Data

The quality-assurance methods employed using the ADAM software are most effective when both automated and manual reviews are conducted on a dataset. For the EDEN data, the automated process is set to run each morning, and once complete, all data that have filter trips on the day of interest will be manually reviewed. Reviewed results are loaded into the EDEN database and used for producing the daily water-surface maps for the previous day. Below is an example of the automated and manual processing using ADAM.

Automated Processing

The ADAM software is currently (2016) being used for automated real-time quality assurance of the EDEN data. The automated review is run every morning using the Windows Task Scheduler, which launches the ADAM_Auto.xlsm application. Once complete, the results are available for uploading into the EDEN database and daily water surfaces will be created whether individuals are available for reviewing the automated results or not; for example, on weekends or holidays. Every day of the week except Mondays, 1 day of data (collected the preceding day) are processed. On Mondays, 3 days of data (collected on Friday, Saturday, and Sunday) are processed to allow manual review of the previously unreviewed weekend data. In addition to the daily quality assurance of the EDEN data, ADAM also is used to evaluate quarterly and annual datasets that have received additional quality-assurance checks by the various agencies operating the EDEN stations. The automated analysis of EDEN real-time data using ADAM consists of the following steps, as illustrated in figure 4:

1. Import the data—Measured data from a daily text file is loaded into the ADAM Microsoft Access database.
2. Filter the data—Univariate filter trip ID (1–14) is automatically assigned to any data point that violates 1 of 14 station-specific, user-defined filter values (table 1).
3. Estimate the data—Two synthetic sets of hydrographs are created for each gaging station to replace missing or erroneous measured data and to provide additional quality assurance of data from each station. At this point, ADAM also can utilize a Dry-Protocol filter (15) and a difference between measured and estimated filter (16) to further quality assure the measured data (table 1). The

Dry-Protocol filter is an optional filter that flags the data when values drop below a user-defined limit, such as a specified ground-surface elevation. To represent dry conditions at a station, the user can assign a constant value (such as 0.5 ft below ground surface) to any data labeled with a filter value of 15. Filter trip 16 evaluates if the difference between the MLR predictions and the measured data exceeds a user-defined tolerance. If the tolerance is exceeded, the measured data are labeled with a filter value of 16 to notify the reviewer that the data may be erroneous.

4. Collate the final dataset—Data that passed the filter test are collated with either SLR or MLR predicted data, as needed.
5. Output the final data—The water-level data are exported in an appropriate format for the EDEN database.

Manual Processing

The ADAM manual review process is initiated by opening the ADAM_Review.xlsm application and selecting the RunID and days of interest in the Control worksheet (fig. 10). Any run can be selected for review, but for real-time quality assurance, the most current run is typically selected. After selection of the RunID of interest, the ADAM_Review.xlsm application allows the user to evaluate measured and predicted data for an individual station and the various filter trips, all of which are displayed on a graph and within a table (fig. 11). During the manual review, a reviewer will evaluate hydrographs of individual station data plotted with filter trip information, estimates of the station data, and adjacent station data (fig. 12) to make any necessary adjustments to the preliminary data prior to exporting the results for upload in the EDEN database to generate the water-surface maps. During this manual review, the reviewer can override any predicted values (to revert back to measured data), adjust the predicted values by shifting the values to better match the measured data, or re-estimate the period of interest by using other means, such as linear interpolation or using data from an adjacent station for estimation. Necessary adjustments can be identified and completed in the data table in the Review_Tab worksheet (figs. 11B and 13) by focusing on the sections of the table and associated graph containing any filter trips (FilterID). Inaccurate data may not be flagged, however, if they fall within the range of acceptable thresholds for the various filter settings. The ADAM_Review process allows the reviewer to evaluate all EDEN stations or only those that have triggered a filter trip on the day(s) of interest.

RunID	Run Start Date	Run Stop Date	File End Date Time	Last Analyzed	Analysis Complete	Type	Initials	Data Qtr	Data Year	Run Date	Review Done	Notes
779	1/30/2014 0:00	2/3/2014 7:00	2/3/2014 7:00	2/3/2014 7:00	TRUE	daily	mdp	2014	2014	2/3/2014	TRUE	
778	1/27/2014 0:00	1/31/2014 7:00	1/31/2014 7:00	1/31/2014 7:00	TRUE	daily	mdp	2014	2014	1/31/2014	TRUE	
777	1/26/2014 0:00	1/30/2014 7:00	1/30/2014 7:00	1/30/2014 7:00	TRUE	daily	mdp	2014	2014	1/30/2014	TRUE	
776	1/25/2014 0:00	1/29/2014 7:00	1/29/2014 7:00	1/29/2014 7:00	TRUE	daily	mdp	2014	2014	1/29/2014	FALSE	
775	1/24/2014 0:00	1/28/2014 3:12	1/28/2014 3:12	1/28/2014 3:00	TRUE	daily	mdp	2014	2014	1/28/2014	TRUE	
774	1/22/2014 0:00	1/27/2014 3:12	1/27/2014 3:12	1/27/2014 3:00	TRUE	daily	mdp	2014	2014	1/27/2014	TRUE	
773	1/20/2014 0:00	1/24/2014 7:00	1/24/2014 7:00	1/24/2014 7:00	TRUE	daily	mdp	2014	2014	1/24/2014	TRUE	
772	1/19/2014 0:00	1/23/2014 7:00	1/23/2014 7:00	1/23/2014 7:00	TRUE	daily	mdp	2014	2014	1/23/2014	TRUE	
771	1/18/2014 0:00	1/22/2014 7:00	1/22/2014 7:00	1/22/2014 7:00	TRUE	daily	mdp	2014	2014	1/22/2014	TRUE	
770	1/17/2014 0:00	1/21/2014 7:00	1/21/2014 7:00	1/21/2014 7:00	TRUE	daily	mdp	2014	2014	1/21/2014	TRUE	
769	1/16/2014 0:00	1/20/2014 7:00	1/20/2014 7:00	1/20/2014 7:00	TRUE	daily	mdp	2014	2014	1/20/2014	FALSE	
768	1/13/2014 0:00	1/17/2014 7:00	1/17/2014 7:00	1/17/2014 7:00	TRUE	daily	mdp	2014	2014	1/17/2014	TRUE	
767	1/12/2014 0:00	1/16/2014 7:00	1/16/2014 7:00	1/16/2014 7:00	TRUE	daily	mdp	2014	2014	1/16/2014	TRUE	
766	1/11/2014 0:00	1/15/2014 7:00	1/15/2014 7:00	1/15/2014 7:00	TRUE	daily	mdp	2014	2014	1/15/2014	TRUE	
765	1/10/2014 0:00	1/14/2014 7:00	1/14/2014 7:00	1/14/2014 7:00	TRUE	daily	mdp	2014	2014	1/14/2014	TRUE	
764	1/9/2014 0:00	1/13/2014 7:00	1/13/2014 7:00	1/13/2014 7:00	TRUE	daily	mdp	2014	2014	1/14/2014	FALSE	

Figure 10. Automated Data Assurance and Management (ADAM) tool in "Review" mode showing part of Control worksheet.

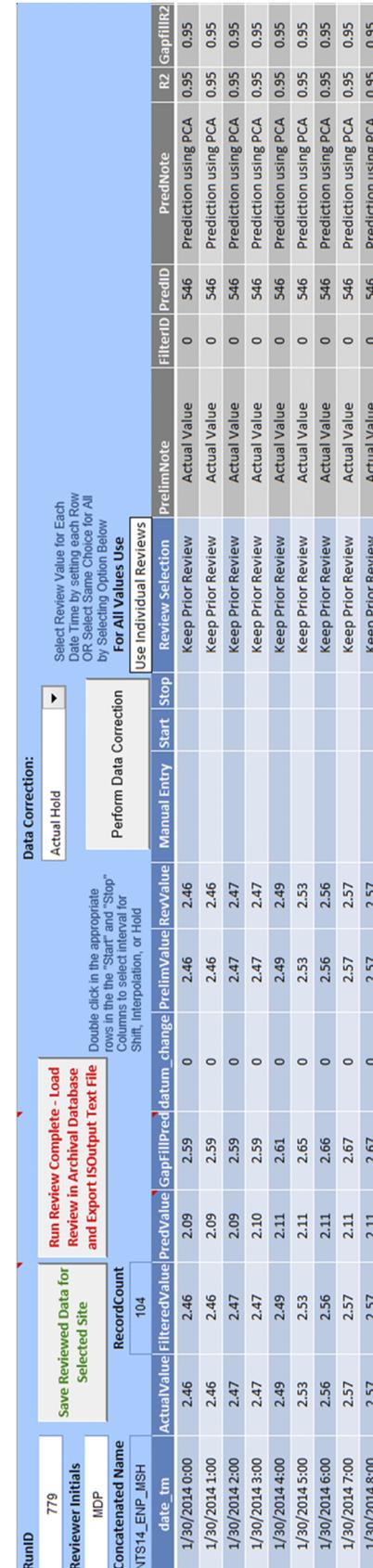
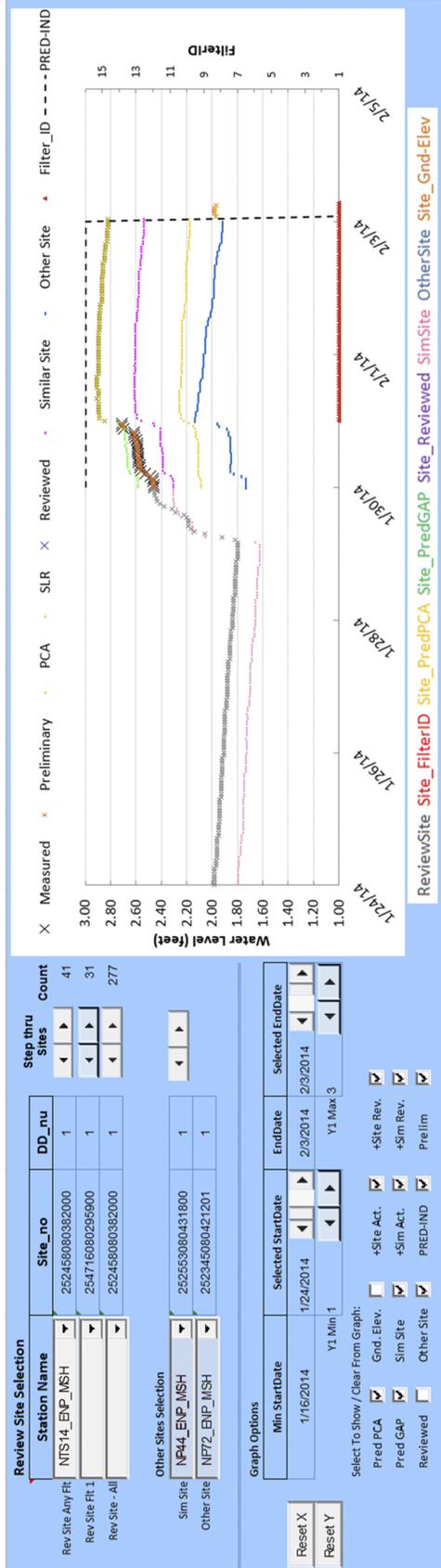


Figure 11. Automated Data Assurance and Management (ADAM) tool in "Review" mode showing, A, Review_Chart and, B, Review_Tab worksheets.

Figures 11, 12, and 13 illustrate how ADAM is used to quality assure real-time EDEN data. For this example, the days of interest are January 30 through February 2, 2015. The left-hand part of the Review_Chart worksheet (figs. 11A and 12A) contains options to select recently measured data for the station of interest (from one of the review site selection dropdown boxes called “Rev Site Any ft,” “Rev Site ftl 1,” and “Rev Site – All”), a similar station (called “Sim Site”), and a second similar station (called “Other Site”). Below the review site selection area are check boxes for displaying estimated data (MLR: Pred PCA; SLR: Pred GAP), and (or) previously measured data (called “+Site Act.,” “+Sim Act.”) and previously reviewed data (called “Site Rev.,” “Sim Rev.”), which can be selected for the station of interest and the similar station, respectively. The right-hand part of the Review_Chart worksheet (figs. 11A and 12B) contains a graph that shows measured data, denoted by large black crosses, for station NTS14 for the dates of interest (January 30 to February 3, 2014) in addition to previously reviewed data, denoted by small black crosses, and previously reviewed and current data for a similar station (Sim_Site: small pink symbols). Red triangles depict the filter trip identification number as listed on the secondary y-axis and indicate that missing data (FilterID 1) is the only filter trip for this station for the period of interest. These overlapping triangles appear as a solid red line along the x-axis. There are no measured data for this station for the period of January 1, 2014, 12:00 a.m., to February 3, 2014, 6:00 a.m., which is the last time stamp in the run file. Ground-surface elevation can be shown as a brown dashed line, but is not shown for this example. Yellow and green symbols depict the MLR and SLR estimates, respectively. (In the graphs generated by the ADAM application, the MLR models are labeled as PCA models.) Black dashes (predictor indicator, PRED-IND) indicates that the stations used for MLR estimation changed February 3, 2014, at 1:00 a.m. Noting a change in predictors is important because it signifies a probable shift in the MLR estimation and fit of estimation (R^2). For example, at the 1:00 a.m. change in MLR predictors the MLR estimate increases from 2.18 to 2.19 ft (fig. 13A). A second change in MLR predictors occurs at 2:00 a.m. on February 3rd, and the estimated water level drops from 2.19 ft to 1.97 ft. The preliminary value (small orange cross) displays the value that will be exported to the EDEN database if no adjustments are made to the data (fig. 12B).

The automated results selected the measured value for all times where measured data were available, SLR estimates for all missing values up to and including the February 3, 2014, 1:00 a.m. value, and MLR estimates for the remaining values (figs. 12B and 13A). Estimates with the higher R^2 are chosen as preliminary values for all replacement data. In this example, R^2 equals 0.952 (0.95 rounded) for the SLR and 0.946 (0.95 rounded) for the MLR estimates prior to February 3, 2014, 12:00 a.m.; therefore, SLR (called “Predicted GAP-FILL” in the PrelimNote column) is chosen as the preliminary value (fig. 13A). After February 3, 2014, 1:00 a.m., all SLR

stations are unavailable and the preliminary value switches to the MLR values (called “Predicted value” in the PrelimNote column) with an R^2 equal to 0.94 for 1:00 a.m. and 0.90 for the remaining values (fig. 13A). The R^2 values for each time stamp can be found listed in the Review_Tab worksheet or accessed by clicking on the PredValue (MLR) or GapFillPred (SLR) value of interest (fig. 13B). To override the preliminary values, the reviewer can either

1. select “Actual Value,” “Predicted Value,” or “Predicted Value – Gapfill” from a dropdown menu in the “Review Selection” column of the Review_Tab worksheet (fig. 13C);
2. make a manual correction using the “Data Correction” dropdown menu, “Perform Data Correction” button, and “Start” and “Stop” columns to apply appropriate interpolations by using a reviewer-selected data type within the Review_Tab worksheet (fig. 13D); or
3. make a manual correction by typing in the corrections in the “Manual Entry” column and selecting the appropriate correction identifier from the dropdown menu in the “Review Selection” column of the Review_Tab worksheet (fig. 13E).

As discussed previously, figure 11A shows that data for station NTS14 were not available for 80 hours of a review period beginning on January 30, 2014. Data for the similar station (NP44, pink symbol), other station (NP72, blue symbol), SLR estimates (green symbol, $R^2=0.952$), and MLR (yellow symbol, $R^2=0.941$) appear to generally match the trend of the measured data when the data are available. The estimates may need to be adjusted, however, in order to avoid any great shifts caused by switching from measured to estimated values. The measured, MLR, and SLR values for January 30, 2014, 11:00 p.m., are 2.71, 2.19, and 2.75 ft above NAVD 88 (fig. 13D), respectively. The preliminary values are set equal to the SLR estimates for the period of missing data because of the higher R^2 value for SLR compared to MLR. To adjust the SLR data to a value that matches the last measured value of station NTS14, a reviewer would enter a formula in the “Manual Entry” column to apply a -0.04 ft shift, which is determined by subtracting the SLR value from the measured value ($2.71 - 2.75$). Using this formula and a “Review Selection” setting of “GAPFILL_with_Shifts” for all of the missing data produces a plausible hydrograph of reviewed values for the period of review (fig. 13F). A similar hydrograph could have been produced in the “Manual_Entry” column by creating a formula using the MLR data; applying a $+0.52$ -ft shift, determined by subtracting the MLR value from the measured value ($2.71 - 2.19$); and selecting “PCA_Pred_with_Shifts” for the “Review_Selection” column. Estimated data after February 3, 2014, at 12:00 a.m. did not have to be adjusted during this review as values for that date were not complete and not part of the review days of interest.

In addition to providing a relatively simple and effective method for reviewing automated quality-assured data, the

A

Review Site Selection

	Station Name	Site_no	DD_nu	Step thru Sites	Count
Rev Site Any Flt	NTS14_ENP_MSH	252458080382000	1	◀ ▶	41
Rev Site Flt 1		254716080295900	1	◀ ▶	31
Rev Site - All		252458080382000	1	◀ ▶	277

Other Sites Selection

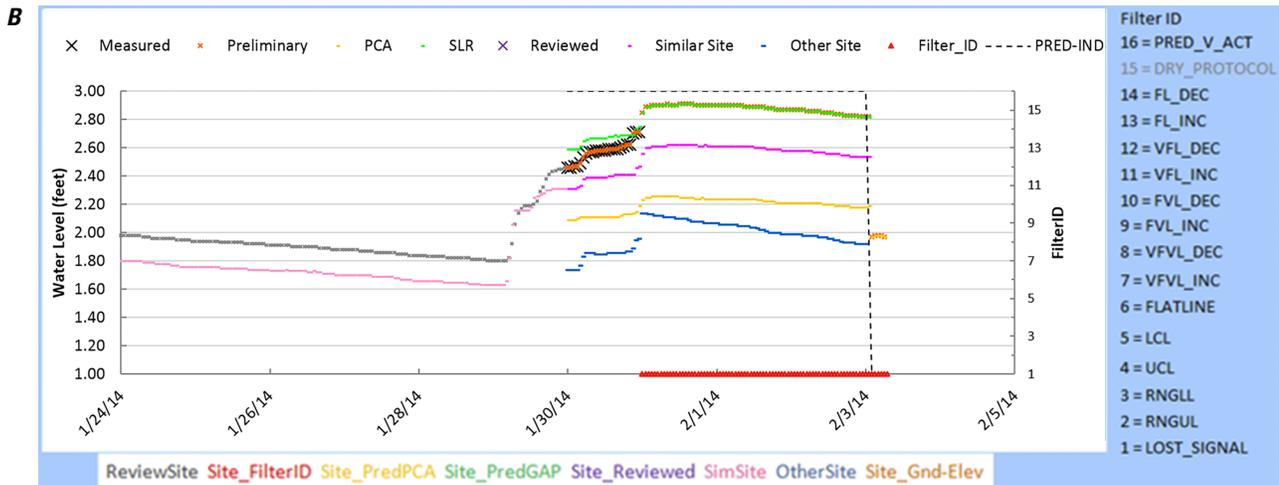
Sim Site	NP44_ENP_MSH	252553080431800	1	◀ ▶
Other Site	NP72_ENP_MSH	252345080421201	1	

Graph Options

	Min StartDate	Selected StartDate	EndDate	Selected EndDate
Reset X	1/16/2014	1/24/2014	2/3/2014	2/3/2014
Reset Y		Y1 Min 1	Y1 Max 3	

Select To Show / Clear From Graph:

Pred PCA	<input checked="" type="checkbox"/>	Gnd. Elev.	<input type="checkbox"/>	+Site Act.	<input checked="" type="checkbox"/>	+Site Rev.	<input checked="" type="checkbox"/>
Pred GAP	<input checked="" type="checkbox"/>	Sim Site	<input checked="" type="checkbox"/>	+Sim Act.	<input checked="" type="checkbox"/>	+Sim Rev.	<input checked="" type="checkbox"/>
Reviewed	<input type="checkbox"/>	Other Site	<input checked="" type="checkbox"/>	PRED-IND	<input checked="" type="checkbox"/>	Prelim	<input checked="" type="checkbox"/>



EXPLANATION

- PCA** Multivariate linear regression estimate
- SLR** Simple linear regression estimate
- Reviewed** Previously reviewed data
- Similar Site** Previously reviewed data and current data for a similar station
- Other Site** Previously reviewed data and current data for an additional similar station
- Filter ID** Filter trip identification number
- PRED—IND** Predictor indicator

Figure 12. Automated Data Assurance and Management (ADAM) tool in “Review” mode showing A, the chart controls and B, graph of the Review_Chart worksheet.

A

RunID	779	Data Correction:		Select Review Value for Each Date Time by setting each Row OR Select Same Choice for All by Selecting Option Below For All Values Use Use Individual Reviews													
Reviewer Initials	MDP	Save Reviewed Data for Selected Site	Run Review Complete - Load Review in Archival Database and Export ISOutput Text File	Actual Hold													
Concatenated Name	NTS14_ENP_MSH	RecordCount	104	Perform Data Correction													
date_tm	ActualValue	FilteredValue	PredValue	GapFillPred	datum_change	PrelimValue	RevValue	Manual Entry	Start	Stop	Review Selection	PrelimNote	FilterID	PredID	PredNote	R2	GapfillR2
2/2/2014 17:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 18:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 19:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 20:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 21:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 22:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 23:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/3/2014 0:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/3/2014 1:00			2.19	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	547	Prediction using PCA	0.94	0.95
2/3/2014 2:00			1.97		0	1.97					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 3:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 4:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 5:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	

B

RunID	779	Data Correction:		Select Review Value for Each Date Time by setting each Row OR Select Same Choice for All by Selecting Option Below For All Values Use Use Individual Reviews													
Reviewer Initials	MDP	Save Reviewed Data for Selected Site	Run Review Complete - Load Review in Archival Database and Export ISOutput Text File	Actual Hold													
Concatenated Name	NTS14_ENP_MSH	RecordCount	104	Perform Data Correction													
date_tm	ActualValue	FilteredValue	PredValue	GapFillPred	datum_change	PrelimValue	RevValue	Manual Entry	Start	Stop	Review Selection	PrelimNote	FilterID	PredID	PredNote	R2	GapfillR2
2/2/2014 17:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 18:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 19:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 20:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 21:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 22:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/2/2014 23:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/3/2014 0:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	546	Prediction using PCA	0.95	0.95
2/3/2014 1:00			2.19	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1	547	Prediction using PCA	0.94	0.95
2/3/2014 2:00			1.97		0	1.97					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 3:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 4:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	
2/3/2014 5:00			1.98		0	1.98					Prelim Value	Predicted Value	1	548	Prediction using PCA	0.90	

Prediction Details

Site: NTS14_ENP_MSH

Prediction Source: Prediction using PCA

R2: 0.95

SimSite1: TSH_ENP_MSH

SimSite2: NP206_ENP_MSH

SimSite3: A13_ENP_MSH

SimSite4:

SimSite5:

Reg StartDate: 8/1/2013 11:45:00 PM

Reg StopDate: 1/29/2014 11:45:00 PM

GAPFILL Site: NP44_ENP_MSH

GAPFILL R2: 0.95

EXPLANATION

- PCA** **Multivariate linear regression estimate**
- SLR** **Simple linear regression estimate**
- Reviewed** **Previously reviewed data**
- Similar Site** **Previously reviewed data and current data for a similar station**
- Other Site** **Previously reviewed data and current data for an additional similar station**
- Filter ID** **Filter trip identification number**
- PRED—IND** **Predictor indicator**

Figure 13. Automated Data Assurance and Management (ADAM) tool in “Review” mode showing parts of the Review_Tab and Review_Chart worksheets and, *A*, example of preliminary results, *B*, prediction detail information, *C*, manual review procedure using the “Review Selection” dropdown menu, *D*, manual review procedure using “Data Correction” dropdown menu, *E*, manual review procedure using “Manual Entry” column and “Review Selection” dropdown menu and, *F*, graphical results of a manual review.

22 Using Inferential Sensors for Quality Control of Everglades Depth Estimation Network Water-Level Data

C

RunID 779	Save Reviewed Data for Selected Site	Run Review Complete - Load Review in Archival Database and Export ISOOutput Text File	Data Correction: Actual Hold		Select Review Value for Each Date Time by setting each Row OR Select Same Choice for All by Selecting Option Below								
Reviewer Initials MDP	Double click in the appropriate rows in the "Start" and "Stop" Columns to select interval for		Perform Data Correction		Use Individual Reviews								
Concatenated Name NTS14_ENP_MSH	RecordCount 104				For All Values Use								
date_tm	ActualValue	FilteredValue	PredValue	GapFillPred	datum_change	PrelimValue	RevValue	Manual Entry	Start	Stop	Review Selection	PrelimNote	FilterID
2/2/2014 17:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1
2/2/2014 18:00			2.19	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1
2/2/2014 19:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1
2/2/2014 20:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1
2/2/2014 21:00			2.18	2.83	0	2.83					Prelim Value	Predicted GAPFILL	1
2/2/2014 22:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1
2/2/2014 23:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1
2/3/2014 0:00			2.18	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1
2/3/2014 1:00			2.19	2.82	0	2.82					Prelim Value	Predicted GAPFILL	1
2/3/2014 2:00			1.97		0	1.97					Prelim Value	Predicted Value	1
2/3/2014 3:00			1.98		0	1.98					Prelim Value	Predicted Value	1
2/3/2014 4:00			1.98		0	1.98					Prelim Value	Predicted Value	1
2/3/2014 5:00			1.98		0	1.98					Prelim Value	Predicted Value	1
2/3/2014 6:00			1.97		0	1.97					Prelim Value	Predicted Value	1

D

RunID 779	Save Reviewed Data for Selected Site	Run Review Complete - Load Review in Archival Database and Export ISOOutput Text File	Data Correction: Actual Hold		Select Review Value for Each Date Time by setting each Row OR Select Same Choice for All by Selecting Option Below							
Reviewer Initials MDP	Double click in the appropriate rows in the "Start" and "Stop" Columns to select interval for Shift, Interpolation, or Hold		Data Correction: GapFill Prorated Shift PCA Prorated Shift Actual Hold Actual Interpolated Manual Interpolated		For All Values Use							
Concatenated Name NTS14_ENP_MSH	RecordCount 104				Use Individual Reviews							
date_tm	ActualValue	FilteredValue	PredValue	GapFillPred	datum_change	PrelimValue	RevValue	Manual Entry	Start	Stop	Review Selection	PrelimNote
1/30/2014 21:00	2.70	2.70	2.14	2.69	0	2.70					Prelim Value	Actual Value
1/30/2014 22:00	2.71	2.71	2.15	2.74	0	2.71					Prelim Value	Actual Value
1/30/2014 23:00	2.71	2.71	2.19	2.75	0	2.71			Start		Prelim Value	Actual Value
1/31/2014 0:00			2.23	2.85	0	2.85					Prelim Value	Predicted GAPFILL
1/31/2014 1:00			2.25	2.89	0	2.89					Prelim Value	Predicted GAPFILL
1/31/2014 2:00			2.25	2.89	0	2.89					Prelim Value	Predicted GAPFILL
1/31/2014 3:00			2.26	2.90	0	2.90					Prelim Value	Predicted GAPFILL
1/31/2014 4:00			2.26	2.90	0	2.90					Prelim Value	Predicted GAPFILL
1/31/2014 5:00			2.26	2.90	0	2.90					Prelim Value	Predicted GAPFILL
1/31/2014 6:00			2.26	2.90	0	2.90					Prelim Value	Predicted GAPFILL
1/31/2014 7:00			2.26	2.90	0	2.90					Prelim Value	Predicted GAPFILL
1/31/2014 8:00			2.26	2.91	0	2.91					Prelim Value	Predicted GAPFILL

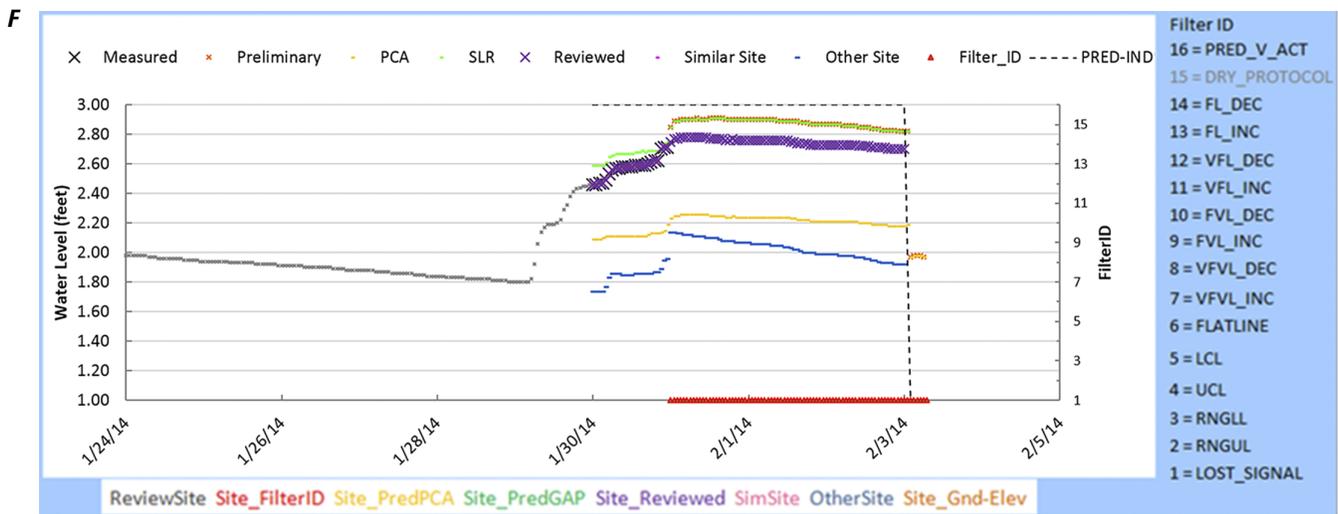
EXPLANATION

- PCA** Multivariate linear regression estimate
- SLR** Simple linear regression estimate
- Reviewed** Previously reviewed data
- Similar Site** Previously reviewed data and current data for a similar station
- Other Site** Previously reviewed data and current data for an additional similar station
- Filter ID** Filter trip identification number
- PRED—IND** Predictor indicator

Figure 13. Automated Data Assurance and Management (ADAM) tool in "Review" mode showing parts of the Review_Tab and Review_Chart worksheets and, A, example of preliminary results, B, prediction detail information, C, manual review procedure using the "Review Selection" dropdown menu, D, manual review procedure using "Data Correction" dropdown menu, E, manual review procedure using "Manual Entry" column and "Review Selection" dropdown menu and, F, graphical results of a manual review.—Continued

E

RunID	779		Data Correction:	Actual Hold	<small>Select Review Value for Each Date Time by setting each Row OR Select Same Choice for All by Selecting Option Below</small>							
Reviewer Initials	MDP		Save Reviewed Data for Selected Site	Run Review Complete - Load Review in Archival Database and Export ISOutput Text File	<small>Double click in the appropriate rows in the "Start" and "Stop" Columns to select interval for Shift, Interpolation, or Hold</small>							
Concatenated Name	NTS14_ENP_MSH		RecordCount	104								
			Perform Data Correction		For All Values Use							
					Use Individual Reviews							
date_tm	ActualValue	FilteredValue	PredValue	GapFillPred	datum_change	PrelimValue	RevValue	Manual Entry	Start	Stop	Review Selection	PrelimNote
1/30/2014 21:00	2.70	2.70	2.14	2.69	0	2.70					Prelim Value	Actual Value
1/30/2014 22:00	2.71	2.71	2.15	2.74	0	2.71					Prelim Value	Actual Value
1/30/2014 23:00	2.71	2.71	2.19	2.75	0	2.71					Prelim Value	Actual Value
1/31/2014 0:00			2.23	2.85	0	2.85		2.81			GAPFILL_with_Shifts	redicted GAPFILL
1/31/2014 1:00			2.25	2.89	0	2.89		2.85			GAPFILL_with_Shifts	redicted GAPFILL
1/31/2014 2:00			2.25	2.89	0	2.89		2.85			Other_Site_with_Shifts	redicted GAPFILL
1/31/2014 3:00			2.26	2.90	0	2.90		2.86			Actual_Interpolation	redicted GAPFILL
1/31/2014 4:00			2.26	2.90	0	2.90		2.86			Hold_Last_Value	redicted GAPFILL
1/31/2014 5:00			2.26	2.90	0	2.90		2.86			Keep Prior Review	redicted GAPFILL
1/31/2014 6:00			2.26	2.90	0	2.90		2.86			Prelim Value	redicted GAPFILL
1/31/2014 7:00			2.26	2.90	0	2.90		2.86			Dry Protocol Value	redicted GAPFILL
1/31/2014 8:00			2.26	2.91	0	2.91		2.87			Linear Interpolation	redicted GAPFILL



EXPLANATION

- PCA** Multivariate linear regression estimate
- SLR** Simple linear regression estimate
- Reviewed** Previously reviewed data
- Similar Site** Previously reviewed data and current data for a similar station
- Other Site** Previously reviewed data and current data for an additional similar station
- Filter ID** Filter trip identification number
- PRED—IND** Predictor indicator

Figure 13. Automated Data Assurance and Management (ADAM) tool in “Review” mode showing parts of the Review_Tab and Review_Chart worksheets and, A, example of preliminary results, B, prediction detail information, C, manual review procedure using the “Review Selection” dropdown menu, D, manual review procedure using “Data Correction” dropdown menu, E, manual review procedure using “Manual Entry” column and “Review Selection” dropdown menu and, F, graphical results of a manual review.—Continued

ADAM tool includes straightforward applications for editing current stations or adding new stations to the ADAM database, modifying station filter settings, assigning SLR equations to the stations, defining MLR settings, running EDEN datasets, and archiving datasets. These applications are accessible from the Microsoft Excel workbook Control worksheet. A general description of ADAM and the methods used to employ these applications are described further in Petkewich and others (2016).

Summary

The Everglades Depth Estimation Network (EDEN) of 247 stations provides hydrologic data for freshwater and tidal areas of the Everglades. Data-quality evaluation and estimation of missing data can be a time-consuming process, especially for a network as large as EDEN. To increase the accuracy of the daily water-level measurements and modeled water surfaces, a Microsoft Excel and Access database tool was created to facilitate quick and accurate quality-assurance review of the real-time data for the EDEN network and to allow estimation or replacement of missing or erroneous data. The ADAM program utilizes inferential sensor technology to effectively and efficiently address data-quality issues by automating many of the processes for quality-assurance checking of data and estimating data when necessary.

The ADAM software uses a sequence of two algorithms for quality-assuring data. The first algorithm provides information about the quality of the data by using a set of 14 univariate filters that identify data that violate user-defined thresholds related to missing, maximum, and minimum values, and value rate of change. The second ADAM algorithm addresses synthesizing estimates to augment actual measurements determined to be erroneous or unreliable using inferential sensors (empirical models) for each station in the network using the subset of good quality data as input. The ADAM software creates inferential sensors for each field sensor using simple linear regression and principal component analysis coupled with multivariate linear regression.

The quality-assurance methods employed using the ADAM software are most effective when automated and manual reviews are conducted on a dataset. The first step is an automated review of the data and the second step is a user-based (manual) review of the automated ADAM review. This review process is user friendly within ADAM, as measured and predicted data for an individual station and the various filter trips are concurrently displayed on a graph and within a table. The ADAM review worksheet allows the reviewer to evaluate hydrographs of individual station data graphed with filter trip information, estimates of the station data, and data for adjacent stations, enabling them to make any necessary adjustments to the preliminary data prior to exporting the results for upload into the EDEN database and use in

the generation of the daily water-surface maps by the EDEN water-surface model. During this manual review, the reviewer can accept the automated results, override any predicted values, adjust the predicted values by shifting the values to better match the measured data, or re-estimate the period of interest by using other means, such as linear interpolation or using data from an adjacent station for estimation. The ADAM database archives the raw data, predicted values, model statistics, filter trips, reviewer's initials, date of review, and all other necessary data associated with the review. Incorporation of ADAM into the daily review process has improved the consistency and utility of the EDEN data. The development and application of inferential sensors is easily transferable to other real-time hydrologic monitoring networks.

References Cited

- Cook, J., Daamen, R.C., Roehl, E.A., Jr., Cho, S., Carlson, K., Byer, D., Byrne, J., and Cline, M., 2008, Distribution system security and water quality improvement through data mining: Report for American Water Works Research Foundation Project no. 3086.
- Jolliffe, I.T., 2002, Principal component analysis (2d ed): Springer Series in Statistics.
- Palaseanu, M., and Pearlstine, L., 2008, Estimation of water surface elevations for the Everglades, Florida: Computers and Geosciences, v. 34, no. 7, p. 815–826, accessed June 2, 2016, at <http://dx.doi.org/10.1016/j.cageo.2007.08.004>.
- 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, Fla., University of Florida, Institute of Food and Agriculture report CIR1521, 18 p., 2 apps.
- Petkewich, M.D., and Conrads, P.A., 2013, Estimation of missing water-level data for the Everglades Depth Estimation Network (EDEN), 2013 update: U.S. Geological Survey Open-File Report 2013–1251, 53 p.
- Petkewich, M.D., Daamen, R.C., Roehl, E.A., and Conrads, P.A., 2016, User's manual for the Automated Data Assurance and Management application developed for quality control of Everglades Depth Estimation Network water-level data: U.S. Geological Survey Open-File Report 2016–1116, 28 p., <http://dx.doi.org/10.3133/ofr20161116>.
- Rencher, A.C., 1998, Multivariate statistical inference and applications: New York, John Wiley & Sons, Inc., 559 p.

- 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 September 3, 2015, at http://sofia.usgs.gov/projects/scopesofwork05/eden_sow05.html.
- 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.
- Telis, P.A., Xie, Zhixiao, Liu, Zhongwei, Li, Yingru, and Conrads, P.A., 2015, The Everglades Depth Estimation Network (EDEN) surface-water model, version 2: U.S. Geological Survey Scientific Investigations Report 2014–5209, 42 p., accessed June 2, 2016, at <http://dx.doi.org/10.3133/sir20145209>.
- U.S. Army Corps of Engineers, 1999, Central and Southern Florida Project, comprehensive review study—Final integrated feasibility report and programmatic environmental impact statement: Jacksonville, Fla., 4 annexes, 15 apps [variously paged].

Manuscript approved June 28, 2016

Publishing support provided by Lafayette and Reston Publishing Service Centers

For additional information regarding this publication, contact:

Director
USGS South Atlantic Water Science Center
Stephenson Center, Suite 129
720 Gracern Road
Columbia, SC 29210-7651
(803) 750-6100
email: dc_sc@usgs.gov

Or visit the USGS South Atlantic Water Science Center Web site at:
<https://www2.usgs.gov/water/southatlantic/>



