Using Inferential Sensors for Quality Control of the Everglades Depth Estimation Network (EDEN)

Matthew D. Petkewich[1], Paul A. Conrads[1], Ruby C. Daamen[2] and Edwin A. Roehl[2] [1]U.S. Geological Survey, Columbia, SC, USA [2]Advanced Data Mining Services, LLC, Greenville, SC, USA

http://www.evergladesplan.org

Background

The Automated Data Assurance and Management (ADAM) software is a Microsoft Excel and Access database® tool[1] created to allow quick and accurate quality assurance review of data and provides estimates or replacements of missing or erroneous data. The ADAM software uses 'inferential sensor' technology similar to that which is often used in industrial applications. Rather than installing a secondary sensor to measure a process, inferential sensors, or virtual sensors, are developed that make very accurate estimates of the process measured by the hard sensor. The advantage of inferential sensors is that they provide redundant signals to sensors in the field without the risk of damage due to the environmental setting. In the event that the sensor does malfunction, ADAM provides an accurate estimate for the period of missing data. The virtual signals can be compared to the measured data and if the difference between the two signals exceeds a predefined tolerance, corrective action can be taken.

The ADAM software was developed for the Everglades Depth Estimation Network (EDEN) of over 290 real-time water-level gages that provide hydrologic data for freshwater and tidal areas of the greater Everglades (fig. 1) (Telis, 2006). A spatially continuous interpolated water surface across the freshwater part of the greater Everglades is generated from daily median water-level values obtained from EDEN (fig. 2). Generation of these daily surfaces is dependent on high-quality data and missing or erroneous data can compromise the quality of the modeled water surfaces. Outliers can easily be detected using minimum and maximum thresholds for each gage, but smaller errors, such as gradual drift of malfunctioning pressure transducers, are more difficult to identify.

[1] *Any use of trademark, product, or firm names is for descriptive purposes and does not imply endorsement by the U.S. Government.*

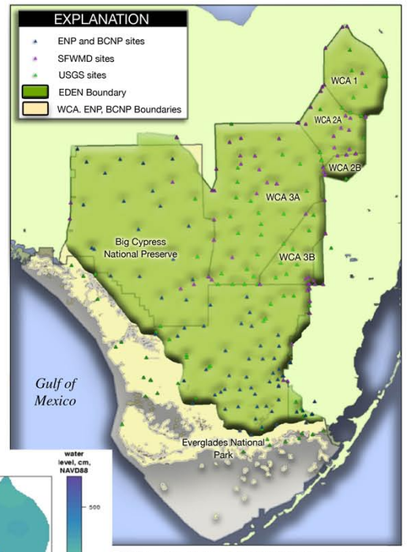

Figure 1. Location of water-level gages in Everglades Depth Estimation Network (EDEN). Water Conservation Areas (WCA) 2 and 3 are subdivided by canals (modified from Pearlstine and others, 2007).

Automated Data Assurance and Management

The ADAM software consists of a Microsoft Excel® workbook (fig. 3) that can be operated in automatic or manual mode to process the data. This workbook is linked to three Microsoft Access® databases that consist of various tables and queries to facilitate loading, processing, and storing the data. The ADAM software uses a two-stage approach for quality assuring the data. The first stage determines the subset of the network data that is good (called filtered data). This is accomplished through a set of 14 univariate filters (table 1) that identify data that violate user-defined thresholds related to maximum, minimum, and rate of change values.

The second stage of ADAM uses the subset of good 'filtered' data as input for developing inferential sensors (empirical models) for each sensor in the network. The ADAM software creates inferential sensors for each field sensor using two types of empirical models: (1) simple linear regression (SLR) and (2) principal component analysis (PCA) coupled with multi-linear regression (MLR). The SLR models are static equations entered into ADAM by the user for each sensor. The PCA/MLR models are dynamic estimates for individual sensors that are calculated during the ADAM analysis using recent filtered data for up to 5 of the most highly correlated sensors to the sensor of interest. The SLR and PCA/MLR predictions also are compared to the measured values for an additional quality assurance check.

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

Filter	
1	No data
2	Data value is greater than maximum historical value
3	Data value is less than minimum historical value
4	Data value is greater than upper control value (90th percentile value)
5	Data value is less than lower control value (10th percentile value)
6	Data has remained unchanged (flat) for a period greater than acceptable
7	Positive rate of change between two measurements exceeds a tolerable rate of change
8	Negative rate of change between two measurements exceeds a tolerable rate of change
9	Positive rate of change over three measurements exceeds a tolerable rate of change
10	Negative rate of change over three measurements exceeds a tolerable rate of change
11	Positive rate of change between two measurements exceeds a rate of change 'warning' threshold
12	Negative rate of change between two measurements exceeds a rate of change 'warning' threshold
13	Positive rate of change over three measurements exceeds a rate of change 'warning' threshold
14	Negative rate of change over three measurements exceeds a rate of change 'warning' threshold

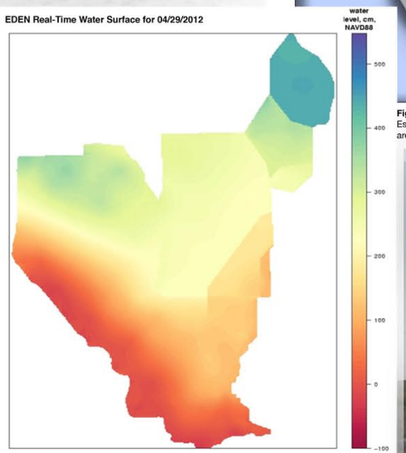

Figure 2. Example of Everglades Depth Estimation Network (EDEN) real-time daily water-surface map (Pearlstine and others, 2007).

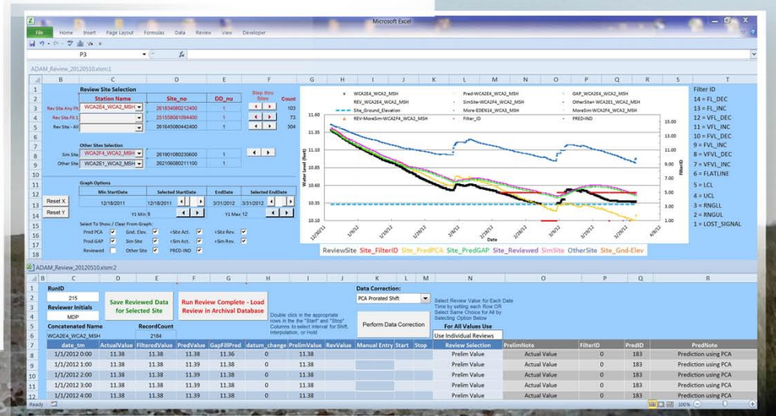

Figure 3. Automated Data Assurance and Management (ADAM) tool in 'Review' mode showing chart and table worksheets.

Using ADAM to Process the EDEN Data

The ADAM software is currently (2012) being used for automated real-time quality assurance of the EDEN data. Analysis of EDEN real-time data using ADAM consists of the following steps as illustrated in figure 4:

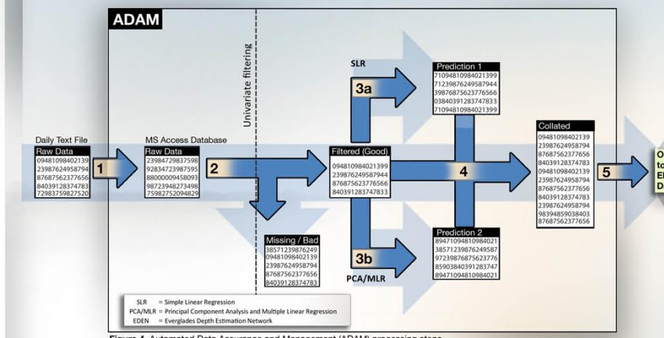

Figure 4. Automated Data Assurance and Management (ADAM) processing steps.

1. Import - raw data from a daily text file is loaded into the Microsoft Access® database.

2. Filter - a filter trip ID (1-14) is automatically assigned to any data point that violates one of fourteen site-specific, user-defined filter values. Questionable data are removed from the filtered data set.

3. Estimate - 2 synthetic sets of hydrographs are created for each gaging station to replace missing or bad raw data. Synthetic data are provided by either (**a**) SLR equations (Conrads and Petkewich, 2009) or (**b**) PCA/MLR (Daamen and others, 2010).

4. Collate - filtered data are collated with either SLR or PCA/MLR predicted data, as needed. The predicted set of data having the highest coefficient of determination (R^2) value is used to replace the missing or bad raw data.

5. Output - once the analysis is complete, the data are exported in an appropriate format for the EDEN database and water-surface model.

The ADAM software also can be operated manually and a hydrologist can over-ride the predicted values, adjust the predicted values, or re-estimate the period of interest by using other means such as linear interpolation or using data from a nearby site. The ADAM database archives the raw data, predicted values, filter trips, reviewer's initials, date of review, and other pertinent information.

Summary

The ADAM software is currently (2012) being used for automated real-time quality assurance of the EDEN data. In addition, ADAM is used to evaluate quarterly and annual data sets. The development and application of inferential sensors is easily transferable to other real-time hydrologic monitoring networks.

Gaging station W11 in Water Conservation Area 3A of the Florida Everglades
Photo by Christa Zweig, Research Assistant, Florida Cooperative Fish and Wildlife Research Unit, University of Florida

References

Conrads, P.A., and Petkewich, M.D., 2009, Estimation of missing water-level data for the Everglades Depth Estimation Network (EDEN): U.S. Geological Survey Open-File Report 2009–1120, 53 p.

Daamen, R.C., Roehl, E., Jr., and Conrads, P.A., 2010 Development of inferential sensors for real-time quality control of water-level data for the Everglades depth estimation network: Proceedings of the 2010 South Carolina Water Resources Conference, held October 13-14, 2010 at the Columbia Metropolitan Convention Center, 4 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.

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. Department of the Interior
U.S. Geological Survey

June 2012