

Everglades Depth Estimation Network (EDEN)

Gap-fill Method for Estimating Missing data
and Review Process

*** Agenda Item #5, Update on Sub-Committee for Dealing with Refuge Stage Data Gaps ***

Eduardo Patino

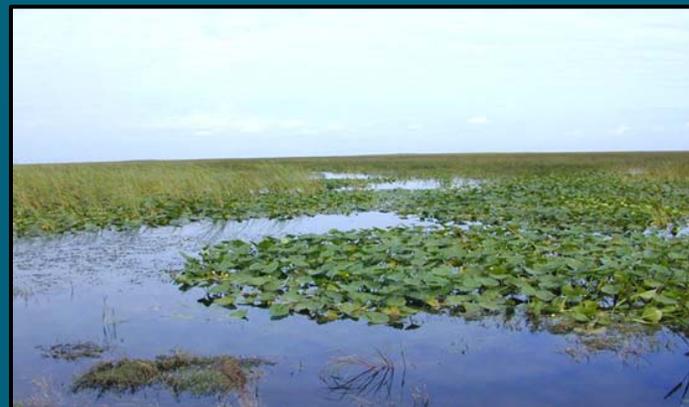
April 26, 2016



<http://sofia.usgs.gov/eden/>

OUTLINE

- ❖ Use of data filters to flag possible erroneous or missing data
- ❖ Use of gap-fill empirical models to estimate missing or erroneous data
- ❖ EDEN Gap-Fill method compared to the Water-Balance method
- ❖ USGS & EDEN review process – real-time, provisional, final
- ❖ Missing data periods for Sites 1-7, 1-8C, & 1-9 (1990 – present)
- ❖ Alternatives moving forward



DATA FILTERS

Filters are used to scan and flag potentially erroneous or missing data.

Filters include several combinations of the following criteria:

- a. Missing values
- b. Exceedance of maximum and minimum value thresholds
- c. Rate of change

As data for any given station is flagged by any of the filters used, it is manually reviewed to determine if in error and if it requires estimation.

*Matthew D. Petkewich and others – **in review**, Using Inferential Sensors for Quality Control of the Everglades Depth Estimation Network Water Level Data.*

GAP-FILL EMPIRICAL MODELS

Gap-fill empirical models were developed using data for the period between 2006 and 2011 and include the following:

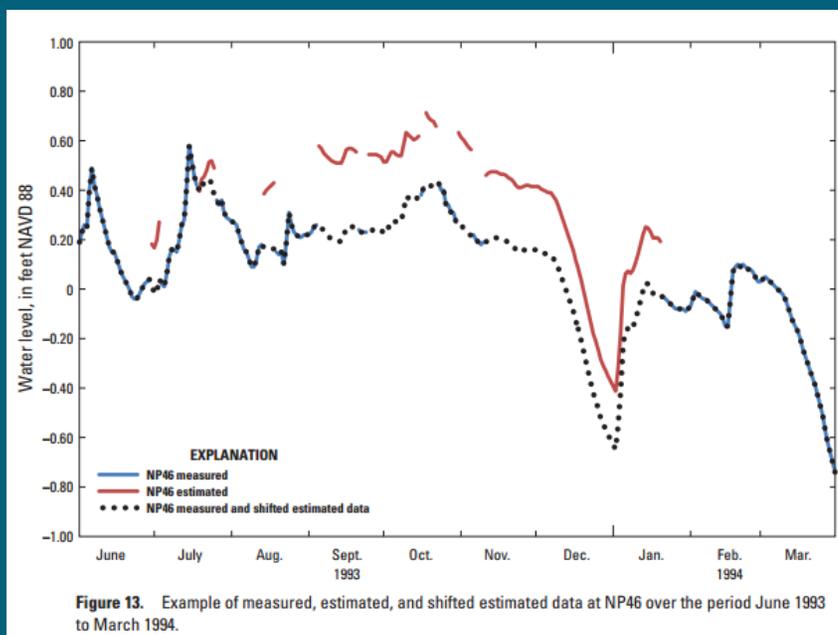
- a. Simple linear regression models
- b. Multivariate linear regression models

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, 45 p.

GAP-FILL EMPIRICAL MODELS

Graph below shows observed data, gap-fill estimates, and shifted final estimated data for NP-46 monitoring station during the period of June 1993 to March 1994.

Gap-fill models capture the general data signature (trend) and shifts applied to the data bring estimates to match observed data prior and after each period of missing data.

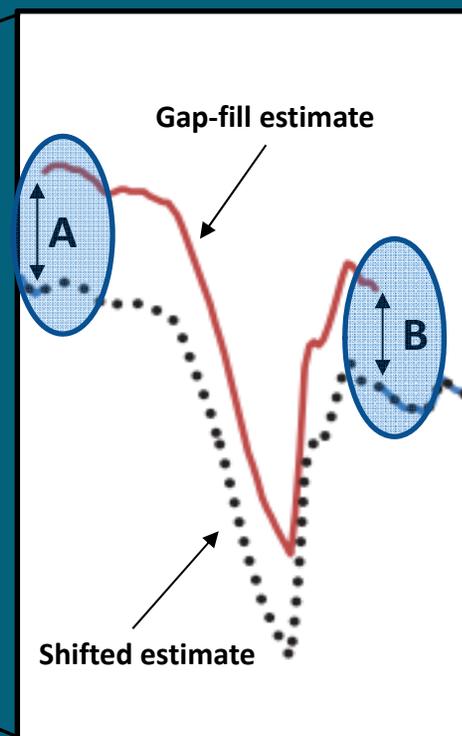
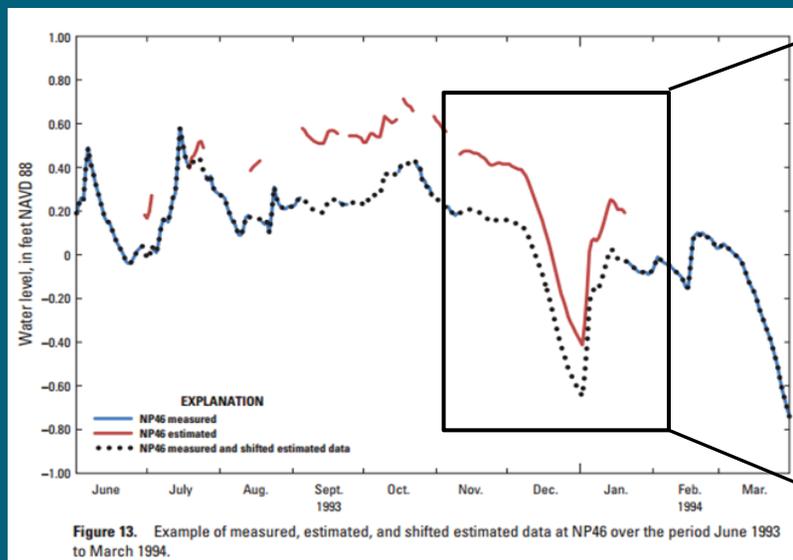


From "Hydrologic Record Extension of Water-Level Data in the Everglades Depth Estimation Network (EDEN), 1991-99. Paul Conrads and others, 2014.

<http://pubs.usgs.gov/sir/2014/5226/pdf/sir2014-5226.pdf>

GAP-FILL EMPIRICAL MODELS

Shifts are calculated from differences between observed data and gap-fill estimates for dates A and B, and prorated in time and magnitude.



GAP-FILL vs WATER-BALANCE, Site 1-7

14-d comparison

Table 4. Statistics of daily error magnitude for each gage (1-7, 1-8C, 1-9) for the period of record (POR: 1/1/99-4/30/15) for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps. [Same data as plotted in Figure 2.]

	EDEN GF			1-d WB			7-d WB			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9
n	5941	5876	5941	5911	5821	5886	5956	5860	5925	5949	5846	5911
Mean	0.09	0.16	0.10	0.01	0.02	0.01	0.02	0.05	0.02	0.03	0.08	0.03
StDev	0.09	0.16	0.10	0.02	0.02	0.02	0.03	0.07	0.03	0.04	0.11	0.04
Skew	2.83	1.58	2.73	3.72	4.69	4.67	3.39	4.07	2.88	3.11	3.40	2.37
1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.03	0.05	0.04	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.01
50%	0.07	0.09	0.09	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.04	0.02
75%	0.12	0.19	0.13	0.01	0.02	0.01	0.03	0.06	0.03	0.04	0.10	0.04
99%	0.50	0.64	0.49	0.10	0.12	0.10	0.15	0.35	0.14	0.22	0.55	0.18

Shifted

	EDEN GF - prorated		
	1-7		
n	5904	count data:	5964
Mean	0.04		
StDev	0.05		
Skew	3.00		
1%	0.00		
25%	0.01		
50%	0.02		
75%	0.05		
99%	0.22		

Table 5. Comparison of median daily error magnitude for each gage (1-7, 1-8C, 1-9) for two periods of record for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps.

Period	EDEN GF			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9
1/1/99-4/30/15	0.07	0.09	0.09	0.02	0.04	0.02
3/1/06-9/30/11	0.07	0.06	0.10	0.02	0.04	0.02

Gapfill - pro-rated shift
Median, all dates
0.02
Median, 3/1/06 - 9/30/11
0.02

GAP-FILL vs WATER-BALANCE, Site 1-8C

14-d comparison

Table 4. Statistics of daily error magnitude for each gage (1-7, 1-8C, 1-9) for the period of record (POR: 1/1/99-4/30/15) for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps. [Same data as plotted in Figure 2.]

	EDEN GF			1-d WB			7-d WB			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9
n	5941	5876	5941	5911	5821	5886	5956	5860	5925	5949	5846	5911
Mean	0.09	0.16	0.10	0.01	0.02	0.01	0.02	0.05	0.02	0.03	0.08	0.03
StDev	0.09	0.16	0.10	0.02	0.02	0.02	0.03	0.07	0.03	0.04	0.11	0.04
Skew	2.83	1.58	2.73	3.72	4.69	4.67	3.39	4.07	2.88	3.11	3.40	2.37
1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.03	0.05	0.04	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.01
50%	0.07	0.09	0.09	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.04	0.02
75%	0.12	0.19	0.13	0.01	0.02	0.01	0.03	0.06	0.03	0.04	0.10	0.04
99%	0.50	0.64	0.49	0.10	0.12	0.10	0.15	0.35	0.14	0.22	0.55	0.18

Shifted

	EDEN GF - prorated		
	1-8C		
n	5801	count data:	5876
Mean	0.05		
StDev	0.07		
Skew	4.37		
1%	0.00		
25%	0.01		
50%	0.03		
75%	0.06		
99%	0.33		

Table 5. Comparison of median daily error magnitude for each gage (1-7, 1-8C, 1-9) for two periods of record for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps.

Period	EDEN GF			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9
1/1/99-4/30/15	0.07	0.09	0.09	0.02	0.04	0.02
3/1/06-9/30/11	0.07	0.06	0.10	0.02	0.04	0.02

Gapfill - pro-rated shift
Median, all dates
0.03
Median, 3/1/06 - 9/30/11
0.03

GAP-FILL vs WATER-BALANCE, Site 1-9

14-d comparison

Table 4. Statistics of daily error magnitude for each gage (1-7, 1-8C, 1-9) for the period of record (POR: 1/1/99-4/30/15) for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps. [Same data as plotted in Figure 2.]

	EDEN GF			1-d WB			7-d WB			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9	1-7	1-8C	1-9
n	5941	5876	5941	5911	5821	5886	5956	5860	5925	5949	5846	5911
Mean	0.09	0.16	0.10	0.01	0.02	0.01	0.02	0.05	0.02	0.03	0.08	0.03
StDev	0.09	0.16	0.10	0.02	0.02	0.02	0.03	0.07	0.03	0.04	0.11	0.04
Skew	2.83	1.58	2.73	3.72	4.69	4.67	3.39	4.07	2.88	3.11	3.40	2.37
1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.03	0.05	0.04	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.01
50%	0.07	0.09	0.09	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.04	0.02
75%	0.12	0.19	0.13	0.01	0.02	0.01	0.03	0.06	0.03	0.04	0.10	0.04
99%	0.50	0.64	0.49	0.10	0.12	0.10	0.15	0.35	0.14	0.22	0.55	0.18

Shifted

	EDEN GF - prorated		
	1-9		
n	5904	count data:	5941
Mean	0.04		
StDev	0.05		
Skew	2.98		
1%	0.00		
25%	0.01		
50%	0.02		
75%	0.05		
99%	0.23		

Table 5. Comparison of median daily error magnitude for each gage (1-7, 1-8C, 1-9) for two periods of record for EDEN Gap-fill (GF) method and Water-balance (WB) method with 1-, 7-, and 14-day data gaps.

Period	EDEN GF			14-d WB		
	1-7	1-8C	1-9	1-7	1-8C	1-9
1/1/99-4/30/15	0.07	0.09	0.09	0.02	0.04	0.02
3/1/06-9/30/11	0.07	0.06	0.10	0.02	0.04	0.02

Gapfill - pro-rated shift
Median, all dates
0.02
Median, 3/1/06 - 9/30/11
0.02

USGS & EDEN REVIEW PROCESS

USGS data review process:

- a. **Real-time** – raw data as received through satellite transmissions (hourly)
- b. **Provisional** – not to exceed 90 days after every field visit – corrections may be applied.
- c. **Final** – Water-year data (Oct. – Sept.) is **finalized by April 1** of the following year, but current efforts are moving towards Continuous Records Processing (CRP) and not to exceed 150 days after every field visit – final corrections applied.

*** No estimated daily water-levels other than for days with partial record ***

USGS & EDEN REVIEW PROCESS... continued

EDEN data review process:

- a. **Real-time** – raw data as received daily from SFWMD, NPS, and USGS, and is posted on EDENWeb.
- b. **Provisional** – quarterly, posted about 45 days after the end of each quarter – replaces real-time data on EDENWeb.
- c. **Final** – yearly, Oct. to Sept. of previous year, posted in October of the following year – replaces provisional data on EDENWeb.

*** Estimated water-levels using gap-fill equations and shifts ***

USGS & EDEN REVIEW PROCESS... continued

It is USGS policy not to approve (finalize) data until after a field inspection has been made at the monitoring station of interest and the data reviewed.

Currently field visits at Site 1-7, Site 1-8C, and Site 1-9 are scheduled for twice per year (6 months), with emergency visits as problems arise at stations.

Missing data periods for Site 1-7, 1-8C, & 1-9

1990 - present

Site 1-7

1-5 days =	16
5-10 days =	2
10-15 days =	4
15-30 days =	3
Over 30 days =	1
-- 2006 --	

Site 1-8C

1-5 days =	20
5-10 days =	1
10-15 days =	0
15-30 days =	0
Over 30 days =	1
-- 2005 --	

Site 1-9

1-5 days =	22
5-10 days =	6
10-15 days =	3
15-30 days =	4
Over 30 days =	0

ALTERNATIVES MOVING FORWARD

- ❖ Reduce the time-frame for approved final data – would require more frequent field visits to sites of interest
- ❖ Reduce the amount of missing data – could install back-up sensors at sites of interest
- ❖ Other alternatives?

QUESTIONS?

