Development of Inferential Sensors for Real-time Quality Control of Water-level Data for the EDEN Network

Ruby Daamen, Advanced Data Mining Int'l
Edwin Roehl, Advanced Data Mining Int'l
Paul Conrads, USGS - SC Water Science Center

GEER 2010 – Naples, FL
July 16, 2010

Presentation Outline

- What is a “Inferential Sensor”?
- Background
  - Industrial application
- Development issues
- Inferential Sensor (IS) development for the EDEN Network
- EDENIS Prototype

Tough Environment to Monitor

- **Emissions regulations require measurements of effluent gases**
- **Smoke stack burns up probes**
- **Need alternative to “hard” sensors**

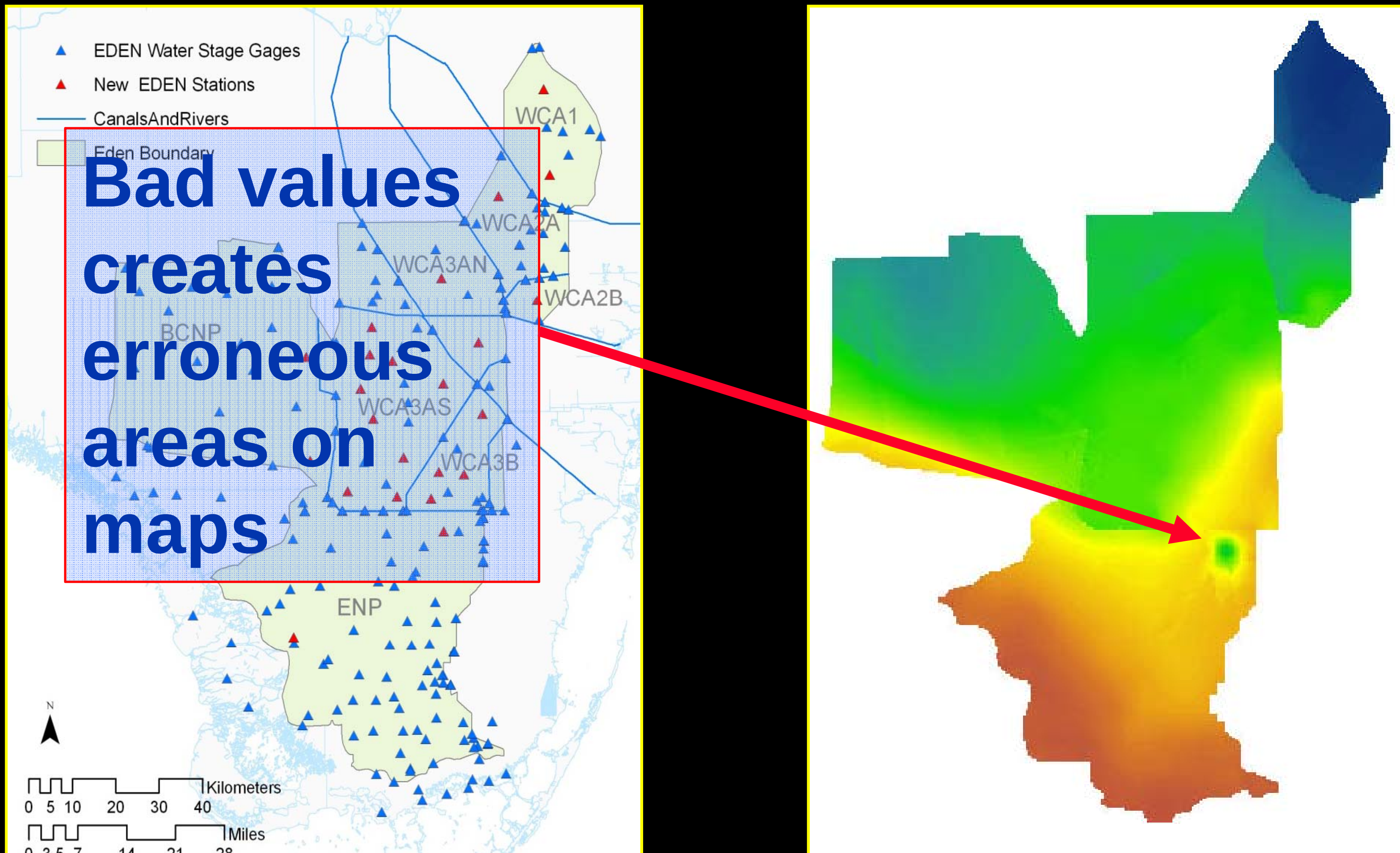
Hard Sensor vs. Inferential Sensor

- Virtual sensor replaces actual sensor
  - Temporary gage smoke stack
  - Operate plant to cover range of emissions
  - Develop model of emissions based on operations
  - Model becomes the “Inferential Sensor’

If it is good enough for Industry...

- Use similar approach for real-time data
- Develop models to predict real-time data
- Use predictions as “inferential-sensor” to:
  - QA/QC hard sensor
  - Provide accurate estimates for hard sensor
  - Provide redundant signal

EDEN Water-Surface Map

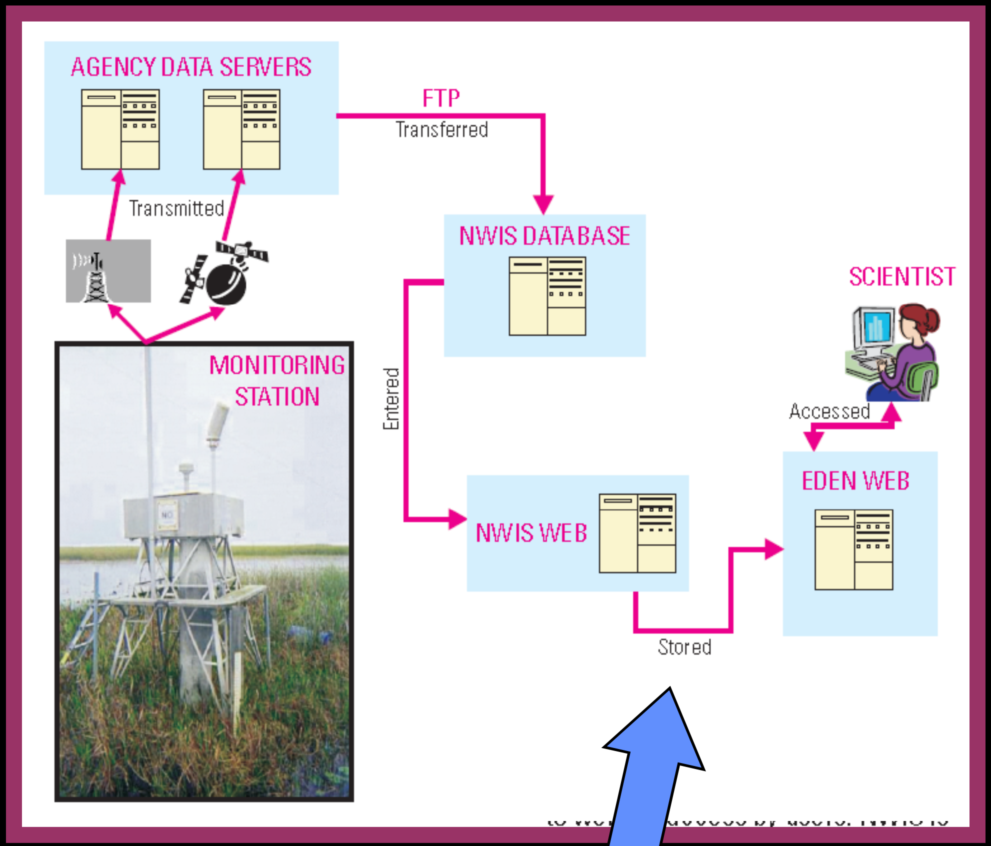


Problem

Need to minimize missing and erroneous data

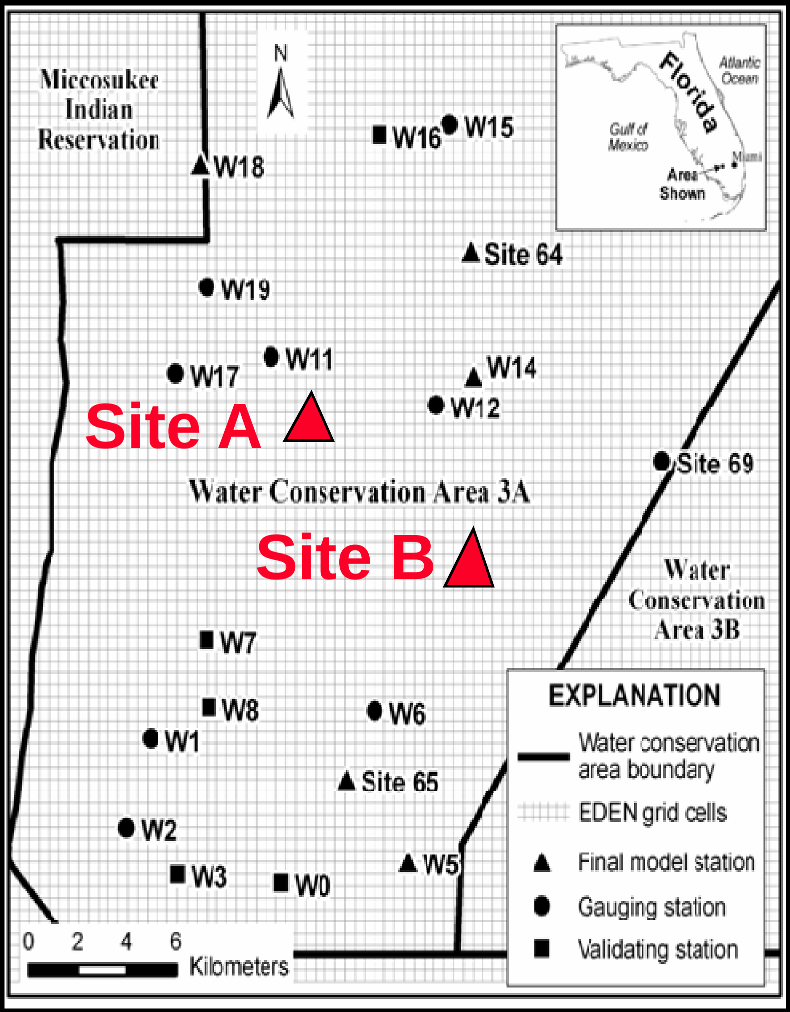
Approach

Develop "inferential" sensors for redundant signal



Inferential Sensor application

Challenges:
Hypothetical Case

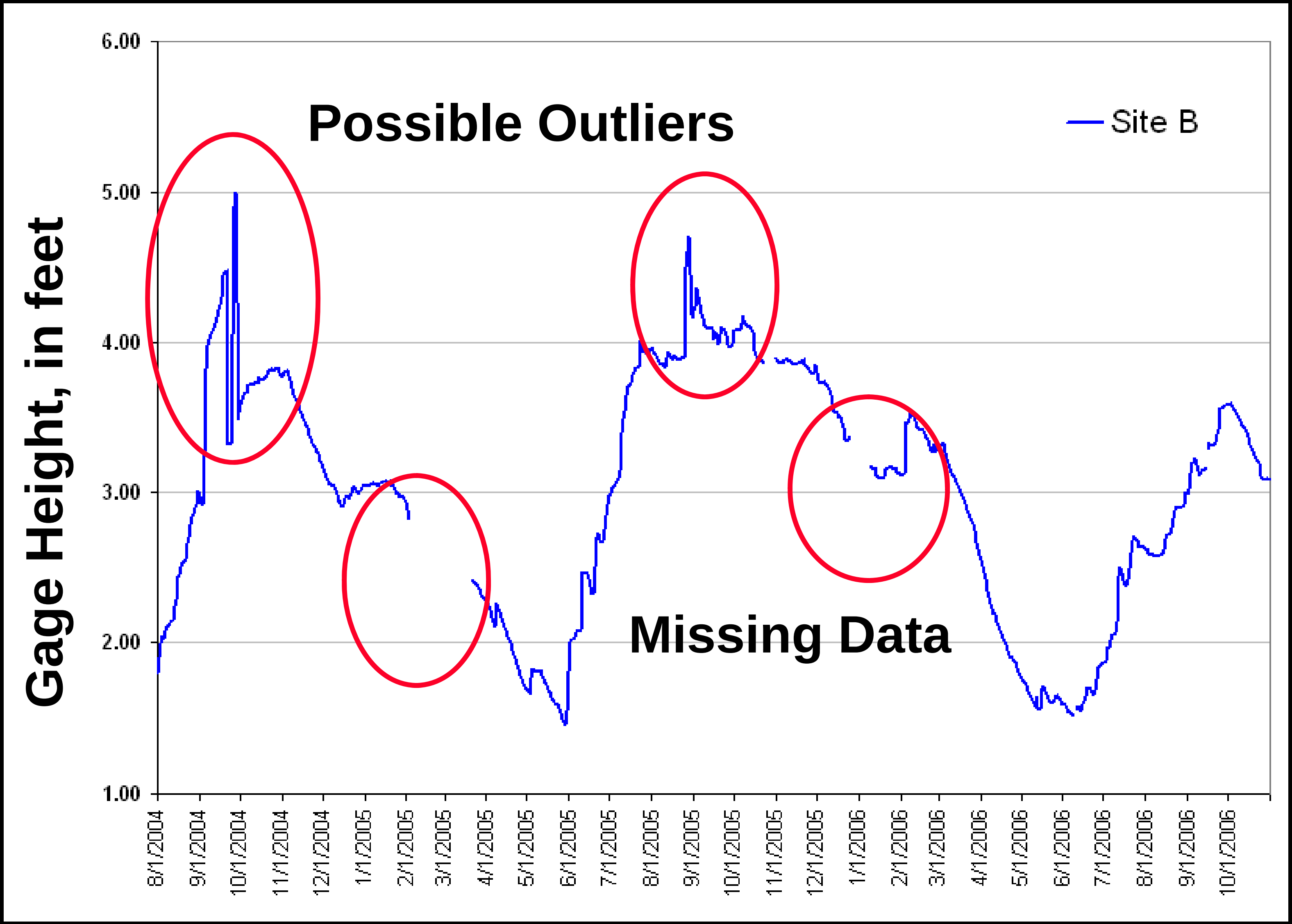


Create model (Inferential Sensor) for Site B using Site A as an input

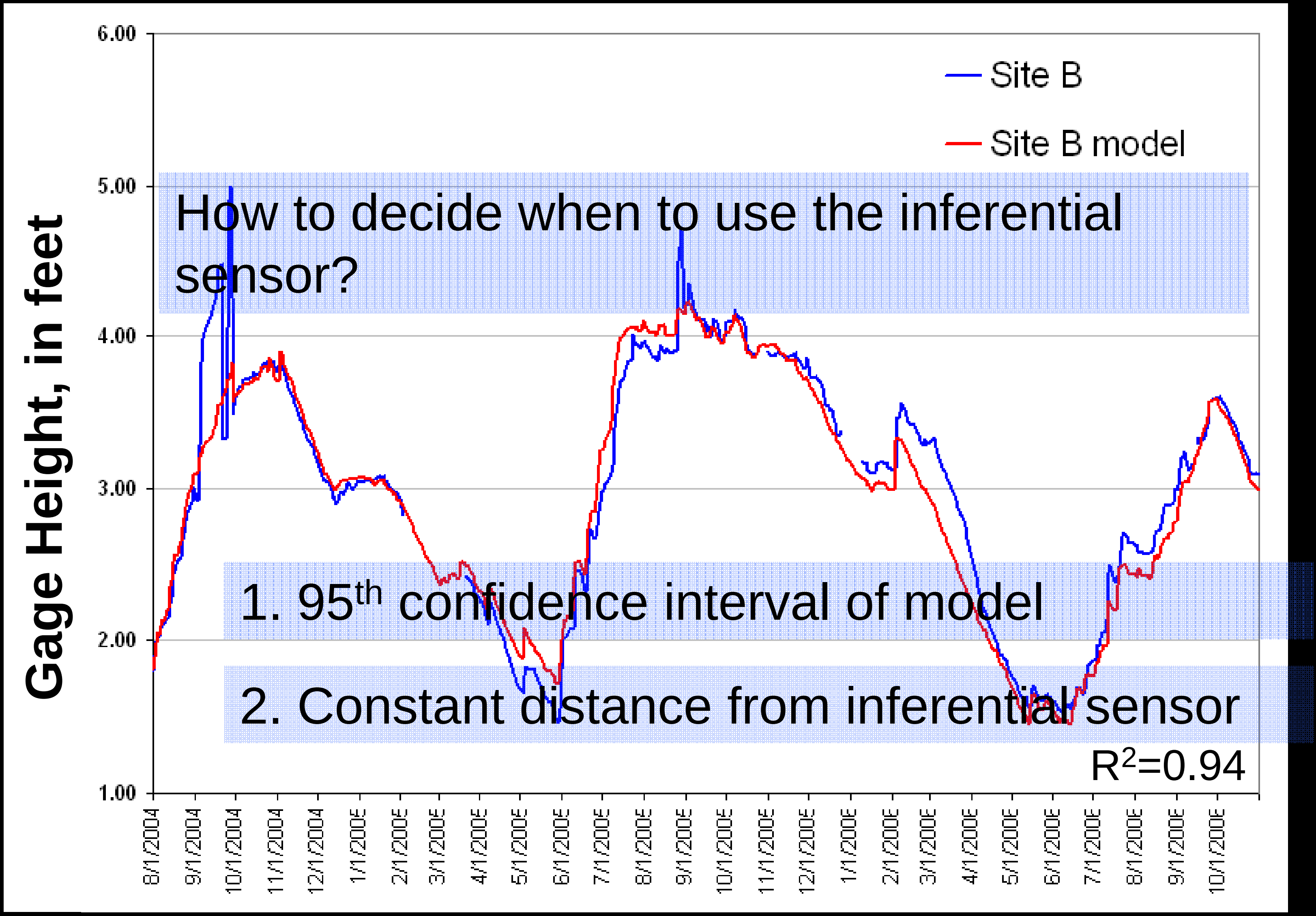
Decide when to use Inferential Sensor instead of gage data

Actual application would be for 253 stations

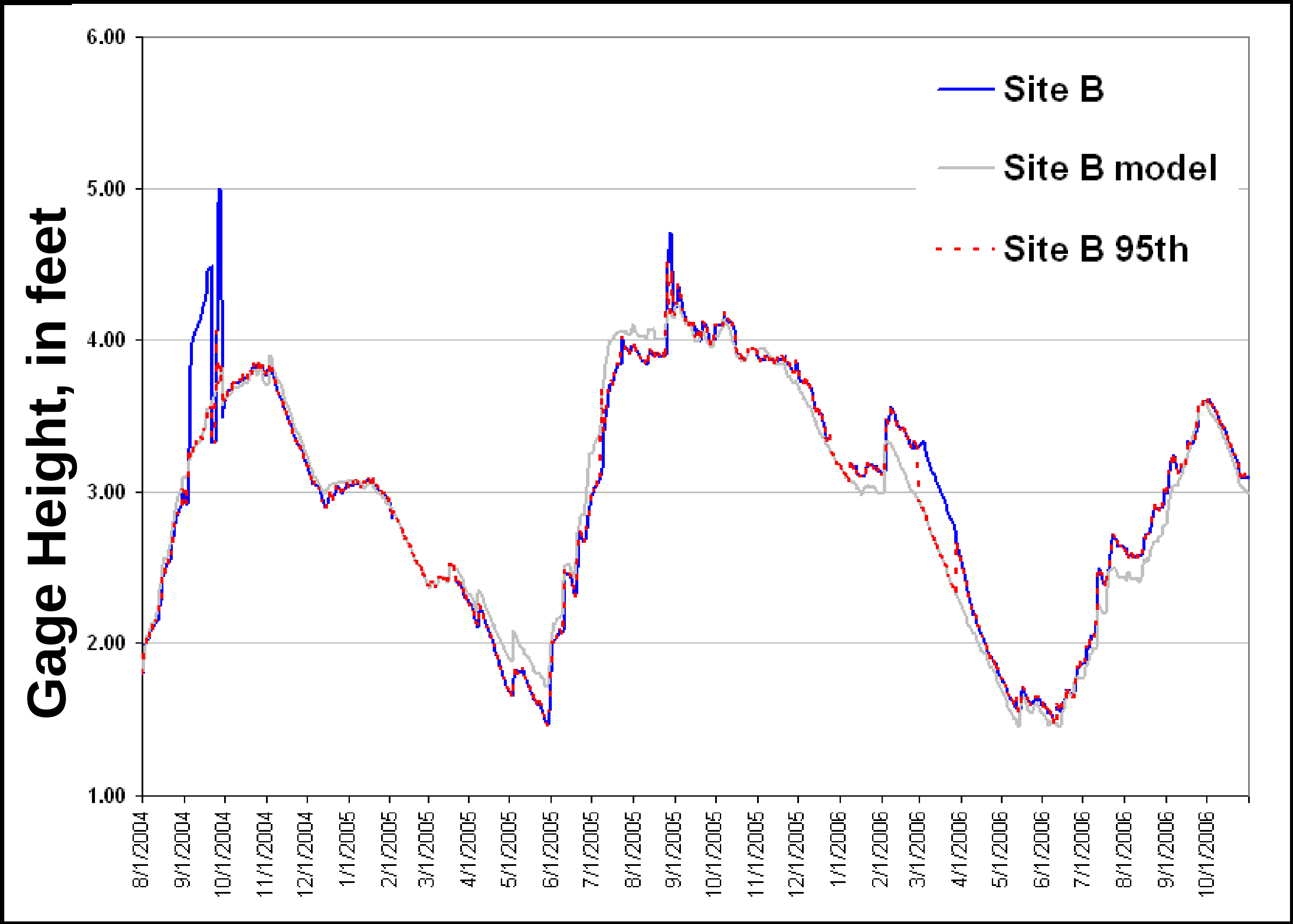
Hypothetical Case: Gage Data



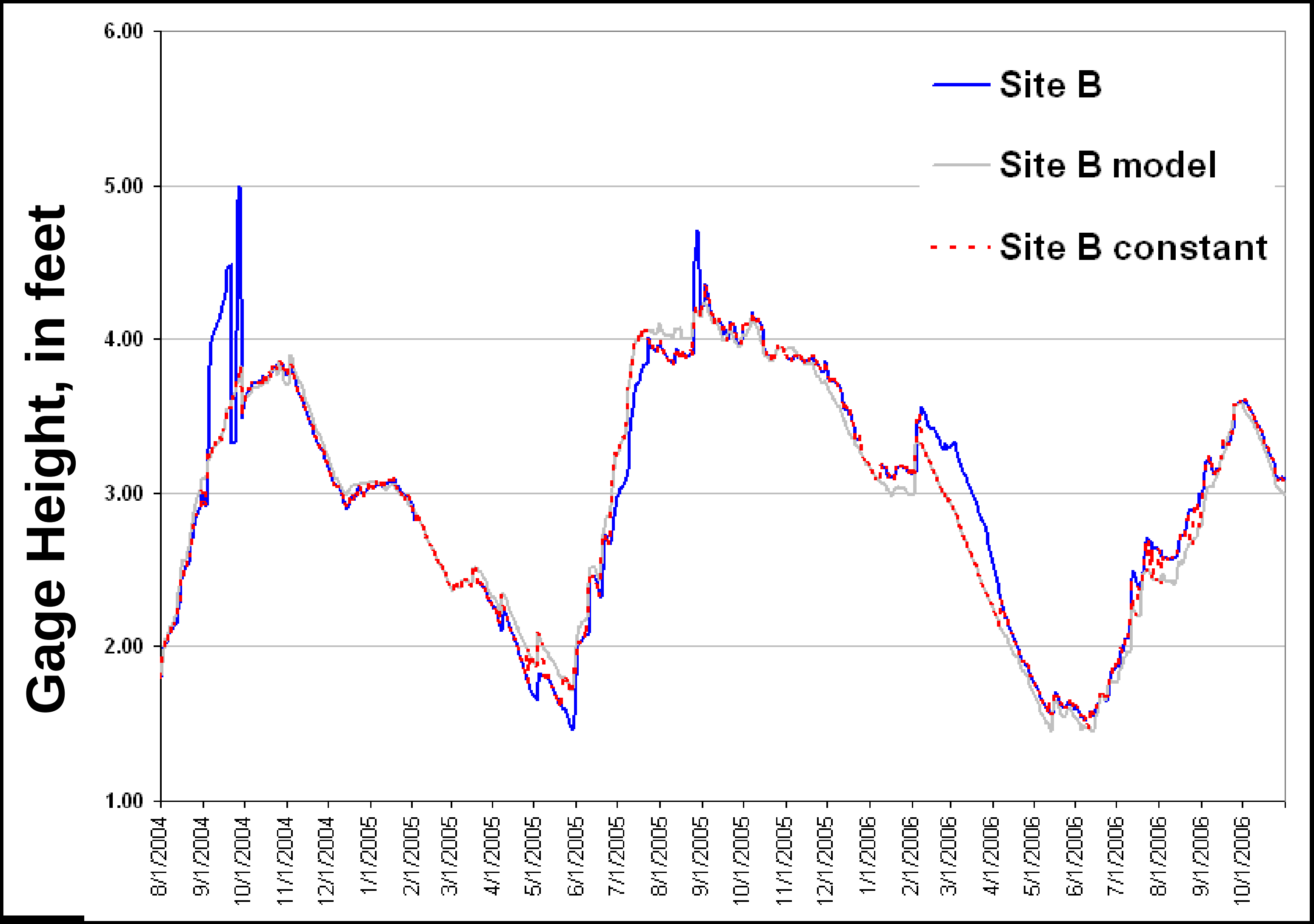
Hypothetical Case: Inferential Sensor



Hypothetical Case:
When to use Inferential Sensor?



Hypothetical Case: When to use Inferential Sensor?



Hypothetical Case - comments

- **Issue of model accuracy**
- **Immediate benefit for missing data**
- **Made the assumption that Site A was correct**
- **What if data for Site A is missing?**
- ***Issues are magnified when dealing with a network of 253 gages***

EDEN Inferential Sensor Development: *3 Phase Project*

- **Phase 1**
  - **Develop application to determine good data**
- **Phase 2**
  - **Develop inferential sensor models**
- **Phase 3**
  - **Integrate inferential sensors into daily real-time EDEN data stream**

Phase 1: Subset of Good Data

- **Create a series of filters to test data**
  - **High & low thresholds**
  - **High & low rate-of-change thresholds**

UNIVARIATE FILTER	CHECK DESCRIPTION	PRECEDENCE	WATER LEVEL LIMIT (ft.)
LOST_SIGNAL	no signal	1	NA
GT_RNG_UL	x(t) > signal range Upper Range Limit	2	15.19
LT_RNG_LL	x(t) < signal range Upper Range Limit	3	6.99
GT_UCL	x(t) > signal Upper Control Limit	4	14.73
LT_LCL	x(t) < signal Upper Control Limit	5	8.56
Sn_LT_L	flatlined: x'(t) = x(t)=x(t-1); SUM[(|x'(t)|,…,|x'(t-n+1)|] < Limit	6	0.00
D1_GT_L_1	vfast vlarge increase: x(t)-x(t-1) > Limit	7	1.92
D1_LT_L_1	vfast vlarge decrease: x(t)-x(t-1) < Limit	8	-2.34
D1Sn_GT_L_1	fast vlarge increase: x'(t)=x(t)-x(t-1); Sum[x'(t),…x'(t-n+1)] > Limit	9	1.98
D1Sn_LT_L_1	fast vlarge decrease: x'(t)=x(t)-x(t-1); Sum[x'(t),…x'(t-n+1)] < Limit	10	-2.52
D1_GT_L_2	vfast large increase: x(t) - x(t-1) > Limit	11	1.69
D1_LT_L_2	vfast large decrease: x(t) - x(t-1)< Limit	12	-0.25
D1Sn_GT_L_2	fast large increase: x'(t)=x(t)-x(t-1); Sum[x'(t),…x'(t-n+1)] > Limit	13	1.98
D1Sn_LT_L_2	fast large decrease: x'(t)=x(t)-x(t-1); Sum[x'(t),…x'(t-n+1)] < Limit	14	-0.27

Phase 2: Inferential Sensor Development
Model Development

- One Approach – Canned Models
  - Create multiple models for a gage
  - Set priority for model to use depending on available data
  - Large number of models
  - Not all combinations of gages would be addressed

Phase 2: Inferential Sensor Development
Model Development

- Second Approach – Model on the Fly
  - Develop models based on available data
  - Address all combinations of gages
  - Issue of correlation of multiple inputs
  - More complex programming than canned equations

Modeling on the Fly

- 12 Sites selected for algorithm development
- Correlations computed for previous 90 days
- Top five correlated stations determined for each site
- Multi-variate regression equations (1-5 stations) computed on the fly

Still problem of correlated inputs

	Rank (Out of 260)	Largest R (Pearson)	10th Largest R
G211T_ENP_CSTR	260	0.3035	0.232
S140H_WCA3_CSTR	257	0.7504	0.6807
S190H_BCNP_CSTR	247	0.8754	0.0081
W15_WCA3_MSH	173	0.9676	0.9204
CR2_ENP_MSH	135	0.9803	0.937469395
E146_ENP_MSH	120	0.9835	0.936
MUD_FBAY_RVR	115	0.9843	0.8423
R3110_ENP_MSH	58	0.9946	0.9125
G119T_PENN_CSTR	35	0.998	0.749
L31N1_ENP_CNL	25	0.9987	0.8001
S10DT_WCA2_MSTR	19	0.9987	0.9551
S343AH_WCA3_MSTR	11	0.999	0.95

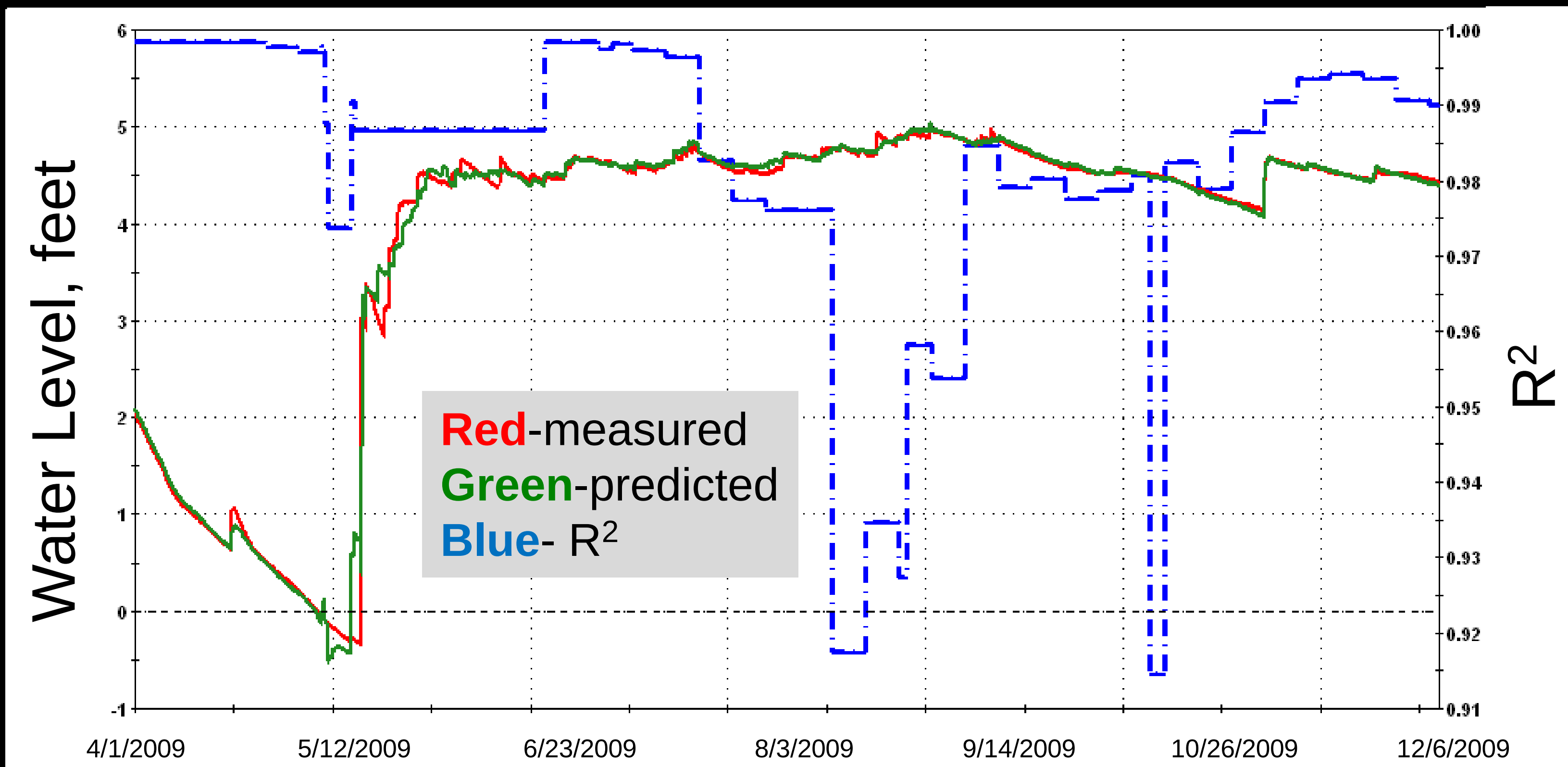
Principal Component Analysis

- Definition: “A mathematical procedure that transforms a set of correlated variables into a smaller number of uncorrelated variables Generate an array of input data”
  - Use the Jacobi eigenvalue algorithm to calculate the eigenvalues and eigenvectors of the covariance matrix.
  - Calculate the Principal Components (PC)
  - PCs are decorrelated from each other
  - Compute multivariate linear regression to predict gage measurement

Example: CR2_ENP_MSH

PCA5_3PC

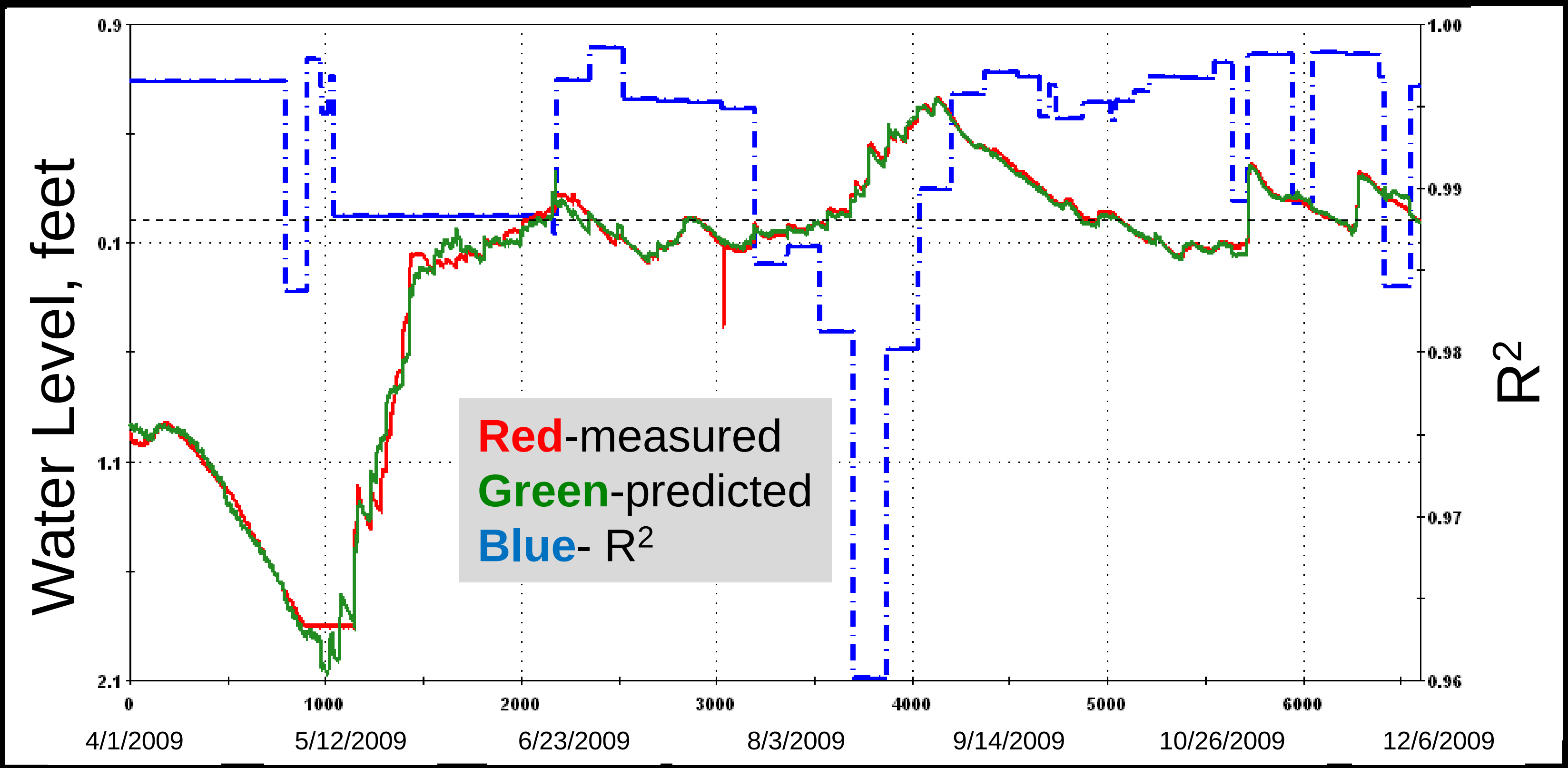
Overall R^2: 0.9953
Max 90 Day R^2: 0.9985
Min 90 Day R^2: 0.9146
RMSE: 0.0762



Example: E146_ENP_MSH

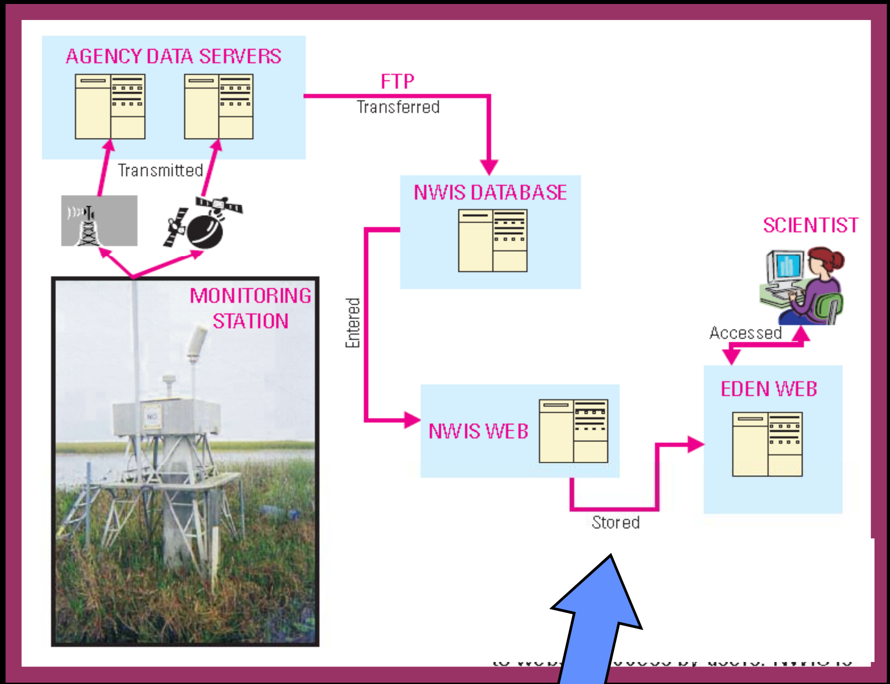
PCA5_5PC

Overall R^2: 0.9976
Max 90 Day R^2: 0.9986
Min 90 Day R^2: 0.9602
RMSE: 0.0461



Phase 3: Operational Application

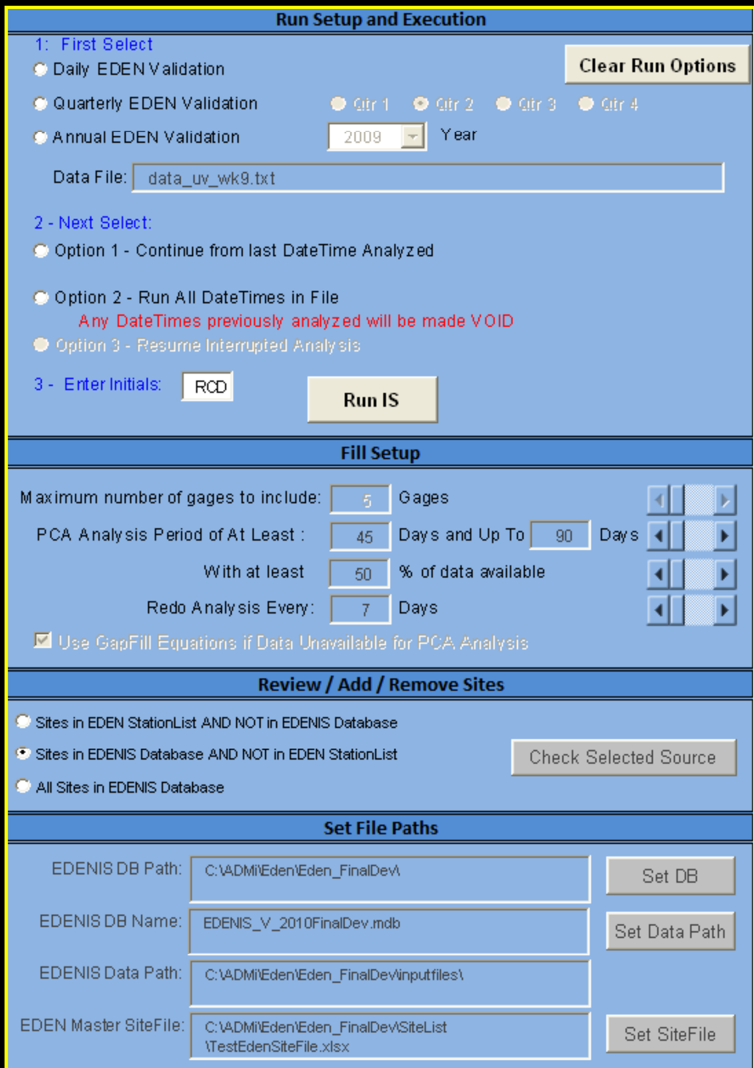
- **Insert inferential sensors in EDEN data stream**
- **Develop application for daily EDEN operations**
- **Provide digital record of daily QA/QC process**



Inferential Sensor application

EDENIS – EDEN Inferential Sensor

Prototype - *Control Worksheet*



Select Daily, Quarterly or Annual Run Analysis

Choose to date to begin analysis

Fill Setup including

Remove , add, or edit sites included in EDENIS

Set Pathnames for files used by EDENIS

EDENIS
Control Worksheet

Status of Current Run

Current DateTime:	5/27/2009 0:00	Run Complete
Run Start DateTime:	5/20/2009 0:15	
Run Stop DateTime:	5/27/2009 0:00	

Run Duration

Run Start:	15:45:44	hh:mm:ss
Run Stop:	16:30:51	hh:mm:ss
Run Dur:	00:45:07	hh:mm:ss

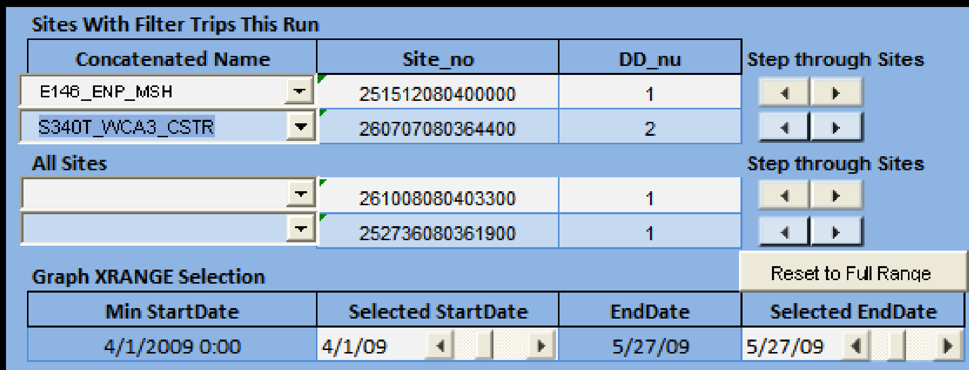
Review Selected Run

Select table row from run list and before selecting "Review Selected Run"

RunID	RunStartDatetime	FileEndDateTime	LastAnalyzed	AnalysisComplete	Type	Initials	DataQtr	DataYear	RunDate
9	5/20/2009 0:15	5/27/2009 0:00	5/27/2009 0:00	TRUE	daily	RCD			7/14/2010
7	5/13/2009 0:15	5/20/2009 0:00	5/20/2009 0:00	TRUE	daily	RCD			7/14/2010
6	5/6/2009 0:15	5/13/2009 0:00	5/13/2009 0:00	TRUE	daily	RCD			7/14/2010
5	4/29/2009 0:15	5/6/2009 0:00	5/6/2009 0:00	TRUE	daily	RCD			7/14/2010
4	4/22/2009 0:15	4/29/2009 0:00	4/29/2009 0:00	TRUE	daily	RCD			7/14/2010
3	4/15/2009 0:15	4/22/2009 0:00	4/22/2009 0:00	TRUE	daily	RCD			7/14/2010
2	4/8/2009 0:15	4/15/2009 0:00	4/15/2009 0:00	TRUE	daily	RCD			7/14/2010
1	4/1/2009 0:00	4/8/2009 0:00	4/8/2009 0:00	TRUE	daily	RCD			7/14/2010

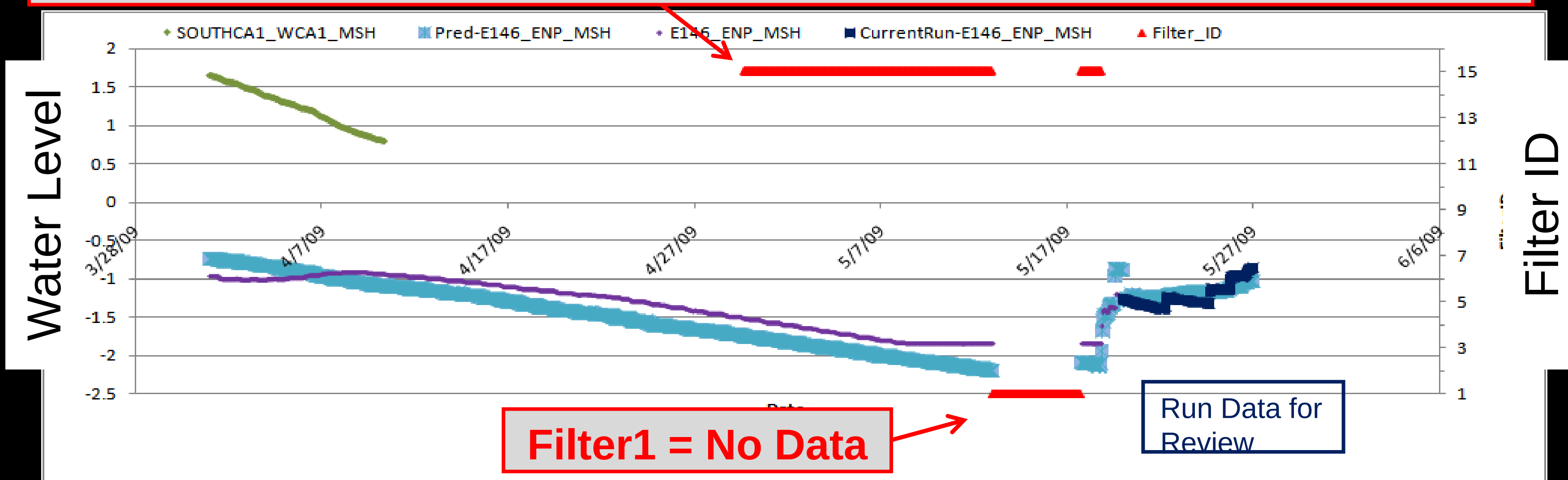
EDENIS

Site_Review Graphics



Select Review Site

Filter15 = Dry Protocol Filter
Measured value is at or below the user set Dry Protocol



EDENIS
Tabular data for review

Concatenated Name	Currently Reviewing	Out of Total Of				Save Reviewed Data	For All Values Use
E146_ENP_MSH	1 - 24	168					Use Predicted Value
date_tm	**ActualValue**	**FilteredValue**	**PredValue**	**RevValue**	**Suggested Value**	**Manual Entry**	**Review Action - Use**
5/20/2009 1:00	-1.26	-1.26	-1.28		-1.26		Actual Value
5/20/2009 2:00	-1.27	-1.27	-1.28		-1.27		Actual Value
5/20/2009 3:00	-1.27	-1.27	-1.27		-1.27		Actual Value
5/20/2009 4:00	-1.28	-1.28	-1.27		-1.28		Actual Value
5/20/2009 5:00	-1.28	-1.28	-1.27		-1.28		Actual Value
5/20/2009 6:00	-1.28	-1.28	-1.26		-1.28		Actual Value
5/20/2009 7:00	-1.29	-1.29	-1.26		-1.29		
5/20/2009 8:00	-1.29	-1.29	-1.26		-1.29		
5/20/2009 9:00	-1.30	-1.30	-1.26		-1.3		
5/20/2009 10:00	-1.30	-1.30	-1.26		-1.3		
5/20/2009 11:00	-1.30	-1.30	-1.23		-1.3		
5/20/2009 12:00	-1.30	-1.30	-1.21		-1.3		
5/20/2009 13:00	-1.31	-1.31	-1.21		-1.31		
5/20/2009 14:00	-1.31	-1.31	-1.23		-1.31		
5/20/2009 15:00	-1.31	-1.31	-1.22		-1.31		
5/20/2009 16:00	-1.32	-1.32	-1.23		-1.32		
5/20/2009 17:00	-1.32	-1.32	-1.24		-1.32		
5/20/2009 18:00	-1.32	-1.32	-1.24		-1.32		
5/20/2009 19:00	-1.33	-1.33	-1.24		-1.33		
5/20/2009 20:00	-1.33	-1.33	-1.25		-1.33		
5/20/2009 21:00	-1.33	-1.33	-1.25		-1.33		
5/20/2009 22:00	-1.34	-1.34	-1.26		-1.34		
5/20/2009 23:00	-1.34	-1.34	-1.26		-1.34		
5/20/2009 0:00	-1.34	-1.34	-1.26		-1.34		Actual Value

Prediction Details

Site	E146_ENP_MSH
Prediction Source:	Prediction using PCA
R2	0.941505765052831
Pearson	
SimSite1	S340T_WCA3_CSTR
SimSite2	S9AT_WCA3_MSTR
SimSite3	S142T_WCA3_MSTR
SimSite4	S142H_WCA2_CSTR
SimSite5	S151H_WCA3_CSTR

Conclusions

- **Inferential Sensor provide:**
  - **Real-time QA/QC**
  - **Redundant signal**
  - **Models developed on the fly**
- **Project scheduled for completion in September**