Development of Empirical Hydrologic and Water Quality Models of the Loxahatchee NWF Using Data-Mining Techniques

Paul Conrads

USGS- SC Water Science Center

Ed Roehl, Jr

Advanced Data Mining International, LLP

Greater Everglade Ecosystem Restoration Conference

July 13, 2010

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

Outline

- Data mining & Data driven models
- Modeling Loxahatchee NWR:
  - Water levels
  - Specific conductance
  - Total phosphorus
- LOXANN Decision Support System (DSS)
- DSS applications
  - Evaluation of flow releases

USGS

What is Data Mining?

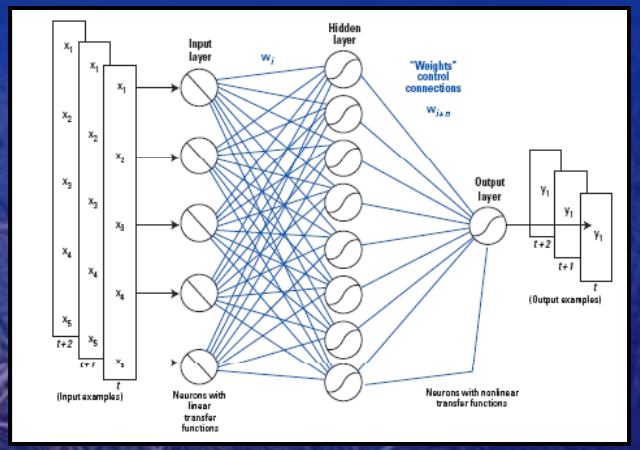
- *Data Mining*: **the search for valuable knowledge in massive volumes of data**
- *An amalgamation of techniques from various disciplines*
- *Data Mining Tool Box*
  - signal processing, statistics, machine learning, chaos theory, advanced visualization
  - Artificial neural networks (ANN) models – one approach to machine learning

Data ⇨ Information ⇨ Knowledge

USGS

Data Driven Models

- Living in an era of “Big Data”
- Modeling – exercise in mapping inputs and outputs
- Empirical models - based on observations rather than on mathematically describable system processes
- Examples:
  - Linear regression: Y = mX + b
  - Artificial Neural Networks:

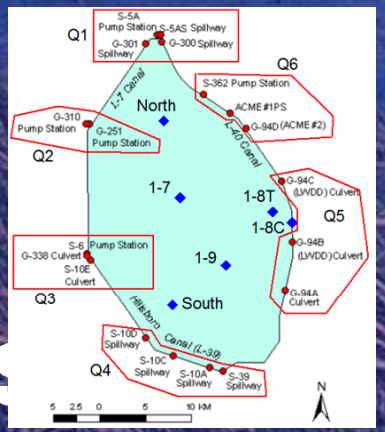


USGS

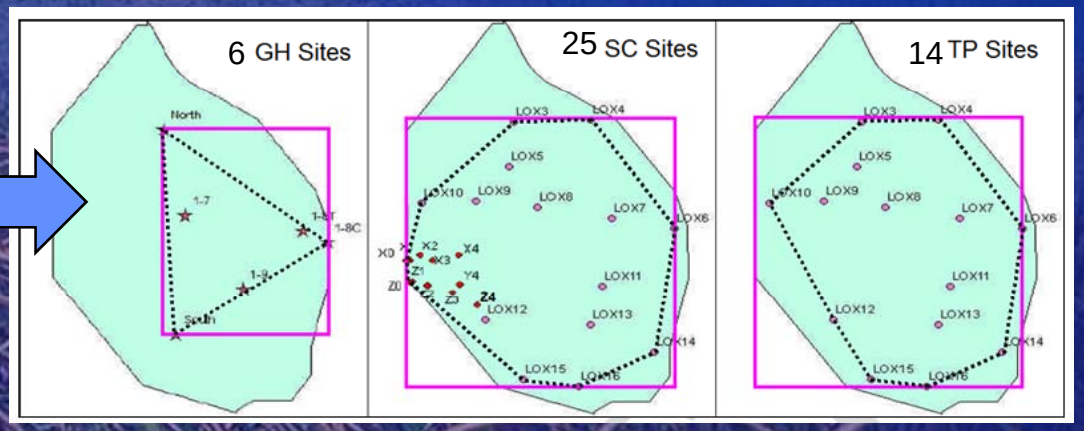
Loxahatchee Empirical Model

- Given inputs of flow, precipitation, and ET
- Create model(s) to simulate:
  - Water levels
  - Specific conductance
  - Total phosphorus

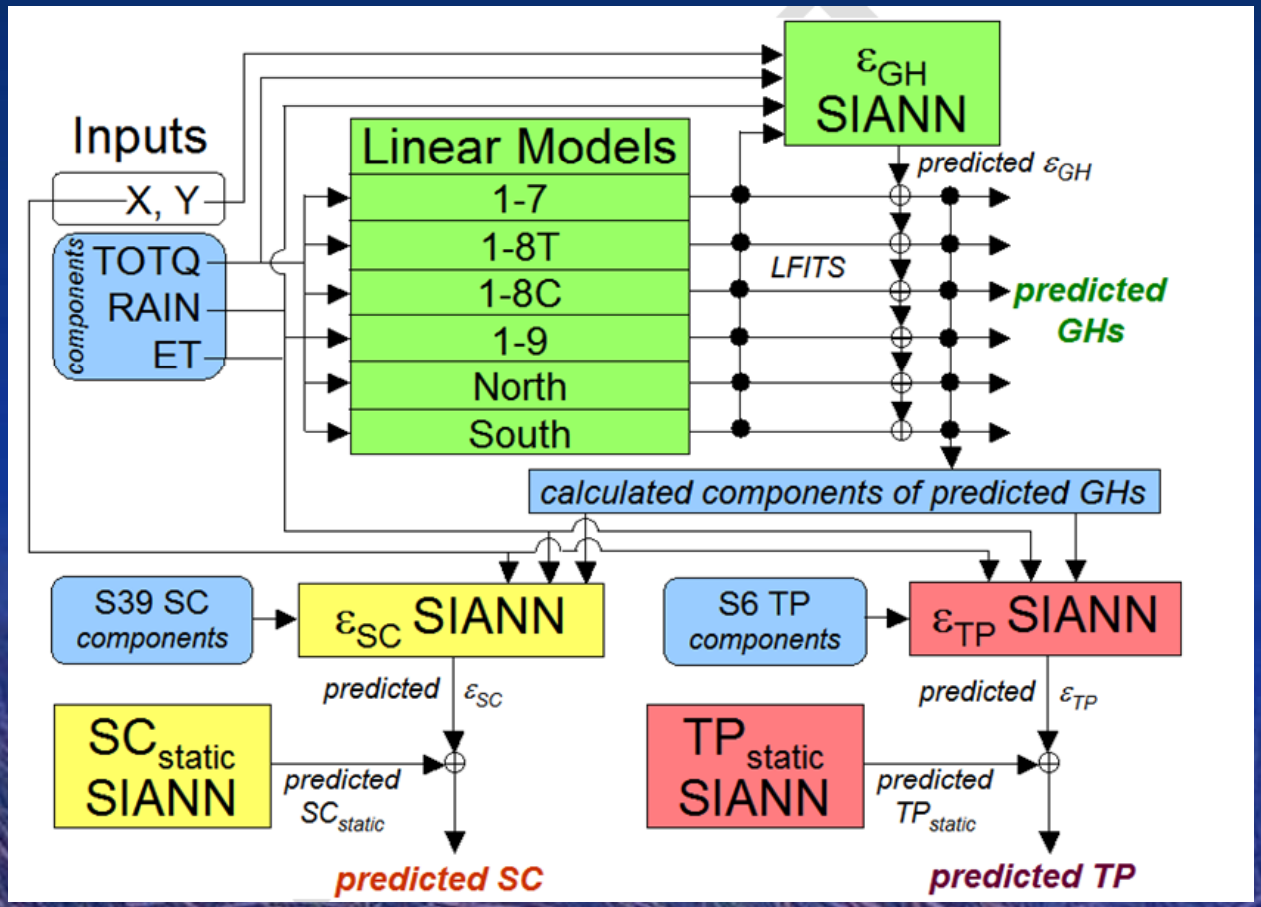
Inputs



Outputs



Model Architecture



Blue – inputs
Green – gage height
Yellow – specific conductance
Pink - phosphorus

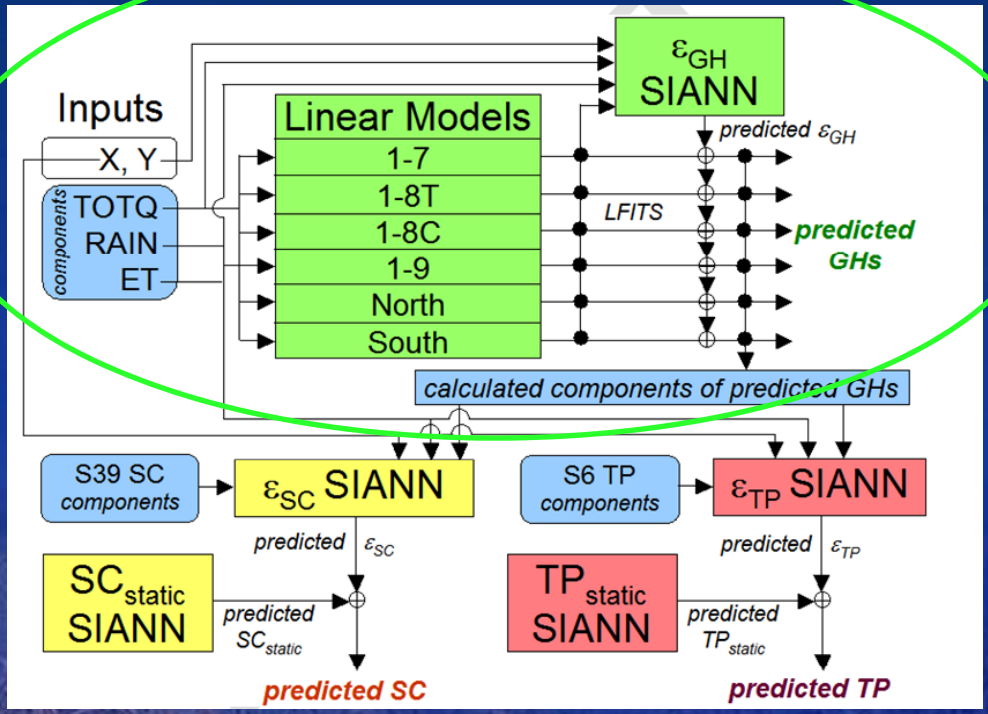
SIANN = spatially interpolating artificial neural network model

USGS

Gage Height Models

Linear models based on optimal time delays and moving window averages of flow, rainfall, and ET

Spatially Interpolating ANNs – error correction models



Final prediction is the sum of the linear model and error prediction models.

USGS

Tau Tool

Excel application to evaluate moving window averages (MWA) and time delays of flow inputs

Station selection

MWA and time delay settings

Rainfall

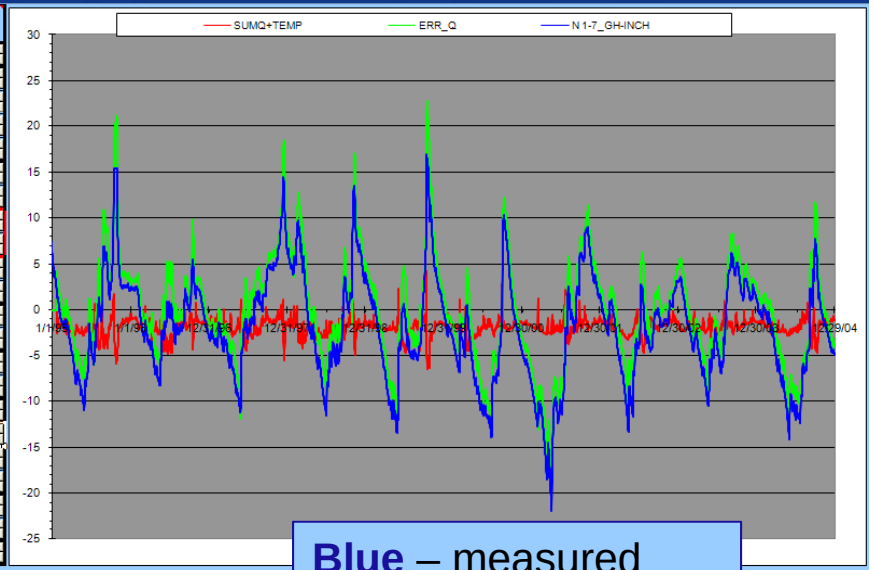
ET

Aggregated flow

param class	parameter select	active	MWsum on/off	d/dt on/off	win	dt	tau
GH Out	1-7_GH-INCH	ON	OFF	OFF	1	1	0
Rain	RAIN_AVG	ON	ON	OFF	1	0	0
	RAIN2	ON			1	0	0
	RAIN3	ON			1	0	0
	RAIN4	ON			1	0	0
ET	STA1W_ET-NEG	ON	ON	OFF	1	0	0
	ET2	ON			26	0	0
SYNC	1		ON	OFF	1	0	0
	2				1	0	0
Q1	Q1-INCH	ON	ON	OFF	1	0	0
sync	Q1.2	ON			4	0	0
Q2	Q2-INCH	ON	ON	OFF	1	0	0
sync	Q2.2	ON			1	0	0
Q3	Q3-INCH	ON	ON	OFF	1	0	0
sync	Q3.2	ON			6	0	0
Q4	Q4-INCH	ON	ON	OFF	4	0	0
sync	Q4.2	ON			5	0	0
Q5	Q5-INCH	OFF	ON	OFF	1	0	0
sync	Q5.2	OFF			1	0	0
Q6	Q6-INCH	ON	ON	OFF	1	0	0
sync	Q6.2	OFF			1	0	0
Tcorr	LOXWS_TEMP	OFF	ON	OFF	1	0	0

SUMQ vs. GH	
R	0.06129
R2	0.00376
Avg Err	-0.04255
Avg |Err|	4.43706

SUMQ+TCORR vs. GH	
R	0.06129
R2	0.00376
Avg Err	-0.04255
Avg |Err|	4.43706



2.124098161 2.02979031
2.066952404 2.01393349
2.061725045 2.00807329
2.060747091 1.89736151

Blue – measured
Red – sum Qs
Green - error

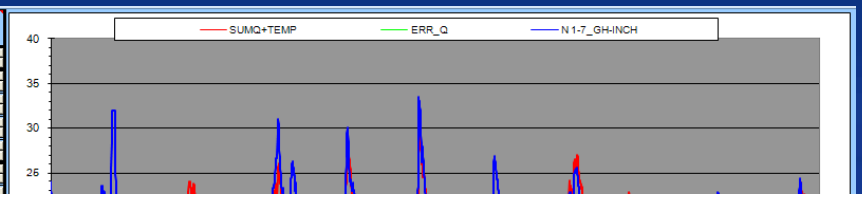
Statistics – R and R^2

Low correlation between untransformed flow input and gage height $R^2 < 0.01$

Linear Models

Adjust MWA and time delays to increase the correlation between inputs and gage heights.

param class	parameter select	active	MWsum on/off	d/dt on/off	win	dt	tau
GH Out	1-7_GH-INCH	ON	OFF	OFF	1	0	0
Rain	RAIN_AVG	ON	ON	OFF	192	0	-1
	RAIN2	ON			12	0	0
	RAIN3	ON			317	0	-1
	RAIN4	ON			1	0	-1
ET	STA1W_ET-NEG	ON	ON	OFF	24	0	0

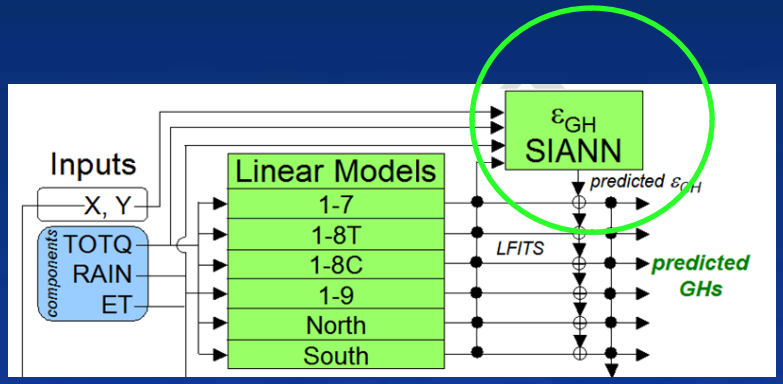


	win&stats	win&stats	win&stats	win&stats	win&stats	win&stats
Output	1-7_GH-INCH	1-8T_GH-INCH	18C_GH-INCH	1-9_GH-INCH	NORTH_GH-INCH	SOUTH_GH-INCH
N	2307	1705	2159	2312	1985	1436
RAIN1	262	286	286	265	289	264
RAIN2	42	82	80	45	51	58
ET	11	55	42	39	5	14
QTOT-in 1	146	206	206	141	140	233
QTOT-in 2	1	45	8	1	1	19
linear m	0.663	0.870	1.182	0.731	0.619	0.993
linear b	181.84	172.39	165.78	181.15	182.88	165.23
linear R2	0.839	0.913	0.815	0.867	0.753	0.933

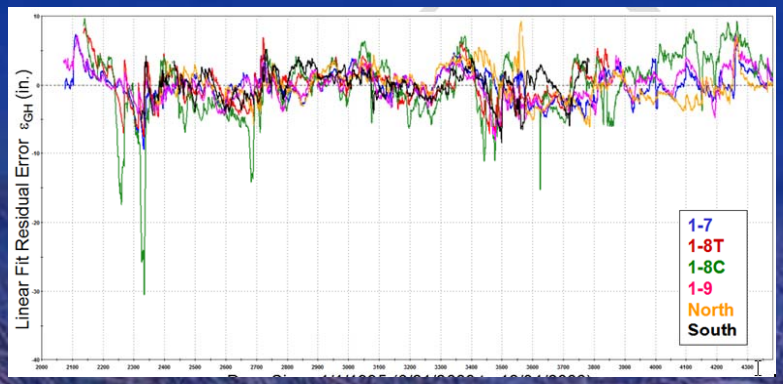
Correlation (R^2) increased from < 0.01 to >0.75

USGS

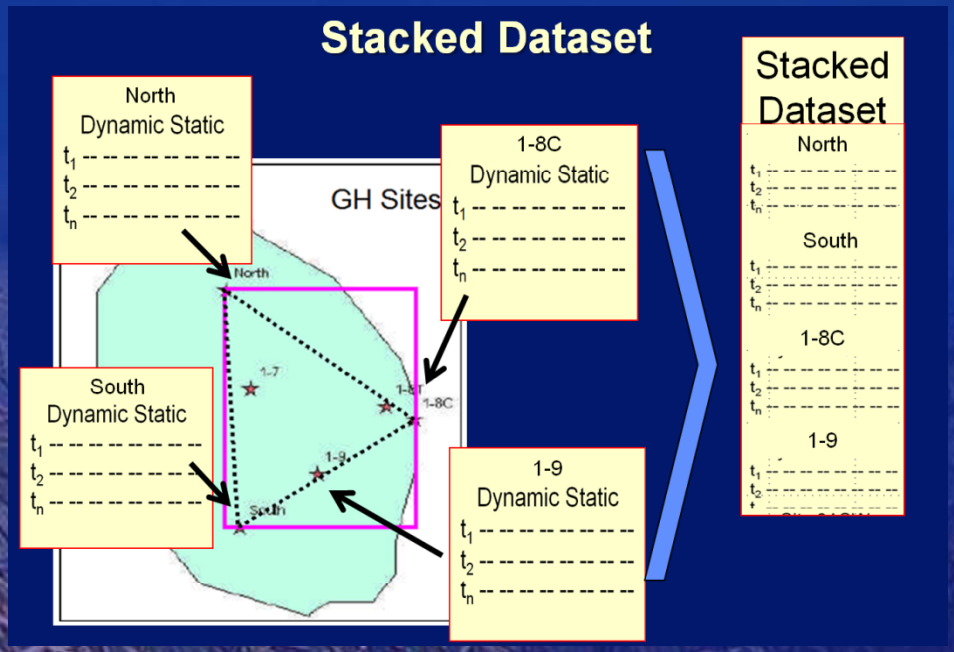
Gage Height Error Correction Model



Model error with a spatially interpolating ANN model

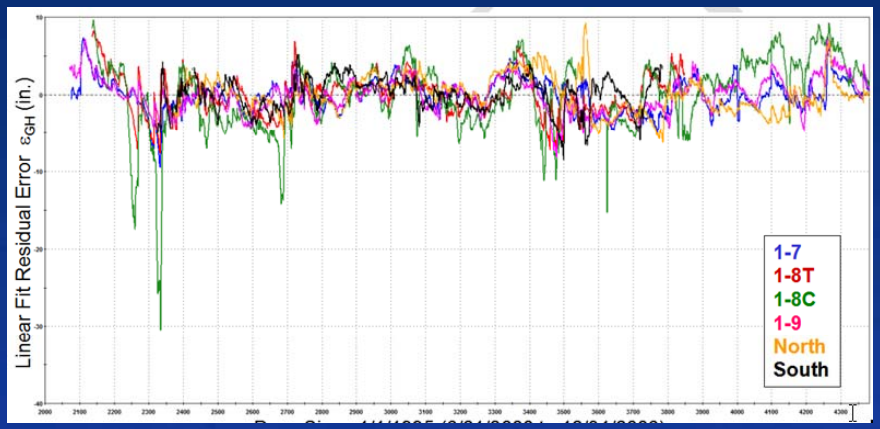


Time series of linear model errors



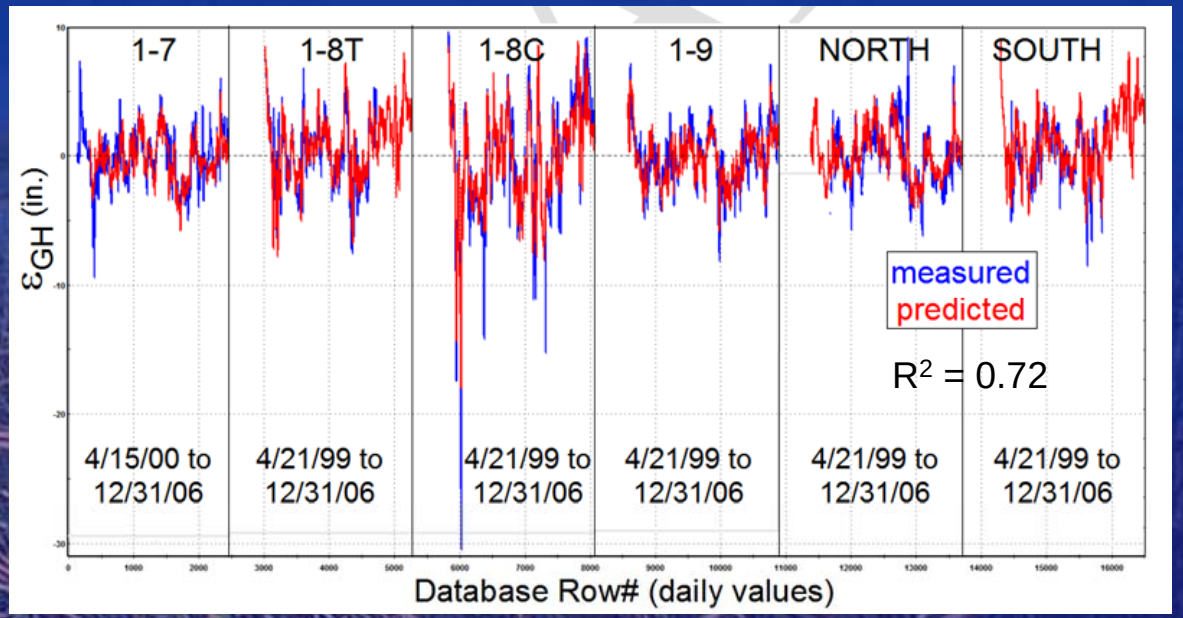
USGS

Gage Height Error Correction Model



Time series of linear model errors

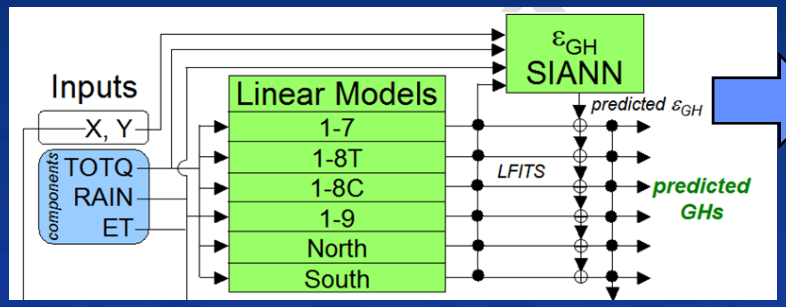
Simulated model error



USGS

Gage Height Predictions

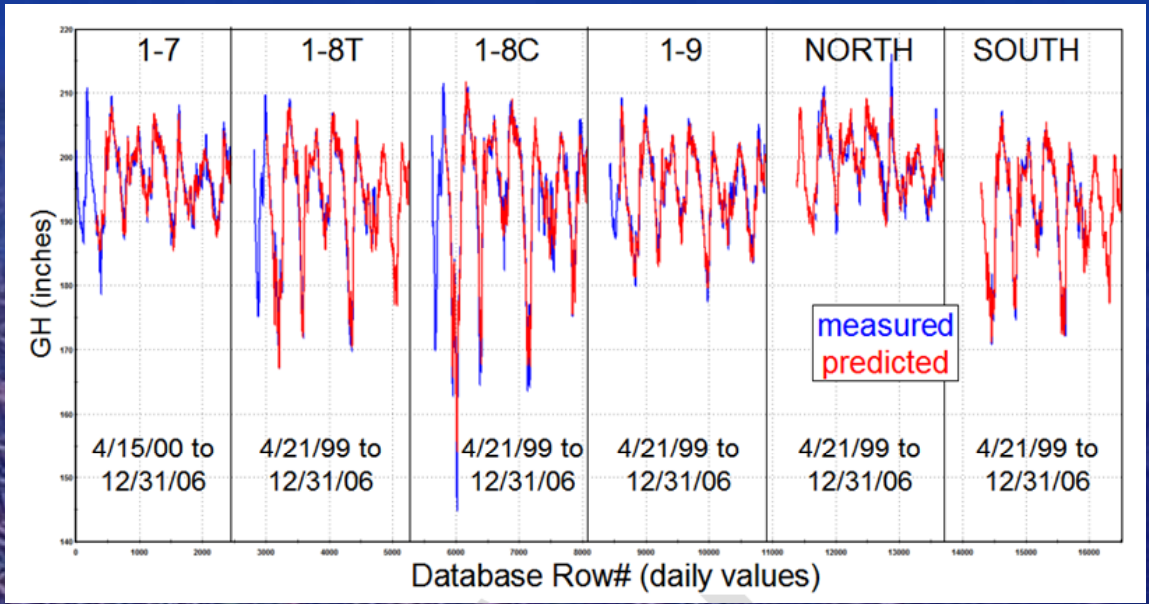
Final gage height prediction is a summation of the linear and error models.



Site	1-7	1-8T	1-8C	1-9	North	South
N	2,123	1,705	2,123	2,312	1,985	1,436
R^2	0.932	0.976	0.932	0.953	0.902	0.972
RMSE (in.)	1.36	1.35	1.36	1.34	1.47	1.38
Min. Value (in.)	178.6	167.3	144.7	177.4	188	170.8
Max. Value (in.)	210.8	209.6	211.6	209.2	216	207.2
Range (in.)	32.3	42.4	66.8	31.8	28	36.4
RMSE/Range (%)	4.21	3.19	2.04	4.21	5.26	3.78

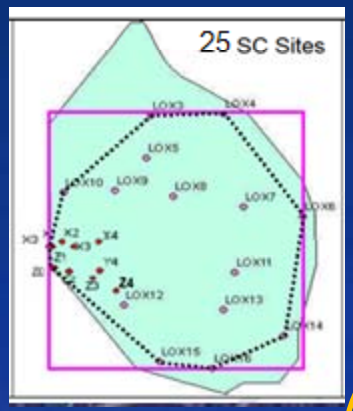
Final gage height Predictions:

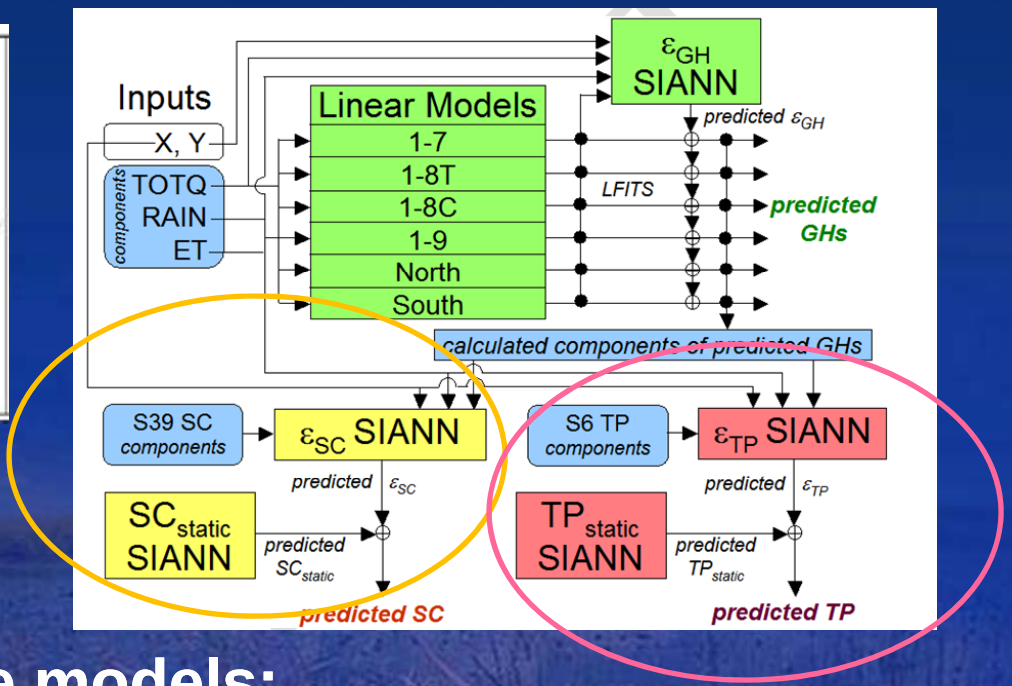
R^2 0.90 – 0.98

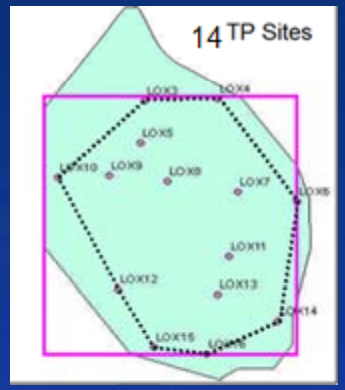


USGS

Simulation of Specific Conductance and Total Phosphorus







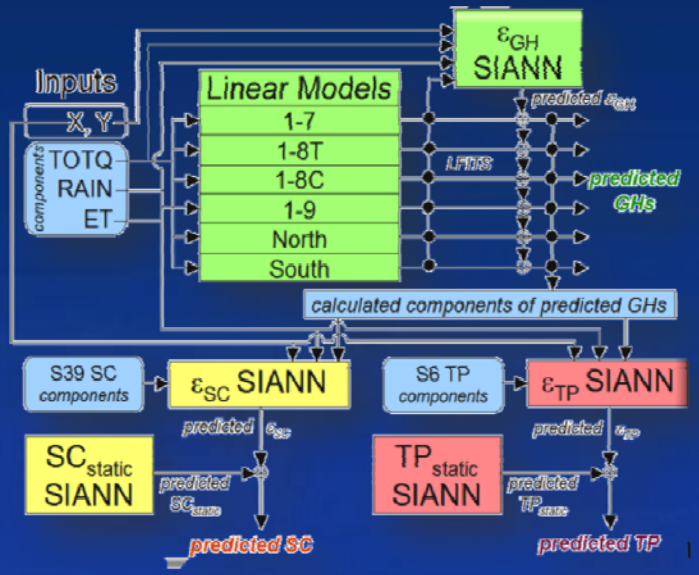
Two stage models:

- **Static model using X, Y and measured data**
- **Dynamic model predict variability about mean**

USGS

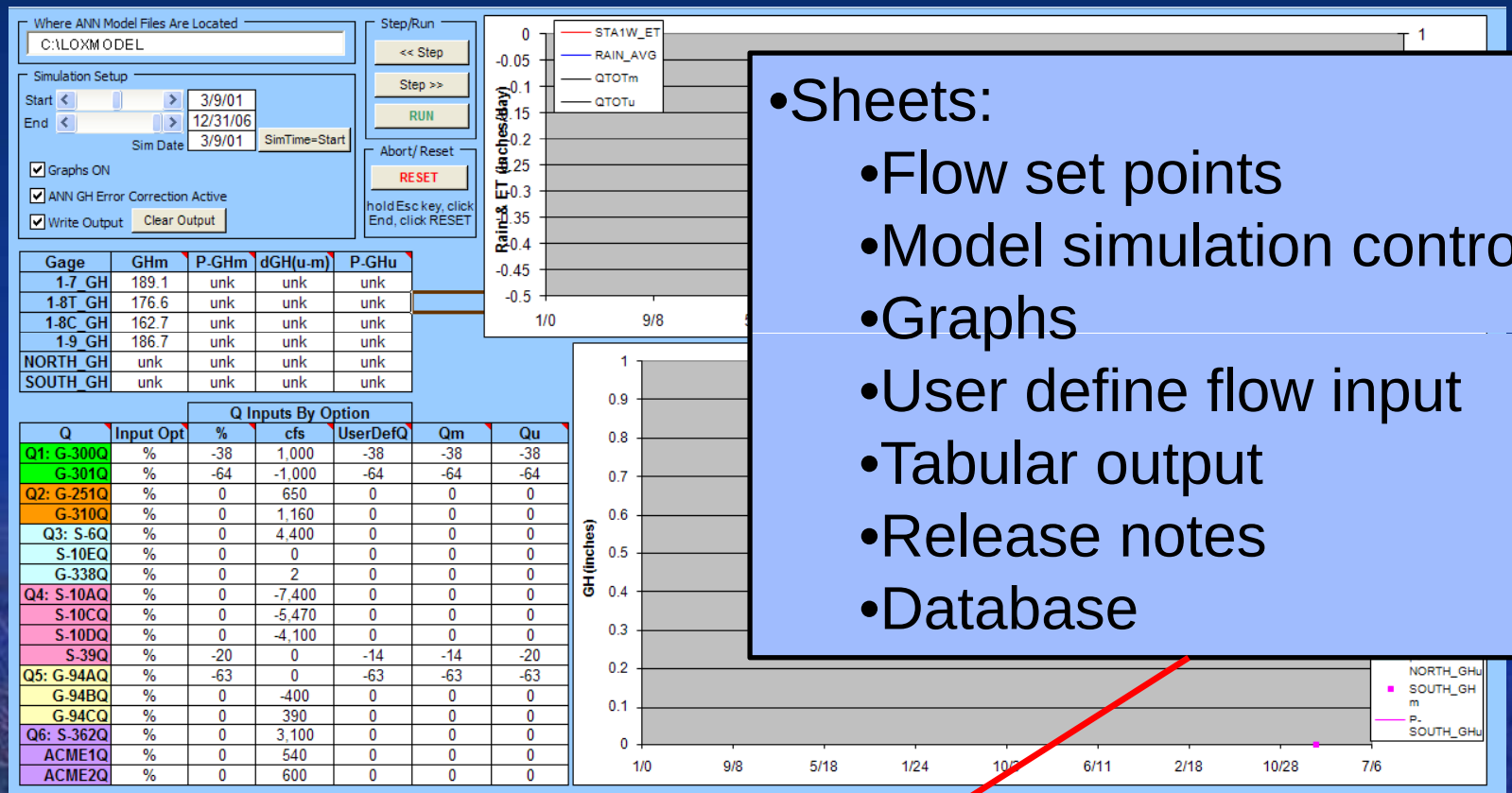
Decision Support System

- Excel application
- Integrates
  - Historical database
  - ANN and regression models
  - Model controls
  - Streaming graphics
  - 3D visualization
  - Model simulation output



USGS

Excel Spreadsheet



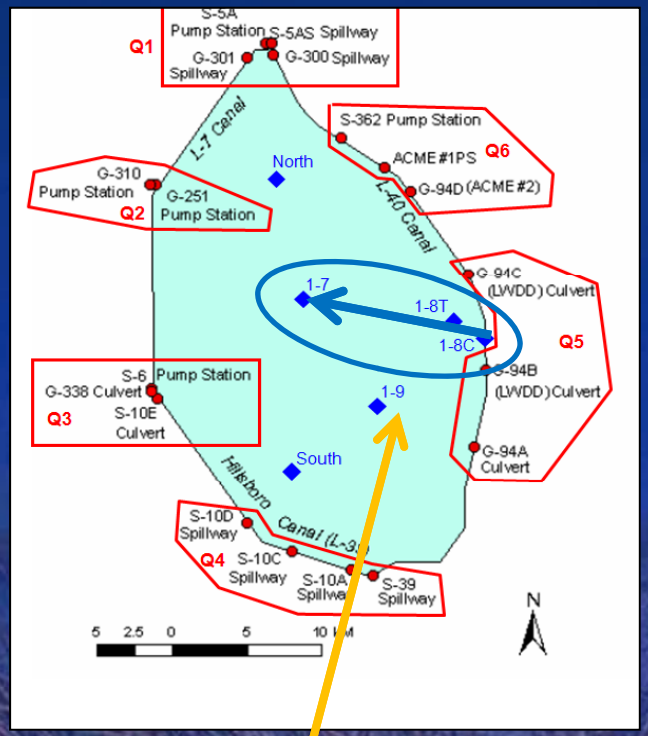
Gage	GHm	P-GHm	dGH(u-m)	P-GHu
1-7_GH	189.1	unk	unk	unk
1-8T_GH	176.6	unk	unk	unk
1-8C_GH	162.7	unk	unk	unk
1-9_GH	186.7	unk	unk	unk
NORTH_GH	unk	unk	unk	unk
SOUTH_GH	unk	unk	unk	unk

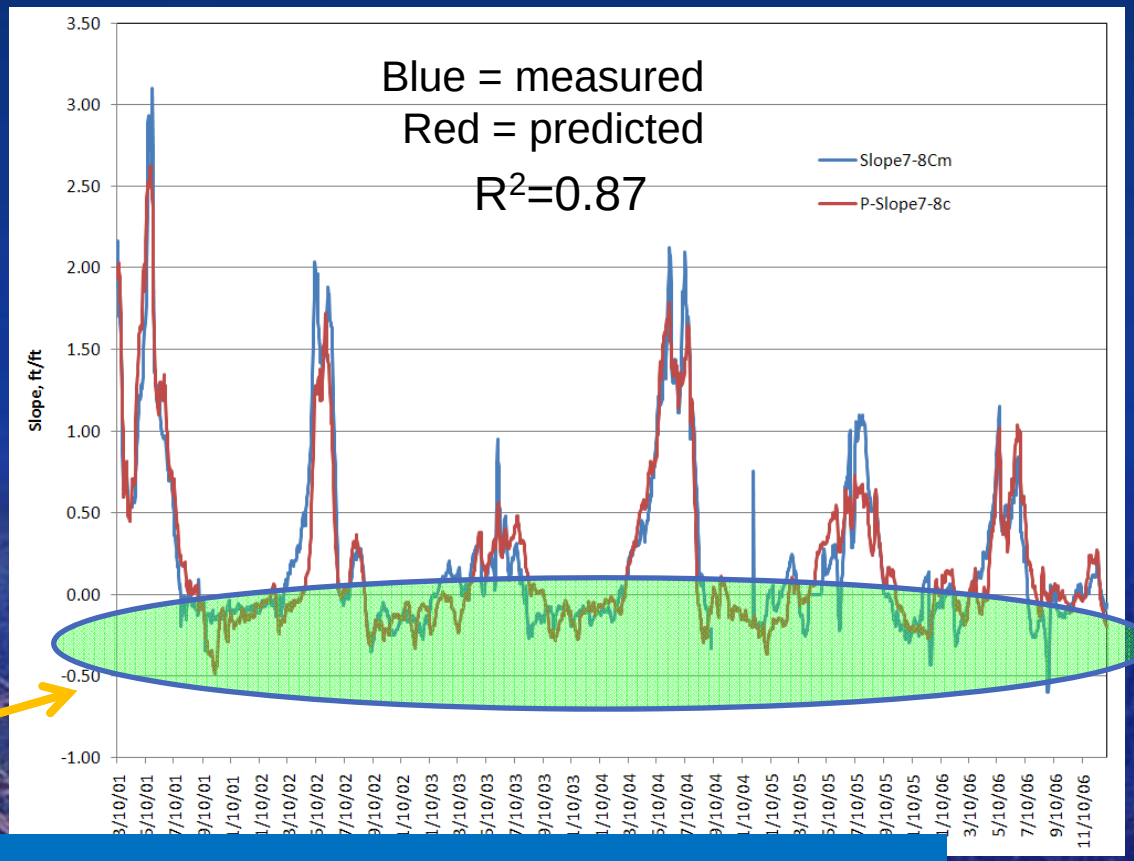
		Q Inputs By Option				
Q	Input Opt	%	cfs	UserDefQ	Qm	Qu
Q1: G-300Q	%	-38	1,000	-38	-38	-38
G-301Q	%	-64	-1,000	-64	-64	-64
Q2: G-251Q	%	0	650	0	0	0
G-310Q	%	0	1,160	0	0	0
Q3: S-6Q	%	0	4,400	0	0	0
S-10EQ	%	0	0	0	0	0
G-338Q	%	0	2	0	0	0
Q4: S-10AQ	%	0	-7,400	0	0	0
S-10CQ	%	0	-5,470	0	0	0
S-10DQ	%	0	-4,100	0	0	0
S-39Q	%	-20	0	-14	-14	-20
Q5: G-94AQ	%	-63	0	-63	-63	-63
G-94BQ	%	0	-400	0	0	0
G-94CQ	%	0	390	0	0	0
Q6: S-362Q	%	0	3,100	0	0	0
ACME1Q	%	0	540	0	0	0
ACME2Q	%	0	600	0	0	0

- Sheets:
  - Flow set points
  - Model simulation controls
  - Graphs
  - User define flow input
  - Tabular output
  - Release notes
  - Database

Info | Q SPs | Controls | InputGraphs | OutputGraphs | UserDefQs | Output | ReleaseNotes | Database

DSS Application:
Canal water intrusion into the marsh

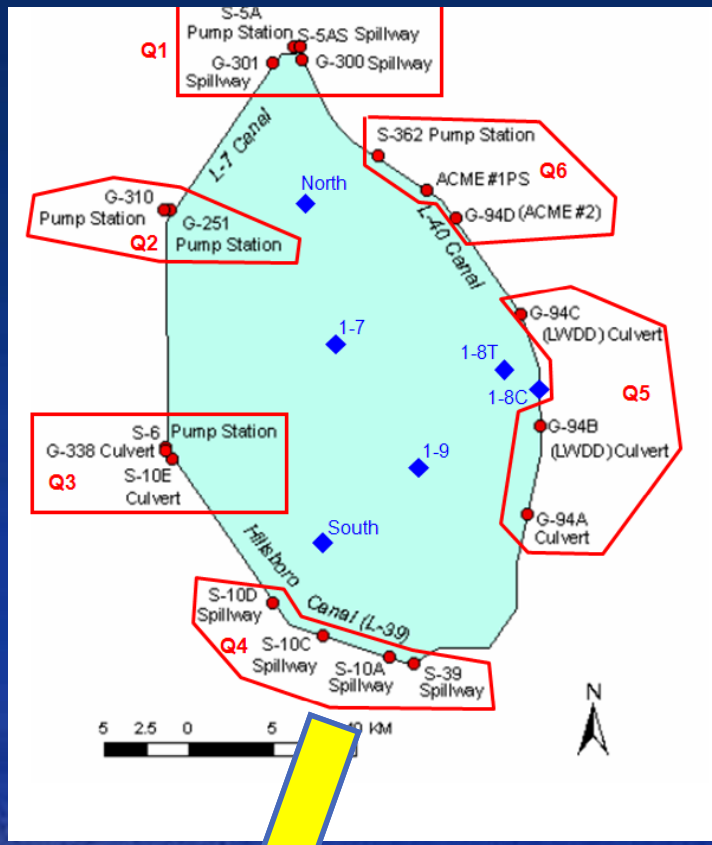




Intrusion events: Canal WL > Marsh WL
Negative slope

Scenario 1

- ***What will be the model respond be to the simulated change in slope if the flow of Q4 (S-10D, S-10C, S-10A, and S-39) is increased by 40 percent?***



Increase flow by 40%

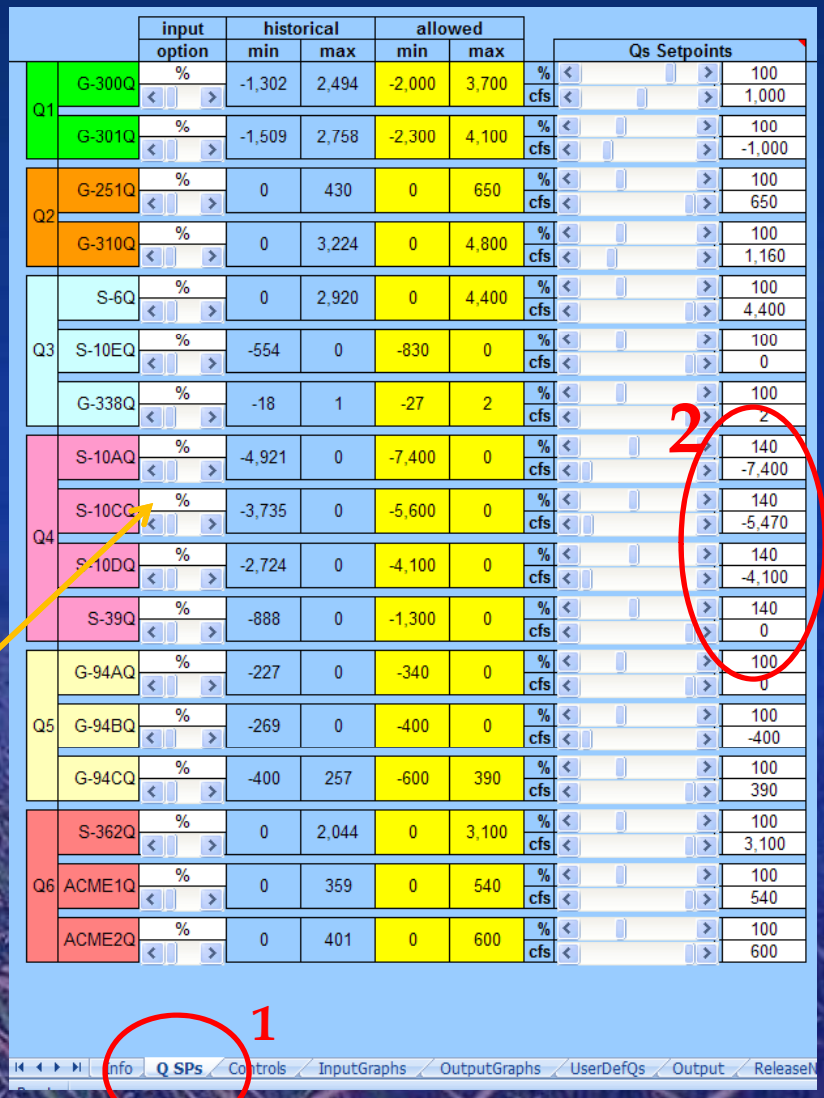
USGS

DSS Application Set-up

- On Flow Set Point sheet[1] (Q SPs) set flows for the Q4 structures to [2]140% of historical flows

Flow input options:
- % historical flow
- Constant flow

All flows are set to % of historical
100% = actual flows



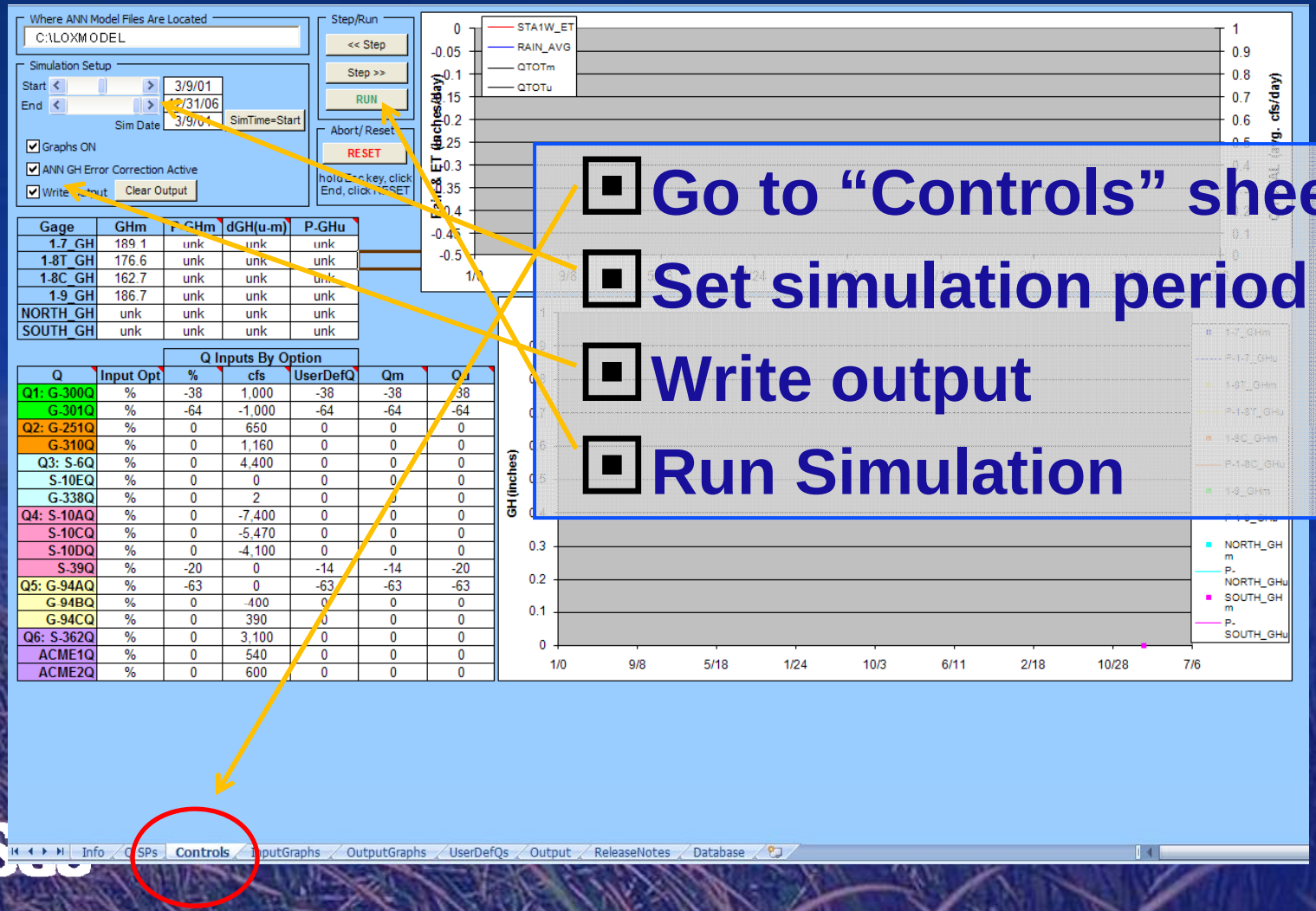
		input option	historical min	historical max	allowed min	allowed max		Qs Setpoints
Q1	G-300Q	%	-1,302	2,494	-2,000	3,700	%	100
							cfs	1,000
	G-301Q	%	-1,509	2,758	-2,300	4,100	%	100
							cfs	-1,000
Q2	G-251Q	%	0	430	0	650	%	100
							cfs	650
	G-310Q	%	0	3,224	0	4,800	%	100
							cfs	1,160
Q3	S-6Q	%	0	2,920	0	4,400	%	100
							cfs	4,400
	S-10EQ	%	-554	0	-830	0	%	100
							cfs	0
	G-338Q	%	-18	1	-27	2	%	100
							cfs	2
Q4	S-10AQ	%	-4,921	0	-7,400	0	%	140
							cfs	-7,400
	S-10CQ	%	-3,735	0	-5,600	0	%	140
							cfs	-5,470
	S-10DQ	%	-2,724	0	-4,100	0	%	140
							cfs	-4,100
	S-39Q	%	-888	0	-1,300	0	%	140
							cfs	0
Q5	G-94AQ	%	-227	0	-340	0	%	100
							cfs	0
	G-94BQ	%	-269	0	-400	0	%	100
							cfs	-400
	G-94CQ	%	-400	257	-600	390	%	100
							cfs	390
Q6	S-362Q	%	0	2,044	0	3,100	%	100
							cfs	3,100
	ACME1Q	%	0	359	0	540	%	100
							cfs	540
	ACME2Q	%	0	401	0	600	%	100
							cfs	600

Info | Q SPs | Controls | InputGraphs | OutputGraphs | UserDefQs | Output | ReleaseN

1

2

DSS Simulation Controls

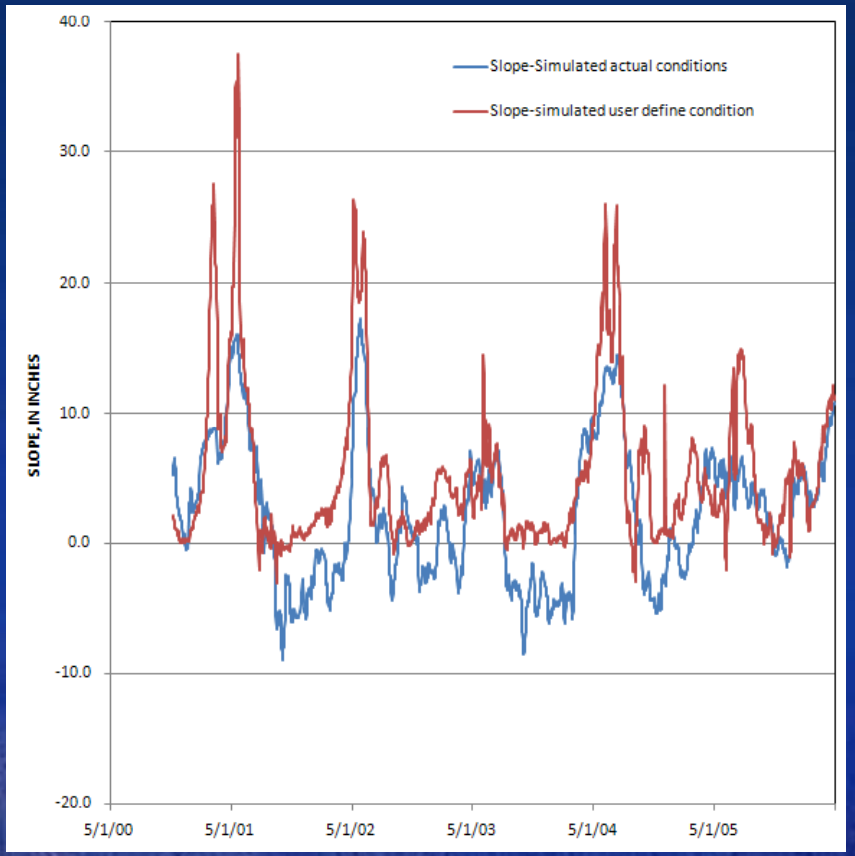


- Go to “Controls” sheet
- Set simulation period
- Write output
- Run Simulation

DSS Scenario Results

40% increase flow Q4

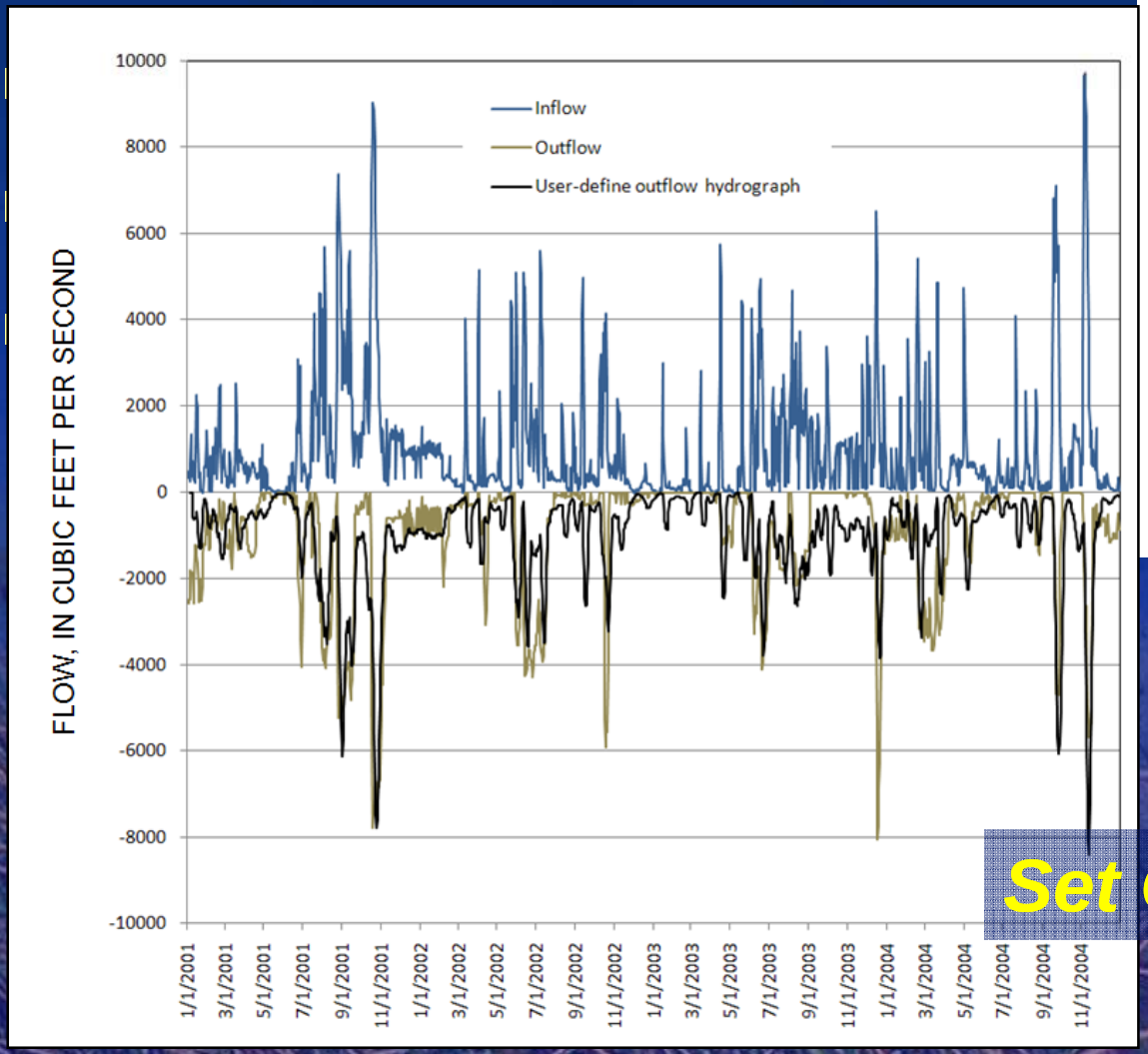
•Blue – Simulated actual slope
•Red – Simulated scenario – increase flow Q4 by 140%

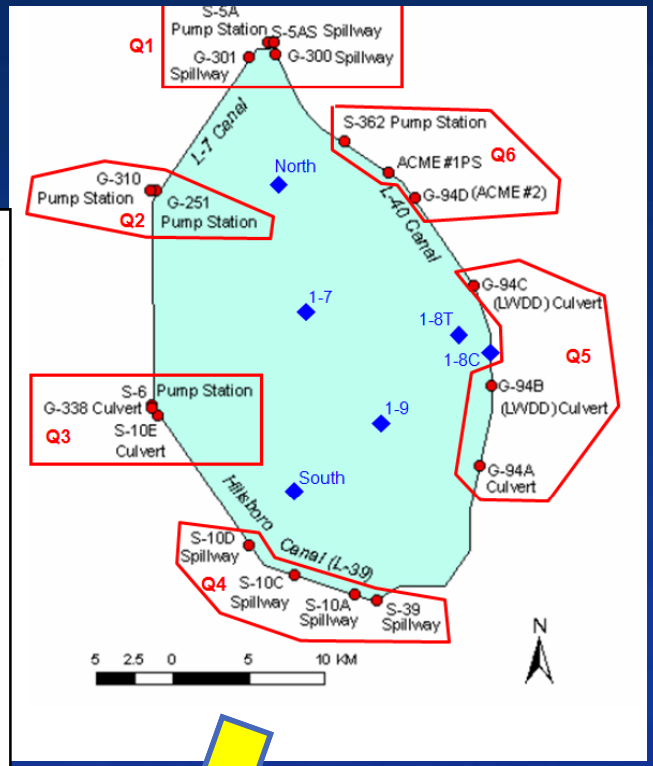


•Increasing the flows did increase the slope.
•Negative slopes overall were minimized
•Positive slopes also increased.

USGS

Scenario 2
Inflow = Outflow



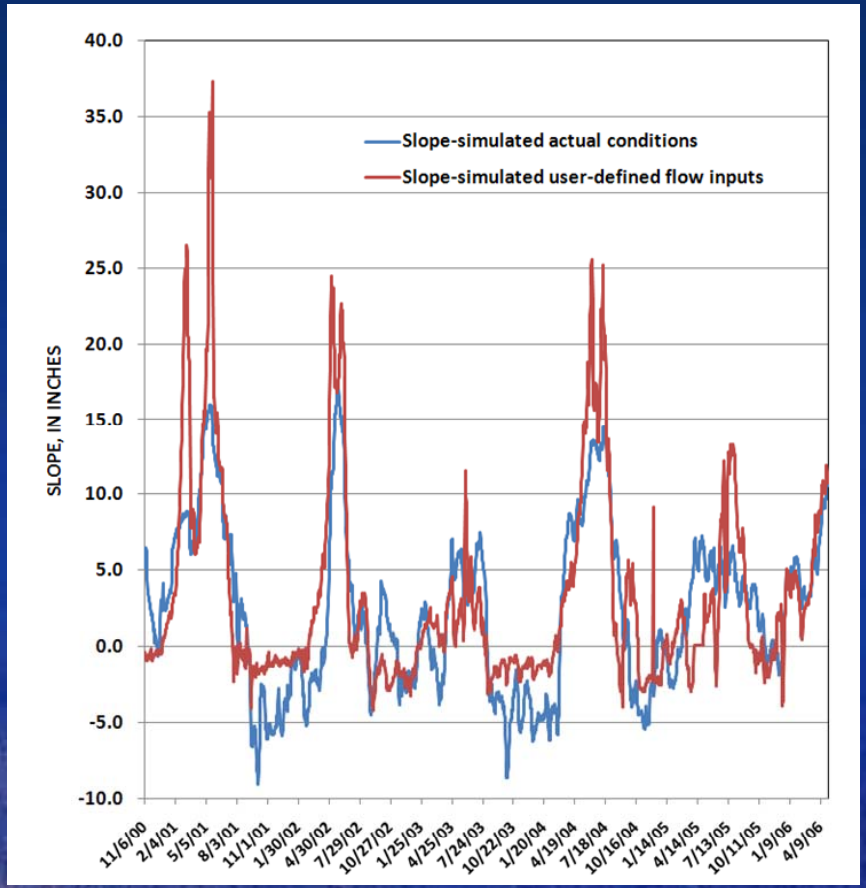


Set Outflow = Inflows

DSS Scenario Results

Outflow = Inflow

- Blue – simulated actual slope
- Red – simulated scenario slope Outflow = Inflow



- **Increasing the outflows did increase the slope.**
- **Negative slopes overall were minimized**
- **Positive slopes also increased.**

USGS

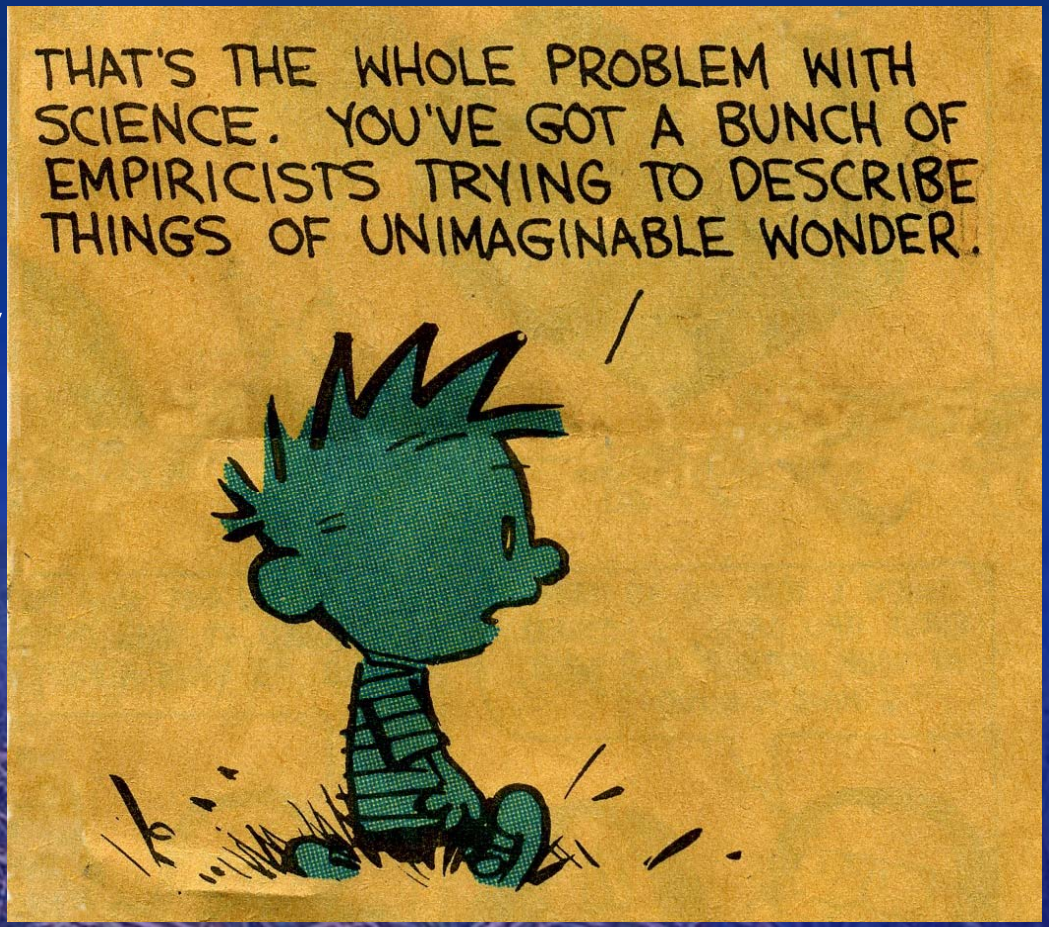
Summary

- Model allows users to evaluate effects of flow releases
- Evaluate short- and long-term flow regimes
- Excel platform for DSS – facilitates dissemination of models user of various technical levels
- DSS database easily updated
- USGS report in final stage of review/production process.

USGS

Questions

Paul Conrads
pconrads@usgs.gov
803 750-6140



USGS