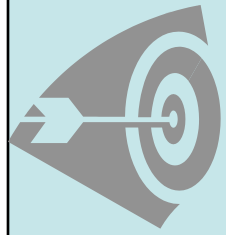


Statistical Methods/ Programs to Support Water Supply Forecasting

CBRFC October 2006
Steve Shumate

The current software package of
inter-related programs was
initiated in 1992
It is called SWS or
Statistical Water Supply

Many renditions later...



SWS – What is it?

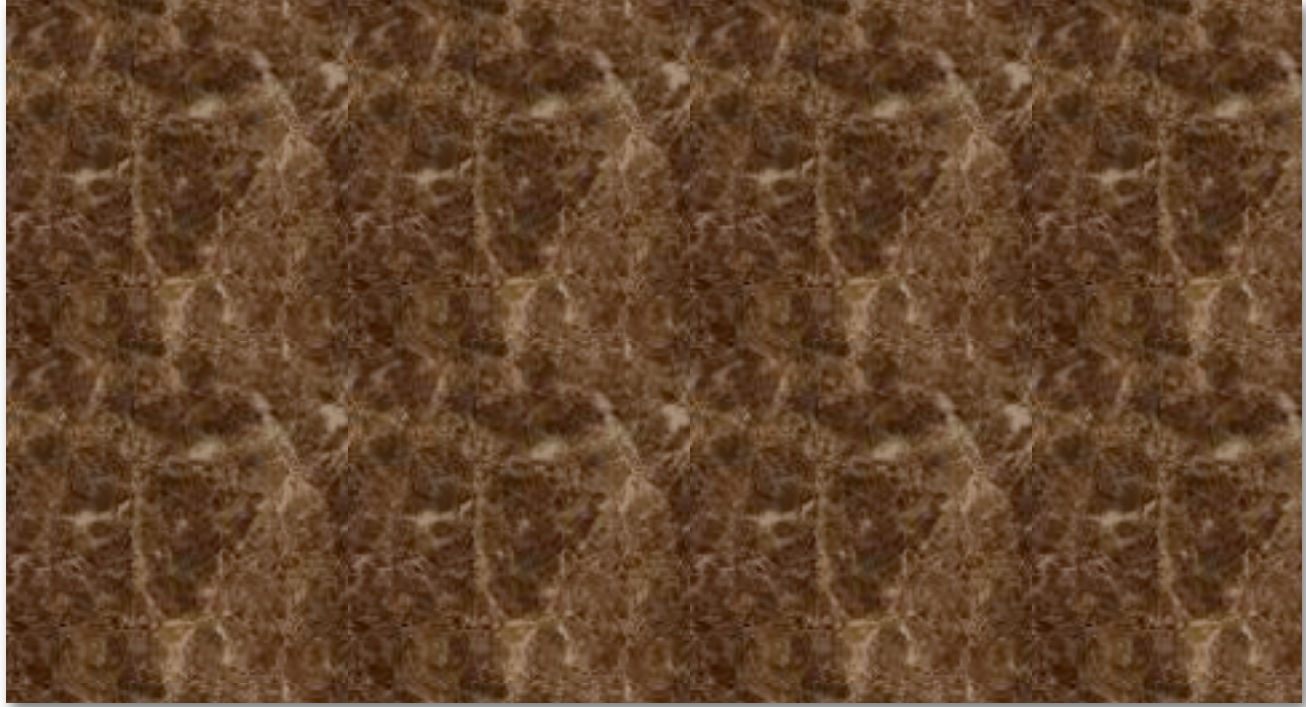
Why is it a good thing?

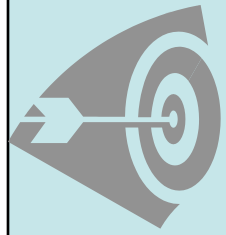
- ⊕ SWS – Statistical Water Supply - a package of inter-related programs to support water supply forecasting
- ⊕ Monthly data – reap the benefits of the relational database (library of functions as well as standard SQL methods)
- ⊕ Ancillary programs – take advantage of many programs to report and manipulate monthly data
- ⊕ Companion to ESP – “Super Ensemble” – one or more models to forecast the same thing (model diversity)
- ⊕ Ease of use has been (and will be) a continuous priority during software development
- ⊕ The often used phrase: “wouldn’t it be nice if...” – features are more easily accommodated/incorporated as the software development environment and working environment are the same



I. Calibration

II. Operation

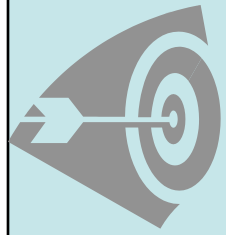




SWS – Calibration Programs

- ⊕ **Regcand** – find candidate variables, calculate correlation matrix
- ⊕ **Regdata** – prepare input file for regcomb
- ⊕ **Regcomb** – find optimal equations (top 20)
- ⊕ **Epal** – Evaluate, Pick and Load equations to the postgres db structure
- ⊕ **Cstats** – Calculate and store period statistics

- Equation # 3
- Y 1 DOLC2/QCMRZZZ, Ap-Jl, DOLORES - DOLORES
- = -5.203
- + 14.859 X4 HSPC2/PPMRZZZ, No-Ja, FORT LEWIS
- + 57.175 X5 DRRC2/QCMRZZZ, Ja, DOLORES - RICO, BLO
- + 10.161 X6 LNCC2/SWIRMZZ, Fe, LONE CONE
- Number of observations used = 30
- Number of principal components used = 1
- CORRELATION COEFFICIENT (R) = 0.709
- STANDARD ERROR = 80.055 (rank = 4)
- JACKKNIFE CORRELATION COEFFICIENT = 0.674
- JACKKNIFE STANDARD ERROR = 84.035
- JACKKNIFE BIAS: above average flow = -56.063 (14 obs.)
- below average flow = 48.137 (16 obs.)
-
- JACKKNIFE JACKKNIFE
- YEAR OBSERVED COMPUTED ERROR COMPUTED ERROR
- 61 186.50 135.21 -51.29 130.37 -56.13
- 62 256.30 248.90 -7.40 247.56 -8.74
- 63 126.40 175.12 48.72 178.03 51.63
- 64 152.60 126.45 -26.15 123.94 -28.66
- 65 335.80 366.18 30.38 368.00 32.20
- 66 206.20 299.66 93.46 304.57 98.37
- 67 123.50 248.37 124.87 252.52 129.02



Regcand - Candidate Variable Search

Lets user visualize spatial relationship of potential predictor variables. Calculates accumulation of Y var. Calculates correlation matrix for each independent variable; accumulations as well as discrete months.

Mercator
Scale 1:2000000

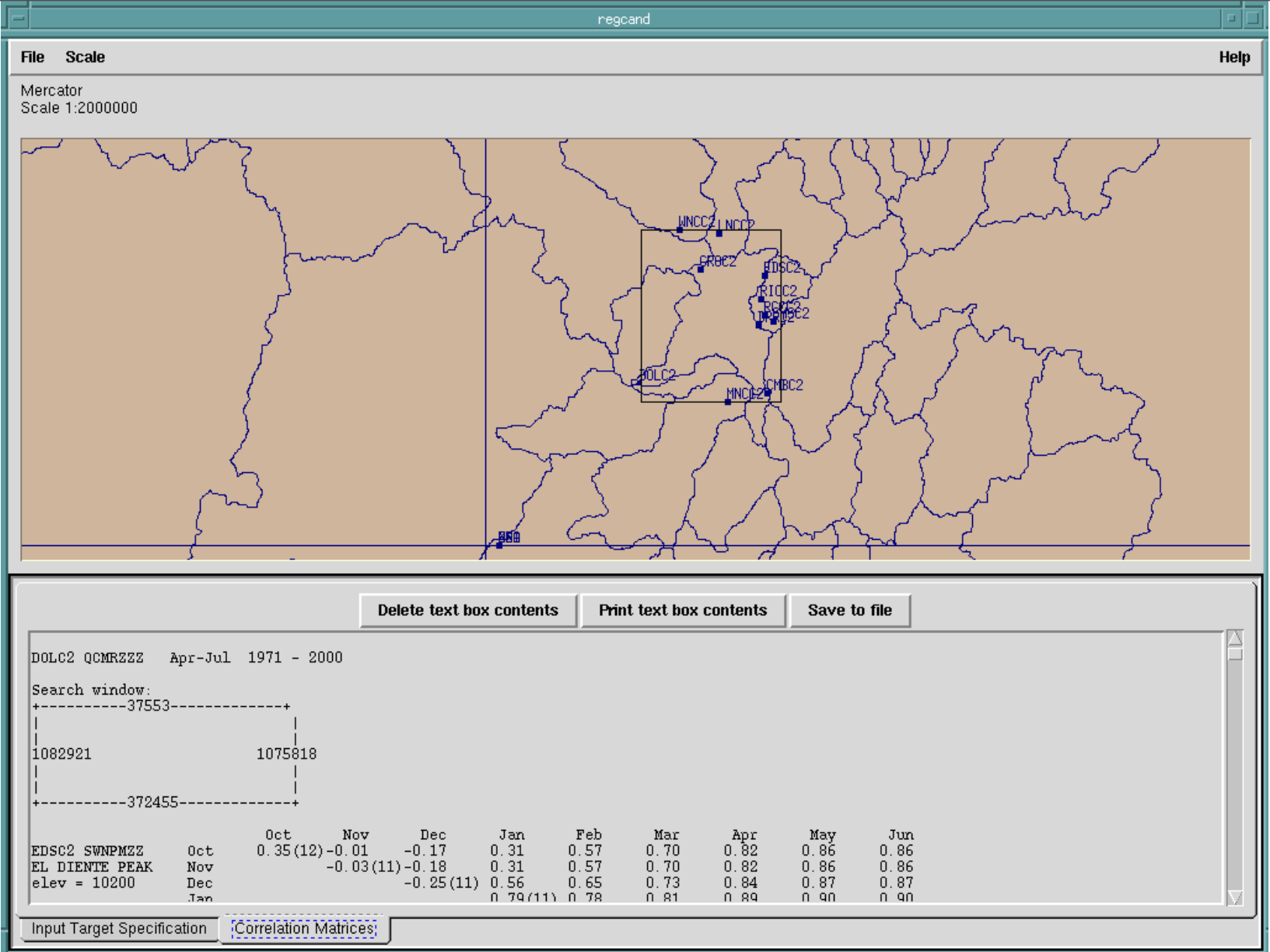
Help

Station search string:

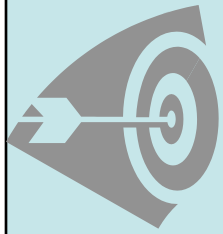
Accumulation period

Beginning Year Ending Year

DOLC2 QCMRZZZ Apr - Jul 1971 - 2000



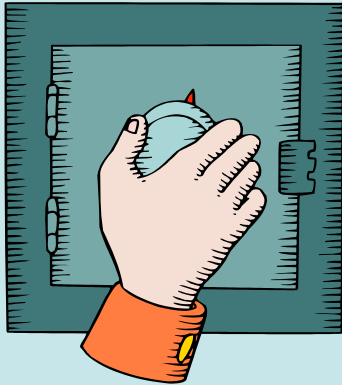
Future software development will include GIS where it is useful



REGCOMB

Combination Analysis

Why? ...there are over 500 million unique combinations of just 30 variables.



Predictors, where A,B,C are stations:

- snow-A, snow-B, snow-C
- precip-A, precip-B, precip-C (Oct-Dec)
- flow-A, flow-B
- ...

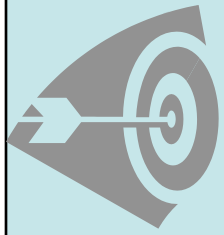
+



$$1. y = mx_1 + mx_2 + mx \dots + b$$

$$2. y = mx_1 + mx_2 + mx \dots + b$$

3...



REGCOMB

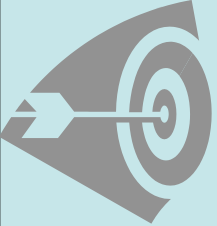
Jack-knife Testing



+

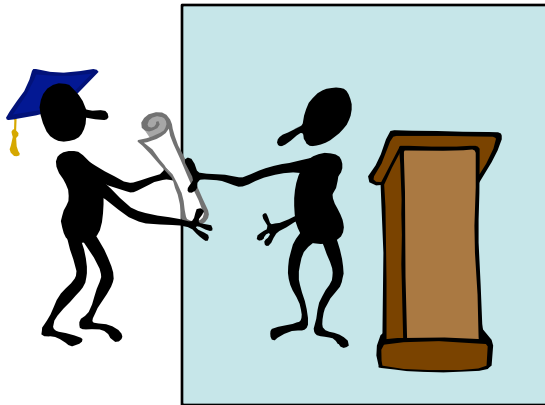


For a given set of predictors, one observation (one year) is deleted from the data set. Optimal coefficients are determined. The equation is then measured as to how well it predicted the selected year. The idea here is to simulate how well the equation will perform in an operational environment where the predictand is not known at the time of equation execution.



REGCOMB

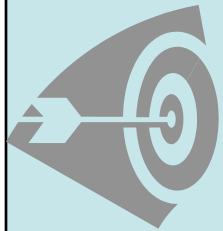
Principal Components



+



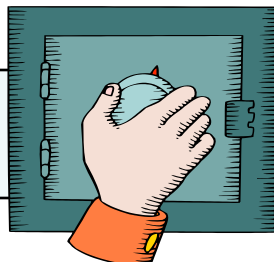
Variables in a water-supply equation tend to have high correlation with each other. This causes problems when trying to determine optimal coefficients via traditional regression techniques. Principal components analysis is a way to determine optimal coefficients while recognizing and addressing the intercorrelation problems.



REGCOMB

It's all good...

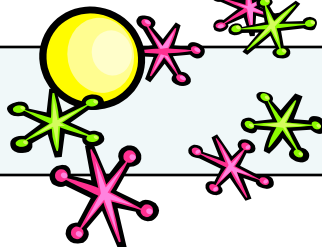
Combination
Analysis



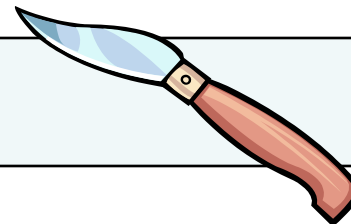
+



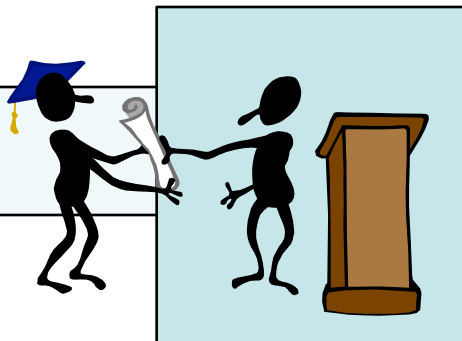
Jack-knife
error
computation



+

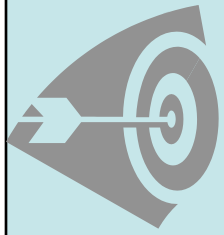


Principal
Components



+





Epal – Evaluation, Pick and Load

This program allows the user to examine the output from REGCOMB and ultimately choose and store an equation to the postgresql database for operational use.

The screenshot displays the Epal software interface, which is used for evaluating and selecting equations. The main window is titled "Evaluate, Pick and Load" and shows the current input file as "/pc1/home/sts/wsup/vkshop/gsfns/mar.out".

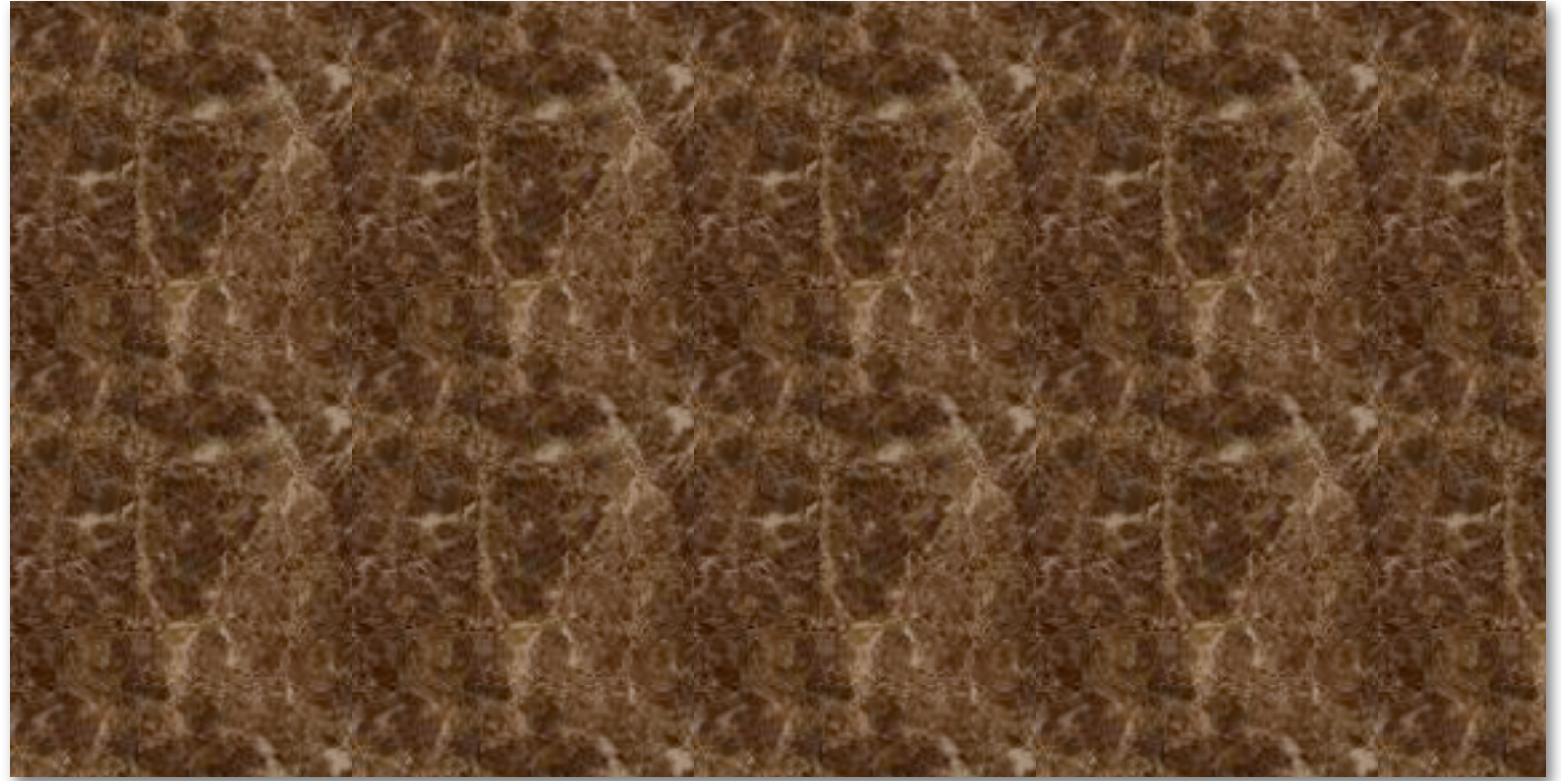
The "EQUATION SUMMARY:" table lists 12 equations with their respective variables, Jackknife Standard Error, Number of Observations (NO. OBS.), and Bias (BIAS ABOVE and BIAS BELOW). The table is as follows:

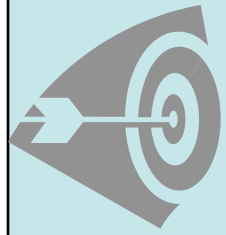
RANK	VARIABLES	JACKKNIFE STANDARD ERROR	NO. OBS.	BIAS ABOVE AVG.	BIAS BELOW AVG.
1	X X X X X X X X X X X X X X X X	0.300	13	0.14	0.01
2	X X X X X X X X X X X X X X X X	0.338	13	0.07	0.03
3	X X X X X X X X X X X X X X X X	0.347	13	0.14	0.03
4	X X X X X X X X X X X X X X X X	0.348	13	0.09	0.03
5	X X X X X X X X X X X X X X X X	0.350	13	0.16	0.02
6	X X X X X X X X X X X X X X X X	0.351	13	0.13	0.01
7	X X X X X X X X X X X X X X X X	0.354	13	0.16	0.02
8	X X X X X X X X X X X X X X X X	0.357	13	0.06	0.02
9	X X X X X X X X X X X X X X X X	0.359	13	0.09	0.03
10	X X X X X X X X X X X X X X X X	0.359	13	0.07	0.02
11	X X X X X X X X X X X X X X X X	0.366	13	0.07	0.02
12	X X X X X X X X X X X X X X X X	0.368	13	0.09	0.02

The "Equation # 8" window shows the selected equation: "GSFNS/QCMRZZZ, Mr-My, SAN FRANCISCO - GLENWOOD, NR". It lists 18 variables (X1 to X18) and provides statistical data: "Number of observations used = 13", "Number of principal components used = 1", "RELATION COEFFICIENT (R) = 0.972", "STANDARD ERROR = 0.321 (rank = 40)", "JACKKNIFE CORRELATION COEFFICIENT = 0.967", "JACKKNIFE STANDARD ERROR = 0.357", "JACKKNIFE BIAS: above average flow = 0.063 (4 obs.), below average flow = 0.022 (9 obs.)".

The "Equation # 9" window shows a graph of "std observed vs. predicted" for "GSFNS/QCMRZZZ, Mr-My, SAN FRANCISCO - GLENWOOD, NR". The graph plots "Observed" (x-axis) against "Computed" (y-axis) values, with a diagonal line representing the identity function. Data points are labeled with their rank: 76, 65, 78, 66, 73.

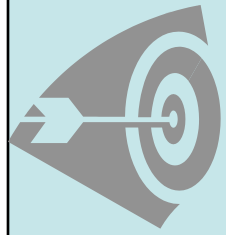
The "Equation # 9" window also shows the selected equation: "GSFNS/QCMRZZZ, Mr-My, SAN FRANCISCO - GLENWOOD, NR". It lists 18 variables (X1 to X18) and provides statistical data: "Number of observations used = 13", "Number of principal components used = 1", "CORRELATION COEFFICIENT (R) = 0.967", "STANDARD ERROR = 0.344 (rank = 105)", "JACKKNIFE CORRELATION COEFFICIENT = 0.968", "JACKKNIFE STANDARD ERROR = 0.358", "JACKKNIFE BIAS: above average flow = 0.089 (4 obs.), below average flow = 0.028 (9 obs.)".





SWS – Operational Programs

- ⊕ **Nextreg** - exercises equation with operational data and allows user to store forecasts in the database
- ⊕ **Nextpub** - extracts forecasts from database and outputs several different formats for various purposes



Nextreg – operational exercise of equations

Equation Output/
Fcst Input tab –
spreadsheet style
output of primary
and secondary
equations

```

DRGC2 QCMRZZZ P Apr-Jul (ANIMAS - DURANGO) JR2: 0.741 # yrs: 30
AVG: 440.000 YTRANS: none
-----
DURANGO WATER RESOURCE DUGC2/PPMRZZZ
May 0.77V 66% * 19.582 = 15.08

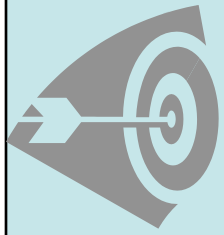
SILVERTON SLVC2/PPMRZZZ (Apr - May):
Apr 1.17V 71%
May 1.03V 55%
-----
2.20 63% * 7.771 = 17.10

RED MOUNTAIN PASS RMPC2/SWIRMZZ
May 17.30Z 64% * 5.966 = 103.21

CASCADE CSCC2/SWIRMZZ
May 0.60Z 16% * 7.216 = 4.33

RED MOUNTAIN PASS RMPC2/SWIRMZZ (May - Jun):
May 17.30Z 64%
Jun 0.70Z 5%
-----
18.00 44% * 2.811 = 50.60
-----
84.055 + 190.31 = 274.37 ( 62%)
-----
DRGC2 QCMRZZZ o Apr-Jul (ANIMAS - DURANGO) JR2: 0.970 # yrs: 30
AVG: 440.000 YTRANS: none
  
```

PSPC2	Coordinated	Model Computed	Comp. w/ Coord.	NWS Preferred.	Other Agency
R. Max	138.00 61%	169.70 75%		138.40 62%	
Most Prob.	110.00 49%	131.30 58%		100.00 44%	
R. Min	62.00 28%	92.90 41%		61.60 27%	



Nextreg – features

Eqn Output/Fcst input

Historical Max/Min appears when entering the edit box for Rmax and Rmin

RED MOUNTAIN PASS RMPC2/SWIRMZZ (May - Jun):

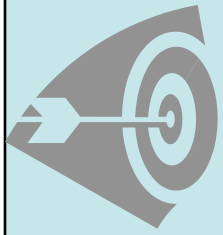
May	17.30Z	64%							
Jun	0.70Z	5%							
	18.00	44%	*	2.811	=	50.60			
				84.055	+	190.31	=	274.37	(62%)

Historical Min and Max
DRGC2 QCMRZZZ **39.54 (1977)** - DURANGO) JR2: 0.970 # yrs: 30
AVG: **455.62 (1941)** S: none

PSPC2	Coorainated	Model Computed	Comp. w/ Coord.	NWS Preferred.	Other Agency			
<input type="button" value="▲"/>	R. Max	138.00 61%	169.70 75%	<input type="text" value=""/>	%	138.40 62%	<input type="text" value=""/>	%
	Most Prob.	110.00 49%	131.30 58%	<input type="text" value=""/>	%	100.00 44%	<input type="text" value=""/>	%
<input type="button" value="▼"/>	R. Min	62.00 28%	92.90 41%	<input type="text" value=""/>	%	61.60 27%	<input type="text" value=""/>	%

Input Specification | **Eqn Output/Fcst Input** | Fcst Point Stats | Eqn Stats | Fcst Performance (Oper) | Fcst Performance (Calib) | Log

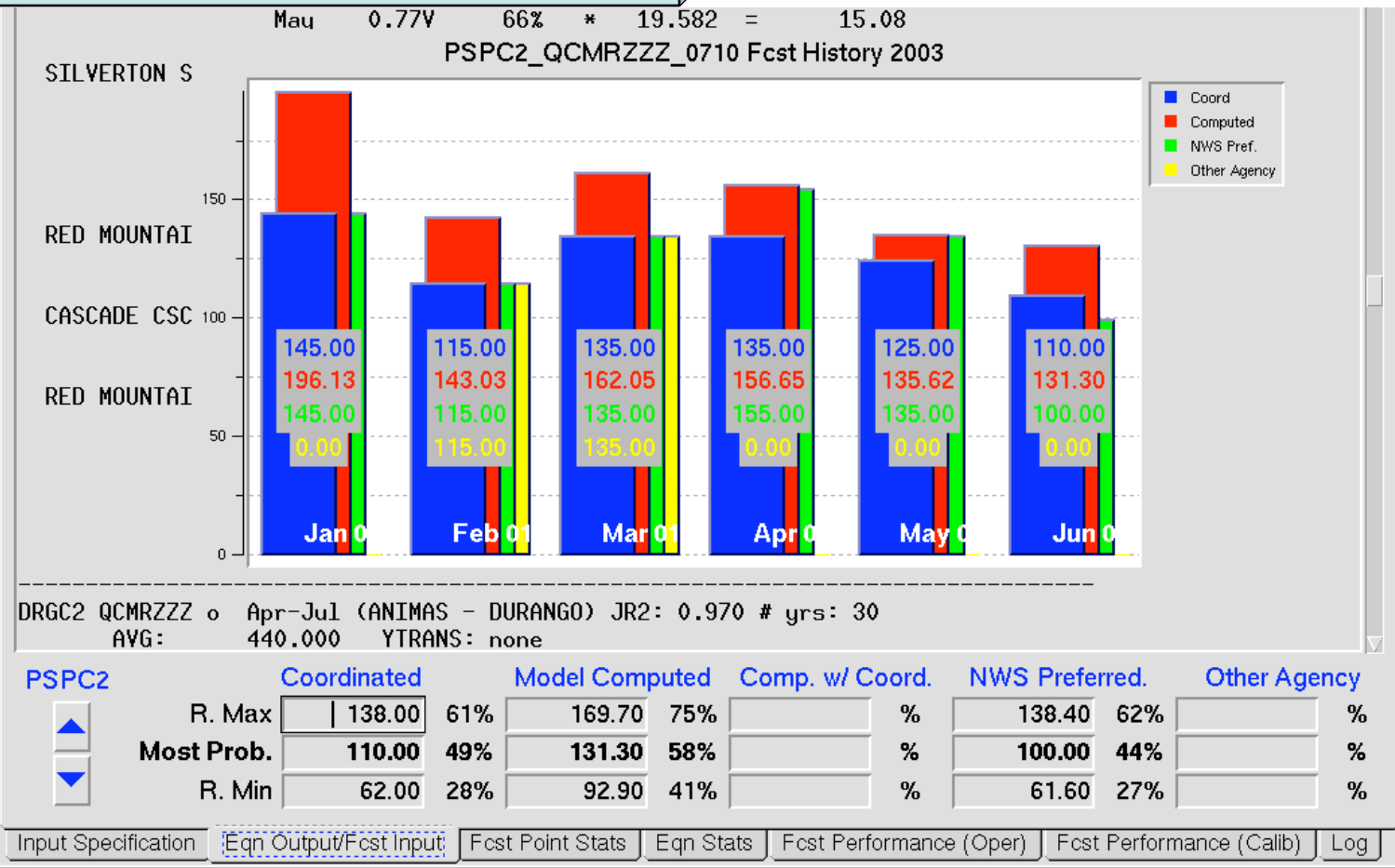
nextreg ver 2.2.4

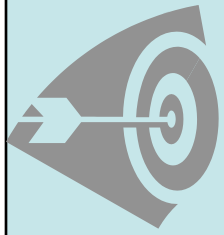


Nextreg – features

Eqn Output/Fcst input

Current forecast history appears when editing the NWS Most Probable number





Nextreg – features

Eqn Output/Fcst input

Per variable X-Y scatterplot activated with right mouse click. Green lines are averages. Zoomable

```

SILVERTON SLVC2/PPMRZZZ (Apr - May):
    Apr  1.17V  71%
    May  1.03V  55%
-----
          2.20  63%  *

RED MOUNTAIN PASS RMPC2/SWIRMZZ
    May  17.30Z  64%  *

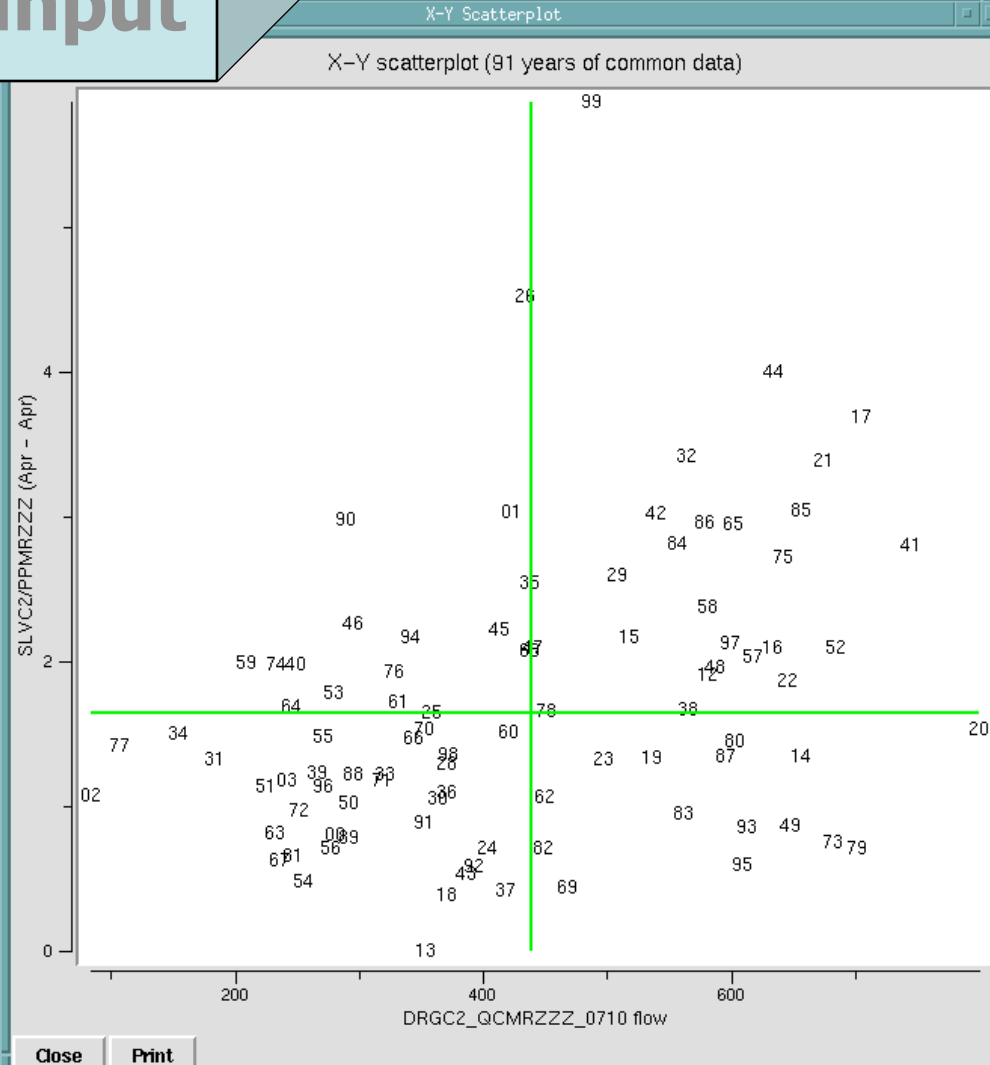
CASCADE CSCC2/SWIRMZZ
    May  0.60Z  16%  *

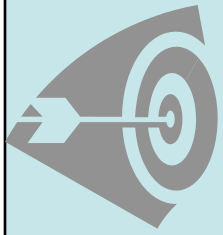
RED MOUNTAIN PASS RMPC2/SWIRMZZ (May - Jun)
    May  17.30Z  64%
    Jun  0.70Z   5%
-----
          18.00  44%  *
    84

DRGC2 QCMRZZZ o Apr-Jul (ANIMAS - DURANGO)
    AVG:  440.000  YTRANS: none
  
```

PSPC2	Coordinated	Model C
R. Max	138.00 61%	169.
Most Prob.	110.00 49%	131.
R. Min	62.00 28%	92.

Input Specification Eqn Output/Fcst Input Fcst Point Stat





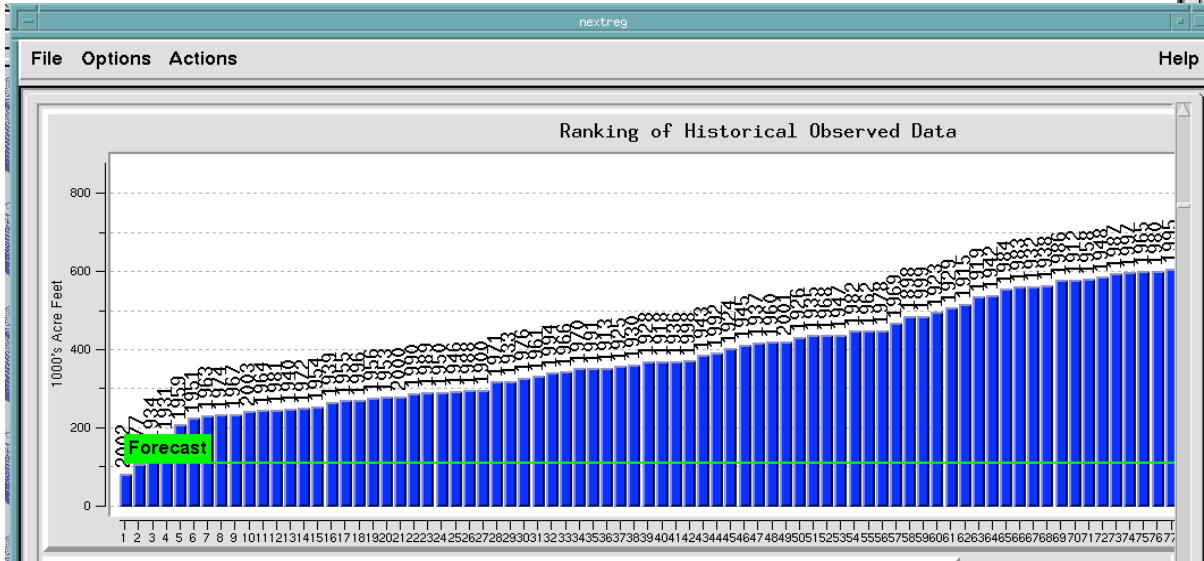
Nextreg – features

Fcst Point Stats tab

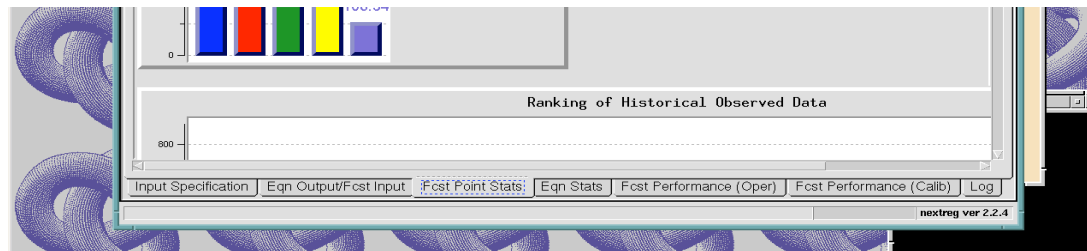
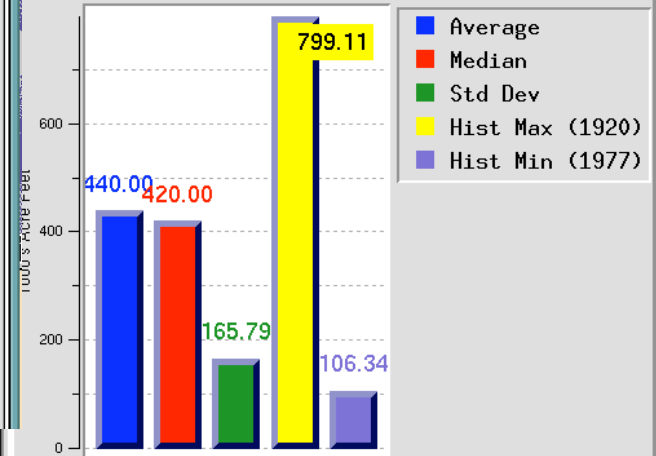
DRGC2 QCMRZZZ

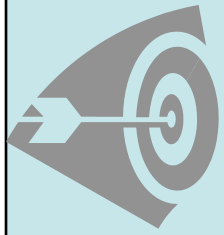
61-90 Average: 440.00
Median: 420.00
Std Dev: 165.79

Hist Max: 799.11(1920)
Hist Min: 106.34(1977)



Ranking of Historical Observed Data Distribution Stats





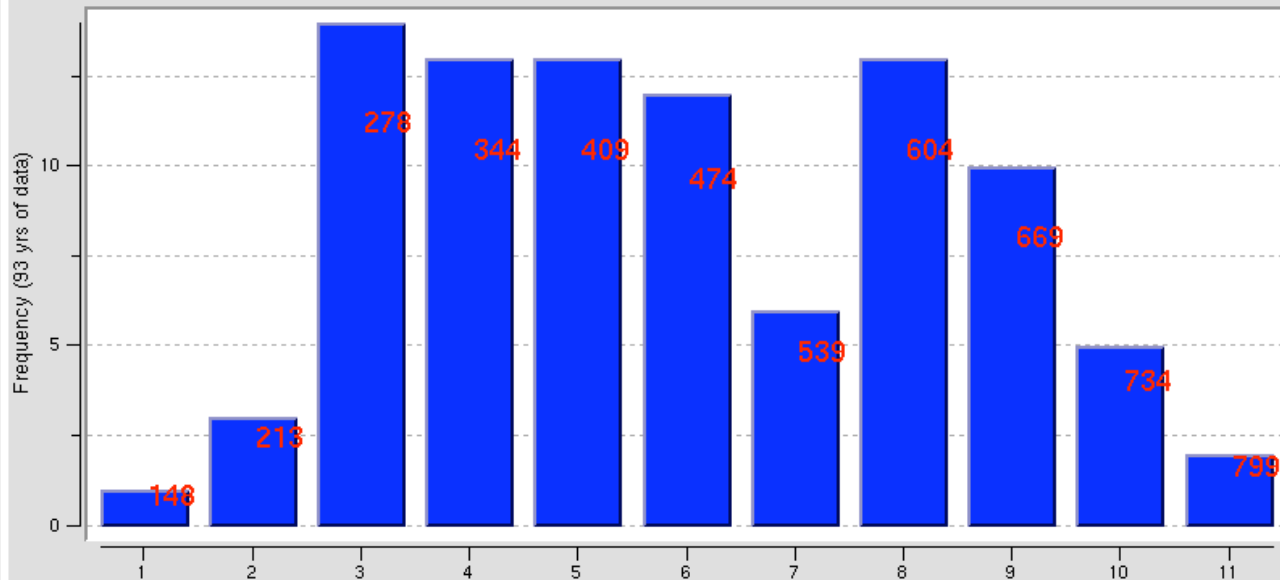
Nextreg – features

Fcst Point Stats tab

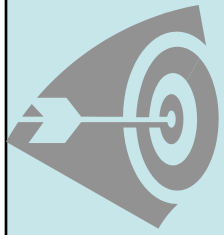
nextreg

File Options Actions

Histogram-Distribution of Historical Observed Data



1	2002	83.27
2	1977	106.3933
3	1934	154.254
4	1931	183.6686
5	1959	208.5734



Nextreg – features

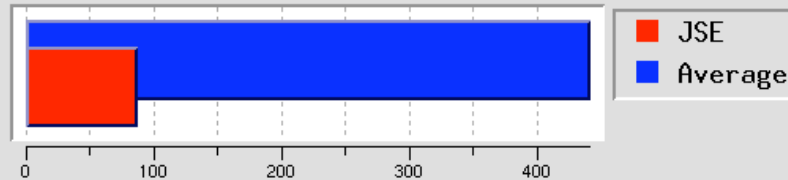
Eqn Stats tab

nextreg

DRGC2_QCMRZZZ_0710 P

Calibration JSE: 86.07
JR2: 0.74
R2: 0.77
n: 30
per: 71-00

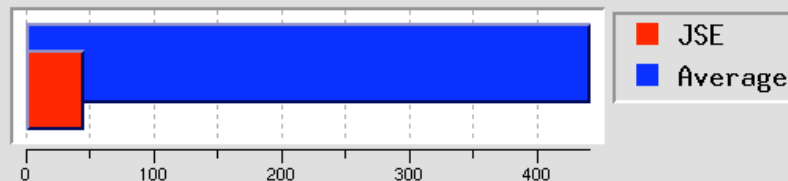
JSE vs. AVG

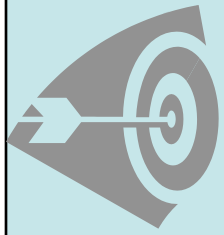


DRGC2_QCMRZZZ_0710 o

Calibration JSE: 44.28
JR2: 0.97
R2: 0.97
n: 30
per: 61-90

JSE vs. AVG





Nextreg – features

Fcst Perf (Oper) tab

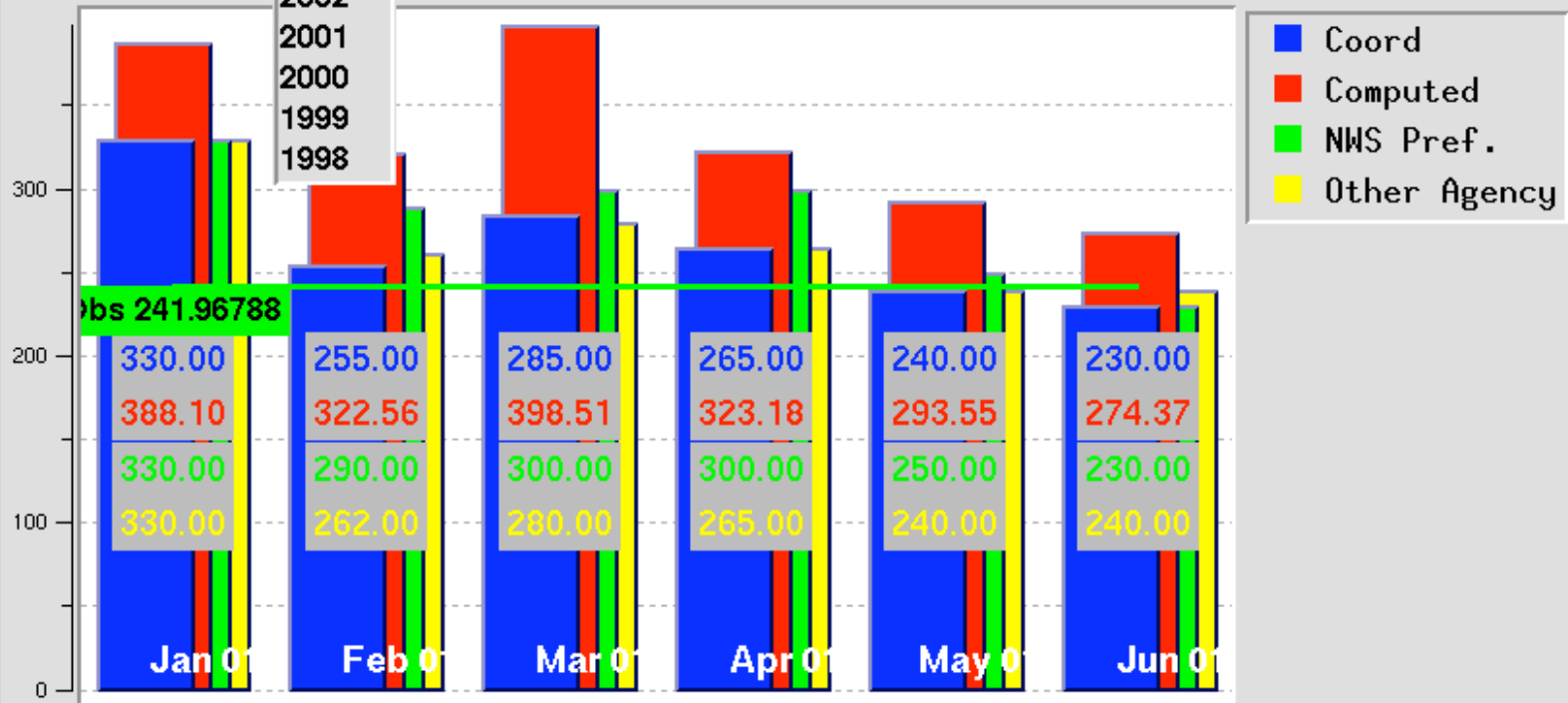
Choose Year:

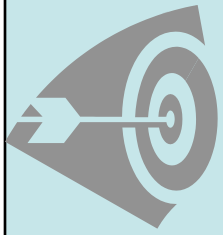
- 2003
- 2002
- 2001
- 2000
- 1999
- 1998

◆ Replace graph

◇ Accumulate graphs

2_QCMRZZZ_0710 Fcst History 2003



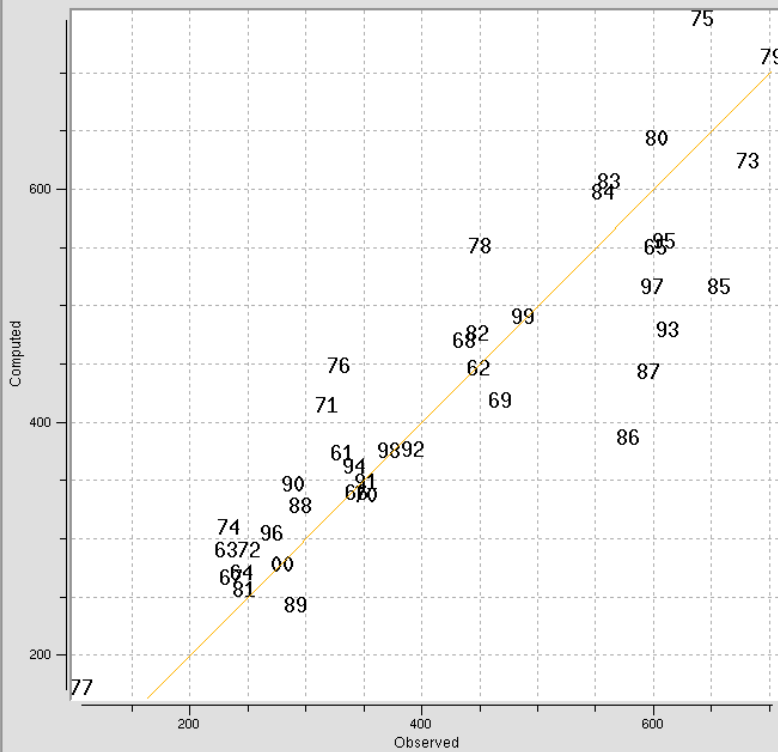


Nextreg – features

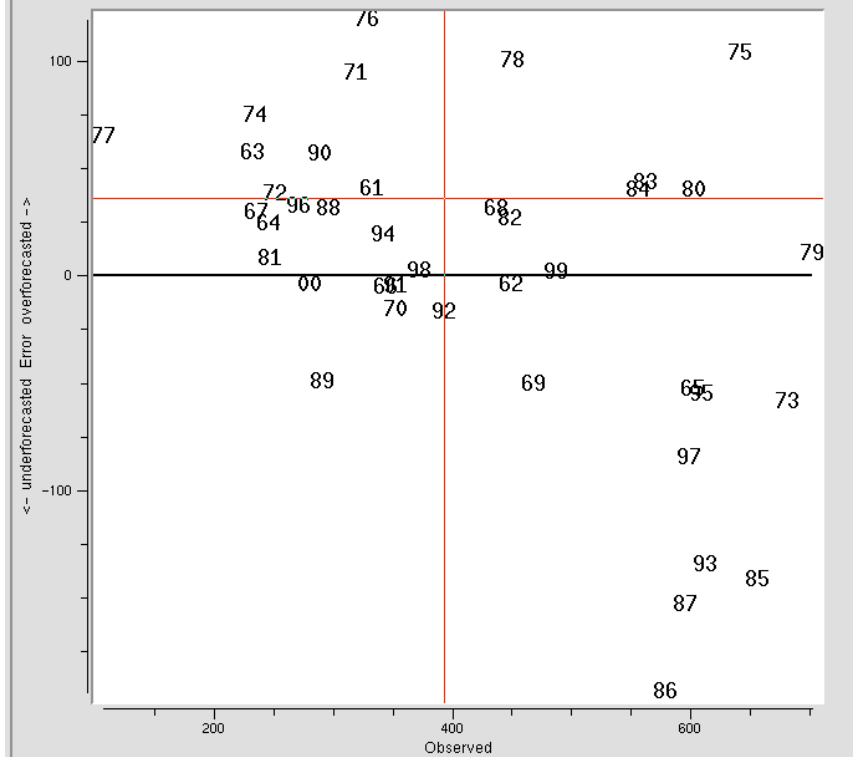
Fcst Perf (Cal) tab

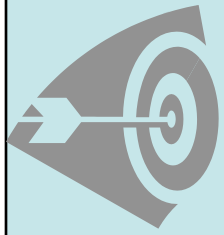
File Options Actions

DRGC2_QCMRZZZ_0710 Primary Eqn Calibration Performance



DRGC2_QCMRZZZ_0710 Primary Eqn Calibration Performance residuals vs. observed

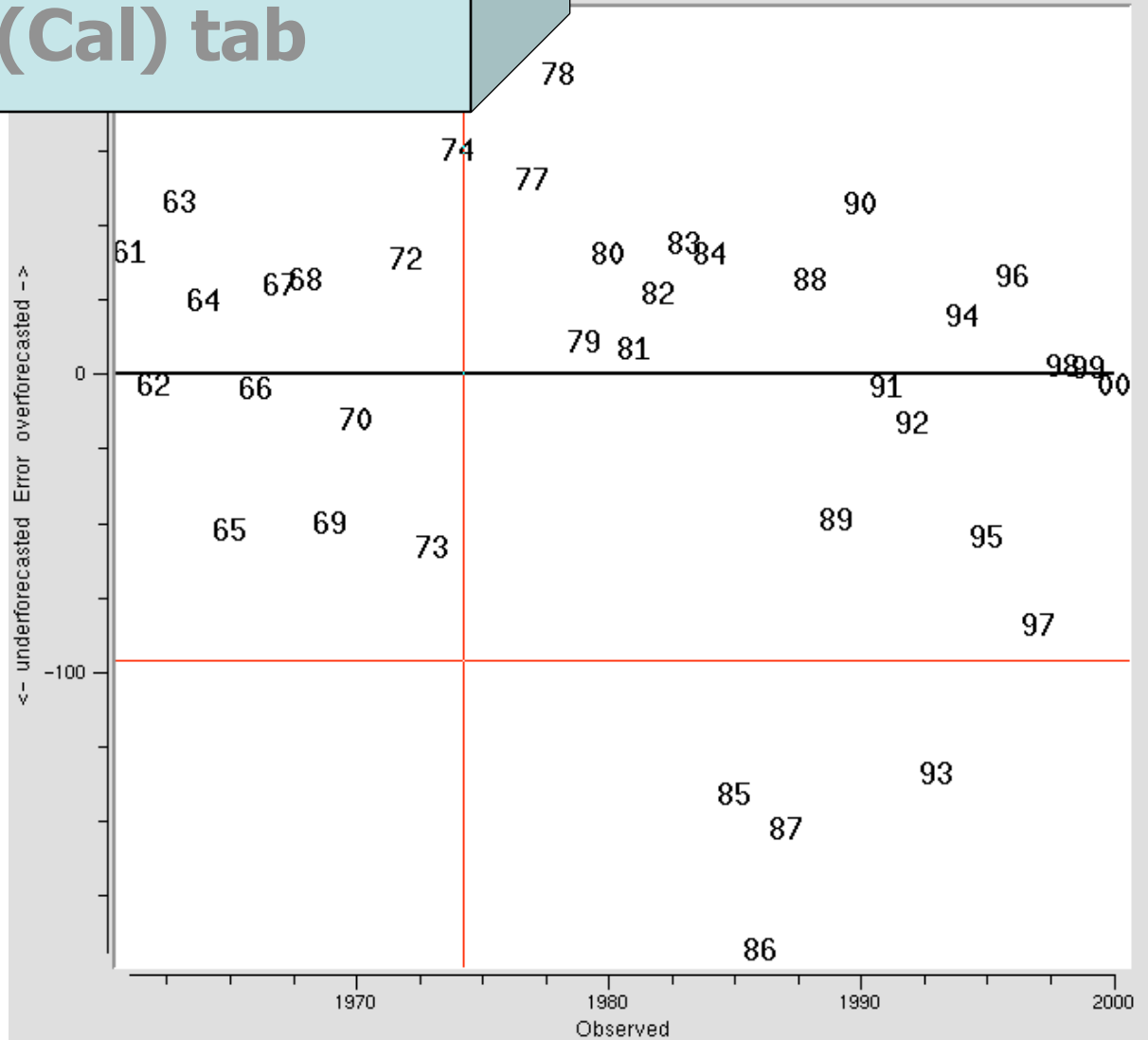




Nextreg – features

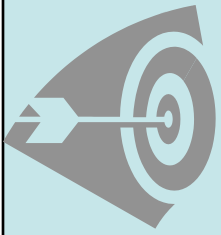
Fcst Perf (Cal) tab

Eqn Calibration Performance
vs. water year





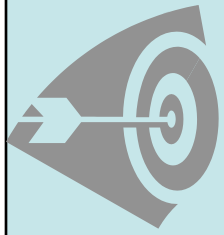
- ⊕ Continue trend of reducing number of programs
- ⊕ GIS ground has been broken with Regcand; continue this approach
- ⊕ Revise database table structures
- ⊕ Get others involved in programming/support
- ⊕ Investigate/incorporate newer statistical techniques and new ways of looking at verification



SWS

Why should I use it?

- ⊕ A package of beginning-to-end integrated programs for water supply forecasting, or really, any kind of statistical forecasting
- ⊕ Monthly data stored in relational database
- ⊕ Other programs that deal with data of a monthly time step
- ⊕ Another way to forecast volume, in addition to ESP
- ⊕ Ease of use
- ⊕ Software has been polished by a lot of “wouldn’t it be nice if...”s



SWS – It is of limited value if...

- ⊕ There is no dominant driving force (like snowmelt)
- ⊕ There is not a substantial period of record of data i.e. 30+ years
- ⊕ **The predictand data set does not closely approximate natural flow**
- ⊕ The predictors used in the equations are not recorded early enough in the month
- ⊕ The predictors used in the equations are not recorded reliably month to month and year to year
- ⊕ The ability to “time distribute” the forecast volume is required

Climate Change Forecasts as Input to CBRFC's Water Supply Models

- Historical MAP and MAT time series are adjusted relative to current climate outlooks before they are used as input into ESP.
 - The probability anomaly shift described by the CPC forecasts must be translated into real physical temperature or precipitation.
 - This is done through the simple method of computing the shift in the medians of the distributions of the temperature and precipitation based on the CPC forecasts and then using this shift to adjust all values of the time series.
 - temperature uses an additive adjustment
 - precipitation uses a multiplicative adjustment

Multi-model Multi-agency Forecast Process

