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?

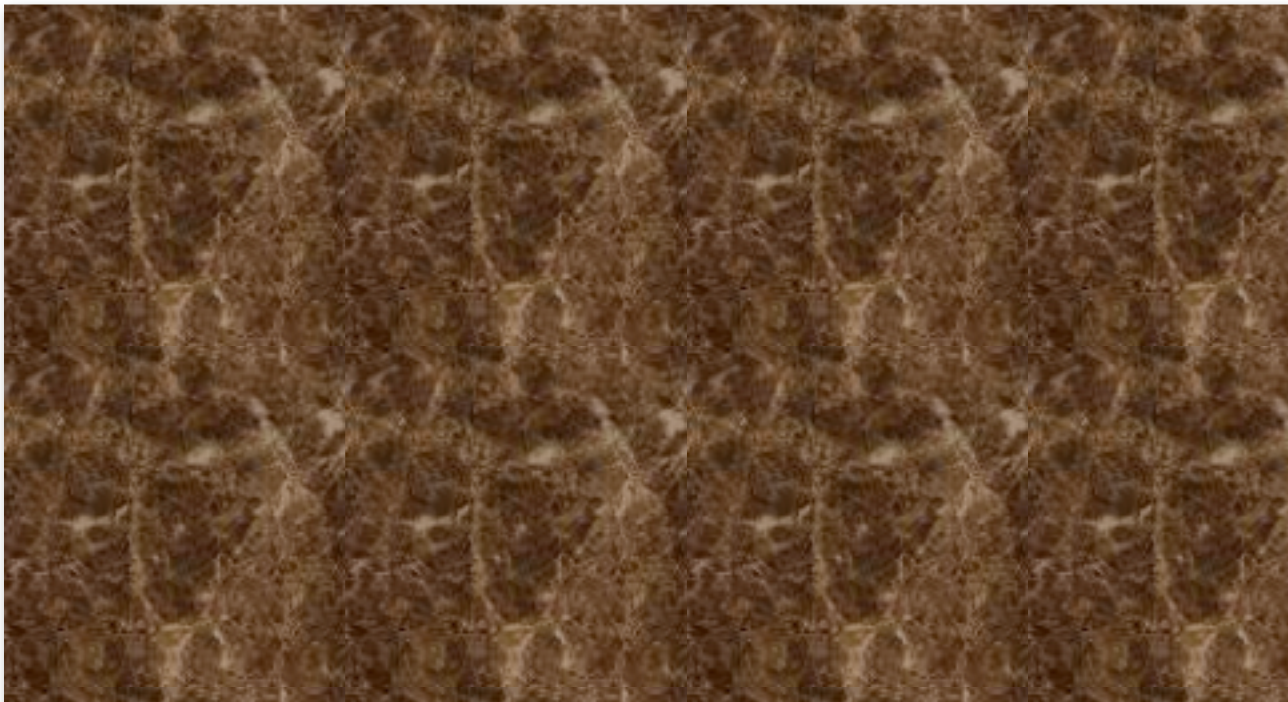
Why is it a good thing?

- 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

SWS

I. Calibration

II. Operation



SWS – Calibration Programs

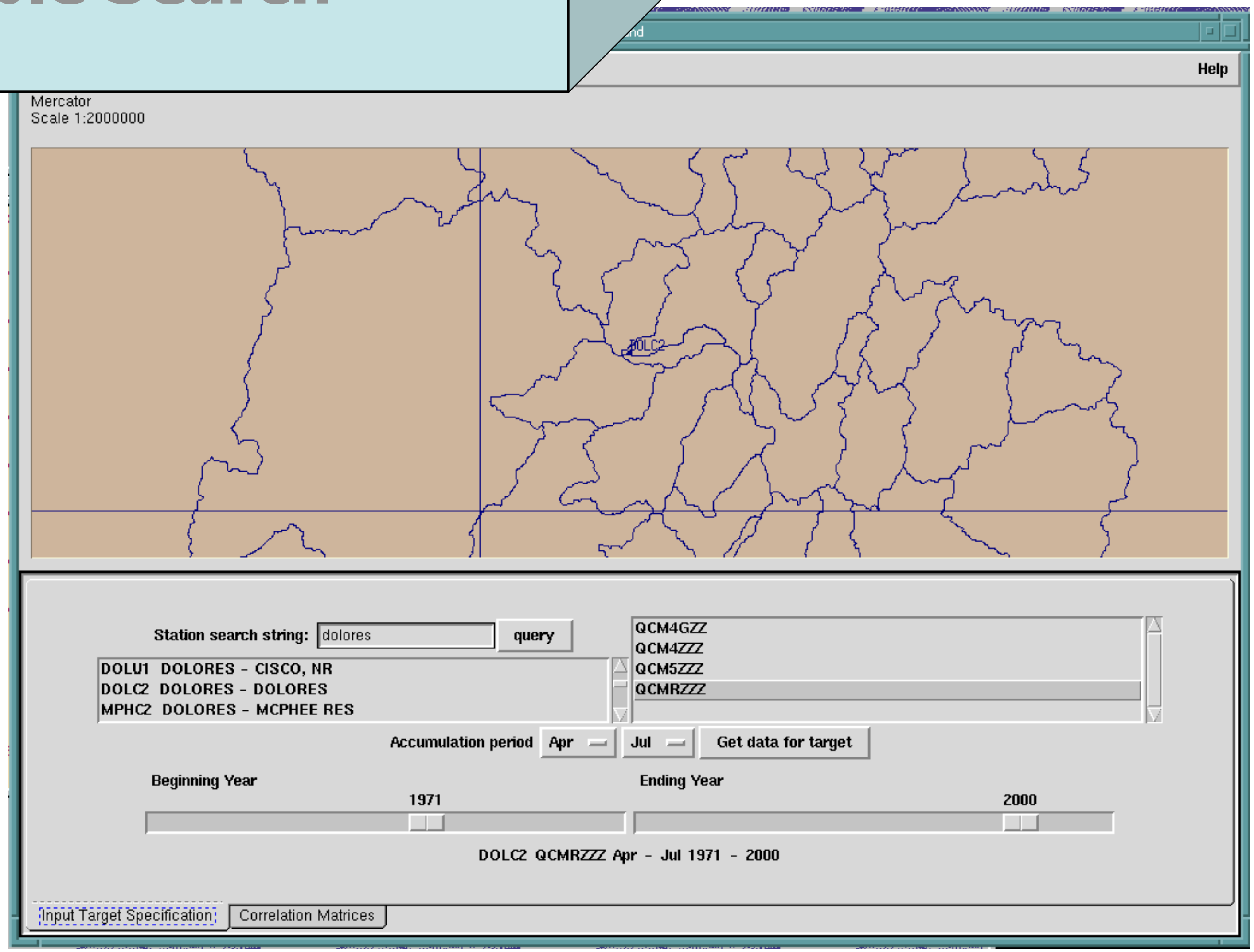
- **Regcand** – find candidate variables, calculate correlation matrix
- **Regdata** – prepare input file for regcomb
- **Regcomb** – find optimal equations (top 20)
- **Epal** – Evaluate, Pick and Load equations to the postgres db structure
- **Cstats** – Calculate and store period statistics

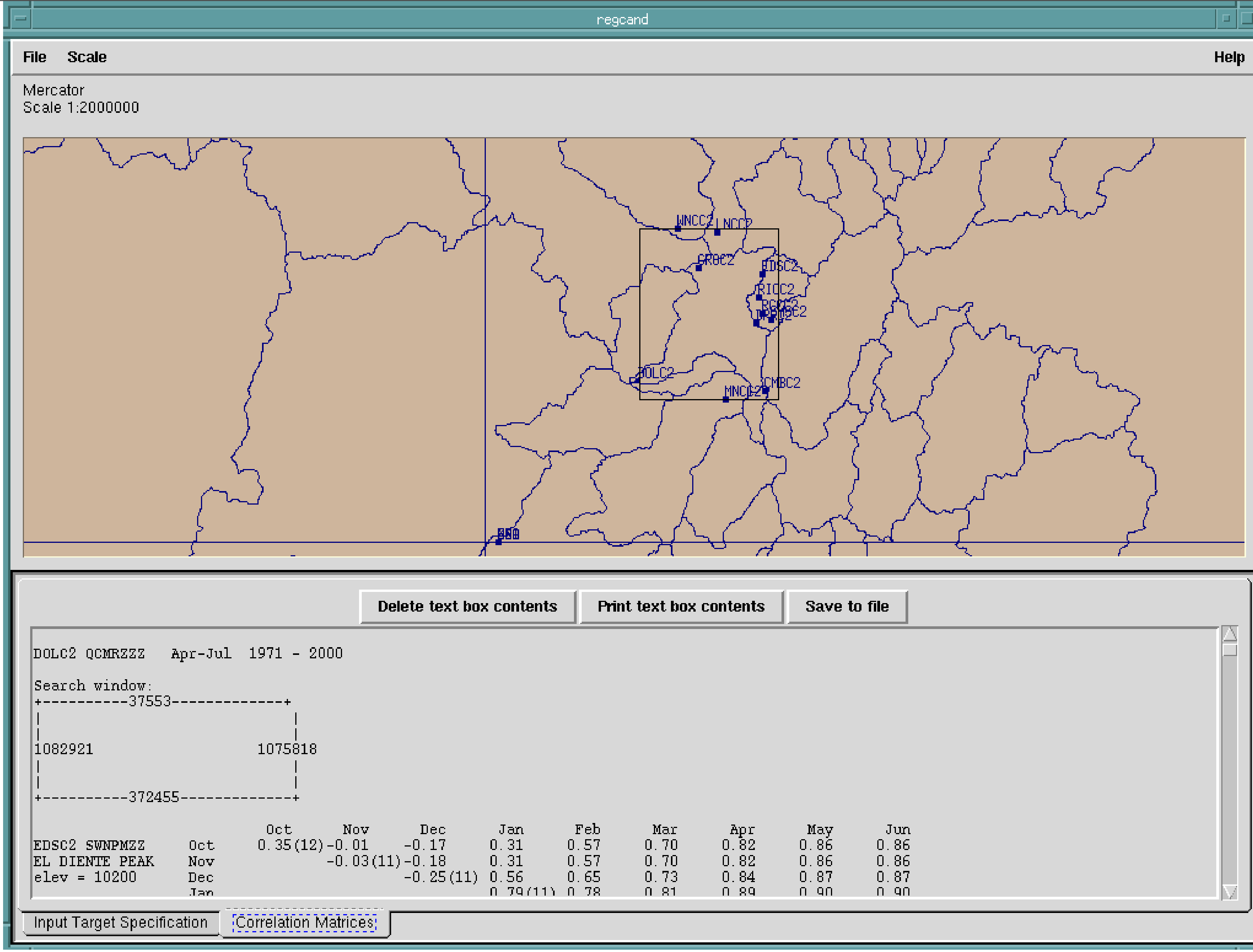
- Equation # 3
- Y 1 DOLC2/QCMRZZZ,Ap-Jl,DOLORES - DOLORES
- = -5.203
- \+ 14.859 X4 HSPC2/PPMRZZZ,No-Ja,FORT LEWIS
- \+ 57.175 X5 DRRC2/QCMRZZZ,Ja,DOLORES - RICO, BLO
- \+ 10.161 X6 LNCC2/SWIRMZZ,Fe,LONE CONE
- Number of observations used = 30
- Number of principal components used = 1
- CORRELATION 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.)

YEAR	OBSERVED	COMPUTED	ERROR	JACKKNIFE COMPUTED	JACKKNIFE 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.

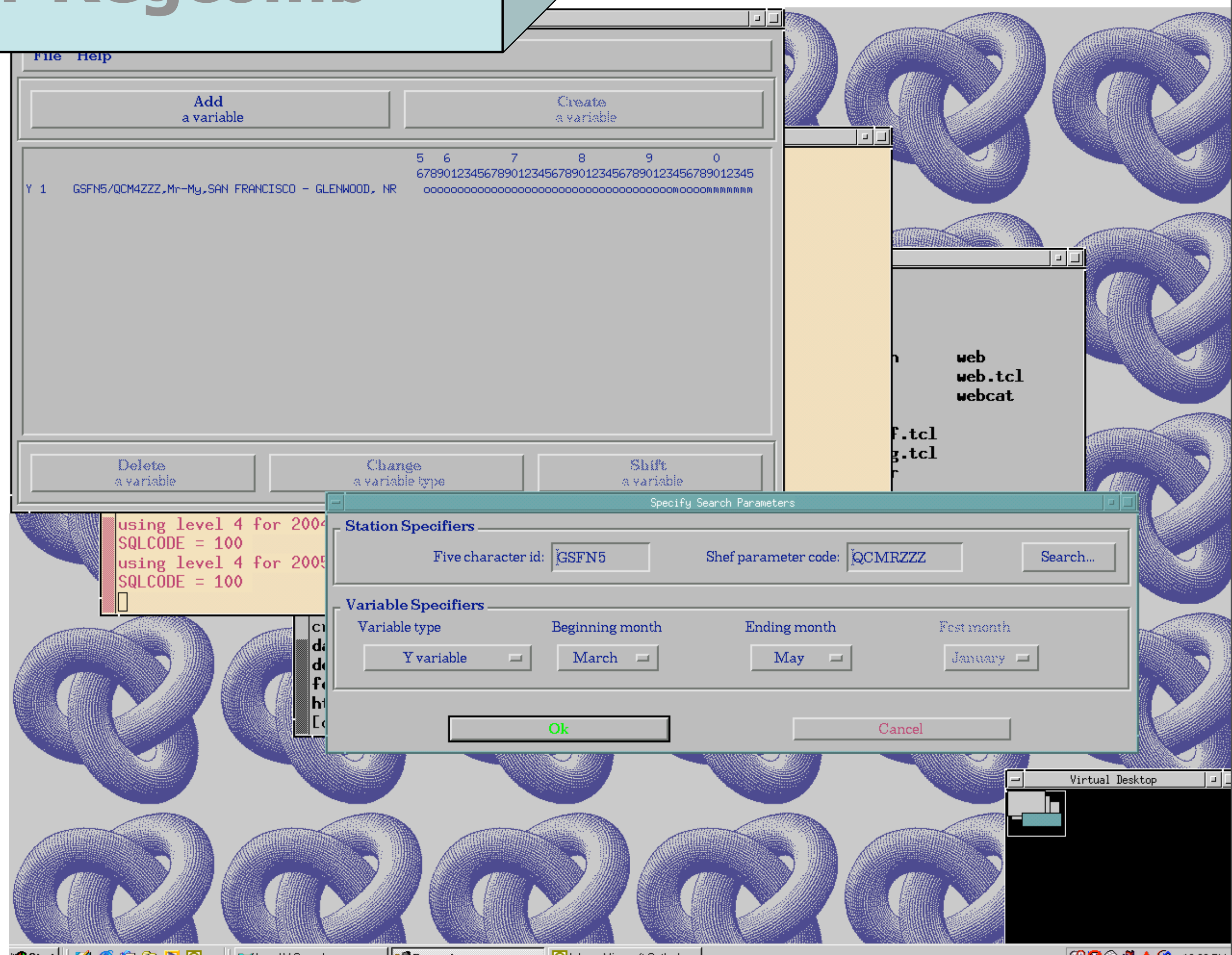




Future software development will include GIS where it is useful

Regdata – Assemble Input for Regcomb

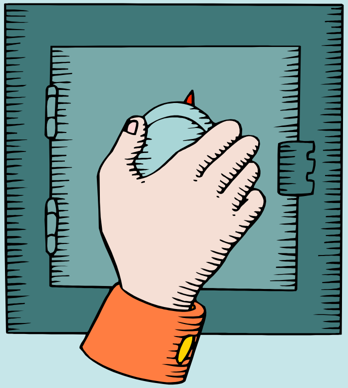
Extracts data from postgresql, accumulates as needed, builds formatted input file for Regcomb



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. $y = mx_1 + mx_2 + mx \ldots + b$
2. $y = mx_1 + mx_2 + mx \ldots + b$
3. ...

REGCOMB

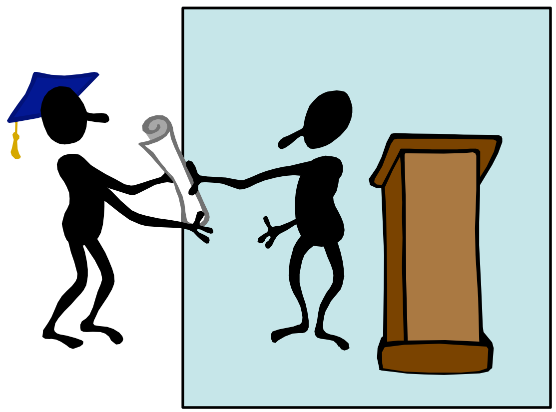
Jack-knife Testing

+

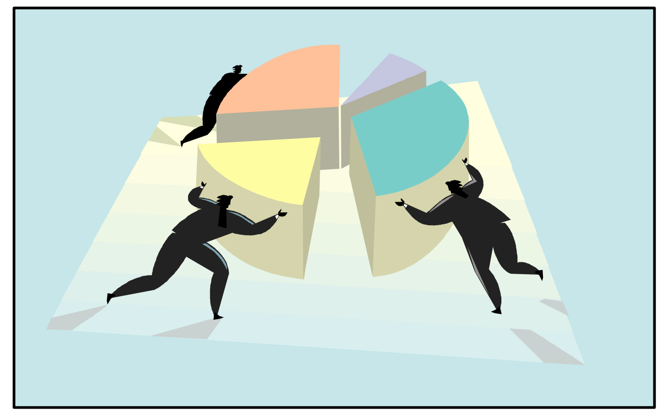
For a given set of predictors, one observation (one year) is deleted from the data set. Optimal coefficients are determined. The equation is then measured as to how well it predicted the selected year. The idea here is to simulate how well the equation will perform in an operational environment where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

REGCOMB

Principal Components



+

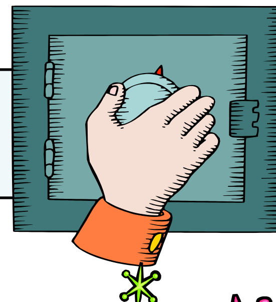


Variables in a water-supply equation tend to have high correlation with each other. This causes problems when trying to determine optimal coefficients via traditional regression techniques. Principal components analysis is a way to determine optimal coefficients while recognizing and addressing the intercorrelation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

REGCOMB

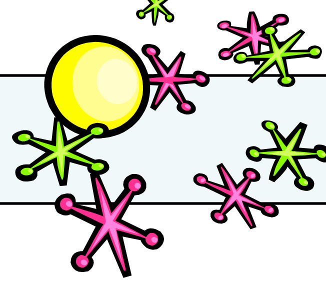
It's all good...

Combination Analysis

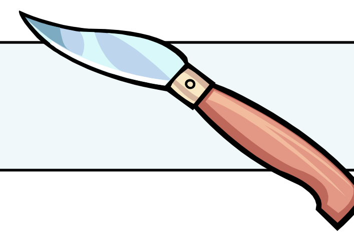


+

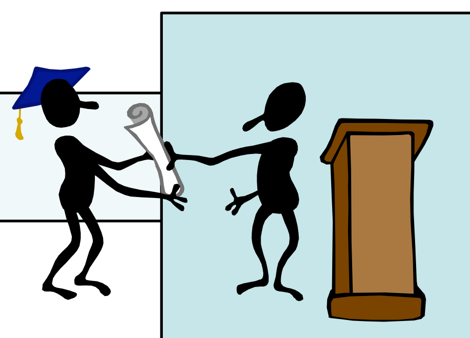
Jack-knife error computation



+



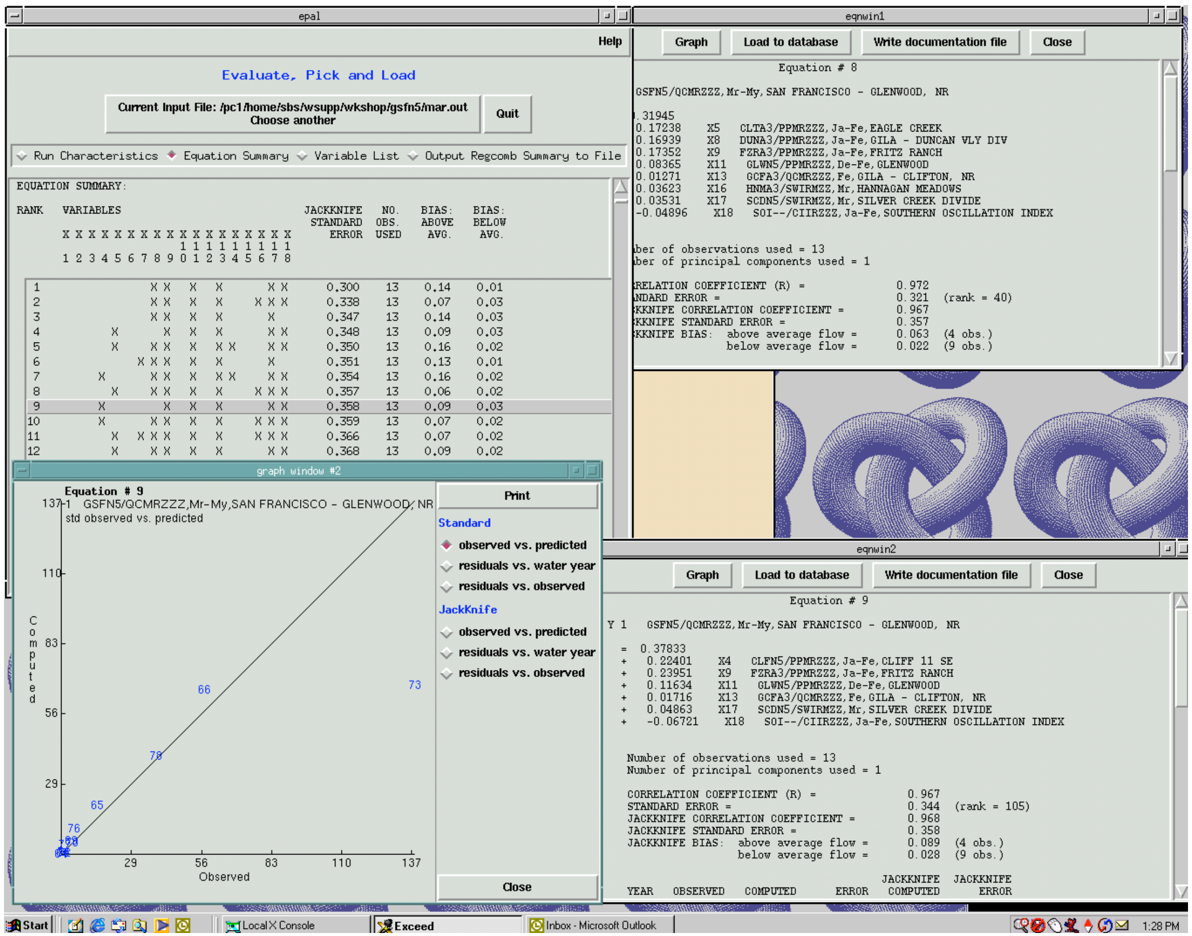
Principal Components



+

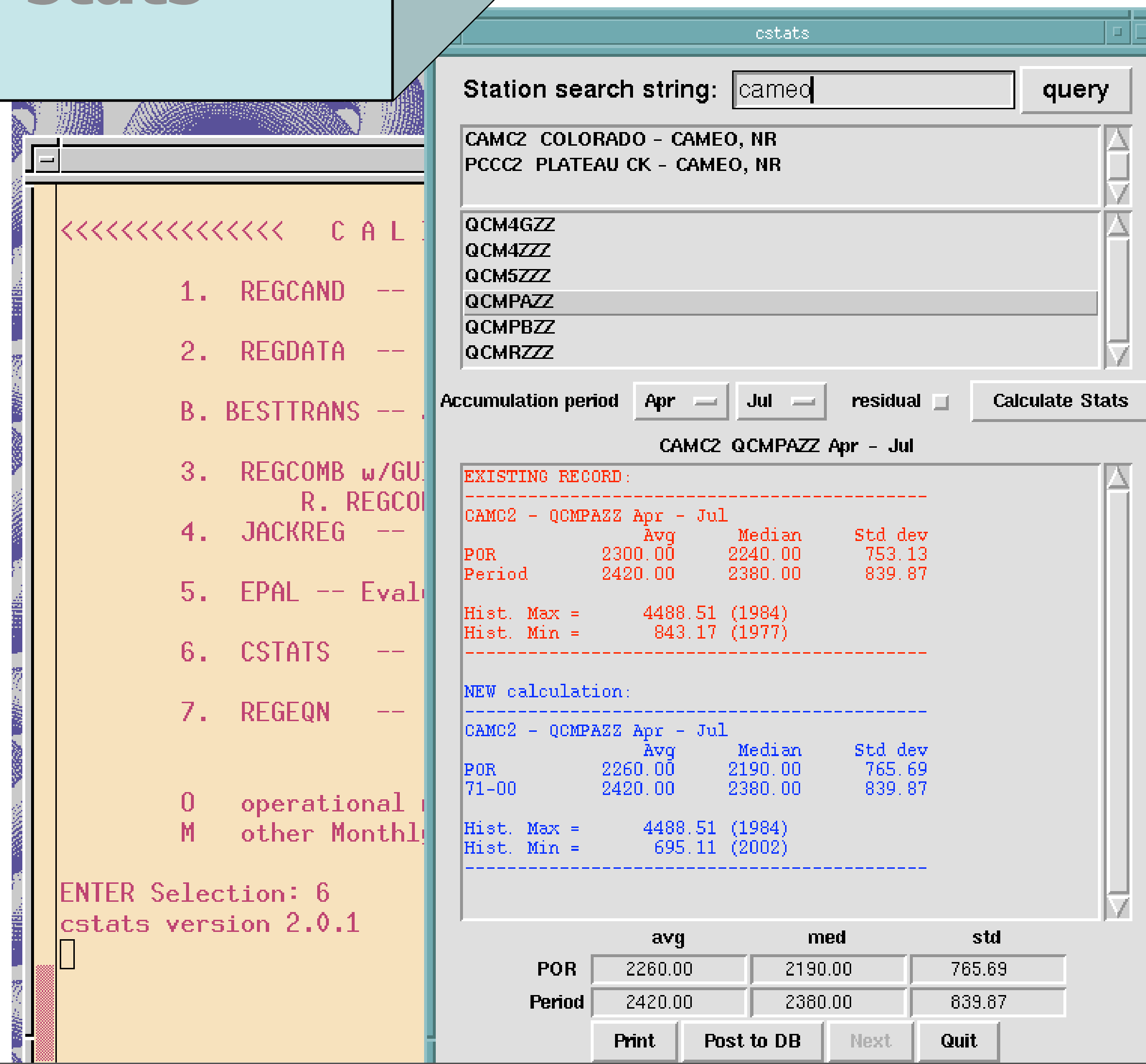
Epal – Evaluation, Pick and Load

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.



Cstats – Calculate and store period stats

Most commonly used to calculate Apr-Jul average, median, and standard deviation for both the conventional “normal” period, currently 1971-2000, as well as the entire period of record.

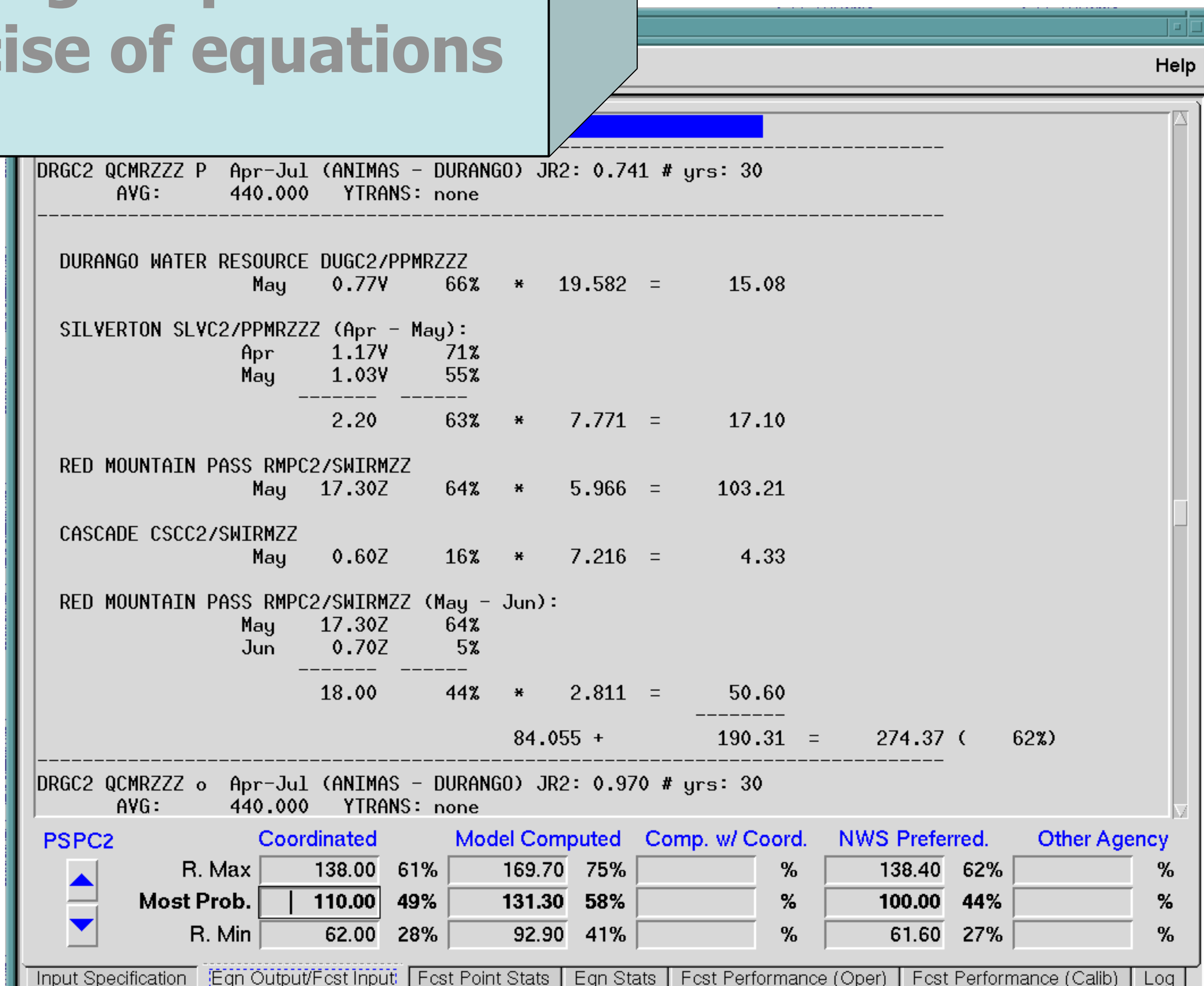


SWS – Operational Programs

- **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

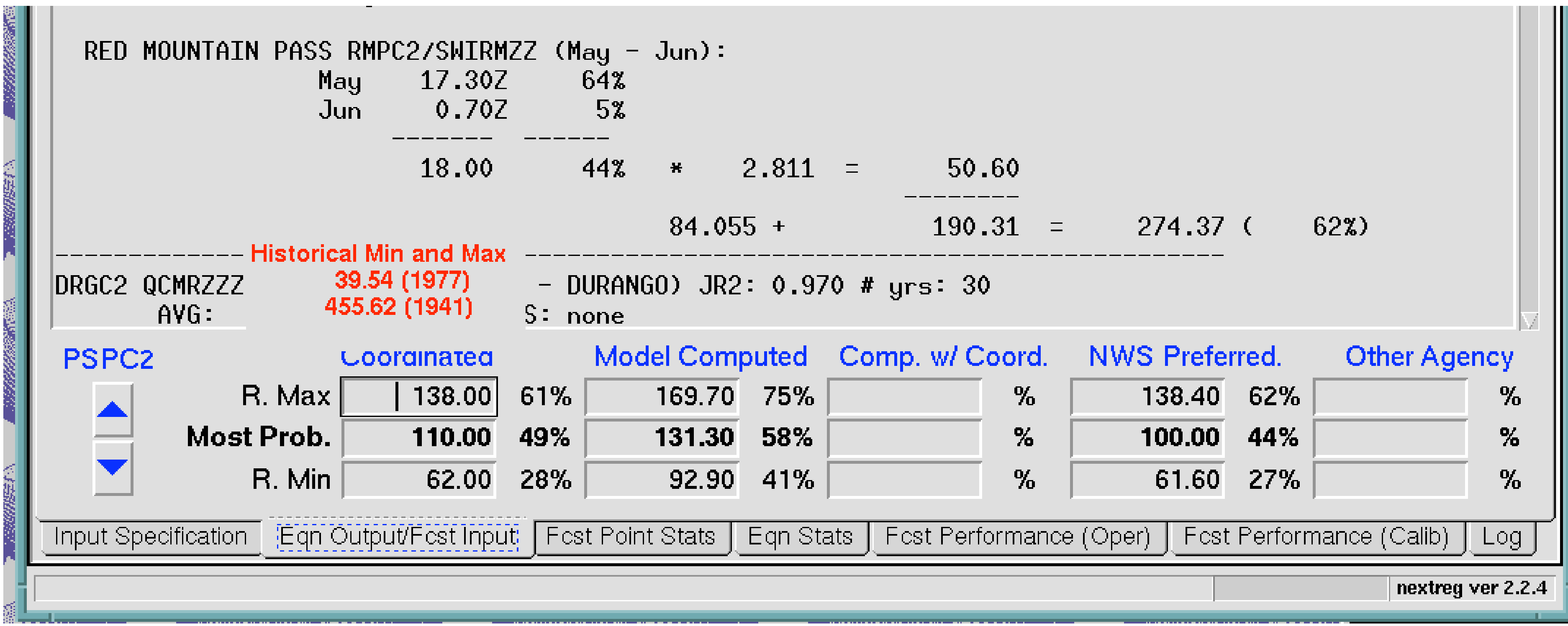
Equation Output/ Fcst Input tab – spreadsheet style output of primary and secondary equations



Nextreg – features

Eqn Output/Fcst input

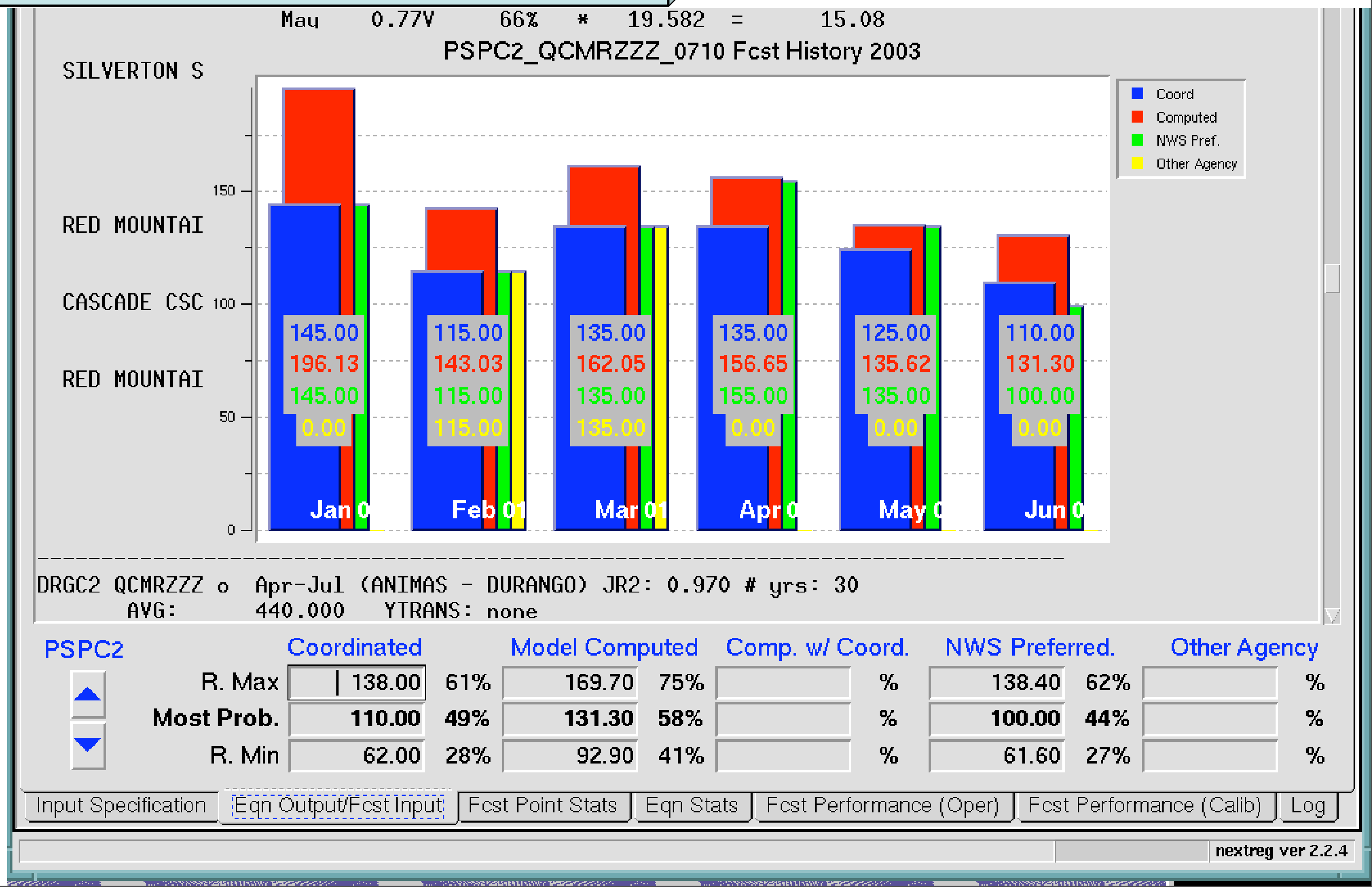
Historical Max/Min appears when entering the edit box for Rmax and Rmin



Nextreg – features

Eqn Output/Fcst input

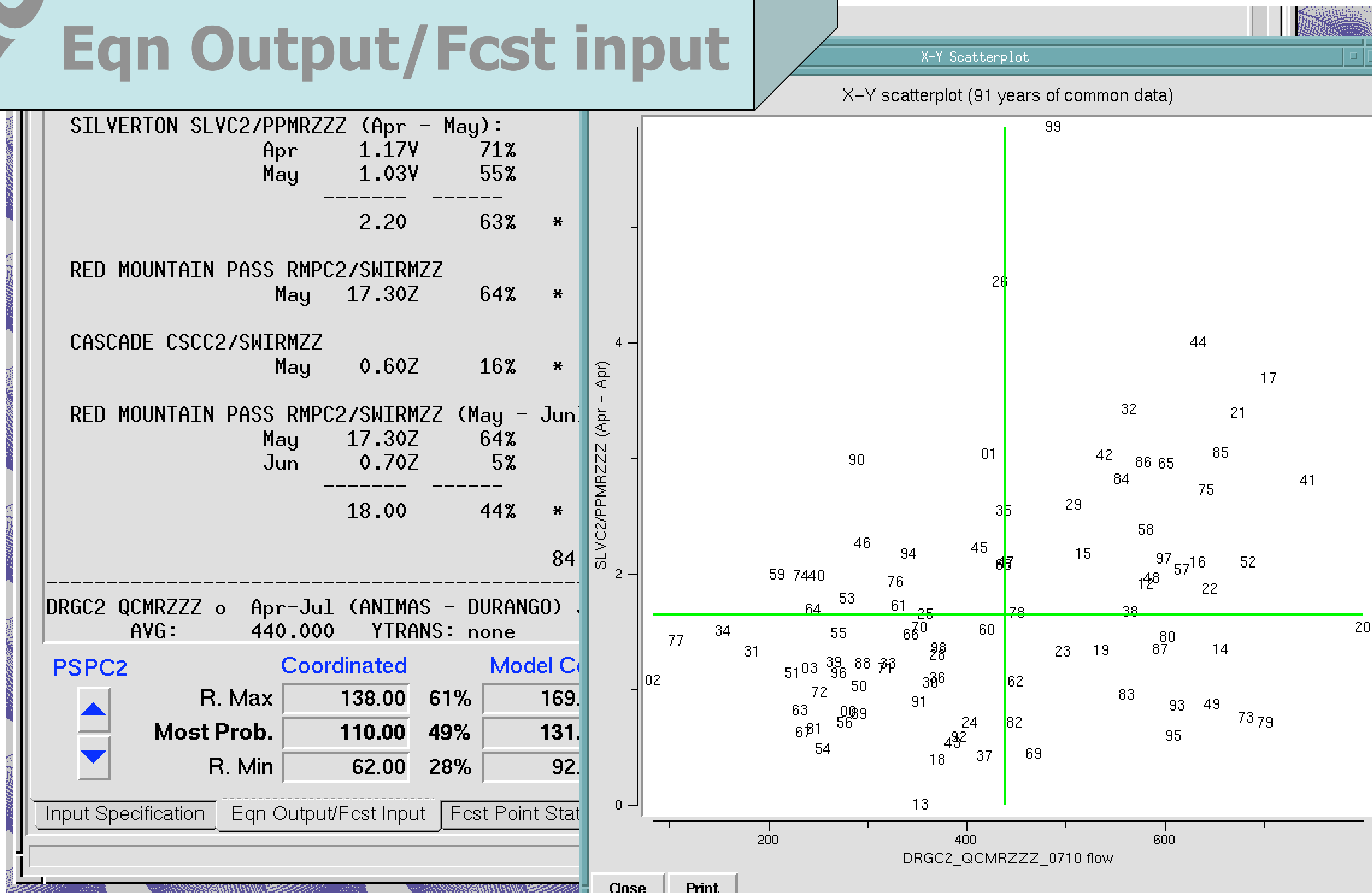
Current forecast history appears when editing the NWS Most Probable number



Nextreg – features

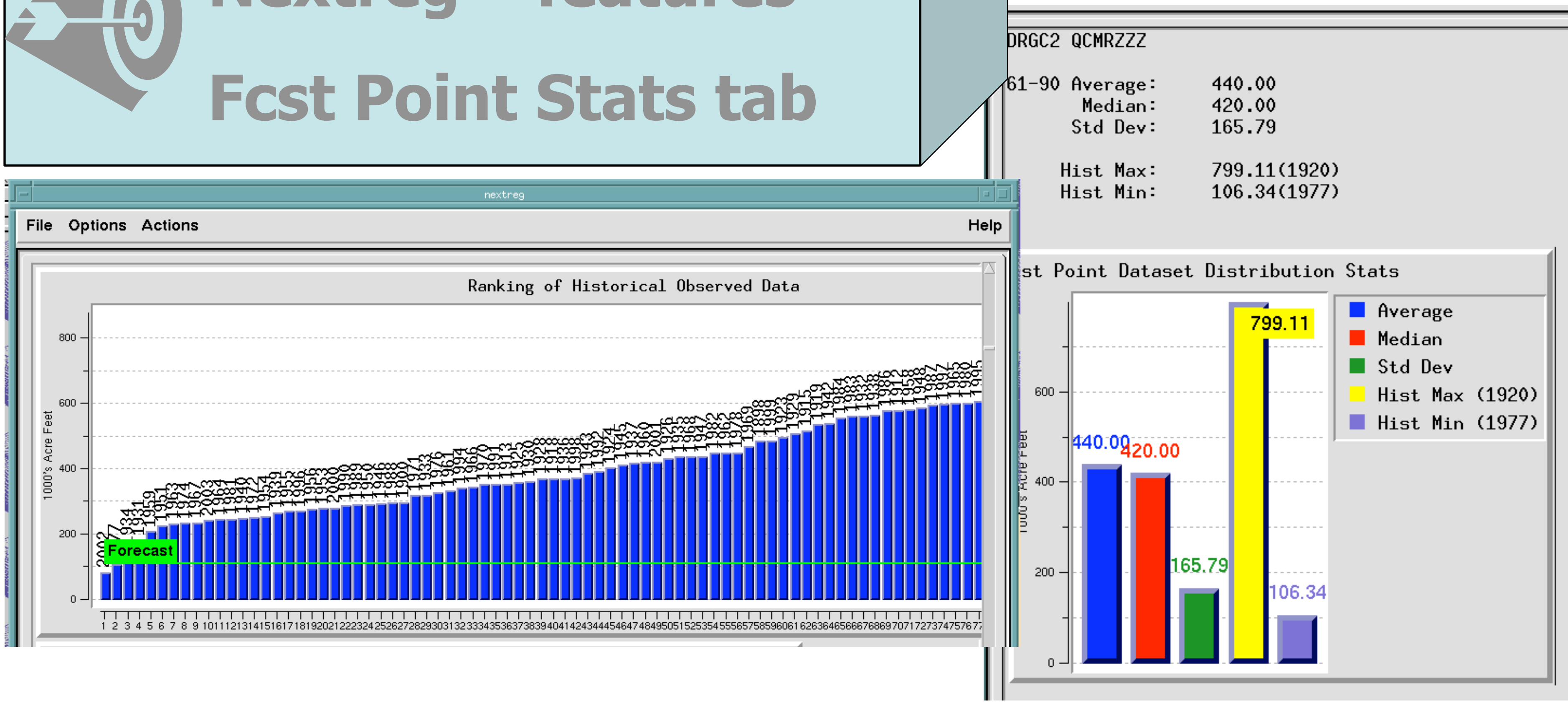
Eqn Output/Fcst input

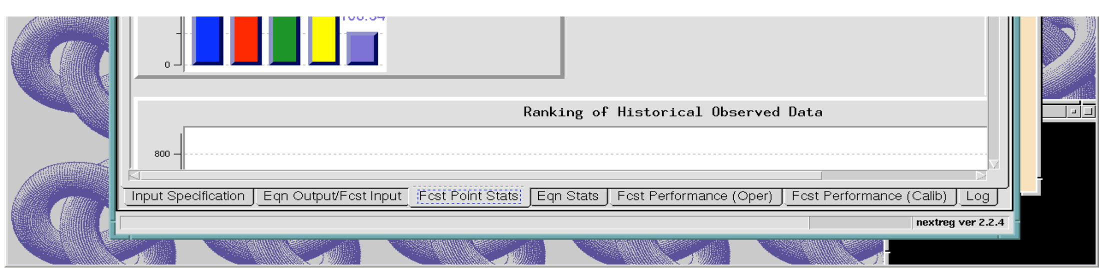
Per variable X-Y scatterplot activated with right mouse click. Green lines are averages. Zoomable



Nextreg – features

Fcst Point Stats tab



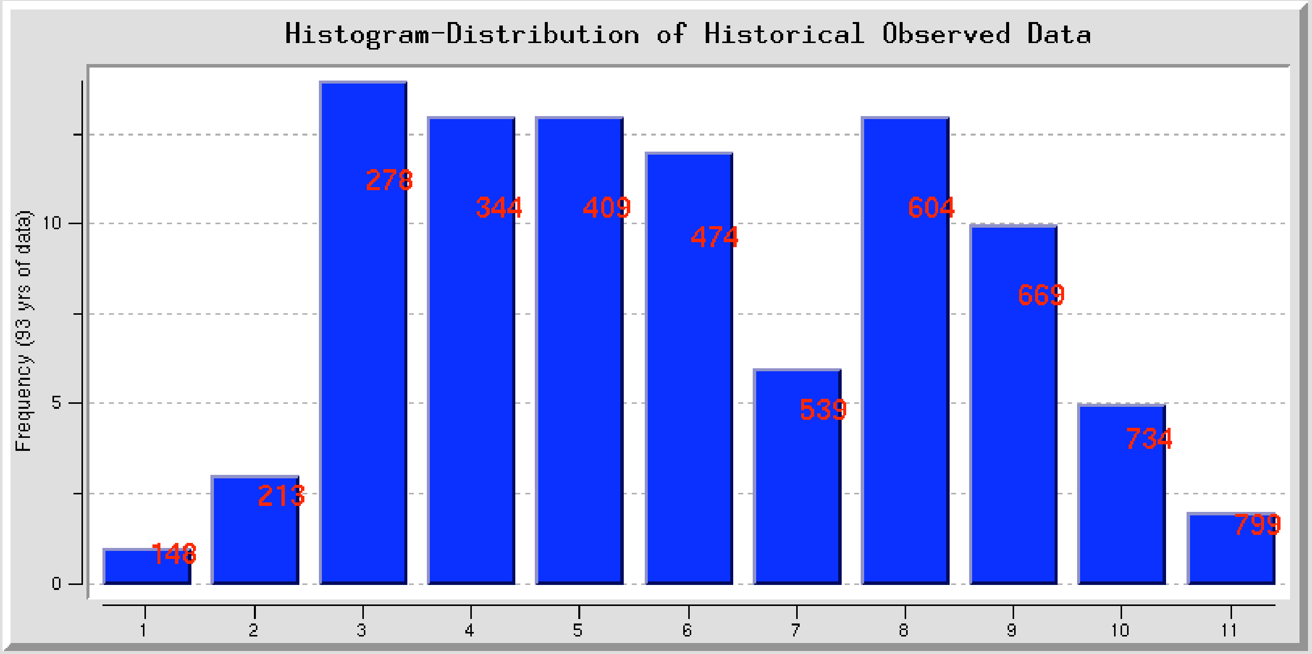


Nextreg – features

Fcst Point Stats tab

nextreg

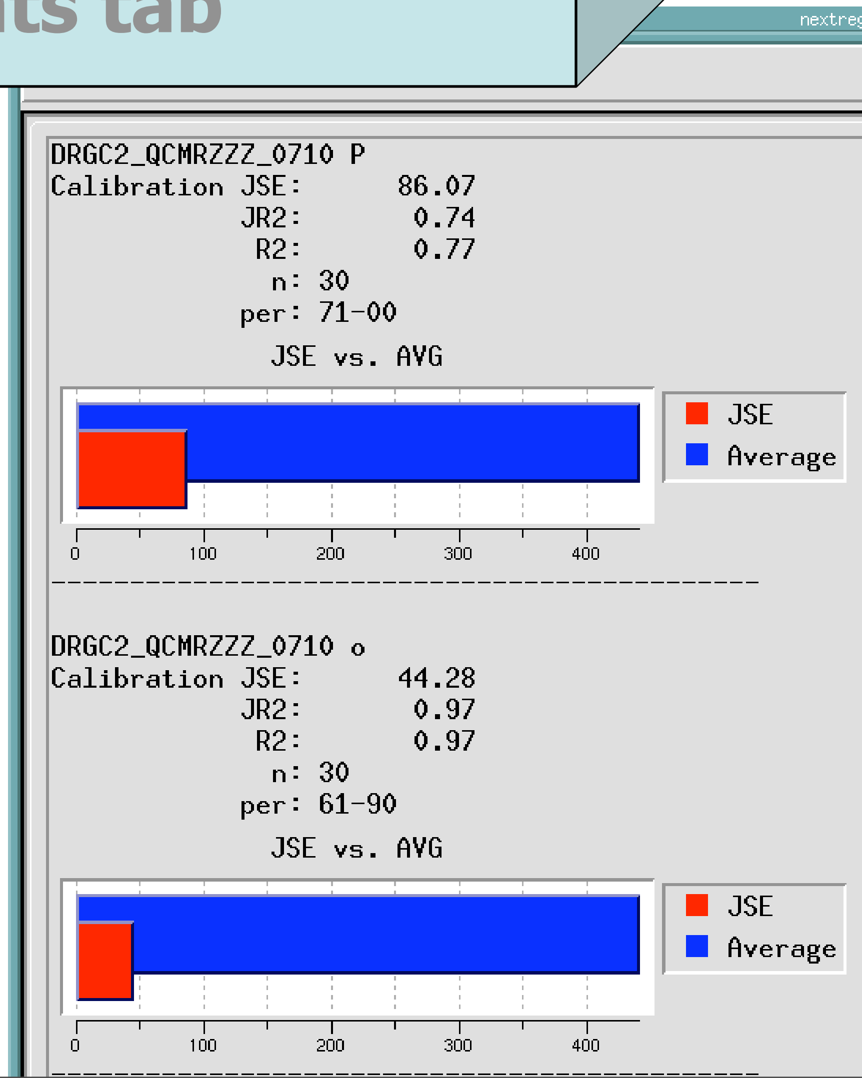
File Options Actions



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

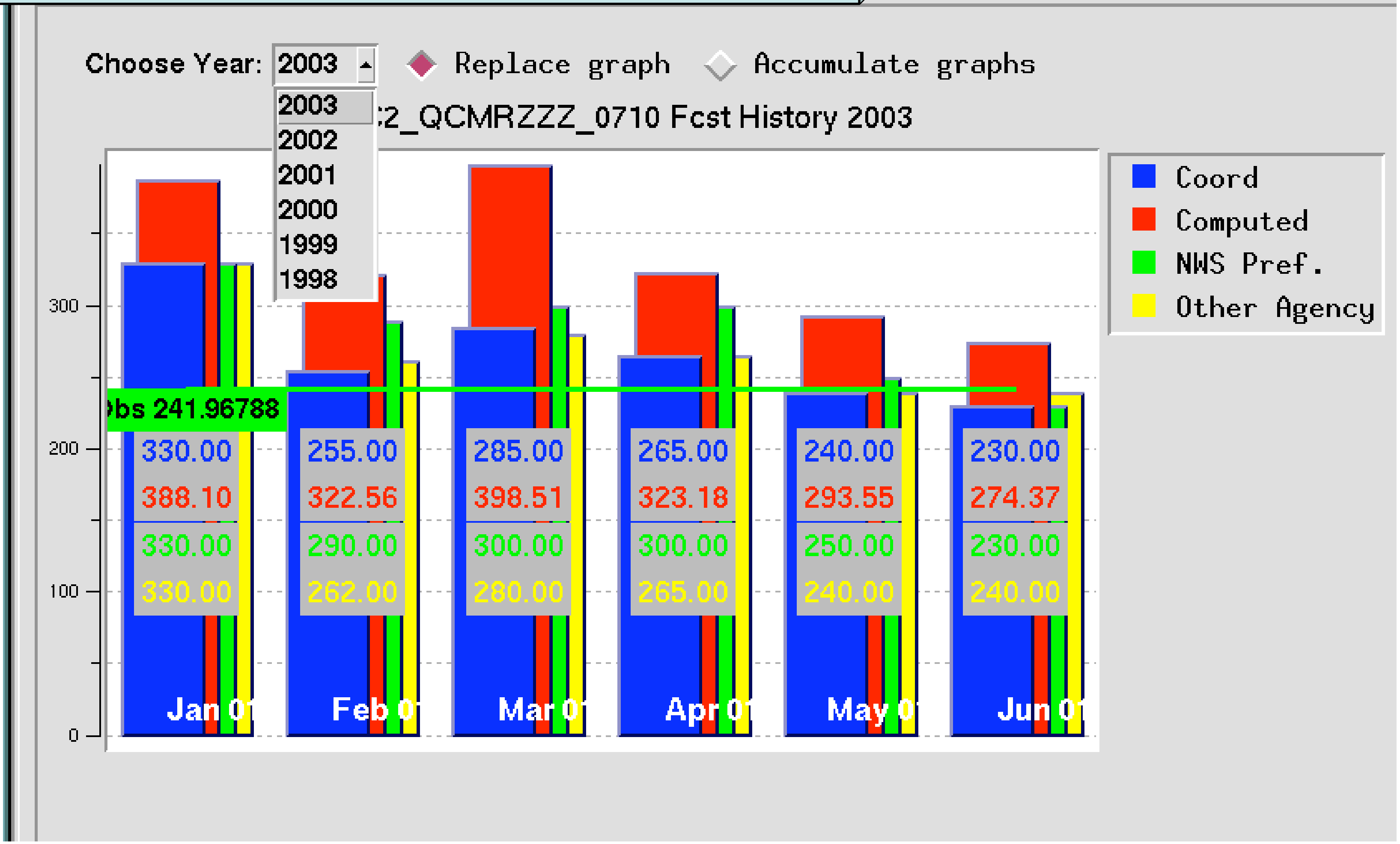
Nextreg – features

Eqn Stats tab



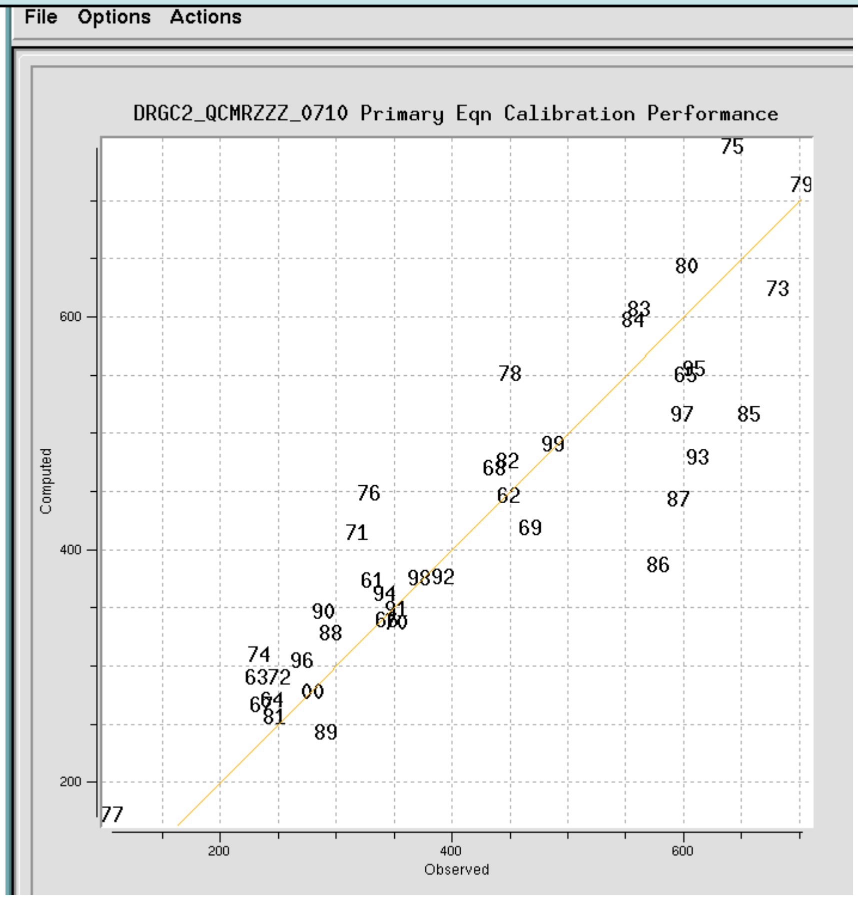
Nextreg – features

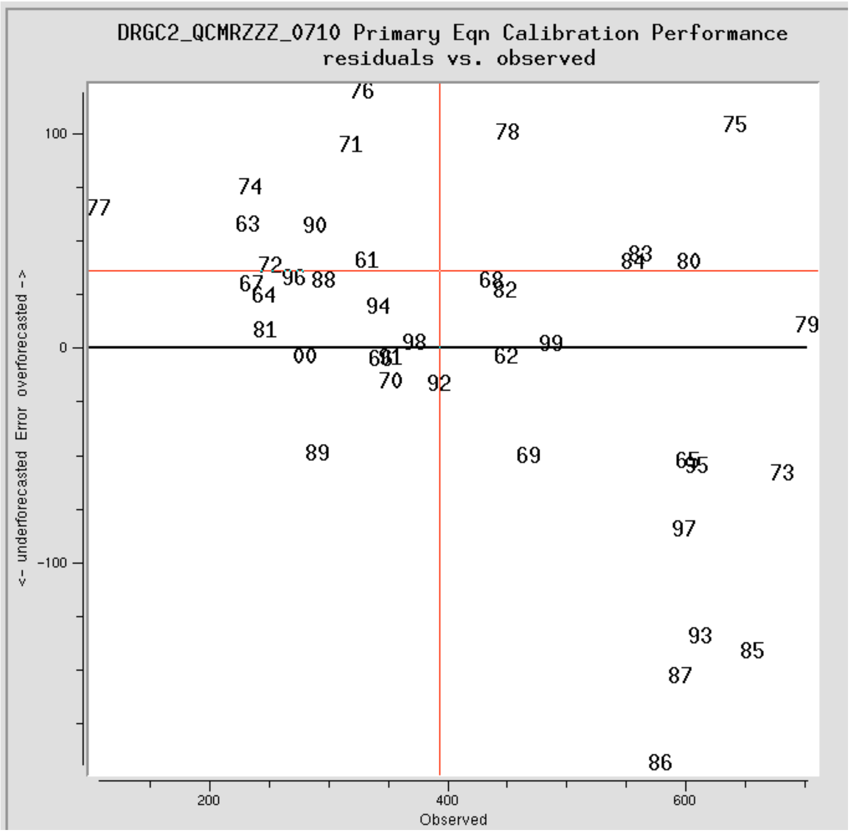
Fcst Perf (Oper) tab



Nextreg – features

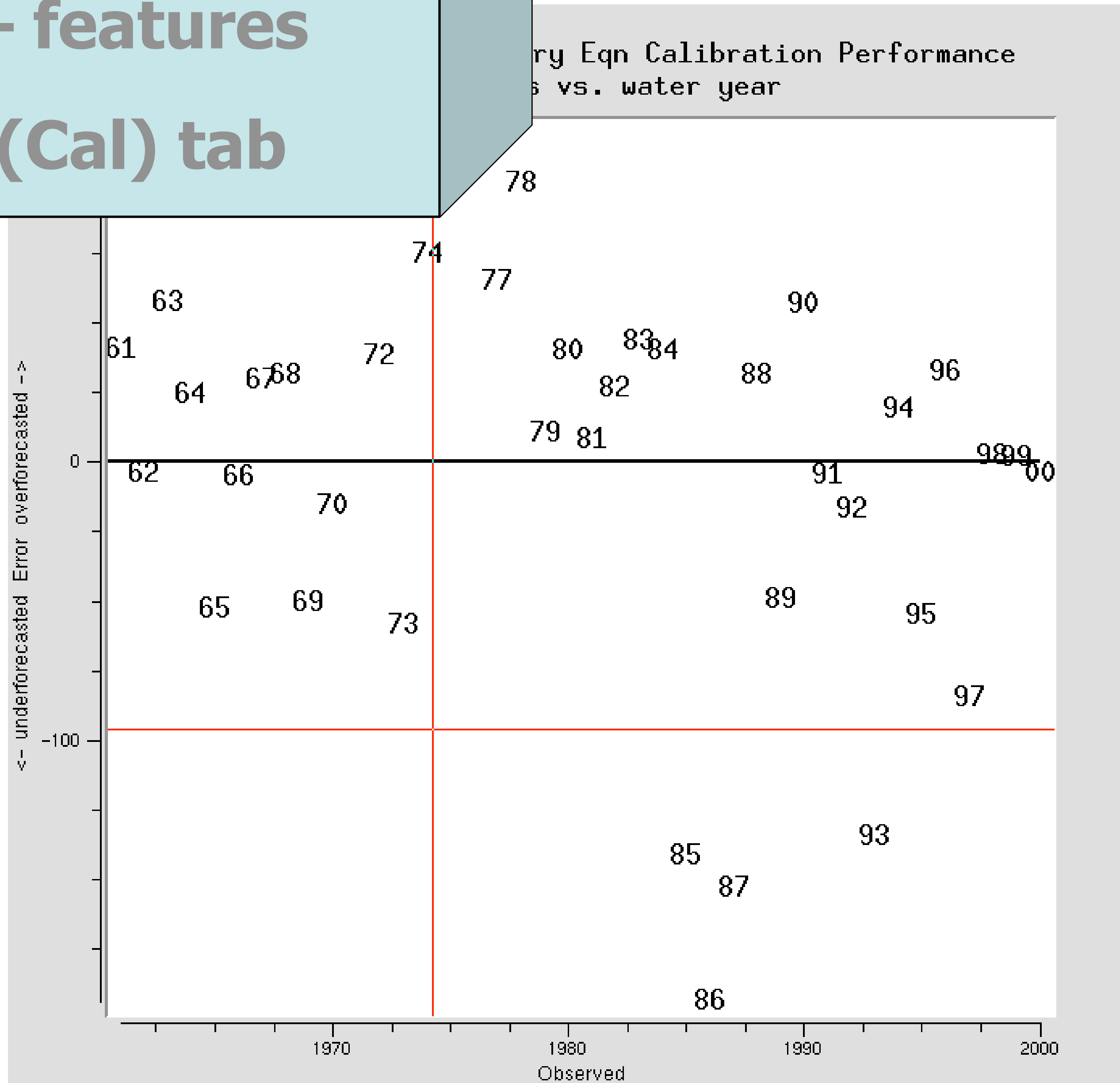
Fcst Perf (Cal) tab





Nextreg – features

Fcst Perf (Cal) tab



SWS Development Plans

- 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 dominate 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