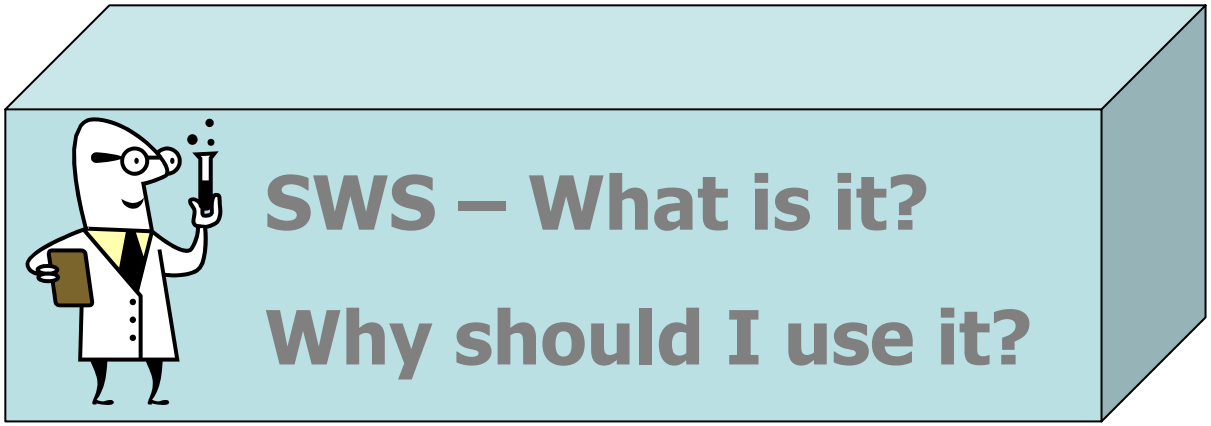
SWS

2003

Overview

Statistical Hydrology Workshop
Missouri Basin River Forecast Center

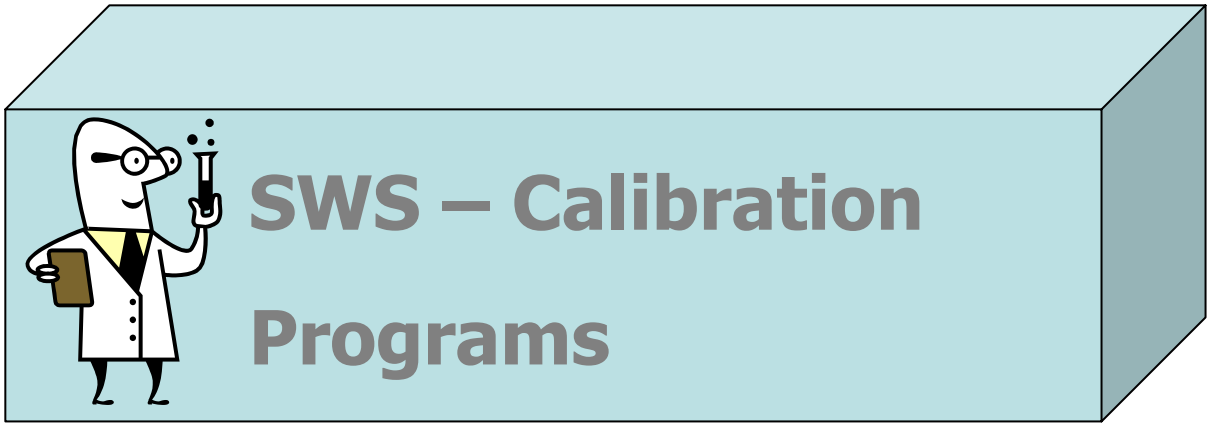


SWS – What is it?

Why should I use it?

- SWS – a package of inter-related programs to support water supply forecasting
- Monthly data – reap the benefits of the Informix 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 Calibration



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 Informix 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)

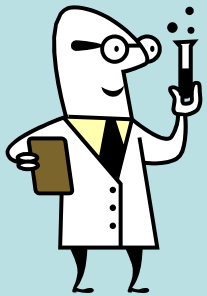
JACKKNIFE CORRELATION COEFFICIENT = 0.674

JACKKNIFE STANDARD ERROR = 84.035

JACKKNIFE BIAS: above average flow = -56.063 (14 obs.)

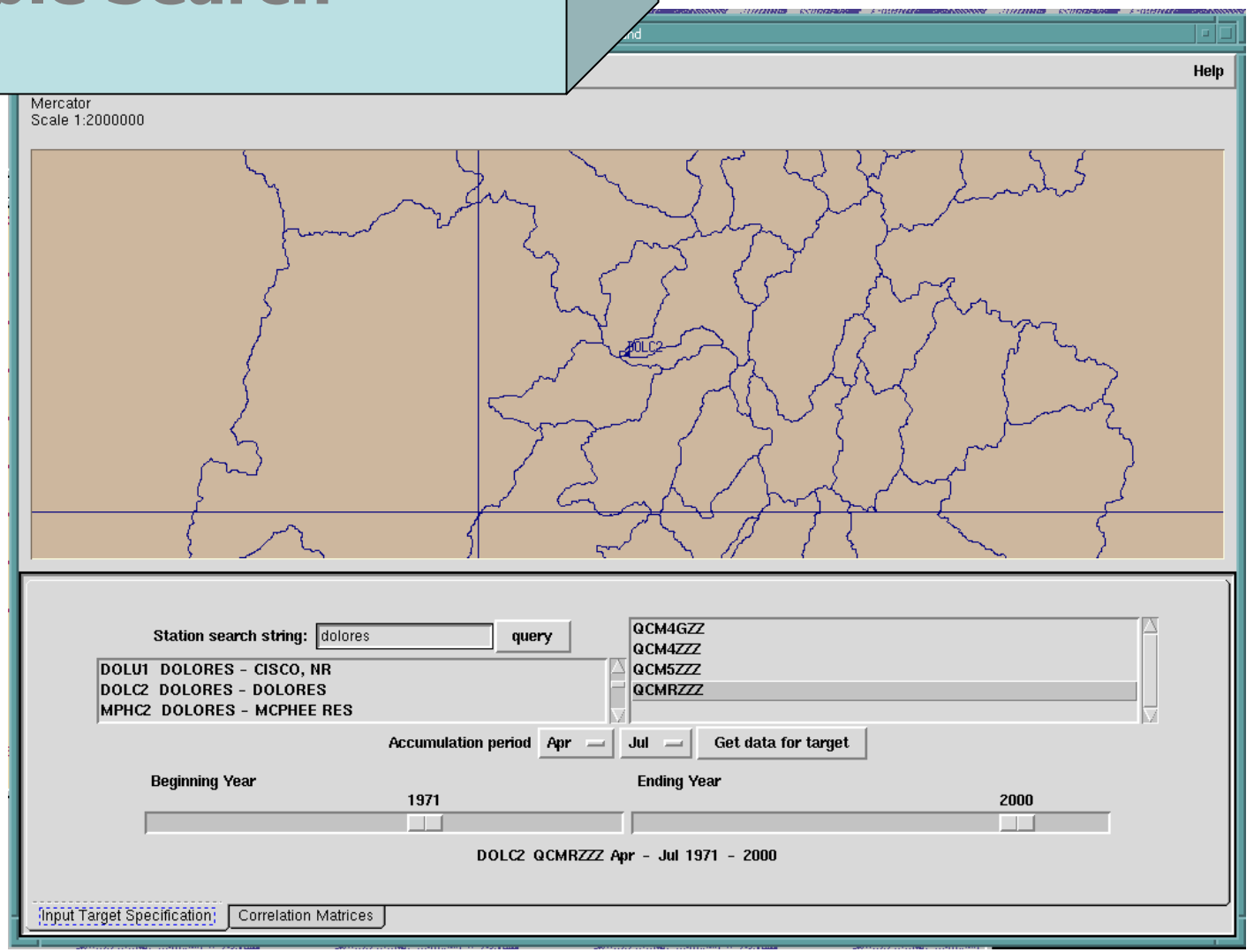
below average flow = 48.137 (16 obs.)

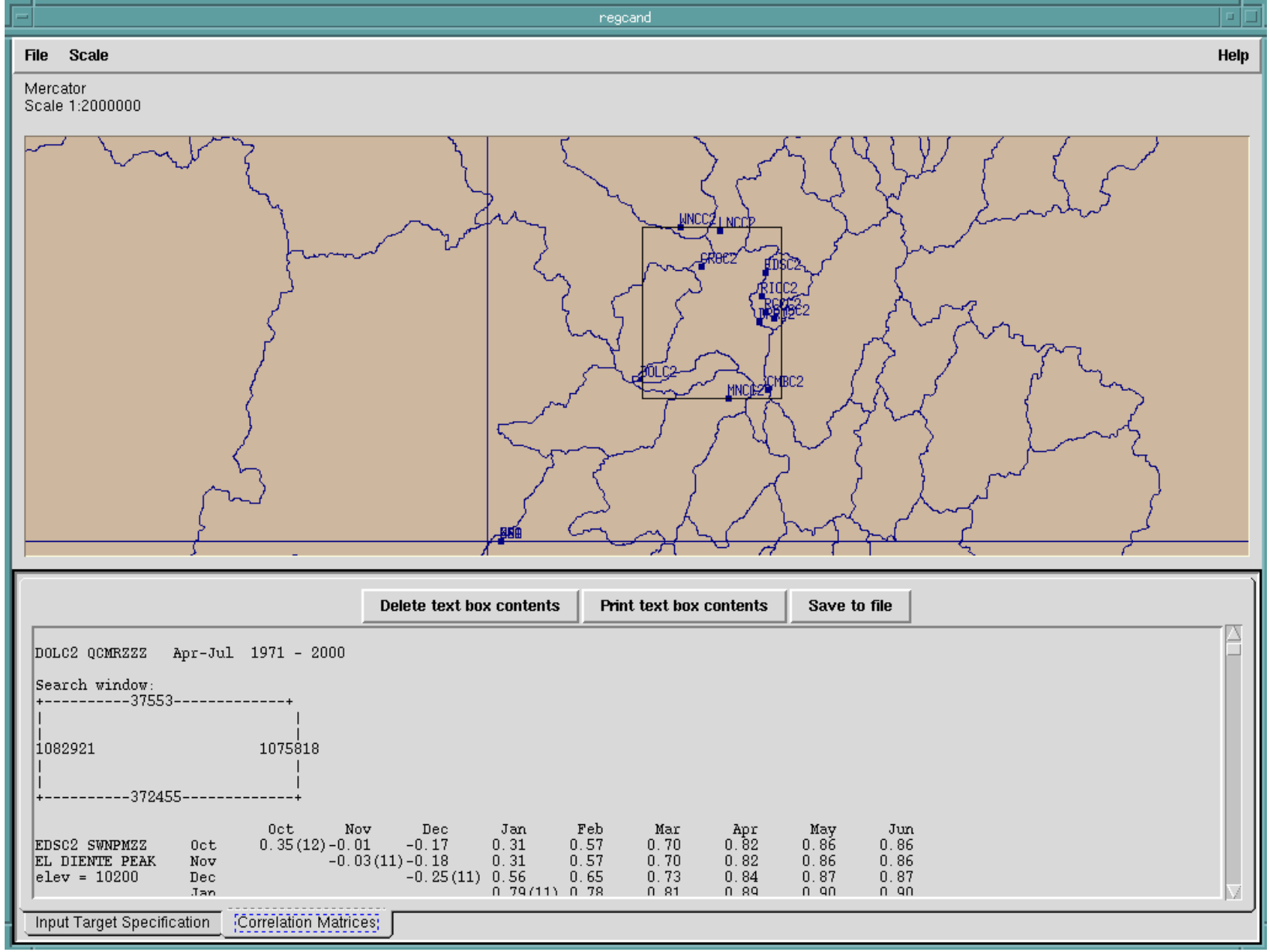
YEAR	OBSERVED	COMPUTED	ERROR	JACKKNIFE COMPUTED	JACKKNIFE ERROR
61	186.50	135.21	-51.29	130.37	-56.13
62	256.30	248.90	-7.40	247.56	-8.74
63	126.40	175.12	48.72	178.03	51.63
64	152.60	126.45	-26.15	123.94	-28.66
65	335.80	366.18	30.38	368.00	32.20
66	206.20	299.66	93.46	304.57	98.37
67	123.50	248.37	124.87	252.52	129.02
68	234.60	227.12	-7.48	226.56	-8.04
69	261.20	276.53	15.33	274.61	13.41
70	205.80	212.61	6.81	212.78	6.98
71	191.70	253.14	61.44	256.85	65.15
72	127.00	285.37	158.37	292.30	165.30
73	422.70	312.42	-110.28	305.55	-117.15



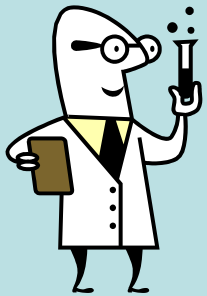
Regcand - Candidate Variable Search

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.



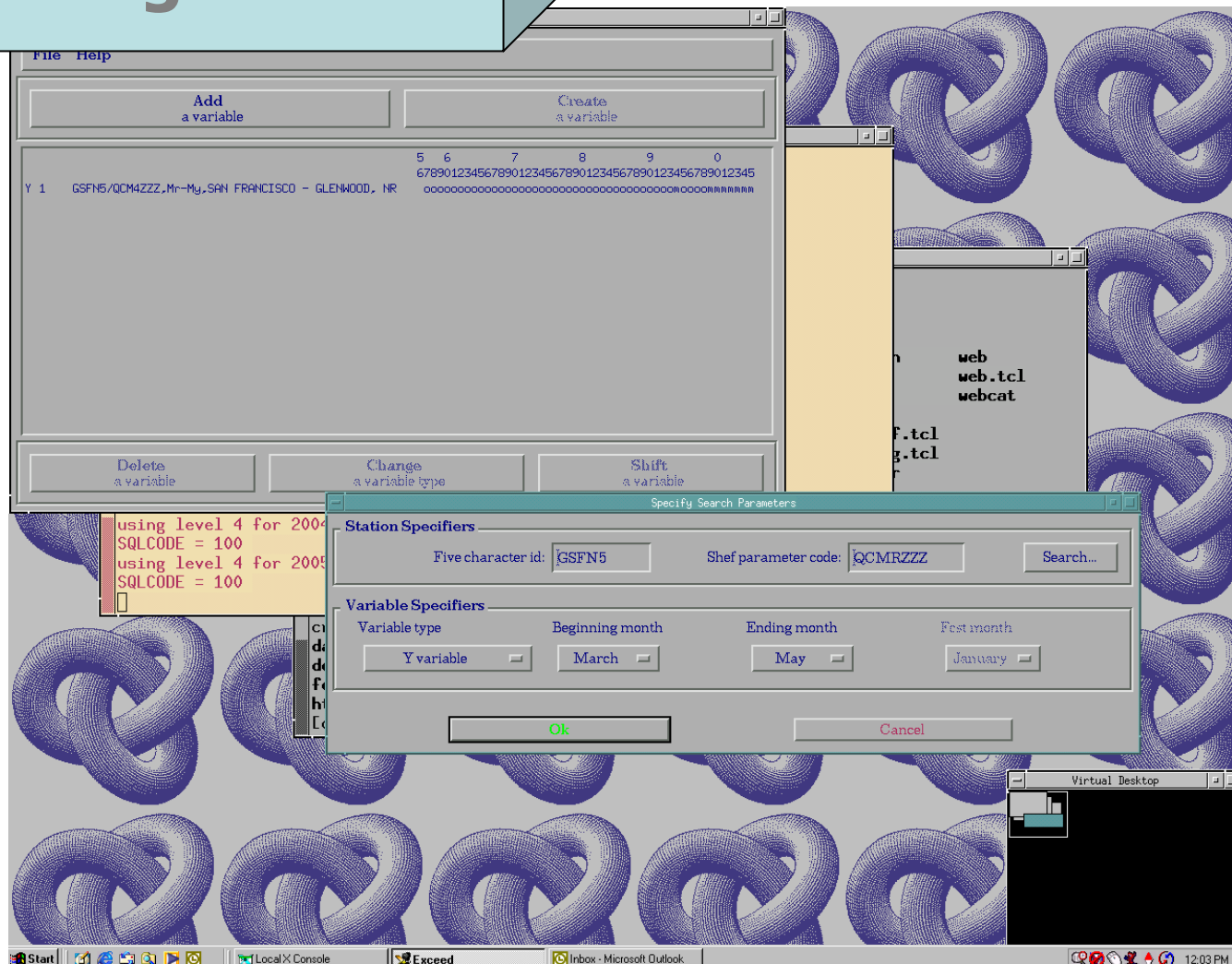


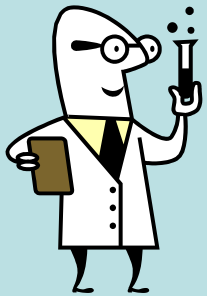
Future software development will include GIS where it is useful



Regdata – Assemble Input for Regcomb

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

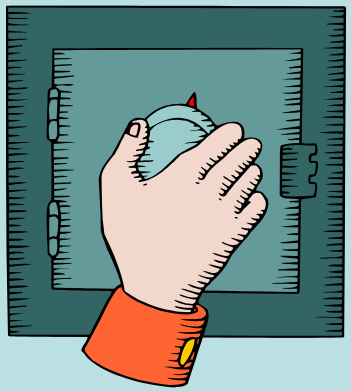




REGCOMB

Combination Analysis

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



Predictors, where A,B,C are stations:

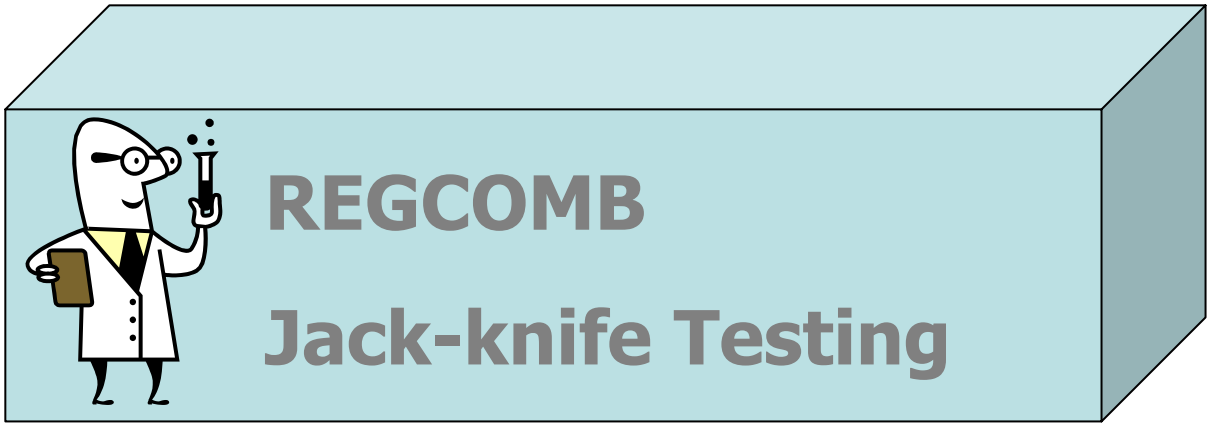
- snow-A, snow-B, snow-C
- precip-A, precip-B, precip-C (Oct-Dec)
- flow-A, flow-B
- …

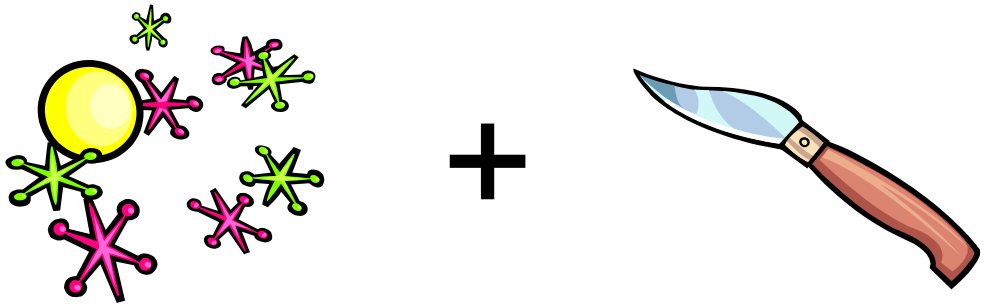
+

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

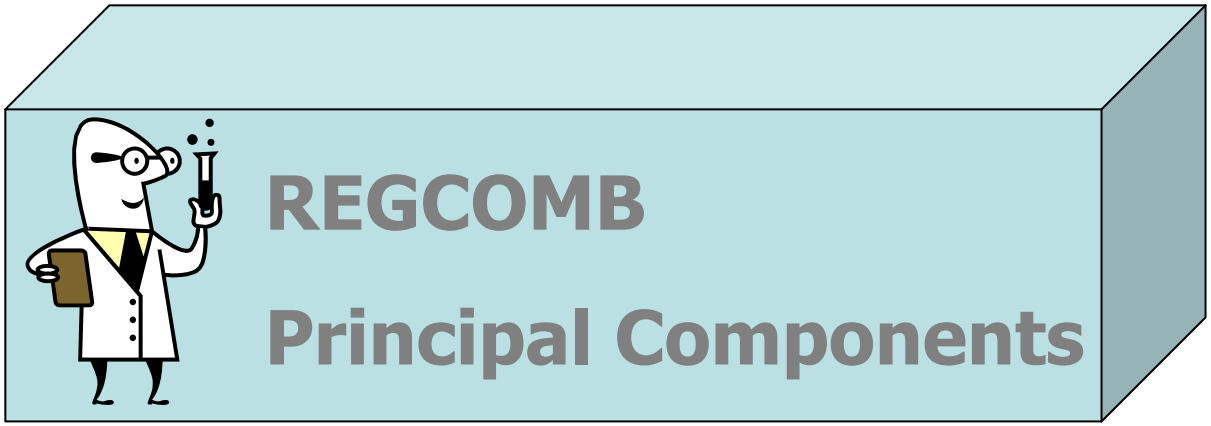
$$2.y = mx_1 + mx_2 + mx \ldots + b$$

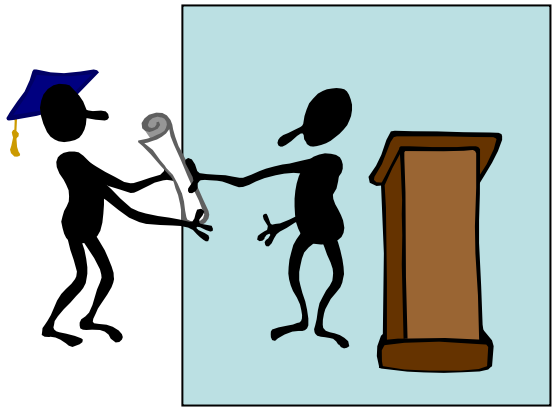
$$3 \ldots$$





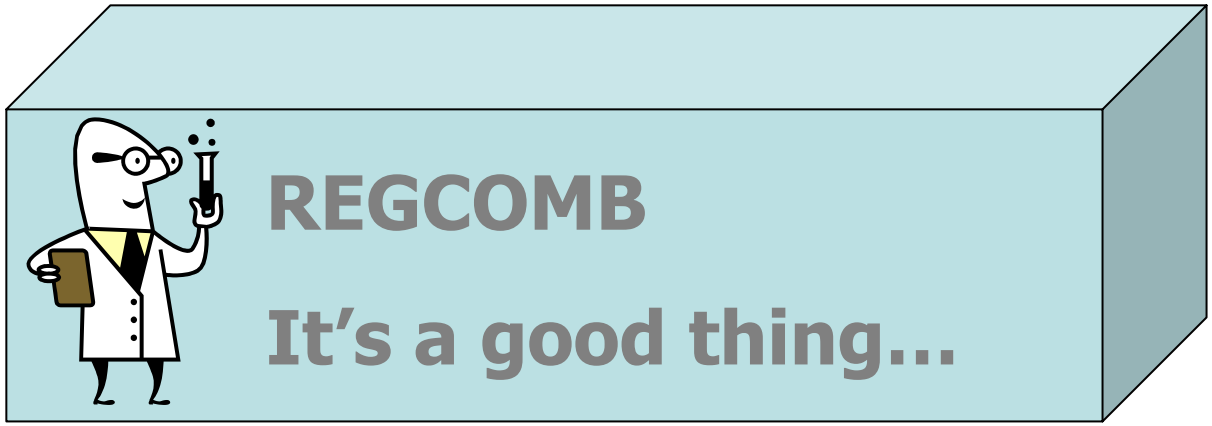
For a given set of predictors, one observation (one year) is deleted from the data set. Optimal coefficients are determined. The equation is then measured as to how well it predicted the selected year. Rinse, repeat. The idea here is to simulate how well the equation will perform in an operational environment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



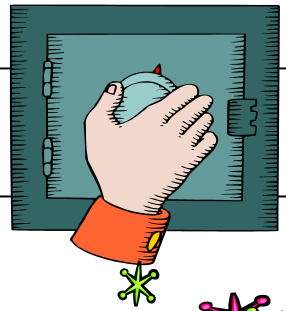


+

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.

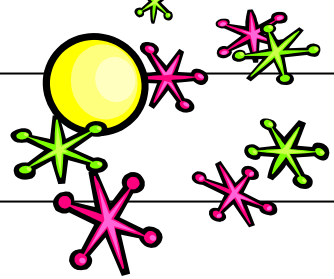


Combination
Analysis

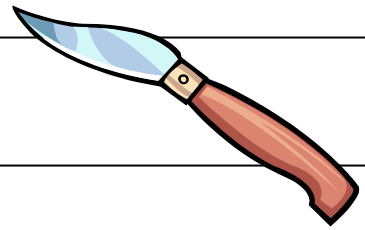


+

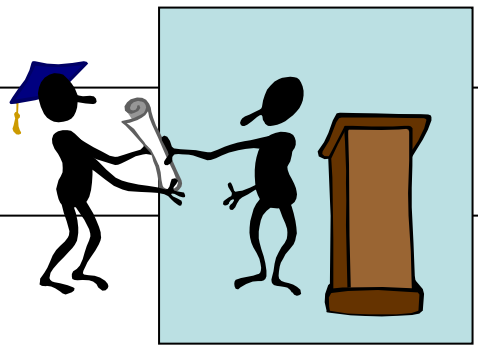
Jack-knife
error
computation



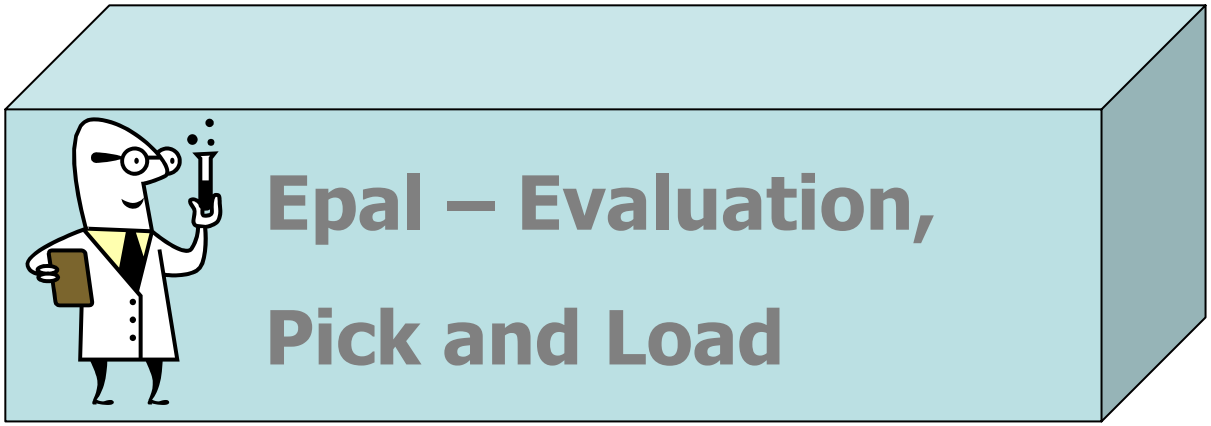
+



Principal
Components

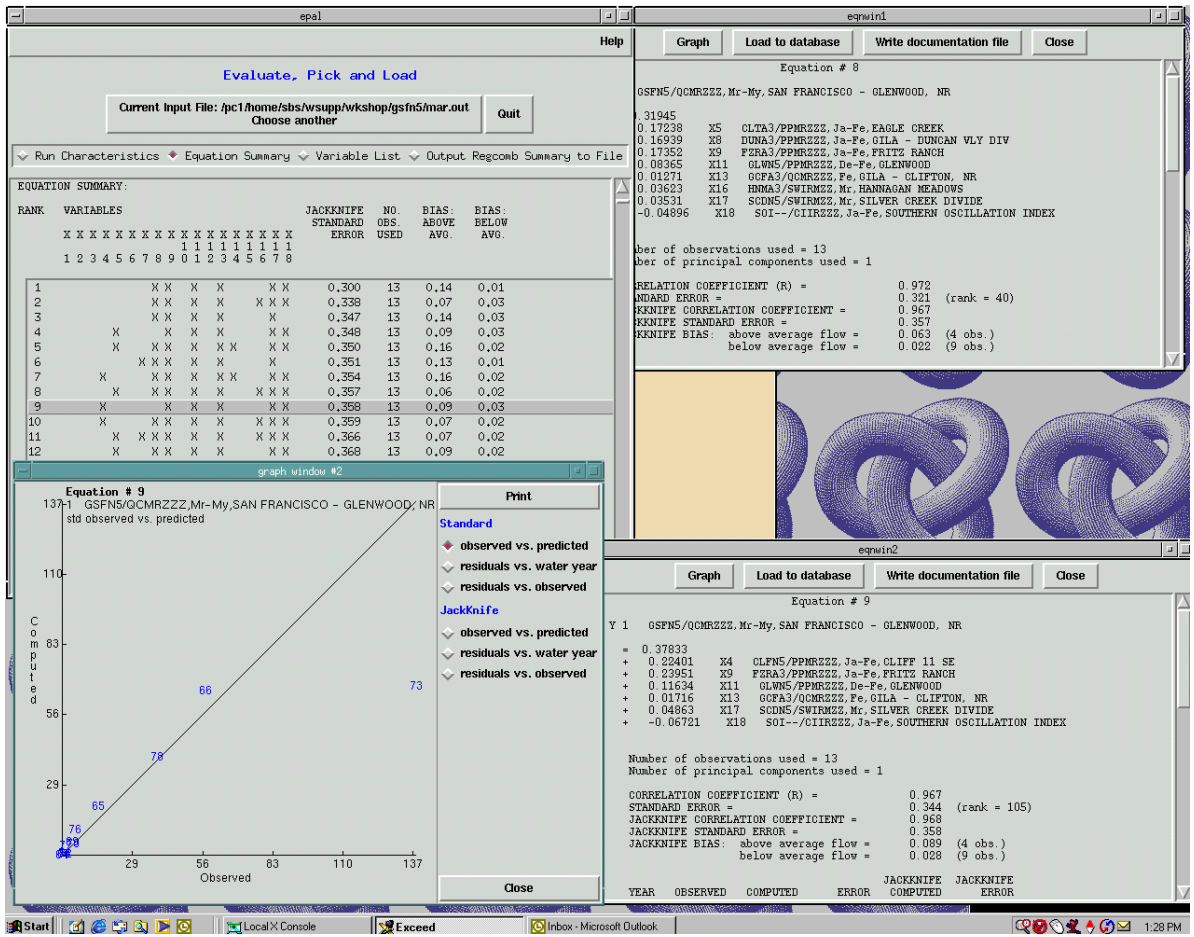


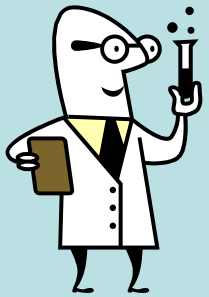
+



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 Informix 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.

```
<<<<<<<<<<<<<<<   C A L

        1.  REGCAND  --

        2.  REGDATA  --

        B. BESTTRANS --

        3.  REGCOMB w/GU
              R. REGCO
        4.  JACKREG  --

        5.  EPAL -- Eval

        6.  CSTATS   --

        7.  REGEQN   --


        O   operational
        M   other Monthl

ENTER Selection: 6
cstats version 2.0.1
```

cstats

Station search string: cameo query

CAMC2 COLORADO - CAMEO, NR
PCCC2 PLATEAU CK - CAMEO, NR

QCM4GZZ
QCM4ZZZ
QCM5ZZZ
QCMPAZZ
QCMPBZZ
QCMRZZZ

Accumulation period Apr Jul residual Calculate Stats

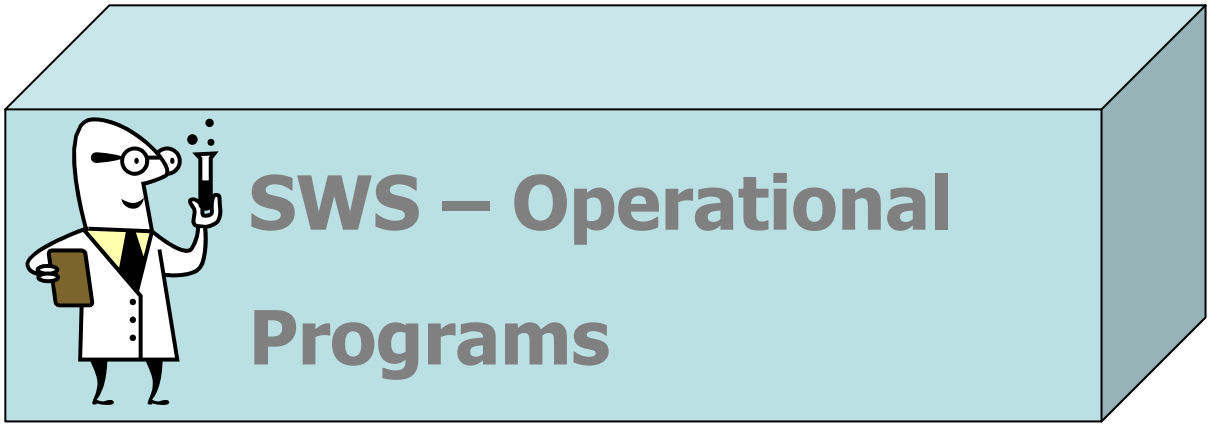
CAMC2 QCMPAZZ Apr - Jul

```
EXISTING RECORD:
-----------------------------------------
CAMC2 - QCMPAZZ Apr - Jul
                Avg      Median     Std dev
POR          2300.00     2240.00      753.13
Period       2420.00     2380.00      839.87

Hist. Max =      4488.51 (1984)
Hist. Min =       843.17 (1977)
-----------------------------------------

NEW calculation:
-----------------------------------------
CAMC2 - QCMPAZZ Apr - Jul
                Avg      Median     Std dev
POR          2260.00     2190.00      765.69
71-00        2420.00     2380.00      839.87

Hist. Max =      4488.51 (1984)
Hist. Min =       695.11 (2002)
-----------------------------------------
```

	avg	med	std
POR	2260.00	2190.00	765.69
Period	2420.00	2380.00	839.87

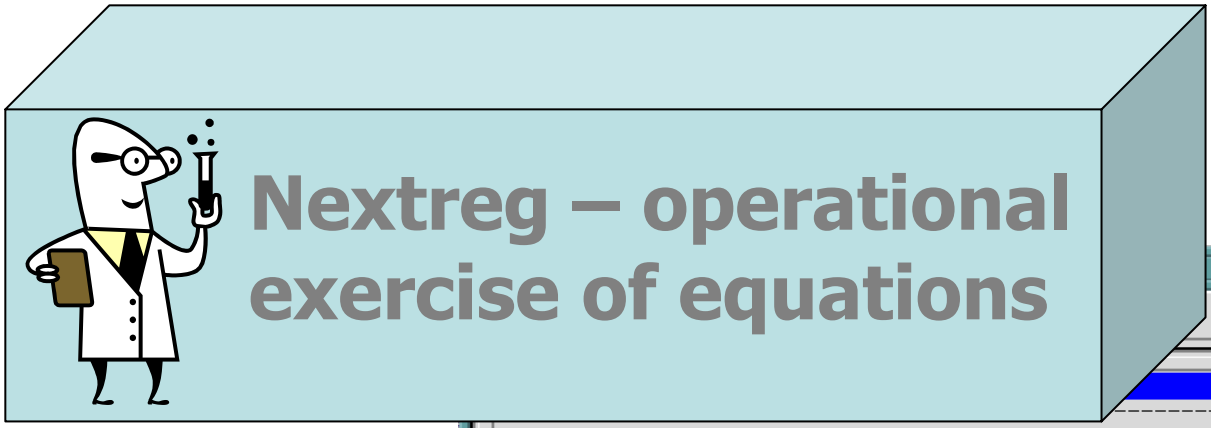
Print Post to DB Next Quit

SWS
Operational Use



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
- Datareq - makes a data availability report



Nextreg – operational exercise of equations

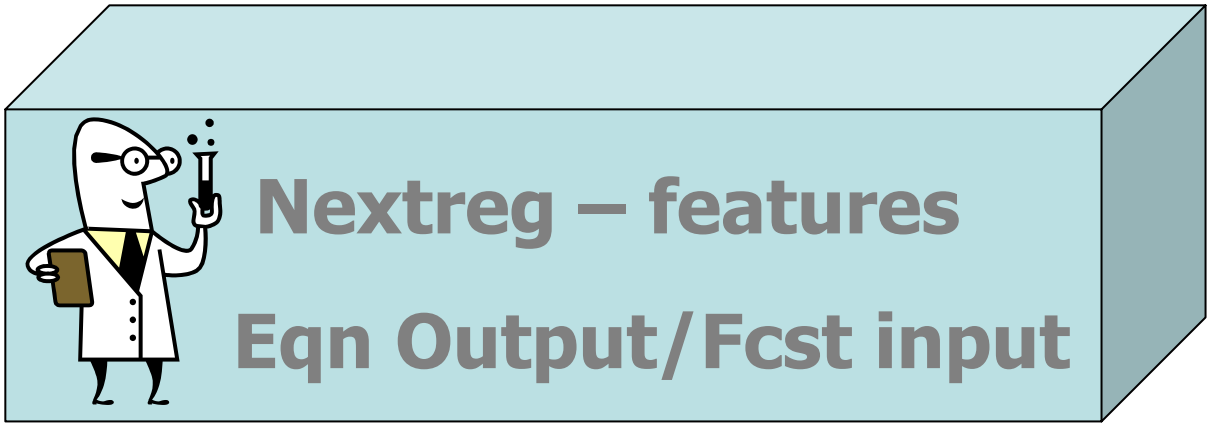
Equation Output/Fcst Input tab – spreadsheet style output of primary and secondary equations

Help

```
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%		%

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



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%)

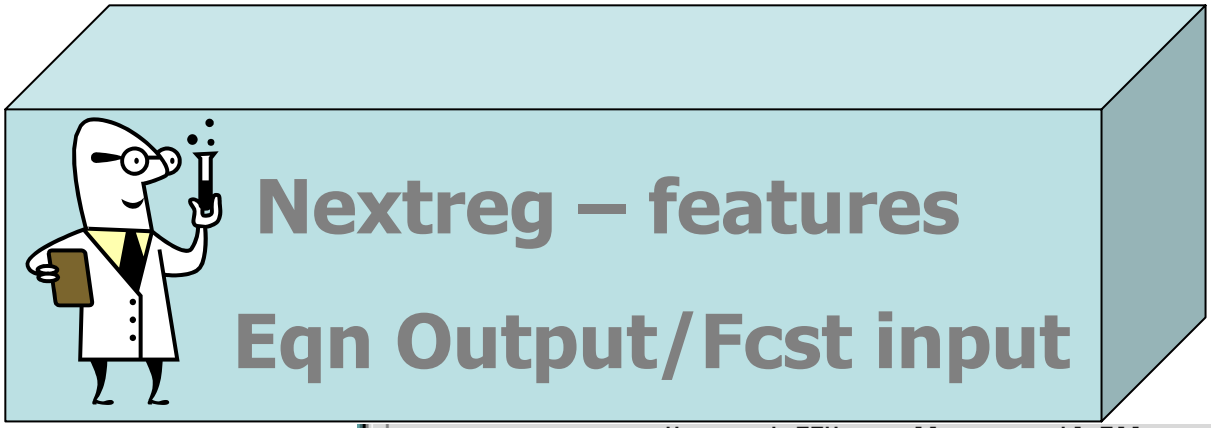
Historical Min and Max
39.54 (1977)
455.62 (1941)

DRGC2 QCMRZZZ - DURANGO) JR2: 0.970 # yrs: 30
AVG: S: 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%		%

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

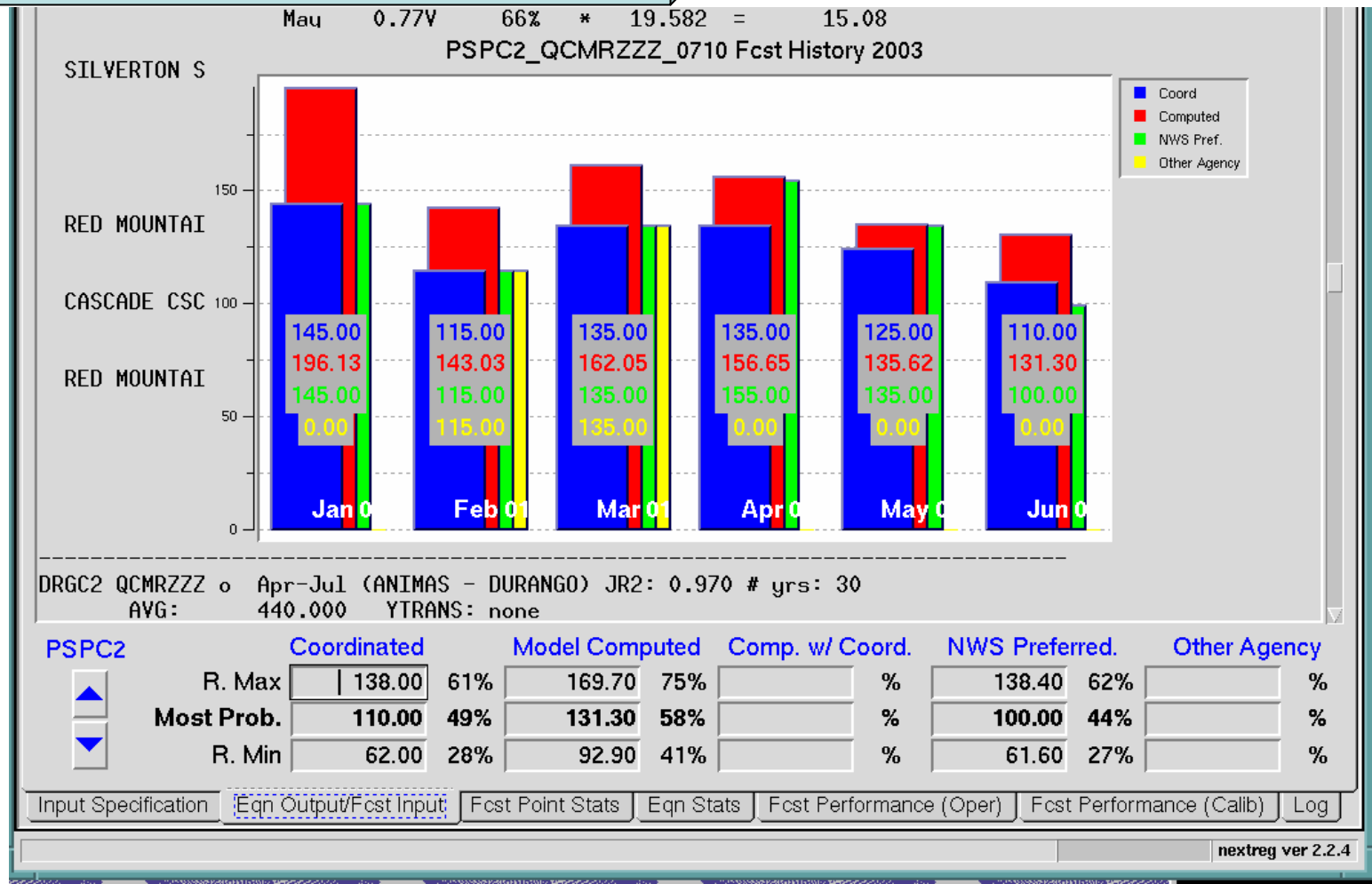
nextreg ver 2.2.4

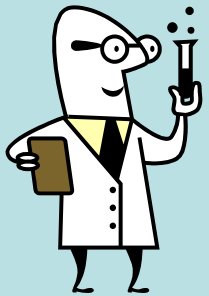


Nextreg – features

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

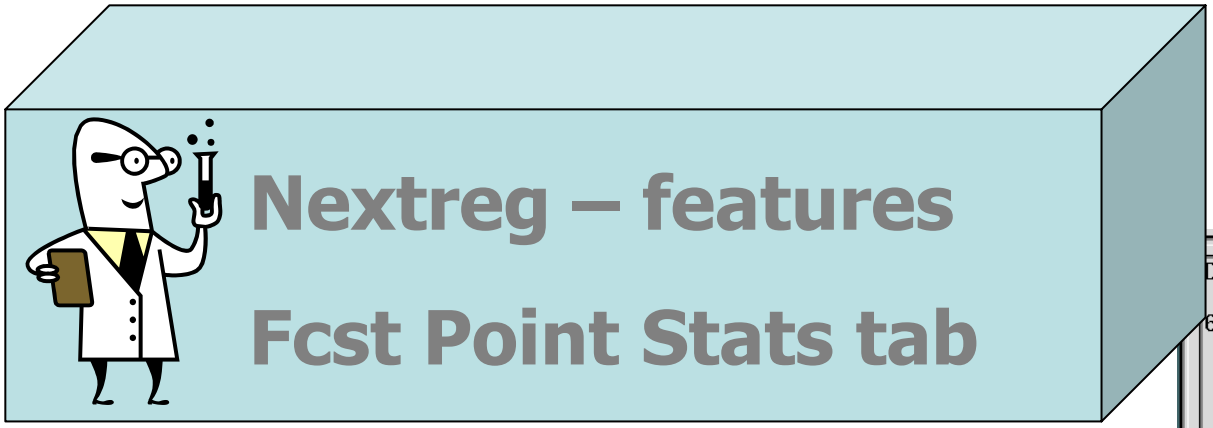
X-Y Scatterplot

X-Y scatterplot (91 years of common data)

SLVC2/PPMRZZZ (Apr - Apr)

DRGC2_QCMRZZZ_0710 flow

Close | Print



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)

nextreg

File Options Actions Help

Ranking of Historical Observed Data

1000's Acre Feet

Forecast

st Point Dataset Distribution Stats

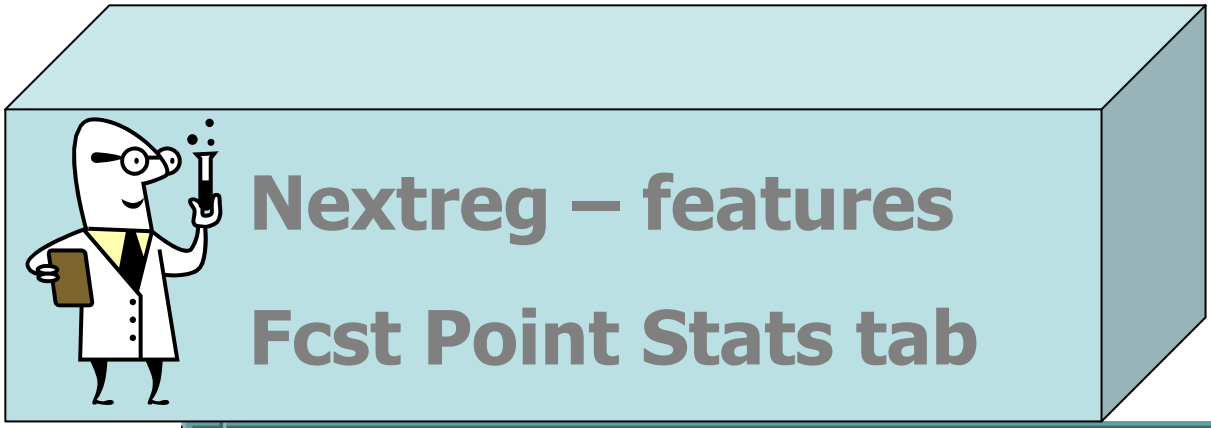
Average
Median
Std Dev
Hist Max (1920)
Hist Min (1977)

440.00
420.00
165.79
799.11
106.34

Ranking of Historical Observed Data

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

nextreg ver 2.2.4

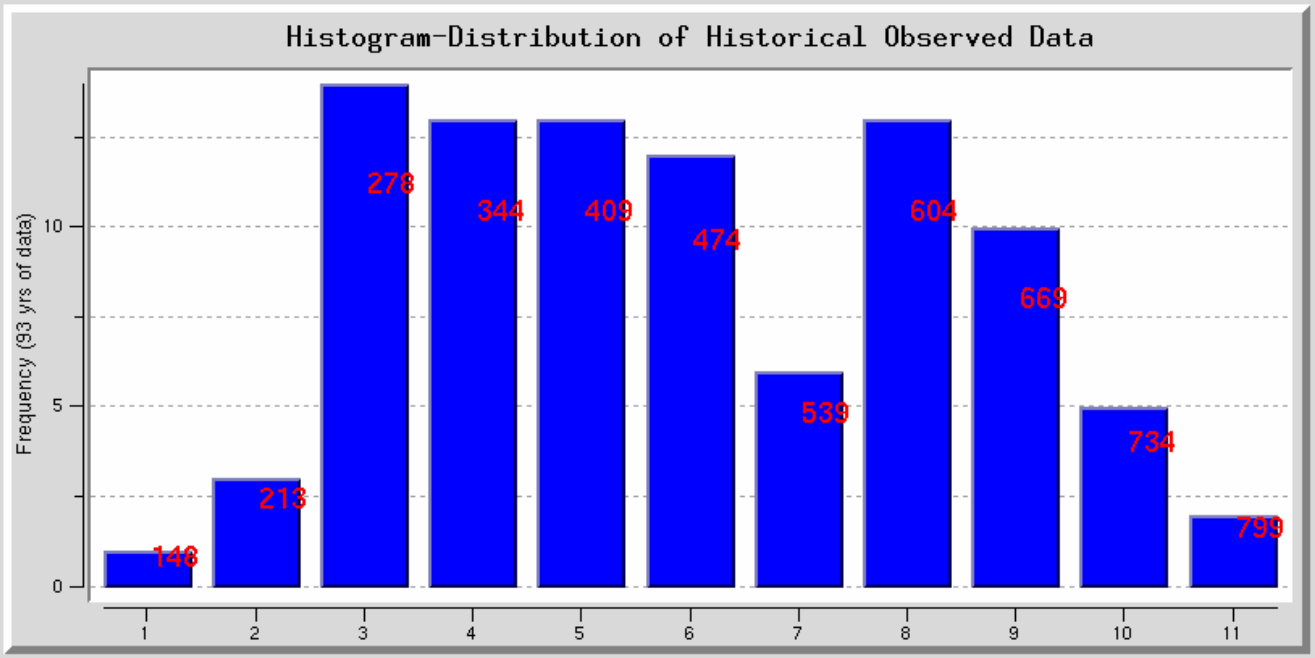


Nextreg – features

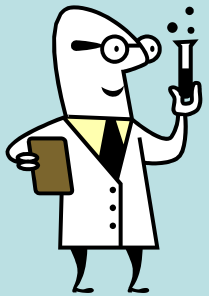
Fcst Point Stats tab

nextreg

File Options Actions




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



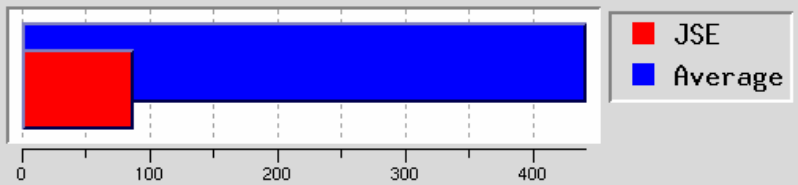
Nextreg – features

Eqn Stats tab

nextreg

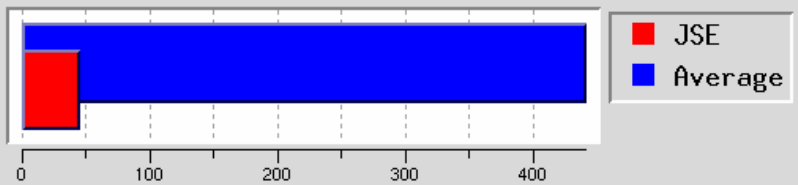
DRGC2_QCMRZZZ_0710 P
Calibration JSE: 86.07
JR2: 0.74
R2: 0.77
n: 30
per: 71-00

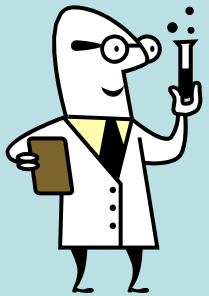
JSE vs. AVG



DRGC2_QCMRZZZ_0710 o
Calibration JSE: 44.28
JR2: 0.97
R2: 0.97
n: 30
per: 61-90

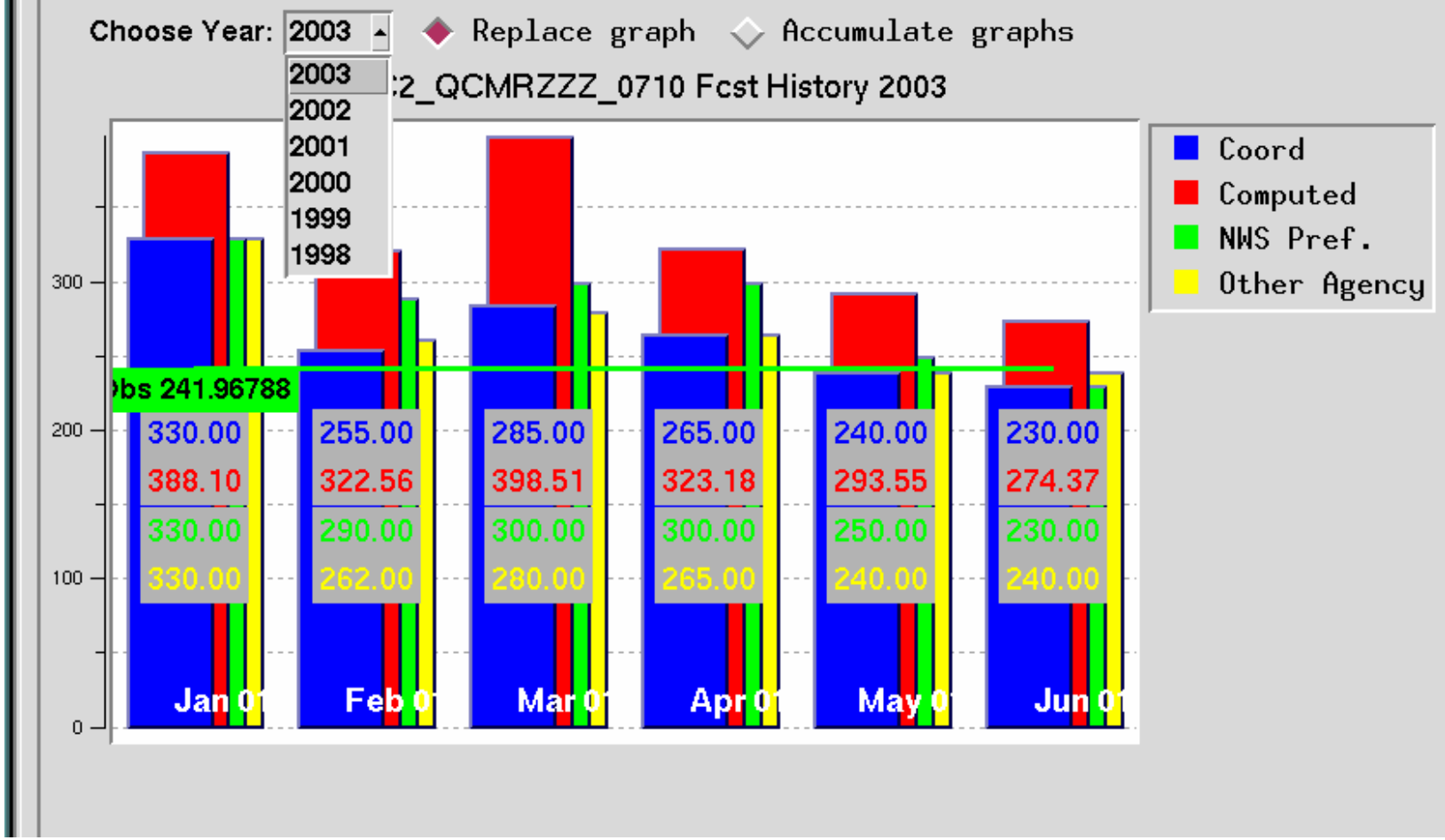
JSE vs. AVG

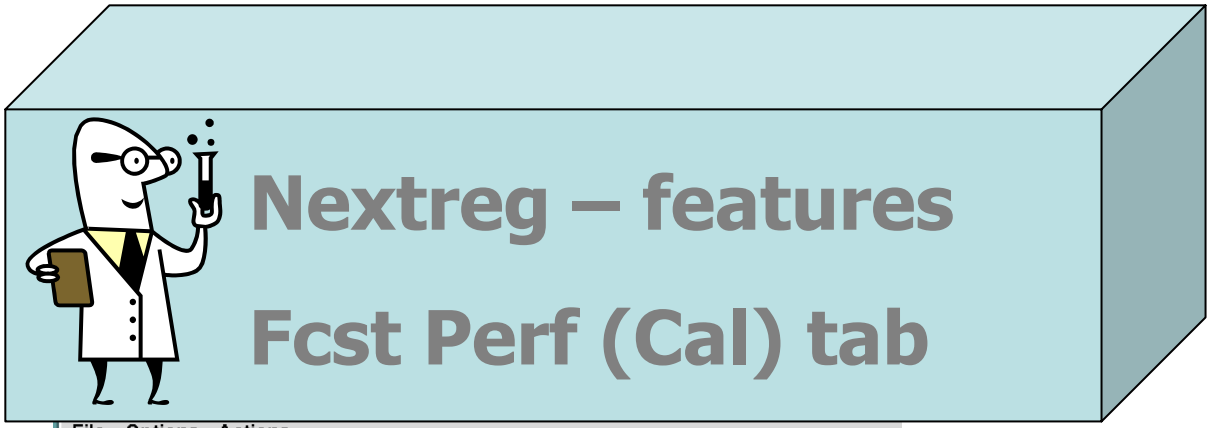




Nextreg – features

Fcst Perf (Oper) tab

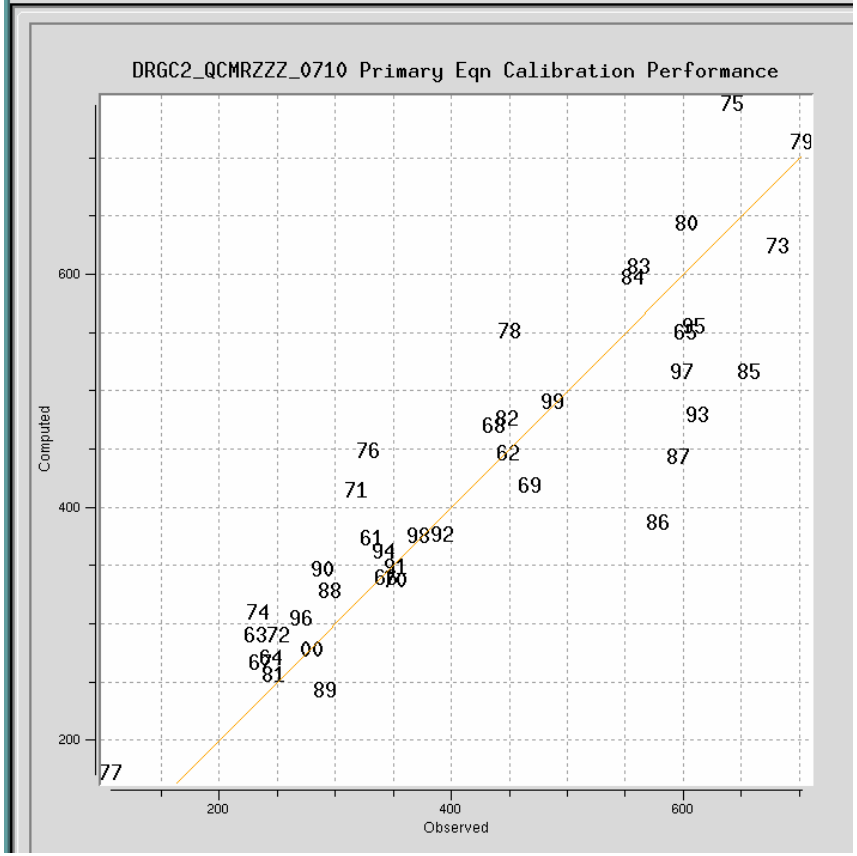


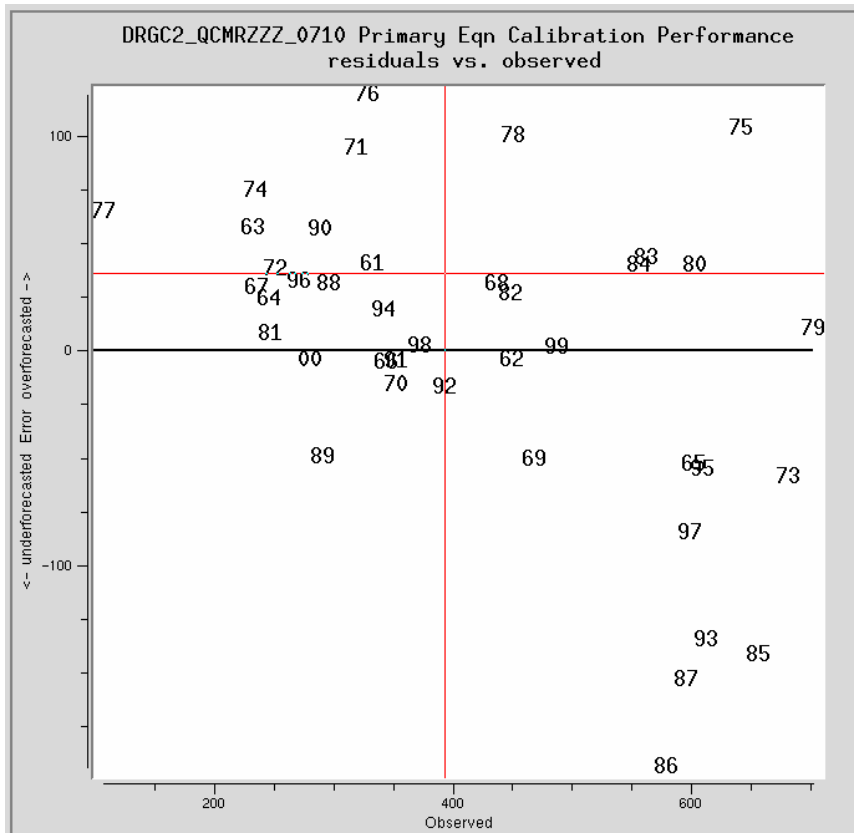


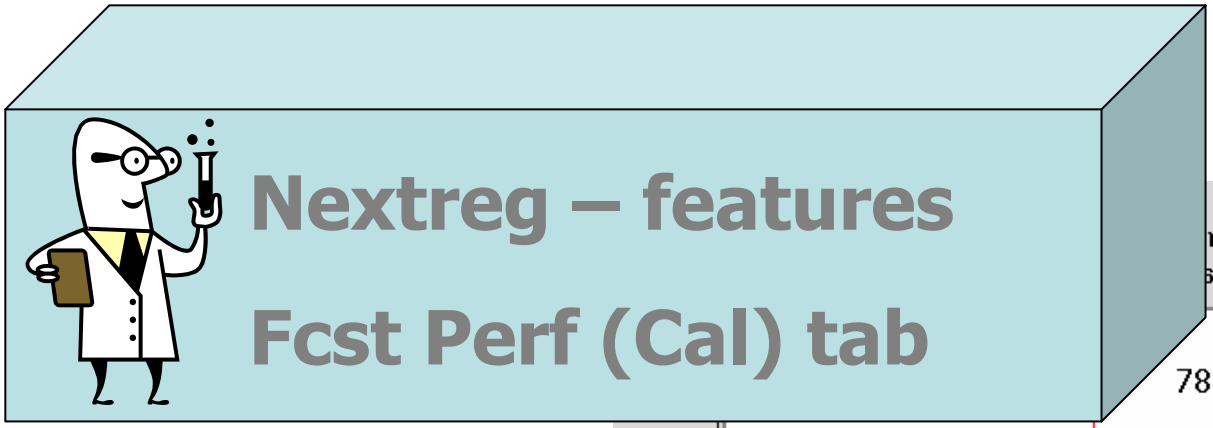
Nextreg – features

Fcst Perf (Cal) tab

File Options Actions

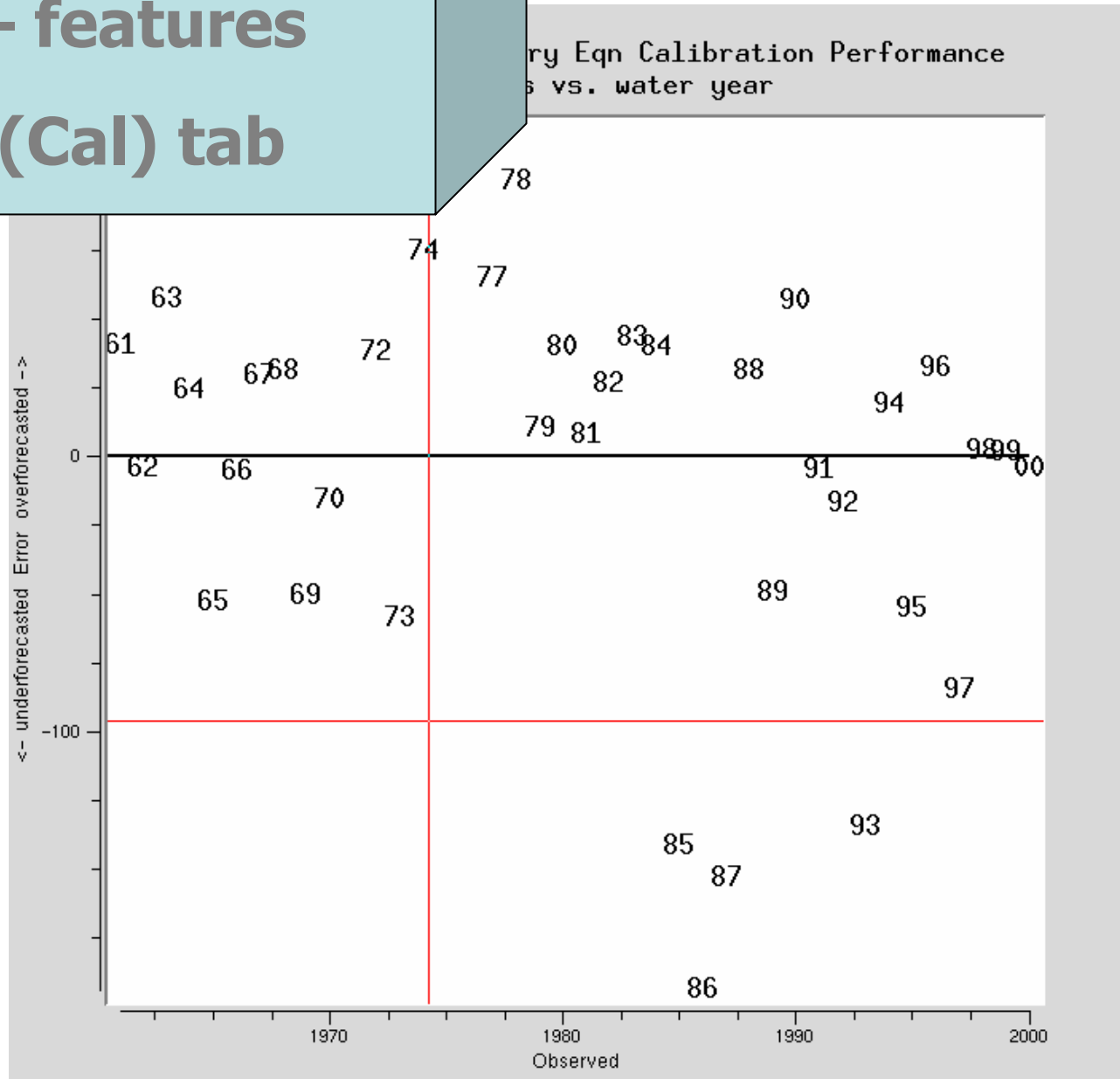


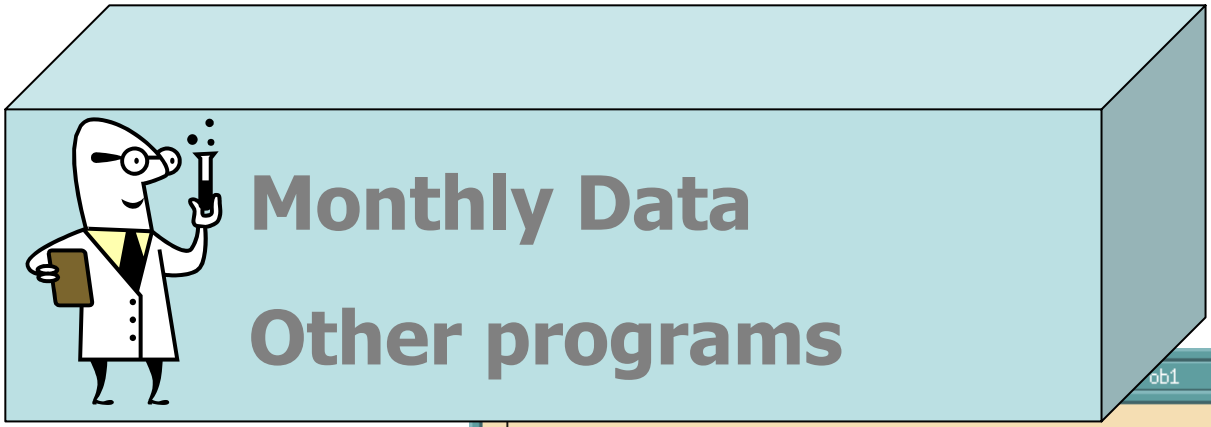




Nextreg – features

Fcst Perf (Cal) tab





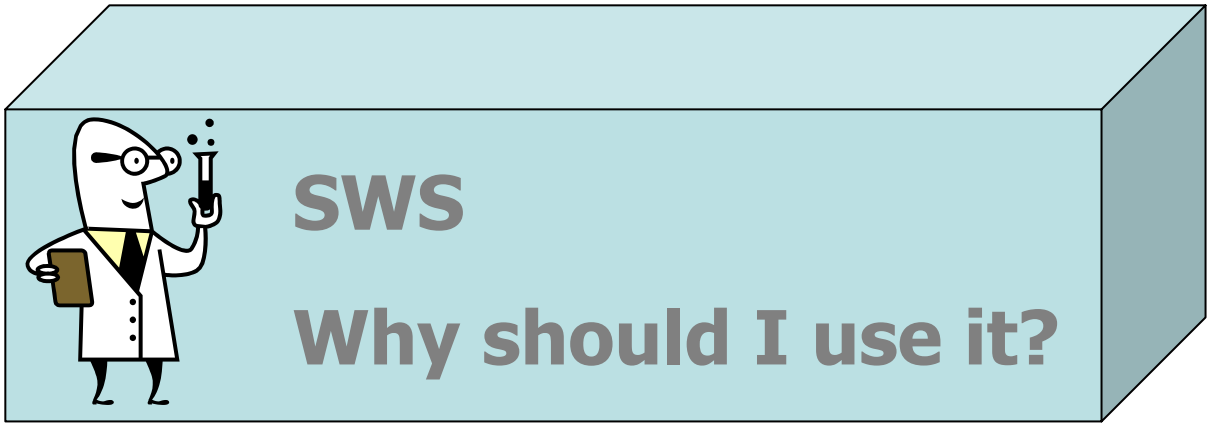
Monthly Data

Other programs

```
<<<<<<<<<<<<<<<   M O N T H L Y     M E N U   >>>>>>>>>>>>>>>>>>>>>

       1.  AVG_CALC  --  populates the avg database table

       2.  Xrank     --  ranks monthly values

       3.  Basum     --  summarizes data types by basin
           P.  Precsum  --  summarizes HUC output from monthly_qc

       4.  Est       --  estimates monthly data

       5.  swest     --  estimates first of (next) month swe
           D.  pcdiff   --  calculates snotel precip from end points

       6.  Reom      --  makes a reservoir end-of-month summary table

       7.  Prod2Prom --  accumulates daily values to monthly values

       8.  Qadjust   --  calculates unregulated flow

       9.  usgsparse -- parses usgs daily files, accums to monthly shef


       O   Operational menu           Q   Quit
       C   Calibration menu

ENTER Selection:
```

SWS Development Plans

- Continue trend of reducing number of programs
- GIS ground has been broken with Regcand; continue this approach
- Revise Informix table structures
- Get others involved in programming/support
- Create better documentation
- Enable more user customization (larger fonts, colors, etc)
- Investigate/incorporate newer statistical techniques and new ways of looking at verification



SWS

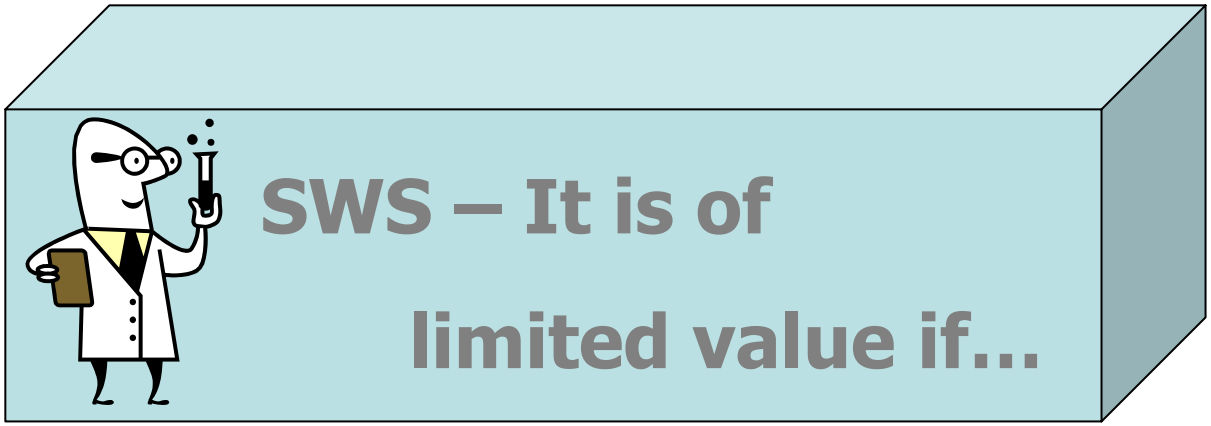
Why should I use it?

- A package of beginning-to-end integrated programs for water supply forecasting, or really, and 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
- A beefy tcl/tk interpreter – BLT, BWidget, cgi, expect, ftp, gd (graphics package),Tkgeomap (gis), tclisql (Informix access), tclx, as well as locally developed library functions



- There is no dominate driving force (like snowmelt)
- There is not a substantial period of record of data e.g. 1971-2000
- **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

- Appropriate? Snow melt? Must be some kind of predominate driving force.
- Setup – gathering monthly data, determining and calculating natural flow
- How long does it take to calibrate?