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 – 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
- **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

\[ Y_1 = \text{DOLC2/QCMRZZZ,Ap-Jl,DOLORES} - \text{DOLORES} \]
\[ = -5.203 + 14.859 \times X_4 \text{HSPC2/PPMRZZZ,No-Ja,FORT LEWIS} \]
\[ + 57.175 \times X_5 \text{DRRC2/QCMRZZZ,Ja,DOLORES} - \text{RICO, BLO} \]
\[ + 10.161 \times X_6 \text{LNCC2/SWIRMZZ,Fe,LONE CONE} \]

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.)

<table>
<thead>
<tr>
<th>YEAR</th>
<th>OBSERVED</th>
<th>COMPUTED</th>
<th>ERROR</th>
<th>COMPUTED</th>
<th>ERROR</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>186.50</td>
<td>135.21</td>
<td>-51.29</td>
<td>130.37</td>
<td>-56.13</td>
</tr>
<tr>
<td>62</td>
<td>256.30</td>
<td>248.90</td>
<td>-7.40</td>
<td>247.56</td>
<td>-8.74</td>
</tr>
<tr>
<td>63</td>
<td>126.40</td>
<td>175.12</td>
<td>48.72</td>
<td>178.03</td>
<td>51.63</td>
</tr>
<tr>
<td>64</td>
<td>152.60</td>
<td>126.45</td>
<td>-26.15</td>
<td>123.94</td>
<td>-28.66</td>
</tr>
<tr>
<td>65</td>
<td>335.80</td>
<td>366.18</td>
<td>30.38</td>
<td>368.00</td>
<td>32.20</td>
</tr>
<tr>
<td>66</td>
<td>206.20</td>
<td>299.66</td>
<td>93.46</td>
<td>304.57</td>
<td>98.37</td>
</tr>
<tr>
<td>67</td>
<td>123.50</td>
<td>248.37</td>
<td>124.87</td>
<td>252.52</td>
<td>129.02</td>
</tr>
</tbody>
</table>
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
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
- ...

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

1. \[ y = mx_1 + mx_2 + mx \ldots + b \]
2. \[ y = mx_1 + mx_2 + mx \ldots + b \]
3...
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.
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
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.
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

<table>
<thead>
<tr>
<th>Equation</th>
<th>Calculation</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>DURANGO WATER RESOURCE</td>
<td>May 0.77V x 19.582 = 15.08</td>
<td></td>
</tr>
<tr>
<td>SILVERTON SLVC2/PPMRZZZ</td>
<td>Apr: 1.17V x 71%</td>
<td>2.60 x 63% x 7.771 = 17.10</td>
</tr>
<tr>
<td></td>
<td>May: 1.03V x 55%</td>
<td>------</td>
</tr>
<tr>
<td>RED MOUNTAIN PASS RMPC2/SWIRMZZ</td>
<td>May: 17.30Z x 64%</td>
<td>3.206 x 5.966 = 103.21</td>
</tr>
<tr>
<td>CASCADE CSC2/SWIRMZZ</td>
<td>May: 0.60Z x 16%</td>
<td>7.216 = 4.33</td>
</tr>
<tr>
<td></td>
<td>Jun: 0.70Z x 5%</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>May-Jun: 17.30Z x 64%</td>
<td>18.00 x 44% x 2.811 = 50.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>84.055 + 190.31 = 274.37 (62%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PSPC2</th>
<th>Coordinated</th>
<th>Model Computed</th>
<th>Comp. w/ Coord.</th>
<th>NWS Preferred</th>
<th>Other Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. Max</td>
<td>136.00</td>
<td>169.70</td>
<td>75%</td>
<td>138.40</td>
<td>62%</td>
</tr>
<tr>
<td>Most Prob.</td>
<td>110.00</td>
<td>131.30</td>
<td>58%</td>
<td>100.00</td>
<td>44%</td>
</tr>
<tr>
<td>R. Min</td>
<td>62.00</td>
<td>92.90</td>
<td>41%</td>
<td>61.60</td>
<td>27%</td>
</tr>
</tbody>
</table>
Nextreg – features
Eqn Output/Fcst input

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

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

- May: 17.30Z, 64%
- Jun: 0.70Z, 5%

18.00 × 2.811 = 50.60

84.055 + 190.31 = 274.37 (62%)

Historical Min and Max:
- DRGC2: 39.54 (1977), 455.62 (1941)
- Durango: JR2: 0.970, yrs: 30

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

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
Fcast Point Stats tab

Ranking of Historical Observed Data

Histogram of Fcast Point Dataset
- Average: 440.00
- Median: 420.00
- Std Dev: 165.79
- Hist Max: 799.11 (1920)
- Hist Min: 106.34 (1977)
Nextreg – features
Fcast Point 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
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
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

1. Data analysis and quality control
2. Run SWS and ESP models
3. Analyze model outputs
4. CBRFC preliminary forecast
5. NRCS preliminary forecast
6. Forecaster insight
7. Final coordinated forecast