SWS 2003 Overview

Statistical Hydrology Workshop
Missouri Basin River Forecast Center

8. 5. 2003
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

\[ Y = -5.203 + 14.859X_4 + 57.175X_5 + 10.161X_6 \]

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>
<tr>
<td>68</td>
<td>234.60</td>
<td>227.12</td>
<td>-7.48</td>
<td>226.56</td>
<td>-8.04</td>
</tr>
<tr>
<td>69</td>
<td>261.20</td>
<td>276.53</td>
<td>15.33</td>
<td>274.61</td>
<td>13.41</td>
</tr>
<tr>
<td>70</td>
<td>205.80</td>
<td>212.61</td>
<td>6.81</td>
<td>212.78</td>
<td>6.98</td>
</tr>
<tr>
<td>71</td>
<td>191.70</td>
<td>253.14</td>
<td>61.44</td>
<td>256.85</td>
<td>65.15</td>
</tr>
<tr>
<td>72</td>
<td>127.00</td>
<td>285.37</td>
<td>158.37</td>
<td>292.30</td>
<td>165.30</td>
</tr>
<tr>
<td>73</td>
<td>422.70</td>
<td>312.42</td>
<td>-110.28</td>
<td>305.55</td>
<td>-117.15</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 Informix, accumulates as needed, builds formatted input file for Regcomb
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
- ...

\[ y = m_1 x_1 + m_2 x_2 + m_x \ldots + b \]

1. \[ y = m_1 x_1 + m_2 x_2 + m_x \ldots + b \]
2. \[ y = m_1 x_1 + m_2 x_2 + m_x \ldots + b \]
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. 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.
REGCOMB
It’s a good thing…

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.
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
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

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

**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

Replace graph
Accumulate graphs

Y-axis: Jobs 241.96788

Jan 01: Coord 330.00, Computed 388.10, NWS Pref. 330.00, Other Agency 330.00
Feb 01: Coord 255.00, Computed 322.56, NWS Pref. 290.00, Other Agency 262.00
Mar 01: Coord 285.00, Computed 398.51, NWS Pref. 300.00, Other Agency 290.00
Apr 01: Coord 265.00, Computed 323.18, NWS Pref. 300.00, Other Agency 265.00
May 01: Coord 240.00, Computed 293.55, NWS Pref. 250.00, Other Agency 240.00
Jun 01: Coord 230.00, Computed 274.37, NWS Pref. 230.00, Other Agency 240.00

Legend:
- Coord
- Computed
- NWS Pref.
- Other Agency
Nextreg – features
Fcst Perf (Cal) tab
Nextreg – features
Fcst Perf (Cal) tab
Monthly Data
Other programs

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

0  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
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
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 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?