MODERNIZATION OF STATISTICAL PROCEDURES FOR WATER SUPPLY FORECASTING

Robert K. Hartman[1] and Arthur F. Henkel[2]

ABSTRACT

The promise of conceptual watershed models and extended streamflow prediction all but terminated statistical water supply forecasting technique development by the National Weather Service in the early 1980's. Statistical procedures have, however, remained the dominant technique and will have a role when conceptual watershed models become dominant in use. Recognizing this, the Colorado Basin River Forecast Center (CBRFC) modernized its' statistical water supply forecasting system to take advantage of current technology and resources.

The system described includes a database, calibration, and operational component. The system was designed to make calibration and operational use fast, simple, and informative. System flexibility allows the hydrologic developer to structure procedures as desired using any information deemed appropriate. As such, weather indices and long range weather forecasts can be used. Application of the system significantly improved the quality of water supply forecasts issued by the CBRFC during water year 1993.

INTRODUCTION

Water resources management is critical in the Western U.S. Water supply forecasts provided by the National Weather Service (NWS) in cooperation with the Soil Conservation Service (SCS) and other cooperators allow resource managers and agricultural interests to scale operations to the available water supply. This service provides significant economic, environmental, and social benefits to the region. Improvements in the quality of forecasts can be directly translated into economic benefits (Castruccio et al., 1981).

As the population of the West has grown, the demands on water resources have, in many cases, exceeded supply. Water is perhaps the principal natural resource that governs the development of the Western U.S. As such, the need for accurate and timely water resource forecasts is important today and will become increasingly important in the future.

Water supply forecasts have been made in the West for nearly 60 years. Early efforts were simple and limited in scale. Starting in the 1950's, the NWS began to develop and issue seasonal volume forecasts. Techniques developed at that time included simple linear regression of lumped indices or multiple linear regression. Each NWS River Forecast Center (RFC) developed a similar yet unique method of calibrating and operationally evaluating water supply procedures. Computer and data resources available in the 1960's and early 1970's limited the development of optimal procedures.

With the development of conceptual (physically based) models and Extended Streamflow Prediction (ESP) within the NWS River Forecast System (NWSRFS) (Twedt et al., 1978; Curtis and Schakke, 1979; Day, 1985), the place of statistically based water supply procedures appeared limited. Clearly a conceptual model is more temporally flexible and more accurately can predict extreme events. Additionally it can provide a variety of hydrologic information about a specific basin. Unfortunately, the vision of this ultimate forecasting system sidelined the natural development and maintenance of the statistically based models. In time, the statistical forecasting system in use became outdated.

[1] Development and Implementation Hydrologist. NWS National Operational Hydrologic Remote Sensing Center. 6301 34th Avenue South, Rm 112, Minneapolis, MN 55450.

[2] Development and Operations Hydrologist. NWS Colorado Basin River Forecast Center. 337 North 2370 East, Salt Lake City, UT 84116.

Paper Presented Western Snow Conference 1994.

Conceptual model calibration and operational application of ESP has not been quick. Batch calibration and poor data resources make the calibration process awkward, time consuming, and non-optimal. In time, the calibration of conceptual models within NWSRFS will be interactive and near-optimal. Currently, the CBRFC and other RFCs use statistical procedures as the principal water supply forecasting tool with ESP as a complementary and time-distributing tool. Additionally, ESP, where calibrated and available, is used to provide intermediate duration (i.e. 2 week) volume and peak forecasts. The CBRFC recognizes that conceptual models represent the future of water resources forecasting and hence the dominant and supportive roles of statistical and conceptual models will eventually be reversed. In the meantime, it is imperative that the NWS provide the best possible guidance to its' users as the move is made towards this new environment.

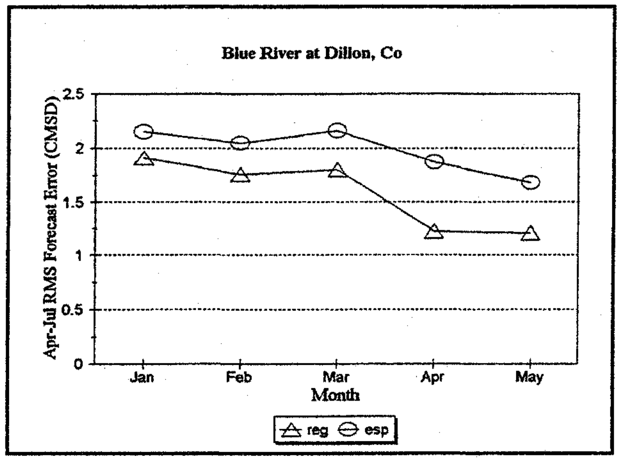

Figure 1. Comparison of regression and ESP based RMS forecast errors for the April-July period (1961-1988), Blue River at Dillon, CO..

Given that conceptual models offer greater flexibility and improved performance at the extremes, do statistical models have a place in a modernized forecasting system? Yes.

Parsimonious application of models is responsible use of public resources. In some cases, the calibration, maintenance, and operation of a conceptual model may not be economically justified. If the only forecast requirement is a seasonal volume, and extreme events are not problematic, statistical procedures may perform better than conceptual models and ESP. Figure 1 shows a comparison of regression and ESP based April-July RMS (root mean squared) forecast errors for the Blue River at Dillon, Colorado. In this case, regression provides better April-July forecasts but the management of the watershed requires shorter duration forecasts (i.e. weekly) that regression procedures cannot provide.

SYSTEM DESCRIPTION

The CBRFC Statistical Water Supply (SWS) system is comprised of three subsystems:

1. Database Subsystem.
2. Calibration Subsystem.
3. Operational Subsystem.

Database Subsystem

The development of the database subsystem preceded the SWS system. The system primarily makes use of the Climate Database (CLIMATDB) portion of the CBRFC system of five linked relational databases structured under the RAPPORT database management system (dbms) on a PRIME minicomputer. The CLIMATDB holds station information, monthy hydrometeorological data, and all statistical model parameters and forecasts. In addition to the CLIMATDB, the system makes use of near real-time data posted to the Processed Database (PROCESSDB) from the Raw Database (RAWDB). See Figure 2.

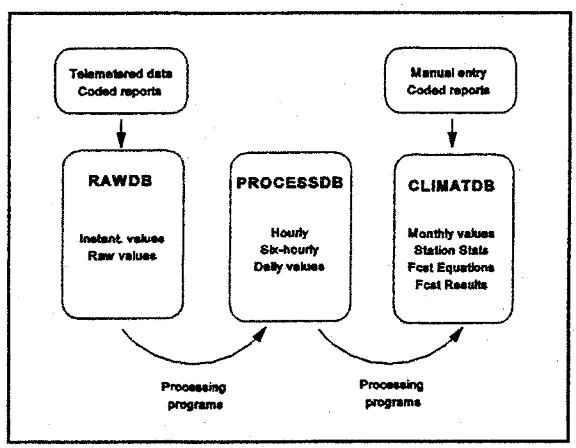

Figure 2. Database components and processing in support of SWS system.

The CLIMATDB contains a variety of data but is dominated by monthly streamflow volume, monthly diversion volume, end of month reservoir storage, monthly precipitation, and first of month snow water equivalent. Historical values were loaded from the most reliable source and new data are entered as they become available. Some stations have records that extend for over 100 years, however, the richest period of data begins in 1961.

Operationally, finalized values for stations are rarely available in time for inclusion in the statistical models. Therefore, provisional values are developed and stored in the CLIMATDB. Stations with automated telemetry are continuously collected in the RAWDB, processed into the PROCESSDB daily and moved to the CLIMATDB upon demand. For example, hourly instantaneous stream gage height might be processed into mean daily flow in the PROCESSDB and further processed to total monthly volume for preliminary storage in the CLIMATDB. Preliminary values are also available for many stations directly from the individuals or the agencies that collect them. These preliminary values are entered directly into CLIMATDB. When final values are available, the preliminary values are replaced. All data values in the CLIMATDB carry qualifiers that indicate if the value is observed, preliminary, estimated (low confidence), calculated (high confidence), or missing.

As will be described later, the operational system makes use of the PROCESSDB when exercising a first of month calibrated equation in a mid-month update.

It is important to note that water supply forecasts are usually made for the natural, unimpaired flow of the river basin at the forecast point, that is, the flow expected if water management activities were not performed upstream of the forecast point. To compute the natural streamflow, the observed streamflow is adjusted for all known and measured diversions and reservoir regulation upstream for which data are available. All data required to develop the natural flows are stored in the CLIMATDB and natural flows are programmatically computed based on the documented adjustments. Each fall, the CBRFC publishes a document that describes the natural flow adjustments made to each forecast point (Colorado Basin River Forecast Center, 1993).

Calibration Subsystem

The calibration subsystem was designed to be easy to use, flexible, and powerful. The system is able to calibrate headwater basins, local basins, and routed points. Calibration consists of:

1. Identification of forecast points.
2. Screening the CLIMATDB for appropriate candidate predictor variables.
3. Development of calibration files with selected candidate predictor variables.
4. Calibration program runs.
5. Selection of primary and alternate (optional) equations.
6. Storage of selected equations and historical results in the CLIMATDB.

At this time, steps 1-3 and 6 are performed on the PRIME minicomputer while steps 4 and 5 are done on desktop PCs operating under the QNX operating system..

Calibrational Approach

A sound and consistently applied calibration philosophy is critical to the development of a functional set of statistical forecast equations. In developing a calibrational philosophy, the user should consider the following issues:

How often will forecasts be generated?
Forecast frequency and timing dictate the type of equations that should be developed. If mid-month updates are required, the user must choose between developing explicit mid-month equations or executing first of month equations in a mid-month mode. If the later configuration is chosen, the system will perform best if daily incremental values are available for the monthly stations used in the equation calibrations.

Are the best predictive data available when forecasts must be generated?
Data availability must be considered in terms of when the forecast equations will be evaluated. If equations must be run on the first of the month, then only stations that are usually available on the first of the month should be considered. There is no value in including stations that will always be missing when forecasts must be generated. Some work suggests that the inclusion of future variables in a forecast equation may degrade the forecast performance (Garen, 1992).

If equation results are to be routed downstream, what base historical period is to be employed?
A river system of regression based forecast equations should exhibit a high level of connectivity. This connectivity requires that the historical calibration period overlap to the greatest extent possible. In some cases, headwater procedures with the lowest errors may not be optimal for the system if their period of record causes problems in routing procedures downstream.

Are missing years to be tolerated?
The calibration program does not deal with missing data in the best possible fashion. A specific combination of variables is evaluated using only the years in which all individual variables are non-missing. Thus, if missing values are included in the calibration input file, the specific years included will vary from one combination to another. This can lead to misleading results if an individual variable is missing in a year that was particularly difficult to forecast. In this case, the resulting CVSE for the combination of variables may be lower because a specific year was omitted rather than because a specific variable was included.

Finding Candidate Variables

Once a forecast point and its' time series have been identified in the CLIMATDB, the developer screens the CLIMATDB for candidate variables. The objective of this process is to find a good set of candidate variables from which the forecast equation(s) can be developed. Prescreening is a crucial calibration step as the inclusion of junk variables slows the calibration program and may contaminate the results.

Specified program inputs include the runoff period (i.e. April through July), the correlation period (i.e. 1961 through 1990), and the geographic region to be considered. The candidate search program identifies all precipitation and snow measurement stations in the geographic region and computes single and contiguous multiple month correlations for precipitation stations and first of month correlations with snow water equivalent stations. Provided correlations include the number of observation pairs. Correlations with current and previous year streamflow periods are also computed to assess the potential benefit of a streamflow carryover variable. A sample of the output provided by the candidate search program is shown in Figure 3.

```
          REGCAND  -  CBRFC NEXTREG CANDIDATE SEARCH PROGRAM

COLORADO - LAKE GRANBY, GRANBY, NR                    CO
LAT:401055 LONG:1055214 ELEV:  8280    DRAINAGE AREA:  12900
FORECAST PERIOD:      APR-JUL

CORRELATION PERIOD:  1961-1990

 Available:      4         5         6         7         8         9
                 1234567890123456789012345678901234567890123456789
 GBYC2 QCMPAZZ                     m000000000000000000000000000000c

[O]bserved [m]issing [e]stimated [p]rovisional [c]alculated [ ]no record

LAT/LONG BOX:    ------4042------
                 |              |
          10610  |              |  10535
                 |              |
                 ------4000------

Found:    4 Precip Stations
          7 SNOTEL Precip Sites
         11 Snow Courses
          7 SNOTEL Pillow Sites

PRECIPITATION SITES:                                 OCT      NOV      DEC      JAN      FEB      MAR      APR      MAY
GDLC2 PPMRZZZ CO GRAND LAKE 6SSW      4011 10552 8288 OCT  0.28(30) 0.48     0.68     0.74     0.74     0.79     0.80     0.85
                                                 NOV           0.45(30) 0.65     0.69     0.69     0.74     0.77     0.81
                                                 DEC                    0.58(30) 0.56     0.62     0.65     0.74     0.74
                                                 JAN                             0.22(30) 0.31     0.37     0.51     0.51
                                                 FEB                                      0.29(30) 0.33     0.45     0.45
                                                 MAR                                               0.16(30) 0.36     0.33
                                                 APR                                                        0.41(30) 0.31
                                                 MAY                                                                 0.00(30)

SNOTEL PRECIPITATION SITES:                          OCT      NOV      DEC      JAN      FEB      MAR      APR      MAY
JWRC2 PPMRMZZ CO JOE WRIGHT           4032 1055310120 OCT  0.38(11) 0.69     0.80     0.72     0.74     0.78     0.86     0.87
                                                 NOV           0.63(11) 0.75     0.59     0.62     0.65     0.79     0.82
                                                 DEC                    0.59(11) 0.35     0.49     0.47     0.73     0.69
                                                 JAN                            -0.20(12) 0.04     0.08     0.54     0.46
                                                 FEB                                      0.27(12) 0.30     0.69     0.61
                                                 MAR                                               0.09(12) 0.76     0.52
                                                 APR                                                        0.90(12) 0.64
                                                 MAY                                                                -0.16(12)

SNOW MEASUREMENT SITES:                              JAN      FEB      MAR      APR      MAY
CPAC2 SWIRZZZ CO CAMERON PASS         4031 10553 10285 0.80(17) 0.70(30) 0.70(30) 0.72(30) 0.73(30)

SNOTEL PILLOW SITES:                                 JAN      FEB      MAR      APR      MAY
WLLC2 SWIRMZZ CO WILLOW CREEK PASS    4021 10606  9540 0.60(12) 0.66(30) 0.64(30) 0.69(30) 0.72(30)

CARRYOVER PERIOD CORRELATIONS:

PREVIOUS WATER YEAR PERIODS:
                          OCT-SEP   0.00 (29)
                          MAR-JUN  -0.12 (29)
                          MAR-JUL  -0.04 (29)
                          MAR-SEP   0.00 (29)
                          APR-JUN  -0.12 (29)
                          APR-JUL  -0.04 (29)
                          APR-SEP   0.00 (29)

CURRENT WATER YEAR PERIODS:                          OCT      NOV      DEC      JAN      FEB      MAR      APR      MAY
GBYC2 QCMPAZZ CO COLORADO - LAKE GRAN 4010 10552 8280 OCT  0.34(29) 0.41     0.48     0.51     0.53     0.55     0.45     0.59
                                                 NOV           0.48(29) 0.62     0.65     0.67     0.68     0.46     0.58
                                                 DEC                    0.72(30) 0.72     0.74     0.74     0.42     0.55
                                                 JAN                             0.53(30) 0.61     0.62     0.33     0.52
                                                 FEB                                      0.54(30) 0.54     0.27     0.51
                                                 MAR                                               0.42(30) 0.22     0.50
                                                 APR                                                        0.17(30) 0.49
                                                 MAY                                                                 0.50(30)
```

Figure 3. Sample output from variable candidate search program (condensed).

By studying the candidate search results, the developer should be able to build a picture of what information should be included in the calibration runs. A set of stations with reasonable correlations should stand out from those stations that have little predictive use.

Building Calibration Input Files

Based on the calibration objectives, the developer then interactively builds the calibration input file(s) from the information provided in the candidate search. In addition to extracting specified data from the CLIMATDB, the file building utility allows the user to create weighted combined variables and time-shifted variables, delete unwanted variables, and append variables from other calibration input files. The user has the option of restricting the temporal range when saving the calibration input file. To facilitate the development of routed procedures, the file building utility allows the user to extract historical equation results from upstream calibrated points.

The file building utility exists on both the PRIME and QNX platforms although data extraction can only be performed on the PRIME. The file building utility options and a sample of the type of data collected are shown in Figure 4.

```
   REGDATA - CBRFC NEXTREG DATAFILE MANAGEMENT PROGRAM

 Options:  1  Load a data file          6  Change var type
           2  Append from a data file   7  Shift a variable
           3  Add Y variable(s)         8  Create a variable
           4  Add X variable(s)         9  Show variables
           5  Remove variable(s)        S  Save
           Q  Quit

 Y 1  GBYC2/QCMPAZZ  Ap-Jl  COLORADO - LAKE GRANBY, GRANBY, NR
 X 1  GBYC2/QCMPAZZ  De-Mr  COLORADO - LAKE GRANBY, GRANBY, NR
 X 2  GDLC2/PPMRZZZ  Oc-Mr  GRAND LAKE 6SSW
 X 3  GLKC2/PPMRZZZ  Oc-Mr  GRAND LAKE 1NW
 X 4  NILC2/SWIRZZZ  Ap     NORTH INLET GRAND LAKE
 X 5  GLAC2/SWIRZZZ  Ap     GRAND LAKE
 X 6  PKVC2/SWIRZZZ  Ap     PARK VIEW
 X 7  CPAC2/SWIRZZZ  Ap     CAMERON PASS
 X 8  GRYC2/SWIRZZZ  Ap     GRANBY
 X 9  WLLC2/SWIRMZZ  Ap     WILLOW CREEK PASS
 X10  PHTC2/SWIRMZZ  Ap     PHANTOM VALLEY
 X11  LKIC2/SWIRMZZ  Ap     LAKE IRENE
```

```
                               4         5         6         7         8         9
                               1234567890123456789012345678901234567890123456789 0
 Y 1 GBYC2 QCMPAZZ Ap-Jl                           00000000000000000000000000000c
 X 1 GBYC2 QCMPAZZ De-Mr                           000000000000000000000000000000
 X 2 GDLC2 PPMRZZZ Oc-Mr                           000000000000000000000000000000
 X 3 GLKC2 PPMRZZZ Oc-Mr                           0000000000000000000000000000c0
 X 4 NILC2 SWIRZZZ Ap                              000000000000000000000000000000
 X 5 GLAC2 SWIRZZZ Ap                              000000000000000000000000000000
 X 6 PKVC2 SWIRZZZ Ap                              000000000000000000000000000000
 X 7 CPAC2 SWIRZZZ Ap                              000000000000000000000000000000
 X 8 GRYC2 SWIRZZZ Ap                              000000000000000000000000000000
 X 9 WLLC2 SWIRMZZ Ap                              cccccccccccccccccc000000000000
 X10 PHTC2 SWIRMZZ Ap                              ccccccccccccccccccc00000000000
 X11 LKIC2 SWIRMZZ Ap                              ccccccccccccccccccc00000000000
```

Figure 4. Sample of file building utility options and data collection.

Calibration Analysis

Once calibration input files are available, the developer is ready to run the calibration program. The calibration program in use at the CBRFC was developed by the SCS Water Supply Staff in Portland, OR (Garen, 1992). For a detailed description of the process, please refer to the referenced article. In a nutshell, the program makes use of three techniques that dramatically improve the quality of developed equations: principal components analysis, cross validation standard error computation, and combination analysis.

Principal components analysis. This technique involves the transformation of the original input variable matrix. The analysis creates new predictor variables that are truly independent and therefore eliminates the problems associated with intercorrelation. Multiple linear regression is performed on the new variables and developed coefficients are expressed in terms of the original input variables.

Cross validation standard error (CVSE) computation. This is variation of a split sample test. Years are withheld one at a time and predicted with the candidate equation calibrated without the benefit of the withheld year. A series of predicted errors are generated from which the CVSE is computed. The CVSE is a better measure of an equation's ability to forecast.

Combination analysis. Selection of variables is a problem that has always plagued developers. Given current computing resources, evaluation of all possible combinations is simply not feasible. The calibration program utilizes a convergence algorithm that provides the developer with the 20 best combinations of candidate variables in a reasonable amount of time. The objective function is minimization of the CVSE.

Aside from the above attributes, the calibration program does not limit the type of information it can utilize. As a result, the system allows the developer to include variables such as the Southern Oscillation Index (SOI) (Koch, 1988) or a categorical 90 day precipitation forecast. A high degree of flexibility allows the developer to lend hydrologic significance to the statistical model through inclusion of appropriate candidate variables.

Equation calibration runs are typically batched as they can take up to six hours to complete depending upon the number of variables, the number of years, and the speed of the computer. Once complete, programs exist to review the results in tabular and graphical form. The developer can review candidate equations to ensure they operate favorably under the required conditions.

Equation Selection

Selection of equations requires the application of hydrologic judgment. Within an individual forecast equation the user must develop a balance between historical performance, as indicated by the CVSE, and the diversity and spatial distribution of station data. Good geographic coverage and variables that address more components of the seasonal water balance have a greater likelihood of performing well under unusual conditions. The equation with the lowest CVSE may only have one or two closely grouped stations and therefore may be unacceptable. If a set of monthly equations is to be developed, month to month consistency should be maintained to the highest level reasonable. Some shifting of stations will naturally occur from month to month but the overall structure of the equations should be consistent. If monthly equations are developed independently, the resulting set of monthly forecasts may be less consistent than they could be. Forecast adjustments from month to month should reflect a change in the watershed state rather than the use of a different forecast equation.

Extracting Calibration Results

Parameters for selected equations are interactively stripped from the calibration run output and loaded into the CLIMATDB. When an equation is loaded into the CLIMATDB, it must be identified as primary or alternate. There can be one primary and up to twenty-six alternate equations for a specific forecast point, forecast time series, forecast period, and forecast date (e.g., Colorado at Kremmling, CO, adjusted streamflow, April through July, February 1). Primary equations normally represent the best balanced equation that meets all of the period of record requirements for routing downstream. Alternate equations are optionally developed to provide additional insight and perspective. Examples of alternate equations might include only snow data, only precipitation data, inclusion of a weather index, or exclusive use of telemetered stations. When a primary equation is loaded, the historical performance of the equation is stored for use in routed equations downstream. When an alternate equation is loaded, the user can overwrite an existing alternate equation or the system will assign a new alternate code. Figure 5 shows a primary and alternate equation as loaded into the CLIMATDB.

```
GBYC2/QCMPAZZ              Apr-Jul  61-90   JSE:  32.91  SE:  30.85     GBYC2/QCMPAZZ              Apr-Jul  61-90   JSE:  36.20  SE:  34.05
COLORADO                   0401P    100292  JR2:   0.73  R2:   0.76     COLORADO                   0401s    100292  JR2:   0.67  R2:   0.71
LAKE GRANBY, GRANBY,                RKH     n  :     30                 LAKE GRANBY, GRANBY,                RKH     n  :     30
COLORADO HEADWATERS                         #pc:      1  #var:    6     COLORADO HEADWATERS                         #pc:      1  #var:    5

     GBYC2  QCMPAZZ   Dec-Mar      3.41                                      NILC2  SWIRZZZ   Apr          3.95
     GDLC2  PPMRZZZ   Oct-Mar      6.78                                      GLAC2  SWIRZZZ   Apr          3.92
     NILC2  SWIRZZZ   Apr          3.44                                      PKVC2  SWIRZZZ   Apr          5.31
     GLAC2  SWIRZZZ   Apr          3.54                                      CPAC2  SWIRZZZ   Apr          1.80
     PKVC2  SWIRZZZ   Apr          4.90                                      PHTC2  SWIRMZZ   Apr          4.27
     CPAC2  SWIRZZZ   Apr          1.71                                      Intercept                     4.45
     Intercept                   -27.95
```

Figure 5. Sample primary and alternate forecast equations as loaded into the CLIMATDB.

Routed Procedures

There are basically two types of equations: headwater and routed. Headwater equations are developed with predictor variables such as precipitation and snow water. Routed procedures are developed from the results of upstream forecast equations. When calibrated equations are loaded into the CLIMATDB, their historical results are also stored. Thus, when developing input data files for downstream routed forecast points, the system can access these data just as they would precipitation or snow water. It is important to develop the routed equations as a function of upstream forecasts rather than upstream observations. This configuration deals with procedural bias in upstream equations and allows the uncertainty associated with upstream forecasts to be properly translated downstream.

Although rather simple, the user has some choices to make in developing routed equations. The choices are associated with which upstream points will be included and whether or not individual weights will be utilized. From a statistical perspective, there is little reason to assign all routed elements the same relative weight. From a hydrologic perspective, this may make the most sense. In most cases, if the calibration program is allowed to assign individual weights, the upstream point that is the most predictive will get the largest relative weight. Situations may occur where an upstream basin that contributes five percent of the downstream volume dominates a forecast equation. This may or may not be acceptable. If it is not, the user can combine the desired upstream forecasts into a single variable using the file building utility. Once equation coefficients have been estimated, they must be applied to each upstream forecast entity prior to loading into the CLIMATDB.

Operational Subsystem

The operational system runs on the PRIME minicomputer and makes use of information stored in the CLIMATDB and PROCESSDB databases. When run, the operational program develops and stores up to five forecast quantities for each evaluated equation.

Pure Computed
This is the raw forecast equation result. For routed points, this reflects the use of purely raw equation results upstream.

Computed with Coordinated
This is only used for routed points and is the result of the routing equation when coordinated upstream forecasts are used as input.

Coordinated
This is the adjusted or agreed-to forecast for the point. These values are forced to adhere to the established rounding convention.

NWS Preferred
If coordination was not required, this is the forecast the NWS would have issued.

Coordinating Agency Preferred
This is similar to the NWS preferred, but is often the coordinating agencies initial or proposed forecast for the point.

It is anticipated that the collection of these five forecast quantities will provide insight into the benefits of judgmental forecast adjustment and coordination.

Connectivity

When statistical water supply equations are developed, they are not explicitly identified as part of a major watershed forecast group. Individual equations stored in CLIMATDB are content independent of their application within a forecast river system.. Normally, a river forecast system is a collection of headwater and routing procedures. Equations are evaluated and routed downstream in a specific order that creates consistency and continuity. Equation Group Driver Datasets (EGDDs) specify the order as well as what procedures are available to the forecast process. There is no limit to the number of EGDDs and a specific equation can be referenced in more than one EGDD.

Operational Analysis

The operational water supply program can be run in either batch or interactive mode. The purpose of the batch program is to quickly evaluate all of the indicated equations within the specified EGDD. The output must be stored on disk and is normally printed for review. The interactive version is used to review and adjust forecasts as a result of additional hydrologic information, judgement, or coordination with another water management agency. When the operational program is invoked, the user must supply the desired EGDD file and the forecast date.

The notion of forecast date is very important because only forecasts with the same date can be routed downstream. First of month forecasts are always assigned the first of month date even if evaluated on the third or forth day of the month. First of month forecasts access monthly data stored in the CLIMATDB.

If a date other than the first of a month is specified, the operation is considered 'mid-month'. Mid-month operations can either use explicitly calibrated mid-month equations or an equation designed for use on the first of a future month. Typically, one would use equations developed for the first of the next month, but this is not required. The user may select any future month but is required to characterize weather conditions between the specified forecast date and the date for which the equations were developed. Additionally, the user must characterize weather conditions between the first of the current month and the forecast date. The mid-month operation then uses the specified weather characterizations, daily observations from the PROCESSDB, and monthly observations and means from the CLIMATDB to estimate the data values required in the forecast equations. Mid-month operations are useful for updating forecast guidance and developing scenario forecasts. It should be noted that the uncertainty associated with this type of mid-month forecast is underestimated because some of the required data must be estimated. Scenario forecasts involve running a future month's set of equations (i.e. June 1) with various assumptions about future moisture input. Although interesting, special care must be taken when interpreting scenario forecasts since the uncertainties and probability distributions are not known.

Figure 6 shows the typical output from the batch operational program for a headwater and routed forecast point. Note that the output provides statistics on observed volumes, historical equation performance, a current equation evaluation, and a history of forecasts made in the current water year. The evaluated equation is broken down by individual data value so that the user can easily see which values are observed, provisional, or estimated and how each contributes to the total computed forecast. The forecast season history helps the user maintain month to month consistency and track procedure performance.

```
GBYC2  QCMPAZZ    Apr-Jul Volume                          Publ. Code:  0301P
Forecast Date:    Mar 1 1994
Major Watershed:  COLORADO HEADWATERS                  61-90 Average:  214.0
River:            COLORADO                                   Median :  211.4
Location:         LAKE GRANBY, GRANBY,                      Std dev:   61.6

Calibration JSE:    40.7                            Hist Max (1984):  342.3
             R2:    0.58                                 Min (1977):  117.2
              n:      30
            per:   61-90

ID      PC       NAME                   MTH       COEF     VALUE   %AVG   %CONT
---------        --------------------   ---       ----     -----   ----   -----
GBYC2/QCMPA      COLORADO - LAKE GRAN   Dec       5.66     3.79p   103.     10.
                                        Jan       5.66     3.90p   105.     11.
                                        Feb       5.66     3.42p   107.      9.
GDLC2/PPMRZ      GRAND LAKE 6SSW        Oct       8.96     2.49p   259.     11.
                                        Nov       8.96     0.75p    80.      3.
                                        Dec       8.96     0.57p    63.      2.
                                        Jan       8.96     1.00p   108.      4.
                                        Feb       8.96     1.00p   132.      4.
PKVC2/SWIRZ      PARK VIEW              Mar       5.79     8.60p   112.     24.
CPAC2/SWIRZ      CAMERON PASS           Mar       2.27    18.60p    84.     20.
                 Intercept                       18.31

                          MP (kaf/%)    EP(%)   RMAX (kaf/%)    RMIN (kaf/%)
Computed (pure)         225.4 / 105.     50.    278.9 / 130.    171.9 /  80.
Coordinated             215.0 / 100.     60.    245.0 / 114.    185.0 /  86.
NWS preferred           205.0 /  96.     69.
Coord. agency pref.     215.0 / 100.     60.

Forecast Season History (KAF/%):
          Jan        Feb        Mar        Apr        May        Jun
Comp
Fcst      217/102    221/103    225/105
Rmax      271/127    273/128    279/130
Rmin      164/ 77    169/ 79    172/ 80

Coord
Adj        -2        -16        -10
Fcst      215/100    205/ 96    215/100
Rmax      270/126    240/112    245/114
Rmin      160/ 75    170/ 79    185/ 86

NWS
Fcst      205/ 96    210/ 98    205/ 96

C.A.
Fcst      215/100    205/ 96    215/100
```

```
EGLC2  QCMPAZZ    Apr-Jul Volume                          Publ. Code:  0301P
Forecast Date:    Mar 1 1994
Major Watershed:  COLORADO HEADWATERS                  61-90 Average: 1362.0
River:            COLORADO                                   Median : 1349.6
Location:         DOTSERO, NR                               Std dev:  478.2

Calibration JSE:   270.0                            Hist Max (1991): 3151.9
             R2:    0.69                                 Min (1977):  563.8
              n:      30
            per:   61-90

Evaluation w/ pure upstream forecasts:

ID      PC       NAME                   MTH       COEF     VALUE    %AVG   %CONT
---------        --------------------   ---       ----     -----    ----   -----
HTSC2/QCMFA      COLORADO - HOT SULPH   Apr-Jul   1.03    390.67p   101.     30.
WFDC2/QCMFA      WILLIAMS FORK - WILL   Apr-Jul   1.03     65.63p    98.      5.
GMRC2/QCMFA      BLUE - GREEN MTN RES   Apr-Jul   1.03    235.51p    90.     18.
GPSC2/QCMFA      EAGLE - GYPSUM, BLO    Apr-Jul   1.03    282.52p    91.     21.
EGLC2/QCMFL      COLORADO - DOTSERO,    Apr-Jul   1.03    341.08p    99.     26.
                 Intercept                      -40.34

                          MP (kaf/%)    EP(%)   RMAX (kaf/%)    RMIN (kaf/%)
Computed (pure)        1309.3 /  96.     50.   1663.8 / 122.    954.8 /  70.

Evaluation w/ coordinated upstream forecasts:

ID      PC       NAME                   MTH       COEF     VALUE    %AVG   %CONT
---------        --------------------   ---       ----     -----    ----   -----
HTSC2/QCMFA      COLORADO - HOT SULPH   Apr-Jul   1.03    380.00c    98.     29.
WFDC2/QCMFA      WILLIAMS FORK - WILL   Apr-Jul   1.03     65.00c    97.      5.
GMRC2/QCMFA      BLUE - GREEN MTN RES   Apr-Jul   1.03    250.00c    95.     19.
GPSC2/QCMFA      EAGLE - GYPSUM, BLO    Apr-Jul   1.03    290.00c    94.     22.
EGLC2/QCMFL      COLORADO - DOTSERO,    Apr-Jul   1.03    340.00n    99.     26.
                 Intercept                      -40.34

                          MP (kaf/%)    EP(%)   RMAX (kaf/%)    RMIN (kaf/%)
Computed (w/coord.)    1319.1 /  97.     49.   1673.6 / 123.    964.6 /  71.
Coordinated            1320.0 /  97.     48.   1670.0 / 123.    970.0 /  71.
NWS preferred          1320.0 /  97.     48.
Coord. agency pref.    1300.0 /  95.     51.

Forecast Season History (KAF/%):
          Jan        Feb        Mar        Apr        May        Jun
Comp
Fcst     1278/ 94   1263/ 93   1309/ 96
Rmax     1682/124   1619/119   1664/122
Rmin      874/ 64    907/ 67    955/ 70

Comp wc
Diff      -34         10         10
Fcst     1244/ 91   1272/ 93   1319/ 97
Rmax     1649/121   1628/120   1674/123
Rmin      840/ 62    917/ 67    965/ 71

Coord
Adj         6         -2          1
Fcst     1250/ 92   1270/ 93   1320/ 97
Rmax     1650/121   1630/120   1670/123
Rmin      850/ 62    910/ 67    970/ 71

NWS
Fcst     1250/ 92   1260/ 93   1320/ 97

C.A.
Fcst     1330/ 98   1250/ 92   1300/ 95
```

Figure 6. Sample of batch program output for basin (left) and routed (right) forecast points.

The interactive version of the operational program is primarily an adjustment and coordination utility. Based again on the forecast date and EGDD, the user interactively displays and enters forecast information. Values entered include the (1) coordinated most probable, reasonable maximum and minimum, (2) NWS preferred most probable forecast, and (3) coordinating agency preferred most probable forecast. Default values are provided where possible to improve the ease of operation.

The EGDD defines the default order of processing but the user can move backward or jump to any point within the EGDD. Options exist to re-evaluate an equation, display full equation results, display and place the current forecast within the historical distribution, and review alternate equations. For routed points, the user can travel upstream and review those points, both routed and headwater, that contribute to the point currently under review. This upstream process can be repeated and allows the user to retrace the process back downstream to the original starting point. As such, a user could theoretically review all of the procedures in a major river basin by starting at the most downstream point.

An example of the information available during an interactive session is shown in Figure 7 for both a headwater and routed forecast point.

```
GBYC2/QCMPAZZ  Apr-Jul Vol      61-90 Average:   214.0   Publ.code :  0301P
COLORADO                              Median :   211.4   Max (1984):  342.3
LAKE GRANBY, GRANBY,                  Std dev:    61.6   Min (1977):  117.2

              Jan    Feb    Mar    Apr    May    Jun  Calib JSE:   40.7
Comp          217    221    225                              R2:   0.58
Coord         215    205    215                               n:     30
NWS           205    210    205                          period:  61-90
C.A.          215    205    215

                       MP (kaf/%)    EP(%)    RMAX (kaf/%)   RMIN (kaf/%)
Computed (pure)        225.4 / 105.    50.    278.9 / 130.   171.9 /  80.
Coordinated            215.0 / 100.    60.    245.0 / 114.   185.0 /  86.
NWS preferred          205.0 /  96.    69.
Coord. agency pref.    215.0 / 100.    60.

Adjust?  (yes [no] display eval info jump back menu top quit):  y
Coordinated MP   [  215.0]:
            RMAX [  245.0]:
            RMIN [  185.0]:
NWS  pref.  MP   [  205.0]:
C.A. pref.  MP   [  215.0]:
```

```
EGLC2/QCMPAZZ  Apr-Jul Vol      61-90 Average:  1362.0   Publ.code :  0301P
COLORADO                              Median :  1349.6   Max (1991): 3151.9
DOTSERO, NR                           Std dev:   478.2   Min (1977):  563.8

              Jan    Feb    Mar    Apr    May    Jun  Calib JSE:  270.0
Comp         1278   1263   1309                              R2:   0.69
Comp wc      1244   1272   1319
Coord        1250   1270   1320                               n:     30
NWS          1250   1260   1320                          period:  61-90
C.A.         1330   1250   1300

                       MP (kaf/%)    EP(%)    RMAX (kaf/%)    RMIN (kaf/%)
Computed (pure)       1309.3 /  96.    50.   1663.8 / 122.   954.8 /  70.
Computed (w/coord.)   1319.1 /  97.    49.   1673.6 / 123.   964.6 /  71.
Coordinated           1320.0 /  97.    48.   1670.0 / 123.   970.0 /  71.
NWS preferred         1320.0 /  97.    48.
Coord. agency pref.   1300.0 /  95.    51.

Adjust?  (yes [no] display eval upstrm info jump back menu top quit):  u

                  Upstream Adjustment for EGLC2/QCMPAZZ

COLORADO                        Computed:    1319.1/ 97% of avg
DOTSERO, NR                     Coordinated: 1320.0/ 97% of avg
                                difference:    -0.9/  0% of avg ( -0.1% diff)

## ID/PC        Per      Location               Comp.Fcst   Coord.Fcst   EP
 1 HTSC2/QCMPA Apr-Jul  HOT SULPHUR SPGS       390p /101    380c / 98   55
 2 WFDC2/QCMPA Apr-Jul  WILLIAMS FORK RES,      65r / 98     65c / 97   51
 3 GMRC2/QCMPA Apr-Jul  GREEN MTN RES          245r / 94    250c / 95   46
 4 GPSC2/QCMPA Apr-Jul  GYPSUM, BLO            282p / 91    290c / 94   45
 5 EGLC2/QCMPL Apr-Jul  DOTSERO, NR            341p / 99    340n / 99   50

Enter # or all ([r]eturn, jump, back , menu, top, quit)
```

Figure 7. Sample dialog for interactive version of operational program. The headwater point on the left shows the adjustment option. The routed point on the right shows upstream option.

Operational Utilities

In addition to the above, the operational system provides for forecast dissemination and publication support as well as data requirements and data availability reports. Utilities exist to develop forecast summaries in the appropriate text format for (1) the "Water Supply Outlook for the Western United States", (2) "CBRFC Water Supply Outlook Reports", and (3) internal NWS products transmitted over AFOS (NWS communications system). Again, the content of these summaries is controlled by EGDDs and some EGDDs are maintained solely for publication and product summary purposes. Direct database generation of forecast information saves time and reduces errors.

The data requirements utility develops a complete summary of all data required by a group of equations for a month, a set of months or an entire forecast season. Again, the group of equations is defined by the EGDD. Data requirements are summarized by type and month. A data requirements list allows the RFC to properly align data collection efforts and advise cooperating agencies of their operational need for specific pieces of information. Figure 8 shows the type of information provided by the data requirements utility.

```
Station Name                CHB5  PEDSTEP  AGCY ID    O N D J F M A M J J A S
------------                ----- -------  --------   - - - - - - - - - - - -
BERTHOUD SUMMIT             BTSC2 PPMRMZZ  05K14S     X X X X X X X X
CASCADE                     CSCC2 PPMRMZZ  07M05S     X X X
LONE CONE                   LNCC2 PPMRMZZ  08M07S     X X X         X X

ASPEN 1 SW                  ASPC2 PPMRZZZ  05-0372    X X X X X X X X
BRECKENRIDGE                BRGC2 PPMRZZZ  05-0909                  X X
COLLBRAN                    CBNC2 PPMRZZZ  05-1741    X X X         X X

COLORADO - LAKE GRANBY, G   GBYC2 QCMPAZZ  09018500       X X X X X X
COLORADO - DOTSERO, NR      EGLC2 QCMPLZZ  09070500                 X X
EAST - ALMONT               ALEC2 QCMRZZZ  09112500                 X X
DOLORES - DOLORES           DOLC2 QCMRZZZ  09166500   X X X         X X

ARROW                       AROC2 SWIRMZZ  05K06S             X X X X
BRUMLEY                     BRMC2 SWIRMZZ  06K40S             X X X X
BERTHOUD SUMMIT             BTSC2 SWIRMZZ  05K14S           X X X X X
BUTTE                       BUTC2 SWIRMZZ  06L11S             X X X X

BLUE RIVER                  BLLC2 SWIRZZZ  06K21              X X X
COCHETOPA PASS              CHPC2 SWIRZZZ  06L06              X X X
CAMERON PASS                CPAC2 SWIRZZZ  05J01              X X X
GRAND LAKE                  GLAC2 SWIRZZZ  05J19              X X X X
```

Figure 8. Sample output from data requirements utility.

The data availability utility assess the current readiness of the CLIMATDB in terms of the data requirements of a group of equations on a specific date. Users will typically run this prior to making a forecast run. The utility provides a summary of what is available and what is still missing. This process helps to focus data collection efforts in anticipation of forecast runs.

SYSTEM IMPLEMENTATION AND RESULTS

After reviewing the issues described under "Calibrational Approach", the CBRFC developed the following guidelines for primary equation development. Alternate equations could be developed without restrictions of any kind.

1. Forecast equations would be developed monthly, January through June for the upper Colorado and Great Basin, and January through May in the lower Colorado.

   Since monthly equations were developed, the use of future data was prohibited. Thus an equation developed for use on January 1 could not reference a February 1 snow water equivalent.

2. The base period of forecasts would be 1961-1990. This would allow for routing of forecast results all through the system.

3. Missing data values would not be permitted. If a value was missing and could not be reliably estimated, the station value would not be used.

Formal development of the Statistical Water Supply (SWS) system began in the spring of 1991. The SWS prototype system was completed in the fall of 1991 and calibrations were made for all water supply points in Arizona as well as the Gunnison Basin in the upper Colorado. The system worked well in water year 1992. All CBRFC water supply points were calibrated in the fall of 1992 and the SWS system assumed full operational status in water year 1993. Scratch calibration (as opposed to recalibration) of all CBRFC regression based water supply procedures is a significant task. The fact that this was accomplished within a period of three months attests to the simplicity, utility, and performance of the calibration system. Currently the system holds more than a thousand calibrated equations.

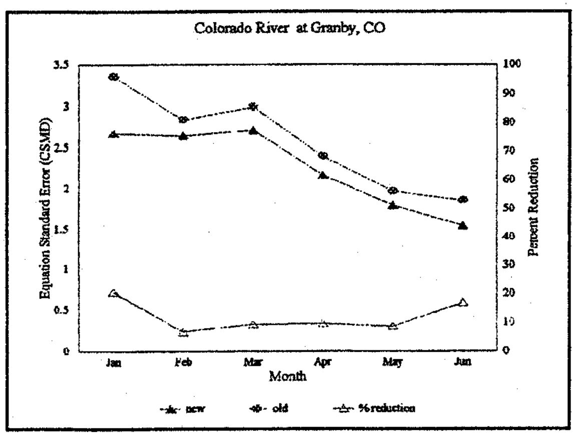

Figure 9. Comparison of old and new forecast equation standard error, Colorado River at Granby, CO.

Although it is not entirely fair to do so, it is interesting to compare the error statistics associated with the old and new statistical water supply forecasting systems. The old system relied upon a single seasonal equation that incorporated lumped indices of fall, winter, and spring precipitation, snow water equivalent (usually April 1), and an antecedent soil moisture term indexed by a period of previous streamflow runoff for the same forecast point. Not all indices were not required, individual station weights could be applied with indices, standard multiple linear regression was utilized, and line segments could be used to overcome non-linearity.

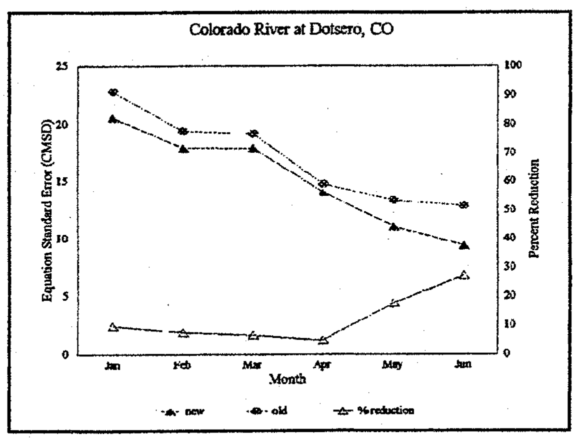

Figure 10. Comparison of old and new forecast equation standard error, Colorado River at Dotsero, CO.

Figures 9 and 10 show a comparison of the equation standard errors for a headwater and routed water supply forecast point in the upper Colorado. Similar results were experienced throughout the CBRFC forecast area (Colorado Basin and eastern Great Basin). Since these are forecast equations, the more appropriate comparative index would be the CVSE. Unfortunately, CVSEs are not available for the older set of equations. Note that the new system provided a significant reduction in the standard error which translates into a higher level of forecast confidence and a narrower distribution of forecast uncertainty.

FUTURE SYSTEM ENHANCEMENTS / MODIFICATIONS

The primary planned system enhancement involves migration of the SWS system to a networked workstation environment. Database and computing resources available at the CBRFC mandated the multi-platform implementation of the prototype system. This configuration severely limits the extension of the SWS system to other NWS River Forecast Centers. Additionally, the calibration program does not have access to the database. This access will streamline the process and eliminate restrictions on the use of future data utilized in single seasonal equations. Porting efforts are currently underway and should be fully complete and operational prior to water year 1996. A limited scale version of the workstation-based system will be tested in water year 1995.

To date, the implementation of the SWS system at the CBRFC has not taken full advantage of the system flexibility. As stated previously, the system is capable of utilizing any information deemed hydrologically appropriate Forecast errors originate from three sources: model error, data error, and weather error (uncertainty associated with future hydrometeorologic conditions). Of these, weather error is dominate, particularly at the beginning of the forecast season (Schaake and Peck, 1985). As such, the inclusion of long range weather forecasts and weather indicies (Koch, 1989) in regression equations may offer significant benefits in terms of forecast accuracy and associated confidence.

SUMMARY

Statistically based procedures have been, and will continue to be the primary tool for forecasting seasonal water supply volumes in the Western U.S. Over time, physically based models will assume a larger role, but in many cases, the effort and cost may not be justified. As such, the NWS has started the process of modernizing the techniques and tools required to provide the best possible guidance to its' user community. The SWS system takes advantage of relational databases, modern statistical techniques, and a systems approach that makes calibration and operation of the system straightforward and informative. Efforts are underway to extend the developed system to other NWS River Forecast Centers.

REFERENCES

Castruccio, P., H. Louts, D.Lloyd, and P. Newman, 1981. Applications Systems Verification and Transfer Project, Volume II: Cost Benefit Analysis for the ASVT on Operational Applications of Satellite Snow-Cover Observations, NASA Technical Paper 1828, Goddard Space Flight Center, NASA, Greenbelt, Maryland.

Colorado Basin River Forecast Center, 1993. "1994 Guide to Water Supply Forecasting", Salt Lake City, Utah.

Curtis, D.C. and J.C. Schaake, Jr., 1979. "The NWS Extended Streamflow Prediction Technique", Presented at the Engineering Foundation Conference: Water Conservation - Needs and Implementation Strategies, Ringe, New Hampshire, 1979.

Day, G.N., 1985. "Extended Streamflow Forecasting Using NWSRFS", Journal of Water Resources Planning and Management, ASCE, 111(2), pp.157-170.

Garen, D.C., 1992 "Improved Techniques in Regression-Based Streamflow Volume Forecasting", Journal of Water Resources Planning and Management, ASCE, 118(6), pp. 654-670.

Koch, R.W., 1988. "Dependence of Seasonal Precipitation and Runoff Volumes in the Pacific Northwest on an Index of the El Nino Southern Oscillation", In: United States / People Republic of China Flood Forecasting Sympoisum/Workshop, Volume 1, Portland, Oregon, pp.129-145.

Schaake, J.C. Jr. and E.L. Peck, 1985. "Analysis of Seasonal Volume Streamflow Forecast Errors in the Western U.S.", Proceedings of the 53rd Western Snow Conference, pp. 44-53.

Twedt, T.M., J.C. Schaake, Jr., and E.L. Peck, 1977. "National Weather Service Extended Streamflow Prediction", Proceedings of the 45th Western Snow Conference, pp 1-8.