Development and Interpretation of Daily Seasonal Water Supply Forecasts

This Technical Note documents the process used to develop the Daily Water Supply Forecast procedures and serves as a guide for helping users interpret the results. This product was developed to supplement official water supply forecasts generated at the beginning of each month during the water supply season. The water resource community requested daily calibrated water supply forecasts to aid in the assessment of increased climate variability. The Natural Resources Conservation Service (NRCS) National Water and Climate Center (NWCC) developed these procedures that use hydroclimatic data collected by the SNOTEL and USGS data networks.

Three graphics are produced: 1) a cross-plot graphic of historical and observed volume forecasts (Fig.1); 2) daily exceedance forecasts vs. historical bounds (Fig. 2); and 3) guidance volume forecasts (percent of normal) vs. skill (Fig. 3). An Excel data file containing all the numbers shown in the graphs and organized by forecast point is also provided. Additionally, a summary file containing an overview of the status and trend for the basins currently calibrated is developed. The entire product suite is accessible from NWCC Web site:
http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html

NWCC archives all products generated by the daily water supply forecast procedure and compares them to the coordinated monthly water supply forecast jointly issued by the National Weather Service and NRCS.

Users are encouraged to direct questions and comments regarding the daily water supply forecast products to Tom Perkins, 503-414-3059 or tom.perkins@por.usda.gov. Snow survey and Water Supply Forecasting programmatic questions or comments may be directed to Mary Greene, 503-414-3058, or mary.greene@por.usda.gov.

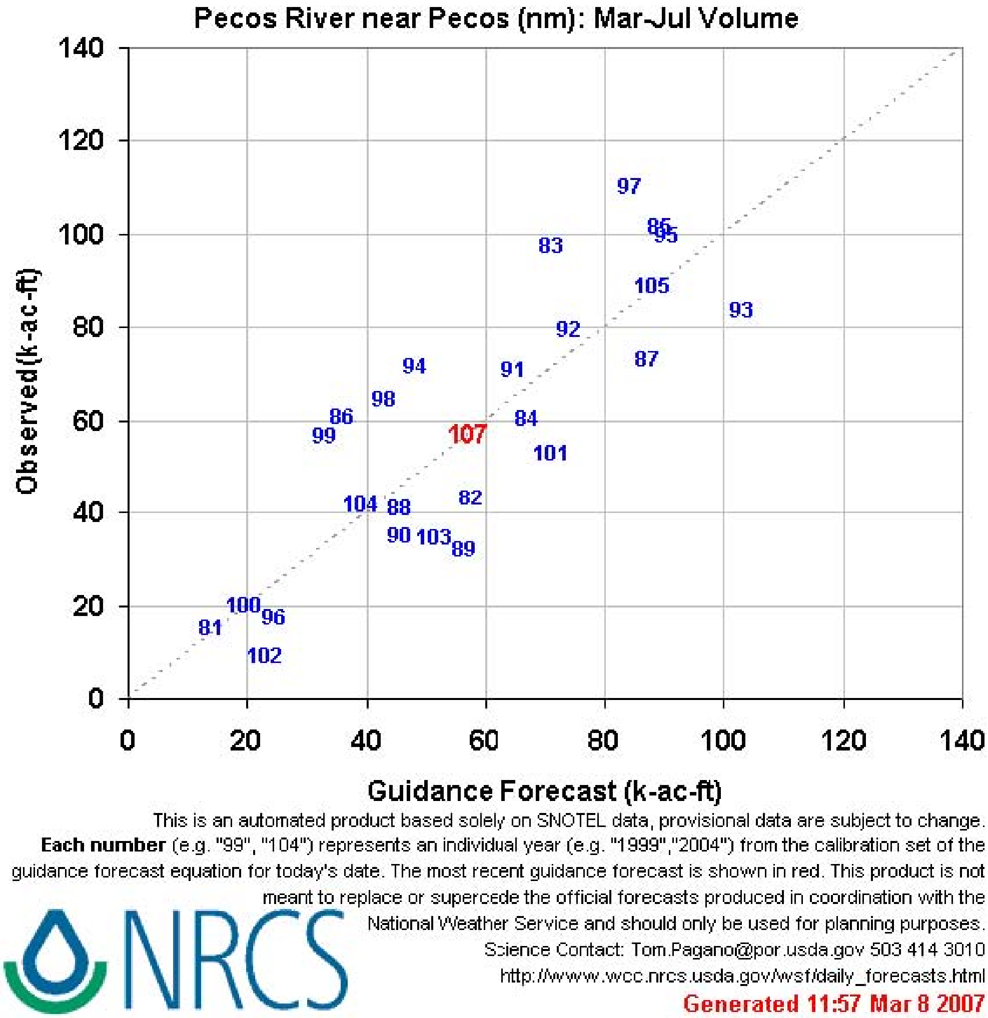

Figure 1. Scatter diagram of historical (calibration) and observed volume forecasts.

Caption reads, “This is an automated product based solely on SNOTEL data; provisional data are subject to change. Each number (e.g. “99”,”104”) represents an individual year (e.g. “1999”,”2004”) from the calibration set of the guidance forecast equation for today’s date. The most recent guidance forecast is shown in red. This product is not meant to replace or supersede the official outlooks produced in coordination with the National Weather Service and should only be used for planning purposes.”

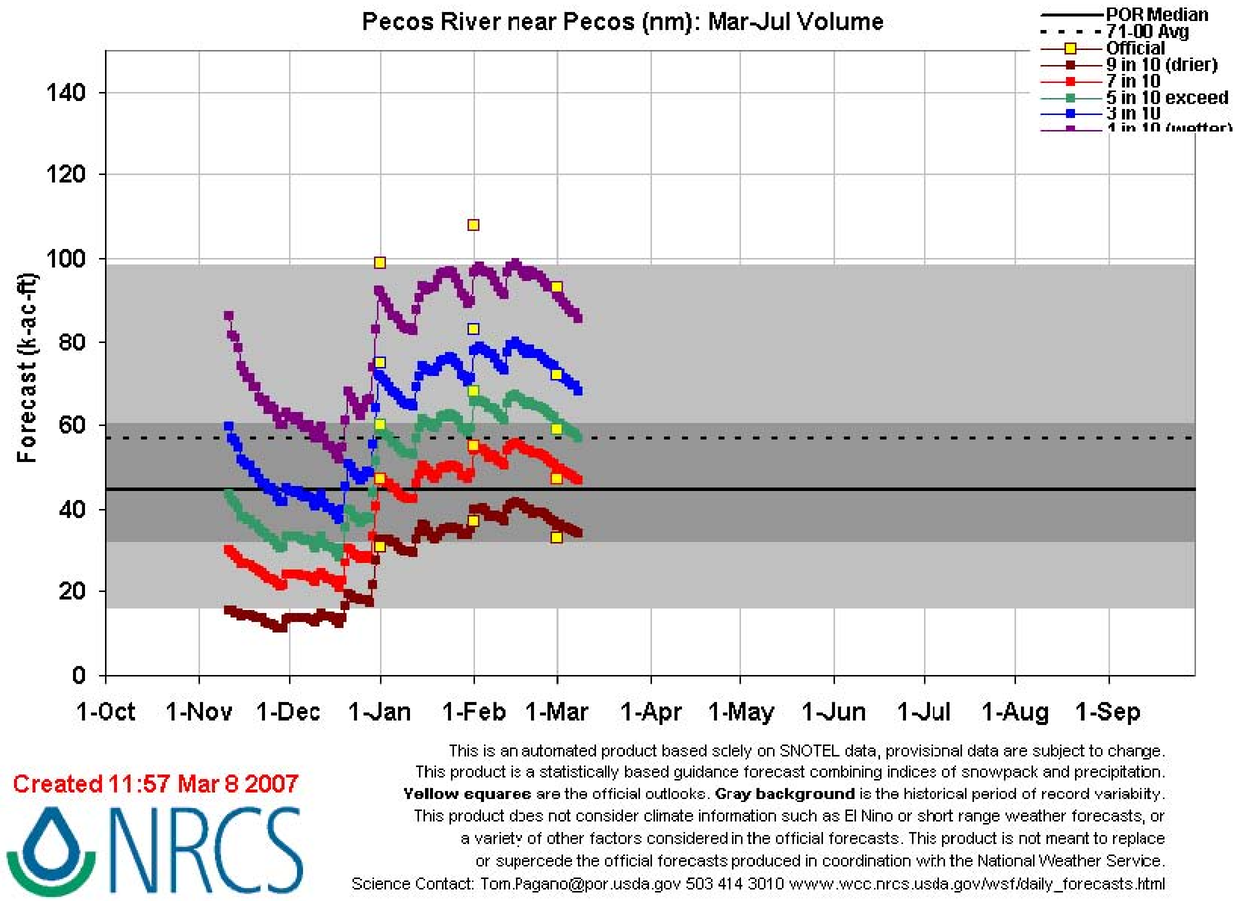

Figure 2. Daily exceedance guidance forecasts (colored lines) versus historical range of variability (gray background) and the official forecasts (yellow squares).

Caption reads, "This is an automated product based solely on SNOTEL data; provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. Yellow squares are the official outlooks. Gray background is the historical period of record variability. This product does not consider climate information such as El Nino or short-range weather forecasts or a variety of other factors considered in the official forecasts. This product is not meant to replace or supersede the official outlooks produced in coordination with the National Weather Service."

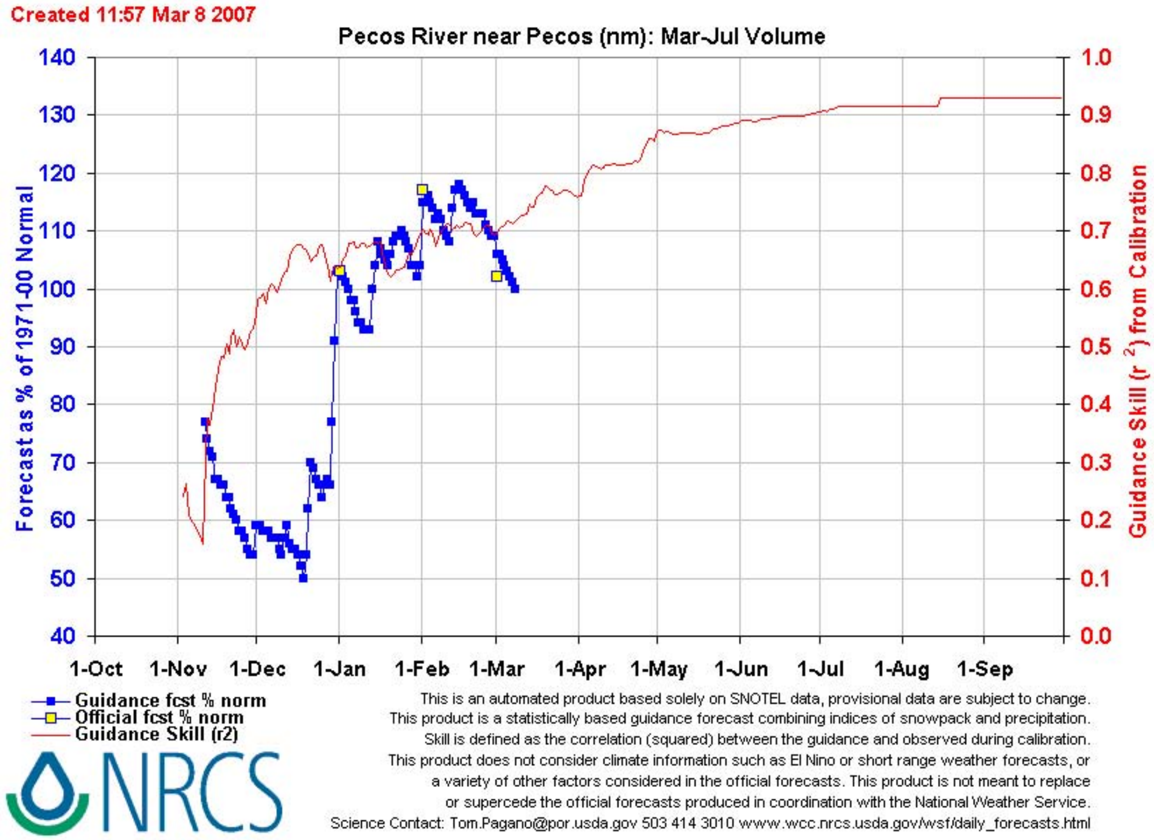

Figure 3. Guidance volume forecast as percent of normal (blue line) versus official forecasts (yellow squares) and guidance skill (r^2) from calibration.

Caption reads, "This is an automated product based solely on SNOTEL data; provisional data are subject to change. This product is a statistically based guidance forecast combining indices of snowpack and precipitation. Skill is defined as the correlation (squared) between the guidance and observed during calibration. This product does not consider climate information such as El Nino or short-range weather forecasts or a variety of other factors considered in the official forecasts. This product is not meant to replace or supersede the official outlooks produced in coordination with the National Weather Service."

Process Overview. The primary inputs to this procedure are historical and real-time daily time series data for SNOTEL snow water equivalent and water year to date precipitation, as well as historical monthly streamflow volume data. Streamflow data are available from the National Water and Climate Center NRCS database, although the program also offers the ability to acquire the data directly from the US Geologic Survey (USGS) Web site, provided that the target forecast point does not have any diversions or regulations that affect the flow, such as reservoirs. The primary product is a seasonal streamflow forecast (e.g. the volume of flow from April to July at a particular location) and ancillary statistics (e.g. measures of the range of historical variability, the expected skill of the forecast at a given lead-time). Forecasts of other hydrograph behaviors (e.g. peak flow) are possible, provided that there is a long time series of historical data. However, the output would only be of a seasonal planning nature, not as specific or skillful as the output of a short-range forecast from a dynamic watershed model.

Establishing a new forecast point is a relatively automated procedure. A hydrologist determines the SNOTEL sites to be used in the analysis and gathers metadata describing the forecast point such as the station name, long-term normal, target period, and so on. The user enters this information into an Excel spreadsheet and executes a macro that draws down the historical snowpack, precipitation and streamflow data. An automatic screen fills small gaps in the meteorological record (no more than 8 days per year by assuming persistence in snowpack or water year precipitation). If more than 8 days are missing in a particular station-year, the entire water year is excluded from analysis. This ensures that shifts in the real-time forecast throughout the season are solely due to meteorological trends and not because of varying length of calibration period of record on different days of the year (e.g. the April 1 forecast should not use data from 1981-2004 if the April 2 forecast uses 1984-2003 data, both should use the same record).

Calibration Process. After inspecting the quality of the data, another macro calibrates 365 forecast equations, one for each day of the year, relating the climate information to streamflow.   Contained within the macro is logic to determine the best predictors for each site, with one predictor based on snow (current snowpack, snowpack on a past date or the peak snowpack to date) and another based on water year precipitation (to date, or to some date in the past).

For example, on April 15, the predictors for station one could be today's snowpack and October 1 - April 14 precipitation, whereas for station two it is peak snowpack to date and October 1 – March 18 precipitation. This locking-in of a particular variable to some date earlier in the season is useful mostly during late spring when snowpack has largely disappeared and convective summer precipitation is not relevant to streamflow.

After calibration is finished, the user reviews the results and adds the Excel spreadsheet into a queue that is run on a scheduler, currently twice a day (7 am and noon Pacific Time). Part of the calibration involves generation of a full set of volumetric hindcasts (retrospective forecasts) for as many days and years as there is available historical SNOTEL data. Real-time operations include recalculating all the forecasts to date for the current water year.

Skill Definition. The program also calculates the expected skill of the forecasts, the r^2 (coefficient of determination) between the hindcasts on each day of the year and the observed. R^2 is the proportion of variability in a dataset that is accounted for by a statistical model. A skill value of 1 indicates perfect correlation between the forecast and observed, 0 indicates no relationship. These skill values are fixed for the calibration set and only vary by lead-time. They are not situation dependent and do not vary from year to year (unless the forecast set is recalibrated using more up to date data). Generally skill is expected to increase as forecast lead-time decreases, although sometimes in practice, skill may appear flat or decrease mid-season. This is likely due to the specifics of the calibration set of data, such as an outlier year that greatly changed its character throughout the season. For example, a year may have started dry, became wet mid-season, dried out and ended the season dry. In this case, the earliest forecasts appear more skillful than the mid-season ones. This effect is most pronounced in the first month or two of the water year when a single storm in an individual year may be a dominant outlier in the calibration set. Therefore, a filter was constructed that censored the data from predictors who had not yet achieved a period of record average on that date that was a particular percentage of the seasonal peak of the period of record average for the variable. This user-editable value is set at a default of 10%, i.e. today’s normal is less than 10% of the seasonal peak of the normal (not to be confused with the normal of the seasonal peak). The general effect of this filter is that forecasts do not begin until mid-November, preventing “whipsawing” forecasts early in the year.

Z-Score Regression. During calibration, various predictors are combined using a technique called Z-Score Regression. Briefly, each predictor is converted into a Z-Score (a standardized anomaly, the value minus the period of record mean divided by the period of record standard deviation). An index is created where each year is a weighted average of each station's Z-Score. The weighting is based on the coefficient of determination (r^2) between the predictor and the predictand), such that better sites are emphasized in the index and the influence of worse sites minimized. Not every station is required to have data available in every year; the weighted average is only based on reporting stations in that year.

Individual stations are grouped by data type and then grouped again across data types. For example, all snowpack variables lumped together into a single index, all precipitation variables into another, then the snowpack and precipitation indices are combined again in a weighted average, again using their respective coefficient of determination (r^2) with streamflow. Finally, a single index summarizing precipitation and snowpack information from all sites is then regressed against the predictand in standard fashion.

Z-Score Evaluation. Pagano (2005) tested this technique on 29 indicator basins across all climates of the western U.S. and generally found comparable skill with the historical official forecasts. The technique has since found widespread use within the NRCS for operational water supply forecasting. The NRCS also uses a principal components based regression technique described by Garen (1992), however many aspects of Z-Score regression make this technique better suited for daily forecasting. For example, given that Z-Score Regression is based on a weighted average of available sites, some sites may be missing in individual years (or in real-time) and the remaining sites will compensate for this missing data and a regression result is still evaluated. In comparison, most other regression techniques require that all data be serially complete and available, a process that would be very difficult to automate in real-time.

Preliminary in-house comparisons of the techniques of Garen (1992) and Pagano (2005) indicate that under most forecasting situations; the skill is comparable, although the convenience of Z-Score Regression in operational forecasting is higher. One of the assumptions of Z-Score Regression is that various predictors combined into a single group represent a single signal. It does not perform well when there are inhomogeneities in the input data versus time (e.g. one poor station with a long period of record mixed in with several good stations with short records, or all stations except the poor station missing at forecast time). The onus is on the forecaster to ensure this is not occurring by selecting the proper mix of stations during calibration. The risk is a calibration with unnecessarily high uncertainty or an overconfident real-time forecast. To help alleviate this risk, if half or more of the stations are missing data for one element (snowpack and/or precipitation), the real-time forecast is not evaluated due to insufficient data.

Product History. This product began in October of 2005 in the headwaters of the Rio Grande and it was put onto an automated operational schedule in December 2005 with 9 forecast points. Since then the product has recently been made public through the NWCC homepage at http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html.

The real-time operations of the forecasts were evaluated by a select subset of volunteer users and NRCS snow survey personnel. The review was highly favorable. There was initial concern that users might confuse this product with the official forecasts, although the concern turned out to be a non-issue. Indeed, several users indicated that the greatest strength of this tool is the relative trend that it indicates, that it could be used as a forecast of the forecast, a mid-month indicator of how the first of month official forecasts may change.

By the end of 2006, 47 basins were being calibrated and run, and a skill evaluation was performed. At lead-times based on April 1, the probabilistic reliability of the forecasts (e.g. when the product says the flow will be below a certain level X% of the time, and this actually occurs X% of the time) was excellent, although there was a slight tendency for overconfidence (i.e. uncertainty bounds too narrow) at the very end of the season. The westwide average forecast error at the end of the year was identical to the official forecasts, although this product was 4% worse than the official forecasts on April 1. This is not entirely unexpected given that it is a completely automated process and uses only a subset of available information to create the forecasts. For example, this product does not include information about antecedent soil moisture, springtime temperatures, or climate forecasts such as those based on El Nino. This product also lacks the human expertise associated with the official forecasts. See section "Relationship to Official Forecasts" below for more details.

Geographic Distribution of Skill. The daily forecasts also display many of the skill characteristics of the official forecasts themselves (Pagano et al. 2004). Their skill versus lead-time is primarily related to the importance of spring precipitation in the annual cycle, so that locations where most of the moisture comes during winter (e.g. the Sierras, Idaho) perform better earlier in the season than those places where the bulk of moisture normally comes during spring and summer (e.g. the Missouri, Front Range). It also performs well in regions that are purely snowmelt dominated (e.g. the highlands of Colorado) and poorly in regions with midwinter melt, high baseflows during winter and complex subsurface geologic processes (e.g. the middle Cascades).

The most skillful forecast set calibrated to date has been the South Fork of the Rio Grande in southern Colorado. A high, purely snowmelt driven basin which experiences dry summers, the end of year skill (i.e. the equation calibrated September 30th, looking backwards to predict an April-September flow) being $r^2 = 0.974$, an average percent of normal error of 5.6%. Of the 130 points calibrated so far, 50% have end of year skill of $r^2 > 0.878$ and 84% have end of year skill of $r^2 > 0.80$. With r^2 of 0.55-0.60, the poorest performers are in the Front Range basins such as the South Platte of Colorado and the Canadian Basin of New Mexico as well as the Musselshell in Montana. None of the temperature dominated basins of the Pacific Northwest (e.g. the Willamette) have been calibrated yet, although they are expected to be of low skill ($r^2 \sim 0.6$), similar to the official forecasts.

Recent Model Enhancements. At the start of 2007, the daily forecast calibration software has been significantly streamlined and improved. While the program is still Excel-based and obtains all its data from the Internet, the data gathering and computation time for calibration was reduced from up to 39 minutes per basin to around 47 seconds. Most of the time is now spent in the human reviewing the data and establishing the sites to be used. Real-time operation takes about 9 seconds per forecast point on a modern desktop computer. This has allowed acceleration in development in forecast points to where the roster is now 130 as shown in Fig.
4.

Previously limited by 8 input SNOTEL sites per basin, this software improvement also allowed the development of forecasts for large basins which may have 100 or more SNOTEL sites available. The procedure is the same as that for a small basin, sometimes referred to as a direct routing. Two large basins have been calibrated and are being run in real-time, the Colorado River inflow to Lake Powell, Arizona (108,335 mi^2 drainage area, 59 SNOTEL sites) and the Salmon River near White Bird, Idaho (13,550 mi^2 drainage area, 19 SNOTEL sites). Computer time of the calibration of the Colorado River basin once the stations were selected was 2 minutes 21 seconds on a modern desktop computer (1 minute and 17 seconds of which was associated with gathering data from the Internet).

Relationship to Official Forecasts. The official water supply outlooks are produced monthly and issued jointly with the National Weather Service. These are available in the State Basin Outlook reports http://www.wcc.nrcs.usda.gov/cgibin/bor.pl and the Westwide Water Supply Outlook http://www.wcc.nrcs.usda.gov/wsf/westwide.html.

The official forecasts are a subjective combination of the output of a variety of objective tools. The output of a set of statistical regression equations based on monthly data is one of many pieces of information brought to the coordination process. These regression equations are expected to capture as many of the runoff-affecting basin processes as possible (e.g. snowpack, autumn rains, climate signals, soil moisture, and temperature effects). In addition to automated data, they also include data that is available in real-time but requires manual collection (e.g. NWS cooperative observer precipitation data, manually measured snow courses). These forecasts also imply an additional level of quality control on the input data. Forecasts may differ significantly from the objective guidance if compelling anecdotal evidence is available from the field if it pertains to a basin process that is known to be important but is difficult to quantify (e.g. sublimation wasting of snow water, quality of low elevation snowpack).

In comparison, this product is the completely automatic and objective output of a single tool. It considers solely daily-resolution SNOTEL snow and precipitation information on water year time scales. Each day, the product re-evaluates all the forecast equations to date for the water year in order to incorporate NRCS Data Collection Office corrections and edits to SNOTEL data. However, the most recent data is only limited by the automated filters built into the SNOTEL system. Users should beware of large jumps in the forecast if a significant meteorological event has not recently occurred.

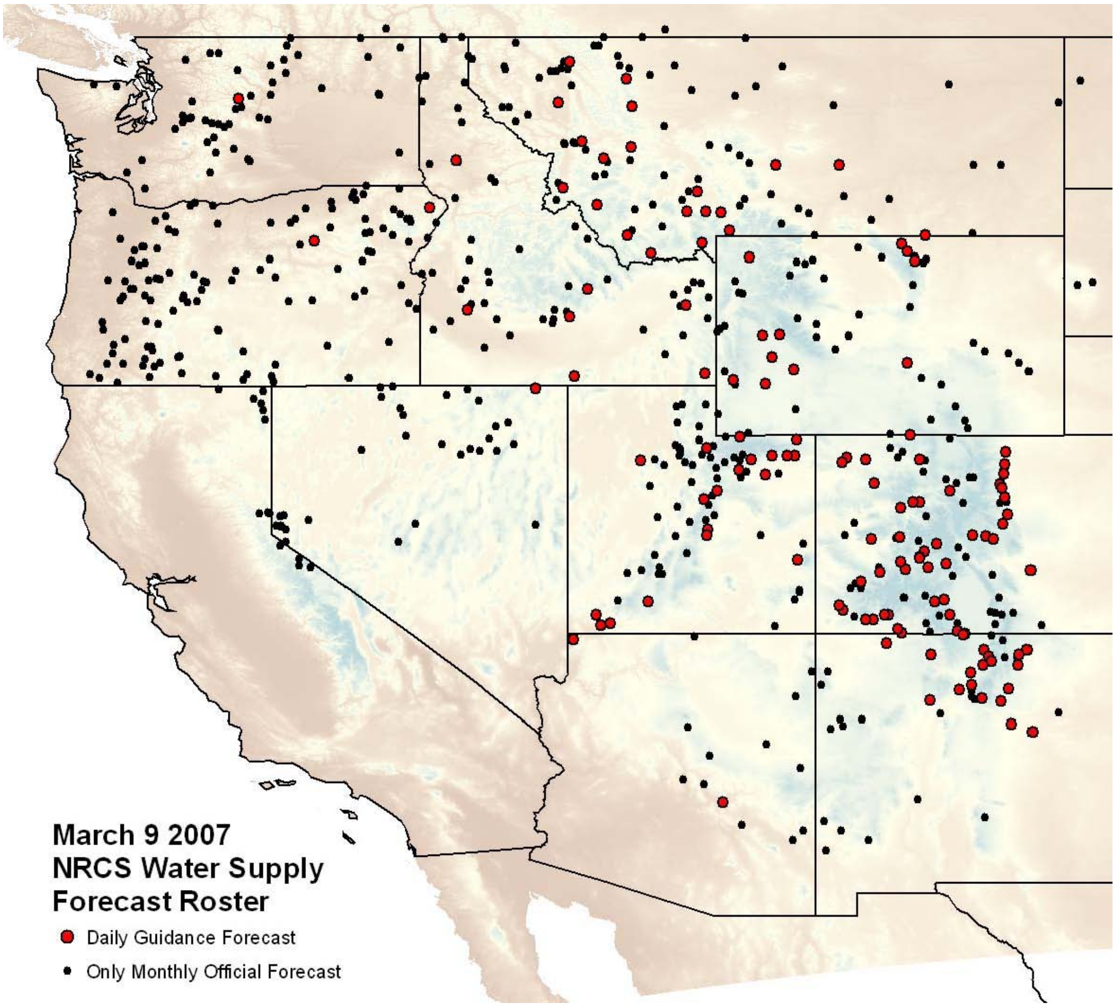

Figure 4. Locations of calibrated daily and monthly water supply forecast basins

If one desires the magnitude of the official forecasts and this product to be similar, it is generally recommended to use the same set of stations and period of record in this product as the monthly regression equations used in the official outlooks. However, it is difficult to build in the effect of non-SNOTEL influences on the forecasts, such as climate signals or baseflow conditions. While more complete in the suite of data considered, the monthly equations are only one of many tools brought to the official forecast generation process; with subjective adjustment and coordination, it may be impossible to develop an automated objective tool that would always exactly match the official outlooks.

Summary. As previously discussed, perhaps the greatest strength of this tool is its ability to track

intraseasonal variations and to fill the gaps between the official outlooks. While factors such as climate signals or baseflow conditions may affect the official forecasts, this effect is generally considered to be constant throughout the season. The primary factor that changes month to month is the accumulation of snowpack and precipitation, something that is captured very well by the SNOTEL network and therefore by this tool.
Nonetheless, it is important to communicate these caveats to users. Therefore, products carry the disclaimer that “this is an automated product based solely on SNOTEL data; provisional data are subject to change.”

Furthermore, “this product does not consider climate information such as El Nino or short range weather forecasts, or a variety of other factors considered in the official outlooks.”

Lastly, “this product is not meant to replace or supersede the official forecasts produced in coordination with the National Weather Service.”

References.

Garen, D. C., 1992. Improved techniques in regression-based streamflow volume forecasting. Journal of Water Resources Planning and Management, American Society of Civil Engineers, 118(6), 654-670.

Pagano, T. C., 2005. The role of climate variability in operational water supply forecasting for the Western United States. Ph.D. dissertation, Department of Hydrology and Water Resources, University of Arizona, 283 pp.

Pagano, T. C., D. C. Garen, and S. Sorooshian, 2004. Evaluation of official Western US seasonal water supply outlooks, 1922-2002. J. Hydrometeorology. 5(5), 683–698.