

## DAILY UPDATING OF OPERATIONAL STATISTICAL SEASONAL WATER SUPPLY FORECASTS FOR THE WESTERN U.S.<sup>1</sup>

Thomas C. Pagano, David C. Garen, Tom R. Perkins, and Phillip A. Pasteris<sup>2</sup>

**ABSTRACT:** Official seasonal water supply outlooks for the western United States are typically produced once per month from January through June. The Natural Resources Conservation Service has developed a new outlook product that allows the automated production and delivery of this type of forecast year-round and with a daily update frequency. Daily snow water equivalent and water year-to-date precipitation data from multiple SNOTEL stations are combined using a statistical forecasting technique ("Z-Score Regression") to predict seasonal streamflow volume. The skill of these forecasts *vs.* lead-time is comparable to the official published outlooks. The new product matches the intra-monthly trends in the official forecasts until the target period is partly in the past, when the official forecasts begin to use information about observed streamflows to date. Geographically, the patterns of skill also match the official outlooks, with highest skill in Idaho and southern Colorado and lowest skill in the Colorado Front Range, eastern New Mexico, and eastern Montana. The direct and frequent delivery of objective guidance to users is a significant new development in the operational hydrologic seasonal forecasting community.

(KEY TERMS: runoff; snow hydrology; streamflow; surface water hydrology.)

Pagano, Thomas C., David C. Garen, Tom R. Perkins, and Phillip A. Pasteris, 2009. Daily Updating of Operational Statistical Seasonal Water Supply Forecasts for the Western U.S. *Journal of the American Water Resources Association* (JAWRA) 45(3):767-778. DOI: 10.1111/j.1752-1688.2009.00321.x

### INTRODUCTION

For close to 70 years, the Natural Resources Conservation Service (NRCS) of the United States Department of Agriculture has published seasonal water supply outlooks for use in natural resource management. The NRCS National Water and Climate Center (NWCC) produces these outlooks once per month from January through June, in partnership with the National Weather Service (NWS) and local

cooperating agencies. The geographic and climatic scope of the forecasts ranges from minor creeks of the semi arid southwestern United States (U.S.) to glaciated basins of Alaska. These forecasts are utilized by a broad spectrum of users for a variety of purposes, including irrigated agriculture, flood control, municipal water supply, endangered species protection, power generation, and recreation.

Near the start of each forecast month, four NWCC forecasters have approximately three to five working days to create, analyze, adjust, coordinate, and issue

<sup>1</sup>Paper No. JAWRA-08-0044-P of the *Journal of the American Water Resources Association* (JAWRA). Received March 10, 2008; accepted December 1, 2008. © 2009 American Water Resources Association. No claim to original U.S. government works. **Discussions are open until six months from print publication.**

<sup>2</sup>Respectively (Pagano), Commonwealth Scientific and Industrial Research Organization (CSIRO), Land and Water Group, Highett Labs, P.O. Box 56, Highett, Victoria 3141, Australia; (Garen, Perkins), United States Department of Agriculture, Natural Resources Conservation Service, National Water and Climate Center, 1201 NE Lloyd Boulevard, Suite 802, Portland, Oregon 97232; (Pasteris) Global Water Resources, CH2M HILL, 2020 Southwest 4th Avenue, Suite 700, Portland, Oregon 97201-4958 (E-Mail/Pagano: Thomas.Pagano@csiro.au).

forecasts for over 180 points apiece simultaneously, for a total of 732 forecast points throughout the region. Maintenance of the forecast environment and production of the outlooks are both human-resource intensive tasks, and this has been a limit on the more frequent updating of the forecasts. Statistical forecasting techniques have also been limited to using the data available on the first of the month [because historically, snow water equivalent (SWE) data were collected at manually measured snowcourses once per month]. Nonetheless, mid-month updates to the forecasts are easily the most frequent request from users, especially if a major weather event has significantly changed the character of hydrologic conditions on the watershed. A recent observed increase in hydrologic variability (Pagano *et al.*, 2004b; Pagano and Garen, 2005) has made dynamic water management all the more critical.

The use of hydrologic simulation models has been proposed as a method to satisfy the demand for more frequent forecasts. These models rely on daily or sub-daily forcing data and therefore are not tied to the monthly schedule and can be run anytime. These models are used extensively by the NWS (Franz *et al.*, 2003). The calibration of these models, however, is extremely time intensive and requires hydrologic modeling expertise. Real-time operation, in practice, also involves manual inspection and adjustment of model states, although some objective “hands-off” hydrologic forecasting systems exist (e.g., Wood and Lettenmaier, 2006). The NRCS has made initial attempts to implement a simulation modeling forecasting system, the most recent effort being the use of the Precipitation Runoff Modeling System within the Object Modeling System (Pagano *et al.*, 2005), but the large resource requirements have been a hindrance for the small staff.

While resource “expensive” to run and maintain, simulation models do not produce significantly more accurate seasonal streamflow volume forecasts than the existing statistical forecasting system (Franz *et al.*, 2003; Pagano *et al.*, 2004a). Furthermore, simulation models have a well-known tendency to produce overconfident forecasts with narrow forecast distributions in part because they ignore model calibration and data errors (Barnston *et al.*, 2003; Franz *et al.*, 2005; Wood and Schaake, 2007). Some simulation models also impose extraordinary requirements on the computers that run them (e.g., specific operating systems, powerful processors, and extensive data storage).

This paper introduces a product that represents a low-cost intermediate solution. These “daily update water supply forecasts” are forced by daily data and therefore can be updated daily and run year-round. Based on a statistical regression technique, the

calibration and operation of the model is almost entirely objective and automated. The calculations are relatively simple, and basic computing resources can be used to complete processes within a demanding operational timeline. A series of western U.S. basins are currently being run operationally; the forecasts are available at [http://www.wcc.nrcs.usda.gov/wsf/daily\\_forecasts.html](http://www.wcc.nrcs.usda.gov/wsf/daily_forecasts.html).

## MODEL DATA

The forecasting technique used to create the product described herein involves statistical regression relating a collection of predictors (SWE, accumulated precipitation) to a predictand (streamflow). This section describes the collection of and basic processing of the data from various sources.

### *Predictand*

The NRCS primarily forecasts seasonal streamflow volume (e.g., April-July total flow) at specific gaging locations or above reservoirs. Many forecast points are regulated, in that the observed streamflow is significantly altered by human activity upstream such as irrigation diversions and reservoir releases. Naturalization of streamflow values to remove human influences is a difficult task, and even the best efforts cannot completely remove human effects. In reality, there are differences between true natural flow and unregulated flow data (which account for a limited number of measured reservoirs and losses). The NRCS maintains a database of unregulated flows (at a monthly time step) that the daily forecast program (described below) can directly access over the Internet. For stations without major human influences, the hydrologist also has the option of using the program to acquire observed streamflow data from the U.S. Geological Survey’s National Water Information System (USGS NWIS) webpage. Acquisition and processing is automated – the hydrologist specifies the station number and instructs the program to obtain the data. If an annual time series of daily values exists, the program can also forecast some aspect of hydrograph behavior besides seasonal volume (e.g., peak flow amount or date). However, that data must be obtained external to the program and entered manually.

If forecasting flow volume, the forecaster then chooses the target season of the predictand (e.g., total flow for April-July). A transformation (i.e., square root, cube root, or natural logarithm) to the

predictand can also be applied. This is useful in the case of a nonlinear relation between predictor and predictand or in the case of heteroscedasticity in the error distribution. The real-time forecasts are ultimately untransformed and expressed as real-world units (e.g.,  $\text{m}^3$ , acre-feet).

### Predictors

The NRCS SNOTEL (SNOW TELemetry) network (<http://www.wcc.nrcs.usda.gov/snow>) consists of over 700 remote data sites throughout the western U.S. Standard sensors include a snow pillow to measure SWE on the ground, a storage precipitation gage that measures water year-to-date (beginning October 1) accumulated precipitation, and an air temperature sensor (some enhanced sites have additional sensors as well). Only the SWE and precipitation data are used as predictors for the daily forecast product.

Most SNOTEL sites are located at moderately high elevations, averaging around 1,400 meters in Washington, 2,000 meters in Idaho, and 3,000 meters in Colorado. Although they vary by region, exposure, and elevation, most NRCS snow measurement sites begin accumulating persistent snow cover by November 1 and reach peak accumulation by mid-April. Approximately half of the sites are typically snow free by the last week of May, although the melt rate is rarely constant, and brief episodes of accumulation during the melt season are common (Serreze *et al.*, 1999). The NRCS began the expansion of the SNOTEL network in earnest in the early 1980s, hence most stations now have approximately 25 years of data. The oldest sites were installed in the mid 1960s, primarily in Montana.

The hydrologist specifies the list of station identifiers to be used as predictors, and the program acquires the data via the Internet and puts it into a conveniently usable format. The program also calculates the maximum SWE to date for each water year for use as a possible predictor. The program automatically fills small gaps in the meteorological record (no more than eight days per year) by assuming persistence in SWE or water year-to-date precipitation. If more than eight not necessarily continuous 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 due to variations in the period-of-record for calibration on different days of the year. For example, 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 for consistency.

## MODEL CALIBRATION

### Predictor Selection and Censoring

The hydrologist begins predictor selection by acquiring a list of candidate SNOTEL stations in the region. This is a somewhat subjective process, attempting to balance complete geographic coverage with adequate station record length and skill in predicting streamflow. A common strategy is to reuse the same stations that are predictors in the once-a-month forecasting equations. The program also provides feedback about the quality of predictors throughout the calibration process, so the hydrologist can easily remove poor predictors and recalibrate if necessary. Given a list of SNOTEL stations, the program next automatically analyzes the available data to determine the best predictors for each station for a given day of the year based on the coefficient of determination ( $r^2$ ) between the predictor and predictand (Figure 1). Two predictors are used for each SNOTEL site, one representing SWE, the other precipitation (real-time SNOTEL air temperature was not considered due to a lack of quality control). For SWE, possible predictors include the station's SWE on the current date, SWE on some date prior to the current date, or the peak SWE to date. For precipitation, the accumulated water year-to-date or water year to some date earlier in the season can be used.

The predictor selection process is summarized in Table 1. Generally, the program assumes that as time passes, more information is learned about the character of the season, leading to less uncertainty and more skillful forecasts. However, late spring or summer precipitation may be less important than precipitation recorded earlier in the season given the higher evaporation rates and spatial variability of convective warm season rainfall. Therefore, the optimal precipitation predictor may end some day before the forecast date. Likewise, SWE during summer is usually zero, hence a poor predictor of flow, in which case, SWE on a past date (e.g., April 4 or the peak for the season) may be a more accurate predictor. To prevent overfitting and rapid switching of variables throughout the season, predictors at a later date are always preferred except in certain cases late in the season. The algorithm finds the day of the year that a predictor has its highest  $r^2$  with the predictand (e.g., April 17,  $r^2 = 0.88$ ). Prior to this date, the most up-to-date data are always used as predictors. However, for subsequent dates, the most up-to-date data are used for prediction until the  $r^2$  falls below a user-specified tolerance (e.g., 0.03) of the maximum  $r^2$  for the season. If the  $r^2$  falls below the tolerance (e.g., May 20,  $r^2 = 0.84$ ), then the data from the last good date (e.g.,

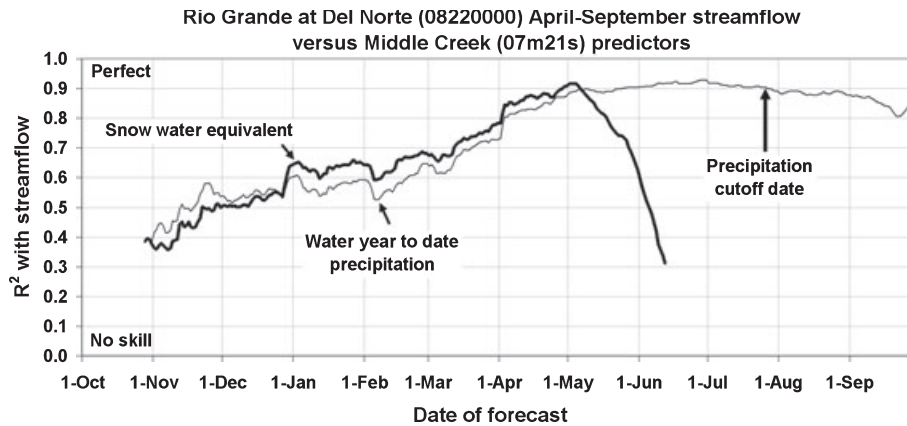


FIGURE 1. Relationship of SWE and Water Year Precipitation With Seasonal Streamflow *vs.* Time for the Rio Grande Near Del Norte. Skill from SWE information increases until early May after which snow begins to melt. Skill from accumulated precipitation decreases in the summer, and the program does not consider new precipitation information after late July (shown by arrow). These correlations are used to determine the relative weight of each predictor in the forecast. Some correlations are not evaluated (SWE before October 28 and after June 12 and precipitation before October 31) due to the censor limit described in the text.

May 19) are chosen as the predictors instead. The assumption of this heuristic is that early season non-monotonic behavior of skill *vs.* forecast lead-time is due to limited sampling of the historical record and that true skill always increases *vs.* time. However, at the end of the season, the loss in skill *vs.* time is due to a change in the meaning of the data (e.g., mid-April SWE is relevant to total basin moisture while late-June is relevant to springtime temperatures and the timing of moisture, the latter being an inferior predictor of total seasonal runoff volume). The default threshold value (0.03) could be changed if the forecaster judges differently about the risk of overfitting *vs.* potential unrealized skill.

Early in the water year, snowpack may be ephemeral, and its correlation with streamflow may be spurious or exaggerated. Likewise, anomalously wet basin conditions early in the season can lead to forecasts that rapidly vary (“whipsawing”) and are well outside the range of historical variability (extrapolation). During the course of the calibration, the program calculates the daily period-of-record average of SWE and water year-to-date precipitation separately. It then finds the day of the year for each element that this daily average is at its maximum (i.e., the peak of the average, as opposed to the average of the peak, the interannual average of each year’s maximum SWE). If the long-term average on a given day for a particular variable (both during accumulation and melt) is less than or equal to a “censor limit” times the average peak, then this predictor is excluded from the analysis. This censor limit can be set by the forecaster, and practice has suggested that 10% is adequate for most situations. For example, suppose the peak of the SWE average for a particular station is 50 cm occurring April 17. On October 15, the period-

of-record average SWE is <5 cm. Therefore, on October 15 of every year, the SWE data for this station will be excluded during the search for candidate predictors. Thresholds are applied to both SWE and water year-to-date precipitation and can come into effect on different days for each element. In the case where all predictors are censored, a forecast is not produced. If a significant early-season moisture anomaly occurs during this period, water managers are still able to acquire raw data from the SNOTEL network and make qualitative assessments on their own.

#### *Predictor Preparation and Filtering Using Z-Score Regression*

Water supply forecasters are faced with several significant statistical challenges, most notably how to address predictor collinearity and missing data. Predictor variables used in forecasting are usually highly correlated with each other. Standard multiple regression has difficulty estimating the significance of and coefficients for each predictor in this situation. Garen (1992) proposed using principal components analysis (PCA) filtering of predictors before use in regression, and this technique has become the standard for operational seasonal streamflow forecasting in the NRCS and elsewhere.

A collection of predictors is also likely to have a variety of periods of record, with relatively young stations intermixed with older stations. In standard regression, the requirement of serial completeness forces the forecaster to discard the younger stations, to discard the non-serially complete years of data, or to estimate every missing data value until completeness is achieved. Operationally, missing



TABLE 1. Pseudo-Code of Predictor Development for a Given Station.

---

```

DEFINE Elements:
  P as water year (beginning October 1) to date accumulated
  precipitation
  S as current snow water equivalent
  M as peak water year snow water equivalent to date

Determine Elements to be used as predictors on each Day:

FOR each Element
  FOR each Day of water year (October 1 - September 30)
    IF current Day's 1971-2000 normal is very small (< 10%)
      compared to peak of average THEN
        Predictor for this Day for this Element will not be developed
      ELSE
        Predictor for this Day for this Element will be developed
        Find day of year that Element has maximum correlation with predictand
      END IF
    NEXT Day
  NEXT Element

Develop predictors:

FOR Elements P and S
  FOR each Day of water year
    IF Element is to be developed THEN
      Calculate skill ( $r^2$ ) of Element for Day
      IF Day is after day of maximum correlation with predictand
        AND IF skill of Element for Day is worse than best skill
          (e.g.,  $\max r^2 - r^2 > 0.3$ ) THEN
            Substitute Element for Day with Element for most recent
            day after day with best skill that still lies above skill
            difference threshold
          ELSE
            Retain Element for Day as predictor
          END IF
        END IF
      NEXT Day
    NEXT Element

Choose between Elements S and M as snow predictor:

FOR each Day of water year
  IF both S and M are developed for Day
    Choose the one with highest  $r^2$  with predictand
  ELSE IF only one of S and M is developed for Day
    Use the one that is available
  ELSE
    Do not use snow predictor for Day
  END IF
NEXT Day

Perform Z-score regression for each Day using predictors available for Day

```

---

real-time data values are also a notable challenge. For example, during March 2003-March 2008, approximately 7.5% and 5.8% of SNOTEL snow and precipitation data, respectively, remained invalid or missing within 24 hours of initial data collection (the higher frequency for snow data is likely due to invalid negative values associated with sensor flutter during snow-free conditions). Without a serially complete real-time dataset, no forecast can be produced even if most of the information is available. To address all these concerns, the program uses a heuristic statistical technique ("Z-Score regression"; Pagano, 2004) to

combine predictors (Figure 2). This technique is a modification of unit-weighted regression (Schmidt, 1971) and takes advantage of predictor collinearity to relax the requirement for serial completeness. The strengths and weaknesses of this technique are described in more detail and numerical examples are provided in NRCS (2007). Generally, in situations of serially complete data, the regression coefficients of PCA and Z-Score are different, but their resulting forecasts are very similar. However, Z-Score regression was chosen for this application because it is adroit at automatically handling a non-serially com-

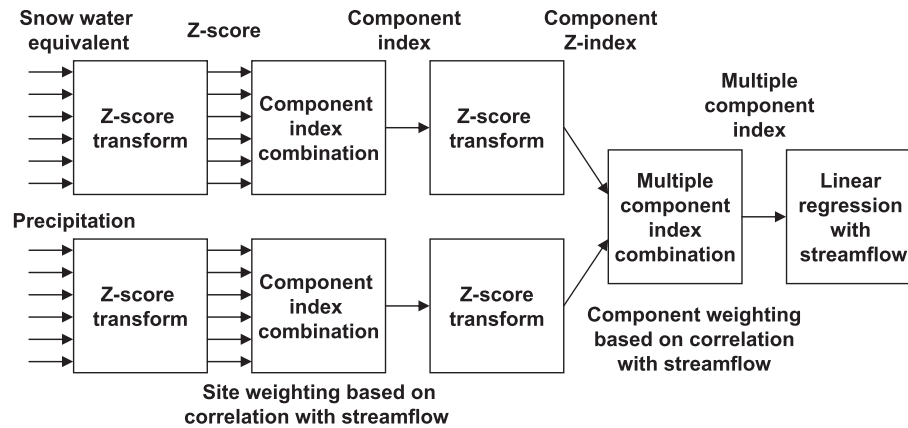


FIGURE 2. Flow Diagram of the Z-Score Regression Process.

plete set of predictors, a common operational real-time forecasting challenge.

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 all of the available stations' Z-Scores. The weighting is based on the  $r^2$  between the predictor and the predictand, such that better sites are emphasized in the index, and the influence of worse sites is reduced. Not every station is required to have data available in every year; the weighted average is only based on reporting stations in that year. Although the Z-Score method can use predictors with a negative correlation with the predictand (by switching the algebraic sign of the predictor time series values), these predictors were not considered for this application given the lack of physical meaningfulness of the result for the types of data considered.

Individual stations are first grouped by data type and then grouped again across data types. For example, all SWE variables are lumped together into a single index and all precipitation variables into another, then the SWE and precipitation indices are themselves combined into a weighted average, again using their respective  $r^2$  with streamflow. Finally, this single composite index summarizing precipitation and SWE information from all sites is regressed against the predictand in standard fashion. A numerical example of the calculations necessary to generate a real-time forecast is included in Table 2.

The end results of model calibration are, for each of 365 days per year:

- (1) Period-of-record mean and standard deviation of each predictor for conversion of station data into Z-Scores.
- (2) The selection, for each station and data type, of which variable is being used as a predictor. For SWE data, the choices are the current SWE

value, a value earlier in the season, or peak-to-date; for water year precipitation data, the choices are accumulation to present or accumulation to a date earlier in the season.

- (3) Weighting factors ( $r^2$ ) for each predictor and for station groups as a whole for use in the combination of predictors during Z-Score regression.
- (4) Slope and intercept of the regression of the predictand with the composite predictor index.
- (5) Regression calibration skill statistics ( $r^2$ , standard error).

It is standard water supply forecasting procedure to use a regression model that is based on the best fit with the data, with a confidence interval whose width is based on the standard error of a jackknife ("leave-one-out") cross-validation. Jackknifing is currently not used to develop the product described herein due to its computational expense (generally at least 20 times the normal expense, possibly much more depending on the strictness of the cross-validation). The *a posteriori* nature of variable selection (as discussed in Predictor Selection and Censoring) and station weighting (as discussed in Predictor Preparation and Filtering Using Z-Score Regression) increases the chance of overfitting, resulting in an inflated estimate of forecast skill and a forecast distribution that is too narrow (i.e., overconfident). In practice, with sufficient years of data in the calibration dataset, the difference between a best-fit standard error and a jackknife standard error is commonly relatively small.

## REAL-TIME FORECASTING AND PRODUCTS

At the NWCC, the forecast models are run twice daily (early morning and mid day, to capture any data quality edits performed by NRCS personnel) for

TABLE 2. Example Calculation of a Real-Time Forecast Issued April 1, 2007 for April-September Runoff Volume (in thousands of acre-feet) for the Rio Grande Near Del Norte (USGS ID: 08220000).

ID	Name	$r^2$	Average	SD	Current Observation	Current Z	$Z \times r^2$
April 1, 2007 snow water equivalent							
07m30s	Slumgullion	0.565	13.28	2.91	12.9	-0.132	-0.075
06m03s	Upper San Juan	0.526	32.38	12.73	18.8	-1.066	-0.561
07m21s	Middle Creek	0.783	18.66	5.82	16.3	-0.405	-0.317
06m23s	Lily Pond	0.415	14.89	6.16	9.8	-0.826	-0.343
07m32s	Beartown	0.566	23.56	7.60	14.7	-1.165	-0.660
	Sum of reporting	2.855					-1.955
Water year to date (October 1-March 31) precipitation							
07m30s	Slumgullion	0.538	12.89	3.14	13.4	0.163	0.088
06m03s	Upper San Juan	0.570	34.57	9.98	33.4	-0.117	-0.067
07m21s	Middle Creek	0.730	22.49	5.72	24.1	0.282	0.205
06m23s	Lily Pond	0.635	19.19	5.13	20.3	0.216	0.137
07m32s	Beartown	0.623	23.49	5.22	25.3	0.346	0.216
	Sum of reporting	3.096					0.579
Composite index		$r^2$	Average	SD	X	Z	$Z \times r^2$
Snow		0.723	0.102	0.980	-0.685	-0.803	-0.580
Precipitation		0.722	0.083	1.012	0.187	0.103	0.074
Sum of reporting		1.445					-0.506
	SE	$r^2$	Slope	Intercept	X	Y	$Y^2$
Regression	2.660	0.738	4.379	22.640	-0.350	21.106	445.457

Notes: In this example, the current data were used for the predictors, although this is not necessarily the case for other lead-times or basins (see Table 1 for more information). The “X” entries for the composite index and final predictor are the sum of reporting  $Z \times r^2$  divided by the sum of reporting  $r^2$ . If a particular station’s data are missing, then it is not included as part of these weighted averages. This nonlinear equation utilized a square root transformation to the predictand, and therefore the regression result (Y) is squared to obtain a forecast in real-world units.

all basins (148 forecasts in 25 min on a standard personal computer of typical office-use specifications). All of the water year-to-date precipitation and SWE data are acquired via the Internet and processed. The real-time predictor data are converted into Z-Scores and combined according to the Z-Score regression technique. If half or more of the predictor data are missing for a basin, the forecast is not evaluated on that date. All of the regressions to date for a forecast point are evaluated, and the 50% chance of exceedance forecast is generated. A confidence interval around this forecast is generated at the 10%, 30%, 70%, and 90% exceedance levels (similar to official NRCS monthly water supply outlooks) by assuming a Gaussian error distribution centered on the regression output and having a width proportional to the calibration standard error. If a transformation was applied to the predictand, the error distribution is applied in transformed space before the exceedance forecasts are untransformed into real-world units (see NRCS, 2007 for more information).

From this information, three graphics are generated per forecast point (Figures 3-5). The data are also shown in map form in a variety of contexts [e.g., forecast as percent of normal, change in forecast over 1-

3-, 7-, and 14-day intervals, and the current day’s expected skill (similar to Figure 6)]. A spreadsheet file is also created containing all the information displayed in the plots and maps as well as additional diagnostic information. In these products, the automated forecast is referred to as the “Guidance volume forecast.” NRCS forecasts are routinely expressed in terms of thousands of acre feet (1 k-ac-ft = 1.23 million m<sup>3</sup>).

#### SKILL CHARACTERISTICS COMPARED TO OFFICIAL FORECASTS

The products display the expected skill of the forecasts, that is, the explained variance ( $r^2$ ) of the regression equation used at each forecast lead-time. A skill value of 1 indicates perfect correlation between the forecast and observed, whereas 0 indicates no relation. 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). One appealing aspect of least-

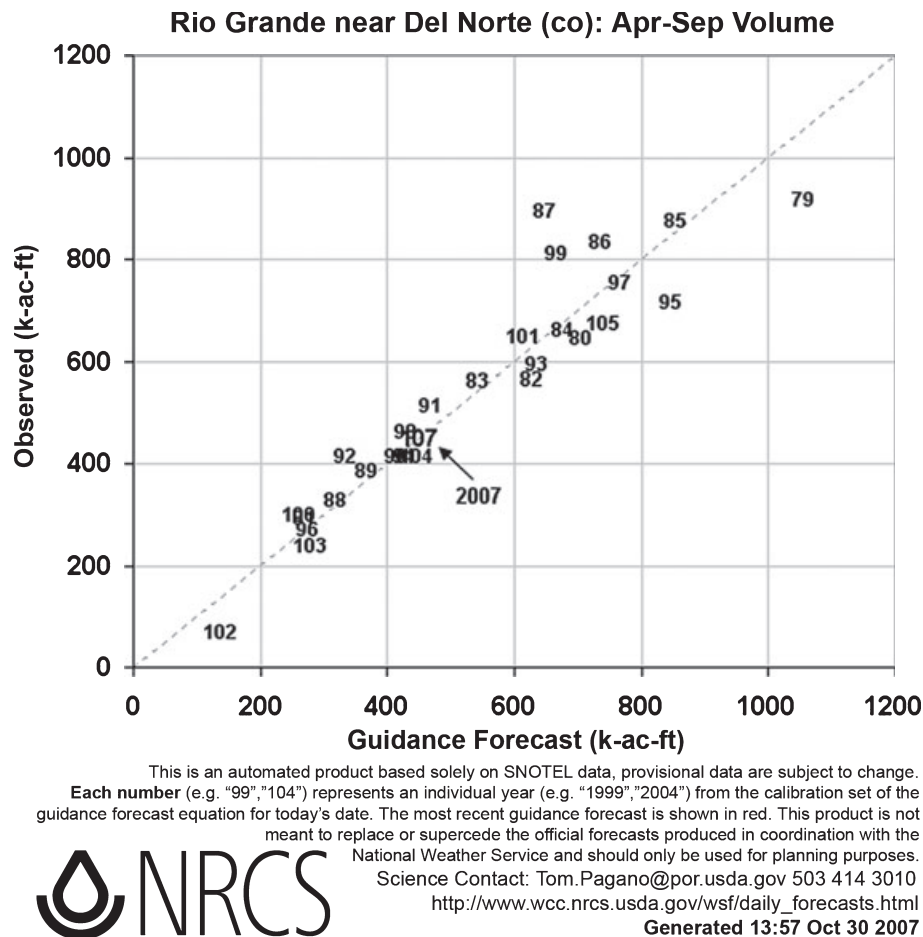


FIGURE 3. Scatter Diagram of Historical (calibration) Forecasts Issued Using April 1 Data *vs.* Observed Streamflow Volume for the Rio Grande Near Del Norte. The forecast using April 1, 2007 data is highlighted. This product was generated retrospectively on October 30, 2007, as indicated by the time-stamp in the figure. In real-time, the forecast for the current day would be shown.

squares regression-based forecasts is that they are constrained to be unbiased by construction, as opposed to forecasts produced from a simulation model. Therefore, in the absence of conditional or unconditional biases, explained variance is an appropriate skill measure.

The daily forecasts exhibit many of the skill characteristics of the official forecasts themselves (Pagano *et al.*, 2004a). Their skill *vs.* lead-time is primarily related to the importance of spring precipitation in the annual cycle, so that forecasts for locations where most of the water year's moisture comes during winter (e.g., Washington, Oregon, Idaho) perform better earlier in the season than those for places where the bulk of moisture normally comes during spring and summer (e.g., the Missouri River basin, the Rocky Mountain Front Range). They also perform well in regions that are snowmelt-dominated (e.g., the highlands of Colorado) and would not be expected to perform well in regions with midwinter melt, high base flows during winter, or complex subsurface geologic

processes (e.g., the west side of the Cascade Mountains in Washington and Oregon, east of the Cascades in central Oregon).

The most skillful April 1 forecast series calibrated to date has been the Boise River near Boise, Idaho. For this high elevation basin in which streamflow during the typically dry summers is driven largely by snowmelt, forecasts of April-July runoff have an  $r^2 = 0.91$  with the observed and an average absolute error of 9.3% of the 1971-2000 normal. Of the 148 points calibrated as of 2007 (Figure 6), half have April 1 skill of  $r^2 > 0.71$ , and half have skill between 0.59 and 0.80. With April 1 forecast *vs.* observed  $r^2$  of 0.2-0.4, the poorest performers are the Front Range basins such as the South Platte River basin of Colorado and the Canadian River basin of New Mexico, as well as the Musselshell River basin in Montana. None of the rainfall-dominated basins of the Pacific Northwest (e.g., the Willamette River basin) have been calibrated yet, although they are expected to be of low skill ( $r^2 \sim 0.4$ ), similar to the official forecasts.



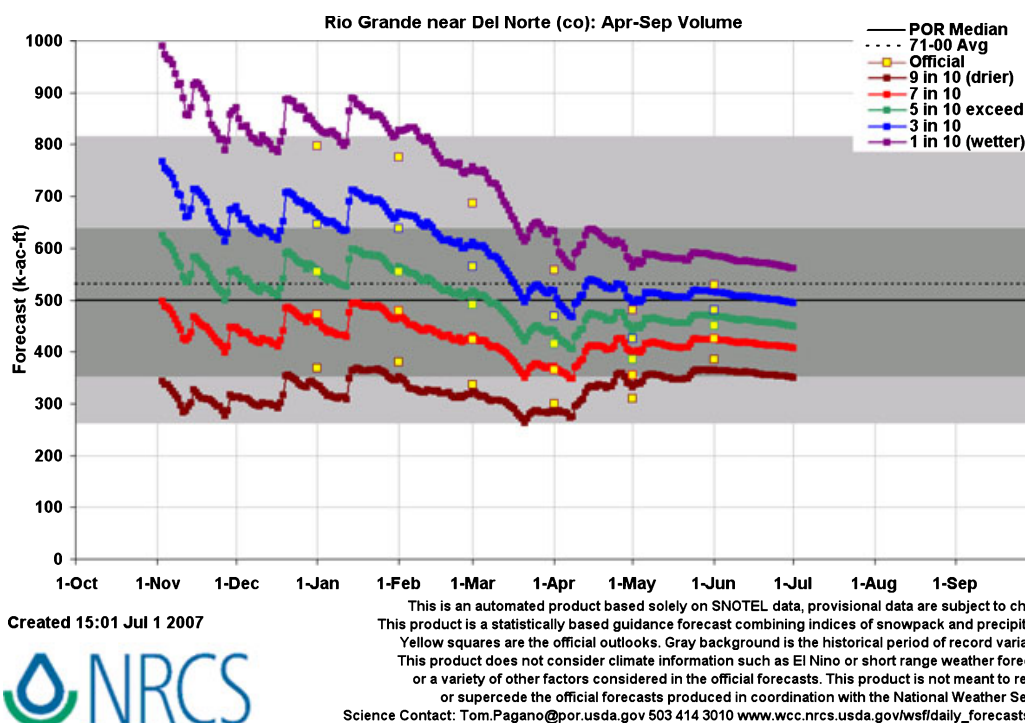


FIGURE 4. Daily Exceedance Guidance Forecasts for 2007 (colored lines) vs. Period-of-Record (POR) Historical Range of Variability (dark gray background is 70-30% exceedance probability interval; light gray background is 90-10% exceedance probability interval) and the Official Forecasts (yellow squares). Observed flow was 580 k-ac-ft. Note that the sudden rises in the forecasts during the winter indicate the occurrence of storms, and the snow declines in the forecasts are during inter-storm periods.

## RECEPTION AND PERFORMANCE

The daily forecast product was originally developed in October 2005 at the request of water managers on the Rio Grande, and early versions were also tested on users and NRCS personnel in Montana, Idaho, and the Upper Colorado River basin. The testing was unstructured and informal, and the subjective response was favorable. There was initial concern that users might confuse this product with the official forecasts, although the concern later proved unfounded (especially if the official forecasts are provided on the graphs for comparison). Indeed, several users indicated that the greatest strength of this tool is that it indicates the relative trend in the outlook since the last official forecast. This trend could be used as a “forecast of the forecast,” a mid-month indicator of how the first of month official forecasts may change. This information is particularly useful if the calendar of water management decisions does not align with the release of the official outlooks. For example, Upper Rio Grande water managers reassess their situation and adjust deliveries every 10 days, and some decisions are required to be made on the first day of the month, even though the official outlooks may not be ready until the fifth day of the month. A common user perspective is that having

daily updates to the water supply forecasts greatly reduces the chance of a major surprise when a new official forecast is released. Users are particularly concerned if a change in forecast conditions will shift them into a different management regime, and they appreciate the early warning of any potential changes.

By the end of 2006, 47 basins were being calibrated and run, and a skill evaluation was performed on the 39 basins with observations available. 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 30% of the time, and this actually occurs 30% of the time) was excellent. At the end of the season (September 30), the forecast reliability remained good (38% of observations fell in the 40% of the distribution between the 30% and 70% exceedance values of the forecasts, and 15% fell in the 20% of the distribution above the 10% exceedance or below the 90% exceedance). A preliminary evaluation of 2007 (98 of 148 observed values available) suggests stronger under-confidence (52% between 30% and 70% exceedance and 10% beyond 10% and 90% exceedance). There is significant spatial correlation of errors across sites, and therefore each basin is not an independent sample. Many years would be necessary to obtain a true assessment of the probabilistic aspects of the forecasts.

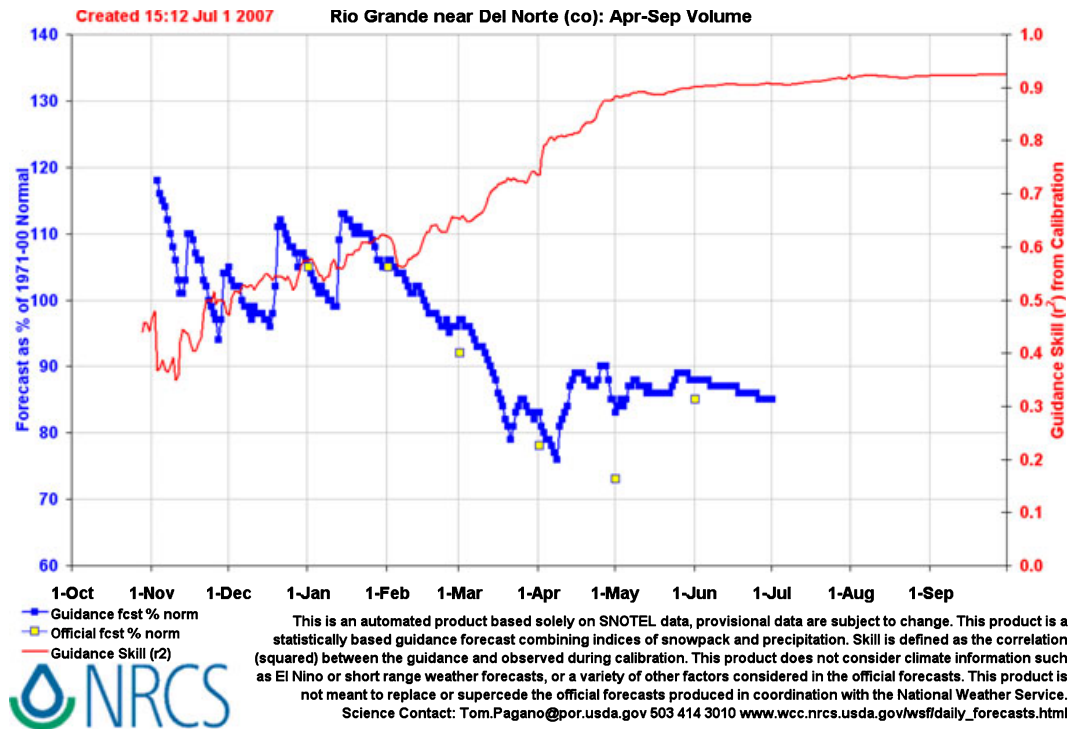


FIGURE 5. Guidance Volume Forecast for 2007 as Percent of Normal (blue line) vs. Official Forecasts (yellow squares) and Guidance Skill ( $r^2$ ) From Calibration. Observed flow was 109% of normal. Notice divergence in official and guidance forecasts in May. Also note that the sudden rises in the forecasts during the winter indicate the occurrence of storms, and the snow declines in the forecasts are during inter-storm periods.

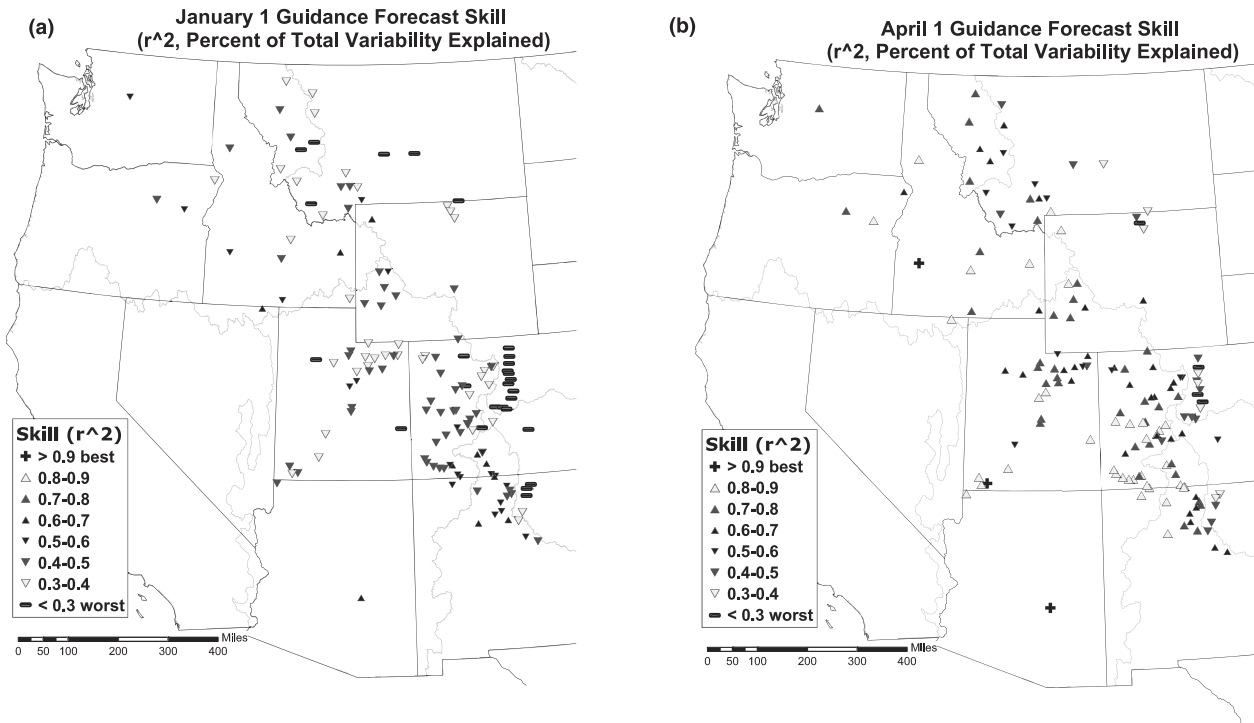


FIGURE 6. Historical Calibration Skill ( $r^2$ ) of the Forecasts Issued January 1 (a) and April 1 (b). Upward triangles indicate relatively high skill, and downward triangles mean low skill, with plus and minus signs highlighting the extremes. Skill on April 1 is highest in Idaho, the Colorado River basin, and parts of the Rio Grande. Skill is lowest throughout the Missouri basin (east flowing rivers in Montana, Wyoming, Colorado) and the Canadian basin in eastern New Mexico.

In 2007, the westwide average of the forecast error as percent of the 1971-2000 normal was 12% on April 1. In comparison, the official forecasts had an April 1 average error of 14%, slightly worse than the objective product. The westwide median forecast error was 10% for the guidance forecasts and 11% for the official forecasts, suggesting that the products are commonly comparable. The variance of the error of the official forecasts is greater than the objective guidance (i.e., where the skill is high, the official forecasts perform best, and when the skill is low, the objective guidance is superior).

Although it is unrealistic to judge the quality of a system based on a single year's performance, the result was unexpected, given that this 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 streamflow or soil moisture, springtime temperatures, or climate forecasts such as those based on the El Niño/Southern Oscillation. In particular, when part of the forecast target (e.g., April-July) is in the past (e.g., a forecast issued May 1), hydrologists know the observed flow to date with complete confidence, whereas that information is not included in this guidance product. This product also lacks the human expertise associated with the official forecasts.

## DISCUSSION

Although it offers many opportunities, the daily forecast tool has faced a few challenges in reaching operational production. First, this product is highly automated and relatively easy to calibrate and therefore requires a minimum amount of human supervision or intervention. Depending on one's perspective, this may be an asset or a liability. The tradeoff between automation and human intervention is a theme that has long been discussed throughout forecasting enterprises. Automation allows the use of a broader palette of tools, and objectivity protects the forecasts from the human forecaster's cognitive biases. On the other hand, improper automation excludes human expertise and gives the process a "black box" feel, potentially eroding the confidence in the product.

The second challenge has been the introduction of a new forecasting and regression technique. Garen (1992) compared the regression results for a basin in Idaho for PCA (described in Predictor Preparation and Filtering Using Z-Score Regression) and the existing forecasting method of the agency at the time. PCA displayed superior performance, and it has been NRCS's official regression technique for close to two

decades now. Although Pagano (2004) found promising results comparing the official historical outlooks to the cross-validated results of Z-Score regression at 29 basins, a systematic comparison between Z-Score and PCA does not currently exist. Therefore, conservative elements within the agency resist the use of Z-Score regression until these tests are completed and its operational behavior is better understood. This requirement is not uncommon; the NWS has indicated that new hydrologic forecasting models must fit within their existing infrastructure and demonstrate improved performance in an independent operational setting over several years (Hartmann *et al.*, 2003). The resources to perform such tests are generally unavailable (to researchers or forecasting agencies), hindering the transfer of research to operations.

The third challenge has been the delivery of raw model guidance directly to users. Myriad climate and weather model outputs are available on the Internet to help users build confidence in the official outlooks or to help them make their own judgments about what the forecast should be. In contrast, much of the inputs into the official water supply forecasts are not available to users because delivering "unprocessed" products is sometimes considered irresponsible and to be avoided. Forecasters are concerned that users may confuse raw model output with the official forecasts (Hartmann *et al.*, 2002). Considering that some forecasts are used to make major hydrologic decisions affecting lives and property, legal requirements, water rights, endangered species needs, and more, caution should be exercised. Although the daily update product described herein contains a lengthy disclaimer, management was asked to halt its release to the public while the legal liability of such raw automated products was considered, perhaps involving review by the Office of the General Counsel, an independent legal agency that reports to the Secretary of Agriculture. Production eventually resumed, in part because federal government forecasters consistently have been shielded from legal claims by users (Klein and Pielke, 2002).

In the early stages of development of this product there were concerns about the role it would play in the official forecast coordination process. With recent technological developments, forecasting agencies have struggled with the mandate from the early 1980s that both agencies must coordinate and agree on the official forecasts so as to "speak with one voice" to users. Does this mean that the agencies cannot release non-coordinated unofficial forecasts, such as raw model output? The NWS has wrestled with this issue in the context of displaying raw Ensemble Streamflow Prediction (ESP) output of the NWS River Forecast System (Day, 1985) on the Internet.

An interagency NWS/NRCS policy meeting on this topic was convened in January 2006. Ultimately it

was determined that, as long as a tool's output was clearly labeled as "guidance" and not an official forecast, it could be distributed to users. Official forecasts issued once a month would continue to be coordinated and issued jointly by the two agencies. The NWS has also begun development of an interactive tool to collectively display the raw output of the several inputs into the official outlooks, including the daily update product described herein.

These activities may allow more transparency in the forecast production process and encourage non-operational entities (e.g., universities) to develop and run their own models in real-time. While the official forecasts remain available for the conservative user in need of an expert assessment, early adopters will have a broader palette of guidance to draw from. The increased diversity of tools available may accelerate the adoption of innovative technologies by the operational agencies. By making research-grade products available to the public, user feedback can also enter into the process at an earlier stage of product development, and users can encourage operational agencies to invest in the most promising new tools.

## SUMMARY AND CONCLUSIONS

Historically, seasonal water supply outlook production has been a human resource intensive task, and forecasts have been issued rarely more than once a month and only during certain times of the year. This paper introduces a technique to produce these seasonal forecasts on a daily basis, year-round, using SWE and accumulated precipitation data from NRCS SNOTEL stations. The skill and character of these forecasts is similar to the official outlooks. This product also provides useful information about the possible relative intra-seasonal variations in the official outlooks. The trend toward allowing direct delivery of objective guidance to users is a significant new development in the operational hydrologic forecasting community, and it could facilitate the increased transfer of research (i.e., new objective methods) into practice at operational agencies. Further steps should also be taken to compare more comprehensively the behavior and skill of this product to the official outlooks as well as other tools from the research community.

## LITERATURE CITED

- Barnston, A.G.B., S.J. Mason, L. Goddard, D.G. DeWitt, and S.E. Zebiak, 2003. Multimodel Ensembling in Seasonal Climate Forecasting at IRI. *Bulletin of the American Meteorological Society* 84:1783-1796.
- Day, G.N., 1985. Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management* 111(2):157-170.
- Franz, K., N. Ajami, J. Schaake, and R. Buizza, 2005. Hydrologic Ensemble Prediction Experiment Focuses on Reliable Forecasts. *Eos, Transactions of the American Geophysical Union* 86(25):239.
- Franz, K.J., H.C. Hartmann, S. Sorooshian, and R. Bales, 2003. Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin. *Journal of Hydrometeorology* 4:1105-1118.
- Garen, D.C., 1992. Improved Techniques in Regression-Based Streamflow Volume Forecasting. *Journal of Water Resources Planning and Management* 118(6):654-670.
- Hartmann, H.C., R.C. Bales, and S. Sorooshian, 2002. Weather, Climate, and Hydrologic Forecasting for the U.S. Southwest: A Survey. *Climate Research* 21:239-258.
- Hartmann, H.C., A. Bradley, and A.F. Hamlet, 2003. Advanced Hydrologic Predictions for Improving Water Management. *In: Water: Science, Policy, and Management*, R. Lawford, D. Fort, H.C. Hartmann, and S. Eden (editors), American Geophysical Union, Washington, D.C., pp. 285-307.
- Klein, R. and R.A. Pielke, Jr, 2002. Bad Weather? Then Sue the Weather Man! Part I: Legal Liability for Public Sector Forecasts. *Bulletin of the American Meteorological Society* 83:1791-1799.
- NRCS (Natural Resources Conservation Service), 2007. Statistical Techniques Used in the VIPER Water Supply Forecasting Software. Engineering-Snow Survey and Water Supply Forecasting Technical Note #210-2. Natural Resources Conservation Service, Washington, D.C. [http://policy.nrcs.usda.gov/tn\\_sswsf\\_2\\_a.pdf](http://policy.nrcs.usda.gov/tn_sswsf_2_a.pdf), accessed October 2007.
- Pagano, T.C., 2004. The Role of Climate Variability in Operational Water Supply Forecasting for the Western United States. Ph.D. Dissertation, University of Arizona, Tucson, Arizona.
- Pagano, T.C., J. Erxleben, and T. Perkins 2005. Operational Simulation Modeling at the NRCS National Water and Climate Center. *In: Proceedings of the Western Snow Conference*, 87-100, Great Falls, Montana. Available at: [ftp://ftp.wcc.nrcs.usda.gov/downloads/factpub/wsf/Pagano\\_et\\_al\\_WSC\\_2005.pdf](ftp://ftp.wcc.nrcs.usda.gov/downloads/factpub/wsf/Pagano_et_al_WSC_2005.pdf), accessed October 2007.
- Pagano, T.C. and D.C. Garen, 2005. A Recent Increase in Western U.S. Streamflow Variability and Persistence. *Journal of Hydrometeorology* 6(2):173-179.
- Pagano, T.C., D.C. Garen, and S. Sorooshian, 2004a. Evaluation of Official Western U.S. Seasonal Water Supply Outlooks, 1922-2002. *Journal of Hydrometeorology* 5(5):683-698.
- Pagano, T.C., P. Pasteris, M. Dettinger, D. Cayan, and K. Redmond, 2004b. Water Year 2004: Western Water Managers Feel the Heat. *Eos, Transactions of the American Geophysical Union* 85:392-393.
- Schmidt, F.L., 1971. The Relative Efficiency of Regression and Simple Unit Predictor Weights in Applied Differential Psychology. *Educational and Psychological Measurement* 31:699-714.
- Serreze, M.C., M.P. Clark, R.L. Armstrong, D.A. McGinnis, and R.S. Pulwarty, 1999. Characteristics of the Western United States Snowpack from Snowpack Telemetry (SNOTEL) Data. *Water Resources Research* 35(7):2145-2160.
- Wood, A.W. and D.P. Lettenmaier, 2006. A Testbed for New Seasonal Hydrologic Forecasting Approaches in the Western U.S. *Bulletin of the American Meteorological Society* 87(12):1699-1712.
- Wood, A.W. and J.C. Schaake, 2007. Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. *Journal of Hydrometeorology* 9(1):132-148.

Barnston, A.G.B., S.J. Mason, L. Goddard, D.G. DeWitt, and S.E. Zebiak, 2003. Multimodel Ensembling in Seasonal Climate Fore-