2.2 EVALUATION OF NDFD AND DOWNSCALED NCEP FORECASTS IN THE INTERMOUNTAIN WEST

Brandon C. Moore[1]*, V.P. Walden[1], T.R. Blandford[1], B. J. Harshburger[1], and K. S. Humes[1]

[1]University of Idaho, Moscow, Idaho

1. INTRODUCTION

Weather forecasts in the Intermountain West are important for a variety of reasons, including public safety, agricultural planning, and water management. Water resource management is becoming more critical as competition for water increases, while water supplies diminish. Weather forecasts have recently been merged with hydrological models to provide more accurate streamflow prediction (Hay et al., 2002; Hay and Clark, 2003). This is being done at the same time that weather forecasts are improving rapidly, and verification of those forecasts is on-going. In addition, weather forecasting is particularly difficult in mountainous terrain where a large portion of streamflow runoff originates.

Various approaches exist for incorporating realistic weather forecasts into streamflow prediction. One approach is to use data from the new National Digital Forecast Database (NDFD) (Glahn and Ruth, 2003) from the National Weather Service (NWS), which issues 7-day forecasts of meteorological parameters two times each day on a 5-km grid. Although this is a useful, new data product, verification of the NDFD has not been widely published yet.

Another approach is to use medium-range forecasts from global models run by the National Center for Environmental Prediction (NCEP). These forecasts have the advantage of going out further in time (up to 15 days), but are quite coarse spatially (usually 2.5° grid cells in latitude and longitude). Model Output Statistics (MOS) is one method that is often used for "downscaling" forecasts to point measurements (Glahn and Lowry, 1972; Murphy, 1999; Antolik, 2000; Clark et al., 2004). Recently, Clark and Hay (2004) have devised a method of "downscaling" medium-range forecasts to point measurements within various watershed basins across the United States. Predictor variables, from a historical archive of global NCEP forecasts, are regressed with actual measurements from individual stations to forecast maximum and minimum temperatures.

In this paper, we examine the quality of temperature forecasts from both the NDFD and downscaled forecasts from the NCEP Experimental week 2 product (Hamill et al, 2004) over mountainous terrain in Idaho and northwestern Montana. The NCEP data are regressed with data from the Natural Resources Conservation Service's (NRCS) SNOpack TELemetry (SNOTEL) archive through 2001. The downscaled NCEP forecasts for 2002 through 2005 for the months of streamflow runoff (April through July) are then compared with actual SNOTEL observations. The NDFD data for 2005 are also compared with the SNOTEL data; the NDFD data from 2005 were the only data available for this study. The mean absolute error (MAE) and the rank probability skill score (RPSS) for both the NDFD and downscaled NCEP forecasts are presented. The downscaled NCEP results exhibit little dependence on elevation, but the accuracy of the NDFD forecasts decreases with increasing elevation.

2. DATA

2.1 Study Area

The study area for this research is shown in Figure 1 and is comprised of the state of Idaho and northwestern Montana. The intersections of the latitude and longitude lines in Figure 1 designate the locations of the NCEP forecasts. The boxes formed by the latitude and longitude lines define twelve grid cells over the study area from 40 N to 50 N and from 110 W to 117.5 W. The black circles represent the locations of the SNOTEL observations.

2.2 Observations

This study uses maximum and minimum temperatures measured at SNOpack TELemetry (SNOTEL) sites across Idaho and northwest Montana (Figure 1) both to build regression equations for downscaling the NCEP forecasts and as a baseline from which to compare recent NDFD and NCEP forecasts in the Intermountain West. Data collected from these stations are available in near real-time from the Natural

* Corresponding author address: Brandon C. Moore, University of Idaho, Department of Geography, Moscow, ID 83844-3021
e-mail: BrandonMoore@uidaho.edu

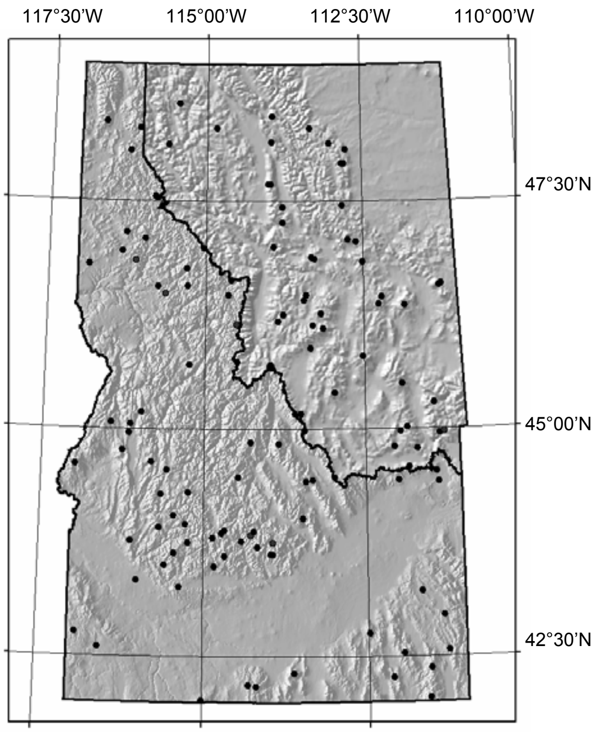

Figure 1: Map of the study area. The intersections of the latitude and longitude lines designate the locations of the NCEP forecasts. The boxes formed by the latitude and longitude lines define twelve grid cells over the study area. The circles represent locations where the SNOTEL observations are made.

Resources Conservation Service (http://www.wcc.nrcs.usda.gov/snotel/). There are currently 127 stations in Idaho and northwest Montana that fall within our study area and have less than 10% of their data missing. The majority of the stations that are currently operational began recording data between 1989 and 1991. Quality control was performed on the SNOTEL data by locating errors using a strict set of guidelines (Reek et al., 1992). Missing daily data values were computed and inserted using a techniques developed by Eisheid et al. (2000).

2.3 Forecast datasets

NDFD. Data from the National Digital Forecast Database (NDFD), produced by the National Weather Service (NWS) (Glahn and Ruth, 2003), is used here to compare with the SNOTEL observations. The 7-day maximum and minimum temperatures have a spatial resolution of 5 km. The data were matched to SNOTEL sites simply by selecting the individual NDFD grid cell that contained the SNOTEL site. Data from the NDFD were downloaded daily via automated FTP directly from the NWS during the 2005 snowmelt season (March-July); the Pacific Northwest data subset is used here.

NCEP. Data from NCEP's Global Forecasting System (GFS) model [formerly the Medium Range Forecast model (MRF)] are used here in two ways. First, the archive of "reforecasted" data (1979-2001) is used to create regression equations that downscale the forecasts to the locations of the SNOTEL sites. Second, recent forecasts during the snowmelt seasons for 2002 through 2005 are downscaled, then compared with the actual SNOTEL temperature observations for those years.

The GFS "reforecasted" data are available via FTP at the following site:

ftp:://ftp.cdc.noaa.gov/Datasets.other/refcst/ensdata/

A lengthy archive of forecasts is needed to produce accurate MOS-based regression equations for the downscaling process. The real-time forecasts can be downloaded via FTP from:

ftp://ftp.cdc.noaa.gov/Public/jsw/refcst/

The NCEP forecasts (both real-time and archived data) were generated using a frozen version of the T62 forecast model using 28-sigmal levels, saving forecasted output every 12 hours out to 15 days (Hamill et al, 2004). The 12-hour output represents the value of the variable from the previous 12 hours. The NCEP forecasts are generated from a 15-member ensemble that contains a control run and 7-paired forecasts centered about the control run, each using a slightly different set of initial conditions. The model provides global forecasts with a horizontal resolution of approximately 2.5°. The forecasted variables included in this study are 2-meter air temperature, mean sea-level pressure, relative humidity, 10-meter meridional wind, 10-meter zonal wind, total column precipitable water, and accumulated water. Clark and Hay (2004) identified this set of variables as important in prediction of 2-meter air temperatures.

3. METHODS

Multiple linear regression with forward selection is used to determine the MOS-based regression equations for downscaling. Unique regression equations are generated for each station, variable, month, and forecast day (1 to 15) using the technique outlined by Clark and Hay (2004).

3.1 Matching NCEP and SNOTEL temperatures

The NCEP forecasts and SNOTEL observations must be matched in both time and space for proper downscaling. Temporally, each NCEP variable is forecasted every 12 hours. That is, the value forecasted is for the previous 12-hour reference time, then out to the thirty forecast leadtimes (15 days). For example, the forecasted temperature at 0000 UTC corresponds to 5 pm (Mountain Standard Time; MST) and 12UTC corresponds to 5 am MST. Figure 2 shows that for Day (N) the maximum temperature will typically be achieved between 12UTC Day (N) and 00UTC Day (N), and the minimum temperature will typically be achieved between 00UTC Day (N-1) and 12UTC Day (N). Therefore, we calculate the downscaled minimum temperature forecasts using forecast leadtimes of 1, 3, 5…29, and the downscaled maximum temperature forecasts are for forecast leadtimes of 2, 4, 6…30. Therefore, daily forecasts of maximum and minimum temperatures can both be computed out to 15 days.

For each forecast day, the variables for the two surrounding forecast leadtimes comprise the predictor variables for the regression, i.e., to generate the maximum temperature for Day+0 (2400 UTC) variables from 1200 UTC and 1200 UTC Day+1 are included. The mean ensemble NCEP values for each variable are matched so that each SNOTEL data value will have 21 corresponding variables for regression, 7 variables for each of the three forecast leadtimes. The first complete year in a given SNOTEL record is used as a starting point for the regression.

The 21 NCEP forecast variables are then interpolated spatially to the location of the SNOTEL observations. This is accomplished by selecting the grid cell that contains a particular SNOTEL site. The four forecasts for each variable, issued for the corners of the grid cell, are then used to determine the forecasted value at the location of the SNOTEL site using Inverse-Distance-Squared-Weighting (IDSW). This interpolation was done for each forecast (all leadtimes) at each SNOTEL station location.

Any days in the SNOTEL data that are missing from a given month are deleted from the NCEP dataset and are, therefore, not used for regression. (Note that the missing values are replaced for the verification datasets from 2002 through 2005, but the missing values typically were less than 3% of the data). The NCEP dataset is also examined for "flatlined" data segments (erroneous constant values that persist within a time series), and those days are then removed from the regression datasets.

Data outliers are removed from both the SNOTEL and NCEP data using the method of Median of Absolute Deviations about the Median (MAD). The equation used to detect outliers with the MAD technique is:

$$S = 1.4826 * median\{|x_i - median(X)|\} \quad (1)$$

The value 1.4826 is used so that S will be approximately equal to the standard deviation for normally distributed data (Pearson, 2001, 2002). A rejection threshold (t*S) is created using t equal to 3, which is similar to the conventional 3-sigma

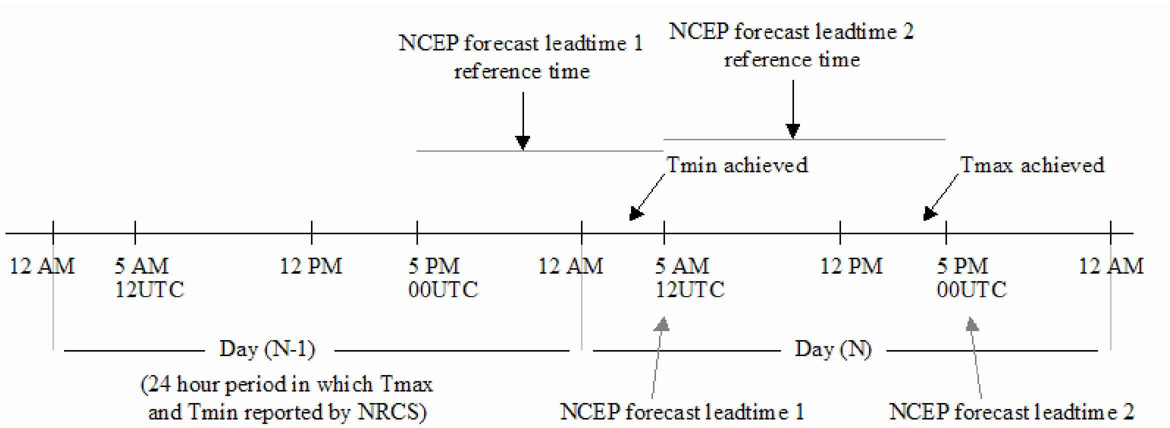

Figure 2: Time line of forecasted maximum and minimum temperatures at Day (N).

rejection technique (Costa et al., 1991). Any value exceeding this threshold is then determined to be an outlier and is removed from the dataset.

3.2 Downscaling Regression Coefficients

MOS techniques were used to create the regression equations for downscaling (Clark et al., 2004; Hamill et al, 2004). By downscaling the global forecasts to weather station locations, any existing bias in the forecasted values is significantly reduced.

The previously created dataset described in section 3.1 is used as input into the regression model. The 21 interpolated NCEP variables are then regressed against the historic SNOTEL temperatures using singular value decomposition to determine a regression equation for the desired month. Each year is successively held out of the regression computation for cross-validation.

Using multiple linear regression with forward selection, the MOS equations are generated using the NCEP variables as predictors. The regression models take on the form of

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon\ , \quad (2)$$

where the kth value represents the final predictor added, β is the regression coefficient, x is the predictor variable, and ε represents the remaining unexplained variance. Using the singular-value decomposition algorithm (Press et al., 1992), the predictor variable that explains the greatest amount of variance is removed. The remaining variables are then regressed once again to find the predictor variable that explains the largest amount of the remaining variance. This process continues until an additional variable explains less than 1% of the variance (Antolik, 2000). The standard deviation of the regression residuals is retained and the resulting regression equation is applied to the year withheld for cross-validation. A correlation coefficient is calculated and the process repeats for each year beginning from the first complete year of the station through 2001. The set of equations that produces the largest correlation coefficient is retained along with the standard deviation of the regression residuals. This procedure is repeated for each station, variable, month, and forecast day.

3.3 Downscaled-Forecast Production

The MOS-based regression equations developed from the historical forecasts and station observations are then used to generate real-time forecasts. The weights obtained from the original interpolation process are used to downscale the 21 variables from each of the four surrounding NCEP forecast points to the station location. The regression coefficients are then applied to the 21 NCEP predictor variables and a new temperature is computed for each forecast day (1 to 15). Daily forecasts are generated for each station, variable (maximum and minimum temperature), and forecast leadtime based on the monthly regressions.

Modeling the standard error of the regression residuals stochastically creates probabilistic forecasts. Using a normal Gaussian distribution with the mean equaling the forecasted temperature and a standard deviation equaling the model error, 100 random values are then generated. This essentially creates 100 different realizations of the temperature on a given day. From this, the probability of exceeding any given temperature can be determined.

4. RESULTS

The forecast skill of the downscaled forecasts using the MOS technique is assessed through the Ranked Probability Skill Score and the Median Absolute Error (Wilks, 1995). Although many various statistical measures are available, the Ranked Probability Skill Score and Median Absolute Error are most commonly used in forecast verification. The Ranked Probability Skill Score (RPSS) is computed from the ranked probability score:

$$RPS = \sum_{m=1}^{J} \left(Y_m - O_m \right)^2 \quad (3)$$

where Y_m is the forecasted probability of the forecasts for category m, and O_m is the cumulative probability of the observed time series for category m. Categories are determined from the observed time series based on equal interval. The cumulative probabilities for each forecast observation pairs are then computed. The RPS for a given forecast is then the squared difference of the forecasted probability and the observed probability squared, summed over all categories. The RPSS is the given as:

$$RPSS = 1 - \frac{\overline{RPS_{forecast}}}{\overline{RPS_{c\lim ato\log y}}} \quad (4)$$

where $RPS_{forecast}$ is the mean ranked probability score for all forecasts, and $RPS_{climatology}$ is the mean ranked probability score for the climatology.

The RPSS for the downscaled temperatures for 2002 though 2005 are shown in Figure 3. The dark line depicts the median RPSS for all stations for the maximum temperature, and the gray line represents minimum temperature. For the duration of the snowmelt season, the downscaled forecasts exhibit a higher degree of predictability compared to the climatology out to approximately 10 to 11 days. The RPSS for T_{min} and T_{max} are comparable for each month.

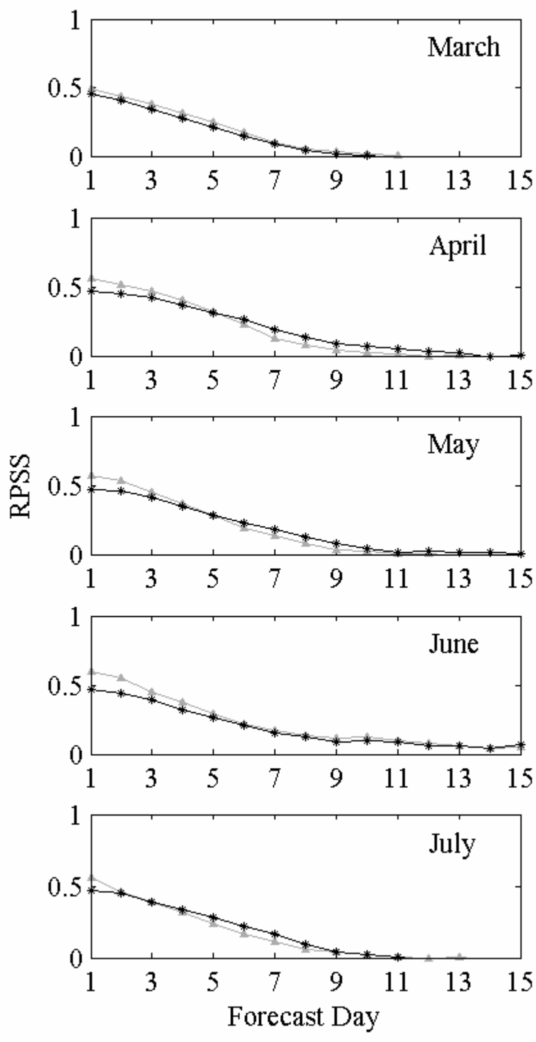

Figure 3: Ranked probability skill scores during the snowmelt season for T_{min} (gray line) and T_{max} (black line) for years 2002 - 2005.

The median absolute error (MAE) is used to compare the raw NCEP forecasts, the downscaled NCEP forecasts, and the NDFD forecasts for 2005 only; NDFD data were only available for that year. Figure 4 illustrates the MAE for the maximum and minimum temperatures during the 2005 snowmelt season. In general, the downscaled NCEP forecasts are comparable to the NDFD forecasts at each leadtime, except for T_{min} in June and July where the downscaled NCEP forecasts are better (lower MAE) than the NDFD forecasts. This may be due to the fact that only a single year is used for computing the MAE for both the NDFD and downscaled NCEP forecasts. So it is possible that these forecasts will have similar MAE values when multiple years are examined. It is also notable that the downscaling process greatly improves the raw NCEP forecasts of the minimum temperature throughout the entire snowmelt season and at short leadtimes for T_{max} in April.

The influence of elevation on forecast ability is also examined. Figure 5 displays the median absolute errors as a function of elevation of the SNOTEL sites. The decreasing trend in the NCEP errors is clearly evident as elevation increases. This is primarily because the downscaling process essentially accounts for elevation by interpolating the surrounding NCEP forecast points to the station location in determining the regression coefficients.

The dependence of the NDFD forecasts on elevation is less clear in Figure 5, but is made more apparent in Table 1. Table 1 shows the mean and standard deviation of the MAE for July 2005 in various elevation zones. While the mean of the NDFD MAE values does not significantly change between the elevation groups, the standard deviation of the forecast errors increases with increasing elevation. The mean of the NCEP MAE values steadily decrease with elevation, while the standard deviation remains fairly constant.

The downscaled forecasts appear to produce smaller errors at higher elevations than the NDFD forecasts. However, the cost of creating the downscaled forecasts needs to be examined in relation to the ease-of-use of the NDFD forecasts. Although easily transferable to other locations once the techniques have been developed, the downscaled forecasts require a large historic forecast dataset, preferably from a frozen version of a forecast model, which requires a significant amount of effort. However, the downscaled forecasts may provide a few extra days of forecasting skill, beyond the NDFD forecasts, from 7 to 10 days.

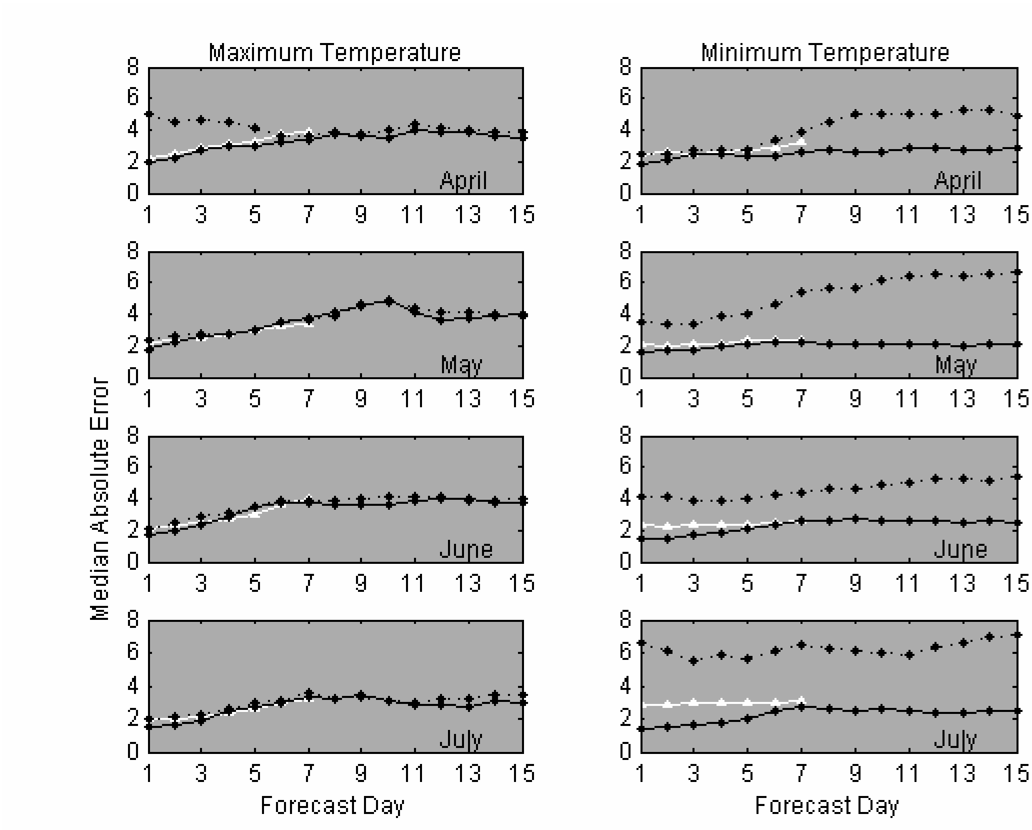

Figure 4: Median absolute error (MAE) at different forecast leadtimes for NDFD forecasts (solid white line), downscaled NCEP forecasts (solid black line), and raw NCEP forecasts (dashed black line).

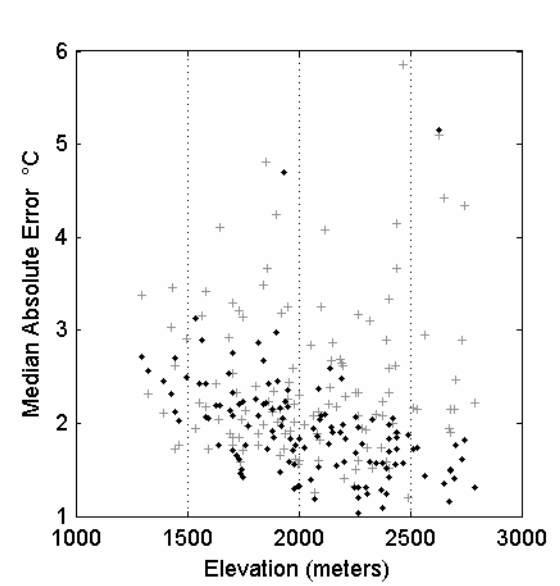

Figure 5: Median absolute error (MAE) as a function of elevation. NCEP errors are displayed as black, NDFD as gray.

Table 1: Mean and standard deviation of the Mean Absolute Error (MAE) for downscaled NCEP forecasts and NDFD forecasts for July 2005 by elevation.

Elevation (m)	NCEP MAE (mean & std)	NDFD MAE (mean & std)
1000 - 1500	2.42 ± 0.25	2.59 ± 0.65
1500 - 2000	2.07 ± 0.44	2.40 ± 0.74
2000 - 2500	1.71 ± 0.35	2.35 ± 0.85
2500 - 3000	1.53 ± 0.21	2.83 ± 1.08

5. Conclusions

As part of a larger project to improve streamflow prediction, weather forecasts for the Intermountain West during the snowmelt season (April-July) are examined from both the National Digital Forecast Database (NDFD) and downscaled from the Global Forecast System (GFS) model from the National Center for Environmental Prediction (NCEP). The downscaled NCEP forecasts at SNOTEL sites in Idaho and northwestern Montana exhibit forecasting skill out to about 10 to 11 days in both maximum and minimum daily temperatures for years 2002 through 2005. The downscaled forecasts are based on regression equations using the historical archives of SNOTEL observations (variable length) and NCEP reforecasted data from 1979 to 2001.

The Mean Absolute Error (MAE) is used to compare the raw NCEP forecasts, downscaled NCEP forecasts, and the NDFD for 2005. The downscaled NCEP forecasts are comparable to the NDFD forecasts in most cases, except for forecasts of minimum temperature in June and July. A closer examination of the MAE values for T_{min} in July 2005 show that the downscaled NCEP forecasts have lower MAE values at higher elevations than the NDFD forecasts.

Although each forecasting method performed well over the Intermountain West, there are certain advantages and disadvantages to each. The NDFD forecasts are readily available and easy to use, although they only extend to 7 days. The downscaled NCEP forecasts require a large initial time investment to generate the regression equations, but they produce forecasts with some skill out to 10 or 11 days.

6. References

Antolik, M. S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *J. Hydrology,* **239,** 306-337.

Clark, M, S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake Shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.*, **5**, 243-262.

Clark, M. and L. Hay, 2004: Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J. Hydrometeor.*, **5**, 15-32.

Costa, R., D. Orgiazzi, V. Pettiti, I. Sesia, and P. Tavella, 2004: Performance comparison and stability characterization of timing and geodetic GPS receivers at IEN. 18th European Frequency and Time Forum (EFTF), Guildford (UK).

Eisheid, J.K., and P.A. Pasteris, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.*, **39**, 1580-1591.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,***11,** 1203-1211.

Glahn, H.R. and D.P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.* **84**, 195-201.

Hamill, T., J. Whitaker, and X. Wei, 2004: Ensemble re-forecasting: Improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review*, **132**, 1434-1447.

Hay, L. E., M. Clark, R. L. Wilby, W. J. Gutowski, R. W. Arritt, E. S. Takle, Z. Pan, and G. H. Leavesley, 2002: Use of regional climate model output for hydrologic simulations. *J. Hydrometeor.,* **3,** 571-590.

Hay, L.E., and M.P. Clark, 2003: Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basin in the western United States. *J. Hydrology*, **282**, 56-75.

Murphy, J., 1999: An evaluation of statistical and dynamical techniques downscaling local climates. *J. Climate*, **12**, 1885-1908.

Pearson, R.K., 2001: Exploring process data. *J. Process Control*, **11**, 179-194.

Pearson, R.K., 2002: Outliers in Process Modeling and Identification. *IEEE Transactions on Control Systems Technology*, **10**, 55-63.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: *Numerical Recipes in Fortran.* 2nd ed. Cambridge University Press, 963 pp.

Reek, T., S. R. Doty, and T. W. Owen, 1992: A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network. *Bull. Amer. Meteor. Soc.,* **73**, 753-762.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction,* Academic Press, 467 pp.