Statistical Downscaling of the Warm Season Precipitation in the Core North America Monsoon Region

Kazungu Maitaria[1,2], David Gochis[3], Steven Mullen[1,4], David Yates[3] and James Shuttleworth[1,2,4]

1. Hydrology and Water Resources Department, University of Arizona, Tucson, AZ 85721. 2. Sustainability of semi-Arid Hydrology and Riparian Areas, University of Arizona, Tucson, AZ 85721.
3. The National Center for Atmospheric Research, Boulder, CO 80301. 4. Atmospheric Sciences Department, University of Arizona, Tucson, AZ 85721.

ABSTRACT

The North America Monsoon System (NAMS) is the principal feature of summer climate of Mexico and the Southwest U.S., and this study explores the development of statistically downscaled estimates of warm season precipitation over the core region of the North American Monsoon Experiment (NAME). Normalized accumulated daily-total summer precipitation anomalies for northwest Mexico for the period 1950 to 1998 are manipulated through a rotated empirical orthogonal function procedure in which three contiguous precipitation regions were realized. Hence, each of these sub regions is studied separately. Using output variables from a 1998 version of the National Centers for Environmental Predictions medium-range forecast (MRF) model, we sought to determine which variables are important predictors of the hydrometeorological and boundary layer features responsible for NAMS rainfall. Choice of MRF predictors is important. The K-Nearest Neighbor algorithm (KNN), an analog-type statistical downscaling technique, is applied to derive local-scale predictions of precipitation from specified MRF model variables. We evaluate the quality of downscaled product in terms of a standard suite of verification metric.

MOTIVATION

Forecasting of warm-season precipitation in the NAMS remains a challenging problem because the subtleties inherent in the local thermodynamic structure of the convective storms are complex. Moreover, it is not well understood how precipitation is forced by larger scale phenomena. Thus, our current research effort is geared towards better comprehension and improvement of forecasting the NAM warm-season precipitation by applying the KNN technique - a comparatively robust data-driven ensemble forecasting technique that indirectly accounts for sources of uncertainty on hydrological forecast.

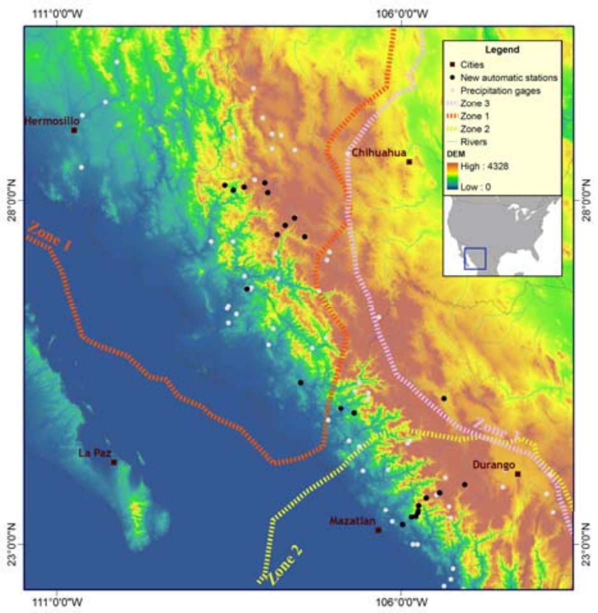

Figure 1. *The core region of the North American Monsoon Experiment (NAME)*

CLIMATOLOGY

The North American Monsoon Experiment (NAME) rain gauge network covers a topographically complex multi-watershed area between 22º and 30º N and 104º and 112º W (see Figure 1). This semi-arid region receives mainly convective precipitation during the warm monsoon season.

To enhance in-depth understanding of the structure of the precipitation regime we examined the seasonal average precipitation amounts for the July-September (JAS) period. Daily precipitation amounts are taken from a gridded, 1ºx1º data set for the period 1948-2002, distributed by the Climate Prediction Center (CPC). Principal component analysis, followed by varimax rotation, identified four regions. Zone 4 is outside of the basin study area (the Baja peninsula). The first four loading factors explained 83% of the total variance in the JAS series. Regions were identified by areas in which a loading factor for a particular principal component was greater than 0.5. This finding forms a basis for developing separate precipitation disaggregation models for the three regions identified in Figure 1.

ENSEMBLE DOWNSCALING

On average, the months of JAS are characterized by high frequency of convective events where studies have proven that moist convection plays significant role in the warm-season precipitation.

Firstly, candidate input fields from MRF model for statistical downscaling model (the KNN) were determined. Subsequently, a reformatted data matrix from these ensembles of daily MRF predictor fields over a much large domain encompassing all sub-watersheds was created. We attempted to run the KNN based on variables found in prior studies as being appropriate for midlatitudes and obtained dismal results. Hence, it was paramount to determine dataset types that corresponded to the many guises of convective phenomena in the JAS monsoon system. To this end we combined existing dataset types and use-determined thermodynamic variables datasets to drive the model.

Effort leading to user-determined thermodynamic variables dataset is enumerated as follows: -

i. operator inversion algorithms in NCAR's Spherepack 3.0 (Adam & Swarztrauber 1999) made it easy to manipulate existing netCDF-based MRF ensemble data to yield complementary variables, i.e., (a.) divergence and vorticity spectral datasets were used to compute both zonal and meridional winds at different levels and (b.) gridded products for vorticity, divergence and temperature at different tropospheric levels were generated from spectral products.

ii. an attempt to resolve the uncertainty necessitated inclusion of the effects of moisture through a consideration of potential atmospheric instability: saturated equivalent potential temperature and its gradient at levels between 700 and 500 mb & 500 and 250 mb respectively.

Note: Because the relative humidity information is only available at 700mb level, this constant has been assumed to be less important than the other thermodynamic constraints.

1. Surface total precipitation
2. Precipitable water
3. Two vertical gradients of saturated potential temperature
4. Average wind vectors at 500 and 250 mb levels
5. Divergence at 250 mb level

Table 1. *Preliminary selected candidate MRF variables for the KNN.*

The KNN is applied to derive local-scale predictions of precipitation from specified MRF model variables. The set of seven chosen predictors (Table 1) for the downscaling algorithm is evaluated against station observation data for the period 1979-1998. For a given grid point, the KNN matches the featured day ensemble fields with archived predictors by employing principal component-based technique to identify a subset of days (K) similar to the feature day. An ensemble of 105 members (15 medium range forecasts x 7 statistical predictions) is used to define an estimate of local precipitation.

PARTIAL RESULTS

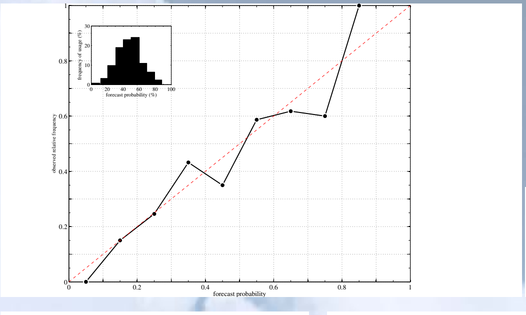

Figure 1. Basin reliability for the warm season (JAS) precipitation in the PC-2 basin. Insert is a histogram indicating frequency of use of the forecasts (refinement plot).

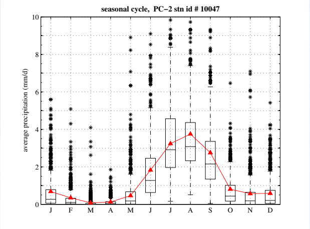

Figure 3. Box plots of total monthly rainfall for 105 ensemble members for selected gaging stations (whiskers show the 25th & 75th interquartile range), fcst lead time at day 3.

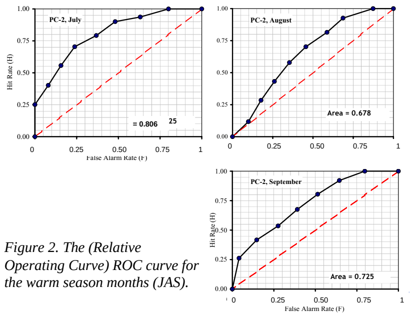

Figure 2. The (Relative Operating Curve) ROC curve for the warm season months (JAS).

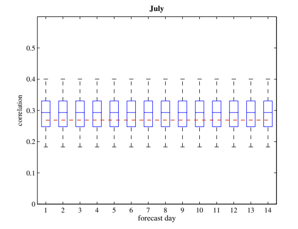

Figure 4. Box plots of spatial autocorrelation from 105 ensemble members between stns 25078 & 18001 for the 14-day fcst lead times.

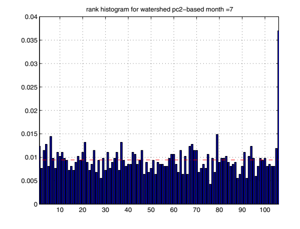

Figure 5. Rank histogram for JAS months (JAS) at 2-day fcst lead time with 105 ensemble members for the PC-2 basin (stations pooled) .

1. Results show that forecasted probabilities match the observed relative frequencies remarkably well (Figure 1).
2. The areas beneath the Receiver Operating Characteristic (ROC) curves suggest that our forecast system has ability to detect the occurrence of a precipitation 'event' (forecast discrimination ability, Figure 2).
3. Statistics of observed data fall within the box implying that KNN captures seasonal variations of precipitation (Figure 3).
4. The box plots of downscaled values adequately brackets the spatial correlation of the JAS precipitation (Figure 4).
5. The rank histograms are relatively flat, demonstrating that the KNN produces realistic ensemble spread (Figure 5).

The following questions will be expanded and clarified over time:
How can the KNN ensemble forecasts be used reliably? What adaptation requirements are needed for the spatial and temporal characteristics of the monsoon system to be integrated into a hydrological ensemble system?