New divisions for monitoring and predicting climate

The following is an adaptation of an article by Klaus Wolter and Dave Allured, University of Colorado at Boulder, CIRES Climate Diagnostics Center, and NOAA-ESRL Physical Sciences Division. It appeared in the June 2007 issue of the Intermountain West Climate Summary .

Climatologists have long questioned the accuracy of the current climate divisions (CDs) in the United States in representing regional climate. To address their concerns, we embarked on a long-term effort in 2003 to create a more rational, statistically-based set of national CDs that would help improve drought monitoring and climate forecasting. The result, thus far, is an experimental map of new CDs that more accurately represents U.S. climate.

Near-real time climate monitoring, long-term climate change assessments, and statistical climate predictions are often based on data aggregated into CDs. CDs come from century-long efforts to organize climate observations across the country to match up with crop reporting districts, county lines, and/or drainage basins; the CD boundaries were finalized in the 1950s. Perhaps surprisingly, given their use, climate, based on objective groupings of long-term observations, was not the primary consideration in determining the CD boundaries (Guttman and Quayle, 1996).

The vast majority of data used in climate analyses comes from stations that are part of the voluntary Cooperative Observer Program (COOP) at NOAA. This network of stations has been collecting daily high and low temperatures, precipitation, and snowfall since 1890. CD data are computed by simply averaging all available, representative COOP station data within each division since 1931 into single monthly values. Data prior to 1931 were derived from statistical relationships between current division data and state-wide averages. CDs

continued on page 4

Climate division methodology

In order to ascertain which climate stations have the tendency to exhibit the same climate anomalies, we performed analyses on temperature (T), precipitation (P), and combined records. We found that the last approach, with combined time series, yielded the best-defined climate regions.

From currently available records for 17,575 COOP stations in the lower 48 states, we selected 4,324 stations with both sufficient P and T records to perform statistical analyses for water years 1979–2006. For much of the U.S., this translates into at least one station per 1,000 square miles. Some less populated regions, such as the deserts in the Interior West, have less dense spatial coverage. We used several thousand more P COOP stations of similar quality for supportive analyses. In addition, there are more than 500 SNOTEL sites in the higher elevations of the western U.S. that have sufficient P records since 1979 to be analyzed as well. However, their T records typically only start in the late 1980s and have been somewhat unreliable. To develop new experimental climate divisions (CDs) we used five steps:

Step1. For every climate station, we computed average T and P totals for every three-month season from October 1978–September 2006. These ‘sliding’ seasons include all three month periods (i.e., October–December, November–January) within the 28-year record. Individual seasonal anomalies were calculated by subtracting the 28-year average for that same season. For missing data, anomaly values were set equal to zero to keep all station anomaly time series to the same length.

Step 2. Multivariate cluster analyses, a statistical method for grouping data in a way that yields a strong degree of association between members of the same cluster and a weak degree of association between members of different clusters (http://www.statsoft.com/textbook/stcluan.html), were used to find out which stations tended to experience climate anomalies of the same sign (i.e., above average or below average), based on correlation matrices among all of them. The two cluster analysis techniques applied here were Average Linkage and Ward’s method, both well-established and superior to other methods (Wilks, 1995, pp. 419-428).

Step 3. Results from both clustering methods were compared against each other and used to group stations with similar T and P anomalies into core regions. A large majority of these cores could be identified by simple overlapping station counts, but some less clear-cut groupings were settled by correlating the respective cluster time series against each other. After this initial classification, core time series were computed based on normalized T and P time series (produced by taking each data value, subtracting from it the long-term average, then dividing by the standard deviation) at the station level. These were used to calculate correlation coefficients between all stations and all cores.

Step 4. The assignment of stations to cores was refined iteratively, until no changes occurred. In particular, if a station was not classified as belonging to a core, but correlated highly with a nearby core, it was admitted to that core; or if a station had been classified as being inside a core, but did not correlate highly with the core time series, it was removed from that core. (This was a rare event in

continued on page 5

Climate divisions, continued

are used in many climate-related monitoring products, like the U.S. Drought Monitor (see page 8), because they allow for an easy calculation of regional averages and a comparison of recent climate anomalies against a century-long record. The NOAA Climate Prediction Center (CPC) has used so-called "mega-divisions," which are based on merging smaller CDs, as targets for climate predictions and for verifying forecasts.

The 344 U.S. CDs allow for up to ten divisions per state; however, they cover the conterminous United States rather unevenly (Figure 1a). Many states, such as Wyoming and Idaho, have ten divisions, but some rather large states do not. Arizona, a large state with complex topography, is represented by only seven CDs, some of which may not accurately represent regional climates. For example, the northeast third of the state, from the Mogollon Highlands and San Francisco Peaks, across the Painted Desert to the Four Corners is represented by a single division. Similarly, the southeastern Arizona CD stretches from the parched deserts of Organ Pipe National Monument, across the lofty Sky Island mountain ranges to the Gila River headwaters. Decisions about how to organize CDs were made on a state-by-state basis rather than from a national perspective (Guttman and Quayle, 1996).

Although the CD data provide a long, consistent, and gap-free record, climatologists have long questioned the assumption that the simple averaging of COOP stations into CDs is optimal for depicting regional climate, especially precipitation. To examine this issue, we correlated individual COOP station data with divisional averages (Figure 1b). Results show that much of the high elevation Interior West, especially along the Rocky Mountains down into New Mexico, is not well represented by divisional averages. During the winter

continued on page 5

Figure 1a. The 344 climate divisions currently in use for the conterminous United States. For more information visit http://www.cdc.noaa.gov/USclimate/USclimdivs.html.

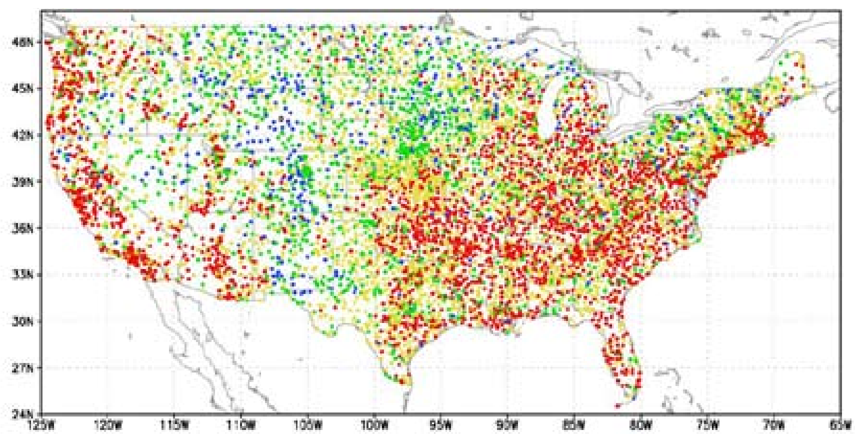

Figure 1b. Seasonal correlations between climate division time series and COOP station time series during January–March 1979–2002. Green and blue dots show that divisional indices account for less than 50 percent of the local seasonal precipitation variance (i.e., values less than 0.5).

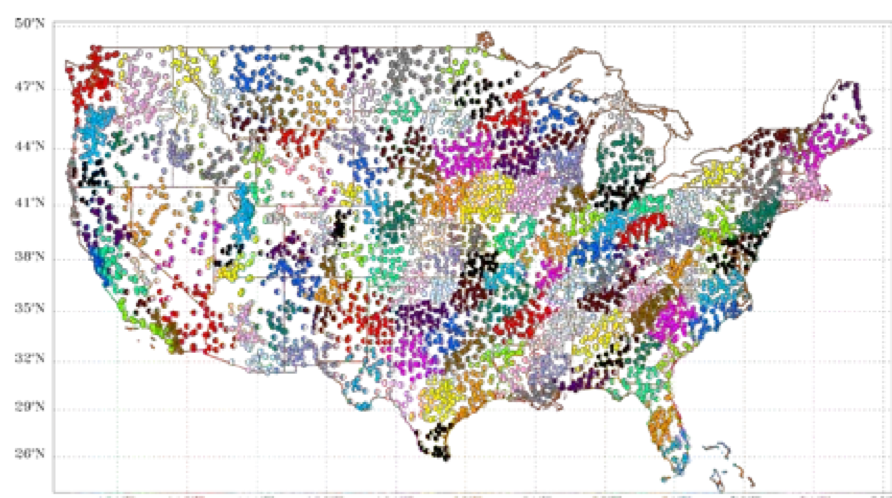

Figure 1c. Near-final map of new climate divisions, based on temperature and precipitation station data. Each dot is a COOP station, and a cluster of dots of the same color represents a new climate division. For more information visit: http://www.cdc.noaa.gov/people/klaus.wolter/ClimateDivisions/

Climate divisions, continued

snow accumulation season in parts of the Interior West, there are poor correlations between individual stations and the associated CD (Figure 1b), and the situation is even worse in the summer.

Low correlations between individual COOP stations and divisional averages translate into poor reliability when large-scale drought assessments or ENSO-related forecasts based on these divisions are scaled down to the station level. This is one reason why drought monitoring and seasonal climate forecasting are difficult in the Interior West. In addition, some of the higher elevation Snowpack telemetry (SNOTEL) sites, operated by the USDA Natural Resource Conservation Service, may correlate negatively with their CD time series. This is due to orographic effects in high mountain areas: during the winter season, strong westerly winds yield large snowfall amounts on the windward side of mountain ranges, while the valleys to the east may experience windstorms and dryness. Because most COOP stations are located in valleys, CD averages may end up with precipitation deficits when compared with long-term averages, whereas SNOTEL-based precipitation assessments may show precipitation surpluses. This type of precipitation pattern is not well captured by the current CDs.

Analogous maps for seasonal temperature correlations do not show the same disparity between station and CD data, most likely because temperature variations are similar over larger regions than precipitation variations. Nevertheless, wintertime regional temperature anomalies are also not well represented by climate divisions in the orographic regions of the Interior West.

In 2003, we launched a project to create a different set of national CDs that would help improve drought monitoring and climate forecasting in the U.S. (see sidebar for methodology). The result is a map of new CDs, based on temperature and precipitation station data, which are no longer bounded by state lines (Figure 1c). Note the divisions along the borders of California, Arizona, Utah, Nevada, New Mexico, and Texas. For example, the map shows divisions that encompass the climatic similarity of the southeast corner of Arizona and the southwest corner of New Mexico. Both have similar ecosystems and year-to-year precipitation variations.

In addition, there is no upper limit of ten divisions per state. One of the goals of this project was to integrate SNOTEL sites into the analysis. We found that SNOTEL data correlates well with the new CDs, and most of the SNOTEL sites match up nicely with the nearest COOP-based CD.

With the creation of the joint temperature and precipitation maps (Figure 1c), this project is almost complete. The remaining stage is to fine-tune the new division boundaries with precipitation data from SNOTEL and precipitation-only COOP stations. For more information on the new climate divisions, visit the NOAA Earth System Research Laboratory web site (Figure 1c). We are working on the additional products, including additional time series of temperature and precipitation averages in each new climate division, from 1978–2006, and from 1948–1978, based on new climate divisions for that period, and final new climate division maps, including boundaries, spatial coverages (in percent of area), and new state-wide averages.

References

Guttman, N.B., and R.G. Quayle, 1996. "A historical perspective of U.S. climate divisions." *Bulletin of the American Meteorological Society*, 77, 293–304.

Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, 467pp.

Methodology, continued

the combined analysis suite, but more common in P analyses). A third scenario involved the transfer of a station from one core to another, if its correlation with the new core was substantially higher than with the old core.

Step 5. While there was some experimentation with correlation thresholds, the basic procedure always remained the same and yielded similar results. Transfers between core regions required at least a 1 percent increase in explained variance for that station, and the drop-correlation threshold had to be lower than the add correlation threshold. The final correlation thresholds were in the 0.55–0.60 range, to allow for virtually all stations to be classified. One final check consisted in correlating all new CD time series against each other to flag regions that were extremely well correlated ($r>0.90$) and thus prime candidates for mergers, as long as the resulting new division did not exceed certain size limitations.

The current version of the new 139 combined core regions (i.e., new CDs) for water years 1979–2006 is shown in Figure 1c. From the pool of 4,324 COOP stations with sufficient temperature and precipitation data, the initial core map classified 3,112 stations as being within 145 initial clusters (Step 3). Using the iterative methodology described above, the remaining stations were gathered into core regions, resulting in a stable classification of all but one station by the seventh iteration in 139 final core regions (Steps 4 and 5; Figure. 1c). While there was no requirement for stations within a core to be spatially adjacent to each other, it is reassuring to see that virtually all of them are indeed neighbors, even in the more diverse terrain of Wyoming, Colorado, Utah, and New Mexico.