2.14 Station density strategy for monitoring long-term climatic change in the contiguous United States

Michael J. Janis*
Southeast Regional Climate Center, SC Department of Natural Resources, Columbia, SC

Kenneth G. Hubbard
High Plains Regional Climate Center, University of Nebraska, Lincoln, NE

Kelly T. Redmond
Western Regional Climate Center, Desert Research Institute, Reno, NV

1. INTRODUCTION

The Climate Reference Network (CRN) is being established to monitor present and future climatic variability across the United States. The initial proposal for the CRN assumed a network of about 250 stations would be sufficient to capture the climatic signal for the nation. This station density was inferred from a methodological examination of Twentieth Century US precipitation trends (Karl and Knight 1998). Their study estimated that a network of 182 stations could reasonably reproduce the 1910-1996 trend in annual precipitation computed from the climate division data set. The purpose of this study is to estimate the spatial density and total number of stations required to reproduce, within predetermined monitoring goals, the observed climatic variability across the contiguous US.

2. BACKGROUND

Previous studies have examined the role of station density in capturing the spatial variability in the regional data sets. For example, Hubbard (1994) found that one station every 60 km in relatively simple terrain was adequate to capture 90% of the spatial variability in daily temperature. Network resolution for capturing daily precipitation variability was an order of magnitude higher (5 km). Based on an 814-station subset of the US Historical Climatology Network (HCN), DeGaetano (2000) found 321 station clusters represented the spatial variability of seasonal precipitation across the contiguous US. Since 101 of 321 total clusters were single-station clusters and the reference network was spatially coarse in the Western United States, at least 321 stations are necessary to resolve the spatial variability of seasonal precipitation across the contiguous United States. A network of 250 stations for the contiguous US, assuming a uniform distribution, is equivalent to approximately one station every 180 km. Comparatively, with 1219 stations for the contiguous US, the HCN has approximately one station every 82 km (Easterling et al. 1996).

**Corresponding author address:* Michael J. Janis, Southeast Regional Climate Center, South Carolina Department of Natural Resources, Columbia, SC 29205; e-mail: janis@dnr.state.sc.us.

3. TEMPERATURE AND PRECIPITATION DATA FOR THE UNITED STATES

High-resolution observed databases are used to represent discrete representations of climatic variability. Temperature and precipitation data are drawn from the 1971-2000 Climatography of the United States sequential database (CLIM81; NCDC 2002). CLIM81metadata are screened for flags indicating estimated or adjusted data. Stations containing more than three years with more than three flags each are culled from the analyses. The resulting networks consist of 3642 temperature stations and 5156 precipitation stations. Sequential temperature and precipitation data are transformed based on the 1971-2000 reference period, into annual temperature anomalies and percent-of-median annual total precipitation.

4. METHODS

The approach is to systematically decrease network resolutions from an initial high-resolution baseline network. Each step involves generating measures of similarity between networks of lower spatial resolution and the baseline network. The ideal density for climate monitoring networks is the number of stations that reproduce trends in the baseline networks within predetermined climate monitoring goals. For example, a monitoring system for climatic change may have the following goals, “temperature change for any location within an area can be represented by a single station with an average mean-absolute-error less than 0.1°C per decade.”

Instead of performing a global spatial sampling of stations in CLIM81, these analyses are performed on weather stations stratified into 2.5° latitude × 3.5° longitude grid cells. Applying a local sampling strategy within each grid cell overcomes any awkward network arrangements that can result from a global random sampling across the entire network. Ensemble average time series of temperature anomalies and precipitation percentiles from either baseline or subnetworks are computed within each grid cell. The required number of stations to satisfy a monitoring goal defines the grid-cell station density. After repeating the process for all grid cells, a total number of stations satisfying the monitoring goal are determined. In this case, the density is variable from region to region although the goal is the same from region to region. Local spatial analyses help determine where higher or lower network densities are needed to satisfy climate monitoring goals.

Within each grid cell, networks of lower spatial density are derived by randomly selecting subsets of stations. The number of stations within each subset is incremented by one (1) so that all subset sizes from one to N-1 are examined (N is the total number of stations within a grid cell). Average or 'ensemble' time series based on individual station time series are generated for each subnetwork. A Monte Carlo resampling procedure is applied to produce 100 different realizations of each subnetwork size. This procedure addresses the effect of multiple network configurations, reduces the effect of poorly distributed subnetworks, and reduces the influence of any undetected inhomogeneous station records. Similar resampling approaches have been successfully employed to examine the influence of sample size on spatial interpolation of annual total precipitation (Willmott et al., 1996) and monthly average temperature (Robeson and Janis, 1998).

Grid-cell ensemble trends of temperature and precipitation are computed from annual average temperature and annual total precipitation. For each N_s-station network, an error statistic for temperature and precipitation trends can be computed (N_s is the number of stations in any subnetwork). Climate monitoring goals for temperature and precipitation trends are determined and the corresponding station density is identified. A Monte Carlo procedure is applied to each grid cell as follows:

1. Randomly sample, without replacement, an N_s-station subnetwork ($1 \le N_s \le N-1$, where N is the total number of stations in a grid cell).
2. Generate ensemble time series of temperature or precipitation from N_s-station subnetwork.
3. Compute thirty-year linear temperature or precipitation trends for each realization of an N_s-station ensemble time series.
4. Repeat steps 1 through 3, with replacement, 100 times to generate multiple realizations of subnetwork trends. For any N_s-station subnetwork, MAE for temperature and precipitation trends is computed as:

$$\mathrm{MAE} = \frac{1}{100} \sum_{k=1}^{100} \left| \frac{\Delta \overline{T}}{\Delta t} - \frac{\Delta \hat{T}_k}{\Delta t} \right|.$$

where 100 is the number of Monte Carlo realizations, $\Delta\overline{T}/\Delta t$ is the temperature or precipitation trend for the baseline time series, and $\Delta\hat{T}_k/\Delta t$ is the temperature or precipitation trend for k^{th} realization of an N_s-station subnetwork.

5. Repeat steps 1 through 4 for all possible subnetwork sizes (N_s = 1, 2, …, N-1).
6. Perform polynomial regression between MAE and subnetwork size (N_s) by:

$$N_s = a_0 + a_1\mathrm{MAE}^1 + \cdots + a_4\mathrm{MAE}^4.$$

A priori assumption is that MAE decreases as the subnetwork density approaches the baseline network density. The selection criteria for resolving the United States climatic variability are based on the relationship between MAE and N_s. For each grid cell, the network density associated monitoring goals are identified.

5. RESULTING STATION DENSITIES

Gridded N_s-station networks corresponding to climate monitoring goals are mapped to provide valuable information regarding regions of the country that are most sensitive to network configuration and network density. Regions that are characterized by high spatial variability display a high degree of variability between Monte Carlo realizations and have overall higher mean-absolute errors.

5.1 *Annual Air Temperature Trends*

Four temperature-trend monitoring goals as well as a null hypothesis (one station per grid cell) are examined. A monitoring goal for annual air temperature trends is 0.05°C per decade. Solving a polynomial equation for N_s with this goal can be interpreted as identifying the network resolution necessary to reproduce annual temperature trends from a baseline network to within 0.05°C per decade. A national network meeting this monitoring goal consists of 627 stations with an average station separation of 148 km (Table 1).

Though the western US often requires more than 6 stations per grid cell, regional patterns of network density are not easily drawn (Fig. 1). More than 40% of grid cells require 4 to 5 stations per cell. Some grid cells require 10 or more stations but rarely fewer than 3. Grid cells requiring as few as 1-2 stations are located along national borders or coastlines. Approximately 60% of grid cells require a 112-162 km spatial separation between stations to meet this monitoring goal.

Table 1 Number of stations satisfying temperature goals. Null represents one station per cell.

Temperature Trend (°C/decade)	Number of Stations
0.050	627
0.075	338
0.100	233
0.125	167
Null	114

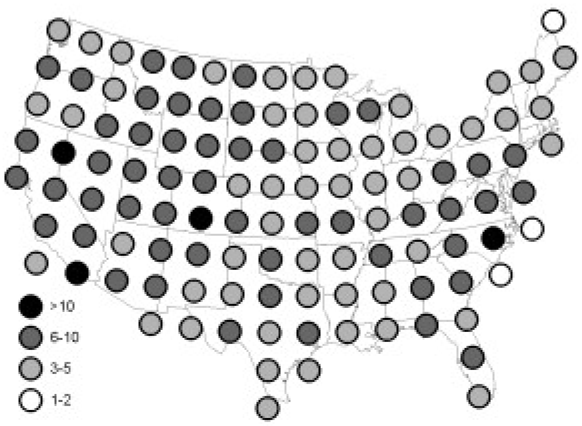

Figure 1 Grid-cell densities of 627 stations satisfying an annual temperature-trend monitoring goal of MAE < 0.05°C per decade.

A national network meeting a monitoring goal of 0.10°C per decade consists of 233 stations with an average station separation of 245 km (Table 1). More than 35% of grid cells require only one station to meet this monitoring goal. With nearly 90% of all cells requiring fewer than 3 stations to meet this monitoring goal, little spatial variability in network density results (Fig. 2). Station spacing increases when monitoring goals are relaxed.

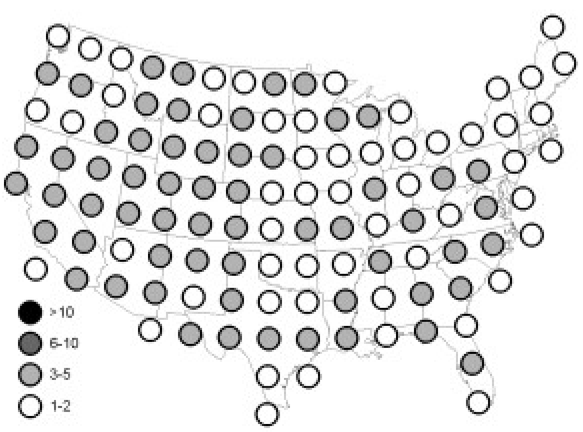

Figure 2 Grid-cell densities 233 stations satisfying an annual temperature-trend monitoring goal of MAE < 0.10°C per decade.

5.2 *Annual Total Precipitation*

Four precipitation-trend monitoring goals as well as a null hypothesis are examined (Table 2). A national network meeting a monitoring goal of 1.00 cm per decade consists of 553 stations with an average station separation of 157 km. Lower network densities are generally found west of the Mississippi River, while the highest network densities are found in the southeast and northwest US (Fig. 3). Although 3 or fewer stations per cell commonly satisfy this monitoring goal, 8 grid cells require more than 10 stations. Approximately 45% of grid cells require a 112-162 km spatial separation between stations.

Table 2 Number of stations satisfying precipitation goals. Null represents one station per cell.

Precipitation Trend (cm/decade)	Number of Stations
1.0	553
1.5	308
1.75	236
2.0	189
Null	115

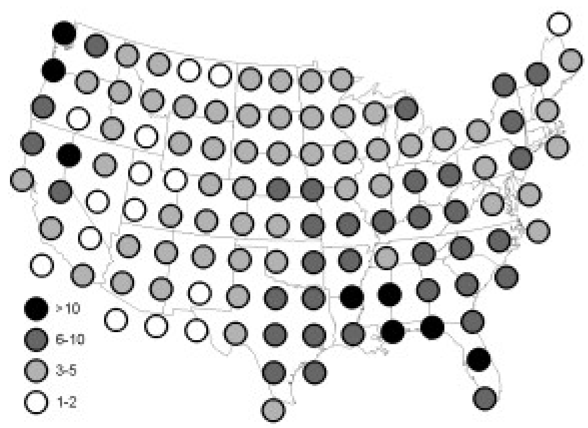

Figure 3 Grid-cell densities of 553 stations satisfying an annual precipitation-trend monitoring goal of MAE < 1.00 cm per decade.

A national network satisfying a monitoring goal of 1.75cm per decade consists of 236 stations. To meet this monitoring goal, higher network densities are necessary in the southeastern US, while lower network densities are necessary in the west with the exception of the coast (Fig. 4). More than 50% of grid cells require only one station to meet this monitoring goal. The most common spatial separation per grid cell is 287-337 km, but higher resolutions (e.g., < 200 km) are necessary for some grid cells.

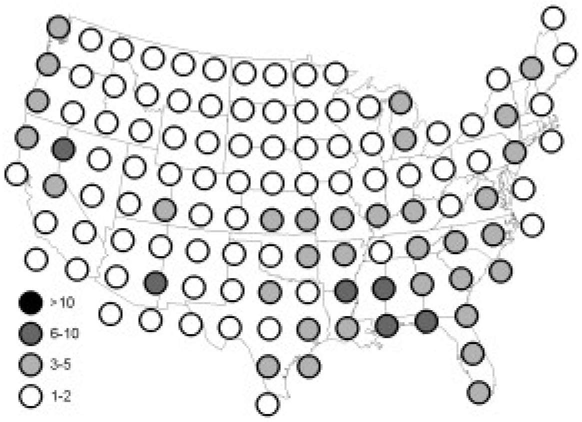

Figure 4 Grid-cell densities of 236 stations satisfying an annual precipitation-trend monitoring goal of MAE < 1.75 cm per decade.

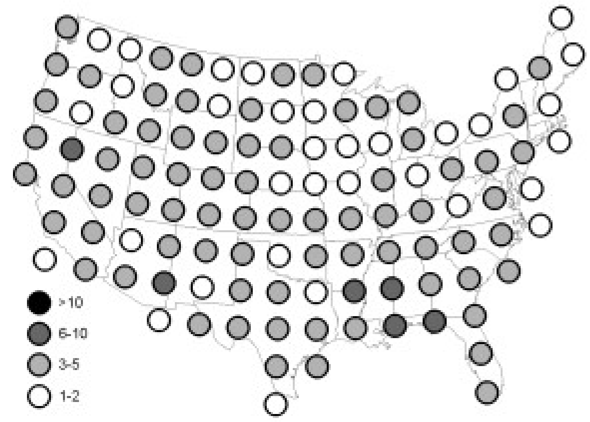

Figure 5 Grid-cell densities of combined 307 stations satisfying monitoring goals of 0.10°C per decade for temperature trend and 1.75 cm per decade for precipitation trend.

6. SUMMARY AND CONCLUSIONS

A goal of this work is to provide a recommendation for CRN station density. Spatial density is examined relative to measured precipitation and temperature from a spatially dense network of existing weather stations. The assumed minimum station density is a uniform spatial distribution of one station per 2.5° latitude × 3.5° longitude grid. The grid-based approach provides regional estimates of network density that satisfy monitoring goals. A stratified local sampling strategy with Monte Carlo resampling techniques applied within each grid cell are used to build information on how trend estimates may diverge with decreasing network density. Regions of the country that require higher station densities to meet climate-monitoring goals are identified.

Two network densities, resulting from 0.10°C per decade and 1.75 cm per decade temperature and precipitation monitoring goals, were superimposed. The maximum number of stations required to meet either monitoring goal determines the resulting network. The result provides a conservative estimate for national climatic change networks (Fig. 5). Although these techniques make an initial assumption regarding uniformity, the resulting subnetworks are invariant with respect to baseline network densities.

7. REFERENCES

DeGaetano, A. T., 2001: Spatial grouping of United States climate stations using a hybrid clustering approach. *Int. J. Climatology*, **21,** 791-807.

Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, and D. P. Bowman, 1996: *United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data.* ORNL/CDIAC-87 NDP-019/R3. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. [Available from National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161.]

Hubbard, K. G., 1994: Spatial variability of daily weather variables in the high plains of the USA. *Agric. Forest Met.*, **68,** 29-41.

Karl, T. R., and R. W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.*, **79,** 231-241.

Robeson, S. M., and M. J. Janis, 1998: Comparison of temporal and unresolved spatial variability in multiyear time-averages of air temperature. *Climate Research*, **10,** 15-26.

Willmott, C. J., S. M. Robeson, and M. J. Janis, 1996: Comparison of approaches for estimating time-averaged precipitation using data from the USA. *Int. J. Climatol.,* **16,** 1103-1115.