Station Density Strategy for Monitoring Long-Term Climatic Change in the Contiguous United States

MICHAEL J. JANIS
Southeast Regional Climate Center, South Carolina Department of Natural Resources, Columbia, South Carolina

KENNETH G. HUBBARD
High Plains Regional Climate Center, University of Nebraska at Lincoln, Lincoln, Nebraska

KELLY T. REDMOND
Western Regional Climate Center, Desert Research Institute, Reno, Nevada

(Manuscript received 12 September 2002, in final form 30 June 2003)

ABSTRACT

The National Oceanic and Atmospheric Administration is establishing the U.S. Climate Reference Network (CRN) to improve the capacity for observing climatic change and variability. A goal of this network is to provide homogeneous observations of temperature and precipitation from benchmark stations that can be coupled with historical observations for detection and attribution of climatic change. The purpose of this study was to estimate the number and distribution of U.S. CRN observing sites. The analysis was conducted by forming hypothetical networks from representative subsamples of stations in an existing higher-density baseline network. The objective was to have the differences between the annual temperature and precipitation trends computed from reduced-size networks and the full-size networks not greater than predetermined error limits. This analysis was performed on a grid cell basis to incorporate the expectation that a greater station density would be required to achieve the monitoring goals in areas with greater spatial gradients in trends. Monte Carlo resampling techniques were applied to stations within 2.5° latitude × 3.5° longitude grid cells to successively lower the resolution compared to that in the reference or baseline network. Differences between 30-yr trends from lower-resolution networks and full-resolution networks were generated for each grid cell. Grid cell densities were determined separately for temperature and precipitation trends. In practice densities can be derived for any parameter and monitoring goal. A network of 327 stations for the contiguous United States satisfied a combined temperature-trend goal of 0.10°C decade⁻¹ and a precipitation-trend goal of 2.0% of median precipitation per decade.

1. Introduction

The National Oceanic and Atmospheric Administration, in order to monitor temporal variability in climate across the United States, is deploying a new Climate Reference Network (CRN; Heim 2001). The initial proposal for the U.S. CRN suggested 250 stations would be deployed across the contiguous United States. This station estimate was inferred from a methodological examination of twentieth-century U.S. precipitation trends (Karl and Knight 1998). They estimated that a network of 182 stations reproduced the 1910–96 trends in annual climate-division precipitation. The purpose of our study is to determine the spatial density and total number of stations required for the U.S. CRN and to provide a relationship between station density and specified criteria.

In this study, we will objectively estimate the number of stations that reproduce, within predetermined limits, observed annual temperature and precipitation trends across the contiguous United States. Previous studies have also examined the role of network density in capturing temporal variability. For example, Hubbard (1994) found that one station every 60 km in relatively simple terrain was adequate to capture 90% of the variability in daily temperature; resolution needed to capture daily precipitation variability was an order of magnitude higher (5 km). Based on an 814-station subset of the U.S. Historical Climatology Network (HCN), DeGaetano (2001) found that 322 station clusters represented the spatial variability of seasonal precipitation across the contiguous United States. Although he found 96 single-station clusters, an average spatial separation between cluster centers was approximately 179 km. Comparatively, a 182-station network for the contiguous...
United States may have a uniform spacing of approximately one station every 239 km while the 1221-station network found in the U.S. HCN has on average one station every 92 km. (Easterling et al. 1996).

2. Air temperature and precipitation data for the United States

To provide initial discrete representations of climatic variability, temperature and precipitation data were drawn from the 1971–2000 climatography of the U.S. sequential database (CLIM81; NCDC 2002). CLIM81 contained 5313 stations with monthly average temperature that underwent adjustments for observation-time differences and 7507 stations with monthly total precipitation. We screened CLIM81 metadata for flags associated with estimated or adjusted data. Flags may arise from missing or anomalous values and typically indicate adjustments made to the data were from neighboring stations. Since multiple adjustments can indicate poorly performing stations, we culled from the study those stations containing more than three years with more than three flags each. The resulting networks, with adjusted data, consisted of 3642 temperature stations (Fig. 1) and 5156 precipitation stations (Fig. 2) with an average spacing of 40–80 km. Sequential temperature and precipitation data are transformed into 30-yr annual temperature anomalies and percent-of-median annual total precipitation based on the 1971–2000 reference period.

3. Methodology

To examine the rate at which representativeness deteriorates, we systematically decreased network resolutions by selectively removing stations from an initial full-density network (e.g., CLIM81). Each step involved generating measures of similarity between the new networks of lower spatial resolution and the full-density network (e.g., Willmott et al. 1996; Robeson and Janis 1998). An ideal density for climate monitoring networks would be the number of stations that reproduced observed trends in the baseline networks to within predetermined thresholds. 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 decade⁻¹.” Although we concentrated on a definition of climatic behavior that emphasized trends, our technique can be fully generalized to other measures of climatic behavior.

Instead of performing a global spatial sampling of all stations in CLIM81, we stratified the weather stations into 2.5° latitude × 3.5° longitude grid cells. This allowed the density to vary from region to region although the goal was the same from region to region. These local spatial analyses determined where higher or lower network densities were needed to satisfy monitoring goals. These analyses also help overcome spatially biased network arrangements that may result from random sam-
pling across the entire network. Stratified in this fashion, the contiguous United States contained 115 grid cells for precipitation and 114 for temperature (Figs. 1 and 2). Since 94% of these grid cells contained more than 10 stations, the analyses began with a critical number of stations in each grid cell. Assuming a minimum density of one station per grid cell, the total number of grid cells was exactly the minimum number of stations that would be recommended by our method.

Following the methods described in von Storch and Zwiers (1999), we applied a Monte Carlo resampling procedure to each grid cell. The specific steps are enumerated below.

1) Randomly sample, without replacement, an \( N_s \)-station lower spatial resolution (LSR) network (\( 1 \leq N_s \leq N - 1 \)), where \( N \) is the total number of stations in a grid cell and \( N_s \) is the number of stations in the grid cell for the subset.

2) Generate ensemble time series of spatially averaged temperature or precipitation from an \( N_s \)-station network. For example, an ensemble time series from the first realization of a randomly drawn (without replacement) 20-station network is computed as

\[
\hat{T}_1 = \frac{1}{20} \sum_{j=1}^{20} T_j
\]

where \( \hat{T}_1 \) is an ensemble time series derived from 20 randomly selected stations and \( T_j \) is a time series from the \( j \)th station.

3) Compute trends for each realization of an \( N_s \)-station ensemble time series. Trends are expressed as degrees Celsius per decade or percent of median annual precipitation per decade. Trends were calculated over the 30 yr contained in CLIM81 (i.e., 1971–2000) and contained more confidence than trends computed over a decade.

4) Repeat steps 1–3, choosing from among all \( N \) stations, 100 times to generate multiple realizations for an \( N_s \)-station network. This addresses the effect of multiple network configurations, reduces the effect of poorly distributed LSR networks, and reduces the influence of undetected inhomogeneous station records. Mean absolute error (MAE) for trends is computed as

\[
\text{MAE} = \left| \frac{1}{100} \sum_{k=1}^{100} \frac{\Delta T}{\Delta t} - \frac{\Delta \hat{T}_k}{\Delta t} \right|
\]

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

5) Repeat steps 1–4 for all possible LSR network sizes (\( N_s = 1, 2, \ldots, N - 1 \)).

6) Perform regression between MAE, and \( N_s \) using a fourth-order polynomial to provide a better fit than lower-order polynomials, especially for small \( N_s \):

\[
N_s = a_0 + a_1 \text{MAE}_1 + \cdots + a_4 \text{MAE}_4.
\]
MAE, decreases as the within-cell density approaches the baseline network density (Fig. 3). MAE decreases rapidly as \( N_e \) increases across small numbers of stations and decreases more slowly as \( N_e \) approaches \( N \) (Fig. 3). This suggests large improvements in network performance are possible with, in this case, increases from one to three stations per grid cell; network designers need to consider this potential benefit in their optimization between resources and scientific goals.

7) Set monitoring goals then solve polynomial model for \( N_e \); in practice these goals will depend on the threshold of detection desired. Inside every grid cell, network densities required to meet the monitoring goals are determined. A monitoring goal for annual air temperature trends of 0.05°C decade\(^{-1}\) applied to a New Mexico grid cell, for example, would lead to a recommendation of eight stations for that location (Fig. 3).

8) Add the stations from all grid cells to obtain the total number of stations to meet the stated goal across the contiguous United States.

4. Resulting station densities

Resulting grid cell densities indicate how many stations are required to capture the spatial variability in the observed network. To identify regions of the country that were most sensitive to network configuration and network density, we analyzed \( N_e \) for several temperature and precipitation monitoring goals. Temperature monitoring goals were defined in comparison to previously observed global temperature trends (Houghton et al. 2001; Karl et al. 1994). Precipitation monitoring goals were based on observed annual trends within ±5% decade\(^{-1}\) across the contiguous United States (Karl and Knight 1998).

a. Annual air temperature trends

We examined networks resulting from four temperature-trend thresholds and a simple one-station-per-grid cell network. Table 1 shows that as goals became more stringent the total number of stations increased, and the range of estimated grid cell trend decreased. A network meeting a goal of 0.05°C decade\(^{-1}\) consisted of 622 stations with an average of 5.5 stations per grid cell and an average station separation of 149 km (Table 1). Figure 4 shows that density estimates for this goal varied across the United States. In the western United States (west of the 100°W) 310 stations or an average of 6.2 stations per grid cell were required. In the eastern United States 311 stations or 4.9 stations per grid cell were required, but higher densities were concentrated in the mid-Atlantic and southeast. Approximately 55% of grid cells required between 4 and 6 stations per cell to meet this goal and 3 grid cells required 10 or more stations (Fig. 5a). Three grid cells requiring one or two stations were located along coastlines. Grid cells with comparatively lower-density estimates, such as northwestern Arizona, had fewer stations in the baseline network. In these cases, density was likely underestimated and contained less certainty. If the stations were to be uniformly distributed across the grid cells, an average spatial separation of 100–175 km would meet this goal in approximately 85% of the grid cells (Fig. 5b). Selection of station locations in each box depends on additional knowledge of climatic behavior outside the scope of this study: the within-grid spatial variability of climate should be addressed before network implementation.

A network meeting a goal of 0.10°C decade\(^{-1}\) consisted of 233 stations with an average station separation of 242 km and an average of 2.1 stations per grid cell (Table 1). Compared with a 0.05°C decade\(^{-1}\) threshold, densities decreased such that 30% of grid cells required only one station and nearly 100% of grid cells required fewer than four stations (Fig. 5c). Assuming a uniform distribution of stations within grid cells, we found an average spatial separation of 175–250 km would meet this threshold for approximately 60% of the grid cells (Fig. 5d). Figure 6 shows that density was uniformly decreased from more stringent goals. For this goal, lower densities (one to two stations per grid cell) were found from New England to the Great Lakes, in parts of the northwest, and across the central United States.

We took three different approaches to interpreting the performance of various LSR networks. The first of these involved mapping the absolute differences between av-
Fig. 4. Grid cell densities of 622 stations satisfying an annual temperature-trend monitoring goal of MAE $< 0.05^\circ$C decade$^{-1}$. Graduated circles represent the number of stations within each grid cell.

Fig. 5. For an LSR network meeting MAE $< 0.05^\circ$C decade$^{-1}$ monitoring goal, frequency distributions are (a) number of stations per grid cell and (b) grid cell station separation. For an LSR network meeting MAE $< 0.10^\circ$C decade$^{-1}$ goal, frequency distributions are (c) number of stations per grid cell and (d) grid cell station separation.
verage temperature trends measured by an \( N \)-station network and those measured by the full-resolution network. The second analysis was actually a specific example of how trend estimates improve as the size of the subsets approaches the full resolution. The third analysis compared LSR networks to a simple one station per grid cell network.

Figure 7 shows that increased network density, the result of more precise monitoring goals, decreased the absolute differences between average LSR network trends and baseline network trends. For a 0.05°C decade \(^{-1}\) threshold, only four cells stand out from the distribution of small trend differences (Fig. 7a). Those along the border of the United States have four or fewer original baseline stations and three or fewer estimated subset stations; thus, larger trend differences may be expected. The southern Nevada grid cell had 11 original stations and 7 estimated stations. Larger trend differences may have been caused by higher spatial variability or data inhomogeneities. For a 0.10°C decade \(^{-1}\) threshold, larger trend differences were found in parts of the West, Midwest, and along the border (Fig. 7b). For interior grid cells, where larger differences suggest more spatial variability, network designers can pay special attention to capturing the observed variability. The largest differences from baseline trend estimates were found for the simple, and as it turns out undesirable, one-station-per-grid cell or equally spaced network (Fig. 7c). Relative to one station per grid cell, estimates of temperature trends were improved as LSR networks met more stringent monitoring goals.

To demonstrate how trend estimates improved as more stringent monitoring goals were applied, we conducted a case study of a single grid cell over north-central New Mexico. The minimum, maximum, and standard deviation of trends were derived from 100 Monte Carlo realizations for each network size. Figure 8 shows that 8 stations per grid cell produced trends within the range of 0.138°C–0.391°C decade \(^{-1}\) with a standard deviation of 0.005°C decade \(^{-1}\) while 16 stations per grid cell produced trends within the range of 0.182°C–0.335°C decade \(^{-1}\) with a standard deviation of 0.003°C decade \(^{-1}\). This example shows that the range of estimated grid cell trends decreased and became closer to the observed grid cell trend as the number of stations in a grid cell increased. In practice, a network designer should apply this performance measure to estimate the observed variability in trend estimates within each grid cell.

In the third analysis, we examined the maximum difference and standard deviation between resulting LSR networks and the full-resolution network. The maximum difference of trend estimates within a grid cell for any subnetwork size is the maximum absolute difference between trends calculated from the subnetwork and the trend calculated from the full-resolution network. This is not to be confused with the mean absolute errors that are average differences and used as thresholds for mon-
old, trend estimates are within 0.40°C decade\(^{-1}\) across only 30% of 114 grid cells. Networks created on the basis of less stringent monitoring goals will, therefore, have less precise trend estimate over some areas. The standard deviations and maximum differences of Monte Carlo trend estimates illustrate that applying more stringent monitoring goals improved the LSR network performance, especially relative to a simple one-station-per-grid cell network (Fig. 9). Selection of LSR network size should be based on relative improvements and acceptable thresholds of climatic variability.

b. Annual precipitation trends

We examined networks resulting from four precipitation-trend thresholds and a simple one-station-per-grid

![Figure 9](image-url)
cell network. A network meeting a goal of 1.5% of median annual precipitation per decade consisted of 490 stations with an average of 4.3 stations per grid cell and an average station separation of 173 km (Table 2). Figure 10a shows that 40% of the grid cells required three or fewer stations per cell to meet this threshold. Assuming a uniform distribution of stations in each grid cell, approximately 70% of grid cells required a 125–200-km spatial separation between stations to meet this monitoring goal (Fig. 10b). Figure 11 shows that higher station densities were required west of the 100°W, with the highest station densities occurring in the southwestern United States. The western United States required 275 stations or an average of 5.4 stations per grid cell while the eastern United States required 215 stations or an average of 3.4 stations per grid cell.

Table 2 shows that a network meeting a 2.0% of median annual precipitation per decade threshold consisted of 293 stations with an average of 2.5 stations per grid cell and an average station separation of 225 km. Thirty percent of grid cells required only one station to meet this monitoring goal (Fig. 10c). Assuming a uniform distribution of stations within grid cells, an average station separation was 150–225 km for approximately 50% of the grid cells (Fig. 10d). Though this average separation is similar to the 1.5% precipitation trend threshold, we found that higher spatial separation (e.g., 250 km or greater) was sufficient across several grid cells for the 2.0% precipitation trend threshold. Figure 12 displays higher network densities in the western United States (168 stations or an average of 3.3 stations per grid cell) and lower network densities in the eastern United States (125 stations or an average of 1.9 stations per grid). As precipitation-trend thresholds were relaxed, overall station density decreased but the regions with relatively higher density remained the same (cf. Figs. 11 and 12).

Figure 13 shows the absolute differences between average trends estimated by N-station networks and the baseline network. Absolute differences increased as monitoring goals were relaxed and reached a maximum with the one-station-per-grid cell network. Although network performance generally improved as more stringent monitoring goals were applied, the maps indicate the greatest improvements due to increased network density were found in the western and southwestern

<table>
<thead>
<tr>
<th>Precipitation trend (% median decade⁻¹)</th>
<th>No. of stations</th>
<th>Avg no. of stations per grid cell</th>
<th>Avg separation between stations (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>899</td>
<td>7.8</td>
<td>128</td>
</tr>
<tr>
<td>1.5</td>
<td>490</td>
<td>4.3</td>
<td>173</td>
</tr>
<tr>
<td>2.0</td>
<td>293</td>
<td>2.5</td>
<td>225</td>
</tr>
<tr>
<td>2.5</td>
<td>218</td>
<td>1.9</td>
<td>259</td>
</tr>
<tr>
<td>One station per grid cell</td>
<td>115</td>
<td>1.0</td>
<td>326</td>
</tr>
</tbody>
</table>
FIG. 11. Grid cell densities of 490-station LSR network satisfying an annual precipitation-trend monitoring goal of MAE < 1.5% median precipitation per decade. Graduated circles represent the number of stations within each grid cell.

FIG. 12. Grid cell densities of 293-station LSR network satisfying an annual precipitation-trend monitoring goal of MAE < 2.0% median precipitation per decade. Graduated circles represent the number of stations within each grid cell.
variability in precipitation trends, especially in those areas described above.

As with temperature trends, we conducted a case study of a single grid cell over north-central New Mexico to demonstrate how trend estimates improved as more stringent monitoring goals were applied. The minimum, maximum, and standard deviation of trends were derived from 100 Monte Carlo realizations for each LSR network size. Figure 14 shows that the range and standard deviation of trend estimates decreased rapidly as the number of stations increased from 1 to 10. A 10-station-per-grid cell network produced trends within the range 5.99% and 11.15% median precipitation per decade with a standard deviation of 0.10% median precipitation per decade. The standard deviation and range of estimated grid cell trends was greater for fewer numbers of stations. Generally, the standard deviation of trend estimates approached zero and the range of trend estimates approached the observed grid cell trend as the number of stations in a grid cell approached the full-network resolution. Although these relationships were typical for most grid cells, the variability in estimated trends should be examined as part of selecting station locations in each grid cell.

Figure 15 shows the cumulative frequency across all grid cells of the maximum difference between LSR network trends and a baseline trend and the cumulative frequency of the standard deviation of trends. These values were derived from the 100 Monte Carlo realizations for each network size in each grid cell. For the 1.5% median annual precipitation per decade threshold, the maximum differences in trend estimates were within 11% of median annual precipitation for all 115 grid cells (Fig. 15a). For the 2.0% decade⁻¹ threshold, only 65% of grid cells were within this range of trend estimates. Application of either monitoring goal beyond a simple 115-station grid substantially decreased the range and standard deviation of trend estimates across the network.
5. Summary and discussion

This work recommends a station density for the U.S. CRN. We examined spatial density based on measured precipitation and temperature trends from a network of existing weather stations. Our method indicated how many stations were required to reproduce the variability in an existing station network. If these stations were inhomogeneous or if the number of stations in each regional sample was not sufficient to capture the actual variability, then our method indicates the number of stations required to meet those observed trends even though they may not be representative of the actual trends. Although we excluded potentially inhomogeneous stations from the analysis and minimized their impact by resampling, their likely effect would be to overestimate grid cell densities. Since our method indicated how many stations were required to replicate the trends in an existing network, undersampled locations in the baseline network and especially Nevada should undergo separate and more detailed examination.

The assumed minimum density was one station per 2.5° latitude × 3.5° longitude grid cell. Our grid-based analysis provided regional estimates of network density that satisfied predetermined monitoring goals. We used a stratified local resampling strategy within each grid cell to build information on how trend estimates diverged with decreasing network density. Regions of the country that required higher station densities to meet climate monitoring goals were identified. Although these techniques made use of a baseline network with different densities among grid cells, Fig. 16 shows that the resulting LSR networks were not dependent on the baseline network densities. Capturing trends in the observed networks was not affected by the fact that, initially, the same numbers of stations were not found within each grid cell.

To provide a minimum density estimate for the U.S. CRN, we superimposed network densities resulting from 0.10°C decade\(^{-1}\) and 2.0% of median annual precipitation per decade monitoring goals. Because the U.S. CRN will measure precipitation and temperature at each station, the larger number of stations within a grid cell for either variable is the resulting recommendation for a grid cell in a combined network. We found this network consisted of 327 stations with an average of 2.9 stations per grid cell. Figure 17 shows the western United States (west of the 100°W) required 168 stations or an average of 3.4 stations per grid cell while the eastern United States required 159 stations or 2.5 stations per grid cell. Comparing the combined network density map (Fig. 17) to the maps for individual monitoring goals (Figs. 6 and 12) shows that precipitation monitoring drove higher densities in the west while densities in the east were slightly higher for temperature monitoring. The primary guidance for the spatial configuration of the U.S. CRN is, however, provided by a triangular grid that captures 98% of the national annual climate signal for temperature and 95% of the national annual climate signal for precipitation (R. Vose 2002, personal communication). This guidance recommends one station per grid node uniformly distributed across the contiguous United States. Our research supplements this guidance by identifying regions of the contiguous United States where climate monitoring could be improved by localized increases in the spatial density of observations.

The implication of superimposing network densities and deriving a single grid cell density for temperature
and precipitation is that temperature monitoring goals may be exceeded in the western United States and precipitation monitoring goals may be exceeded in the eastern United States. An optimization whereby temperature and precipitation thresholds would be applied simultaneously would likely lead to fewer total stations. Instead of increasing monitoring potential, optimization may increase one variable’s threshold, but at the expense of the other variable. We feel that combining density estimates after independent calculation provides a more robust density estimate, especially with inevitable station closures.

Acknowledgments. We thank Richard Heim, Russ Vose, John Jensen, and Michael Helfert of the National Climatic Data Center for their various contributions to this work. We acknowledge the thoughtful and constructive comments of three anonymous reviewers. This work was partially supported under the NOAA Cooperative Agreement NA17RJ1222.

REFERENCES


