The Quality Control of Long-Term Climatological Data Using Objective Data Analysis

JON K. EISCHEID

Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado

C. BRUCE BAKER AND THOMAS R. KARL

National Climatic Data Center, NOAA, Asheville, North Carolina

HENRY F. DIAZ

Climate Diagnostics Center, NOAA/ERL, Boulder, Colorado

(Manuscript received 30 August 1994, in final form 15 June 1995)

ABSTRACT

One of the major concerns with detecting global climate change is the quality of the data. Climate data are extremely sensitive to errant values and outliers. Prior to analysis of these time series, it is important to remove outliers in a methodical manner.

This study provides statistically derived bounds for the uncertainty associated with surface temperature and precipitation measurements and yields a baseline dataset for validation of climate models as well as for a variety of other climatological uses. A two-step procedure using objective analysis was used to identify outliers. The first step was a temporal check that determines if a particular monthly value is consistent with other monthly values for the same station. The second step utilizes six different spatial interpolation techniques to estimate each monthly time series. Each of the methods is ranked according to its respective correlation coefficients with the actual time series, and the technique with the highest correlation coefficient is chosen as the best estimator. For both temperature and precipitation, a multiple regression scheme was found to be the best estimator for the majority of records. Results from the two steps are merged, and a combined set of quality control flags are generated.

1. Introduction

A major concern regarding the detection of global climate change has to do with data quality. Climatic trends are very sensitive to errant values and outliers arising from a variety of sources. The general sources of error in monthly mean and temperature and total precipitation are well known and have been addressed by many authors (e.g., Karl et al. 1989). In general our concern here is in the development of an automated objective analysis algorithm to flag possible errors in the database. Hence, prior to the analysis of climatic time series, it is important to remove data outliers in a methodical manner. Currently, very few long-term climate datasets with global coverage being used to detect climate change can provide reproducible results starting with originally reported values. In this paper an objective quality control analysis scheme examines global climate data for outliers temporally and spatially.

The raw data used in the implementation of the quality control scheme is a subset of the Global Historical Climate Network (GHCN) (Vose et al. 1992) and consists of mean monthly temperature and precipitation. Each station selected from the GHCN was required to have at least 20 years of record within the 1951–80 period. This criteria resulted in 3467 temperature stations (Fig. 1) and 5899 precipitation stations (Fig. 2).

A temporal check and a spatial check for outliers are described in the next section. When combined, these two measures of quality control provide a comprehensive set of uncertainty flags to determine the validity of a particular monthly value. This procedure is being implemented to enable us to maintain the basic GHCN dataset in near real time (for monthly data, soon after the end of the data month in question), in order to continue our efforts to monitor climatic variations on regional to global scales.

Corresponding author address: Dr. Jon K. Eischeid, CIRES, University of Colorado, Boulder, CO 80309.

2. Temporal tests

The temporal check for outliers for a particular station is based on the premise that an individual monthly value should be "similar" (in a statistical sampling sense) to values for the same month for other years. In order to make as few assumptions as possible re-

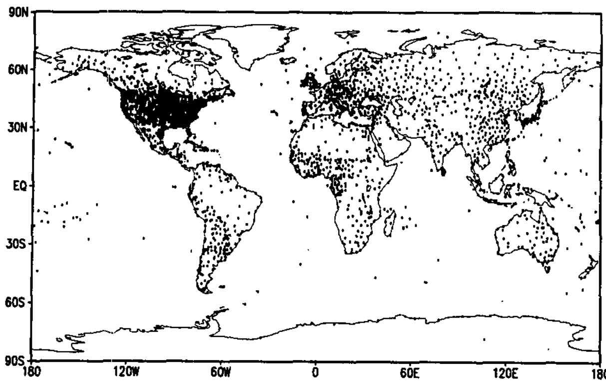


FIG. 1. Global distribution of mean monthly temperature measurements (3467 stations).

garding the wide range of data to be checked, outliers were identified utilizing the sample distribution of each calendar month separately for each station. Extreme values are flagged based on limits determined from a multiple of the interquartile range (IR, 75th percentile minus the 25th percentile) calculated for each station/month. This procedure is common in exploratory data analysis procedures. An outlier is flagged when

$$X_i - q50 > f\mathrm{IR}, \quad (1)$$

where X_i is the monthly mean of year i, $q50$ is the median (or the 50th percentile), and f is the multiplication factor.

A typical value of f used to identify extreme outliers in this context is 3 (Velleman and Hoaglin 1981). Rather than use a fixed multiplier, a characteristic value for each dataset (temperature and precipitation) was determined by examining the relationship between the percentage of data that would be flagged for a range of multipliers of the interquartile range. Figure 3 is an example showing the percentage of all station months flagged for various values of f over a latitudinal range for the month of January for both temperature and precipitation. Similar curves were examined for each month and each variable. The dependence of latitude on the selection of f was deemed negligible. The cutoff for outliers was determined by calculating the slope for each curve and choosing the multiple of the interquartile range where the slope was sufficiently near zero. A value of 2.75 was used for temperature and 4.00 for precipitation, as depicted in Fig. 3. Figure 4 illustrates the percentage of data flagged for each calendar month for both precipitation and temperature using these sets of values. Thus, the temporal check is designed to determine whether or not the month in question is consistent with the sample population of other such months for that station. Typically the total number of station months identified as outliers is on the order of 0.1% for temperature and 0.5% for precipitation. There is a relative increase in the number of values rejected for northern summer precipitation, as illustrated in Fig. 4. Since the bulk of these stations is located in the Northern Hemisphere, the increase is probably due to the greater likelihood of extreme rainfall totals during the summer months associated with severe convective storms.

3. Spatial quality control—Interpolation methods

The spatial quality control of monthly temperature and precipitation data includes the use of nearby simultaneous values to calculate an estimated value at a target station over the period of time for which adequate data are available. In the context of this paper, a value is taken to be an actual monthly record from which its long-term mean has been subtracted. Threshold limits are computed based on the distribution of the differences between the observed and estimated values. Any particular value is determined to be an outlier based on the magnitude of this difference compared with the previously established threshold. The efficiency, or accuracy, of the estimates over a long period of time gives guidance in determination of appropriate threshold limits used to assess the quality of the monthly data value in question.

There are numerous spatial interpolation methods available for point estimation with irregularly spaced data. Typically, the choice of methodology is dependent on several factors: the meteorological variable under consideration; the geographical area; the spatial distribution of surrounding observations; and the month/season for which the target station is to be estimated (Schlatter 1975; Bennet et al. 1984; Thiebaux and Pedder 1987). Since estimates are required for each month separately over a variety of terrain with a differing number of available surrounding observations, we have chosen several different methods for testing. This section will focus on six methods of spatial interpolation. These are defined as the 1) normal ratio method (NR), 2) simple inverse distance weighting (IDW), 3) optimal interpolation (OI), 4) multiple regression using the least absolute deviation criteria (MLAD), 5) single best estimator (SBE), and 6) me-

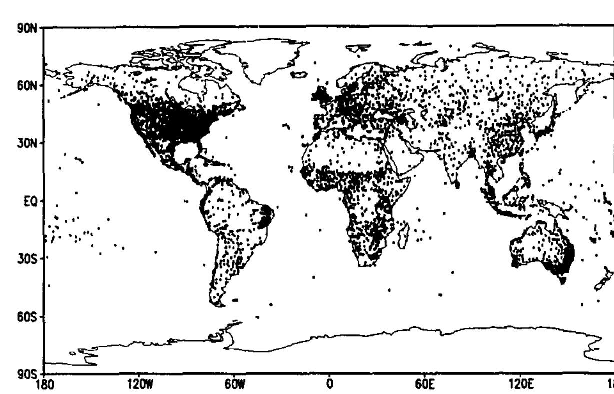


FIG. 2. Global distribution of total monthly precipitation measurements (5899 stations).

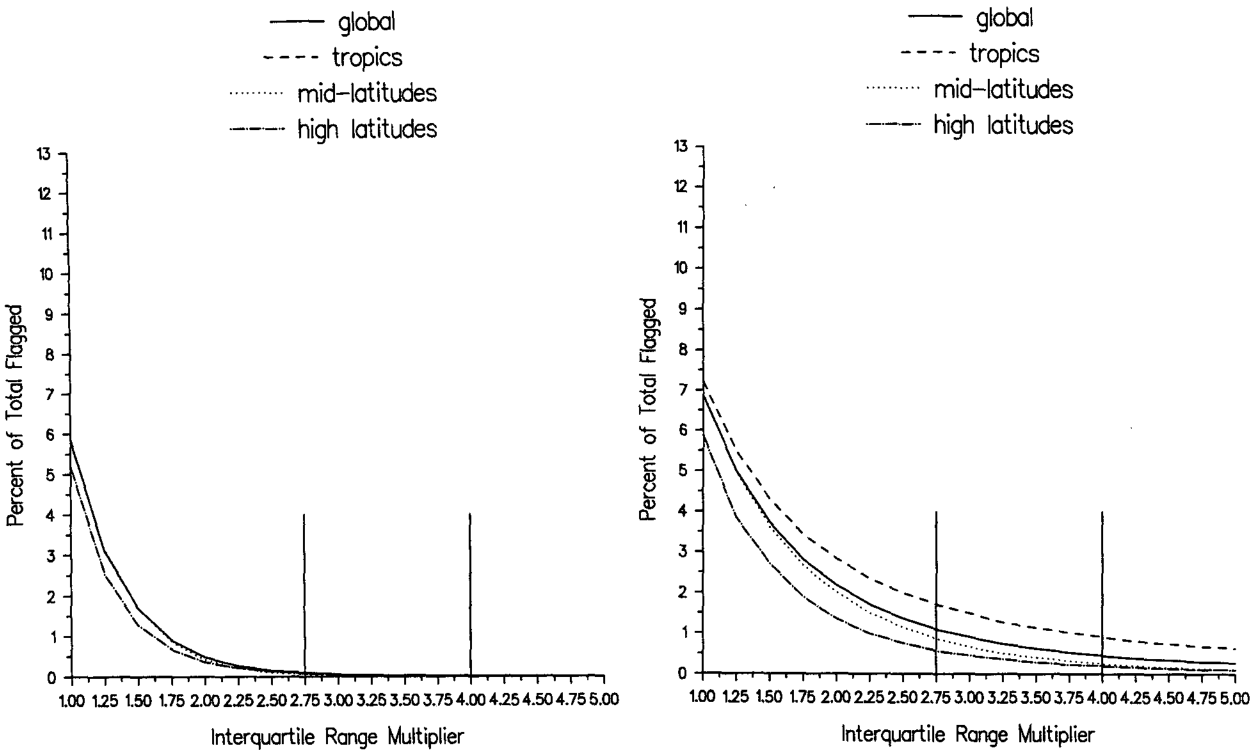


FIG. 3. (a) Relationship between the percentage of temperature station months that would be flagged over selected latitudinal ranges given different multipliers of the interquartile range. The two vertical lines represent the threshold for temperature (2.75) and precipitation (4.00). (b) Same as (a) but for precipitation.

dian (MED) of the previous five methods. Each of the six methods is compared for each month for each station, and the one with the highest correlation to the target station was chosen for the spatial consistency check.

In any spatial interpolation scheme the selection and quantity of surrounding stations are critically important to the results of the interpolations. Problems arise when using climatological data because of missing values and the varying availability of stations through time. In order to determine which stations are to be used, surrounding stations are preselected based on their relationship with the target station. A first difference series ($X_{t+1} - X_t$) is computed for each monthly time series from the raw dataset that has not been quality controlled. This removes any trends from the time series and helps minimize unwanted spatial inhomogeneities that may affect the correlation between the target station and its neighbors. The 10 closest stations are identified for each target station and ranked by the value of the correlation coefficient between the candidate station and its neighbors. The stations with the largest positive correlation coefficients, where the minimum criterion is an r of at least 0.35, are subsequently used in the estimation procedures. A minimum of one station is needed to compute the estimate at the target station with a maximum of four. Tests have shown that inclusion of more than four stations does not significantly improve the interpolation and may in fact degrade the estimate. The number (never greater than four) of neighboring stations meeting the criteria is not fixed in time. It varies depending on available station data for the year/month in question. As such, the in-

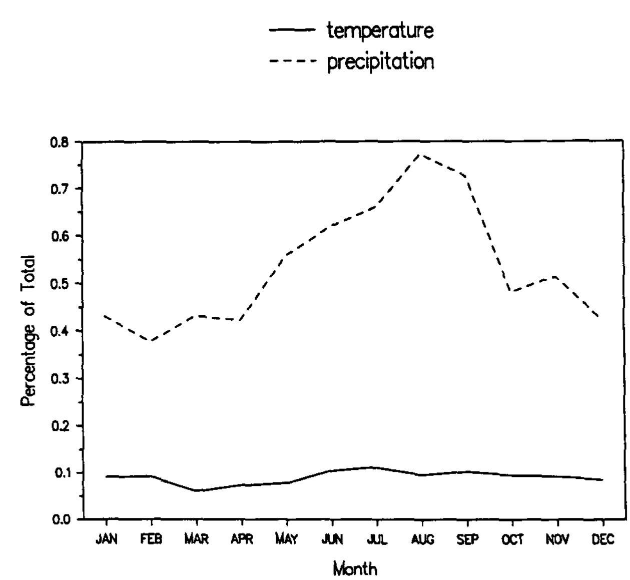
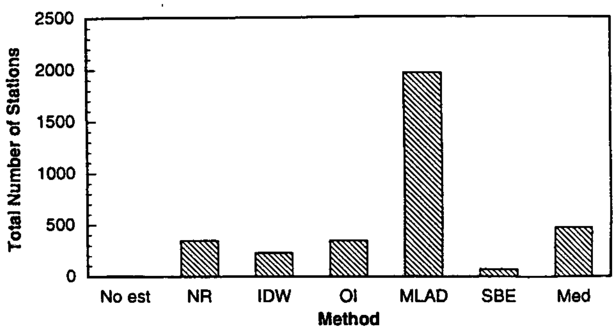
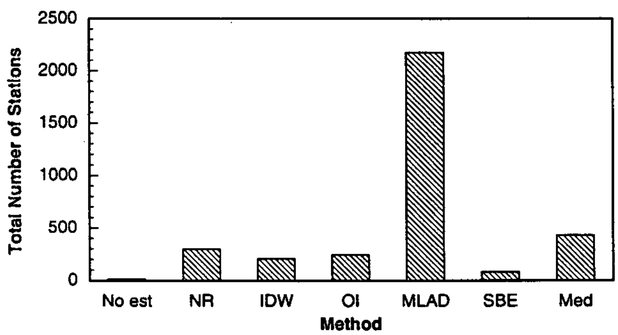


FIG. 4. The percentage of data flagged for each dataset.



	NR	IDW	OI	MLAD	SBE	Med
$\bar{x}$	.94	.92	.95	.90	.92	.93
σ	.12	.13	.09	.12	.11	.10
cv	12.8%	14.1%	9.5%	13.3%	12%	10.8%



	NR	IDW	OI	MLAD	SBE	Med
$\bar{x}$	.85	.86	.89	.85	.84	.88
σ	.20	.13	.11	.14	.17	.12
cv	23.5%	15%	12.4%	16.5%	20.2%	13.6%

FIG. 5. (a) Frequency distribution of the number of times a particular interpolation method was selected for temperature in January. The mean correlation, standard deviation, and coefficient of variation is shown below for each interpolation method. (b) Same as (a) but for July temperatures.

terpolation models may also change in time. Moreover, the surrounding stations that may be optimal for a particular calendar month (e.g., January) may not be optimal for a different month (e.g., July); thus, the station selection procedures are computed for each calendar month separately. The selected stations are identical, however, for each of the six interpolation methods. A brief description of each method is given below.

a. Normal ratio method

The normal ratio (NR) method of spatial interpolation was first proposed by Paulhus and Kohler (1952). A modified version is used in the present analysis described by Young (1992). Weights for the surrounding stations used in the estimation algorithm are found according to

$$W_i = \frac{r_i^2(n_i - 2)}{1 - r_i^2}, \qquad (2)$$

where r is the correlation coefficient for each monthly time series between the target station and the ith surrounding station, n is the number of points used to derive the correlation coefficient, and W is the resultant weight.

b. Inverse distance method

The inverse distance method (IDW) is a simple distance-weighted "area average" estimate of the value at the target station. The assumption here is that surrounding stations are related to the target station by their proximity to the target station. This procedure is given by

$$\hat{Z} = \frac{\sum_{i=1}^{n} W_i Z_i}{\sum_{i=1}^{n} W_i}, \qquad (3)$$

where Z_i is the particular monthly anomaly at the ith surrounding station and the weight function W_i is derived from the inverse of the distance from the target station to the ith surrounding station.

c. Optimal interpolation

Early uses of optimal interpolation (OI) in meteorology may be traced back to Gandin (1963). Since that time it has had wide usage in climatology and meteorology. In most applications OI is used to estimate values at a target site, for example, a grid point.

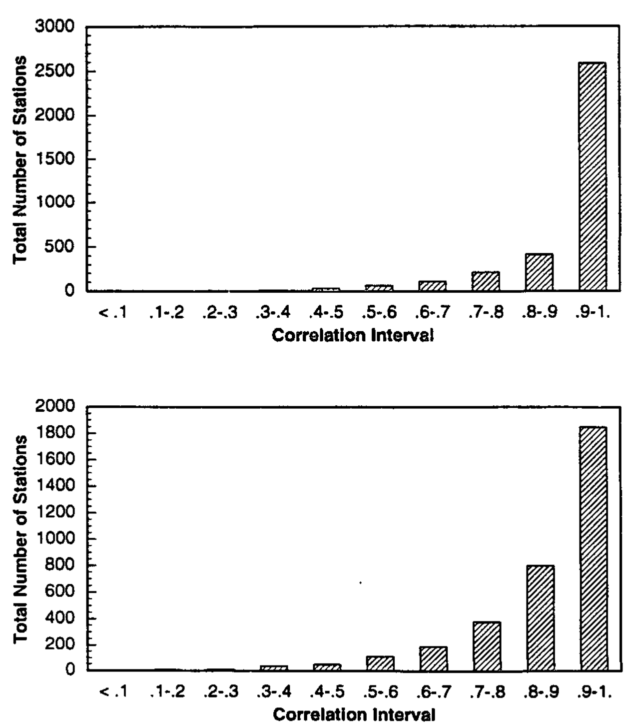


FIG. 6. (a) Frequency distribution of the number of station correlations for the best estimation method for January temperatures. (b) Same as (a) but for July temperatures.

TABLE 1a. The mean error and the standard deviation of the mean estimation error for temperature (°C) by month for each of the interpolation methods.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Mean error						
NR	0.013	0.002	−0.021	−0.042	−0.029	−0.037	0.014	0.007	0.007	−0.025	−0.021	−0.022
IDW	−0.025	−0.036	−0.013	−0.049	0.000	0.016	−0.067	0.027	0.004	−0.036	0.000	0.027
OI	−0.051	−0.021	−0.042	0.001	−0.033	0.000	−0.008	−0.011	−0.005	−0.004	−0.010	−0.048
MLAD	0.002	0.005	−0.002	−0.001	0.000	−3.003	0.000	−0.001	0.003	0.002	0.004	0.008
SBE	−0.018	−0.044	−0.037	−0.035	0.064	−0.029	0.005	0.027	−0.096	−0.063	0.028	−0.092
MED	0.046	−0.021	−0.007	0.000	0.002	−0.001	−0.002	0.018	0.016	−0.019	−0.005	0.003
Mean	−0.005	−0.019	−0.020	−0.021	0.001	−0.009	−0.009	0.011	−0.013	−0.024	−0.001	−0.020
						Standard deviation of mean error						
NR	0.893	0.887	0.800	0.834	0.850	0.805	0.780	0.733	0.773	0.790	0.836	0.846
IDW	1.011	1.011	0.875	0.824	0.793	0.798	0.843	0.774	0.814	0.848	0.879	0.925
OI	1.030	1.010	0.874	0.831	0.794	0.783	0.781	0.785	0.825	0.828	0.897	0.991
MLAD	0.869	0.851	0.796	0.715	0.709	0.720	0.697	0.706	0.689	0.717	0.791	0.840
SBE	0.890	0.974	0.788	0.778	0.704	0.684	0.793	0.759	0.767	0.736	0.793	0.887
MED	0.965	1.003	0.877	0.817	0.805	0.794	0.760	0.792	0.764	0.818	0.904	0.988
Mean	0.943	0.956	0.835	0.800	0.775	0.764	0.775	0.758	0.772	0.790	0.850	0.913

Here we use a univariate OI to estimate values at a known station location.

The univariate OI is a spatial interpolation technique that assigns weights to the observed difference values (observed minus first guess) at the selected neighboring station locations:

$$\hat{Z} = Z_f + \sum_{i=1}^{n} W_i (Z_{oi} - Z_{fi}), \qquad (4)$$

where Z_{oi} and Z_{fi} are the observed and first-guess values at the ith neighbor station, respectively, and Z_f is the first-guess value at the target station location being estimated. The weighting coefficients W_i are determined in an objective manner such that the root-mean square error of the analyzed difference values at each target location is minimized over the spatial domain. The weights are dependent upon the spatial autocorrelations among the surrounding observation values and are typically modeled mathematically as a function of distance separating the neighbors and the target location. Rather than model each spatial domain (the area surrounding each target station) individually, and to compensate for possible anisotropic fields, we use the actual relationships among all observations by directly using the calculated correlation coefficients.

Once the correlation coefficients are known, the weights needed to solve (4) are given by the solution of a system of linear algebraic equations. In matrix form this can be written as

$$\mathbf{W}_i \mathbf{C}_{ir} = \mathbf{G}_r, \qquad (5)$$

where $i = 1, 2, \cdots, N$ (number of surrounding observations), whose coefficients are given by the correlation with selected neighboring station observations ($\mathbf{C}$, $n \times n$) and correlation coefficients from the target station to each of the surrounding stations ($\mathbf{G}$, $1 \times n$ vector). As the empirical correlation coefficients are used, the solution matrix could fail to be positive definite, but in practice this was a rare occurrence.

d. *Multiple regression, least absolute deviations criteria*

The method of multiple regression using the least absolute deviations criteria (MLAD) is a robust version of the general linear least squares estimation. The method of least squares is an effective method when the errors are normally distributed and independent. However, for precipitation data especially, the as-

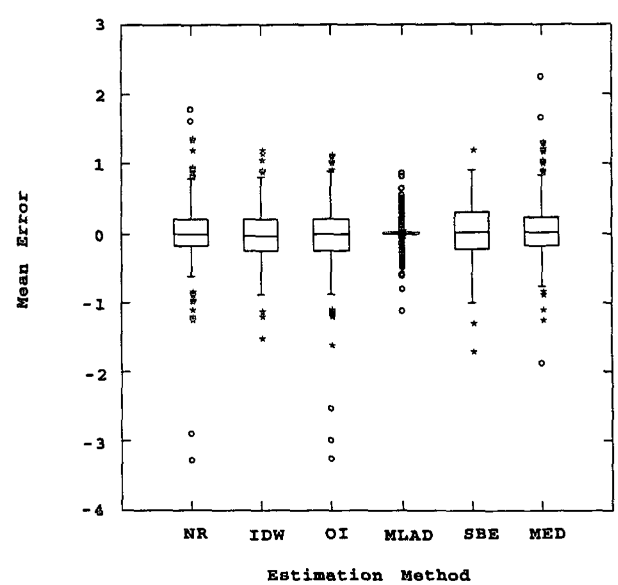
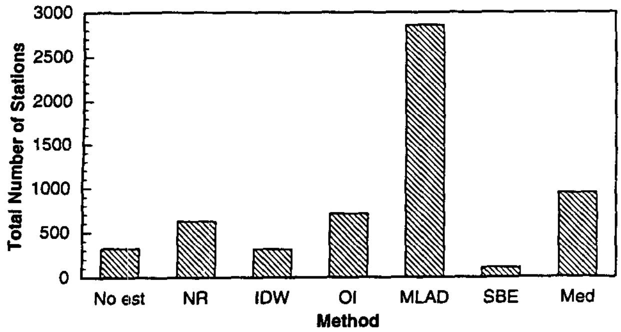
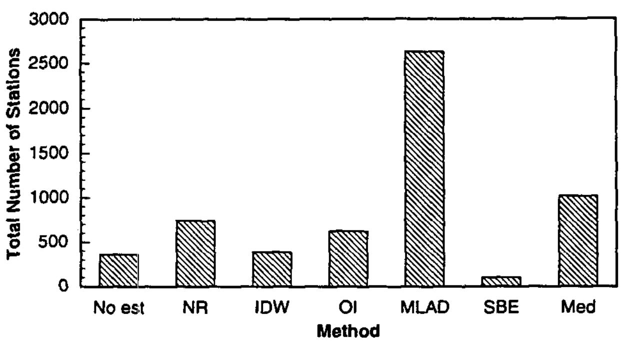


FIG. 7. Distribution of the mean estimation error for temperature (°C).



	NR	IDW	OI	MLAD	SBE	Med
$\bar{x}$	.71	.80	.85	.81	.72	.81
σ	.23	.14	.12	.14	.18	.14
cv	32.9%	17.5%	14.1%	17.3%	25%	17.3%



	NR	IDW	OI	MLAD	SBE	Med
$\bar{x}$	.62	.70	.77	.75	.70	.75
σ	.22	.15	.12	.15	.20	.13
cv	35.5%	21.4%	15.6%	20%	28.6%	17.3%

FIG. 8. (a) Frequency distribution of the number of times a particular interpolation method was selected for precipitation in January. The mean correlation, standard deviation, and coefficient of variation is shown below for each interpolation method. (b) Same as (a) but for July precipitation.

sumption of normality over the wide range of situations can lead to poor estimations. The principal advantage of least absolute deviations is its resistance to outliers and to overemphasis of large tailed distributions (Barrodale and Roberts 1973).

MLAD estimates the unknown parameters in a stochastic model so as to minimize the sum of absolute deviations of the neighboring station observations from the values predicted by the model. Regression coefficients b are calculated so as to minimize

$$\sum_i \left| \sum_j x_{ij} b_j - Y_i \right|, \qquad (6)$$

where x, $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$ denote a set of n measurements on m surrounding stations (independent variables), and y, $i = 1, 2, \cdots, k$ denote the associated measurement on the dependent (target station) value. The linear programming techniques of Barrodale and Roberts (1973) are used to accomplish this task.

e. Single best estimator

The single best estimator (SBE) is simple and analogous to using the closest neighboring station as an estimate for the target station. The target station is estimated using the neighboring station's value that has the highest positive correlation with the target station.

f. Median

The median method (MED) is not a true interpolation model but is simply the median value obtained from the above five estimation methods. By using the median we allow for the estimation formula to change over time, which may yield a better long-term estimate.

4. Comparison of interpolation methods

A sampling replacement procedure is used to test the interpolation methods. Estimates for each station and month are determined and comparisons among the techniques are based on the correlation coefficient R between the actual anomaly at the target station location and each of the corresponding estimated anomalies. The method that exhibits the largest R is considered to be the most representative and is used to determine the levels at which the data are accepted or rejected for all monthly values. In general, stations may, and quite often do, require the use of different models depending on the month in question.

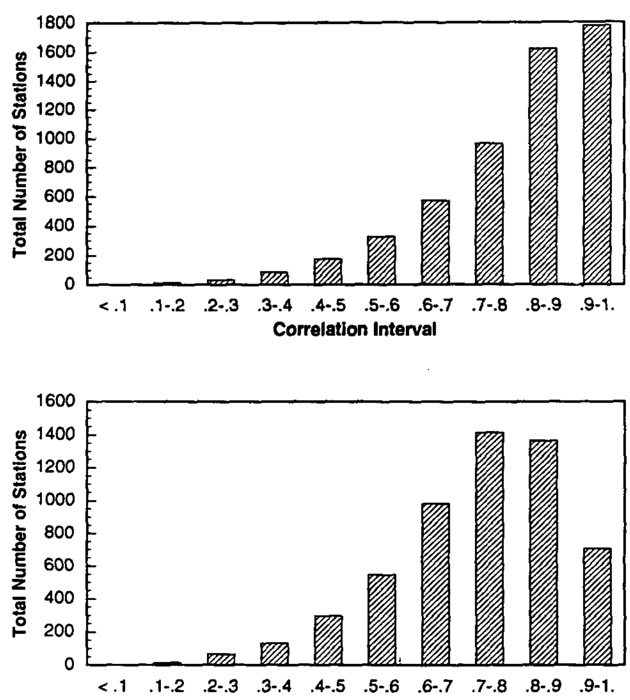


FIG. 9. (a) Frequency distribution of station correlations for the best estimation method for January precipitation. (b) Same as (a) but for July precipitation.

The frequency at which each of the various methods was found to be "best" for temperature for the months of January and July is shown in Figs. 5a and 5b, respectively. Overall, the MLAD method is observed to be superior nearly one-half of the time for both these months. The results are similar for the remaining calendar months. The differences among the other methods are not significant except for the single best estimator that always exhibits lower total numbers. In addition, there is no indication of a spatial bias (not shown) for any of the methods; that is, no method was observed to perform better in a particular geographic region. The mean correlation for each of the methods chosen as the best is in the range 0.90–0.95 for January and 0.84–0.89 for July. Figure 6a summarizes the efficiency of the best estimation method via the range of correlations found for all stations. Ninety-three percent of the stations have an R^2 of 0.7 or greater for January, while July shows 88% of stations have R^2 values of 0.7 or greater. No significant bias was detected in any of the methods with respect to over- or underestimating observed values.

The tendency of the interpolation methods to be consistent among stations and months is summarized in Table 1a, which gives the mean error of the estimate and its standard deviation for temperature. The mean error, as opposed to the mean squared error, is presented as a measure of the relative positive or negative bias of each technique. The standard deviations of the mean error are all relatively similar between methods (slightly lower in July as compared to January), though the MLAD method generally shows smaller values. The distribution of the errors shown for each station are illustrated in Fig. 7 for January. It can be seen that the MLAD method exhibits far less variability between stations compared to other methods.

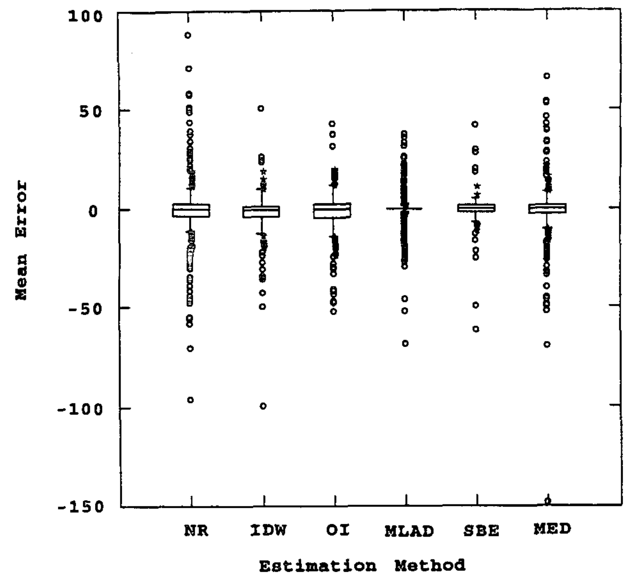


FIG. 10. The distribution of the mean estimation error for precipitation (mm).

Analysis of monthly precipitation data produced results similar to the temperature series. The MLAD method was clearly the best estimator for both January (Fig. 8a) and July (Fig. 8b). The mean correlations for each of the methods (0.71–0.85 for January, 0.62–0.77 for July) are much lower than those shown for temperature but follow the same pattern of smaller July values compared to January. The number of stations that could not be estimated is also considerably larger for precipitation as opposed to temperature due in large part to the lack of a sufficient network of stations surrounding stations in arid regions and over topographically diverse terrain. The total number of precipitation

TABLE 1b. The mean error and the standard deviation of the mean estimation error in precipitation (mm) by month for each of the interpolation methods.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
						Mean error						
NR	−0.807	−1.031	−0.696	−0.809	−2.170	−1.307	−1.958	−1.589	−0.874	−1.711	−1.749	−1.358
IDW	−2.024	−2.336	−2.515	−1.774	−1.640	−3.144	−2.140	−3.152	−2.338	−1.579	−1.292	−1.988
OI	−1.598	−1.392	−1.752	−1.593	−1.543	−1.414	−2.180	−2.353	−1.451	−0.903	−1.214	−1.077
MLAD	−0.595	−0.638	−0.510	−0.728	−0.751	−0.926	−0.978	−0.950	−0.868	−0.690	−0.517	−0.572
SBE	−1.186	−1.069	−1.961	0.090	−0.108	−1.066	−0.380	−0.930	0.789	−0.320	−1.287	−2.747
MED	−1.566	−1.491	−1.432	−1.347	−1.580	−1.553	−2.075	−2.212	−1.613	−1.844	−1.320	−1.493
Mean	−1.296	−1.326	−1.478	−1.027	−1.300	−1.568	−1.618	−1.864	−1.059	−1.175	−1.230	−1.539
						Standard deviation of mean error						
NR	53.884	49.937	55.310	58.504	60.090	64.185	67.647	68.411	69.452	62.937	52.580	56.522
IDW	40.435	41.668	50.523	42.338	51.308	57.931	59.372	52.222	56.098	46.313	42.308	43.683
OI	40.881	37.047	39.029	36.532	40.763	43.492	48.348	48.264	39.618	34.694	35.124	37.036
MLAD	33.617	31.723	33.527	34.545	37.365	41.740	44.972	43.701	39.728	34.745	32.343	33.094
SBE	44.070	56.943	49.346	45.240	65.050	65.693	61.150	59.223	61.850	55.332	51.243	50.856
MED	41.535	38.187	42.185	39.112	42.857	46.428	50.206	47.852	42.058	40.723	38.379	42.155
Mean	42.404	42.584	44.987	42.711	49.572	53.244	55.283	53.279	51.467	45.790	41.996	43.891

TABLE 2a. Summary of the percentage of the temperature data by month for each flag.

Flag	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	94.5676	94.5458	94.5848	94.5591	94.4856	94.2584	94.3160	94.3327	94.4356	94.3574	94.5834	94.4323
1	2.1011	2.2038	2.1580	2.1540	2.1337	2.1609	1.9642	2.0403	1.9119	2.0829	2.0415	2.0725
2	0.0474	0.0519	0.0198	0.0257	0.0231	0.0282	0.0316	0.0248	0.0286	0.0333	0.0433	0.0340
3	3.2407	3.1592	3.1974	3.2144	3.3037	3.4768	3.6086	3.5323	3.5502	3.4653	3.2839	3.4111
4	0.0021	0.0051	0.0038	0.0059	0.0055	0.0084	0.0139	0.0076	0.0055	0.0072	0.0047	0.0064
5	0.0410	0.0342	0.0362	0.0408	0.0484	0.0673	0.0657	0.0623	0.0682	0.0540	0.0433	0.0438

stations that could not be estimated, and thus cannot undergo a spatial quality control check, is approximately 5% of the total number of stations in any one month. The range of correlations between the estimated and observed values for the best estimation method (Fig. 9) reveals that 88% have an R^2 greater than 0.6 for January and 80% have an R^2 greater than 0.6 for July.

The mean of the errors of the estimate for the monthly precipitation values are listed in Table 1b. They are also smallest for the MLAD method though the differences are not significant. The tendency for all of the methods to have a negative bias is indicative of the nature of precipitation distributions to be positively skewed (interpolated values will tend to cluster about the median error rather than the mean). It can be seen in Fig. 10 that the distribution of mean precipitation errors is for the most part symmetric in nature, so the bias is relatively small. As with temperature, the mean error for the MLAD method indicates much greater consistency; that is, the standard deviation of the mean error is less than that found for the other methods (Table 1a and Fig. 10).

It is interesting to note that the optimum interpolation method for both temperature and precipitation is not always "optimum" in the sense of having the minimum standard error in comparison with the other techniques. This may be due to the preselection process, outlined previously, resulting in surrounding stations that are not statistically independent. Since these station intercorrelations are important for deriving the weights used for this method, there may be some bias against the optimum interpolation technique.

5. Determination of error limits

The determination of limits at which observed values should be flagged as suspect should directly reflect the efficiency of the estimation/interpolation procedure. The upper/lower limits are then defined by utilizing a multiple of the interquartile range of the difference between the observed and the estimated values presented in Figs. 7 and 10. The limits are unique for each station and month, dependent on the efficiency of the estimation process. The best estimation technique for a particular monthly series will tend to have the smaller limits (smaller interquartile range) for which observed values will be flagged as outliers. These are the stations and months when confidence in the spatial quality control is greatest. The distribution of observed minus estimated differences define the appropriate error limits.

If the difference between the observed and the estimated value is within the upper–lower limit, then the observed value is in agreement with its neighbor(s) and "passes" the spatial consistency check. If, however, the difference is large and outside its defined limit, either the observation being checked is bad or one of the surrounding observations used in the interpolation is bad. To determine which is the case, we reanalyze to the target location eliminating one of the surrounding stations at a time. If successively eliminating each neighbor produces the same result (i.e., the observation minus estimate exceeds our limits), then the target observation is considered suspect and flagged as failing the spatial consistency check. Alternatively, if eliminating one of the surrounding stations results in an estimate that agrees with the observation, then the observation is likely good and the neighbor is labeled as suspect. The iterations proceed until each year/month for each station has been analyzed and can be labeled with a simple set of flags: the data value in question can either pass or fail the spatial consistency check or cannot be checked due to a lack of neighboring stations.

6. Summary and conclusions

Final preparation of the dataset(s) involves combining the flags from the temporal (ICC) and spatial (HCC) error-checking procedures to produce a final set of flag codes. These flags are coded for each year/month as follows:

flag = 0 pass both ICC and HCC
flag = 1 pass ICC, fail HCC
flag = 2 fail ICC, pass HCC
flag = 3 pass ICC, unable to perform the HCC
flag = 4 fail ICC, unable to perform the HCC
flag = 5 fail ICC, fail HCC.

In addition, a flag code of 6 is included to indicate a missing or indeterminate value (i.e., there was insufficient data to establish the value of the flag value). The flag codes are summarized as a percentage of the total number of nonmissing values in Tables 2a and 2b for temperature and precipitation, respectively.

TABLE 2b. Summary of the percentage of the precipitation data by month for each flag.

Flag	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	90.4392	90.4258	91.3531	91.7011	91.8753	89.6479	88.5366	89.6239	89.8661	92.0142	91.2089	90.5981
1	0.9425	0.8813	0.7909	0.7405	0.9519	0.9379	0.9793	1.1465	1.3261	1.0314	1.0343	0.8235
2	0.2228	0.1659	0.2090	0.1938	0.2765	0.2736	0.3189	0.3667	0.3169	0.2501	0.2516	0.2059
3	8.1870	8.3156	7.4242	7.1363	6.6161	8.7947	9.8251	8.4572	8.0808	6.4729	7.2428	8.1547
4	0.0567	0.0700	0.0676	0.0634	0.0602	0.0806	0.0599	0.0713	0.0768	0.0338	0.0547	0.0583
5	0.1517	0.1415	0.1553	0.1650	0.2200	0.2654	0.2802	0.3343	0.3333	0.1976	0.2078	0.1595

The results shown for temperature (Table 2a) indicate that approximately 94% of the data passes both the ICC and the HCC tests. The percentage of data that fails either test or both is on the order of 2.5%. Approximately 3.5% of values could not be checked by the HCC. For precipitation (Table 2b), roughly 90% passes both tests, whereas the failure rate is of the order of 1%–2%, and 8%–9% of the data were unable to be checked spatially.

The goal of any quality control procedure is to provide the end user with as much information as possible such that he/she can make an informed choice whether to accept or reject a particular monthly value. The results of our analysis provide the end user with a set of flags, noted above, and with summary statistics of the efficacy of several estimation procedures.

In the aggregate, the MLAD method outperforms the other five methods nearly one-half of the time. This is true for both temperature and precipitation for all months. Analysis of the mean error as an indicator of estimation bias showed that, except for a small negative bias for precipitation, none of the methods exhibited tendencies to over- or underestimate observed values. Additionally, no significant spatial bias was found for any of the methods for either temperature or precipitation.

This study should bound the uncertainty on the available set of surface monthly mean temperature and monthly precipitation measurements; it also provides a baseline dataset for validation of climate models as well as for a variety of other climatological uses. The result of this analysis provides a high quality updateable baseline climatological dataset with global coverage. This is the first step in methodically analyzing in a reproducible manner trends in temperature and precipitation in order to assess, quantitatively, climate change on different spatial and temporal scales.

In an analysis of temperature by Karl et al. (1994) there was a rather broad range of values obtained for the surface trends. They found that on a global scale for a surface trend calculated from 1850 to 1993 the difference in the magnitude between a 5° × 5° grid and Hansen and Lebedeff's (1988) original calculation could be as much as 0.17°C per century, depending on the size of the grid boxes, the technique used for the grid interpolation, the method used to obtain the linear regression, and where one starts in the time series. We feel that there is a great need to perform a careful study of the various methodologies used to develop surface temperature trends to determine the variance induced simply by changing the analysis technique. There may be a unique set of techniques for different spatial and temporal scales as well. To accomplish such analysis, one needs to start with a well-documented quality controlled baseline climatological dataset.

REFERENCES

Barrodale, I., and F. D. K. Roberts, 1973: An improved algorithm for discrete L1 approximation. *SIAM J. Num. Anal.*, **10,** 839–848.

Bennet, R. J., R. P. Haining, and D. A. Griffith, 1984: The problem of missing data on spatial surfaces. *Ann. Assoc. Amer. Geogr.*, **74,** 138–156.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Gidrometeorologicheskoe Izdatel'stvo, Leningrad. Translated from Hansen, J. and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.*, **92,** 13 345–13 372.

Karl, T. R., J. D. Tarpley, R. G. Quayle, H. F. Diaz, D. A. Robinson, and R. S. Bradley, 1989: The recent climate record: What it can and cannot tell us. *Rev. Geophys.*, **27,** 405–430.

——, and Coauthors, 1994: Global warming update. 74th Annual American Meteorological Society Meeting, 24–28 January 1994.

Paulhus, J. L. H., and M. A. Kohler, 1952: Interpolation of missing precipitation records. *Mon. Wea. Rev.*, **80,** 129–133.

Schlatter, T. W., 1975: Some experiments with a multivariate objective analysis scheme. *Mon. Wea. Rev.*, **103,** 246–257.

Thiebaux, H. J., and M. A. Pedder, 1987: *Spatial Objective Analysis with Applications in Atmospheric Science.* Academic Press, 299 pp.

Vellum, P. F., and D. C. Hoaglin, 1981: *Applications, Basics, and Computing of Exploratory Data Analysis.* Duxbury Press, 354 pp.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. K. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041. Carbon Dioxide Analysis Center, 100 pp.

Young, K. C., 1992: A three-way model for interpolating monthly precipitation values. *Mon. Wea. Rev.*, **120,** 2561–2569.