2.4

IDENTIFYING INHOMOGENEITIES IN PRECIPITATION TIME SERIES

Edward I. Tollerud
NOAA Research-Forecast Systems Laboratory
Boulder, Colorado

Barbara G. Brown and Tressa L. Fowler
Research Applications Program, National Center for Atmospheric Research
Boulder, Colorado

1. INTRODUCTION

The identification of systematic nonmeteorological changes in precipitation observations is critical to many users, including those in the power-generating and transportation industries as well as researchers doing climate change studies. At gage sites, these changes commonly involve location, elevation, exposure, or instrumentation, leading to systematic increase or decrease in precipitation amounts. Other, more subtle effects involve changes in observation time, method, consistency, etc. These kinds of changes may have a greater effect on precipitation frequency, timing, variability, and completeness than on the actual precipitation totals. They may also be more difficult to find.

Although procedures exist to identify “change points” in individual time series (Jaruskova 1996, 1997; Potter 1981), the large natural temporal variability in precipitation seriously complicates methods that rely solely on long-term characteristics from a single observing location. Comparison with other sites that are likely to share the same precipitation characteristics can provide additional information to the search for subtle changes. To facilitate this process, it is useful to formulate time-dependent measures of spatial correlation. Ultimately, time series of these measures can then be subjected to change-point analyses that may identify times at which a station’s relationship to its neighbors has significantly changed. In this paper we describe indices based on verification scores (specifically, bias and the equitable threat score) ordinarily used to evaluate numerical forecasts. We then assess their usefulness to climate station monitoring as pursued in the National Climatic Data Center (NCDC) “Health of the Network” project by applying them to daily precipitation observations at U.S. cooperative observer network sites in Iowa.

Corresponding author address: Edward Tollerud, NOAA/FSL, R/FS1, 325 Broadway, Boulder, CO 80305-3328. E-mail: tollerud@fsl.noaa.gov. Tel: 303-497-6127.

2. IDENTIFYING INHOMOGENEITIES

The extraction of often subtle climate change signals from noisy climatological records requires very careful monitoring of the performance of climate stations. The immediate objective of the work presented here is to establish statistical procedures that can effectively and accurately identify changes in the characteristics of precipitation gage observations that adversely affect the validity of temporal change data from the network. In addition to degradation in data quality, changes to be identified include unreported station moves and instrumentation variations. The ultimate goal of the project is to develop an automated system that will provide “early warning” indicators of possible problems associated with data reported by identified stations, so that data users can be aware of such problems at an early stage.

Numerous approaches have been developed to identify and correct inhomogeneities in historical climate time series (*e.g.*, Peterson *et al.* 1998). However, these approaches generally have been designed for retrospective studies and are able to take advantage of relatively long time series of observations both before and after the occurrence of the inhomogeneity (*e.g.*, Alexandersson 1986; Hanssen-Bauer and Forland 1994). In contrast, the goal of this study is to develop methods to monitor and identify inhomogeneities as soon as possible after they occur or develop. Thus, it may be necessary to develop methods that are sensitive to more subtle changes in the series than is required for retrospective analyses.

Precipitation station data quality monitoring has one distinct disadvantage that renders many methods unfeasible: extreme spatial variability. This variability prevents, for instance, any reliance on evaluation of subtle differences in rainfall totals, even over time periods of a month or more. It also necessitates more innovative use of nearest-neighbor analyses than a simple comparison of temporal totals. Thus, it will likely be necessary to focus on a variety of tools to identify possible changes. Possible tools include time series approaches *(e.g.*, change point analyses, Kalman filter) as well as spatial relationships among distributional characteristics of the observations. Ultimately, we anticipate that a variety of tools will be applied and blended to formulate a warning indicator that could be used to evaluate data as they arrive at NCDC. Since they are tuned to be sensitive to unique precipitation characteristics, verification utilities like those applied here may also serve as sensitive indicators of changes in a station's spatial relationship to neighboring observations.

Gonzalez-Rouco *et al.* (2001) describe an approach for detection of inhomogeneities in precipitation data which involves comparisons of time series of observations from neighboring stations. This approach is based on the method described by Alexandersson (1986) and relies on the use of the normal distribution to model deviations in ratios of values from a single station to values from the neighbors. Unfortunately, several years of data are required to allow detection of an inhomogeneity, so the method is better applied for retrospective analyses, rather than as a near-real-time warning indicator.

3. COOPERATIVE PRECIPITATION DATA

Although it may become necessary to use longer-period rainfall to control temporal and spatial variability, we have chosen to first analyse daily precipitation data. These gage data are from the TD3200 dataset observed by the United States cooperative observer network and available on CD-ROM from NCDC.

The results described here are from observing sites in Iowa. For these preliminary analyses, this choice of state removes the complicating influence of extreme topography and allows some presumption of spatial homogeneity. In addition, the observing sites in Iowa are well distributed across the state and the data themselves appear to be of good quality. In contrast to these simplifying characteristics, however, we chose to analyse June 1994 observations, under the assumption that summertime convective precipitation typical of this month would provide a rigorous test of the analyses.

Figure 1 depicts the rainfall in 60 of these stations during June. Since the time series are ordered such that proximate series are constructed from observations at gagesites that are close neighbors geographically, the consistency from series to series suggests data of generally high quality. This is particularly true, for example, during the statewide heavy rainfall on June 23. However, A few problem sites appear. The three timeseries with no rain days during the entire month, for instance, are stations for which the station history information indicated incorrect periods of operation; in fact, no observations exist in the dataset for these station during this month (ramifications of this inconsistency between data and station history will be discussed in the next section). Similarly, as indicated by the lighter shading of several of its daily bars, station 133589 (index 13) fell prey to the common systematic problem of rainfall accumulated over more than one day before being reported. The large total precipitation on June 23 of more than 5 inches is a result of this reporting problem. Other stations, like 131962 (index 3), appear to be out of lockstep with the majority of other nearby sites.

These data and their idiosyncracies illustrate the difficulty inherent in this endeavor: except for the few time series with recognized systematic problems, it is not obvious which, if any, of the stations are observing inaccurately. In a set of reasonable-looking time series which still exhibit noteworthy differences from neighbor to neighbor, what kind of test is likely to pull out from these measurements those stations at which observations have experienced some change?

4. APPLICATION OF VERIFICATION SCORES

Using verification scores as proxy measures of correlation between stations has some distinct advantages over other measures. Since these scores are tuned to be sensitive to the unique characteristics of rainfall (its on-off nature and relative infrequency, for instance), they are also more likely to capture the variability of those characteristics.

Verification scores used here are the bias computed with precipitation rate and the equitable

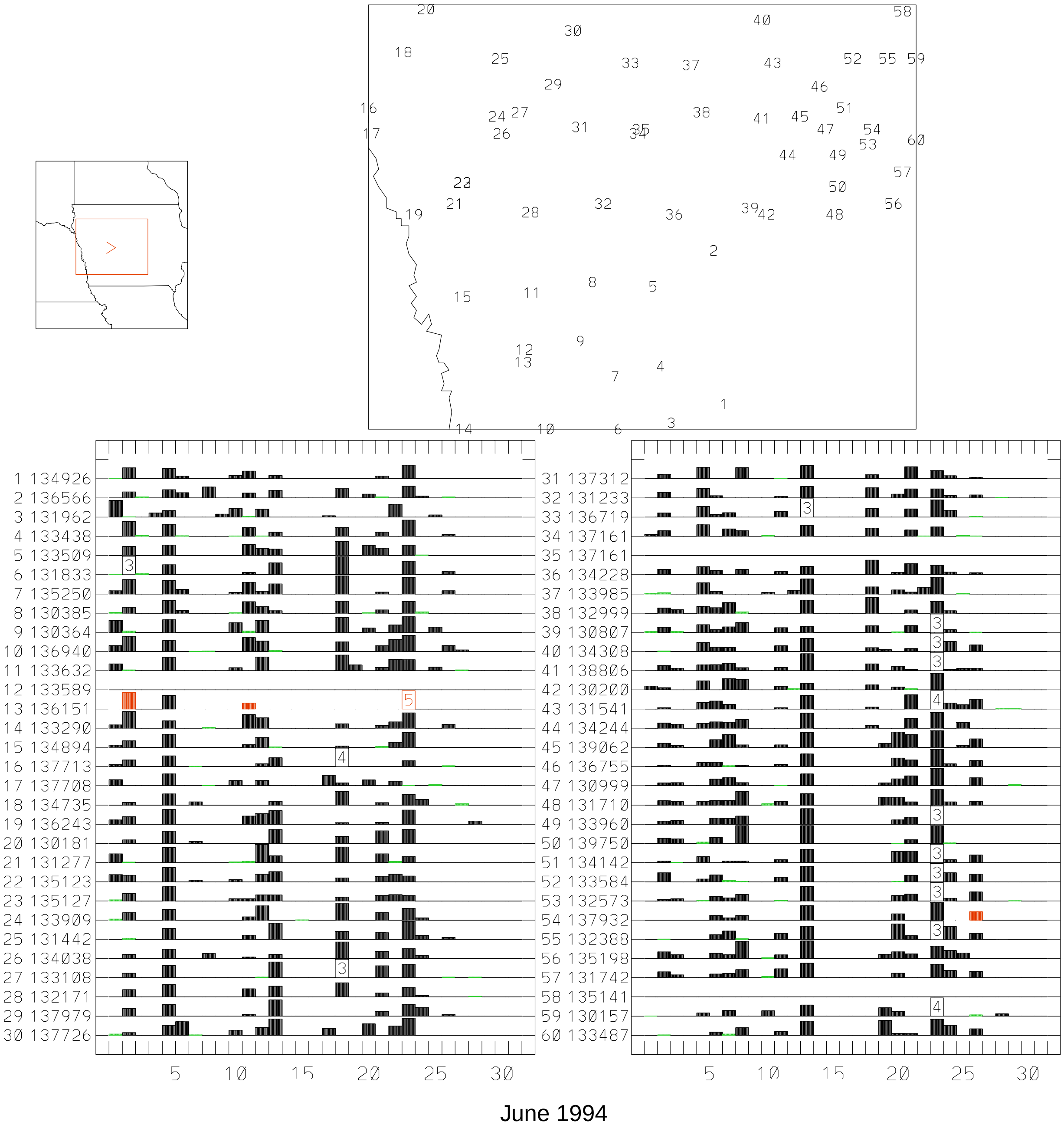

Fig. 1. Precipitation at selected cooperative observer stations in Iowa for each day during June 1994. Solid bar heights are logarithmically scaled such that the height between series axes corresponds to three inches of precipitation. Rainfall greater than 3 inches is indicated by the number in the open bars. Locations of gage stations (identified by index numbers and station IDs along the vertical axes) are indicated by index numbers plotted on map insets. Dotted lines (e.g., station index 13) indicate missing observations; bars with lighter interior shading were flagged as inaccurate.

threat score (ETS). These and other precipitation scoring algorithms are described in Wilks (1995). Bias is fundamentally a ratio of total precipitation at a target station or set of stations (or, more commonly, model gridpoints) to total precipitation observed at the same set of locations by a reference (verification) network. Hence, it is sensitive to inaccuracies of measurement. The ETS, on the other hand, measures how often the target and reference networks have "hits", that is, both observe precipitation over a certain threshold. In our study, the threshold is low (.01 in), so a hit essentially means that both networks have recorded nonzero precipitation. Thus, the ETS is sensitive to correlations of rainfall *occurrence* between networks but not to correlations of precipitation *amount*.

Typical uses of verification scores are to evaluate model performance. However, they can also be used as measures of correlation between pairs of observations. In our application, we assemble pairs of observations for each Iowa gagesite by finding its 10 closest neighboring stations on each of the 30 days in June 1994 when either reported. Thus, each station-specific score computation includes 300 (in some cases of missing reports, a few less than 300) comparison pairs of daily precipitation observations.

The distribution of bias scores thus calculated is displayed on the panels of Fig. 2. As the legend explains, the distribution in Fig. 2a was produced before recognition of an inconsistency between station history information in a separate metacode file and the actual data. The result was a set of 8 stations that did not make precipitation reports during this month (a few reported temperature) but which were assumed in the computations to have observed zero precipitation on each day. The stations thus affected show up in the distribution as a secondary peak at bias 0. However, their greatest impact is to inflate bias scores at neighboring target sites by contributing zero precipitation to reference site totals. In Fig. 2b the inconsistency is corrected by removing the affected stations as possible neighbors. They are, however, retained as target stations so that the secondary peak at bias 0 remains. Comparison of the two panels reveals that the principal result of this correction is to narrow the range of scores by moving some large values back toward the median.

In a sense, the scenario of Fig. 2a represents a kind of inhomogeneity, albeit an ultimately correctible one. Although not due to an observational problem *per se*, it is a type of mistake that is not uncommon even for knowledgeable users, and a technique that would identify it along with more purely observational changes would be useful. The secondary peak at zero bias is certainly a powerful clue in this case.

For purposes of station monitoring, these distributions would be used to flag stations whose scores were clearly separated from the bulk of scores or far enough out in either tail to be considered significantly worse than the general population of stations. To test the bias scores in this way, we created a “contaminated” station by doubling

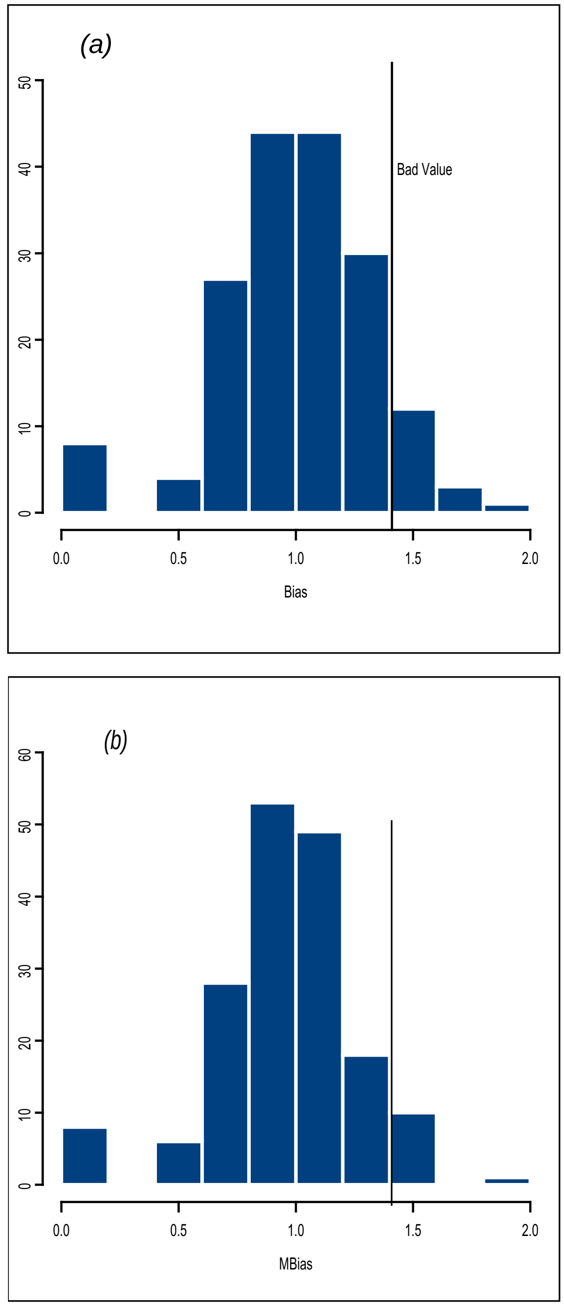

Fig. 2. Distributions of the bias score (see text) for 173 stations in Iowa during June 1994. In (a), the eight nonobserving stations with scores of zero were improperly included as neighbors with zero precipitation on each day of the month. In (b), these eight stations are correctly eliminated as neighbors although their own bias scores of zero are retained. The vertical line in each panel indicates the bias scores of a contaminated time series created by doubling each nonzero observation of the original time series.

one station's precipitation. This station is clearly toward the large-score tail of the distribution in Fig. 2b, but there are still several stations with higher biases. Several of these "bad" stations are shown on Fig. 1, and their situations are illuminating. For instance, station 137726 (index 30) appears to score poorly because it reported rain on more days than its neighbors (it also appears to be out of step with its neighbors, but this fact in itself would not produce a high bias score). It might in fact be the case that this site should be examined more closely for possibly inconsistent reporting. Station 133108 (index 27), on the other hand, reports rain on much the same days as its neighbors but still has a high bias score. The apparent explanation in this case is that on a couple of days of large precipitation it overdoes its report (June 18 is an example) relative to its neighbors. This points out a weakness of this bias computation, which is sensitive to a few large errors on days of heavy precipitation. Longer periods (a season, perhaps) of observation would help to alleviate this effect.

Figure 3 illustrates an application that appears more successful. Since the ETS tallies "hits" and would be unaffected by a simple increase of precipitation on the same day, we created a contaminated station with another kind of "inhomogeneity" in which the majority of rainfall occurs a day earlier than reported in the original time series. This station's hits with its neighbors should therefore be reduced, and indeed, this station is found far enough into the low-score tail of

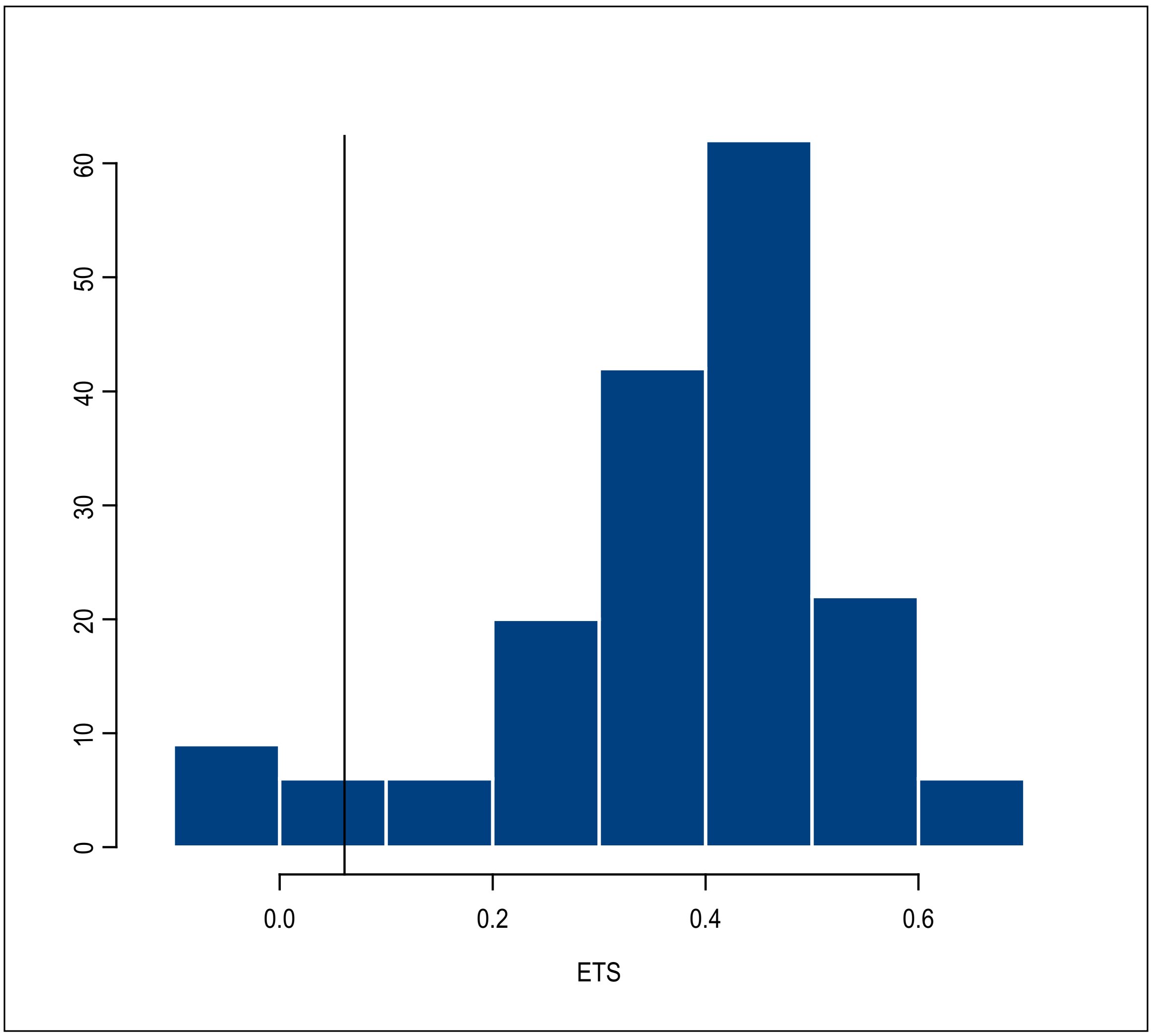

Fig. 3. As in Fig. 2b except for the distribution of the equitable threat score (ETS). The vertical line indicates the score of a contaminated station series created by displacing the majority of daily nonzero reports by one day.

the ETS distribution to be a candidate for screening. As before, the zero ETS peak contains the 8 nonreporting stations.

5. CONCLUSIONS

As suggested previously, these comparisons are just the first step toward a system to monitor precipitation station performance. The location of an individual station in these static distributions are almost certainly less illuminating than would be *changes* in their location. For instance, for reasons of topography or natural variability, a station might exhibit differences from its neighbors that result in its location farther out in the tail of the distribution. These differences might have nothing to do with measurement inhomogeneities. If this station suddenly changes location in the distribution, however, then an unreported change in gage characteristics or observation practice is suggested and should be checked out. By this reasoning, it may be more valuable to monitor the relative change in performance (as given by distribution location) than the absolute magnitude of the measure of performance.

The position of the contaminated observations toward the tails of the distributions suggests some usefulness of verification scores as proxy measures of correlation; examination of time series of these scores will determine just how useful they might be as monitors for station performance. It is also clear, however, that different scores are sensitive to different types of potential station inhomogeneities. Perhaps in the end a combination of several measures of correlation will be necessary.

Acknowledgments. This work was completed with funding from the NCDC Health of the Network Project. We thank Adrian Marroquin for his helpful review.

6. REFERENCES

Alexandersson, H.A., 1986: A homogeneity test applied to precipitation data. *Journal of Climatology*, **6**, 661-675.

Gonzalez-Rouco, J.F., J.L. Jimenez, V. Quesada, and F. Valero, 2001: Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate*, **14**, 964-978.

Hanssen-Bauer, I., and E. J. Forland, 1994: Homogenizing long Norwegian precipitation series. *Journal of Climate*, **7**, 1001-1013.

Jaruskova, D., 1997: Some problems with application of change-point detection methods to environmental data. *Environmetrics*, **8**, 469-4833.

Jaruskova, D., 1996: Change-point detection in meteorological measurement. *Mon. Wea. Rev.*, **124**, 1535-1543.

Peterson, T.C., D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D, Parker, 1998: Homogeneity adjustments of *in situ* atmospheric climate data: A review. *International Journal of Climatology*, **18**, 1493-1517.

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean precipitation series. *Mon. Wea. Rev.*, **109**, 2040-2045.

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