2.9 ESTIMATED OBSERVATION TIMES FOR COOPERATIVE STATIONS

Karen Andsager* and Kenneth E. Kunkel
Illinois State Water Survey, Champaign, IL

1. INTRODUCTION

Observation times for the Cooperative Observer Network (COOP) have not been digitized for observations taken before 1982, yet they are needed for various climatological applications. Differences in observation time introduce a nonclimatic bias in the temperature record of a station (see DeGaetano 2000). DeGaetano (1999, 2000) has developed a technique to simulate observation time based on day-to-day temperature variations and has applied it to the stations in the daily Historical Climatological Network. DeGaetano's technique can be applied to a single station, and produces annual simulated observation times.

Monthly observation times were needed to assist with quality control of NCDC's new Summary of the Day TD-3206 data set, which covers the entire U.S. for the first half of the 20th Century. A new technique has been developed to estimate monthly observation times based on the correlation of the maximum temperatures for a station with surrounding stations. This technique relies on a sufficient number of surrounding stations, which is now available in TD-3206.

2. ESTIMATION TECHNIQUE

The estimated observation times are based on the correlation of the maximum temperatures for a station with the surrounding stations on a monthly basis. Estimated observation times are distinguished as either AM or PM, with midnight-observing stations falling in the PM category. Knowing the category of observation time is usually sufficient for most climatological applications, although some applications require the exact hour of observation. The term observation time is used here to refer to the two categories of AM and PM. In general, the correlation between two stations with the same observation time (i.e., AM with AM, or PM with PM) is higher with no lag, while the correlation between two stations with different observation times (i.e., AM with PM) is higher with a lag of one day. The technique identifies AM observing stations with maximum temperatures digitized shifted back one day as PM stations.

The observation time estimation technique was developed and tuned using data from Illinois stations. The technique starts with a set of known (digitized) observation times for a set of Illinois stations. The observation times are then estimated for other continental U.S. stations in order of increasing distance from Urbana, Illinois. Since the technique works through the set of stations in a bootstrap fashion, it is also possible to start the estimation process by assuming a set of nearby stations are PM, which they generally are for the Cooperative Observer Network. The observation times are estimated twice for all stations; on the second pass through the stations, the observation times may be estimated for those stations that were assumed to be PM.

The technique has been applied to all U.S. stations with at least 10 years of digitized maximum temperatures for the period 1898-1947 to assist with the quality control of TD-3206. There are over 4500 stations which meet this criteria. The technique could also be applied to stations with shorter records, including ones with only a single month of data.

3. U.S. OBSERVATION TIMES

For the period 1898-1947, the average number of digitized COOP stations with AM observing times digitized as such was 10% for the continental U.S., with the rest PM observing times and AM stations digitized shifted back one day. Application of a technique for identifying AM-shifter stations suggested that 75 stations have at least three years of AM-shifted maximum temperatures. These stations were not evenly distributed across the U.S.

The estimated observation times indicate over 130 stations switched between AM observation times in the summer months and PM observation times in the winter months for at least one year.

The observation time estimation technique also distinguished individual months for some AM observing stations in which the editor failed to mark the shifting of the maximum temperatures back one day, resulting in stray months digitized "correctly" as AM in a series of months otherwise digitized shifted to "PM".

4. ACCURACY OF ESTIMATED OBSERVATION TIMES AND STATION QUALITY

The estimated observation time is based on the relative fraction of neighboring stations which, when correlated with a station, suggest that that station's observation time is AM or PM. If there is not a clear majority of stations suggesting one or the other observation time for a particular month, the month is flagged as indeterminant. If the correlations give no guidance at all for an estimate, then the same observation time as the previous month is used, as it is

* *Corresponding author address*: Karen Andsager, Illinois State Water Survey, University of Illinois, 2204 Griffith Drive, Champaign, IL 61820; e-mail: andsager@uiuc.edu.

most likely that the observation time did not change. Since the technique uses correlations with neighboring stations, the rate of flagging for a particular station may be considered as a measure of the quality of the data for that station. Stations with high rates of flagged estimated observation times are suspect.

For the Midwest, 80% of the stations have fewer than 5% of their estimated monthly observation times flagged by the estimation technique. Only one of the 982 Midwestern stations has more than 30% of its estimated monthly observation times flagged. For the entire U.S., 50% of the stations have fewer than 5% of their estimated monthly observation times flagged. Of the 4583 U.S. stations, 3% have more than 30% of their estimated monthly observation times flagged. These stations are in the mountainous West and desert Southwest. In general, the rate of flagging is higher in the western third of the U.S. and Texas, as well as Maine and Florida. In the case of the later two states, the rate of flagging is most likely due to the lower number of nearest neighbors available. In the western U.S., the station density is lower, so fewer nearest neighbors may be available there as well. Also, in the mountainous western U.S., stations may not correlate as highly, resulting in a higher flagging rate.

Original records for stations both within Illinois and outside the Midwest have been examined in an attempt to quantify the accuracy of the estimated observation times. In doing so, it is necessary to assume that the observer knew what time he was observing at and accurately recorded it on the form. Since the estimated observation times are based on the digitized records, it is also necessary to know how the keyers were to interpret and key AM-shifting of the maximum temperatures by either the observer or by later editors. For TD-3206, as well as TD-3200 and TD-3205, the editors' AM-shifting were accepted, and are therefore reflected in the digitized data. It is then necessary to examine the original forms not only for the observation times but also for any other marks or comments about possible AM-shifting or other observational irregularities.

The accuracy rate for the Illinois stations on which the estimation technique was developed is very high, better than 98%. These stations include several stations which switch between PM observations in the winter months and AM observations during the growing season, during the period from about 1912 through 1918. The estimation technique generates accurate observation times for these stations as they switch between AM and PM.

The original records for several California stations were examined to try to determine the accuracy of the estimation technique in a region far removed and climatologically different from the relatively flat Midwest. The monthly observation times were taken from the original records for thirteen stations for the period 1915-1919. Of the thirteen stations, six were PM stations, one was an AM station, and six were AM-shifter stations. For the six PM stations, the accuracy rate is very high, better than 98%. For the one AM station, the accuracy rate is also better than 98%. The estimation technique accurately catches the change in observation time of this station from AM to PM in May of 1919. For the AM-shifter stations, the estimated observation time should be PM, which gives an accuracy rate of better than 80% (that is, 80% of the estimated monthly observation times are PM for the AM-shifter stations). A large portion of the lower accuracy rate for the AM-shifter stations comes from 3 years of observations for one station which are marked as AM-shifted but which all have AM estimated observation times. It is unclear whether the estimation technique or the observer and editor are responsible for this discrepancy.

Two California stations with long records and flagging rates of over 40% were examined in detail. For one of the stations (047253 Randsburg), the original forms indicate that the observation time was always AM for the period 1937-1947. The estimated observation times are PM for 75% of the months and AM for the other 25%. There is no indication of AM-shifting on the forms. For the other station (044223 Imperial), the original forms indicate that the observation time was AM for 1912-1941, then switched to PM in 1942. Beginning in October 1942, the forms include both AM and PM observation times and the editor's comments "thermometer set twice daily" and "maximum temperature set back". This continues through 1947. As with the other station, about 75% of the estimated observation times are PM and 25% are AM. It is unclear how these stations should be included in a calculation of the estimation technique's accuracy. The ability of the estimation technique to accurately estimate the observation times for stations with relatively low flagging rates suggests that for stations with high flagging rates, the observers were following irregular observing practices.

5. ACKNOWLEDGMENTS

This work was supported by the NOAA Office of Global Programs, Climate Change Data and Detection Program. The views expressed in this document are those of the authors and do not necessarily reflect those of NOAA.

6. REFERENCES

DeGaetano, A.T., 1999: A method to infer observation time based on day-to-day temperature variations. *J. Climate*, **12**, 3443-3456.

DeGaetano, A.T., 2000: A serially complete simulated observation time metadata file for U.S. Daily Historical Climatology Network stations. *Bull. Amer. Meteor. Soc.*, **81**, 49-67.