A Method to Infer Time of Observation at U.S. Cooperative Observer Network Stations Using Model Analyses

Brian N. Belcher and Arthur T. DeGaetano

Northeast Regional Climate Center

Cornell University

Ithaca, NY 14853

Revision submitted to :

International Journal of Climatology

January 2005

Corresponding author address:

Dr. Arthur T. DeGaetano, Northeast Regional Climate Center,

Cornell University, 1119 Bradfield Hall, Ithaca, NY 14853

Email: atd2@cornell.edu

Phone: 607-255-0385

FAX: 607-255-2106

Abstract

A method to estimate the time of observation employed at U.S. Cooperative Observer Network stations has been developed using Rapid Update Cycle model analyses. This method uses the day-to-day variability in model temperature biases to estimate observation schedules on a timescale of weeks, making it ideal for use in "real-time" applications. Observation time estimates from a 2-category system (morning and "non-morning") and 3-category system (morning, afternoon and midnight) were both evaluated. The performance of the 2-category system was compared with existing techniques that employ this system on monthly timescales. The results were comparable, showing dependence on season and climatological characteristics, but reveal an ability to reach high levels of accuracy (>90% of stations have observation schedules correctly estimated) over similar time periods (10-50 days). To our knowledge, the evaluation of 3-category estimation performance for the timescales investigated has not been documented. Accuracy remained high for morning and midnight stations (>90%), and decreased for stations with afternoon observation schedules (85%-65%). Additionally, the 3-category estimation technique was extended to 4 categories in order to identify observers who shift temperature records temporally. The accuracy of detecting shifted records within the context of the 4-category estimation technique was comparable to the performance of the 3-category system, with shifted observations correctly identified more than 75% of the time in most cases.

Keywords:

Observation time; metadata; Rapid Update Cycle; Cooperative Observer Network;

Daily temperature; Bias

1. Introduction

Since 1890, daily observations of maximum and minimum temperatures and precipitation have been recorded at numerous locations across the United States as part of the U.S. National Weather Service's Cooperative Observer Network (CON). Given the volunteer nature of the CON, over 11,000 participating stations have specific, but different, observation times (OBT). When deciding on an appropriate OBT for their particular station, each observer takes into consideration times that are suggested by the National Weather Service, as well as times that best fit into their daily schedule. While these observation times are generally consistent from day-to-day and recorded by the observer on a monthly basis, this information is sometimes undocumented or incorrectly archived.

Each OBT biases a station's temperature record to a different degree (Baker 1975; Blackburn 1983; Karl et al. 1986; DeGaetano and Knapp 1993; Janis 2002). Minimum temperature biases are largest for morning observation schedules while maximum temperature biases are largest for afternoon observation schedules. These biases are largest when daily extreme temperatures occur near the time of observation, as the temperature occurring at the beginning of a 24-hour cycle may be the highest (maximum temperature) or lowest (minimum temperature) during this period. Such timing results in 'carry-over' of extreme daily temperature reports from one day to the next. Additionally, maximum temperatures occurring on day x are frequently recorded on day $x+1$ at stations that follow morning observation schedules since the highest daily temperature often occurs during the afternoon of the previous calendar day. As a result, maximum temperature time series reported by morning observers appear shifted by one day when compared to time series of maximum temperatures recorded by afternoon or midnight observers (Figure 1).

It is important to maintain an accurate record of OBT in order to correctly account for such observational differences when data are used in climate research and applications. Recognizing potential errors in a timely fashion minimizes their propagation in the data record. Efforts to alleviate OBT problems in the CON metadata include the inference of observation times on an annual timescale (DeGaetano 1999, 2000) as well as on a weekly to monthly timescale that is more suitable for 'real-time' applications (Andsager and Kunkel 2002; Belcher and DeGaetano 2003).

Andsager and Kunkel (2002) recently developed their method of estimating monthly observation times in order to assist with the quality control of the National Climatic Data Center's (NCDC) new Summary of the Day TD-3206 dataset. Using this method, stations were assigned one of two observation schedules, morning or afternoon, with midnight-observing stations falling in the afternoon category. OBT estimates were based on the correlation of the maximum temperatures for a station with surrounding stations. The method was tested on over 4500 CON stations over the period 1898-1947 with at least 95% accuracy at about 50% of the stations, and less than 70% accuracy at fewer than 3% of the stations. Belcher and DeGaetano (2003) identified patterns of observation-time bias in a CON station's empirical temperature probability distribution function to develop a method to detect OBT changes between morning and afternoon categories on a timescale of weeks to months. Using this method, some regions across the U.S. required 3 to 6 months of data to achieve detection rates of over 95% while limiting false alarm rates to less than 10%. Both of these methods had limited accuracy in mountainous and coastal regions, particularly California and Florida. Additionally, Andsager and Kunkel (2002) attribute low station density to performance problems in Maine.

Reporting errors are also sometimes introduced into the daily temperature records by CON observers who shift observations by one day. Often this practice is associated with observers using a morning observation schedule. These observers do not follow recording procedures outlined by the CON, but rather record maximum temperatures based on the calendar day in which they perceive the temperatures occurred (as opposed to the date that the observation was made). As with correct OBT characterization, identification of shifting problems in a timely fashion would allow for higher-quality temperature data through correction of the shifted data and retraining of the observer to prevent shifting in the future.

This paper describes a method that has been developed to determine time of observation at CON stations by utilizing Rapid Update Cycle (RUC) model analyses (Benjamin et al. 2002). Continuous improvements in model physics, resolution, and data assimilation techniques have increased the accuracy of both model analyses and forecasts, providing a useful tool for this application. This method determines OBT on a timescale of weeks, improving upon both the accuracy and temporal resolution of Belcher and DeGaetano (2003), without depending on the potentially uncertain metadata of surrounding stations and inadequate station density that plague the technique of Andsager and Kunkel (2002). This new technique extends beyond 2-category OBT estimates (morning and afternoon) to include midnight OBT estimates (3-category) and identification of observers who report maximum temperatures erroneously shifted in time (4-category).

2. Data

Daily maximum and minimum temperatures from 5 subsets of CON stations (Fig. 2) were obtained for the period March 1, 2003 through March 1, 2004. Each subset was selected for evaluation of the OBT estimation method in diverse climate regions. These regions include locations in which previous methods (DeGaetano 1999; DeGaetano 2000; Andsager and Kunkel 2002; Belcher and DeGaetano 2003) were found to have problems determining OBT (e.g. coastal, mountainous) and regions in which OBT was determined with good accuracy (e.g. Midwest).

The availability of hourly analyses from the latest version of the RUC model (RUC20; Benjamin et al. 2002) on an approximately 40-km lambert conformal grid covering the contiguous United States dictated the 1-year analysis period. Surface air temperature (2 meters) as well as temperature and height at 1000 hPa, 950 hPa, 900 hPa and 850 hPa levels were the only RUC analysis variables necessary for OBT estimation. CON observations are not operationally assimilated in the RUC initialization. RUC analyses were obtained from NCDC/NOMADS (http://nomads.ncdc.noaa.gov/data-access.html).

Evaluation of this method was based on the comparison of OBT estimates with observation times reported in the metadata. Digital metadata files used for this purpose were obtained from NCDC. These monthly files include the temperature observation time at each station and are supposed to be consistent with publications of *Climatological Data*. Some inconsistencies do exist however between these two sources of metadata, as some stations have different observation times documented in each reference. These differences are utilized during the evaluation of our method to show that such inconsistencies can be revealed and corrected.

3. Method of Observation Time Estimation

The method of characterizing each CON station's observation schedule involves the comparison of interpolated (from RUC analyses at the grid point closest to the station of interest) and observed (reported from the CON) daily maximum and minimum temperatures over 10, 20, 30, 40 and 50-day periods. In order to mimic data recorded at CON stations, daily extreme temperatures were determined at each RUC grid point by retaining the maximum and minimum hourly 2-meter temperature values occurring within 24-hour periods that correspond to each of the three most common observation schedules (0700, 1700 and 2400 LST). These three sets of interpolations were selected to represent morning (AM), afternoon/early evening (PM), and late evening/midnight (MID) observation schedules, respectively. This resulted in 1 observed (recorded by the station observer) and 3 RUC interpolated (AM, PM and MID) daily maximum and minimum temperature time series for each CON station.

a. Using RUC Interpolated Temperature Time Series to define OBT

Maximum temperatures interpolated from RUC analyses tend to have a negative (RUC - observed) bias and minimum temperatures tend to have a positive bias, however median biases for these variables are generally less than 1°C and consistent throughout the annual cycle. Despite the presence of model biases in the temperature data, the variance of the daily temperature bias should be relatively small if the simulated observation schedule is close to the actual observation schedule. Larger variances should characterize the other simulated observation schedules. Therefore, each of the 3 interpolated temperature data sets from the RUC analyses were compared separately to the observed temperature data through the variance calculation below:

$$T_{BIAS} = T_{RUC} - T_{OBS} \quad ; \tag{1}$$

$$Var_{BIAS} = \frac{\sum_{i=1}^{N}\left[T_{BIAS_i} - \overline{T_{BIAS}}\right]^2}{N-1} \quad , \tag{2}$$

where,

T_{RUC} = RUC interpolated maximum or minimum temperature;

T_{OBS} = CON observed maximum or minimum temperature;

T_{BIAS} = RUC temperature bias;

N = number of days (10, 20, 30, 40 or 50) within the period of analysis;

Var_{BIAS} = sample variance of the RUC temperature bias.

Two different OBT estimates, one based on maximum temperature biases and one based on minimum temperature biases, were obtained. The observation schedule associated with the simulated time series that resulted in the lowest Var_{BIAS} determined the OBT estimate for the station. Often, these two OBT estimates were consistent and determined the final OBT estimate for the station (~ 80% consistency). When inconsistencies occurred between the two OBT estimates, most often the OBT estimate based on maximum temperature bias was the most accurate. As a result, OBT estimates were simply based on maximum temperature biases alone. A flowchart shows the procedure described above graphically (Figure 3) when distinguishing between 2 or 3 observation times.

b. Evaluating OBT Estimation Performance

The performance of the OBT estimation technique was evaluated by assessing the consistency between OBT estimates and the corresponding reported observation times available from the stations' metadata. The ranges of hours that define each of the 3 OBT categories are 0300 – 1100 LST (AM), 1200 – 2000 LST (PM) and 2100 - 0200 LST (MID). When the ability to distinguish between only 2 OBT categories was evaluated, the ranges of hours for PM and MID were combined to create a "non-AM" category.

Two standard performance measures (Wilks 1995) were used to numerically summarize the results of this technique in each region. The probability of detection (POD) is defined as the percentage of stations that had their reported observation times (according to the metadata) correctly estimated. The false alarm rate (FAR) is defined as the percentage of observation time estimates that were not consistent with the metadata. Other performance measures such as the Heidke or Kuiper's skill score have additional benefits over the POD and FAR when applied to meteorological forecast verification (particularly when forecasting rare events), but provide limited additional utility for this application. The POD and FAR provide specific information about separate skill characteristics that cannot be extracted from a single concise skill score.

c. Normalizing Results

Results presented in subsequent sections were adjusted to reflect a uniform distribution of observation schedules across all stations within a region. This was done since the FAR is sensitive to an unequal distribution of observation times. Since each of the analyzed regions had a different number of stations using each observation schedule (Table 1), normalizing the results allowed for a direct performance comparison among different regions and OBT categories.

To illustrate the adjustment technique, a contingency table is constructed for simple 2-category OBT definitions (Table 2). From this table, the POD and FAR are defined as

$$POD_{AM} = \frac{a}{M_{AM}}; \quad POD_{nonAM} = \frac{d}{M_{nonAM}}; \quad FAR_{AM} = \frac{c}{E_{AM}}; \quad FAR_{nonAM} = \frac{b}{E_{nonAM}} \quad , \tag{3}$$

where subscripts for *POD* and *FAR* are the OBT categories represented and all other variables refer to frequencies defined in Table 2. Adjustments to each of the frequencies in the numerators of Equation 3 were performed by finding the OBT category with the largest number of stations (based on metadata):

$$M = \max(M_{AM}, M_{nonAM}) \quad , \tag{4}$$

before making adjustments to frequencies in each OBT category based on a common value of M:

$$a_{ADJ} = a * \left(\frac{M}{M_{AM}}\right); \quad b_{ADJ} = b * \left(\frac{M}{M_{AM}}\right); \quad c_{ADJ} = c * \left(\frac{M}{M_{nonAM}}\right); \quad d_{ADJ} = d * \left(\frac{M}{M_{nonAM}}\right). \tag{5}$$

The POD for each category remains unchanged by these adjustments as the ratio of frequencies is preserved (i.e. $a/M_{AM} = a_{ADJ}/M$; $d/M_{nonAM} = d_{ADJ}/M$) owing to the adjusted row sums (e.g. $c_{ADJ} + d_{ADJ}$) equaling M in the denominator. Modifications to b and c (Eq. 5), as well as to E_{AM} (= $a_{ADJ} + c_{ADJ}$) and E_{nonAM} (= $b_{ADJ} + d_{ADJ}$), resulted in adjusted FARs that could be directly compared between OBT categories. A similar adjustment technique was employed when evaluating OBT estimates consisting of more than 2 categories.

4. Results

a. 2-Category Performance

The ability of this OBT estimation technique to distinguish between AM and non-AM observation schedules is presented in Table 3 for each region and season. A period of 30 days was chosen for presentation of results since it represents the median period length that was analyzed while also corresponding to the monthly resolution of reported observation times from CON observers. Sensitivity to period length is addressed in subsequent sections. Results in Table 3 reflect the median OBT estimation performance from nine separate tests that each utilize 30 sequential days. The location of each 30-day time period within the seasons was randomly chosen, but each period was required to be completely within the season of interest. While many of these test periods overlap to some degree within each season, the median results were representative of overall season performance as 30-day test periods were not confined to calendar months. It is apparent that results were quite good in most regions. The probability of detection is comparable to that achieved by Andsager and Kunkel (2002) on the monthly timescale. Using 30 days, this method outperforms the 4-week accuracy (both POD and FAR) achieved by the methods of Belcher and DeGaetano (2003). The weakest performance occurs in region WC, which contains numerous mountainous and coastal stations. The spatial distribution of stations with false alarms shows little spatial uniformity within each region (Fig. 4).

The sensitivity of these results to missing data was investigated by performing these tests on 30 *randomly-selected* days within each season rather than 30 *sequential* days. Performance using randomly-selected days was comparable to that presented in Table 3 for 2-category OBT estimates, as well as for results using 3-category OBT estimates in the subsequent section (not shown).

b. 3-Category Performance

The ability of this method to estimate OBT from the 3 most frequently used observation schedules (AM, PM and MID) was generally best in the Northeast and Midwest, and worst on the Gulf Coast and West Coast (Table 4). While MID estimates were comparable in accuracy to

AM estimates, the ability to correctly determine PM OBTs barely reached two-thirds for some instances in coastal regions. The performance in detecting these afternoon OBTs does show the most improvement however when additional temperature data (i.e. periods longer than 30 days) were used in all regions except GC (Fig. 5).

The false alarms for the 3-category test increased over those in the 2-category test since PM and MID were separate categories. Despite the increase in false alarms, there continued to be very little spatial uniformity in their locations within each region (not shown). Both morning and midnight false alarms were generally estimated as afternoon observation schedules, independent of season or region, while afternoon false alarms had an equal tendency towards occurring at stations with morning or midnight OBTs (Table 5). These biases are likely due to the proximity in time of each of the observation times. For instance, the three 24-hour observation "windows" (0700 – 0700 LST; 1700 – 1700 LST; 2400 – 2400 LST) each overlap by a specific number of hours. The PM window overlaps with the AM window by 14 hours and with the MID window by 17 hours. The AM and MID observation periods only overlap by 7 hours, during a period in the diurnal cycle when the maximum temperature is less likely to occur, particularly when daily solar heating is more dominant than temperature advection over a region. It is therefore expected that AM observation schedules would be falsely estimated as PM more frequently than they would be falsely estimated as MID.

c. Influence of using modified interpolation techniques

In previous sections, OBT estimates were obtained by comparing observed and RUC daily extreme temperature time series at nearby locations. Such a comparison ignored any differences between the station and its nearest model grid point, such as elevation and horizontal distance. It was possible that a decrease in model bias and an improvement in OBT estimates could result, particularly in mountainous regions, by accounting for these differences. Spatial and temporal interpolation techniques were incorporated into this analysis in order to determine their effect on OBT estimation accuracy. Subsequently, the method that utilizes these interpolation techniques is referred to as the *FULL interpolation method*. The method used in previous sections, without any interpolation, will be referred to as the *BASIC interpolation method.*

i. FULL interpolation

The *FULL* interpolation technique begins with adjustments for elevation differences between RUC grid points and stations. The 2-meter temperature at each RUC grid point is reflective of a temperature extrapolation to a “minimum” topography field used by the model (Benjamin et al. 2002). This is performed by RUC to give values that are more representative of valley locations in mountainous areas, where surface stations are usually located. Before any horizontal interpolation from RUC grid points to CON station locations was performed, 2-meter RUC temperatures were placed on a constant-height surface based on the elevation of the station of interest. This was achieved by first calculating the temperature lapse rate between the two closest isobaric levels of the RUC analysis that encompassed both the station and RUC grid point elevations. This lapse rate was then used to adjust the 2-meter temperature at each grid point based on the difference between the station elevation and each RUC “minimum” topography grid value. The end result is a RUC temperature field on a constant-height surface for each analyzed hour.

Next, daily maximum and minimum temperatures were simulated at RUC grid points by fitting a cubic spline to the RUC hourly temperature values to mimic data recorded at CON stations for AM, PM and MID observation schedules. Performing this interpolation between hourly values enabled a more accurate assessment of extreme daily temperatures since analyses were no longer restricted to only temperatures recorded at the top of each hour.

Finally, the simulated daily maximum and minimum temperature fields on the RUC grid were interpolated to the station latitude and longitude using a method of multiquadric interpolation (Nuss and Titley 1994). This resulted in 4 temperature time series at each CON station: 1 observed time series and 3 RUC time series (each based on a different observation time) that were adjusted to coincide with the station's location and elevation.

Employing *FULL* interpolation techniques resulted in a decrease in RUC model temperature bias, particularly in regions with spatially varying topography. Despite these improvements in the magnitude of the model bias, the consistency of the bias from one day to the next was relatively unchanged. Since the variability of model bias was similar between the *BASIC* and *FULL* methods, utilizing the more sophisticated and computationally-intensive technique had a very small influence on both 2- and 3-category OBT estimation performance. The only region that benefited from *FULL* interpolation by more than 1.5% was WC, where POD increased by 6% for 3-category PM estimates, but improved by only 1% for AM and MID.

ii. BASIC 3-hourly interpolation

Unfortunately, archived RUC analyses are only available since 1998 at various horizontal resolutions (60, 40 and 20 km). This short period of availability limits the usefulness of this technique. Some archived model analyses, such as the new North American Regional Reanalysis (NARR) from the National Center for Environmental Prediction (Mesinger et al. 2004), are available over longer time periods and thus could be used to estimate or verify historical observation time metadata (from 1970s to present). However, these analyses are only available on a 3-hourly basis. In order to test the effects of using this coarser temporal resolution, time series of RUC analyses were sub-sampled to simulate the same 3-hour availability as the NARR. *BASIC* interpolation techniques were employed with one exception: a cubic spline was used to interpolate between the 3-hourly temperature values and the resulting fit values used to determine the daily temperature extremes. The use of 3-hourly data had limited effects on the accuracy of 2-category OBT estimates. Thus, this technique provides a viable means of differentiating between AM and non-AM observation times in historical metadata. However the POD and FAR values associated with the 3-hourly data were consistently less favorable than the 1-hour values for 3-category OBT estimates. The errors introduced by this low-temporal-resolution data were less consistent and thus had detrimental effects on this technique of comparing the variance of maximum temperature bias. Decreases in 3-category POD were largest for MID estimates in each region, particularly in region MW (drop from 92% to 59% annually) and WC (drop from 81% to 48% annually). Increases in FAR were largest for PM estimates, where each region had at least a 15% increase over the entire annual cycle.

d. Results reflective of current station distribution

The results presented in previous sections are representative of OBT estimation performance for a uniform distribution of observation times among all stations within a region. While the POD is not dependent upon the distribution of observation times, the FAR is. For instance, if a region had 10 times more AM stations than MID stations, the FAR for OBT estimates would be much larger for MID estimates than it would be for AM estimates. Performance measures representative of uniform OBT distributions preserve the statistical integrity of comparisons between regions, seasons and observation times, but are unrepresentative of the true OBT distribution.

In practice OBT distributions within each region are very non-uniform (Table 1). This strongly influences false alarm rates for both 2-category and 3-category OBT estimates when the OBT estimation procedure is applied operationally. Since stations with morning observation schedules vastly outnumber those with other schedules, incorrectly estimated AM observation times result in a disproportionate number of PM and MID false alarms. On a percentage basis, this results in low false alarm rates for AM estimates and high false alarm rates for estimates of other OBTs. In the NE region for example, the current distribution of observation schedules results in lower FAR percentages for AM estimates (one-third of values in Table 3) and higher FAR percentages for non-AM estimates (twice that of values in Table 3).

e. Identifying shifts in temperature recording: A 4-category procedure

Reported temperatures that are shifted in time introduce problems into a dataset that are similar to those that result from incorrectly characterized OBTs. A daily maximum temperature time series that is shifted by an AM observer (AM-shifter) often resembles that of a station following a midnight observation schedule. Assuming that all of the AM observing stations maintain a constant OBT through the analysis period, the *BASIC* 3-category OBT estimation technique was applied to two maximum temperature time series at each AM station. The first time series represented the maximum temperatures correctly reported as occurring on the day of observation. The second time series was identical to the first, but shifted back in time by one day to simulate shifted observations. The OBT estimation routine classified an overwhelming majority of the shifted time series as MID. In Table 6, false alarms represent (unshifted) AM time series that were incorrectly classified as MID. Successfully detected shifts represent shifted time series that were classified as MID.

Although Table 6 indicates that shifted observations appear to follow a MID OBT schedule, this feature does not offer a means to distinguish between an AM-shifter and a station that simply follows a MID observation schedule. The time series of maximum temperature are often identical for these two cases. Minimum temperatures provide additional information that allows for this distinction, since minimum temperature time series are seldom shifted. Up to this point, minimum temperatures have not been utilized to estimate OBT using the 2- or 3-category technique. For each station with an estimated OBT of midnight (based on maximum temperature), an additional OBT estimate based on minimum temperature was now determined. Stations with consistent OBT estimates (both MID) were assumed to be stations that follow midnight observation schedules. Stations with inconsistent OBT estimates were assumed to have shifted maximum temperature data and follow AM observation schedules, even when OBT estimates from minimum temperature biases were PM. While non-intuitive, MID/PM OBT estimate combinations (resulting from T_{MAX}/T_{MIN} temperature biases) could be used along with MID/AM combinations to identify AM-shifters due to the infrequent appearance of the MID/PM combination during applications of the 3-category technique (<2% of all OBT estimate combinations were MID/PM). By contrast, MID/PM OBT estimate combinations characterized AM-shifter stations on a more frequent basis (e.g. 27% of the time in NE). Thus, the MID/PM combination corresponded to AM-shifters with a considerably higher frequency than it did to PM or MID observation schedules. MID/AM OBT estimate combinations characterized nearly 67% of AM-shifter stations in NE. An additional flowchart is provided to graphically represent the steps of using this 4-category technique (Figure 6).

The ability to correctly detect AM-shifters using these assumptions within the context of a new 4-category (AM, PM, MID, AM-shifter) OBT estimation procedure is given in Table 7. The inclusion of this fourth OBT category does not influence the POD for AM or PM observation schedules. Any small differences that do exist between these 3-category (Table 4a,b) and 4-category (Table 7a,b) PODs are due to the variability in the test periods that were randomly chosen. Substantial decreases in the POD of MID observation schedules occur in mountainous and coastal regions (Table 7c) when compared to the 3-category procedure (Table 4c), however detection ability remains high in the Northeast and Midwest. The rate of PM and MID false alarms increased by including an additional category (Table 7b,c) however a shifted AM OBT was seldom mistaken for an unshifted AM OBT, resulting in relatively small changes in the AM false alarm rate (Table 7a). Performance results of detecting shifted AM observations (Table 7d) are comparable to the other 3 OBTs included in this 4-category procedure.

5. Discussion

As presented during the evaluation of this technique, some OBT estimates were inconsistent with the corresponding reported observation times in digital metadata files. After closer inspection of these inconsistent estimates within two diverse regions during a particular season (57 stations in region NE and 108 stations in region WC during Autumn 2003), each station falls into one of three general categories:

1) The reported observation time in the metadata file was incorrect and could be verified by an inconsistency with the observation time reported in the publication of *Climatological Data* (October 2003). For the 2-category method, approximately 72% (41 out of 57) of the OBT estimates in region NE and 27% (29 out of 108) of OBT estimates in region WC

that were initially labeled as incorrect were consistent with OBTs reported in *Climatological Data*. For the 3-category method, about 10% fewer incorrect OBT estimates fall in this category.

2) Visual comparisons between the RUC simulated maximum temperature time series and observations from 2 nearby stations (1 "AM" and 1 "non-AM") confirm that the OBT estimates are likely correct (Fig. 1). Although subjective, this verification method confirmed about one-third of the OBT estimates that could not be verified by *Climatological Data*.

3) OBT estimation was likely incorrect given the *Climatological Data* OBT and visual comparison. About 15% of the OBT estimates that were initially labeled as incorrect in region NE fall in this category (45% in region WC).

Given the above results, it appears that rough estimates of error rates in the digitized metadata are around 6% for both of these regions. It is extremely difficult to assess a precise estimate of this error rate without knowing the error rate of observation times appearing in *Climatological Data*. It can be concluded, however, that metadata in *Climatological Data* was more reliable than the digitized metadata during the period of analysis.

The verification of OBT estimates that are inconsistent with metadata is very promising for this procedure's use in future applications. Enhancements of this procedure for operational use could combine OBT estimates with metadata from stations at which observation times are certain. These include stations that employ automated systems such as the U.S. National Weather Service's Automated Surface Observing Systems (ASOS), at which daily CON observation times are midnight. Incorporating this information would decrease false alarm rates for AM, PM and AM-shifted OBT estimates since none of the documented MID stations would be incorrectly classified (assuming automated stations will not return to manual observing practices). For 3-category estimates, AM and PM false alarms would be expected to decrease by about 20% and 50%, respectively (Table 5). For 4-category estimates, false alarms occur with the highest frequency in the WC region. FAR would decrease by about 10% for AM and PM estimates and nearly 30% for AM-shifted estimates in this region if unquestionable MID OBTs are utilized and assumed to be correct.

Various factors likely contribute to incorrectly estimated OBTs, including missing observations (especially during periods of high temperature variability), meteorological conditions that are highly persistent over the period of analysis, and the timing of the daily maximum temperature during the afternoon hours. The former two factors make it difficult to distinguish between observation times since periods with highly variable temperatures may be missing or non-existent. The latter factor is influential if a region often observes its maximum temperature late in the afternoon, since the afternoon observation window used for estimations in this technique was 1700 – 1700 LST. Slightly lower OBT simulation performance during summer months was likely a result of higher rates of meteorological persistence and later daily maximum temperature occurrence. Three additional afternoon observation schedules (1500, 1600 and 1800 LST) were included in the simulations along with the original observation schedules (0700, 1700 and 2400 LST) to test the influence of capturing variability in maximum temperature timing. This had a minimal effect on the overall performance, however both the POD and FAR for afternoon estimates increased slightly.

6. Summary

A new procedure used to estimate observation times at CON stations using Rapid Update Cycle model analyses and CON daily extreme temperature observations was presented (summarized in Figures 3 and 6). This procedure allows the user to distinguish between 2 (AM, non-AM), 3 (AM, PM, MID) or 4 (AM, PM, MID, AM-shifted) categories of observation times on a weekly to monthly timescale. Evaluation of the procedure over 30-day test periods in 5 regions across the U.S. produces the following results:

- The probability of correctly estimating observation times between 2 possible categories was found to be comparable with the performance of Andsager and Kunkel (2002), while high performance levels (> 90% accuracy) were reached in a much shorter time period (~30 days) when compared to those achieved by Belcher and DeGaetano (2003). Consistent with these previous studies, limitations in OBT estimation accuracy occurred in mountainous and coastal regions.

- Application of a 3-category OBT estimation system on such small timescales (~30 days) was not previously documented, to our knowledge. In these tests, afternoon and midnight observation times were distinguished from the non-AM observation times used in the 2-category tests. Estimates of midnight OBTs had a higher probability of detection than afternoon OBTs, but still less than that achieved by morning OBT estimates.

- While only maximum temperatures were employed for OBT estimation, minimum temperatures were also included to facilitate the identification of observers who temporally shift temperature records. This extended the 3-category procedure to 4-categories. The performance of detecting shifted observations is comparable to the performance achieved when classifying other OBT categories in the context of a 4-category system.

- Introducing horizontal interpolation of temperature to station locations and vertical adjustments of temperature to correct for elevation differences had almost no influence on the performance of the OBT estimation procedure in most areas evaluated, and only minimal influence (improvement) in mountainous regions. It is possible that greater improvements at stations in mountainous terrain may be achieved by using RUC analyses with higher vertical resolution (RUC hybrid levels). RUC hybrid levels are highly resolved (~10 mb) terrain-following coordinates near the model surface. Employing such levels would increase the likelihood of capturing inversions that occur frequently in such regions.

- Observation time estimates obtained using 3-hourly RUC analyses maintained relatively high levels of accuracy for 2-category OBT estimates, but decreased well below levels achieved with 1-hourly data when OBT was estimated from 3 categories. This implies that the 2-category OBT estimates can be extended back to early 1970s using archived model analyses of lower temporal resolution such as the NARR.

7. Acknowledgments

This work was supported by USDA CREES Grant VE2003-6193 and NOAA Grant NA17RJ1222. Thanks are extended to Jeff Arnfield of the National Climatic Data Center, Asheville, North Carolina for providing digitized metadata files used in evaluating our techniques.

REFERENCES

Andsager, K. and K. E. Kunkel, 2002: Estimated observation times for cooperative stations. Preprints, *13th Conf. On Applied Climatology*, Portland, OR, Amer. Meteor. Soc., 47-48.

Baker, D. G., 1975: Effect of observation time on mean temperature estimation. *J. Appl. Meteor.*, **14**, 471-476.

Belcher, B. N. and A. T. DeGaetano, 2003: A Method for Operational Detection of Daily Observation-Time Changes. *J. Appl. Meteor.,* **42**, 1823-1836.

Benjamin, S. G., J. M. Brown, K. J. Brundage, D. Dévényi, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, T. L. Smith, S. Weygandt and G. S. Manikin, 2002: RUC20 – The 20-km version of the Rapid Update Cycle. *NWS Technical Procedures Bulletin No. 490*, National Oceanic and Atmospheric Administration, U. S. Department of Commerce, 30 pp.

Blackburn, T., 1983: A practical method of correcting monthly average temperature biases resulting from differing times of observation. *J. Climate Appl. Meteor.*, **22**, 328-330.

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.

DeGaetano, A. T. and W. W. Knapp, 1993: Standardization of weekly growing degree day accumulations based on differences in temperature observation time and method. *Agric. For. Meteor.*, **66**, 1-19.

Janis, M. J., 2002: Observation-time-dependent biases and departures for daily minimum and maximum air temperatures. *J. Appl. Meteor.*, **41**, 588-603.

Karl, T. R., C. N. Williams Jr., P. M. Young and W. M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. *J. Climate Appl. Meteor.*, **25**, 145-160.

Mesinger, F. and coauthors, 2004: North American Regional Reanalysis. *15th Symposium on Global Change and Climate Variations*, Seattle, WA, 11-15 Jan 2004.

Nuss, W. A. and D. W. Titley, 1994: Use of Multiquadric Interpolation for Meteorological Objective Analysis. *Mon. Wea. Rev.,* **122**, 1611-1631.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 464 pp.

FIGURE CAPTIONS

Figure 1. Daily maximum temperature time series (September 2003) for three adjacent stations in California. Both Donner Memorial State Park, CA (042467) and Sagehen Creek, CA (047641) follow afternoon observation schedules, while Truckee Ranger Station, CA (049043) follows a morning observation schedule. Circled areas show a clear one-day difference in the reporting of temperatures at the morning station.

Figure 2. Map showing 5 regions across the U.S. used for assessment of OBT estimation performance.

Figure 3. Schematic characterizing the steps of the 2- and 3-category observation time estimation procedures. Dotted boxes indicate OBT estimates that result from each procedure.

Figure 4. Spatial distribution of false alarms resulting from the 2-category OBT estimation technique during summer (JJA 2003). Open circles represent AM false alarms and closed circles represent non-AM false alarms.

Figure 5. Dependence of OBT estimation performance on the number of days utilized in a) NE, b) GC, c) MW, d) RM and e) WC regions. Results represent annual averages.

Figure 6. Schematic characterizing the steps of the 4-category observation time estimation procedure (continuing from the 3-category results in Figure 3). Dotted boxes indicate OBT estimates that result from each procedure.

TABLE CAPTIONS

Table 1. The total number of stations analyzed in each region and the distribution of 3-category observation schedules among the stations of each region (as of March 2004).

Table 2. A 2-category contingency table used for defining the construction of POD and FAR, as well as the calculation of their adjusted (normalized) values. The frequencies in which OBT estimates are consistent (*'a'* and *'d'*) and inconsistent (*'b'* and *'c'*) with metadata reports make up the components of the contingency table. Column (E_{AM} and E_{nonAM}) and row (M_{AM} and M_{nonAM}) totals represent the number of estimates assigned to each category and the number of stations documented as following each observation schedule, respectively.

Table 3. Median 2-category performance results for various seasons (MAM, JJA, SON, DJF) and regions (NE, GC, MW, RM, WC). Results are provided for stations with (a) morning observation schedules and (b) observation schedules other than morning. Results are representative of performance when observation times at stations within a region are uniformly distributed across these 2 possible categories. Bold values signify POD ≥ 90 or FAR ≤ 10.

Table 4. As in Table 3, but for 3-category performance. Results are provided for (a) morning, (b) afternoon and (c) midnight observation schedules, and are representative of performance when observation times at stations within a region are uniformly distributed across these 3 observation schedules.

Table 5. Distribution of false alarm estimates among categories (% of total false alarms).

Table 6. The probability of characterizing shifted AM maximum temperature time series as MID (POD) and falsely characterizing unshifted AM maximum temperature time series as MID (FAR).

Table 7. As in Table 4, but for 4-category performance. Results are provided for (a) morning (unshifted), (b) afternoon, (c) midnight and (d) morning (shifted) observation schedules, and are representative of performance when observation times at stations within a region are uniformly distributed across these 4 observation schedules.

Table 1. The total number of stations analyzed in each region and the distribution of 3-category observation schedules among the stations of each region (as of March 2004).

	NE	**GC**	**MW**	**RM**	**WC**
Total Number of Stations	835	395	1220	512	797
	-----	-----	-----	-----	-----
AM schedules (% of total stations)	71.5	75.0	72.0	60.9	51.9
PM schedules (% of total stations)	13.2	12.4	16.1	31.7	36.9
MID schedules (% of total stations)	15.3	12.6	11.9	7.4	11.2

Table 2. A 2-category contingency table used for defining the construction of POD and FAR, as well as the calculation of their adjusted (normalized) values. The frequencies in which OBT estimates are consistent (*'a'* and *'d'*) and inconsistent (*'b'* and *'c'*) with metadata reports make up the components of the contingency table. Column (E_{AM} and E_{nonAM}) and row (M_{AM} and M_{nonAM}) totals represent the number of estimates assigned to each category and the number of stations documented as following each observation schedule, respectively.

	AM (Estimated)	**non-AM (Estimated)**	**Total (Metadata)**
AM (Metadata)	a	b	$M_{AM} = a+b$
Non-AM (Metadata)	c	d	$M_{nonAM} = c+d$
Total (Estimated)	$E_{AM} = a+c$	$E_{nonAM} = b+d$	

Table 3. Median 2-category performance results for various seasons (MAM, JJA, SON, DJF) and regions (NE, GC, MW, RM, WC). Results are provided for stations with (a) morning observation schedules and (b) observation schedules other than morning. Results are representative of performance when observation times at stations within a region are uniformly distributed across these 2 possible categories. Bold values signify POD ≥ 90 or FAR ≤ 10.

(a) Morning

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**96**	86	**97**	**94**	84	12	12	**9**	**4**	**7**
JJA	89	86	88	88	82	11	**10**	**8**	**6**	**9**
SON	**96**	**91**	**98**	**96**	88	13	**10**	**9**	**10**	15
DJF	**95**	**97**	**96**	89	81	13	11	11	**9**	13

(b) Non-Morning

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	88	89	**91**	**96**	**94**	**5**	14	**3**	**6**	15
JJA	89	**91**	**93**	**94**	**92**	11	14	12	11	16
SON	86	**90**	**91**	**90**	85	**4**	**10**	**3**	**5**	13
DJF	86	88	88	**91**	88	**6**	**3**	**4**	**10**	18

Table 4. As in Table 3, but for 3-category performance. Results are provided for (a) morning, (b) afternoon and (c) midnight observation schedules, and are representative of performance when observation times at stations within a region are uniformly distributed across these 3 observation schedules.

(a) Morning

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**96**	86	**97**	**94**	84	11	21	15	**5**	11
JJA	89	86	88	88	82	**9**	19	13	**7**	12
SON	**96**	**91**	**98**	**96**	88	12	18	15	14	20
DJF	**95**	**97**	**96**	89	81	13	20	19	12	18

(b) Afternoon

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	83	67	83	74	66	**4**	24	**5**	12	21
JJA	76	63	73	66	66	18	37	20	34	29
SON	79	67	83	83	72	**4**	15	**4**	12	22
DJF	86	65	76	75	66	**4**	**3**	**5**	27	40

(c) Midnight

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**96**	88	**94**	**95**	**90**	13	15	**6**	19	27
JJA	**92**	76	**90**	74	85	16	20	17	29	25
SON	**95**	**96**	**95**	**90**	87	12	13	**5**	**6**	13
DJF	**95**	**98**	**95**	80	69	**6**	12	**7**	17	24

Table 5. Distribution of false alarm estimates among categories (% of total false alarms).

	AM False Alarms		**PM False Alarms**		**MID False Alarms**	
Estimated as:	PM	MID	AM	MID	AM	PM
NE	81	19	50	50	5	95
GC	100	0	66	34	15	85
MW	75	25	58	42	13	87
RM	82	18	32	68	18	82
WC	90	10	43	57	28	72

Table 6. The probability of characterizing shifted AM maximum temperature time series as MID (POD) and falsely characterizing unshifted AM maximum temperature time series as MID (FAR).

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**96**	86	**97**	**97**	**90**	**2**	**4**	**1**	**1**	**6**
JJA	89	65	88	79	79	**3**	**2**	**3**	**3**	**3**
SON	**99**	**91**	**98**	**98**	88	**1**	**2**	**1**	**1**	**5**
DJF	**99**	**99**	**99**	80	66	**2**	**1**	**1**	**1**	**4**

Table 7. As in Table 4, but for 4-category performance. Results are provided for (a) morning (unshifted), (b) afternoon, (c) midnight and (d) morning (shifted) observation schedules, and are representative of performance when observation times at stations within a region are uniformly distributed across these 4 observation schedules.

(a) Morning

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**97**	86	**97**	**95**	83	**6**	23	15	**5**	11
JJA	**90**	86	87	87	82	**4**	19	13	**8**	12
SON	**98**	**90**	**98**	**95**	88	**8**	19	15	11	21
DJF	**97**	**96**	**98**	**90**	79	**9**	20	19	15	24

(b) Afternoon

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	86	58	83	74	64	**6**	37	**8**	17	32
JJA	78	67	73	64	64	23	51	30	45	43
SON	78	71	82	85	70	**4**	26	**6**	**10**	28
DJF	88	63	78	72	58	**4**	**8**	**4**	36	56

(c) Midnight

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**93**	70	83	75	70	**10**	25	**10**	23	35
JJA	87	60	72	41	48	19	26	19	45	44
SON	**95**	78	**90**	82	54	**7**	**6**	**6**	13	22
DJF	**97**	**96**	**92**	77	57	**8**	**7**	**4**	12	23

(d) Morning (shifted)

	POD					FAR				
Season	NE	GC	MW	RM	WC	NE	GC	MW	RM	WC
MAM	**90**	71	**91**	84	64	12	32	12	26	39
JJA	77	51	78	55	53	21	32	27	53	51
SON	**97**	87	**96**	**90**	78	13	21	**7**	16	36
DJF	**96**	63	**98**	79	61	**1**	11	**5**	17	32

Figure 1. Daily maximum temperature time series (September 2003) for three adjacent stations in California. Both Donner Memorial State Park, CA (042467) and Sagehen Creek, CA (047641) follow afternoon observation schedules, while Truckee Ranger Station, CA (049043) follows a morning observation schedule. Circled areas show a clear one-day difference in the reporting of temperatures at the morning station.

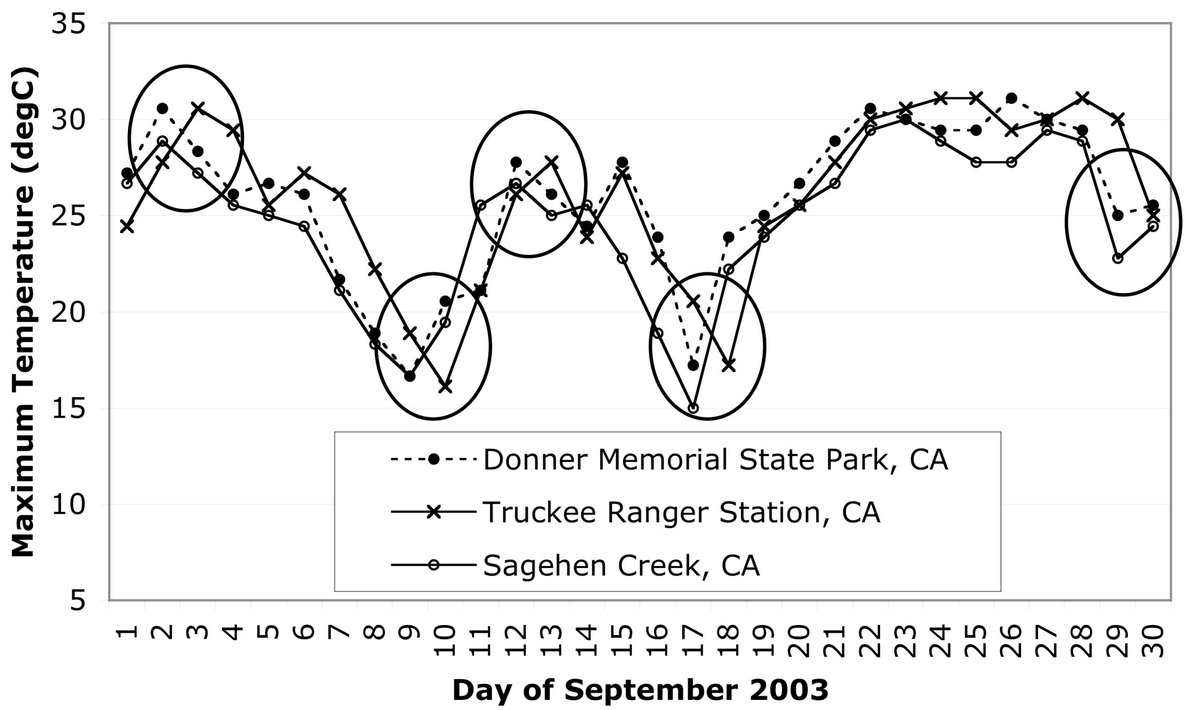

Figure 2. Map showing 5 regions across the U.S. used for assessment of OBT estimation performance.

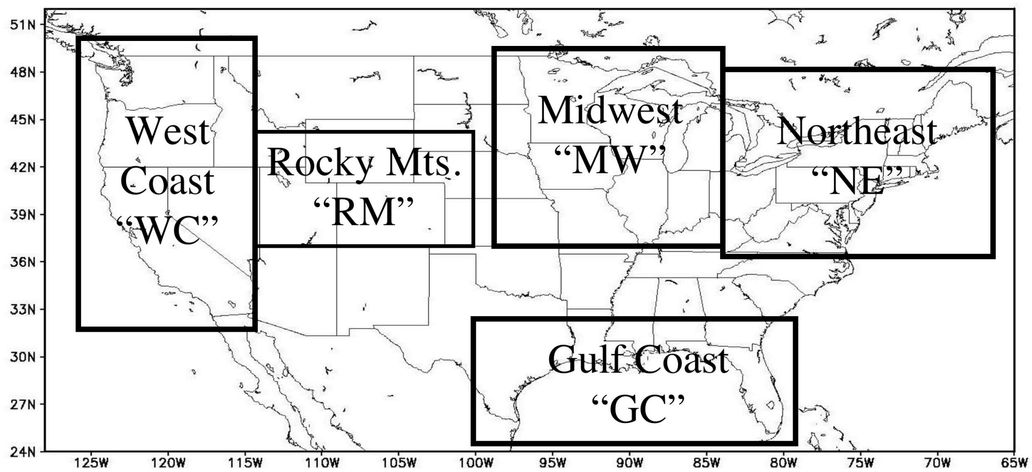

Figure 3. Schematic characterizing the steps of the 2- and 3-category observation time estimation procedures. Dotted boxes indicate OBT estimates that result from each procedure.

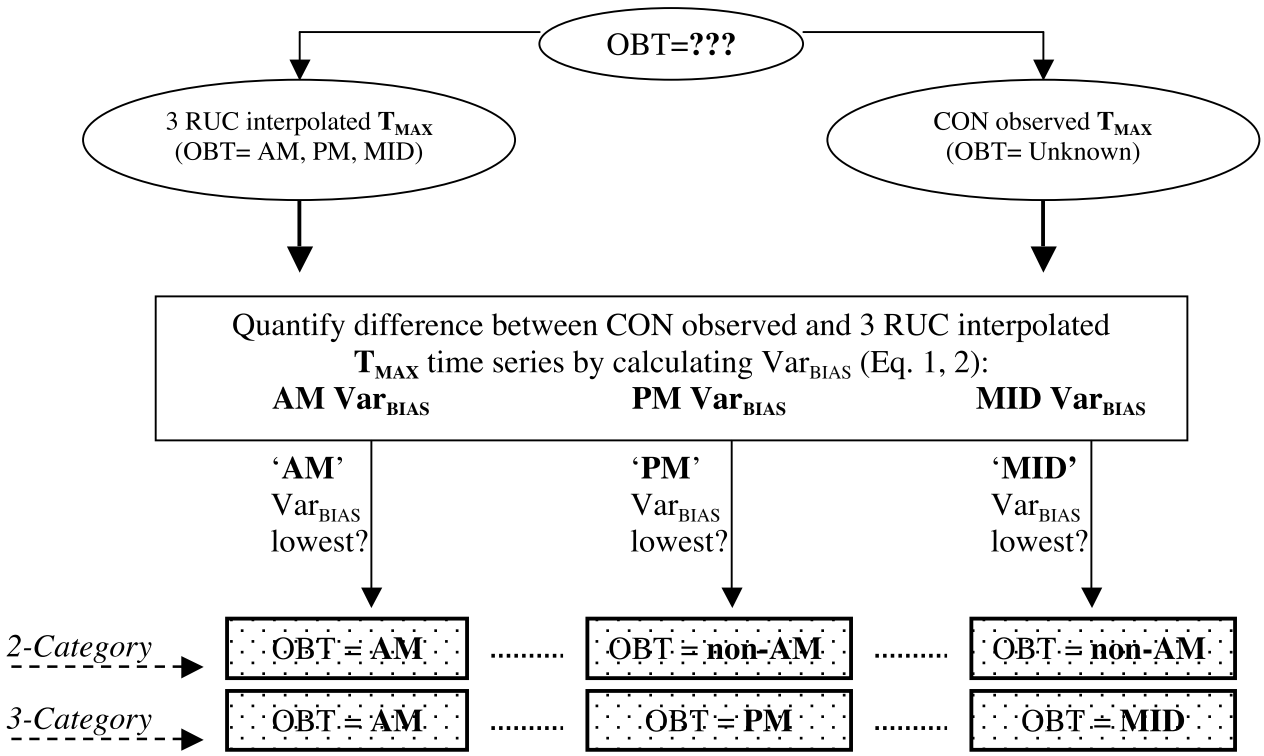

Figure 4. Spatial distribution of false alarms resulting from the 2-category OBT estimation technique during summer (JJA 2003). Open circles represent AM false alarms and closed circles represent non-AM false alarms.

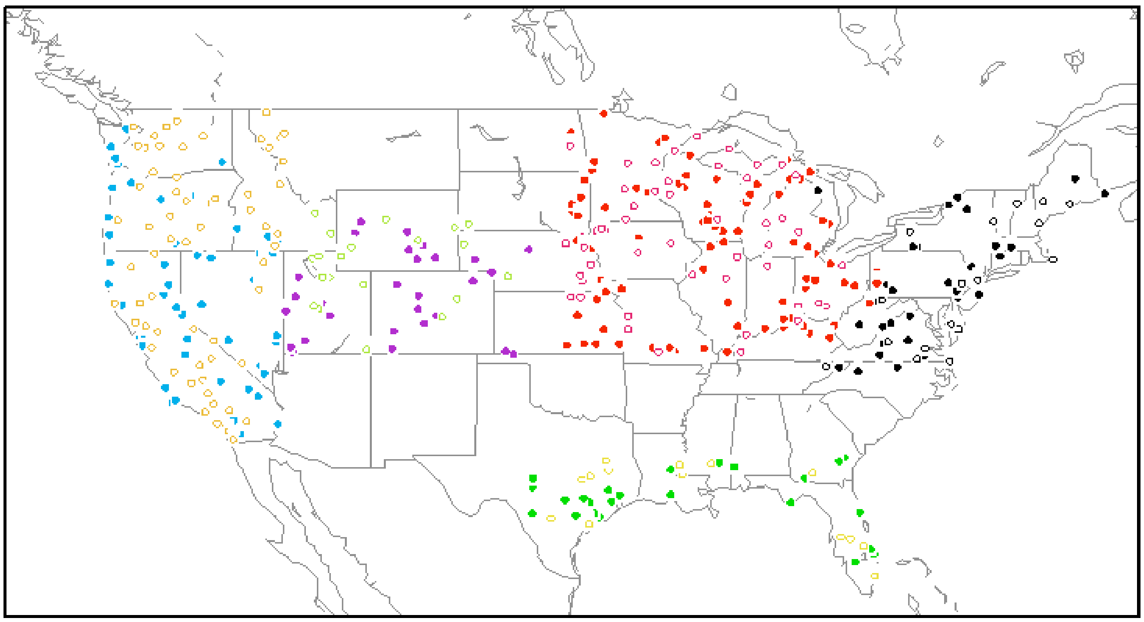

Figure 5. Dependence of OBT estimation performance on the number of days utilized in a) NE, b) GC, c) MW, d) RM and e) WC regions. Results represent annual averages.

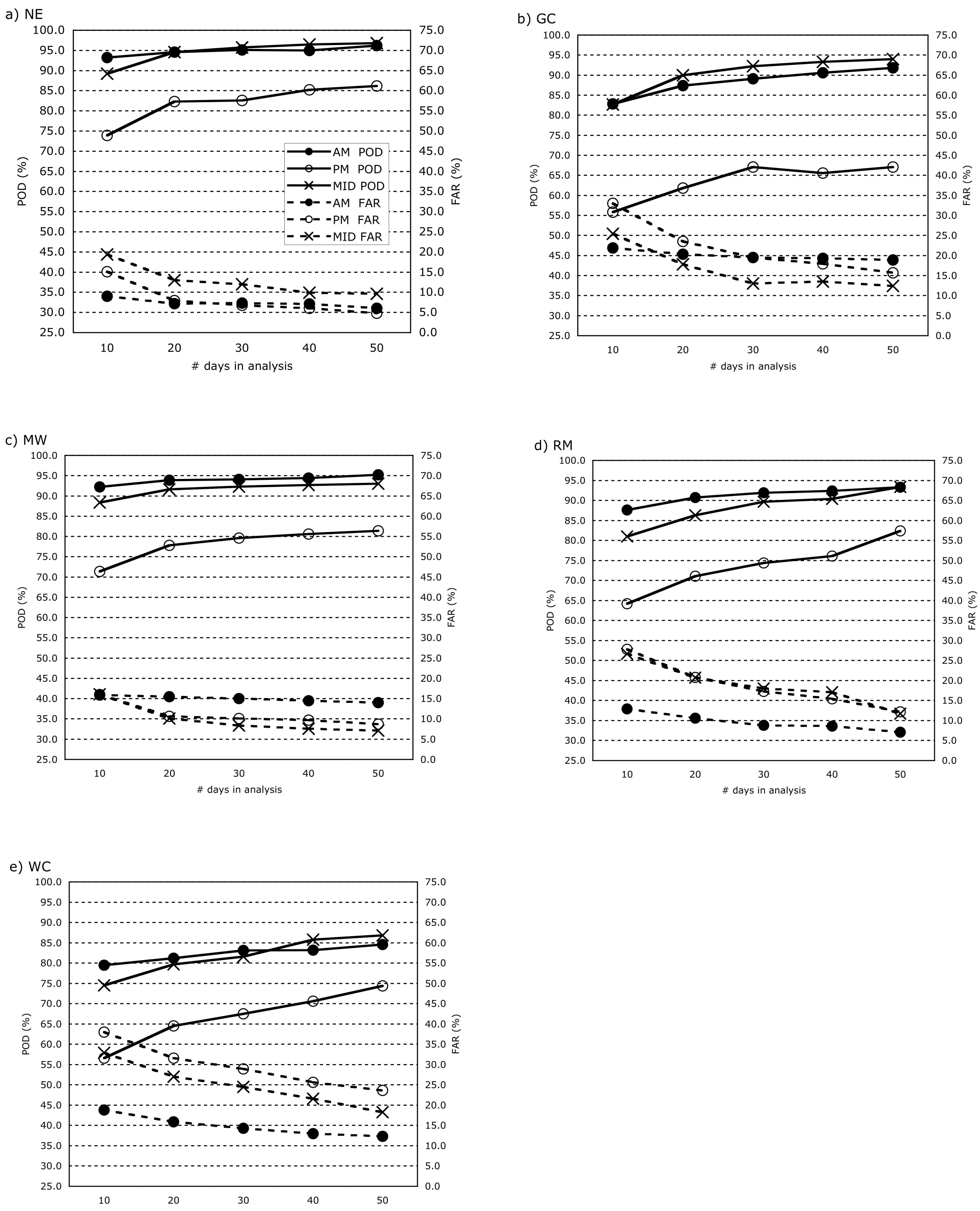

Figure 6. Schematic characterizing the steps of the 4-category observation time estimation procedure (continuing from the 3-category results in Figure 3). Dotted boxes indicate OBT estimates that result from each procedure.