Oyler Jared (Orcid ID: 0000-0002-9676-6734)

Time of Observation Adjustments to Daily Station Precipitation May Introduce Undesired Statistical Issues

Authors: Jared W. Oyler[1*], Robert E. Nicholas[1]

Short Title: Time of Observation Adjustments to Daily Precipitation Observations

Keywords: time of observation, precipitation observations, extremes, GHCN-D, NLDAS-2

Sponsors: National Science Foundation (NSF)

National Oceanic and Atmospheric Administration (NOAA)

Affiliations:

[1]Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania, USA.

[*]Correspondence to:

The Pennsylvania State University

Earth and Environmental Systems Institute

The Network for Sustainable Climate Risk Management

2217 EES Building

University Park, PA 16802

Email: jared.oyler@psu.edu

Phone: 215-260-4487

Abstract

Interstation observations of daily precipitation are often temporally misaligned due to differences in station time of observation. Several time of observation adjustment methods have historically been applied to improve interstation temporal alignment, but the efficacy of such adjustments has not been fully tested. Here, we examine the ability of several time of observation adjustments to improve observation compatibility under three different adjustment scenarios: adjusting morning observations to midnight, adjusting afternoon observations to midnight, and adjusting afternoon observations to morning. We find that all adjustment methods provide necessary improvements to the temporal alignment of daily precipitation observations, especially with respect to morning versus midnight time of observation totals. However, for a majority of the adjustments, improved temporal alignment comes at the cost of significantly altering observed precipitation intensity, frequency, and extremes. We also find adjustments have the potential to overcorrect and increase both general and extreme event spatiotemporal coherence.

1 Introduction

Rain gauge-based observations of daily precipitation are a key input for a broad range of hydrological and environmental analyses. Due to operational requirements and reporting practices, daily precipitation observations from various stations and networks often have different times of observation. Station time of observation is the ending accumulation time of the 24-hour period for a reported daily precipitation total. For instance, a daily total with a 0700 LT observation time (i.e. 7:00 am local time) is the total from 0700 LT on the previous day to 0700

LT on the reporting day. Time of observation inconsistencies cause interstation temporal misalignment in reported daily precipitation totals. This misalignment affects interstation data compatibility thus limiting the applicability of gauge-based observations for spatiotemporal analyses and the development of gridded products (Hopkinson *et al.*, 2011).

Time of observation inconsistencies are a well-known issue, but the efficacy of different adjustment methods has not been fully tested. Time of adjustment methods have ranged from shifting morning observation precipitation back a calendar day (Holder *et al.*, 2006) to disaggregating daily precipitation totals to hourly and then aggregating back to a different daily observation time (Holder *et al.*, 2006; Kim and Pachepsky, 2010; Maurer *et al.*, 2002). Disaggregation can be performed using actual hourly observations (Holder *et al.*, 2006) or under the assumption that a daily total is distributed uniformly across all hours in a 24-hour period (Kim and Pachepsky, 2010; Maurer *et al.*, 2002). While time of observation adjustment methods have been shown to improve interstation temporal correlations (Holder *et al.*, 2006; Kim and Pachepsky, 2010), such adjustments have only been tested under the context of adjusting morning observations to midnight and it is unclear how effective they are under different adjustment scenarios (e.g. afternoon to morning). Additionally, it is unknown how time of observation adjustments impact observed precipitation summary statistics (e.g. average intensity) and the coherence of interstation precipitation observations in both space and time. The objective of this study is to determine the scenario-specific effects of time observation adjustments on (1)

the temporal alignment of daily precipitation time series with different observation times; (2) observed precipitation statistics; and (3) interstation spatiotemporal coherence.

2 Materials and Methods

2.1 Time of Observation Adjustment Methods

We examine three time of observation adjustment methods: the 1-day *shift* method of Holder *et al.* (2006), disaggregation-reaggregation under the *uniform* distribution assumption (Kim and Pachepsky, 2010; Maurer *et al.*, 2002), and disaggregation-reaggregation with *hourly* observations from a gridded dataset.

2.1.1 Shift Method

The shift adjustment method shifts all daily totals with a morning observation time back one calendar day to better conform them with midnight-to-midnight and afternoon/evening observation times. The shift method assumes that, because a majority of the 24-hour period for a morning observation time is actually in the previous calendar day (Figure 1), a simple backward shift of the daily totals will improve temporal alignment with midnight-to-midnight totals (Holder *et al.*, 2006). The shift method can also be applied in reverse where midnight-to-midnight and afternoon/evening observation totals are shifted forward one calendar day to improve alignment with morning observation times.

2.1.2 Uniform Method

To adjust a reported precipitation total for a specific date to a different time of observation, the uniform assumption disaggregation-aggregation method reapportions reported daily totals from a three-day moving window surrounding the target date:

$$P_d^t = (P_{d-1} \cdot F_{d-1}) + (P_d \cdot F_d) + (P_{d+1} \cdot F_{d+1}) \quad (1)$$

where P_d^t is the adjusted estimate for the target time of observation on target date, d; P_{d-1}, P_d, and P_{d+1} are the original reported daily totals for the previous, target, and next dates, respectively; and F_{d-1}, F_d, and F_{d+1} are the fractions of P_{d-1}, P_d, and P_{d+1}, respectively, to include in the estimate of P_d^t. Because the uniform method assumes that a reported daily total is distributed uniformly across all hours within its respective 24 hour period, F_{d-1}, F_d, and F_{d+1} are determined directly by the number of hours of overlap between the 24-hour periods represented by P_{d-1}, P_d, and P_{d+1}, and the new P_d^t (Figure 1).

2.1.3 Hourly Method

The hourly time of observation method also uses equation (1) to estimate midnight-to-midnight precipitation totals. In contrast to the uniform method, the hourly method does not just use overlapping hours to determine static values for F_{d-1}, F_d and F_{d+1}. Instead, the three Fvalues are daily varying and based on both overlapping hours and corresponding hourly precipitation observations. First, hourly observations are aggregated to the 24-hour time periods defined by the original time of observation. Second, the hourly observations are divided by their

corresponding 24-hour aggregations to produce hourly fraction values that represent the fraction of the 24-hour totals that fell in each hour. Values for F_{d-1}, F_d and $F_{d+1}$are then the sum of hourly fractions for the corresponding hours of overlap between P_{d-1}, P_d, and P_{d+1}, and the new P_d^t (Figure 1). If hourly observations have direct correspondence with the daily totals, the hourly method will result in a perfect correction.

Because hourly observations must be available at the station being adjusted or from a neighboring station, the hourly method does not have wide applicability. To address this limitation, we examine the application of hourly precipitation data from the North American Land Data Assimilation System Phase 2 (NLDAS-2; Xia *et al.*, 2012). NLDAS-2 provides 1/8th degree resolution (~12 km) hourly precipitation from 1979 to present for a region surrounding the conterminous U.S. NLDAS-2 daily precipitation is derived from a topographically adjusted (Daly *et al.*, 2008) version of the Climate Prediction Center (CPC) unified daily gauge analysis (Chen *et al.*, 2008). To obtain hourly precipitation totals, NLDAS-2 temporally disaggregates the CPC-derived daily totals using one of four different hourly datasets dependent on data availability and the following priority order: hourly radar observations (Lin and Mitchell, 2005), microwave satellite observations (Joyce *et al.*, 2004), gridded hourly gauge-based observations (Higgins *et al.*, 1996), and reanalysis (Mesinger *et al.*, 2006). The disaggregation data source used for a specific day and grid cell is not available in the NLDAS-2 metadata, but hourly gauge-based data are mainly used from 1979 to July 1996 and hourly radar observations from July 1996 to present (Ferguson and Mocko, 2017). Because temporal inconsistencies in NLDAS-2 have

been associated with the switchover from gauge to radar-based disaggregation (Ferguson and Mocko, 2017), we focus our use of NLDAS-2 on the period of radar-based disaggregation from 1997 to present. When applying the hourly adjustment to a station, we extract NLDAS-2 hourly precipitation from the closest grid cell to the station location.

2.2 Validation

2.2.1 Overview

We validate the adjustment methods in three main areas: temporal alignment, summary precipitation statistics, and spatiotemporal coherence. For temporal alignment, we examine how well the different adjustment methods improve alignment between daily precipitation time series with different observation times. In the second validation, we analyze if and by how much the adjustments modify observed precipitation statistics like average precipitation amount and intensity. An effective adjustment should not only improve temporal alignment, but produce a daily precipitation time series whose summary precipitation statistics are similar to those of the target time of observation time series. Lastly, we examine how the adjustments modify interstation spatiotemporal coherence. A reliable adjustment method should modify station time series with disparate times of observation, so that, when viewed as whole, they better represent the true spatiotemporal coherence of the precipitation field. In all analyses, we validate the adjustment methods in a region surrounding the U.S. Mid-Atlantic (32.0° to 45.25°N; 71.4° to 84.0°W; Figure 2) under three different adjustment scenarios: (1) adjustment of daily

precipitation from a morning to midnight-to-midnight observation time (AM' midnight); (2) adjustment of daily precipitation from an afternoon to midnight-to-midnight observation time (PM' midnight); and (3) adjustment of daily precipitation from an afternoon to morning observation time (PM' AM).

2.2.2 Temporal Alignment

To validate temporal alignment, we use high-quality 2006 to 2015 hourly precipitation observations from U.S. Climate Reference Network (USCRN) stations (Diamond *et al.*, 2013) within the U.S. Mid-Atlantic spatial domain (n=15; Figure 2a) . The Mid-Atlantic USCRN stations are spatially sparse, but allow for the straightforward creation of co-located daily time series with different times of observation. For each USCRN station, we create three 2006 to 2015 time series of daily 24-hour aggregated precipitation totals to encompass the AM' midnight, PM' midnight, and AM' PM adjustment scenarios: midnight-to-midnight aggregations, 0700-to-0700 LT aggregations, and 1700-to-1700 LT aggregations. We select 0700 and 1700 LT to represent AM and PM observations, respectively, because they are the most common morning and afternoon observation times used historically. To examine impacts on temporal alignment, we compare the Pearson correlation coefficient (r), Heidke skill score (HSS), and normalized mean absolute error (NMAE) between the three daily time series at each station both before and after adjustment. Here, we define NMAE as the ratio of mean absolute error to mean daily precipitation. We use r and NMAE to quantify temporal alignment of daily precipitation amounts and HSS to quantify alignment of precipitation occurrence.

We also validate temporal alignment with the tail dependence measure, χ (Coles *et al.*, 1999). In contrast to *r*, which quantifies the correlation of two station time series across their entire joint distribution, χ quantifies dependence in the tail of their joint distribution above a set quantile, *u* (Weller *et al.*, 2012). As such, χ allows for examination of how the time of observation adjustments affect the temporal alignment of extreme events. Under a common marginal assumption, for two X and Y time series, χ at quantile *u* is defined as:

$$\chi(u) = P(Y > u \mid X > u) \quad (2)$$

We calculate χ using the *R* (R Core Team, 2016) *extRemes* package (Gilleland and Katz, 2016), which uses the χ formulation of Reiss and Thomas (2007). We follow Weller *et al.* (2012) and set u= 0.95. The χ(0.95) parameter for two time series can be generally interpreted as the probability that one time series observation is extreme given that the corresponding observation in the other time series is extreme (Coles *et al.*, 1999).

2.2.3 Summary Statistics

As in the temporal alignment analysis, we use 2006 to 2015 daily aggregations of hourly observations from USCRN stations to validate summary precipitation statistics. We calculate and compare 5 summary precipitation statistics before and after adjustment of the daily USCRN time series: average precipitation amount, intensity, frequency, extremes, and lag 1 temporal autocorrelation (r_{lag1}). For instance, under the AM' midnight adjustment scenario, we compare the average intensity of a specific station's AM time series with that of its corresponding midnight time series both before and after the AM time series is adjusted. Average precipitation

amount is the mean precipitation value over all wet and dry days. Precipitation intensity is the mean precipitation value over wet days only. Frequency is the total number of wet days over the 2006 to 2015 time period. Our extreme statistic is defined as the mean of observed values > the 95th percentile.

2.2.4 Spatiotemporal Coherence

A reliable analysis of daily precipitation spatiotemporal coherence requires a sufficient station density. Because the Mid-Atlantic USCRN stations are spatially sparse, we instead use daily precipitation observations from the denser Global Historical Climatology Network-Daily (GHCN-D; Menne *et al.*, 2012) for the spatiotemporal coherence analysis (Figure 2b). To match the radar era of the NLDAS-2 dataset (see Section 2.1.3), we conduct the analysis over the 1997 to 2015 time period.

The spatiotemporal coherence validation also requires stations that have accurate time of observation metadata for each daily observation and maintain a relatively consistent time of observation from 1997 to 2015. To meet these criteria, we apply a strict series of requirement filters to the GHCN-D stations. First, we require each GHCN-D station to have at least 10 years of observations in each month over the 19-year 1997 to 2015 time period. A total of 1361 Mid-Atlantic GHCN-D stations meet this requirement. Second, because time of observation metadata can sometimes be unreliable (DeGaetano, 1999, 2000), we quality assure each station's metadata using the hourly NLDAS-2 data from the grid cell closest to each station location. For each station and year, we create 24 NLDAS-2 daily time series corresponding to each time of

observation hour. We then predict the time of observation for a specific station year by finding the daily NLDAS-2 time series that correlates most closely with the station's observations for the year. For each station, we compare the predicted times of observation with those in the station's metadata and only keep those stations whose mean absolute error between predicted and recorded times of observation is d 4 hours. A total of 1208 GHCN-D stations meet our time of observation quality assurance requirement. For missing time of observation metadata (10.3% of remaining daily observations), we use the nearest non-missing time of observation in a specific station's record. Next, to better ensure time of observation consistency at each station, we drop 93 stations whose time of observation changes by more than 4 hours over the 1997 to 2015 time period. In a final step, we classify remaining stations (n=1115) as AM, PM, midnight or "other" observers based on each station's most frequent time of observation. To maintain consistency with USCRN-based analyses, we limit the AM classification to those stations with an exact 0700 LT time of observation (n=510 stations) and the midnight classification to stations with an exact 2400 LT time of observation (n=223 stations). Due to a limited number of stations with an exact 1700 LT time of observation (n=28 stations), we expand the PM classification to stations with a time of observation between 1600 and 2000 LT (n=63 stations), which is the common 4-hour period used for PM classification in previous analyses (e.g. DeGaetano, 2000). This means that the PM classification has slightly greater time of observation variability than AM and midnight classifications, but the 4-hour classification window is necessary to ensure reliable quantification of spatiotemporal coherence. We classify all stations outside the AM, PM, and midnight

classifications as "other" (n=319) and remove them from the analysis. To establish a balanced number of observations in each classification, we only keep AM and midnight stations that are nearest neighbors to the 63 PM stations. In the end, this gives us 189 input stations for the spatiotemporal coherence validation, 63 in each time of observation classification (Figure 2b).

Using the 189 GHCN-D stations, we quantify spatiotemporal coherence both in terms of the full distribution of precipitation values and extremes by calculating separate decorrelation lengths (*x0*) for r and $\chi(0.95)$. The *x0* metric is defined as the spatial distance at which the dependence parameter (r or $\chi(0.95)$) between pairs of station time series decays to a threshold value of $1/e$ (Briffa and Jones, 1993; Gervais *et al.*, 2014; Osborn and Hulme, 1997). To quantify *x0* for a set of stations, temporal r or $\chi(0.95)$ is first calculated for all unique station pairs. Average r or $\chi(0.95)$ is then plotted as a function of distance between station pairs and a nonlinear function is fit to the resulting decay curve to determine *x0* (e.g. Figure 3). Although an exponential decay function is traditionally used in this regard (Briffa and Jones, 1993; Gervais *et al.*, 2014; Osborn and Hulme, 1997), we use an exponential variogram model because we found it to produce much better fits (Figure 3).

Using the *x0* metric, we perform two separate spatiotemporal coherence validations for each adjustment method and scenario. We perform the validations for both r and $\chi(0.95)$. In the first validation, we ask: how does the adjustment change the spatiotemporal coherence of stations with the different observation times in the scenario? For instance, in the AM' midnight scenario, what is the *x0* value for AM-midnight station pairs both before and after adjustment of

the AM stations? We assess whether the adjustment-driven change in *x0* is an improvement by calculating the percent error in *x0* for AM-midnight station pairs relative to the *x0* values for midnight-midnight station pairs before and after adjustment of the AM stations. We refer to this validation as the "target *x0*" validation, because it examines the ability of the adjustment to match the spatiotemporal coherence of stations with the target time of observation. In the second validation, we ask: how does the adjustment change the spatiotemporal coherence of stations with the original observation time in the scenario? For instance, in the AM' midnight scenario, what is the *x0* value for AM-AM station pairs both before and after adjustment of the AM stations? An effective adjustment should have little impact on the spatiotemporal coherence of stations with the original time of observation. We refer to this validation as the "original *x0*" validation, because it examines the ability of the adjustment to maintain the spatiotemporal coherence of stations with the original time of observation.

3 Results and Discussion

3.1 Temporal Alignment

Before adjustment, the different USCRN daily precipitation aggregations displayed various degrees of temporal alignment (Figure 4). Unadjusted AM' midnight had the weakest correlation (r = 0.37; Figure 4a) while correlations for PM' midnight (r = 0.78) and PM' AM (r = 0.70) were stronger (Figures 4b and 4c). These differences can be attributed to the number of hours of overlap between the different 24-hour time periods: AM (i.e. 0700 LT) and midnight

observations only have 7 hours of overlap on a specific date compared to 17 hours for PM (i.e. 1700 LT) and midnight, and 14 hours for PM and AM (Figure 1).

The ability of the different adjustment methods to improve temporal alignment was also variable. The NLDAS-2 hourly method produced the greatest and most consistent increases in temporal alignment across the three adjustment scenarios (Figure 4). With the NLDAS-2 hourly adjustment, temporal alignment parameters for AM' midnight, PM' midnight, and PM' AM all improved to similar, strong values despite pre-adjustment differences (Figure 4). While not to the extent of the NLDAS-2 hourly method, the uniform and shift adjustment methods also produced a clear improvement in AM' midnight temporal alignment (Figure 4a). However, the uniform and shift adjustments produced inconsistent improvements in the temporal alignment of the other adjustment scenarios and, in some cases, were detrimental (Figures 4b and 4c). Most notably, the uniform method weakened frequency temporal alignment for PM' midnight (unadjusted HSS = 0.72; uniform adjusted HSS = 0.64; Figure 4b) and the shift adjustment greatly weakened the overall temporal alignment of PM' AM (Figure 4c). Unlike AM' midnight, where a simple shift increases the number of hours of overlap on a specific date, a shift of PM' AM actually decreases the number of hours of overlap from 14 to 10 (Figure 1).

3.2 Summary Statistics

Summary statistics for the three different USCRN daily precipitation aggregations were nearly identical before adjustment (Figure 4). Absolute value percent error in average amount, intensity, frequency, and extremes was d 1.15% for all three adjustment scenarios (Figure 4) and

absolute percent error in r_{lag1} was < 10%. The shift adjustment also had little impact on precipitation statistics (Figure 4).

The uniform adjustment produced large changes in observed summary statistics (Figures 4 and 5). Across all scenarios, the uniform adjustment artificially decreased precipitation intensity (percent error = -33 to -34%) and extremes (percent error = -21 to -25%), and increased precipitation frequency (percent error = +50 to +51%) and temporal autocorrelation (r_{lag1} percent error = +282 to +368%; Figure 5). In absolute terms, the exceedingly large percent increases in temporal autocorrelation correspond to a +0.31 to +0.41 overestimation of r_{lag1} . For all times of observation, mean r_{lag1} ranges from 0.11 to 0.12 for unadjusted time series whereas mean r_{lag1} for uniform-adjusted time series ranges from 0.43 to 0.53. The changes in precipitation statistics introduced by the uniform method are due to the reapportioning of daily precipitation totals. Because the uniform method assumes an equal distribution of precipitation across every hour, it will always reapportion a single day total to two days. This doesn't change the average precipitation amount, but, as evident in the results here, it will increase frequency and temporal autocorrelation, and decrease intensity and extremes (Figures 4 and 5).

While not as large as the uniform adjustment errors, the NLDAS-2 hourly adjustment also consistently decreased precipitation intensity (percent error = -15 to -17%) and increased precipitation frequency (percent error = +18 to +21%; Figure 5). This is likely due to a scale mismatch where the hourly NLDAS-2 data conform the local point-based station observations to the hourly precipitation spatial patterns of the ~12-km NLDAS-2 grid. In grid-based datasets like

NLDAS-2, grid cell precipitation values are typically areal averages. As such, compared to point-based observations, hourly spatiotemporal variability will tend to be smoother with decreased intensity and increased frequency (e.g. Tustison *et al.*, 2001). At USCRN station locations, 2006 to 2015 average NLDAS-2 hourly precipitation intensity is -34% less than station-observed intensity, and NLDAS-2 hourly frequency is +44% greater than station-observed frequency. While not to the extent of the uniform method, this will cause NLDAS-2 to often reapportion single day totals to two days and decrease daily intensity and increase daily frequency.

3.3 Spatiotemporal Coherence

Before adjustment, pairs of GHCN-D stations with different times of observations displayed negative biases in interstation spatiotemporal coherence across all adjustment scenarios (Figure 5). Percent errors in target *x0* for correlation ($x0_r$) and target *x0* for tail dependence ($x0\chi$) were most negative for unadjusted AM-midnight station pairs (target $x0_r$ = -90%; target $x0\chi$ = -100%; Figure 5). The percent error of -100% for target $x0\chi$ indicates that the average $\chi(0.95)$ tail dependence between unadjusted AM-midnight station pairs never exceeded the $1/e$ decorrelation threshold at any distance. Percent errors for target $x0_r$ and $x0\chi$ for unadjusted PM-midnight and PM-AM station pairs were of smaller magnitude, but still consistently negative (target $x0_r$ = -15 to -24%; target $x0\chi$ = -27 to -44%; Figures 5b and 5c).

The NLDAS-2 hourly adjustment provided the greatest and most consistent improvement in target $x0_r$ and $x0\chi$ percent error (Figures 4 and 5). After application of the NLDAS-2

adjustment, target $x0_r$ and $x0\chi$ absolute value percent error fell below 3% across all three adjustment scenarios (Figure 4). However, the NLDAS-2 adjustment did produce a +5 to +11% error in original $x0_r$ and $x0\chi$ (Figure 5) suggesting that the adjustment has the potential to artificially increase interstation spatiotemporal coherence for stations with the original time of observation. We examine this issue in more detail in a seasonal analysis (Section 3.4).

The uniform adjustment also notably reduced percent error in target *x0* spatiotemporal coherence metrics, but exhibited several keys differences from the NLDAS-2 hourly adjustment (Figures 4 and 5). The ability of the uniform adjustment to reduce target $x0\chi$ percent error was not as substantial as the NLDAS-2 hourly adjustment. With the uniform adjustment, target $x0\chi$ percent error remained consistently negative ranging from -18 to -26% across the three adjustment scenarios (Figure 5). The uniform adjustment also produced larger magnitude positive percent error in original $x0_r$ than NLDAS-2, but did not produce a consistent positive bias in original $x0\chi$ (Figure 5). For stations with the original time of observation, this suggests that the uniform adjustment has a greater potential to artificially increase spatiotemporal coherence with respect to interstation correlation, but not with respect to extreme events.

In contrast to the NLDAS-2 and uniform adjustments, shift adjustment improvements to target *x0* spatiotemporal coherence were limited to the AM' midnight adjustment scenario. For this scenario, the shift adjustment reduced percent error for target $x0r$ and $x0\chi$ to -13% and -24%, respectively. The shift adjustment had no effect under the PM' midnight scenario because it is not applied when adjusting PM to midnight (Figures 4b and 5b). Similar to the temporal

alignment results, the shift adjustment increased the percent error in target $x0r$ and $x0\chi$ for the PM' AM scenario (Figures 4c and 5c). Nonetheless, one advantage of the shift adjustment over the uniform and NLDAS-2 adjustments was that it had very little impact on the spatiotemporal coherence of stations with the original time of observation ($x0r$ and $x0\chi$ percent error ≅ 0%; Figure 5). This is because the shift adjustment does not partially reapportion daily totals like the other adjustment methods. Similarly, for unadjusted observations, original $x0_r$ and $x0_x$ percent errors were all 0% because no adjustment is yet applied to the stations with the original time of observation (Figure 5).

3.4 Seasonal Analysis

Here, we examine the validation metrics in Figures 4 and 5 from a seasonal perspective. We again assess the adjustments in three main areas: temporal alignment (Figure 6), summary precipitation statistics (Figures 7 and S1), and spatiotemporal coherence (Figures 8 and 9). A seasonal analysis is important to determine whether the performance of the adjustments varies by the seasonal precipitation regime, most notably the organized synoptic scale precipitation systems of winter compared to the generally smaller scale and more spatially random convective precipitation patterns of summer (Osborn and Hulme 1997; Hofstra and New 2009; Hutchinson and McKenney 2009; Gervais et al. 2014).

Monthly temporal alignment metrics further support the result that the NLDAS-2 hourly adjustment is the best performing adjustment with respect to temporal alignment (Figure 6). Across all months and adjustment scenarios, the NLDAS-2 adjustment improved temporal

alignment more than the uniform and shift adjustments (Figure 6). The closest a different adjustment came to matching NLDAS-2 performance was for the temporal alignment of summer precipitation occurrence under the AM' midnight adjustment scenario; July HSS values for the NLDAS-2 and shift adjustments were 0.81 and 0.80, respectively (Figure 6g). Surprisingly, under the AM' midnight scenario, monthly temporal alignment performance of the simpler shift method was near or better than the uniform adjustment (Figures 6a, 6d, 6g, and 6j), particularly for precipitation occurrence (Figure 6g).

Seasonal variability in temporal alignment was evident in all adjustment scenarios, but the PM' midnight scenario had the most distinctive winter versus summer seasonal pattern (Figure 6). Here, unadjusted (shift = unadjusted for this scenario) and uniform adjusted PM' midnight data generally exhibited better temporal alignment between the two times of observation in winter than in summer (Figures 6b, 6e, 6h, and 6k). This is most likely indicative of the stronger diurnal variation of summer precipitation. During the summer, precipitation in the validation region is more frequent during the late afternoon (Dai *et al.*, 1999). Because this late afternoon timing closely coincides with the PM 1700 LT time of observation, it likely weakens the temporal alignment of the PM and midnight observations. Overall, even with the weakened temporal alignment of summer, the unadjusted PM and midnight observations displayed temporal alignment that was similar to or even better than their temporal alignment with a uniform adjustment (Figures 6b, 6e, 6h, and 6k). Under the PM' midnight scenario, correlation

was the only temporal alignment metric in which the uniform adjustment noticeably outperformed unadjusted data across all months (Figure 6b).

Seasonal variability in validation metrics for summary statistics was similar across the three adjustment scenarios. As such, we only present monthly results for the AM' midnight adjustment scenario (Figure 7). Results for the other adjustment scenarios can be found in Figure S1. The seasonal summary statistics results further confirm that the uniform adjustment produces a significant positive bias in r_{lag1} due to its reapportioning of all single-day precipitation totals to two days (Figure 7e). Out of all the validation metrics for summary statistics, percent error in r_{lag1} for the uniform adjustment had the largest season swings ranging from +752% in January to +291% in October (Figure 7e). Because the monthly range of r_{lag1} for uniform adjusted data (0.38 to 0.45) is so much higher than unadjusted data (0.06 to 0.15), these large seasonal changes correspond to only small seasonal changes in r_{lag1} error when viewed in absolute terms. For instance, the +752% in January corresponds to a +0.32 overestimation of r_{lag1} whereas the +291% error in October corresponds to +0.30 overestimation of r_{lag1}, a difference of +0.02.

The monthly summary statistics results again indicate that the uniform and NLDAS-2 hourly adjustments produce non-trivial artificial changes in intensity (Figure 7b) and frequency (Figure 7c), and that the uniform adjustment also produces substantial artificial changes in extremes (Figure 7d). Compared to the uniform adjustment, the intensity and frequency errors introduced by the NLDAS-2 adjustment exhibited a more discernable seasonal cycle (Figures 7b and 7c). The NLDAS-2 adjustment intensity error reached its most negative value in mid-to-late

summer (July to September percent error = -19 to -20%) and least negative value in mid-to-late winter (January to March percent error = -15%; Figure 7b). Correspondingly, NLDAS-2 adjustment frequency error reached its most positive value in mid-to-late summer (July to September percent error = +24 to +25%) and least positive value in mid-to-late winter (January to March percent error = +17 to +18%; Figure 7c). As discussed in Section 3.2, this is most likely a scale mismatch between NLDAS-2 gridded data and local point-based station observations that becomes more accentuated by the smaller scale spatial patterns of summer convective precipitation.

Monthly *x0* spatiotemporal coherence metrics for unadjusted AM-AM, PM-PM, and midnight-midnight station pairs clearly illustrate the validation region's yearly seasonal change from a winter to summer precipitation regime (Figure 8). Both $x0r$ and $x0\chi$ reached their highest values in winter months and their lowest values in summer months (Figure 8). With these distinct seasonal changes in the spatiotemporal coherence of the precipitation regime (Figure 8), seasonal variability was also evident in the spatiotemporal coherence validation metrics (Figure 9). Although the NLDAS-2 adjustment was the only adjustment to produce percent errors in both target xr and $x0\chi$ that hovered near 0% across all months and adjustment scenarios (Figures 9a-9c and 9g-9i), NLDAS-2 adjustment percent errors in original xr and $x0\chi$ were generally positive and had a strong seasonal pattern (Figures 9d-9f and 9j-9l). This provides further evidence that the NLDAS-2 hourly adjustment has the potential to artificially increase interstation spatiotemporal coherence for stations with the original time of observation,

particularly during non-winter months when positive percent errors are greater for both $x0r$ and $x0\chi$ and can reach values of +10 to +22% (Figures 9d-9f and 9j-9l). Analogous to the NLDAS-2 adjustment's errors in frequency and intensity, this is likely due to a scale discrepancy between the NLDAS-2 dataset and station observations. The smoother spatial variability of the NLDAS-2 grid cells will tend to increase the spatial coherence relative to that of point-based station observations (e.g. Maraun 2013). For instance, at a daily timestep, $x0r$ for the NLDAS-2 dataset itself is 506-km, which is +71% greater than the $x0r$ for AM-AM station pairs (Figure 3).

Similar to general spatiotemporal coherence analysis (Section 3.3; Figures 4 and 5), seasonal spatiotemporal results for the uniform adjustment had several key similarities and differences relative to the NLDAS-2 adjustment. As with the NLDAS-2 adjustment, the uniform adjustment percent error for target $x0r$ was near 0% across all months and adjustments scenarios (Figures 9a-9c). However, percent error for original $x0r$ was even more positive and had a strong seasonal pattern with higher error values of +18 to +29% during the summer (Figures 9d-9f). Although the uniform adjustment's percent error in original $x0\chi$ had some seasonal variability, it was not as consistently positive as the NLDAS-2 adjustment (Figures 9j-9l). For stations with the original time of observation, this again suggests that the uniform adjustment has less potential to add a consistent positive bias to the spatiotemporal coherence of extreme events. However, unlike the NLDAS-2 adjustment, the uniform adjustment had a consistent negative bias for target $x0\chi$ across all months that was enhanced during the summer for the PM' midnight (Figure 9h) and PM' AM scenarios (Figure 9i).

In contrast to the other adjustments, shift adjustment improvements to target spatiotemporal coherence were limited to the AM' midnight scenario (Figures 9a and 9g). For PM' midnight the shift adjustment was not applicable (i.e. shift = unadjusted; Figures 9b and 9h) and under the PM' AM scenario it was detrimental across all months (Figures 9c and 9i). Although shift adjustment percent errors were consistently larger for target *x0r* and $x0\chi$ than the other adjustments for the AM' midnight scenario (Figures 9a and 9g), it still provided a clear improvement over unadjusted data across all months and had the added benefit of having negligible impacts on original *x0r* and $x0\chi$ (Figures 9d and 9j).

4 Conclusions

Time of observation adjustments for daily precipitation are important for improving interstation data compatibility, but they are not without drawbacks. Based on the results shown here and those from previous analyses (Holder *et al.*, 2006; Kim and Pachepsky, 2010), we make the following conclusions: (1) Time of observation adjustments can clearly improve temporal alignment, especially between AM and midnight observations (Figures 4a, 6a, 6d, 6g, 6j). (2) Although they improve temporal alignment, adjustment methods that reapportion daily precipitation observations have the potential to artificially increase precipitation frequency and temporal autocorrelation, and decrease average intensity and extremes (Figures 5 and 7). (3) Adjustment methods that reapportion daily precipitation observations also have the potential to

artificially increase the interstation spatiotemporal coherence of the stations being adjusted (Figures 5, 9d-f, 9j-l).

The issues introduced by the time of observation adjustments complicate the selection of an appropriate adjustment method. While much will depend on the specific application, we make several suggestions. First, if observations are dominated by a certain time of observation category (e.g. AM), we recommend only adjusting observations that are not in the dominant category unless a specific application requires a different observation time. This limits the number of observations that have to be adjusted and helps to avoid the introduction of widespread biases. Second, we recommend that the uniform method not be used as a time of observation adjustment. Although the uniform method has an ability to improve certain aspects of temporal alignment (Figures 4 and 6) and spatiotemporal coherence (Figures 5 and 9), especially with respect to the AM' midnight adjustment scenario, it can be detrimental to the temporal alignment of precipitation occurrence (Figures 4 and Figures 6h-i), and also substantially bias precipitation statistics (Figures 5 and 7) and the spatiotemporal coherence of the stations being adjusted (Figures 5 and Figures 9d-f). An appropriate alternative is to use the shift adjustment, but only in adjustment scenarios where a one-day shift increases the number of hours of overlap between the two times of observations. In such scenarios, the shift adjustment can improve temporal alignment similar to the uniform adjustment (Figures 4a, 6a, 6d, 6g, 6j), but without biasing precipitation statistics (Figures 5a and 7) and spatiotemporal coherence (Figures 5a, 9d, 9j). A simple method to determine the appropriateness of the shift adjustment for

two times of observation is to calculate their difference in hours. If the difference is > 12 hours, then a simple shift in one of the time series will likely improve temporal alignment because it increases the number of hours of overlap between the two times of observation. For instance, because the hour difference of the AM' midnight scenario (0700 LT vs. 2400 LT) is 17 hours (Figure 1a), a 1-day forward shift of the AM time series or a 1-day backward shift of the midnight time series improves temporal alignment (Figures 4a, 6a, 6d, 6g, 6j). In contrast, because the hour difference of the PM' AM scenario (1700 LT vs. 0700 LT) is 10 hours (Figure 1c), a shift adjustment decreases the number of hours of overlap and is not appropriate (Figures 4c, 6c, 6f, 6i, 6l).

Before applying hourly adjustments based on gridded hourly precipitation datasets like NLDAS-2, we recommend that possible spatial scale discrepancies between the gridded data and point-based station observations be carefully assessed. While the NLDAS-2 hourly adjustment has the ability to improve temporal alignment (Figures 4 and 6), and aspects of spatiotemporal coherence (Figures 5, 9a-9c, 9g-9i) more than any other adjustment, scale discrepancies can artificially modify frequency and intensity (Figure 5), and bias the spatiotemporal coherence of the stations being adjusted (Figures 5, 9d-9f, 9j-9l). The temporal consistency of the spatial scale discrepancies in the gridded products also need to be examined. NLDAS-2 uses two main data sources to produce hourly precipitation estimates: station-based hourly observations from 1979 to July 1996 and hourly radar observations from July 1996 to present (Ferguson and Mocko 2017). Although we focused our analyses on the radar era, we did find evidence that errors in

spatiotemporal coherence introduced by NLDAS-2 were modified by the switch to radar in 1996. For instance, under the AM' midnight adjustment scenario, average percent error in original *x0r* was +16% before 1996, but then abruptly dropped to an average of 6% after radar was introduced (Figure S2). As a result, if the NLDAS-2 hourly adjustment is used across the time period of the switch to radar, it will likely introduce temporal inconsistencies in the spatiotemporal coherence of the precipitation field represented by the adjusted station observations.

It is important to note that while time of observation disparities can cause significant temporal alignment issues with daily precipitation, they become less of an issue at coarser temporal resolutions. For example, average temporal correlation between unadjusted AM and midnight USCRN time series increases from 0.37 for daily precipitation to 0.88 for pentad precipitation (Figure S3). The noted drawbacks of the adjustment methods also diminish with coarser temporal resolutions. When changing from a daily to pentad resolution under the AM' midnight adjustment scenario, coinciding percent errors in frequency and intensity caused by the uniform adjustment decrease in magnitude from +51% and -34% to +6% and -5%, respectively (Figure S4). Therefore, if coarser temporal resolutions are appropriate for the application, degradation of the temporal resolution can be an effective alternative for minimizing the introduction of biases or eliminating the need for time of observation adjustments altogether.

In the end, time of observation adjustments can be essential or questionable depending on the mix of stations and application of interest. Without adjustment, a mix of AM and midnight

stations will significantly decrease the spatiotemporal coherence of the daily precipitation field both with respect to the entire distribution of precipitation observations and extreme events (Figures 5a, 9a, 9g). However, in situations where temporal misalignment is not as significant (e.g. PM' midnight scenario; Figure 4b), the introduction of biases in precipitation frequency and intensity could be more detrimental than minimal temporal misalignment. Any adjustment-introduced increase in the spatiotemporal coherence of extremes (Figures 5 and 9j-9l) will also exaggerate the concurrent spatial coverage and impact of specific events. If considered necessary, all daily precipitation time of observation adjustments need to be applied with a full understanding of not only improvements to temporal alignment, but also impacts to precipitation statistics and the spatiotemporal coherence of the precipitation field.

5 Acknowledgements

We thank Dr. Murali Haran for invaluable feedback on a draft version of the manuscript. This work was supported by the National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and by the National Oceanic and Atmospheric Administration (NOAA) through the Mid-Atlantic Regional Integrated Sciences and Assessments (MARISA) program under NOAA grant NA16OAR4310179.

6 Supporting information

Figure S1: Validation results for summary precipitation statistics by month for different time of observation adjustment methods and scenarios: percent error in average amount for **(a)** AM adjusted to midnight-to-midnight daily precipitation (AM' midnight), **(b)** PM adjusted to midnight-to-midnight daily precipitation (PM' midnight), and **(c)** PM adjusted to AM daily precipitation (PM' AM); percent error in average intensity for **(d)** AM' midnight, **(e)** PM' midnight, and **(f)** PM' AM; percent error in frequency for **(g)** AM' midnight, **(h)** PM' midnight, and **(i)** PM' AM; percent error in extremes (mean of values > the 95th percentile) for **(j)** AM' midnight, **(k)** PM' midnight, and **(l)** PM' AM; and percent error in lag 1 temporal autocorrelation (r_{lag1}) for **(m)** AM' midnight, **(n)** PM' midnight, and **(o)** PM' AM.

Figure S2: Percent error in original Pearson correlation coefficient (r) decorrelation length ($x0_r$) by year from 1979 to 2015 for the NLDAS-2 adjustment method and the AM adjusted to midnight-to-midnight daily precipitation adjustment scenario. Original *x0* error metric is defined in Section 2.2.4.

Figure S3: Temporal correlation between AM and midnight-to-midnight U.S. Climate Reference Network (USCRN) precipitation time series as a function of temporal resolution. Pearson correlation coefficient (r) is shown for both unadjusted and uniform adjusted AM data.

Figure S4: Percent error in frequency and intensity introduced by the uniform adjustment as a function of temporal resolution for the AM to midnight adjustment scenario.

7 References

Briffa KR, Jones PD. 1993. Global surface air temperature variations during the twentieth century: Part 2 , implications for large-scale high-frequency palaeoclimatic studies. *Holocene* **3**(1): 77–88. DOI: 10.1177/095968369300300109.

Chen M, Shi W, Xie P, Silva VBS, Kousky VE, Higgins RW, Janowiak JE. 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research, D: Atmospheres* **113**(4): 1–13. DOI: 10.1029/2007JD009132.

Coles S, Heffernan J, Tawn J. 1999. Dependence measures for extreme value analyses. *Extremes* **2**(4): 339–365. DOI: 10.1023/A:1009963131610.

Dai A, Giorgi F, Trenberth KE. 1999. Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *Journal of geophysical research* **104**(D6): 6377–6402. DOI: 10.1029/98JD02720.

Daly C, Halbleib M, Smith JI, Gibson WP, Doggett MK, Taylor GH, Curtis J, Pasteris PP. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* **28**(15): 2031–2064. DOI: 10.1002/joc.1688.

DeGaetano AT. 1999. A method to infer observation time based on day-to-day temperature variations. *Journal of climate* **12**(12): 3443–3456. DOI: 2.0.CO;2.">10.1175/1520-

0442(1999)012<3443:AMTIOT>2.0.CO;2.

DeGaetano AT. 2000. A serially complete simulated observation time metadata file for U.S. daily historical climatology network stations. *Bulletin of the American Meteorological Society* **81**(1): 49–67. DOI: 2.3.CO;2.">10.1175/1520-0477(2000)081<0049:ASCSOT>2.3.CO;2.

Diamond HJ, Karl TR, Palecki MA, Baker CB, Bell JE, Leeper RD, Easterling DR, Lawrimore JH, Meyers TP, Helfert MR, Goodge G, Thorne PW. 2013. U.S. Climate Reference Network after one decade of operations: status and assessment. *Bulletin of the American Meteorological Society* **94**(4): 485–498. DOI: 10.1175/BAMS-D-12-00170.1.

Ferguson CR, Mocko DM. 2017. Diagnosing an artificial trend in NLDAS-2 afternoon precipitation. *Journal of Hydrometeorology* **18**(4): 1051–1070. DOI: 10.1175/JHM-D-16-0251.1.

Gervais M, Tremblay LB, Gyakum JR, Atallah E. 2014. Representing extremes in a daily gridded precipitation analysis over the United States: impacts of station density, resolution, and gridding methods. *Journal of climate* **27**(14): 5201–5218. DOI: 10.1175/JCLI-D-13-00319.1.

Gilleland E, Katz R. 2016. extRemes 2.0: an extreme value analysis package in R. *Journal of statistical software* **72**(1): 1–39. DOI: 10.18637/jss.v072.i08.

Higgins RW, Janowiak JE, Yao Y. 1996. *A gridded hourly precipitation data base for the United States (1963-1993)*. NCEP/Climate Prediction Center.

Holder C, Boyles R, Syed A, Niyogi D, Raman S. 2006. Comparison of collocated automated

(NCECONet) and manual (COOP) climate observations in North Carolina. *Journal of Atmospheric and Oceanic Technology* **23**(5): 671–682. DOI: 10.1175/JTECH1873.1.

Hopkinson RF, McKenney DW, Milewska EJ, Hutchinson MF, Papadopol P, Vincent LA. 2011. Impact of aligning climatological day on gridding daily maximum–minimum temperature and precipitation over Canada. *Journal of Applied Meteorology and Climatology* **50**(8): 1654–1665. DOI: 10.1175/2011JAMC2684.1.

Joyce RJ, Janowiak JE, Arkin PA, Xie P. 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology* **5**(3): 487–503. DOI: 2.0.CO;2.">10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

Kim J-W, Pachepsky YA. 2010. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of Hydrology* **394**(3–4): 305–314. DOI: 10.1016/j.jhydrol.2010.09.005.

Lin Y, Mitchell KE. 2005. The NCEP stage II/IV hourly precipitation analyses: development and applications. *19th Conf. Hydrology*. paper presented at the 19th Conf. Hydrology. American Meteorological Society.

Maraun D. 2013. Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *Journal of climate* (D): 2137–2143. DOI: 10.1175/JCLI-D-12-00821.1.

Maurer EP, Wood AW, Adam JCL, Nijssen B. 2002. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *Journal of climate* **15**(22): 3237–3251. DOI: 2.0.CO;2.">10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2.

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. 2012. An overview of the Global Historical Climatology Network-Daily database. *Journal of Atmospheric and Oceanic Technology* **29**(7): 897–910. DOI: 10.1175/JTECH-D-11-00103.1.

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jovi D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W. 2006. North American regional reanalysis. *Bulletin of the American Meteorological Society* **87**(3): 343–360. DOI: 10.1175/BAMS-87-3-343.

Osborn TJ, Hulme M. 1997. Development of a relationship between station and grid-box rainday frequencies for climate model evaluation. *Journal of climate* **10**(8): 1885–1908. DOI: 2.0.CO;2.">10.1175/1520-0442(1997)010<1885:DOARBS>2.0.CO;2.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reiss R-D, Thomas M. 2007. *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser Basel. DOI: 10.1007/978-3-7643-7399-3.

Tustison B, Harris D, Foufoula-Georgiou E. 2001. Scale issues in verification of precipitation forecasts. *Journal of geophysical research* **106**(D11): 11775–11784. DOI: 10.1029/2001JD900066.

Weller GB, Cooley DS, Sain SR. 2012. An investigation of the pineapple express phenomenon via bivariate extreme value theory. *Environmetrics* **23**(5): 420–439. DOI: 10.1002/env.2143.

Wood AW, Maurer EP, Kumar A, Lettenmaier DP. 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research, D: Atmospheres* **107**(20): 1–15. DOI: 10.1029/2001JD000659.

Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, Luo L, Alonge C, Wei H, Meng J, Livneh B, Lettenmaier D, Koren V, Duan Q, Mo K, Fan Y, Mocko D. 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of geophysical research* **117**(D3): D03109. DOI: 10.1029/2011JD016048.

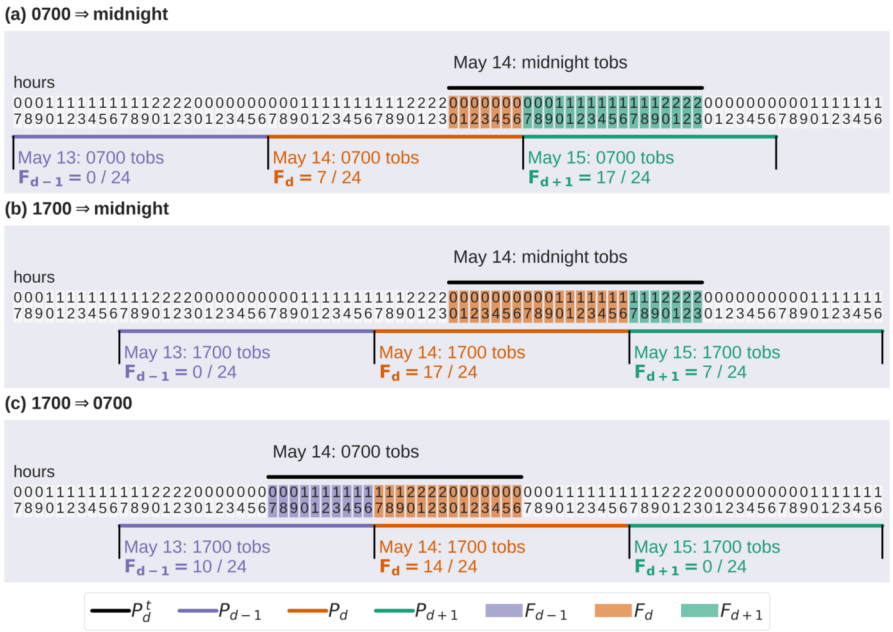

Fig01.tif

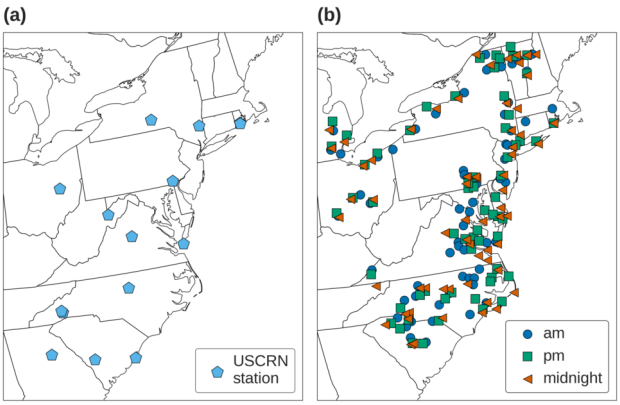

Fig02.tif

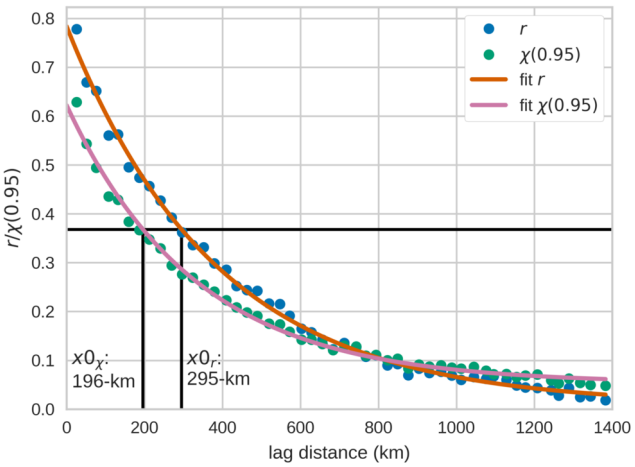

Fig03.tif

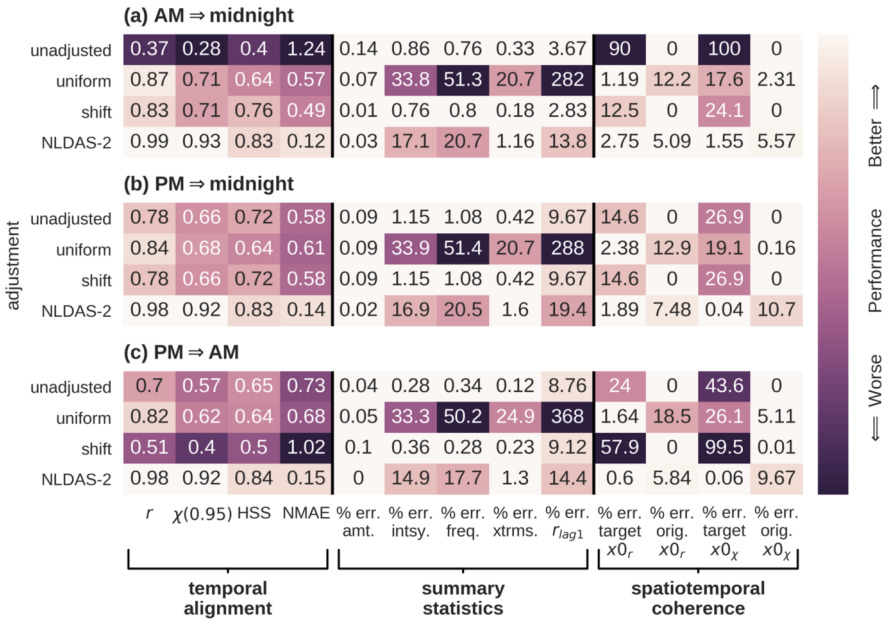

Fig04.tif

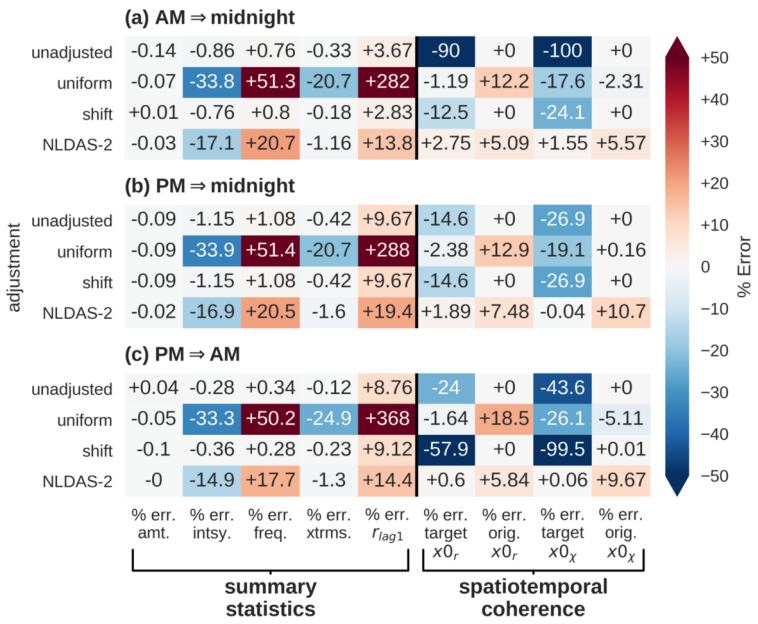

Fig05.tif

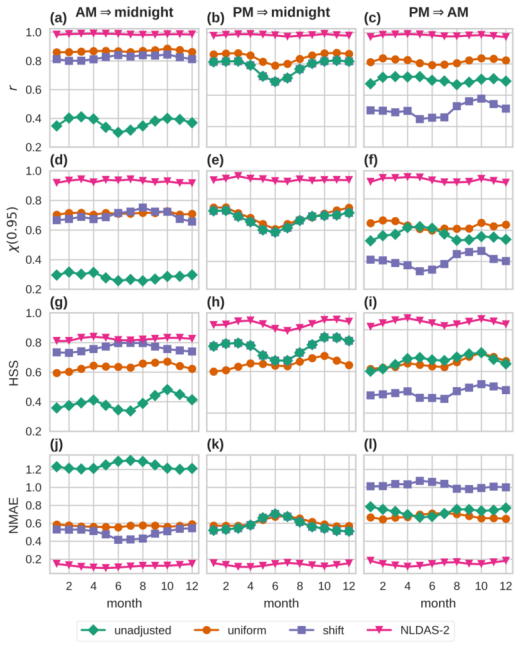

Fig06.tif

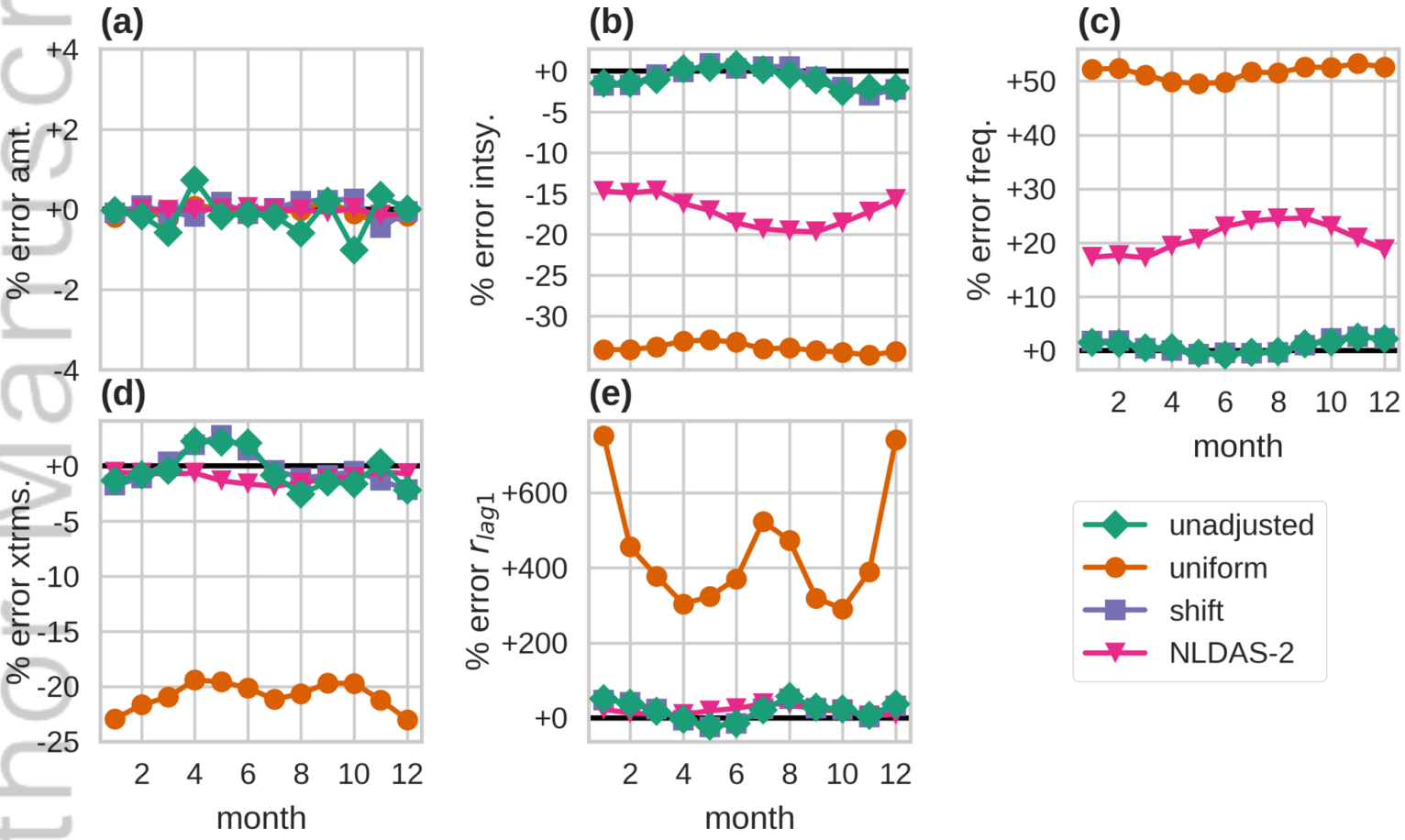

Fig07.tif

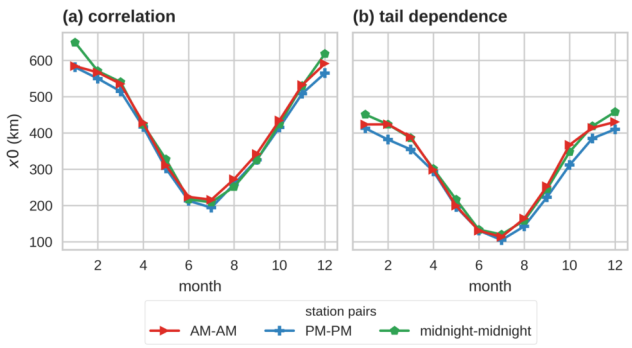

Fig08.tif

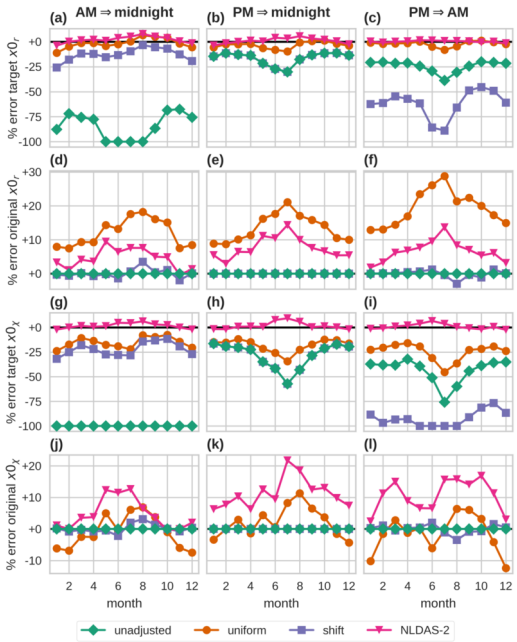

Fig09.tif

Title: Time of Observation Adjustments to Daily Station Precipitation May Introduce Undesired Statistical Issues

Authors: Jared W. Oyler*, Robert E. Nicholas

Interstation observations of daily precipitation are often temporally misaligned due to differences in station time of observation. Time of observation adjustments for daily precipitation are important for improving interstation data compatibility, but they are not without drawbacks. We find that adjustment methods can artificially increase precipitation frequency and decrease intensity, while also artificially increasing interstation spatiotemporal coherence.