Accounting for missing data in monthly temperature series: Testing rule-of-thumb omission of months with missing values

Conor I. Anderson | William A. Gough

Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

Correspondence
Conor I. Anderson, Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C1A4, Canada. Email: conor.anderson@utoronto.ca

The “3/5 rule” is a commonly used rule-of-thumb for dealing with missing data when calculating monthly climate normals. The rule states that any month that is missing more than three consecutive daily values, or more than five daily values in total, should not be included in calculated monthly climate normals. We quantify the impact of missing data in a given year–month for between 1 and 25 missing values. As such, we describe the error the “3/5 rule” (and a related rule that we have dubbed the “4/10 rule”) permits. We tested the statistical robustness of these rules using observed temperature data from a temperate station and a tropical station. We show that, for observed data, the “3/5 rule” permits an average of between 0.06 and 0.07 standard deviations of error in the calculated monthly mean ($\epsilon$) when three consecutive or five random values are missing. For its part, the “4/10 rule” permits a maximum $\epsilon$ of between 0.07 and 0.09 when four consecutive values are missing, or up to 0.10 when 10 random values are missing. The proportional impact of missing values was similar across variables. We performed a correlation analysis and show that each additional missing value from a year–month of data increases $\epsilon$ by between 0.008 and 0.018 for up to 19 missing values. There is a significant relationship between the lag-1 autocorrelation of a year–month, and $\epsilon$. $\epsilon$ can be reduced by simple linear interpolation when values are missing at random and the year–month exhibits lag-1 autocorrelation. Overall, we find that the application of any “rule-of-thumb” should be based on the particular characteristics of the source data and the goals of the research project.

KEYWORDS
data gaps, missing values, monthly series, temperature, 3/5 rule

1 | INTRODUCTION

Climate research is based on long-term trends in meteorological phenomena. Data gaps, or missing data—a common occurrence in climatological time series—present a challenge to researchers. Indeed, it stands to reason that the analysis of any data series that is missing data may not be as statistically robust as a series of fully intact data.

Myriad studies have described and implemented gap-filling and homogenization techniques for long-term climate series, many of which require the use of a neighbouring station or stations (e.g., Price et al., 2000; Jeffrey et al., 2001; Hofstra et al., 2008; Mekis and Vincent, 2011; Vincent et al., 2012; Trujillo et al., 2015). All of these interpolation methods involve the creation of approximated data, potentially introducing other challenges for the statistical robustness of a study.

In other studies, data series that are missing values are omitted from analysis based on some generally specified threshold or “rule-of-thumb,” for instance, in Zhai et al. (1999): no more
than 20 daily precipitation values missing in a year, and no stations with three consecutive missing years or five randomly missing years; in Zubieta et al. (2017): no more than 5% of daily values missing in a month; in Chapman and Walsh (1993), Klein Tank et al. (2002), and Haylock et al. (2008): no more than 20% missing daily, monthly, or annual values, respectively. These “rules-of-thumb” are often study-specific and may lack formalization or specific statistical justification.

Perhaps the most formalized “rule-of-thumb” is the so-called “3/5 rule.” We sought to identify the origin of this rule, which appeared in the technical literature at least as early as 1989:

When arithmetic means are to be calculated for each month of each year from daily data the following rule (hereafter referred to as the “3/5 rule”) applies. If more than three consecutive daily values are missing or more than five daily values in total in a given month are missing, the monthly mean should not be computed and the year–month mean should be considered as missing. The number of days for which monthly means are calculated is \( N \) where \( N \) can vary from 23 to 31 (World Meteorological Organization, 1989, p. 7).

This rule remained in place within the World Meteorological Organization and was still recommended methodology in more recent WMO reports (Plummer et al., 2003). The “3/5 rule” was included as the de facto standard practice for calculating climate normals, employed by agencies around the world, including Environment and Climate Change Canada, formerly Environment Canada (2016). While the rule was set out for the definition of climate normals, it is a common tool in climate scientists’ quality-control procedures for other types of analysis such as time series analysis and trend detection.

There is no formal reference to the “3/5 rule” in either the second (1983) or third (2011) editions of the Guide to Climatological Practices published by the World Meteorological Organization. Instead, the manual recommends omitting any month that has more than 10 missing daily values. Likewise, it recommends that any month that has five or more consecutive missing values should be omitted. This recommendation was not modified in the revision of section 4.8 of the second edition, published in January 2016 (World Meteorological Organization, 2016). While the definition of the “3/5 rule” states that there can be up to a maximum of three (i.e., \( k \leq 3 \)) consecutive missing values, the Guide to Climatological Practices is explicit that there should not be five or more consecutive missing values (i.e., \( k < 5 \), thus, \( k \leq 4 \)). Therefore, we choose 4 as the upper limit on the number of allowable, consecutive missing values. For the sake of consistency, we will stylize this alternative guideline as the “4/10 rule” in this paper.

Few studies seek to describe the impact, if any, that missing values may have on an analysis. Stooksbury et al. (1999) examined the impact of missing values in maximum and minimum temperature series in the continental United States. Those authors defined the impact as the standard deviation of 30 simulations of the departure from the true mean when 1–10 consecutive values are removed artificially from a month of intact daily data. They found spatial and temporal patterns in the impact of missing data—higher impact in continental regions during the winter months, and lower impact in the American southeast during the summer. Trewin (2007), on the other hand, assessed the impact of daily missing values as the change in the width of the confidence interval of a calculated monthly mean when data are missing. That author found that missing values will have little impact on the confidence of calculated monthly averages. In fact, at the limits of the “3/5 rule” (five missing values of which three are consecutive), the confidence interval for the calculation of the monthly mean is roughly the same width as the error introduced when normals from the period 1961–1990 were used to predict values for 1991–2000. The most recent, and perhaps most rigorous treatment of missing values in a temperature time series was by Massetti (2014). Massetti constructed a model based on the artificial removal of one to five consecutive and random missing values from observed temperature data at 720 complete data sets. Massetti showed that consecutive values have a larger impact than random missing values due to the tendency of temperature data series to be autocorrelated, finding a statistically significant relationship between lag-1 autocorrelation and error due to consecutive missing values. The modelled error in a monthly series varied between 0.5 and 3.7°C for 10 missing values.

The “3/5 rule” (“4/10 rule”) ostensibly protects the statistical integrity of calculated monthly climate means from the errors that were described by Stooksbury et al. (1999), Trewin (2007), and Massetti (2014); however, we have not found any formalism in the literature to explain why the numbers 3 (4) and 5 (10) were chosen. We sought, therefore, to quantitatively describe the error that missing data may introduce to an analysis at the upper limits of these rules. To that end, we analysed 20 years of daily temperature data from two climate stations. We will describe the impact of missing data on calculated monthly means by removing daily values from the month of study, recalculating the monthly mean, and measuring the error in the calculated monthly mean as a result of this manipulation. While these rules apply to the calculation of normals, in the present study, we focus on individual year–months, as the overall temperature error that a single erroneous year–month will contribute to a calculated monthly normal is typically substantially smaller than the error in the year–month mean derived from incomplete daily values. We report the error as a proportion of the standard deviation of the daily values of a year–month under study. We compare these results in the context of a range of missing values, with single year–month data sets missing between one and 25 data points. Our findings can be generalized to other threshold-based “rules-of-thumb” around data omission.
2 | METHODOLOGY

In order to estimate the practical implications of the “3/5” and “4/10” rules, we analysed 20 years of observed data with no missing values from both a temperate station, and a tropical station. We performed our analyses in the R language for statistical computing (R Core Team, 2017).

For our temperate station, we acquired climate data from the “Toronto,” later “Toronto City” climate station (No. 5051 and 31688, 43.67°N, 79.4°W) for the period 1841–2015 from Environment and Climate Change Canada’s Historical Climate Data archive. We used the canadaHCD package (Simpson, 2016) in R to perform this downloading. Toronto City is located in downtown Toronto, Canada, on the University of Toronto’s St. George Campus. It is a temperate station that is affected by a strong urban heat island (Mohsin and Gough, 2012), and is located along the typical path of the northern polar jet, which causes a great deal of variation in the city’s temperature. When calculated across all months with no missing values \((n = 2,091)\), the average standard deviation of daily temperature data in a month was 4.03°C for \(T_{\text{mean}}\), with a range of between 1.31 and 8.00°C.

For our tropical station, we acquired data from the “Requena” station (No. 000280, 5.04°S, 73.84°W) for the period from 1960 to 2015. These data were acquired from the Peruvian National Meteorology and Hydrology Service (Senamhi), using the senamhiR package (Anderson, 2017). Requena is a small city in the heart of the lowland Peruvian Amazon. The station is located near the Pacaya-Samira National Reserve. It experiences a warm, humid climate, with little variation in daily temperatures; the average standard deviation of the daily values within a month for all months with no missing values \((n = 545)\) was 1.03°C for \(T_{\text{mean}}\), with a range of 0.15–2.86°C.

Toronto has relatively continuous, intact data, while Requena contains myriad missing values throughout the time series. As such, it was impossible to find any single long-term period that was free from missing values. Instead, we identified shorter, continuous periods where there were no missing values at either station, and selected four 5-year periods: 1974–1978, 1989–1993, 1998–2002, and 2004–2008, for a total of 240 year–months over 20 years.

For each of the 240 year–months in our data series, we calculated the mean \((\mu)\) and standard deviation \((\sigma)\) of the daily values in each year–month for each of \(T_{\text{mean}}, T_{\text{max}},\) and \(T_{\text{min}}\). We proceeded to eliminate between 1 and 25 consecutive \((k)\) or random \((j)\) data points and re-calculate the mean value. The difference between the original mean and the “modified” mean \((\mu_{\text{mod}})\) is reported as proportion of the standard deviation of the daily values of the intact year–month. We will use the Greek letter epsilon \((\epsilon)\) to symbolize this metric (Equation (1)). Expressing \(\epsilon\) as a proportion of \(\sigma\) allowed us to account for the variability in a given data set, and facilitates comparison between different months or stations, especially when comparing year–months with higher internal variability, such as transition months in temperate climates. As such, in this paper, when we refer to \(\epsilon\) as a proportion of the standard deviation, we refer to the standard deviation of the daily values of the intact year–month.

\[
\epsilon = \frac{|\mu_{\text{mod}} - \mu|}{\sigma}. \tag{1}
\]

We employed slightly different sampling methodologies for the consecutive and random tests. When performing consecutive sampling, we tested all possible permutations of a month with \(k\) missing values, and eliminated data points in consecutive order from the first day of the month to day number \((n - k + 1)\), where \(n\) is the number of days in the month. Hence, we performed a total of 110,625 samples at each station over the 240 year–months and 25 levels of \(k\).

When performing random sampling, we choose a random sample of \(j\) values from 1 to \(n\) without replacement. It is considerably more difficult to test all possible combinations of \(j\) random variables. There are a total of 98,280 possible combinations of five missing values for a 28-day February; this value rises to 142,506 combinations for a 30-day month, and 169,911 combinations for a month of 31 days. In order to make efficient use of computational resources, we repeated our experiment at each value of \(j\) 1,000 times for each year–month and value of \(j\), for a total of 6 million tests at each station. While this will not account for all possible combinations of missing values, we are confident that it provides a representative sample of the sort of data integrity issues that are common in the field of climatology. The code used to perform this analysis is freely available on the GitLab platform as Anderson (2018). Given that our sampling was fully random, a given sample may contain periods of consecutively missing values. We also have a small number of instances of repeated samples at low or high values of \(j\). For February, for instance, there are only 28 truly distinct samples of a single day \((j = 1)\). Likewise, with a random seed set to 404, there were 998 distinct samples (two samples were repeated) at \(j = 21\). At \(j = 25\), there were 862 distinct samples. We did not duplicate these samples in our analysis.

2.1 | On the influence of lag-1 autocorrelation

The temperature of any given day is often influenced by the temperature of the previous day or days (Gough, 2008). To assess differences between the impact of a one-unit increase of \(k\) or \(j\), we sought to determine if there is a relationship between serial autocorrelation and \(\epsilon\). We used the Durbin–Watson test (Fox and Weisberg, 2011) to detect serial autocorrelation in our 240 year–months, and performed a linear regression between the average \(\epsilon\) of each year–month and variable, and the lag-1 autocorrelation for the corresponding time series.
2.2 | Regression analysis

We performed a knotted linear regression to test the relationship between the average $\epsilon$ of each station–variable pair and the values of $k$ or $j$ from 1 to 25. Using the aggregate $\epsilon$ of each station–variable pair, averaged across the 240 year–months, we fit degree-1 (linear) polynomials with a knot (or elbow) at values of $k$ or $j$ between 2 and 24 and selected the model that produced the highest $R^2$ value. A knotted linear regression was selected because our data showed an apparent change-point around $k$ or $j$ of 19.

2.3 | Interpolating missing values

We sought to test whether our calculated monthly means would be more robust if we sought to approximate missing values. We used two simple interpolation methods as “toy” methods to approximate the values that we had artificially removed. The first was linear interpolation, which joins adjacent points linearly and interpolates missing values from the function of that line. We also tested fitting by cubic splines. Both of these methods are implemented by Zeileis and Grothendieck (2005). We used these computationally cheap, within-station methods due to the scale of our analysis. For an applied climatological study, however, we would recommend more robust interpolation and homogenization methods, such as those referenced in the introduction of this paper. For all of our tests, we treated each year–month as an isolated unit; as such we did not perform interpolation by either method when a missing value occurred at the start or end of the month, as these cases performed very poorly. In practice, these cases are unrealistic as year–months will typically be considered in the context of a longer time series.

We calculated the impact of those missing values using Equation (1) both before, and after interpolation. We evaluated the interpolation skill by comparing the $\epsilon$ before interpolation with $\epsilon$ after interpolation. A visual example of these interpolation methods are provided in Figure 1.

2.4 | Comparing months

Finally, we examined whether $\epsilon$ varies across months. We created aggregate series for $k \in \{3, 4\}$ and $j \in \{5, 10\}$ for each station–variable pair, averaging each month across the 20 years in our data set. Given that our samples did not meet the assumptions of ANOVA, specifically, the sample size of our consecutive samples varies based on month length, we used the nonparametric Kruskal–Wallis test to identify whether there were any significant differences in the impact of missing values between months at either station. The Kruskal–Wallis test is a one-way analysis of variance test to detect if any 1 month shows stochastic dominance over the others. That is, the test will detect cases where there was a statistically significant difference in the median $\epsilon$ between months. We also performed the all-pairs Nemenyi test with $\chi^2$ approximation to detect pairwise differences between months. Both the Kurskal–Wallis test and Nemenyi test were implemented by Pohlert (2018).

3 | RESULTS AND DISCUSSION

3.1 | The impact ($\epsilon$) of removing daily values

3.1.1 | Eliminating $k$ consecutive values

The average error after eliminating $k$ consecutive values is illustrated in Figure 2a. The curves for the overall average $\epsilon$,
and for the first and third quartiles are shown. We detected significant relationships (at the 0.001 level) between the number of consecutive missing values ($k$) and the error in the calculated mean for all variables. The knotted linear regression (summarized in Table 1) performed best with a knot at $k = 19 \pm 1$, with $R^2$ values exceeding 0.99 for all variables at both stations. The linear regression showed an increase in $\epsilon$ of approximately 0.011 to 0.018 for each additional missing value of $k$ before the knot. The slope of the relationship increased to between 0.022 and 0.026 $\epsilon$ for each $k$ after the knot location. At both stations, the pre-knot slope of the line was lowest for $\epsilon_{T_{max}}$. The pre-knot slope of the relationship was roughly 50% steeper at Toronto than at Requena for all variables. We suspect that this stronger relationship is due to the higher degree of autocorrelation at Toronto, which we explore further in section 3.1.2. There was a smaller difference in the post-knot slope. We note that the knot location in all of our consecutive tests exceeds what we would consider to be a practical level of missing values, and thus we suggest that the pre-knot slope is more useful for the purposes of this study.

As for our “rules-of-thumb,” when $k$ was set to 3, the results for all three variables showed $\epsilon$ values of between 0 and 0.30 at Requena and 0 and 0.25 at Toronto when we consider all samples individually. On aggregate, Requena showed an average $\epsilon$ of 0.06 across variables, with median values of 0.05. Toronto exhibited average $\epsilon$ values of 0.07 across variables, with median values of 0.06. Toronto suffered a greater impact to the calculated average due to three consecutive missing values than Requena, which is consistent with our regression analysis.

![FIGURE 2](https://wileyonlinelibrary.com)  
**Average impact of missing values on average $\epsilon_{T_{max}}$, $\epsilon_{T_{mean}}$, and $\epsilon_{T_{min}}$ for values of (a) $k$ and (b) $j$ from 1 to 25. Solid line presents the average of all samples (random: $n = 240,000$; consecutive: $n = 20 \times (N-k+1)$, where $N$ is the number of days in the month). The dashed lines show the first, and third quartiles of each series [Colour figure can be viewed at wileyonlinelibrary.com]

### TABLE 1 Linear regression analysis

| Station | Variable | Knot | $R^2$ | Pre-knot slope | Post-knot slope |
|---------|----------|------|-------|----------------|-----------------|
| Consec. Requena | $\epsilon_{T_{max}}$ | 18 | 0.999 | 0.011 | 0.023 |
| | $\epsilon_{T_{mean}}$ | 18 | 0.999 | 0.012 | 0.022 |
| | $\epsilon_{T_{min}}$ | 19 | 0.999 | 0.013 | 0.023 |
| | $\epsilon_{T_{max}}$ | 19 | 1.000 | 0.017 | 0.024 |
| | $\epsilon_{T_{mean}}$ | 20 | 1.000 | 0.018 | 0.026 |
| | $\epsilon_{T_{min}}$ | 20 | 1.000 | 0.018 | 0.026 |
| Random Requena | $\epsilon_{T_{max}}$ | 19 | 0.997 | 0.008 | 0.021 |
| | $\epsilon_{T_{mean}}$ | 19 | 0.997 | 0.008 | 0.021 |
| | $\epsilon_{T_{min}}$ | 19 | 0.997 | 0.008 | 0.021 |
| | $\epsilon_{T_{max}}$ | 19 | 0.997 | 0.008 | 0.021 |
| | $\epsilon_{T_{mean}}$ | 19 | 0.997 | 0.008 | 0.021 |
| | $\epsilon_{T_{min}}$ | 19 | 0.997 | 0.008 | 0.021 |

**Note.** The above are the results of a knotted linear model between $k$ or $j$ missing values and $\epsilon$. The model fits a degree-1 (linear) polynomial to the aggregate average $\epsilon$ before and after the knot. The knot location that maximized $R^2$ was chosen. $k$ and $j$ range from 1 to 25. All of the detected trends were significant at the 0.001 level.
results. Perhaps the most notable characteristic of our analysis, however, was that at each station, $\epsilon$ was similar across all three temperature variables.

Eliminating $k = 4$ consecutive daily data points resulted in $\epsilon$ values of between 0 and 0.31 at Requena and 0 and 0.35 at Toronto. Four missing values led to a range in $\epsilon$ of between 0 and 0.31 at Requena and 0 and 0.35 at Toronto. Eliminating consecutive values, which is consistent with the results we report above.

To ease comparison, the average value for each variable and station are presented in Table 2, along with monthly values, which we address in section 3.3.

### 3.1.2 Eliminating $j$ random values

The results of our 1,000 repetitions testing random missing values for each year–month are shown in Figure 2b. The results are averaged across the 240,000 samples for each level of $j$.

Our knotted linear regression showed the highest $R^2$ (0.997) for all variables at both stations with a knot at $j = 19$. Each increment in the value of $j$ contributed less to $\epsilon$ than a similar increment in $k$. The results were identical across our stations and variables. The linear regression showed a pre-knot increase in $\epsilon$ of 0.008 for each increment in $j$ for all three variables at both Requena and Toronto, with a post-knot slope of 0.021. The similarity between our two stations is interesting, as random sampling is less affected by autocorrelation as there are less consecutive missing values at each level of $j$. Nonetheless, the difference between the pre- and post-knot slopes can be explained, at least partially, by our analysis of consecutive values. Figure 3 shows the number of the 240,000 samples at each station that were free of consecutive missing values. As $j$ increases, the number of randomly missing values that fall on subsequent days increases. At $j = 12$, only 33 of the samples at Requena and 420 of the samples at Toronto were free of data gaps of two

| $k = 3$ | $k = 4$ |
|---|---|
| $e_{T_{\text{max}}}$ | $e_{T_{\text{mean}}}$ | $e_{T_{\text{min}}}$ |
| $e_{T_{\text{max}}}$ | $e_{T_{\text{mean}}}$ | $e_{T_{\text{min}}}$ |
| Jan | 0.07 (0.06) | 0.07 (0.06) | 0.07 (0.06) | 0.09 (0.07) | 0.09 (0.08) | 0.09 (0.07) |
| Feb | 0.07 (0.06) | 0.08 (0.06) | 0.08 (0.06) | 0.09 (0.07) | 0.10 (0.08) | 0.10 (0.08) |
| Mar | 0.07 (0.05) | 0.07 (0.05) | 0.07 (0.05) | 0.09 (0.06) | 0.09 (0.06) | 0.09 (0.06) |
| Apr | 0.07 (0.05) | 0.08 (0.05) | 0.08 (0.06) | 0.09 (0.06) | 0.10 (0.06) | 0.10 (0.07) |
| May | 0.07 (0.05) | 0.07 (0.06) | 0.07 (0.05) | 0.09 (0.07) | 0.09 (0.07) | 0.09 (0.07) |
| Jun | 0.07 (0.06) | 0.08 (0.06) | 0.08 (0.06) | 0.09 (0.07) | 0.10 (0.07) | 0.10 (0.07) |
| Jul | 0.06 (0.05) | 0.07 (0.06) | 0.07 (0.06) | 0.08 (0.07) | 0.08 (0.07) | 0.08 (0.07) |
| Aug | 0.06 | 0.07 (0.06) | 0.07 (0.06) | 0.08 (0.07) | 0.08 (0.07) | 0.08 (0.07) |
| Sep | 0.07 (0.05) | 0.07 (0.06) | 0.07 (0.05) | 0.09 (0.06) | 0.09 (0.07) | 0.09 (0.07) |
| Oct | 0.07 (0.05) | 0.07 (0.06) | 0.07 (0.05) | 0.09 (0.06) | 0.09 (0.07) | 0.09 (0.07) |
| Nov | 0.07 (0.06) | 0.07 (0.06) | 0.07 (0.06) | 0.09 (0.07) | 0.09 (0.07) | 0.09 (0.07) |
| Dec | 0.07 (0.05) | 0.07 (0.05) | 0.07 (0.05) | 0.09 (0.06) | 0.09 (0.06) | 0.09 (0.06) |
| Avg. | 0.07 (0.06) | 0.07 (0.06) | 0.07 (0.06) | 0.09 (0.07) | 0.09 (0.07) | 0.09 (0.07) |

**Note.** The mean overall impact of missing data on the calculated monthly average $T_{\text{max}}, T_{\text{min}}$ and $T_{\text{mean}}$ is presented for each sample method. The values are expressed as the average absolute difference between the calculated mean of the intact month and the calculated mean of same month with missing values, expressed as a proportion of the standard deviation for the intact year–month. Values are for Toronto (Requena). A single value in a cell indicated that the average value was the same at both stations.
The similarity of the results for random elimination across stations and variables is encouraging, as it supports our use of the standard deviation of the daily values in a year–month in our calculation of $\epsilon$. We refer the reader again to Table 2 for summary averages.

### 3.2 The influence of autocorrelation on our results

Massetti (2014) found a positive relationship between the error due to missing values and the lag-1 autocorrelation of a time series. Where there is autocorrelation between adjacent days, contiguous periods of missing values will have a greater impact than randomly missing values. In Table 3, we present the total number of the 240 year–months for which the Durbin–Watson test detected significant ($\alpha = .05$) autocorrelation at lags 1, 2, and 3. Of the 240 year–months studied, 227 exhibited lag-1 autocorrelation at Toronto, compared to only 86 at Requena. As per the slopes of the regression analysis, the number of autocorrelated months was lowest for $T_{\text{max}}$, and highest for $T_{\text{mean}}$ at both stations, with a higher incidence of lag-1 and lag-2 autocorrelation at Toronto than at Requena.

We performed a linear regression analysis for each of our 240 year–months, to determine whether the lag-1 autocorrelation for a particular year–month and variable pair has an impact on the average $\epsilon$ of the respective year–month and variable for $k \in \{3, 4\}$ and $j \in \{5, 10\}$. The relationship is presented in Figure 4, with the results of the analysis in Table 4. For both Toronto and Requena, there was a significant (at the 0.001 level) relationship between $\epsilon$ and lag-1 autocorrelation across our three variables when sampling was performed consecutively. The strongest relationships were found at Requena, rather than Toronto. While the relationship between lag-1 autocorrelation and $\epsilon$ is stronger at Requena than at Toronto, the much higher proportion of year–months that exhibit lag-1 autocorrelation at Toronto can explain the larger impact of each additional consecutive missing datum at Toronto.

We did not find any strong relationships between lag-1 autocorrelation and $\epsilon$ when sampling was performed randomly. When $j = 10$, the linear regression detected a weak relationship (at the 0.05 level) for $\epsilon$ $T_{\text{mean}}$ at Requena and for $\epsilon$ $T_{\text{max}}$ and $\epsilon$ $T_{\text{min}}$ at Toronto (at the 0.01 and 0.05 levels, respectively). These relationships, however, show shallow, negative slopes, with $R^2$ values of between 0.02 and 0.04. We do not expect that these relationships would hold if replicated with other stations or year–months.

### 3.3 Interpolating missing values

While we sought to quantify the impact of missing days in a month of data, we were also interested in determining whether it is better to leave missing data as they are, or to try to interpolate them. We implemented two computationally cheap, within-series methods: linear interpolation, and cubic

### TABLE 3 Number of year–months exhibiting Durbin–Watson autocorrelation

| Station | Variable | Lags |
|---------|----------|------|
|         |          | 1    | 2    | 3 |
| Requena | $T_{\text{max}}$ (°C) | 64   | 19   | 25 |
|         | $T_{\text{mean}}$ (°C) | 86   | 28   | 19 |
|         | $T_{\text{min}}$ (°C)  | 67   | 24   | 19 |
| Toronto | $T_{\text{max}}$ (°C) | 182  | 34   | 22 |
|         | $T_{\text{mean}}$ (°C) | 227  | 49   | 26 |
|         | $T_{\text{min}}$ (°C)  | 214  | 46   | 32 |

*Note. The number of year–months (out of 240) which exhibited statistically significant ($\alpha = .05$) autocorrelation for each station and variable. Toronto exhibited substantively more months with lag-1 autocorrelation than Requena.*

| Number of samples (n = 240,000) |
|---------------------------------|
| 0                              |
| 5                              |
| 10                             |
| 15                             |
| 20                             |
| 25                             |

| FIGURE 3 | The number of the 240,000 random samples made for each level of $j$ that contained no consecutive missing days for Toronto and Requena. As $j$ increases, the number of samples that contained no missing values decays rapidly; none of the 240,000 samples at either station were free of consecutive missing values at levels of $j$ above 12 [Colour figure can be viewed at wileyonlinelibrary.com] | or more days. There were no samples without at least two consecutive missing values for values of $j$ above 12. As such, as $j$ crosses this threshold, the influence of consecutive missing values increases, causing an upward curve in the values of our aggregate values. The knot location for random sampling again exceeded what we consider practical levels of missing values, so we highlight the pre-knot slope as the more useful metric. At $j = 5$, we observed once again that the average $\epsilon$ was similar across all six station–variable pairs, at 0.06. These values ranged from a minimum of 0 to a maximum of 0.33 at Toronto and 0.37 at Requena. The median $\epsilon$ value was also 0.06 across stations and variables. Overall, the average $\epsilon$ of 1,000 repetitions of five random data was rather similar to the results from the elimination of three consecutive values.

Eliminating 10 random data points, resulted in the highest overall $\epsilon$. Minimum $\epsilon$ was 0, with a maximum of 0.55 at both Toronto and Requena. Overall, the average $\epsilon$ in all six station–variable sets was 0.10, or 10% of one standard deviation, which is only 1–2% higher than the impact of half as many consecutive missing values.
spline approximation. We found that, on average, simple linear interpolation decreased $\epsilon$ for $k \in \{3, 4\}$ and $j \in \{5, 10\}$ at Toronto, but was less effective at Requena (see Table 5). The only positive result at Requena was a slight decrease in $\epsilon_{T\text{mean}}$ when $j$ was 5.

Interpolation by cubic splines improved $\epsilon$ in all three variables at Toronto for $j = 5$, and in two variables ($\epsilon_{T\text{mean}}$ and $\epsilon_{T\text{min}}$) at Toronto at $j = 10$. In all of these cases, however, linear interpolation outperformed interpolation by cubic splines.

Figure 5 shows the difference in $\epsilon$ before and after linear interpolation. We have omitted interpolation by spline curves from the figure due to its poor performance. When the data set was missing values consecutively, linear interpolation improved the error at Toronto up to values of $k$ of 6, 10, and 9 for $\epsilon_{T\text{max}}$, $\epsilon_{T\text{mean}}$, and $\epsilon_{T\text{min}}$, respectively. Linear interpolation outperformed interpolation by cubic splines up to values of $j$ of 18, 20, and 20.

The Requena data were, again, less improved by the interpolation of missing data. $\epsilon_{T\text{mean}}$ was the most improved, with marginal decreases up to levels of $j = 6$. Both $\epsilon_{T\text{max}}$ and $\epsilon_{T\text{min}}$ were again only improved at $k = 1$, but these values were only 0.02 and 0.01% of a standard deviation, respectively.

Interpolation by cubic splines increased average $\epsilon$ in all cases at Requena. The fact that linear interpolation—which assumes a perfect linear relationship between the days on either side of a data gap—would perform better at Toronto than at Requena can again be attributed to the proportion of months with significant lag-1 autocorrelation at Toronto. We find that, for a year–month that exhibits lag-1 autocorrelation, when data are missing at random, even simple linear interpolation can slightly improve the accuracy of a calculated mean for relatively high values of $j$. We concede, however, that our
results showed only modest improvement (maximum values of 0.012 to 0.017 for consecutive values and 0.018 to 0.027 for random values); as such, \( e \) while improved, may still be higher than is statistically acceptable at moderate to high values of \( j \).

For the purposes of this study, we used only computationally cheap, within-series methods for interpolation. In practice, it is advisable to use more powerful between-station methods, such as those reviewed by Hofstra et al. (2008) and Trujillo et al. (2015), or more advanced methods such as multiple imputation (e.g., Schneider, 2001), or artificial neural networks (e.g., Infante et al., 2008).

### 3.4 Monthly analysis

We calculated aggregate \( e \) at \( k \in \{3, 4\} \) and \( j \in \{5, 10\} \) for each month (January–December), averaged across our 20 years of fully-intact data from both stations. In all of our tests, the month-to-month range in the average monthly \( e \) for a given station and variable was between 0.01 and 0.02. The individual monthly \( e \) values were rather similar to the overall averages presented in section 3.1, falling within 0.01 of the overall average \( e \) for each level of \( k \) and \( j \). In the case of consecutive sampling, \( e \) was again found to be slightly greater at Toronto than at Requena. The results of the Kruskal–Wallis test are presented in Table 6. The null hypothesis that the median error is similar between months was rejected for all variables at both stations for all tests except for \( k = 3 \) in \( T_{\text{max}} \) and \( T_{\text{mean}} \) at Toronto, indicating that there was a significant difference in the median \( e \) between months at these sampling levels. These results notwithstanding, the post hoc all-pairs Nemenyi test, which performs one-to-one comparison between each month, did not identify any significant differences between any 2 months for \( k \in \{3, 4\} \). When our missing values were removed randomly, however, the Nemenyi test, did identify a number of significant differences within pairs of months. We note, however, that all of these differences occurred in pairs of one 31-day month with a shorter month, or pairs in which one of the months was February. We speculate, therefore, that the differences detected are solely due to the shorter length of February and the 30-day months, as opposed to some climatological phenomenon. The right side of Table 6 shows the group assigned to each month based on the results of the Nemenyi test for each station, variable, and level of \( k \) or \( j \). There was no significant difference (at \( \alpha = .05 \)) for months that share the same group number. These groupings support our assertion that month length is the main influence on \( e \) for randomly missing values. For \( j \in \{5, 10\} \) the 31-day months were always grouped together and February was always grouped alone. Any other groups were due to significant

---

**TABLE 4** Influence of lag-1 autocorrelation on \( e \)

| Station  | Variable | \( k = 3 \) | \( k = 4 \) | \( k = 5 \) | \( k = 10 \) |
|----------|----------|-------------|-------------|-------------|-------------|
|          |          | \( R^2 \)    | Slope       | \( R^2 \)    | Slope       |
| Requena  | \( e T_{\text{max}} \) | 0.599       | 0.034*      | 0.504       | 0.047*      |
|          | \( e T_{\text{mean}} \) | 0.286       | 0.030*      | 0.245       | 0.047*      |
|          | \( e T_{\text{min}} \) | 0.621       | 0.035*      | 0.531       | 0.050*      |
| Toronto  | \( e T_{\text{max}} \) | 0.254       | 0.031*      | 0.231       | 0.051*      |
|          | \( e T_{\text{mean}} \) | 0.564       | 0.035*      | 0.508       | 0.051*      |
|          | \( e T_{\text{min}} \) | 0.245       | 0.029*      | 0.221       | 0.047*      |

**TABLE 5** Error in calculated means before and after approximation of missing data

| Station | Variable | \( k = 3 \) | \( k = 4 \) | \( j = 5 \) | \( j = 10 \) |
|---------|----------|-------------|-------------|-------------|-------------|
|         |          | Miss  | Lin  | Spl | Miss  | Lin  | Spl | Miss  | Lin  | Spl |
| Requena | \( e T_{\text{max}} \) | 0.06  | 0.07 | 0.13 | 0.07  | 0.09 | 0.19 | 0.06  | 0.07 | 0.09 |
|         | \( e T_{\text{mean}} \) | 0.06  | 0.07 | 0.12 | 0.07  | 0.09 | 0.19 | 0.06  | 0.07 | 0.08 |
|         | \( e T_{\text{min}} \) | 0.06  | 0.07 | 0.12 | 0.07  | 0.08 | 0.19 | 0.06  | 0.07 | 0.09 |
| Toronto | \( e T_{\text{max}} \) | 0.07  | 0.06 | 0.09 | 0.09  | 0.08 | 0.14 | 0.06  | 0.05 | 0.06 |
|         | \( e T_{\text{mean}} \) | 0.07  | 0.06 | 0.08 | 0.09  | 0.08 | 0.12 | 0.06  | 0.04 | 0.05 |
|         | \( e T_{\text{min}} \) | 0.07  | 0.06 | 0.08 | 0.09  | 0.08 | 0.12 | 0.06  | 0.04 | 0.05 |

**Note.** The mean overall error in the calculated monthly mean for each test is presented above. Values are presented before (Miss) and after linear interpolation (Lin) or interpolation by cubic splines (Spl) of missing values. Cases where values were missing at the start or end of the year–month are omitted. Values marked in bold show an overall improvement in the calculated mean after the approximation of missing values.

---
FIGURE 5  Change in the error in calculated mean after approximating between 1 and 25 missing values using linear interpolation. Values are shown for the calculated average maximum temperature, mean temperature, and minimum temperature for values of (a) \( k \) and (b) \( j \) from 1 to 25. Negative values indicate an overall improvement in the error (less error). Solid line presents the average of all samples (random: \( n = 240,000 \); consecutive: \( n = 20 \times (N–k + 1) \), where \( N \) is the number of days in the month). The dashed lines show the first, and third quartiles of each series [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 6  Kruskal–Wallis results and Nemenyi post hoc all-pairs groupings

| Station | Variables | Test | values | Nemenyi all-pairs test groupings |
|---------|-----------|------|--------|----------------------------------|
|         |           |      |        | J      | F      | M      | A      | M      | J      | J      | A      | S      | O      | N      | D      |
| Requena | \( \epsilon T_{\text{max}} \) | \( k = 3 \) | 24.0*  | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 30.1** | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 345.2*** | 1 2 3 4,5 6 4,5 1,3,4,5,6 | 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 10 \) | 321.1*** | 1 | 2 3 4,1,3 | 4 | 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         | \( \epsilon T_{\text{mean}} \) | \( k = 3 \) | 33.3*** | 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 47.0*** | 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 266.0*** | 1 | 2 3 1,3 1,3 | 1,3 | 1,3 5 | 1,3 | 1,3 1,3 |
|         |           | \( j = 10 \) | 314.6*** | 1 | 2 3 1,3 | 3 | 1 | 1 1 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         | \( \epsilon T_{\text{min}} \) | \( k = 3 \) | 30.7** | 1 | 1 1 1 1 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 45.6*** | 1 | 1 1 1 1 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 248.8*** | 1 | 2 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 10 \) | 371.24*** | 1 | 2 1,3 | 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
| Toronto | \( \epsilon T_{\text{max}} \) | \( k = 3 \) | 15 | 1 1 1 1 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 30.1** | 1 | 1 1 1 1 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 256.74*** | 1 | 2 1,3 | 1,3 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 10 \) | 353.2*** | 1 | 2 1,3 | 1,3 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         | \( \epsilon T_{\text{mean}} \) | \( k = 3 \) | 16.6 | 1 | 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 27.7** | 1 | 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 252.9*** | 1 | 2 1,3 | 1,3 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 10 \) | 335.0*** | 1 | 2 1,3 | 1,3 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         | \( \epsilon T_{\text{min}} \) | \( k = 3 \) | 23.4* | 1 | 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( k = 4 \) | 36.2*** | 1 | 1 1 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 5 \) | 282.6*** | 1 | 2 1,3 | 1,3 1,3 1,3 1,3 | 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |
|         |           | \( j = 10 \) | 313.4*** | 1,2,3 4 1,2 1,2 1,2 1,2 1,2 | 1,2 1 | 1 | 1 1 1 | 1 1 1 | 1 1 1 |

Note. Kruskal–Wallis test results for the elimination of \( k \) or \( j \) values for each month in 20 years of data are presented. The detection of a significant difference between months is marked according to statistical significance: *at the 0.01 level; **at the 0.05 level; and ***at the 0.001 level. The results of the post hoc Nemenyi all-pairs comparison test are included (\( \alpha = .05 \)). Months are grouped by number; months that share a number showed no significant difference in the impact of missing values.
differences between some of the 30- and 31-day months, but these differences were less consistent.

4 | RELEVANCE OF THESE RESULTS

The University of Toronto Climate Lab undertakes research on climate-related issues around the world. However, our research has primarily focused on Canada, and especially our home city of Toronto. Using the “3/5 rule” and other such “rules-of-thumb” is largely unnecessary for Toronto. Of the 176 years of data recorded at Toronto since 1840 and, later, Toronto City (2,110 year–months of data), only 19 year–months contain any missing data at all for $T_{\text{mean}}$. All of those 19 year–months are admissible under the conditions of the “3/5 rule.”

As we begin research on a major project to describe climate change impacts in the Peruvian Amazon, controlling for missing data becomes increasingly important. At Requena, for instance, data have been collected by Peru’s National Meteorology and Hydrology Service since March 1960. Of the 670 year–months in this 56-year data set, only 545 contain no missing values for $T_{\text{mean}}$, and 94 year–months are completely missing. At the levels of missing values permissible under the “3/5 rule” or the “4/10” rule, 560 year–months are available for analysis. With one additional consecutive missing value (5/10), 563 year–months would be included in our analysis. As we allow more missing values, we make more year–months of data available: at 7/10: 566; at 8/10: 567; at 12/12: 571; at 15/15: 574. We are only limited, therefore, by the maximum error we are willing to accept in our analysis.

In our prior analysis, we found that allowing consecutive missing values of $k$ between 7 and 12 would result in $\epsilon T_{\text{mean}}$ of between 0.10 and 0.16. At Requena, when calculated for all 545 year–months with no missing data (and after correcting three decimal-place recording errors in the original Senamhi data), the mean standard deviation of the daily values in each year–month was 1.03°C (min: 0.15°C, max: 2.86°C). At our tropical station, therefore, we could include an additional 3 year–months of data, if we are willing to accept an average error of approximately 0.10°C of error in the calculated means for the months that we have included. Likewise, if we consider 0.16°C of error acceptable, we can include 11 additional year–months in our analysis. The Peruvian climate record is relatively short and sparse, especially for some rainforest regions that were affected by terrorism in the 1980s. For data-scare regions such as these, “data rescue” techniques, including the admittance of more months with incomplete records, may be necessary to ensure our study contains sufficient baseline data.

As we have seen in our analysis, the application of such data rescue techniques is highly dependent on the degree of serial autocorrelation in each year–month. Additionally, while we have seen that $\epsilon$ is similar between stations and variables, it is a proportion of the standard deviation of the intact daily values of the year–months in our studies. As such, the internal variability of the station under study will influence the real temperature change. At Toronto, for instance, when calculated for all 2091 year–months without any missing data, the mean standard deviation of the daily values for each year–month at Toronto is approximately 4.03°C (min: 1.31°C, max: 8.00°C). If we were to, hypothetically, include year–months with between 7 and 12 consecutive missing values, as above, this decision could result in an average miscalculation in the monthly average $T_{\text{mean}}$ of between 14 and 23% of a standard deviation, or 0.56 to 0.93°C. This would introduce a climatologically relevant error to our calculated monthly mean, threatening the accuracy of our results.

In addition to the inclusion of months with more missing values, we have shown that, while remaining cognizant of the potential error that we are introducing into our analysis, where data are missing at random, simple linear approximation will improve our calculated monthly means at low values of $j$ when the daily values of a year–month exhibit significant autocorrelation.

5 | CONCLUSIONS

In our trials, we have shown that, on average, the “3/5 rule” permits between 0.06 and 0.07 standard deviations of error ($\epsilon$) in the calculation of a monthly mean when three consecutive or five random values are missing from a month of daily data. This error would contribute a true absolute error of $\epsilon \sigma /N$ to the monthly normal, where $N$ is the number of year–months used in the calculation of the normal, and will vary from 23 to 31 (World Meteorological Organization, 1989). For its part, the “4/10 rule” permits $\epsilon$ of between 0.07 and 0.09 when four consecutive values are omitted, or up to 0.10 when 10 random values are omitted from a complete year–month. Of great importance is the fact that the proportional impact of missing values was reasonably similar across variables and stations, with a maximum difference of 0.02 when averaged across all 240 year–months. At any level of $k$ or $j$ missing values below 19, each additional missing value from a monthly data set contributes $\epsilon$ of between 0.008 and 0.018. February is disproportionately affected by missing values compared to other months; as a shorter month, each missing value represents a higher proportion of the overall data. Likewise, months with 31 days also tended to be less affected by the missing values.

Given the above, we have shown that the application of any “rule-of-thumb” around the exclusion of year–months with missing data should be carefully considered. In many cases $\epsilon$ at $k \in \{3, 4\}$ and $j \in \{5, 10\}$ represents a small proportion of a standard deviation. If we consider the true error to be $\epsilon \star \sigma$, the change is likely to be lost to rounding when calculating temperature for a climate normal. In this sense,
we consider that the lost information from an omitted year–month may prove to be as—or more—statistically harmful than the inclusion of a year–month with missing values. Indeed, we can consider an idealized example with a relatively high value of $j$. In our 20 years of analysis, the overall average maximum February temperature at Toronto was 0.55°C. Removing 15 random values from February 2004 (sampled with a seed of 287) results in $\epsilon T_{\text{max}}$ of 0.18 (a true error of 0.70°C for the average maximum temperature of that year–month). Including this year–month in the calculation of our period average would give us an overall calculated mean of 0.58°C. The magnitude of this error is, in fact, roughly equivalent to omitting the same month, which results in an average $T_{\text{max}}$ of 0.51°C.

We have shown that small changes to the limits of these “rules-of-thumb” can allow for the inclusion of more year–months of data, particularly in contexts where missing data is commonplace, and such “data rescue” techniques become necessary.

In addition, we found that even our “toy” interpolation method, simple linear interpolation, was somewhat effective at improving the estimated monthly mean at moderate values of $j$ when a station is characterized by lag-1 autocorrelation. More advanced interpolation and gap-filling techniques, particularly between-station methods, are likely to infer a greater reduction in $\epsilon$, and could, therefore further reduce the statistical risk of including months with missing data in monthly temperature series and the calculation of monthly average temperature for climate normals.

In this study, we examined $k$ consecutive values and $j$ random values independently, however, in practice researchers are unlikely to encounter a data set in which missing values are exclusively consecutive or exclusively scattered. Indeed, in our own study, no samples contained exclusively non-consecutive missing values at levels of $j$ above 11. As such, future study could build on our method and results to examine the interplay between $k$ and $j$, describing the impact of changes in the proportion of $k$ consecutive missing values of a total of $j$ missing values in a given year–month. Further, the results of the present study could be used to generate a model to optimize the threshold of missing values based on the lag-1 autocorrelation at a given station.

The results of the present study suggest that limits on the permissible number of missing daily values in a given year–month will be subject to the amount of error that a researcher is prepared to admit into a study, and should be based on the particular characteristics of the source data and the goals of the research project.

ACKNOWLEDGEMENTS
Thank you to Dr. Ken Butler for the suggestion to scale to the standard deviation of the daily values to ease comparison between stations. Thank you to Mr. Adam Dickinson for tips on post hoc tests in R. Thank you to the anonymous reviewers who provided insightful suggestions for improvement to this paper.

ORCID
Conor I. Anderson http://orcid.org/0000-0003-2619-1207

REFERENCES
Anderson, C.I. (2017) senanuhR: a collection of functions to obtain Peruvian climate data. R package version 0.4.2. Available at: https://gitlab.com/ConorA/senanuhR/. [Accessed 25 January 2018].
Anderson, C.I. (2018) claut: functions from the University of Toronto Climate Lab. R package version 0.1.3. Available at: https://gitlab.com/ConorA/claut/. [Accessed 25 January 2018].
Chapman, W.L. and Walsh, J.E. (1993) Recent variations of sea ice and air temperature in high latitudes. Bulletin of the American Meteorological Society, 74(1), 33–48. https://doi.org/10.1175/1520-0477(1993)074[0033:RVOSIA]2.0.CO;2.
Environment Canada. (2016) World Meteorological Organization (WMO) standards for “climate normals.” Available at: http://climate.weather.gc.ca/glossary_e.html. [Accessed 21 July 2016].
Fox, J. and Weisberg, S. (2011) An R Companion to Applied Regression, 2nd edition. Thousand Oaks, CA: Sage. Available at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion. [Accessed 25 January 2018].
Gough, W.A. (2008) Theoretical considerations of day-to-day temperature variability applied to Toronto and Calgary, Canada data. Theoretical and Applied Climatology, 94(1–2), 97–105. https://doi.org/10.1007/s00704-007-0346-9.
Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D. and New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. Journal of Geophysical Research: Atmospheres, 113(D20), D20119. https://doi.org/10.1029/2008JD010201.
Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008) Comparison of six methods for the interpolation of daily, European climate data. Journal of Geophysical Research: Atmospheres, 113(D21), D21110. https://doi.org/10.1029/2008JD010100.
Infante, S., Ortega and C. Cedeño, F. (2008) Estimación de datos faltantes en estaciones meteorológicas de Venezuela via un modelo de redes neuronales. Revista de Climaología, 8, 51–70.
Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R. (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environmental Modelling and Software with Environment Data News, 16(4), 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1.
Klein Tank, A.M.G., Wijngaard, J.B., Können, G.P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejlekli, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pääsdötter, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantas, A., Aberfeld, R., van Engelen, A.A.F.V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., Antonio López, J., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L.V. and Petrovic, P., C.L. (2014) Analysis and estimation of the effects of missing values on temperature and precipitation for 1950–2006. Theoretical and Applied Climatology, 117(3–4), 511–519. https://doi.org/10.1007/s00704-013-1024-8.
Mekis, E. and Vincent, L.A. (2011) An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. Atmospheric Ocean, 49(2), 163–177. https://doi.org/10.1080/07055900.2011.583910.
Mohsin, T. and Gough, W.A. (2012) Characterization and estimation of urban heat island at Toronto: impact of the choice of rural sites. Theoretical and Applied Climatology, 108(1–2), 105–117. https://doi.org/10.1007/s00704-011-0516-7.
Plummer, N., Allsopp, T., Lopez, J. and Lianso, P. (2003) Guidelines on climate observation networks and systems. Geneva: World Meteorological Organization. Technical report: WMO/TD No. 1185.
Pohlert, T. (2018) PMCMRplus: calculate pairwise multiple comparisons of mean rank sums extended. R package version 1.0.0. Available at: https://CRAN.R-project.org/package=PMCMRplus. [Accessed 25 January 2018].

Price, D.T., McKenney, D.W., Nalder, I.A., Hutchinson, M.F., and Kesteven, J.L. (2000) A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. Agricultural and Forest Meteorology, 101(2–3), 81–94. https://doi.org/10.1016/S0168-1923(99)00169-0.

R Core Team. (2017) R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/. [Accessed 25 January 2018].

Schneider, T. (2001) Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. Journal of Climate, 14(5), 853–871. https://doi.org/10.1175/1520-0442(2001)014h0853:AOICDE2.0.CO;2.

Simpson, G.L. (2016) canadaHCD: Canadian historical climate data. R package version 0.0-1. Available at: https://github.com/gavinsimpson/canadaHCD/. [Accessed 25 January 2018].

Stooksbury, D.E., Idso, C.D. and Hubbard, K.G. (1999) The effects of data gaps on the calculated monthly mean maximum and minimum temperatures in the continental United States: a spatial and temporal study. Journal of Climate, 12(5), 1524–1533. https://doi.org/10.1175/1520-0442(1999)012(1524:TEODGO2.0.CO;2).

Trewin, B.C. (2007) The role of climatological normals in a changing climate. Geneva: World Meteorological Organization. Technical report: WCDMP-No. 61, WMO/TD No. 1377.

Trujillo, A.M.T., Ramírez, R.A., Peña, M.A.V. and Castillo, L.A.I. (2015) Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño [Fill of daily series of precipitation, minimum and maximum temperature from the northern region of Urabá Antioquia]. Revista Mexicana de Ciencias Agrícolas, 6(2), 577–588.

Vincent, L.A., Wang, X.L., Milewska, E.J., Wan, H., Yang, F. and Swail, V. (2012) A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. Journal of Geophysical Research: Atmospheres, 117(D18), D18110. https://doi.org/10.1029/2012JD017859.

World Meteorological Organization. (1983) Guide to Climatological Practices, Vol. 100, 2nd edition. Geneva: World Meteorological Organization.

World Meteorological Organization. (1989) Calculation of monthly and annual 30-year standard normals. Washington, DC: World Meteorological Organization. Technical report: WMO-TD No. 341, WCDP-10.

World Meteorological Organization. (2011) Guide to Climatological Practices, Vol. 100, 3rd edition. Geneva: World Meteorological Organization.

World Meteorological Organization. (2016) Normals. In: Guide to Climatological Practices, Vol. 100, 3rd edition. Geneva: World Meteorological Organization.

Zeileis, A. and Grothendieck, G. (2005) Zoo: S3 infrastructure for regular and irregular time series. Journal of Statistical Software, 14(6), 1–27. https://doi.org/10.18637/jss.v014.i06.

Zhai, P., Sun, A., Ren, F., Liu, X., Gao, B. and Zhang, Q. (1999) Changes of climate extremes in China. Climatic Change, 42(1), 203–218. https://doi.org/10.1023/A:1005428602279.

Zubieta, R., Saavedra, M., Silva, Y. and Giráldez, L. (2017) Spatial analysis and temporal trends of daily precipitation concentration in the Mantaro River basin: central Andes of Peru. Stochastic Environmental Research and Risk Assessment, 31(6), 1305–1318. https://doi.org/10.1007/s00477-016-1235-5.

How to cite this article: Anderson CI, Gough WA. Accounting for missing data in monthly temperature series: Testing rule-of-thumb omission of months with missing values. Int J Climatol. 2018;38:4990–5002. https://doi.org/10.1002/joc.5801.