atmosphere

MDPI

Article

Analyzing Error Bounds for Seasonal-Trend Decomposition of Antarctica Temperature Time Series Involving Missing Data

Chun-Fung Kwok [1,†], Guoqi Qian [1,*,†] and Yuriy Kuleshov [2,3]

[1] School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia
[2] Bureau of Meteorology, Docklands, VIC 3008, Australia
[3] SPACE Research Centre, School of Science, Royal Melbourne Institute of Technology (RMIT) University, Melbourne, VIC 3000, Australia
* Correspondence: qguoqi@unimelb.edu.au
† These authors contributed equally to this work.

Abstract: In this paper, we study the problem of extracting trends from time series data involving missing values. In particular, we investigate a general class of procedures that impute the missing data and then extract trends using seasonal-trend decomposition based on loess (STL), where loess stands for locally weighted smoothing, a popular tool for describing the regression relationship between two variables by a smooth curve. We refer to them as the imputation-STL procedures. Two results are obtained in this paper. First, we settle a theoretical issue, namely the connection between imputation error and the overall error from estimating the trend. Specifically, we derive the bounds for the overall error in terms of the imputation error. This subsequently facilitates the error analysis of any imputation-STL procedure and justifies its use in practice. Second, we investigate loess-STL, a particular imputation-STL procedure with the imputation also being performed using loess. Through both theoretical arguments and simulation results, we show that loess-STL has the capacity of handling a high proportion of missing data and providing reliable trend estimates if the underlying trend is smooth and the missing data are dispersed over the time series. In addition to mathematical derivations and simulation study, we apply our loess-STL procedure to profile radiosonde records of upper air temperature at 22 Antarctic research stations covering the past 50 years. For purpose of illustration, we present in this paper only the results for Novolazaravskaja station which has temperature records with more than 8.4% dispersed missing values at 8 pressure levels from October/1969 to March/2011.

Keywords: imputation; local polynomial regression; smoothing; time series; trend extraction

Citation: Kwok, C.-F.; Qian, G.; Kuleshov, Y. Error bounds for Seasonal-Trend Decomposition Based on Loess with Missing Data. *Atmosphere* **2023**, *14*, 193. https://doi.org/10.3390/atmos14020193

Academic Editor: Zuntao Fu

Received: 11 December 2022
Revised: 10 January 2023
Accepted: 13 January 2023
Published: 17 January 2023

1. Introduction

Extracting trends from time series data is a central task in many fields, including economics, geophysics, climatology and engineering. Extensive research has been done on trend extraction methods, and these methods can be roughly divided into two groups, the smoothing-based approach and the non-smoothing-based approach. The division is due to the dominant role smoothing-based methods historically played in the context of trend extractions. Research along the smoothing-based path has yielded fruitful results and gained much popularity. Some well-known methods are Henderson filters [1], seasonal-trend decomposition based on loess (STL) [2], Hodrick-Prescott filters [3] and X12-ARIMA [4], which was updated to X-13-ARIMA in 2013. All these methods give a set of weights that are applied to the data as an averaging operator to give the underlying trend, and they differ mainly in the class of functions used for fitting and the smoothness criterion. They are also referred to as linear filters. For nonlinear filters, various methods have been suggested, see, for example, the optimal order statistic filter [5], the stack filters [6] and the median filters [7].

The non-smoothness-based approach has received more attention in later years. Two methods, originating from the field of signal processing, have made their way into the field

of time series analysis. They are the Singular Spectrum Analysis (SSA) and the Empirical Mode Decomposition (EMD). SSA is based on the idea of factorisation: it performs the Singular Value Decomposition to the covariance matrix of trajectory matrices to give a trend [8]. SSA has been actively developed [9–14] and applied to various kinds of data, e.g., climate data [15,16], financial data [17,18] and geophysical data [11]. On the other hand, EMD is related to the idea of orthogonal projection: the method decomposes signals into finite, nearly-orthogonal components that admit Hilbert transforms [19]. While the method works in the time domain, it can also be interpreted as a special case of the Wavelet methods which work in the frequency domain [20]. Due to its high adaptiveness to nonlinear and non-stationary data, EMD has been widely studied and applied, cf. [21–23]. For a comprehensive review of trend extraction methods, readers are referred to Alexandrov, Bianconcini, Dagum, Maass and McElroy [24].

Out of all the methods noted, we investigate in this work seasonal-trend decomposition based on loess, aka STL, in the context of trend extractions when there is missing data. We study STL because, compared with other more recent and mathematically sophisticated methods, STL has a broader user base. For one, STL is one of the few early methods that give a full decomposition of time series data (into the trend, seasonal and residual components) with almost no assumptions made about the data. Secondly, STL is easy to use and has good properties, e.g., it can handle non-stationary data and has fast convergence. The method also has the capacity to handle missing data, however, it seems this feature has not been implemented in practice. This poses a challenge to practitioners as the missing data problem is often encountered. A typical approach to circumvent the difficulty of missing data is to make complete the data with imputation methods. Numerous questions arise immediately upon the decision of doing so. To name a few, do imputations introduce bias? How reliable the STL estimates are after the imputation? To what extent are imputation methods able to recover missing data in the context of trend extraction with STL? It is the goal of this paper to settle these questions.

Moreover, we remark that imputing missing data before applying STL is not merely to avoid re-implementing a version of STL that can handle missing data. In fact, it addresses a problem that was not fully considered in the work of Cleveland et al. [2] on proposing STL. Cleveland et al. [2] suggested handling missing data with loess smoothing (in the cycle-subseries smoothing step of the work), but this is not possible when the missing data form large gaps in the observed data. By considering imputations before applying STL, we open up ourselves to many imputation methods so that different types of missing data can be handled, after which STL can be applied.

Regarding imputation methods and the general framework for handling missing data, the major theoretical issues were settled in [25–27]. Later developments were mostly about clarifications [28], implementations [29,30], and practical concerns [31,32]. One notable imputation method is the multivariate imputations by chained equation (MICE) [33] which deals with multivariate missing data (both categorical and numerical) and has raised some revised interest in recent years [34,35].

In this work, we study the problem of extracting trends from time series data when some data are missing. In particular, we investigate a general class of procedures that impute the missing data and then extract trends using STL. We refer to them as the imputation-STL procedures. Working under the settings given in [2], we derive an error bound for the extracted trends in terms of imputation errors. This answers the questions we posed earlier regarding the impact of imputations on the trend estimates. More importantly, this provides a framework for analysing errors of the trend extracted with any imputation-STL procedure. Apart from the theoretical results, we also examine a special case, the loess-STL procedure through simulation studies. We demonstrate that loess-STL provides reliable trend estimates when the ground-truth trend is smooth and the missing data disperse over the time series. This, together with the theoretical results, justifies the use of the procedure in practice. We also present an application to real data; specifically, we apply the loess-STL procedure to the Antarctic upper air temperature data and make available a profile of temperature trends for further climatological analysis.

The structure of this paper is as follows. In Section 2, we review the methods loess and STL for time series data and define some terminology. In Section 3, we present an error analysis of the imputation-STL procedures in the context of trend extraction with missing data. In Section 4, we present simulation studies with loess-STL procedures. In Section 5, we apply the loess-STL procedure to a real dataset of radiosonde records of upper air temperature at 22 Antarctic research stations covering the past 50 years, and we conclude in Section 6.

2. Methods Review and Terminology

In this section, we first review the methods of loess and STL by summarising the work of Cleveland, Cleveland and McRae [2]. This can be skipped by readers who are familiar with the methods. Then we define some terminology which we will use in the rest of the paper.

2.1. Loess

Locally weighted regression, aka loess, is a nonparametric method in regression analysis. It models the dependent variable as a smooth function of the independent variables; the smooth function is estimated by fitting the data locally with polynomials. The method adapts well to data and has several advantages. First, the flexible form incorporates a wider class of relationships beyond the linear. Second, no prior knowledge about the data is required (other than that the data are a representative sample), so subjective judgement can be avoided when little is known about the relationship between the variables. Third, it is useful for explanatory analysis, e.g., it can serve as a baseline for searching for good parametric models; it can also act as a benchmark against parametric models during model evaluations. However, these advantages come at a cost: like other nonparametric methods, loess requires more data than parametric models to get the same precision for the estimates. In the following, we detail the assumptions and the fitting procedure.

2.1.1. Assumptions

Loess assumes a data generating process of $Y_j = f(X_j) + \epsilon_j$, $j = 1, 2, \ldots, N$, where Y_j are observations of the dependent variable Y, X_j are those of the independent variable X, N is the total number of observed data points and ϵ_j are independent normal random variables with mean 0 and variance σ^2. The function f specifies the functional relationship between the dependent variable Y and the independent variable X and is assumed to be smooth. This justifies the use of Taylor's theorem, which gives grounds for approximating functions locally by polynomials. The normality assumption about the data-generating process allows the distribution of residuals, fitted values and residual sum of squares to be represented by some known parametric families of distributions. In particular, given the assumption, the residuals and the fitted values are normally distributed provided that σ^2 is known, and the residual sum of squares follows a chi-squared distribution [36]. The distributional results make it possible to assess the uncertainty in these quantities. Loess also assumes that the estimate $\hat{f}$ approximates f with no bias. The assumption is not unrealistic as it is shown on p. 62 in [37] that under some mild conditions, the estimate is asymptotically unbiased. Each assumption we saw is associated with a particular feature of the method, and it is possible to forgo some of the properties for greater generality of the model. For instance, ref. [38] relaxed the normality assumption using the idea of robust regression by Huber [39]. In this case, a Monte Carlo simulation is then needed to assess the standard error.

2.1.2. Fitting Procedure

Loess approximates the functional relationship f by fitting a polynomial locally at each point x (in the domain of f) using points in the neighbourhood of x. The fitting uses weighted least square(WLS) regression. Overall, three quantities are needed in this procedure, the degree of polynomials to be fit, the size of the neighbourhood and the

weights for performing the WLS regression. Regarding the degree of polynomials to be fit, first-degree or second-degree polynomials are commonly used and are usually sufficient as long as the functional relationship is not too erratic. Quadratic fitting is generally preferred over linear fitting near extrema [37]. Alternatively, the degree of polynomials can be chosen using M-plots as suggested in [36]. Regarding the neighbourhood size, as it directly controls the smoothness of the estimates, the choice should be made based on the research context. The neighbourhood size is chosen such that the resulting estimate answers the research question in some optimal sense. But in cases where one wants to avoid subjectivity, the neighbourhood size can be chosen through data-driven techniques like cross-validation. Regarding the weights for the WLS regression, we will calculate them based on the tricube weight function

$$W(u) = \begin{cases} (1-u^3)^3 & \text{for } 0 \le u < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where u is a dummy variable. Concretely, suppose we have N data points, the degree of polynomial to be fit is d and the size of the neighbourhood is q, and we want to fit a polynomial locally at the point (X_1, Y_1). First, we identify the q data points, denoted by $(a_1, b_1), (a_2, b_2), \ldots, (a_q, b_q)$, so that $a_1, \ldots, a_q$ are nearest to X_1. Next, to each of these points, we assign a weight,

$$v_i(X_1) = W\left(\frac{|a_i - X_1|}{\max_i |a_i - X_1|}\right),\ i = 1, \ldots, q$$

where $W(\cdot)$ is the tricube weight function and the maximum is taken over the points in the neighbourhood, i.e., $(a_1, b_1), (a_2, b_2), \ldots, (a_q, b_q)$. Then we fit a degree-$d$ polynomial, denoted by $p_{d,X_1}(x)$, to these points using weighted least square regression. In the following step, the fitted value for Y_1 at X_1 is given by $p_{d,X_1}(X_1)$. Finally, the above procedure is repeated for each (X_j, Y_j), $j = 1, \ldots, N$, with each Y_j being fitted by $p_{d,X_j}(X_j)$.

2.2. STL

Seasonal-trend decomposition based on loess, aka STL, is a method of decomposition of time series. Through iterative smoothing of the data, it decomposes a time series into three components: the trend component, the seasonal component, and the remainder component. From a frequency analysis point of view, what STL does is filter out signals of different frequencies; the signal with the lowest frequency is regarded as the trend, the one with the medium frequency is the seasonal component, and the ones with the highest frequencies are the remainder (also named the noise in STL). Several advantages are using STL for time series decomposition. First, it only makes weak assumptions about the data-generating process, so it handles a wide class of data. Second, the computation is fast, and it can handle missing data and outliers. Third, prior knowledge about the components can be incorporated into the model.

To apply STL, six parameters need to be specified, they are the number of outer loops n_o, the number of inner loops n_i, the number of cycle-subseries n_p, the neighbourhood size for seasonal smoothing n_s, the neighbourhood size for trend smoothing n_t, and the neighbourhood size for seasonal trend smoothing n_l. Cleveland et al. [2] recommend the following choice of parameters. $n_o = 1, n_i = 2$ if resistance to outliers are not needed, and $n_o = 5, n_i = 1$ otherwise; n_p depends on the application, for example, 12 would be an appropriate choice for monthly climate data; n_s is specified by the user to incorporate prior knowledge of the regularity of the seasonal pattern. This must be an odd integer ≥ 7; $n_t = \left[1.5n_p/(1 - 1.5n_s^{-1})\right]_{odd}$ and $n_l = [n_p]_{odd}$, where $[\,\cdot\,]_{odd}$ is an operator such that $[x]_{odd}$, for any real number x, equals the smallest odd integer greater than or equal to x. Readers are referred to the original paper for full details on the reasoning behind these choices.

Suppose we have a time series of N data points $\{Y_\nu, \nu = 1, \ldots, N\}$, and the parameters of the STL have been chosen, then the STL procedure works as follows. First we initialise the robustness weight $\rho_\nu = 1, \nu = 1, \ldots, N$ and the trend $T_\nu = 0, \nu = 1, \ldots, N$. Then we feed them into the inner loop, which does the following three things.

1. Detrend the data: The procedure subtracts the inputted trend from the data, i.e., $Y_v^{detrend} = Y_v - T_v,\ v = 1, 2, \ldots, N$.
2. Extract the seasonal component: this involves 7 minor steps. Denoting the data $Y_1, Y_2, ..., Y_N$ by $\mathbf{Y}$, the steps are:
   (a) Break into n_p cycle-subseries $\mathbf{Y}^{detrend} \rightarrow (\mathbf{Y}^{\text{sub-1}}, \mathbf{Y}^{\text{sub-2}}, \ldots, \mathbf{Y}^{\text{sub-}n_p})$.
   (b) Smooth each cycle-subseries using loess with a degree of the polynomial set to 1 and neighbourhood size set to n_s, i.e., $\mathbf{Y}^{\text{sub-i-smooth}} = loess(\mathbf{Y}^{\text{sub-i}}, d = 1, q = n_s), i = 1, \ldots, n_p$. Note that smoothed values are computed from the position just prior to the first point to the position just after the last point.
   (c) Combine smoothed cycle-subseries to get the seasonal component $\mathbf{C}$, i.e., $(\mathbf{Y}^{\text{sub-1-smooth}}, \mathbf{Y}^{\text{sub-2-smooth}}, \ldots, \mathbf{Y}^{\text{sub-}n_p\text{-smooth}}) \rightarrow \mathbf{C}$.
   (d) Run a moving average filter of length n_p through $\mathbf{C}$ twice.
   (e) Then a moving average filter of length 3 once. The result is still denoted as $\mathbf{C}$.
   (f) Extract the seasonal trend vector $\mathbf{L} = (L_1, \ldots, L_N)$ from the smoothed seasonal component vector $\mathbf{C}$, i.e., apply loess smoothing with a degree of the polynomial set to 1, and neighbourhood size set to n_s.
   (g) Detrend the seasonal component, i.e., compute $\mathbf{S} = \mathbf{C} - \mathbf{L}$, with $S_v = C_v - L_v,\ v = 1, 2, \ldots, N$.
3. Deseasonalise the data and extract a new trend: The deseasonlising is done by subtracting from the data the detrended seasonal component estimated in the last step, i.e., $Y_v^{Deseasonalised} = Y_v - S_v,\ v = 1, 2, \ldots, N$, and the extraction of the new trend is done by performing loess smoothing on the deseasonalised data with a degree of the polynomial set to 1 and neighbourhood size set to n_t, i.e., $T_v = loess(Y_v^{Deseasonalised}, d = 1, q = n_t)$.

After a single pass of the inner loop, we get a revised trend and a seasonal component. If the data has not passed through the inner loop for n_i times, then we feed the revised trend into the inner loop again. Otherwise, both the revised trend and the seasonal component are fed into the outer loop, which does the following.

1. Compute the remainder component with $R_v = Y_v - T_v - S_v, v = 1, \ldots, N$.
2. Update the robustness weights with $\rho_v = B\left(\frac{|R_v|}{6 \cdot \text{median}_{i=1,\cdots,N}\{|R_i|\}}\right),\ v = 1, \ldots, N$, where $B(u) = (1 - u^2)^2 \cdot \mathbf{1}_{0 \le u < 1}$ is the bisquare weight function suggested in [2], with $\mathbf{1}_{0 \le u < 1} = 1$ if $u \in [0, 1)$, and $= 0$ if $u \notin [0, 1)$. The idea is to give little weight to any point far apart from the rest; this is to remove distortions of the result by outliers.

After a single pass of the outer loop, we get a full decomposition of the data (i.e., the trend, seasonal and remainders components) and revised robustness weights. If the data has not passed through the outer loop for n_o times, then we feed everything into a new round of inner loops. Otherwise, the procedure ends and returns the full decomposition. For readers' convenience, we give in Figure 1 a schematic representation of the algorithm.

2.3. Terminology

We will use the following terminology throughout the remaining sections. We first talk about data. We define

- the *missing dataset* to be the observed dataset that has some data missing,
- the *complete dataset* to be the dataset without any parts missing—assuming it actually exists in the first place, and
- the *imputed dataset* to be the dataset we get after applying imputation methods to the missing dataset.

Next, we talk about trends. The trend is the long-term low-frequency signal, obtained by our procedure, which in the simplified form is deseasonalising the data and then smoothing the result to remove short-term fluctuation. We define

- the *complete trend* to be the trend estimated with the complete dataset,

- the *imputed trend* to be the trend estimated with the imputed dataset, and
- the *true trend* to be the true underlying trend.

In Section 3, we relate the imputed trend to the complete trend in terms of imputation errors. In Section 4, we verify the result through simulations, demonstrating that the imputed trend can approximate the true trend well. Combining the results from both sections, we see a more complete picture of how missing data affects the trend estimate. This helps us identify the situations where imputation-STL procedures can give reliable trend estimates.

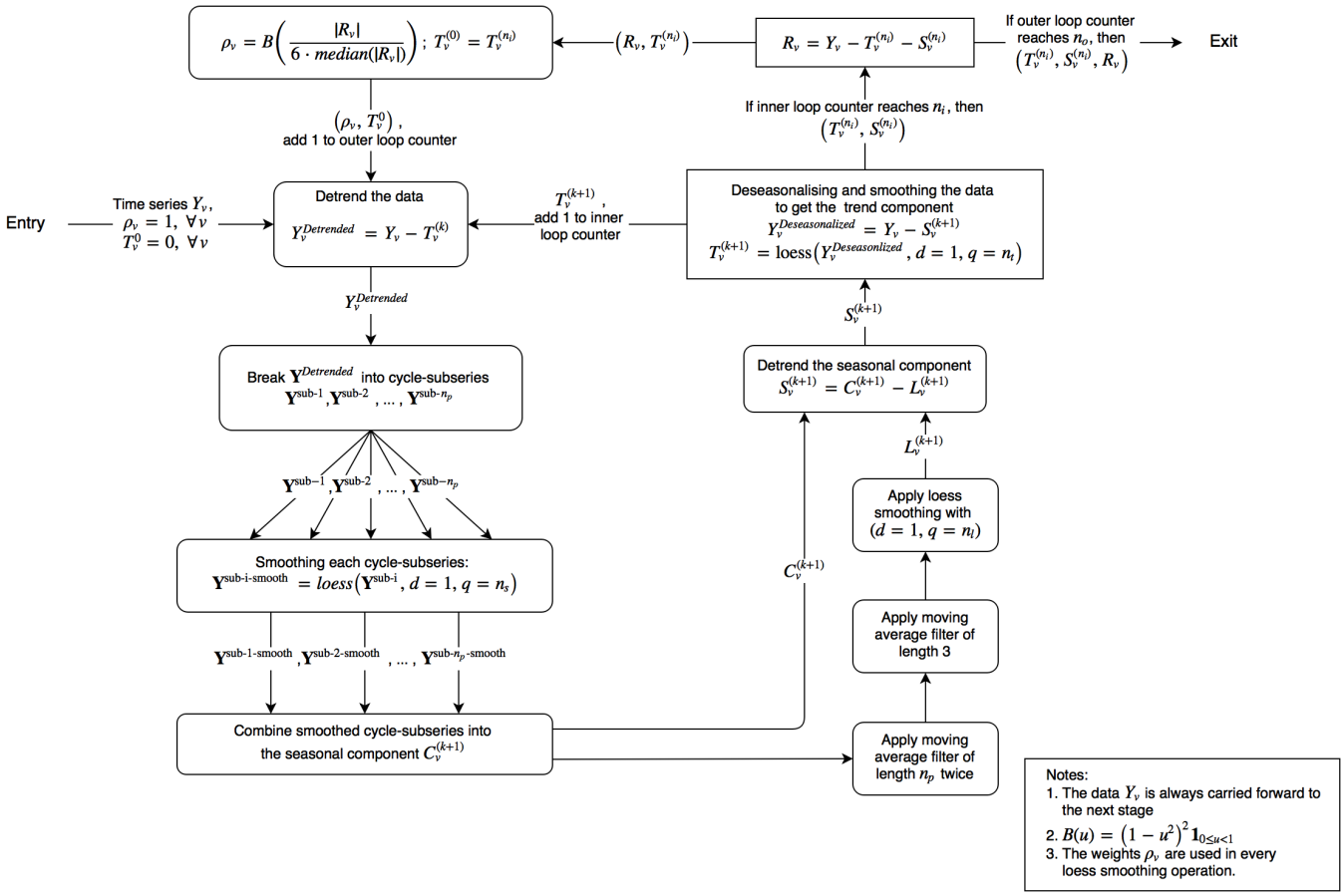

Figure 1. Schematic representation of STL, where k is a running integer index in the inner loop starting from 0 to n_i.

3. Error Analysis of STL with Imputations

In this section, we first analyse the errors of the trend estimates from the imputation-STL class of procedures. Then we investigate a particular case, the loess-STL procedure. Lastly, we conclude the section with some remarks.

3.1. Error Bound for the Estimated Trend from an Imputation-STL Procedure

We first define the terms *trend error* and *imputation error*, and then we present our results. The trend error is defined to be the mean of the squared differences between the complete trend and the imputed trend. For complete dataset $\{Y_j,\ j = 1, \ldots, N\}$, denote the complete trend by $\{z_j,\ j = 1, \ldots, N\}$. For imputed dataset $\{Y_j^{imputed},\ j = 1, \ldots, N\}$, denote the corresponding imputed-trend by $\{z_j^{imputed},\ j = 1, \ldots, N\}$. Then the trend error MSE_{trend} is given by

$$MSE_{trend} = \frac{\sum_{j=1}^{N} (z_j - z_j^{imputed})^2}{N} = \frac{\|\mathbf{z} - \mathbf{z}^{imputed}\|^2}{N},$$

where $\mathbf{z}, \mathbf{z}^{imputed}$ are z_j and $z_j^{imputed}$ $(j = 1, \ldots, N)$ in vector notation, and $\|\cdot\|$ denotes the Euclidean norm. Similarly, the imputation error is defined to be the mean of the squared differences between the complete dataset and the imputed dataset. With the above introduced notations, the imputation error $MSE_{imputation}$ is given by

$$MSE_{imputation} = \frac{\sum_{j=1}^{N}(Y_j - Y_j^{imputed})^2}{N} = \frac{\|\mathbf{Y} - \mathbf{Y}^{imputed}\|^2}{N},$$

where $\mathbf{Y}, \mathbf{Y}^{imputed}$ are Y_j and $Y_j^{imputed}$ $(j = 1, \ldots, N)$ in vector notation.

Our results can be stated plainly as follows. Assuming the settings given in Cleveland et al. [2], the trend estimated by an imputation-STL procedure has an error bounded above by a constant multiplying the imputation error. This is useful in two ways. First, one can now explicitly assess the trend error with the imputation error. In other words, our result answers the question "How much error do I get in my estimated trend if my imputations are wrong by X amount?". Second, given a desired accuracy level, our result specifies the amount of imputation error a data set can tolerate. In addition, it also specifies how much improvement the estimate could have when new data points become available. Further discussion is given at the end of this section; now we present the mathematical statement.

Theorem 1. *Suppose a time series is circular, and the parameters of STL are chosen according to Section 2.2, then the trend produced by an imputation-STL type of procedure satisfies that*

$$MSE_{trend} \leq L \cdot MSE_{imputation},$$

where $L = (2n_i)^2$ *with* n_i *being the number of inner loops chosen.*

We first state the theoretical settings and consequences given in Cleveland et al. [2] and then several lemmas, before we can prove Theorem 1.

Settings: The data $\{Y_1, \ldots, Y_N\}$ is assumed coming from a circular time series $\{Y_i, i = 1, 2, \ldots\}$ of period length N; namely $Y_i = Y_j$ if $i \equiv j \ (mod\ N)$, i.e., $i - j$ is divisible by N. Also, the parameter choices follow the recommendation as given in Section 2.2.

Consequences: Denote the operator matrices associated with the operations in steps 2 and 3 of Section 2.2 by $\mathcal{S}$ and $\mathcal{T}$ respectively. To be clear, $\mathcal{S}$ is the $N \times N$ operator matrix that takes the input $\mathbf{Y}^{detrend}$ and outputs the seasonal component $\mathbf{S}$, and $\mathcal{T}$ is the $N \times N$ operator matrix that takes the input $\mathbf{Y}^{Deseasonalised}$ and outputs the revised trend $\mathbf{T}$. Given the above and by [40], we have

(C1) $\{Y_1, \ldots, Y_N\}$ being from a circular time series implies $\mathcal{S}$ and $\mathcal{T}$ can be augmented to be circulant matrices;

(C2) Enforcing the parameter choices in Section 2.2 implies that all eigenvalues of $\mathcal{S}$ and $\mathcal{T}$ are inside or on the unit circle, and $\mathcal{S}$ and $\mathcal{T}$ each have, at most, one eigenvalue on the unit circle.

The following definition and lemmas are taken from pages 104 and 113–114 in [40], except Lemma 1(a) which follows directly from the usual definition of induced operator norms.

Lemma 1. *Let* $\mathcal{A}$ *be a matrix and consider* $\mathcal{A} : x \mapsto \mathcal{A}x, x \in \mathbb{R}^m$ *as an operator, then*

(a) $||\mathcal{A}x|| \leq ||\mathcal{A}||_{op}||x||$, *where* $||\cdot||$ *is the Euclidean norm, and* $||\cdot||_{op}$ *is the induced operator norm;*

(b) $||\mathcal{A}||_{op} = \sqrt{\lambda_{(G)}(\mathcal{A}^\top \mathcal{A})}$, *where* $\lambda_{(G)}(\mathcal{A}^\top \mathcal{A})$ *denotes the greatest eigenvalue of* $\mathcal{A}^\top \mathcal{A}$.

Definition 1. *An* $n \times n$ *matrix of the form*

$$circ(a_0, a_1, ..., a_{n-1}) = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-3} & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-4} & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_0 & a_1 \\ a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_0 \end{bmatrix}$$

is named a circulant matrix.

Lemma 2. *If $\mathcal{A} = circ(a_0, a_1, ..., a_{n-1})$ and $\mathcal{B} = circ(b_0, b_1, ..., b_{n-1})$ are circulant matrices, then*

(a) *$\mathcal{A} + \mathcal{B}$ and $\mathcal{AB}$ are circulant;*

(b) *All eigenvalues of $\mathcal{A}$ are given by $\lambda_j = \sum_{l=0}^{n-1} a_l \exp(2\pi l j n^{-1}\sqrt{-1})$, $j = 0, 1, \ldots, n-1$;*

(c) *$\lambda_j(\mathcal{AB}) = \lambda_j(\mathcal{A})\lambda_j(\mathcal{B})$ and $\lambda_j(\mathcal{A} + \mathcal{B}) = \lambda_j(\mathcal{A}) + \lambda_j(\mathcal{B})$, where $\lambda_j(\mathcal{A})$ is the j-th eigenvalue $\mathcal{A}$, and $\lambda_j(\mathcal{B})$ and $\lambda_j(\mathcal{AB})$ are similarly defined.*

Now we present the proof of Theorem 1.

Proof. Denote the complete data by $\mathbf{Y} = (Y_1, \ldots, Y_N)^\top$ and the resultant imputed data by $\mathbf{Y}^{imputed}$. Also, we denote the operator matrix of the trend filter of STL after k iterations of the inner loop by $\mathcal{T}_k$; the expression for $\mathcal{T}_k$ can be shown to be

$$\mathcal{T}_k = \sum_{m=1}^{2k} (-1)^{m-1} \mathcal{B}_m, \text{ where } \mathcal{B}_m = \begin{cases} (\mathcal{TS})^{m/2} & \text{for } m \text{ even} \\ (\mathcal{TS})^{(m-1)/2}\mathcal{T} & \text{for } m \text{ odd} \end{cases} \tag{1}$$

We want to compare the trend extracted from the complete data and the trend extracted from the imputed data. Hence, we compute the mean squared difference between the two trends, $MSE_{trend} = ||\mathcal{T}_k\mathbf{Y} - \mathcal{T}_k\mathbf{Y}^{imputed}||^2/N$. By Lemma 1(a), we have

$$\frac{||\mathcal{T}_k\mathbf{Y} - \mathcal{T}_k\mathbf{Y}^{imputed}||^2}{N} = \frac{||\mathcal{T}_k(\mathbf{Y} - \mathbf{Y}^{imputed})||^2}{N} \leq \frac{||\mathcal{T}_k||_{op}^2 \cdot ||\mathbf{Y} - \mathbf{Y}^{imputed}||^2}{N} \tag{2}$$

Now we evaluate $||\mathcal{T}_k||_{op}$. First from (1), we have

$$\begin{aligned} \mathcal{T}_k &= \sum_{m=1}^{2k} (-1)^{m-1} \mathcal{B}_m \\ &= \mathcal{T} - \mathcal{TS} + (\mathcal{TS})\mathcal{T} - (\mathcal{TS})^2 + (\mathcal{TS})^2\mathcal{T} - (\mathcal{TS})^3 + ... + (\mathcal{TS})^{k-1}\mathcal{T} - (\mathcal{TS})^k \\ &= \left[\mathcal{T} + (\mathcal{TS})\mathcal{T} + (\mathcal{TS})^2\mathcal{T} + ... + (\mathcal{TS})^{k-1}\mathcal{T}\right] - \left[\mathcal{TS} + (\mathcal{TS})^2 + (\mathcal{TS})^3 + ... + (\mathcal{TS})^k\right] \\ &= \left[\mathcal{I} + \mathcal{TS} + (\mathcal{TS}^2 + (\mathcal{TS})^3 + ... + (\mathcal{TS})^{k-1}\right]\mathcal{T} - \left[\mathcal{TS} + (\mathcal{TS})^2 + (\mathcal{TS})^3 + ... + (\mathcal{TS})^k\right] \\ &= \left[\mathcal{I} + \mathcal{TS} + (\mathcal{TS})^2 + (\mathcal{TS})^3 + ... + (\mathcal{TS})^{k-1}\right](\mathcal{T} - \mathcal{TS}). \end{aligned} \tag{3}$$

Note that $\mathcal{T}$, $\mathcal{S}$ (by **(C1)**) and $\mathcal{I}$ are circulant matrices, so Lemma 2(a) implies $\mathcal{T}_k$ is circulant. It then follows from Lemma 2(c) that $\lambda_j(\mathcal{T}_k^\top\mathcal{T}_k) = \lambda_j(\mathcal{T}_k^\top)\lambda_j(\mathcal{T}_k)$. Next, note that $\mathcal{A} = circ(a_0, a_1, ..., a_{n-1}) \Rightarrow \mathcal{A}^\top = circ(a_0, a_{n-1}, a_{n-2}, ..., a_1)$ and denoting the eigenvalues of $\mathcal{A}$ by λ_j's and the eigenvalues of $\mathcal{A}^\top$ by $\hat{\lambda}_j$'s, we have by Lemma 2(b),

$$\lambda_j = a_0 + \sum_{l=1}^{n-1} a_l \exp\left(\frac{2\pi\sqrt{-1}jl}{n}\right)$$

and

$$\hat{\lambda}_j = a_0 + \sum_{l=1}^{n-1} a_{n-l} \exp\left(\frac{2\pi\sqrt{-1}jl}{n}\right) = a_0 + \sum_{l=1}^{n-1} a_l \exp\left[\frac{2\pi\sqrt{-1}j(n-l)}{n}\right] = \lambda_{n-j}.$$

Setting $\mathcal{A} = \mathcal{T}_k$, it follows $\lambda_j(\mathcal{T}_k^\top \mathcal{T}_k) = \lambda_j(\mathcal{T}_k^\top)\lambda_j(\mathcal{T}_k) = \lambda_{N-j}(\mathcal{T}_k)\lambda_j(\mathcal{T}_k) = |\lambda_j(\mathcal{T}_k)|^2$. We will need this later to get $||T_k||_{op}$. Continuing from (3) and applying Lemma 2(c) give

$$\begin{aligned}\lambda_j(\mathcal{T}_k) &= \lambda_j\Big(\mathcal{I} + \mathcal{TS} + (\mathcal{TS})^2 + (\mathcal{TS})^3 + \cdots + (\mathcal{TS})^{k-1}\Big) \cdot \lambda_j(\mathcal{T} - \mathcal{TS}) \\ &= \lambda_j(\mathcal{I}) + \lambda_j(\mathcal{TS}) + \lambda_j((\mathcal{TS})^2) + \lambda_j((\mathcal{TS})^3) + \cdots + \lambda_j((\mathcal{TS})^{k-1})) \cdot \left[\lambda_j(\mathcal{T}) - \lambda_j(\mathcal{TS})\right] \\ &= \left[1 + t_j s_j + (t_j s_j)^2 + \cdots + (t_j s_j)^{k-1}\right] t_j\,(1 - s_j).\end{aligned}$$

where $\lambda_j(\mathcal{T}) = t_j$ and $\lambda_j(\mathcal{S}) = s_j$ are the j-th eigenvalues of $\mathcal{T}$ and $\mathcal{S}$ respectively, and $\lambda_j(\mathcal{I}) = 1$. Note that the indexing strictly follows that of Lemma 2(b). Now we take modulus on both sides to get

$$|\lambda_j(\mathcal{T}_k)| = \left|\left[1 + t_j s_j + (t_j s_j)^2 + ... + (t_j s_j)^{k-1}\right] t_j\,(1 - s_j)\right|.$$

Let M be the index where $|\lambda_j(\mathcal{T}_k)|$ is maximised, i.e., $\max_j |\lambda_j(\mathcal{T}_k)| = |\lambda_M(\mathcal{T}_k)|$ and let t and s be the M-th eigenvalues of $\mathcal{T}$ and $\mathcal{S}$, respectively. Then we have

$$\begin{aligned}|\lambda_M(\mathcal{T}_k)| &= \left|\left[1 + ts + (ts)^2 + ... + (ts)^{k-1}\right] t\,(1 - s)\right| \\ &\leq \left|1 + ts + (ts)^2 + ... + (ts)^{k-1}\right| |t|\,|(1 - s)| \\ &\leq \left[1 + |t||s| + (|t||s|)^2 + ... + (|t||s|)^{k-1}\right] |t|\,|(1 - s)| \leq 2k\end{aligned}$$

where the last inequality above is implied by **(C2)**.

Finally, by Lemmas 1(b) and the result we referenced previously, we have $||\mathcal{T}_k||_{op}^2 = \lambda_{(G)}(\mathcal{T}_k^\top \mathcal{T}_k) = |\lambda_M(\mathcal{T}_k)|^2$. Substituting this back into (2), we have

$$MSE_{trend} \leq (2k)^2 \cdot \frac{||\mathbf{Y} - \mathbf{Y}^{imputed}||^2}{N} \leq L \cdot MSE_{imputation} \tag{4}$$

Now the proof is completed noting that by definition $MSE_{imputation} = N^{-1}||\mathbf{Y} - \mathbf{Y}^{imputed}||^2$ and $k \leq n_i$ with n_i being the inner loop size in STL, and $L = (2n_i)^2$.

As a particular case, if we enforce the parameter choice from Section 2.2, then we have

$$MSE_{trend} \leq \begin{cases} 4 \cdot MSE_{imputation} & \text{if } n_i = 1 \\ 16 \cdot MSE_{imputation} & \text{if } n_i = 2. \end{cases} \tag{5}$$

□

With the expression we derived, we continue our discussion about the result. First, we now have an upper bound for the error of the estimated trend in terms of the squared imputation error. So if the imputation error is known, then we know how large the trend error would be in the worst-case scenario. In practice, the imputation error is however unknown; we will discuss how to estimate the imputation error in the next section. Second, if we examine closely the right hand side of (4), we see two quantities affecting the upper bound, namely the squared total imputation error $||\mathbf{Y} - \mathbf{Y}^{imputed}||^2$ and the number of data points "originally" available, N. They have the following implications:

1. Since the constant L is fixed beforehand, its influence on the upper bound becomes negligible when N is far larger than L. For the same reason, large imputation errors for a few data points would not cause too much trouble.
2. The expression suggests the imputation errors at individual data points can grow, e.g., we have a faulty machine that goes wrong consistently once in a while and produces missing data which are then imputed with some errors. Remarkably, we know precisely how fast the total imputation error can grow before it is no longer possible to keep the trend error small. Expression (4) says as long as $||\mathbf{Y} - \mathbf{Y}^{imputed}||^2$

grows at a rate (strictly) slower than N, then we will not lose much precision in our estimated trend.

3.2. Error Bound for STL with Loess Imputation

In this section, we investigate the loess-STL procedure, which is a particular case of the imputation-STL class of procedures. We made this choice because smoothing methods are widely used in different branches of statistics. Furthermore, interpolation by smoothing is a common way to handle missing data. The goal of this section is twofold. One, we illustrate concretely how to apply oerror-boundund results in practice. Two, this prepares us for the simulation studies we conduct in the next section.

Consider a case where data is missing every other point, i.e., out of the complete data $Y_1, Y_2, Y_3, ..., Y_N$, we have $Y_2, Y_4, Y_6, ...$ missing. One such example is that, for a monthly mean temperature time series, the data is missing every second month. We study this case because it corresponds to the worst-case scenario of the dispersing missing pattern. By dispersing missing data patterns, we mean there are no consecutive missing data points. In other words, the missing-ness incidences do not cluster and form any gap of size greater than or equal to 2.

We use the same notation as the previous section, i.e., denote the time series data by $\mathbf{Y}$ and the data with imputed values by $\mathbf{Y}^{imputed}$. We also write $Y_j = f(X_j)$, where X_j can be regarded as the time at which Y_j is observed, $j = 1, 2, ..., N$. As it only makes sense to apply loess smoothing when the underlying trend is smooth, we assume f to be twice differentiable. For mathematical convenience, we also assume $f''(X)$ to be bounded, i.e., $|f''(X)| \leq D$ for some constant $D > 0$. (Further justification is given at the end of this section.)

In the following, we find the imputation error at a point, i.e., we find $|Y_j - Y_j^{imputed}| = |f(X_j) - \hat{f}(X_j)|$. First, by Taylor's expansion theorem, we have

$$f(X_j + h) = f(X_j) + f'(X_j)h + \frac{f''(\zeta)}{2}h^2,$$

where $\zeta \in (X_j, X_j + h)$ and h is a small increment. Next, since loess is a linear smoother, we can express it in a form of equivalent kernels. We denote the kernel weight associated with a data point X_i by $w_{i|j}$ when smoothing $f(X_j)$; then the loess estimator is given by

$$\hat{f}(X_j) = \sum_{X_i \in \mathbb{N}(X_j)} w_{i|j} f(X_i) \tag{6}$$

$$= \sum_{X_i \in \mathbb{N}(X_j)} w_{i|j} \left[f(X_j) + f'(X_j)(X_i - X_j) + \frac{f''(\zeta_{i|j})(X_i - X_j)^2}{2} \right], \tag{7}$$

where $\sum_i w_{i|j} = 1$, $\zeta_{i|j} \in \big(\min(X_i, X_j), \max(X_i, X_j)\big)$ and $\mathbb{N}(X_j)$ is the set containing the neighbours of X_j.

Note that as we have time series data, we know data points are equally spaced of 1-unit distance in the time domain. This together with our assumption of the missing data pattern implies that we have symmetric neighbourhoods and weights for the loess imputation. Hence, $\sum_{X_i \in \mathbb{N}(X_j)} w_{i|j} f'(X_j)(X_i - X_j) = 0$, and we have

$$\begin{aligned} |\hat{f}(X_j) - f(X_j)| &= \left| \sum_{X_i \in \mathbb{N}(X_j)} w_{i|j} \frac{f''(\zeta_{i|j})(X_i - X_j)^2}{2} \right| \leq \sum_{X_i \in \mathbb{N}(X_j)} \left| w_{i|j} \frac{f''(\zeta_{i|j})(X_i - X_j)^2}{2} \right| \\ &\leq \sum_{X_i \in \mathbb{N}(X_j)} D \left| w_{i|j} \frac{(X_i - X_j)^2}{2} \right| = \sum_{X_i \in \mathbb{N}'(X_j)} D \left| w_{i|j} (X_i - X_j)^2 \right| \end{aligned} \tag{8}$$

where $\mathbb{N}'(X_j)$ is the one-sided neighbourhood of X_j. Now suppose the size of the neighbourhood of X_j is $2l$. Then the kernel weights $w_{i|j}$ is given by

$$W(|X_i - X_j|/|X_{l+j} - X_j|) / \sum_{\ell=-l}^{l} W(|X_{\ell+j} - X_j|/|X_{l+j} - X_j|), \quad i = -l+j, -l+1+j, \ldots, l+j,$$

where $W(\cdot)$ is the tricube weight function. Inequality (8) cannot be directly applied to the aforementioned complete data $\{Y_j,\ j = 1, \ldots, N\}$ which is recorded at $\{X_j = j,\ j = 1, 3, 5 \ldots\}$ but missing at $\{X_j = j,\ j = 2, 4, 6 \ldots\}$. However, with slight variation in the proof we obtain the following

$$|\hat{f}(X_j) - f(X_j)| \leq \sum_{i=1}^{l} D \left| \frac{W((2i-1)/(2l-1))}{\sum_{\ell=-l+1}^{l} W((2\ell-1)/(2l-1))} (2i-1)^2 \right|;$$

and with the assistance of computer algebra software such as Maple 2020 we have shown that the expression on the right-hand side is asymptotically

$$D \cdot \left[\frac{217}{1000} - \frac{7l}{25} + \frac{7l^2}{60} + O\left(\frac{1}{l}\right) \right].$$

We have thus found an upper bound for the imputation error at the point X_j. We can use the bound directly as a conservative estimate of the imputation error. To get the total squared imputation error over the whole time series, we square the individual errors and sum up all the missing data points. In our current setting, we have $N/2$ points missing(ignoring odd-even parity). Assuming for simplicity the neighbourhood size is the same for all points, the total squared imputation error is

$$||\mathbf{Y} - \mathbf{Y}^{imputed}||^2 = \frac{ND^2}{2} \cdot \left[\frac{217}{1000} - \frac{7l}{25} + \frac{7l^2}{60} + O\left(\frac{1}{l}\right) \right]^2$$

Now we can substitute this back to (5) and get

$$MSE_{trend} \leq \begin{cases} 2D^2 \cdot \left[\frac{217}{1000} - \frac{7l}{25} + \frac{7l^2}{60} + O\left(\frac{1}{l}\right) \right]^2 & \text{if } n_i = 1 \\ 8D^2 \cdot \left[\frac{217}{1000} - \frac{7l}{25} + \frac{7l^2}{60} + O\left(\frac{1}{l}\right) \right]^2 & \text{if } n_i = 2 \end{cases} \tag{9}$$

We arrive at the expression that tells us how large the trend error can be when loess-STL is applied. Overall, we illustrated how to apply our result from Section 3.1 when a particular imputation method is considered.

3.3. Remarks and Some Practical Concerns

Here we present remarks about Sections 3.1 and 3.2 in the order of descending importance.

- How to estimate D, the upper bound of $|f''(x)|$? One way is to smooth the data with any choice of smoother, then differentiate the resulting curve twice to get D. There are various packages in R [41] that can handle this, e.g., the *fda* package. In some cases, eyeballing can give a rough but quick estimate. One simply traces the curve with a tangent line, records the maximum slope and the minimum slope, and then takes the difference.
- What to do if the estimated D gives a loose bound? If D gives a loose bound, or if one feels that the bounded-second-derivative condition is too strong, one can redo the derivation with an alternate form of Taylor's theorem, given by

  $$f(X_j + h) = f(X_j) + f'(X_j)h + \frac{f''(X_j)}{2} h^2 + o(h^2),$$

  where $o(h^2)$ is the remainder such that $\lim_{h \to 0} = o(h^2)/h^2 = 0$. Then one can replace $f''(\zeta_i)$ in the derivation by $f''(X_j)$, which can also be estimated by what we suggested

in the previous point. In this case, one may directly evaluate the imputation error and stop at the first inequality in Equation (8).
- In the later section, we will see loess smoothing is applied to subseries instead of the whole time series. In that case, the dispersing missing data assumption should hold for each subseries. This is, however, a relaxation rather than a tightening of assumption.
- Is bounded second derivative well justified? It is a reasonable assumption to make unless one expects the trend changes direction with infinite acceleration. But again, if the condition does not hold, or D is too large to be useful, one can still refer to the second point above to get the result.
- It turns out from Section 3.2 that expression (9) also provides some guidelines on how to pick the parameter of the imputation method considered. In particular, the (half) neighbourhood size l can be chosen such that the upper bound is tight.
- In principle, any missing data pattern can be analysed in the same way as shown in our derivation. The approach we used is quite standard in analysing linear smoothers. See, for example, Hastie and Tibshirani [42], Fan and Gijbels [37].
- A minor technical remark: The bound of D only needs to hold over the domain where the smoothing is performed.

4. Simulation Studies

In this section, we present simulation studies of the loess-STL procedure. Our goals are to verify the theoretical results from Section 3 and to assess the applicability of loess-STL to real data. The simulation studies are conducted under 40 settings; the settings are detailed in Section 4.1. For each setting, 10,000 simulations are run. Each simulation consists of three steps. First, we simulate a dataset and remove some proportions of points from it. Second, we apply the loess-STL procedure to impute the missing data and extract a trend from it. Third, we evaluate the estimated trend. These steps are detailed in Section 4.2. We present the simulation result in Section 4.3 and its consistency with the theoretical result in Section 4.4.

4.1. General Settings

The simulation studies are conducted with under 40 settings. Each setting specifies how a time series is generated and how many data points are removed to create missing data. The time series is made up of three components, the trend component, the seasonal component and the remainder component. Two approaches are available for simulating these components: the data-based approach and the model-based approach. In the former approach, we extract the components from some real data with STL (using the `stats::stl` function in R [41]) and use them directly as the simulated components; in the latter approach, we simulate the components from a model we specify. We consider the data-based approach because our goal is to assess the loess-STL procedures' applicability to real data, so we want the simulated data to share similar characteristics as the real ones. On the other hand, we consider the model-based approach to show that loess-STL not only works on a particular dataset, but could also generalise to other situations. Full details of the two approaches are given in the next section.

An outline of the 40 settings and some information about the real dataset we use are given as follows. We consider:

- a model-based approach for the trend component.
- a model-based or data-based approach for each of the seasonal components (which is of 12-month frequency) and the remainder component, and
- ten missing data proportions, ranging from 5% to 50% at a step of 5%.

These give a combination of $1 \times (2 \times 2) \times 10 = 40$ settings which is summarised in Table 1.

Table 1. Simulation settings (The details of the two approaches were given in Section 4.1).

Configurations	1	2	3	4
Trend component	Model-based			
Seasonal component	Model-based	Data-based	Model-based	Data-based
Remainders component	Model-based	Model-based	Data-based	Data-based

For the data-based approach, we consider the Antarctic upper air temperature data which were obtained from the Integrated Global Radiosonde Archive (IGRA) available at https://www.ncei.noaa.gov/products/weather-balloon/integrated-global-radiosonde-archive (accessed on 23 December 2022). The IGRA is a comprehensive radiosonde dataset which consists of radiosonde and pilot balloon observations from more than 2800 globally distributed stations. Observations are available at standard and variable pressure levels; meteorological variables include pressure, temperature, geopotential height, relative humidity, dew point depression, and wind direction and speed. For this study, we select the temperature data from 22 Antarctic stations at 16 standard pressure levels for a period of 50 years. Radiosonde observations are usually performed twice a day, at noon and midnight respectively.

The IGRA radiosonde data have undergone quality assurance procedures, including, most notably, the basic checks on the elapsed time and relative humidity as well as an improved selection of a single surface level within soundings in which multiple levels are identified as surface; further information could be found at https://www.ncei.noaa.gov/data/integrated-global-radiosonde-archive/doc/igra2-readme.txt (accessed on 23 December 2022).

Out of the 704 $(= 22 \times 16 \times 2)$ noon or midnight time series that would have been fully observed under ideal conditions, only 526 of them are available with each having at least one observation but possibly many missing values. As the quantity of interest is the macro movement of the temperature data, we aggregate the data using averaging to get monthly noon or midnight data; this also helps reduce the noise (i.e., the local fluctuations) in the data. Note that during the aggregation, a monthly average noon (or midnight) temperature will be missing only when there were no radiosonde observations at all across all noons (or midnights) of that month.

In the next session, we detail the simulation procedure.

4.2. Details of Simulation Studies

For each of the 40 settings, 10,000 simulations are run. Each simulation consists of three steps. First, we simulate an artificial monthly mean temperature time series and remove some proportions of data from it. Second, we apply the loess-STL procedure to impute the missing data and extract a trend from it. Third, we evaluate the estimated trend. Details are given in the remainder of this section.

4.2.1. Simulating a Dataset to Have Missing Data

To simulate a time series, we first search through the 526 time series in the real dataset and collect those with no missing data. Then we sample one time series from this pool of 'perfect' time series randomly and apply STL to extract the trend, seasonal (of 12-month frequency) and remainder components, denoted by $\hat{T}_t$, $\hat{S}_t$ and $\hat{R}_t$ respectively. These components are used directly as the simulated components in the data-based approach, or they are used to estimate the model parameters in the model-based approach, which we detail in the following.

Trend. We generate a piece-wise linear function and then apply loess smoothing to give a smooth trend. The number of pieces follows a discrete uniform distribution $Dunif(1, 10)$. Each piece occupies the same amount of time (except the last piece may contain extra points when exact division is not possible). Each slope is sampled from the normal distribution $Normal(\mu, \sigma^2)$, where the parameters μ and σ^2 are estimated by the method of moment,

i.e., we compute $D_t = \hat{T}_t - \hat{T}_{t+1},\ t = 1, \cdots, n-1$, then we set $\hat{\mu} = \bar{D} = \sum_{t=1}^{n-1} D_t/(n-1)$ and $\hat{\sigma}^2 = \sum_{t=1}^{n-1}(D_t - \bar{D})^2/(n-1)$, where n is the size of the corresponding monthly mean temperature time series used in the current simulation. As for the intercept, the first piece starts at 0, and the subsequent pieces start where the previous pieces end.

Season. We use a sine curve with random magnification at the stationary points. The size of magnification follows a uniform distribution, $U(0.8,1.2)$. To ensure the component is smooth, the points in the neighbourhood of the stationary points are scaled by the same factor; the radius of the neighbourhood is set to be a quarter of the wavelength of the sine curve.

Remainder. We use the normal distribution $Normal(\mu_r, \sigma_r^2)$, where μ_r and σ_r^2 are estimated by matching the moments of the remainder component of the sampled data set, $\hat{R}_t$.

After a time series is simulated, as above, we randomly remove from it a proportion of points assuming an equal chance of removal for each point. The missing proportions we consider are 5% to 50% at a step of 5%.

4.2.2. Loess-STL

Once the time series with missing data is ready, we proceed to impute the missing data with our procedure, i.e., we apply loess smoothing (using a neighbourhood size = 0.75 × No. of available points; the choice of 0.75 follows the default setting in `stats::stl` function in R [41]) to the cycle-subseries of the time series and interpolate the missing points. By cycle-subseries, we mean the subseries formed by partitioning the series according to the cycle implied by the research context. For example, for the artificial monthly temperature data simulated here, we partition the data according to months to form 12 subseries. The first series contains the January temperature data over the years; the second series contains the February temperature data over the years and so on. We consider the imputation as successful if all the missing points are imputed and failure if there is any point that cannot be imputed, owing to, for example, having an insufficient number of data points.

Next, we apply STL to extract a trend from the imputed dataset. Six parameters need to be specified. Five of them can be chosen automatically as given in Section 2.2. The only one left is the neighbourhood size for seasonal smoothing, n_s. n_s is a tunable parameter that incorporates expert knowledge about the seasonal component into the analysis, but the flexibility opens up a gap to be filled when little is known about the data. We suggest two ways to choose n_s in such a situation. The first is to note that choosing n_s is related to the bias-variance trade-off in finding the best curve that describes the seasonal effect. Hence, one can choose a value that minimises the smoothness-penalised least square error as n_s. This approach makes sense theoretically but requires a large effort to implement. The second way is more ad hoc. Noting that STL uses loess to smooth the data and loess in R uses a span of 0.75 by default, one can choose $n_s = 0.75 \cdot \{\text{length of the time series}\}/\{\text{number of subseries}\}$. Admittedly, the value of 0.75 is somewhat arbitrary; the general idea is to avoid extreme cases like 0 and 1 when one does not have much information. For our simulation, we suggested using $n_s = 0.5 \cdot \{\text{length of the time series}\}/\{\text{number of subseries}\}$, reflecting no preference is given to any of 0 and 1 over the other one.

4.2.3. Evaluation Measures for the Estimated Trend

For each simulation, we compute two measures to evaluate the estimated trend. The first quantity is the mean squared difference between the complete trend and the imputed trend. We refer to it as the trend error MSE_{trend}. For the second quantity, we first perform an ordinary least square(OLS) fit on both the complete trend and the imputed trend versus the time, and then we take the modulus of the difference of the two slope coefficients. We refer to it as the slope error. Note that the slopes express the average (temperature) change per unit time (month) change over the entire timeframe. We consider this quantity because it is frequently used in the context of climate data (see for example, Turner, Lachlan-Cope, Colwell, Marshall and Connolley [43]; Zhang [44]; Steig, Schneider, Rutherford, Mann, Comiso and Shindell [45]).

4.3. Results of Simulation Studies

For each of the 40 simulation settings (4 configurations $\times$ 10 missing proportions), 10,000 simulations are run; we summarise the results using boxplots in Figures 2 and 3. The configurations referred to in the figures are given in Table 1.

In Figure 2 we present boxplots of the trend errors (defined in Section 4.2.3) against the different proportions of missing data under the four configurations. In addition, we label the averages with grey squares. In the following, we summarise our findings.

1. The average, the medium, the interquartile range and the maximum/minimum (excluding the outliers) of the trend errors in each configuration show a near-linear increasing pattern as the proportion of missing points increases. Similar patterns are observed over the four configurations.
2. Averages of the trend errors are significantly below 0.2 squared degree Celsius over all the settings, with the maximum being at the 50% missingness setting. At this 50% missingness level, the trend errors only go as high as 0.181 for the average, 0.718 for the maximum(including outliers), and 0.110 for the interquartile range. The results are satisfactory given that the amount of missing data is substantial.
3. At the 50% missingness level, a few outliers show abnormally large errors in configurations 2 and 4. Two facts contribute to this phenomenon. First, the missingness proportion 50% is a critical point beyond which a dispersing missing-data pattern cannot form. At the critical level, large gaps can sometimes form; this prevents loess-STL from extracting trends accurately and gives rise to large errors. Second, both configurations 2 and 4 use the data-based approach to generate the seasonal component. These seasonal components are relatively more irregular, so it is harder to get an accurate estimate. The large missingness proportion further worsens the situation, leading to large errors in the extracted trend.
4. Relating to the true trend, in all 40 settings, the 10,000 mean squared differences between the (smoothed) complete trend and the true trend have an average of less than 0.048 and a maximum of less than 0.34. The corresponding figures for the 10,000 mean squared differences between the (smoothed) imputed trend and the true trend are 0.075 and 0.485 respectively. The first two statistics suggest STL can produce reliable trend estimates while the last two statistics suggest when loess imputation is used, STL can be robust against missing data. Figure 4 gives a visual impression of the comparison between the complete trend, the imputed trend, and the true trend for a randomly simulated time series of length 498 months using Configuration 2 and with a random removal of 40% data.
5. As a supplementary note, the 95%-quantiles of the trend errors are below 0.324 in all 40 settings. Moreover, in each setting the 95%-quantile of the 10,000 trend errors is less than 0.6 times the corresponding maximum error. This suggests that the typical case is generally much better than the worst case.

In Figure 3, we present boxplots of slope errors (defined in Section 4.2.3) against the different proportions of missing data under the four configurations. Again, we label the averages using grey squares. Our findings can be summarised as follows:

1. The average, the medium, the interquartile range and the maximum/minimum (excluding the outliers) of the slope errors show a near-linear increasing pattern as the proportion of missing points increases. Similar patterns are observed over the four configurations.
2. Averages of the slope errors are below 0.001 over all the settings. At the 50% missingness level, the slope errors only go as high as 0.00071 for the average, 0.00385 for the maximum (including outliers), and 0.00074 for the interquartile range.
3. As a supplementary note, the 95%-quantiles of the slope errors are below 0.00172 in all 40 settings. In each setting, the 95%-quantile of the 10,000 slope errors are less than 0.54 times the corresponding maximum error, again suggesting that the estimate one typically gets is generally much better than that in the worst-case scenario.

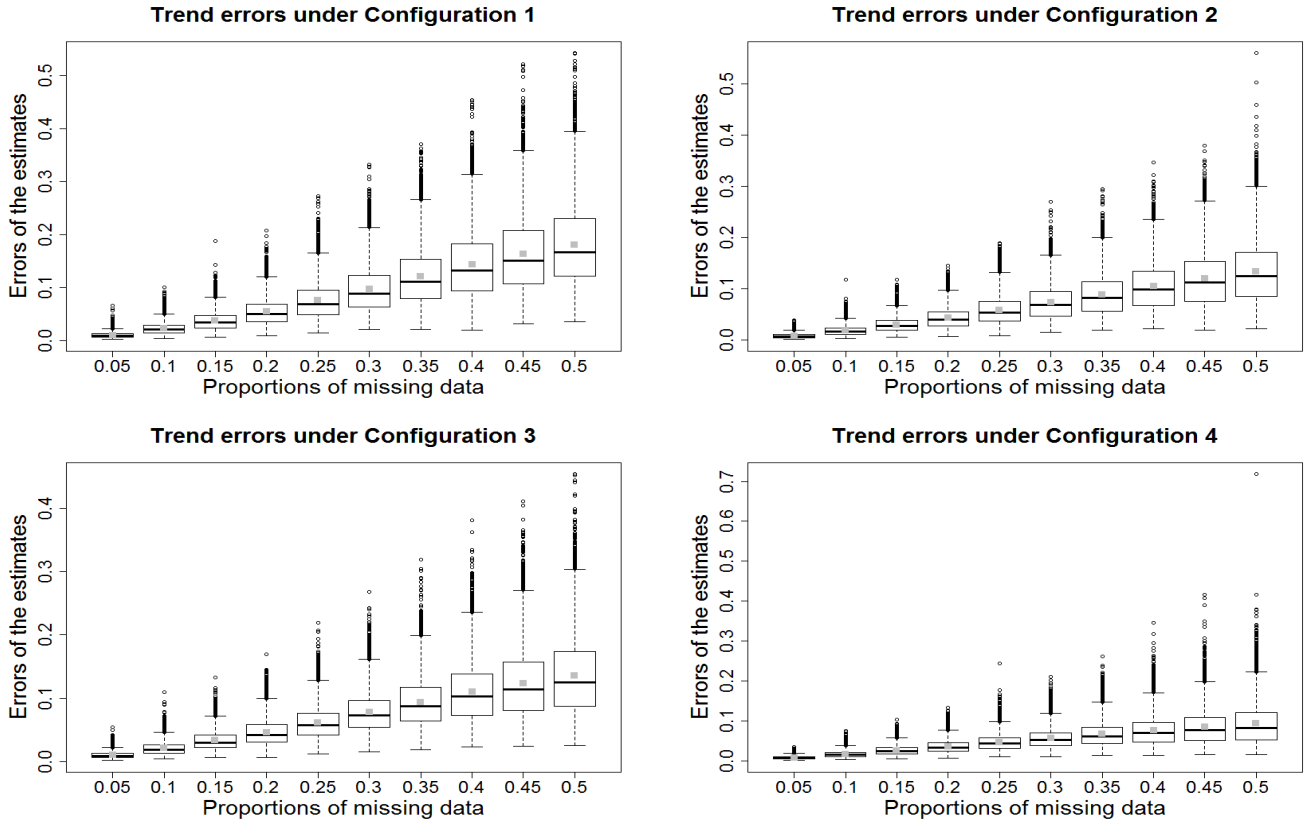

Figure 2. Boxplots of the trend errors (in squared degree Celsius) under different settings.

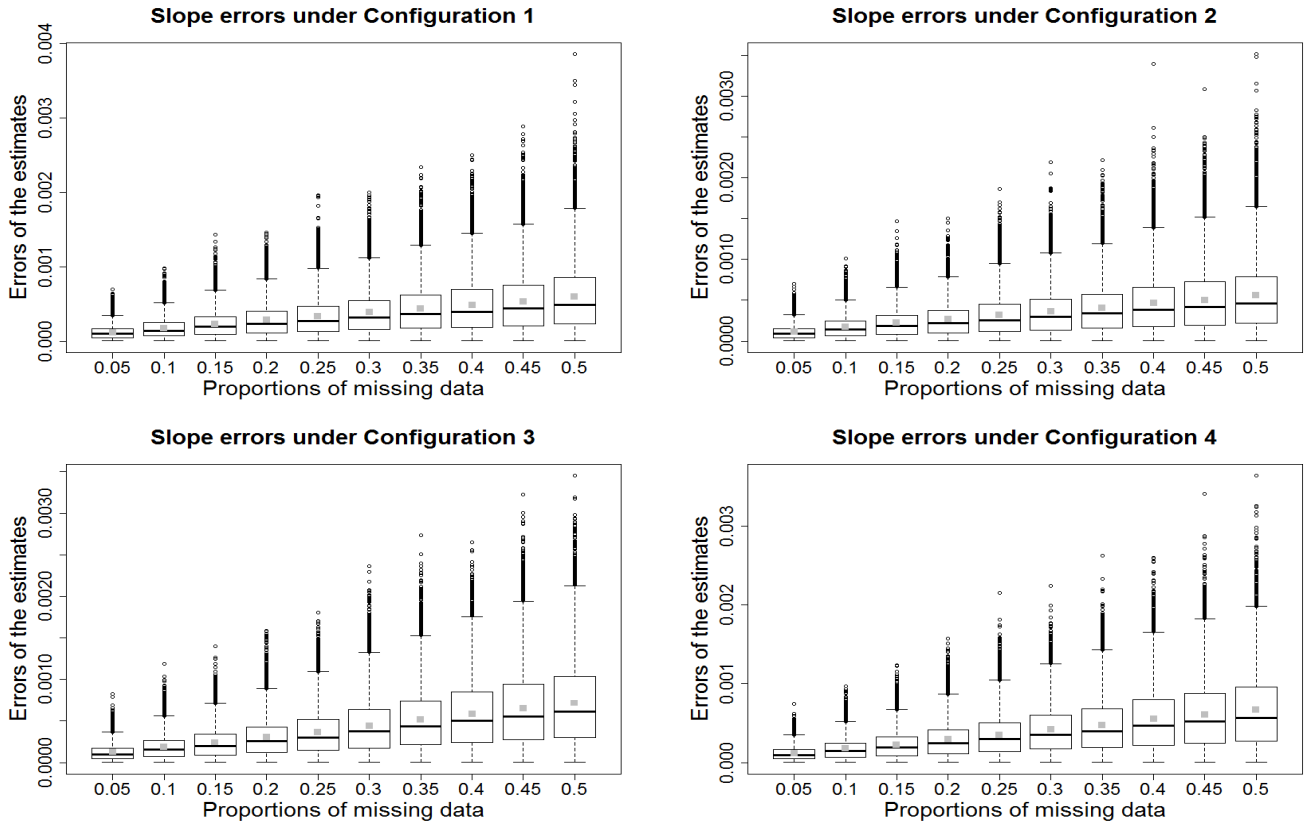

Figure 3. Boxplots of the slope errors (in degree Celsius per month change) under different settings.

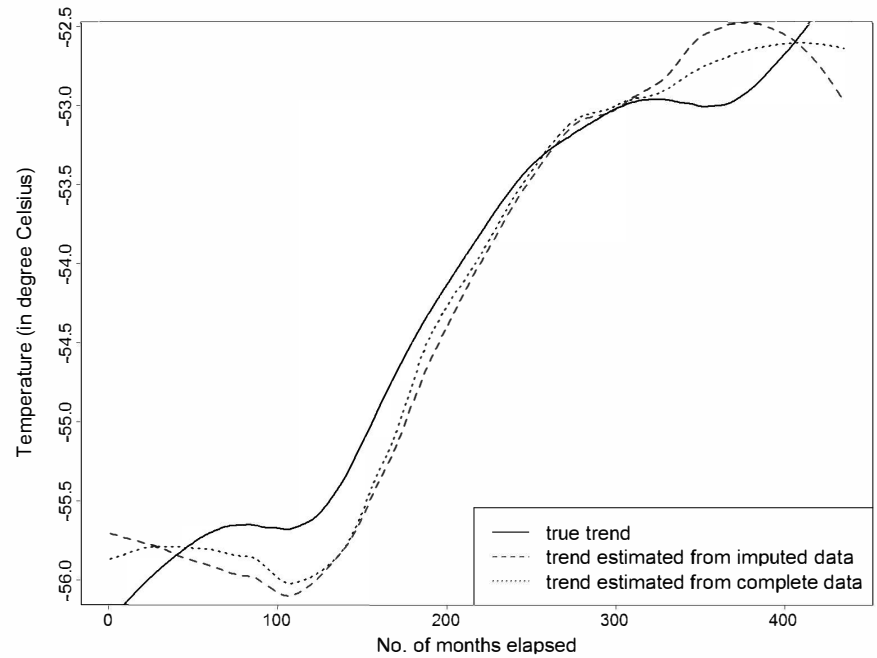

Figure 4. A comparison of the true trend and the estimated trends for a randomly simulated time series of length 498 months using Configuration 2 and with a random removal of 40% data.

4.4. Consistency with Theoretical Results

Our simulation studies have shown that applying loess-STL to incomplete data (up to 50% missing) can produce trend estimates that are close to the ones from applying STL to complete data. Now we relate these results to the theoretical results from Section 3. We first compute theoretical upper bounds for the trend errors, then we check if the bounds actually hold. Since we know the original data in the simulation studies (as we generated them), we can directly compute the imputation errors and apply Equation (4) to find the theoretical bounds. The maximum of the trend errors (over 10,000 simulations in each setting) and the theoretical bounds are given in Tables 2 and 3 respectively.

Table 2. Maximum of trend errors (in squared degree Celsius) over 10,000 simulations.

Config.	Proportions of Missing Data									
	0.05	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**
1	0.0664	0.1006	0.1878	0.2072	0.2721	0.3323	0.3708	0.4539	0.5212	0.5424
2	0.0388	0.1179	0.1175	0.1460	0.1891	0.2703	0.2937	0.3458	0.3791	0.5588
3	0.0538	0.1093	0.1330	0.1692	0.2195	0.2677	0.3195	0.3811	0.4119	0.4543
4	0.0354	0.0761	0.1034	0.1335	0.2446	0.2105	0.2622	0.3455	0.4170	0.7176

Table 3. Theoretical upper bounds.

Config.	Proportions of Missing Data									
	0.05	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**
1	0.4486	0.9182	1.4050	1.8780	2.3821	2.8884	3.4318	3.9511	4.5352	5.1445
2	0.4011	0.8296	1.2630	1.6855	2.1358	2.5936	3.0501	3.5747	4.0772	4.6522
3	0.5074	1.0415	1.5777	2.1024	2.6729	3.2470	3.8160	4.4323	5.0864	5.7861
4	0.4762	0.9760	1.4852	1.9986	2.5099	3.0434	3.6271	4.2040	4.7851	5.4598

Comparing Tables 2 and 3, we see that the theoretical bounds are effective, showing consistency of the numerical results with our theoretical results. However, we remark that

these bounds inevitably become loose as the proportion of missing data gets large. This is because the imputation error only gives information about the mean squared individual imputation errors but not the signs of the individual errors. Thus, in the worst-case scenario where all the individual imputation errors have the same sign, the trend extracted with STL can indeed have a large bias. The bias is more pronounced when the proportions of missing data are large, therefore, the bounds for those proportions are loose.

5. Application

Deriving an accurate trend in meteorological data (e.g., temperature) is important for the detection and attribution of climate change. To derive plausible trends, long-term time series, preferably over several decades, are used. However, these time series often have missing data (e.g., due to failure of instruments) which impact the accuracy of the estimated trend. In remote areas like the Arctic and the Antarctic, the proportion of missing data could be particularly large, but at the same time, accurate trend estimates over these areas are of great importance. For example, the response of the Arctic to global warming is one of the major indicators of climate change, cf. IPCC [46].

In this section, we apply the loess-STL procedure to the Antarctic temperature data introduced in Section 4.1. In particular, we apply the loess-STL to the midnight temperature time series collected at the Novolazaravskaja station at 8 different pressure levels from October 1969 to March 2011. We made this choice because the missing data in these time series show a high degree of dispersed-ness, hence it fulfils the assumptions we proposed. In Figure 5, we show as an illustration the monthly average midnight temperature time series at 150 hPa pressure level over Novolazaravskaja station before and after imputations. The circles are the original data points, and the rhombuses are the imputed data points. The time series is 498 months long and has 42 data points missing, which is equivalent to a missing proportion of 8.4%. Upon the imputation, STL is applied to extract the trend from the time series. We do this to each of the 8-time series collected at the 8 pressure levels and generate a profile plot displayed in Figure 6. In the figure, the bars and the dots represent the average temperature change per decade (in degree Celsius) at the corresponding pressure level. As a reminder, the average change is the slope coefficient (multiplying 120 months) of the OLS line fit on the extracted trend. The ± 2 standard errors are provided using error bars to represent approximately 95% confidence intervals, and a smoothed line is plotted to show the dynamics of the average temperature change over the different pressure levels. Figure 6 confirms the climatologists' understanding that radiosonde observations over Antarctica become warming in the lower troposphere between 850 and 400 hPa, and strong cooling in the upper troposphere between 250 and 150 hPa over the past 5 decades.

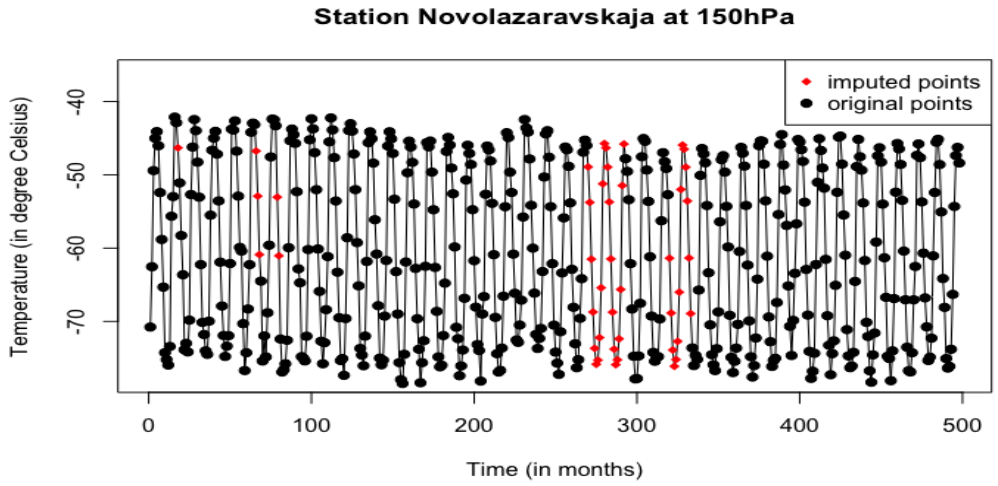

Figure 5. An illustration of the Novolazaravskaja time series at 150 hPa before and after imputation.

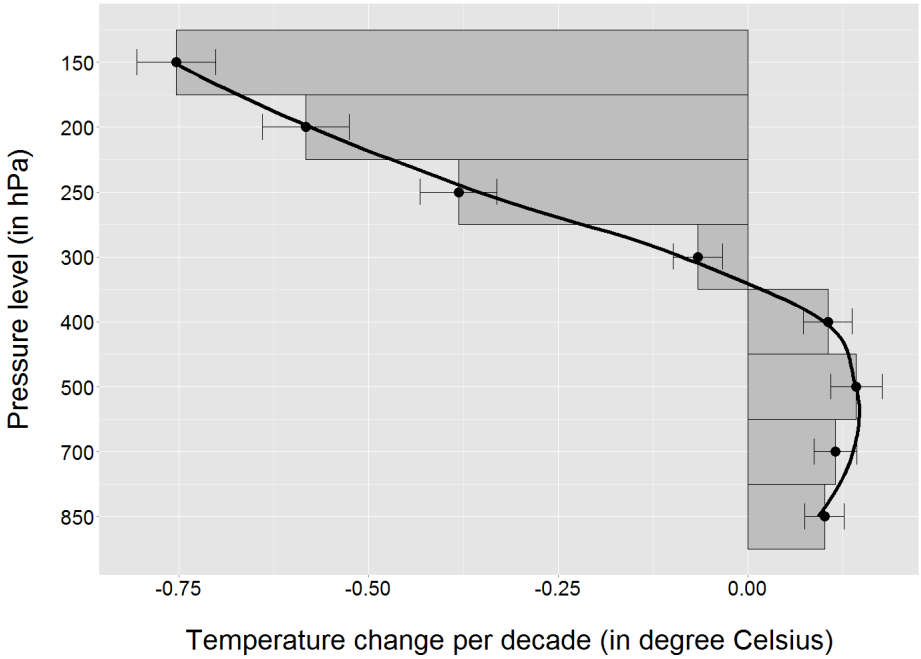

Figure 6. Midnight temperature change at the Novolazaravskaja station from October 1969 to March 2011.

6. Conclusions

In this paper, we studied the problem of trend extraction when there are missing data in time series. Specifically, we investigated and derived analytic results for a general class of procedures, the imputation-STL procedures. The results provide insight into the effect of imputation errors on the trend estimates and justify the use of the procedures in practice. We also examined a particular case, the loess-STL procedure, and evaluated its performance through simulated time series data and the Antarctic upper air temperature time series. A set of conditions under which the procedure can give reliable trend estimates is identified: the underlying trend needs to be smooth and the missing data needs to be dispersed over the whole time series. The simulation studies and the theoretical results showed consistency with each other and overall, they both support strongly the use of loess-STL procedures when the conditions are satisfied. Finally, we apply loess-STL to the upper air temperature data from station Novolazaravskaja between October 1969 and March 2011, with the results on temperature change per decade being displayed in Figure 6.

Author Contributions: Conceptualization, G.Q., C.-F.K. and Y.K.; methodology, G.Q. and C.-F.K.; software, C.-F.K.; validation, G.Q. and C.-F.K.; formal analysis, G.Q. and C.-F.K.; investigation, G.Q. and C.-F.K.; data curation, Y.K.; writing—original draft preparation, C.-F.K.; writing—review and editing, G.Q., C.-F.K. and Y.K.; visualization, C.-F.K.; supervision, G.Q. and Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available from the Integrated Global Radiosonde Archive (IGRA) at https://www.ncei.noaa.gov/products/weather-balloon/integrated-global-radiosonde-archive (accessed on 23 December 2022).

Acknowledgments: We thank the two anonymous reviewers for valuable comments which helped us to improve the quality of the manuscript and its presentation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Henderson, R. Note on graduation by adjusted average. *Trans. Actuar. Soc. Am.* **1916**, *17*, 43–48.
2. Cleveland, R.B.; Cleveland, W.S.; McRae, J.E.; Terpenning, I. STL: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **1990**, *6*, 3–73.
3. Hodrick, R.J.; Prescott, E.C. Postwar US business cycles: An empirical investigation. *J. Money Credit Bank.* **1997**, *29*, 1–16. [CrossRef]
4. Findley, D.F.; Monsell, B.C.; Bell, W.R.; Otto, M.C.; Chen, B.C. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *J. Bus. Econ. Stat.* **1998**, *16*, 127–152.
5. Bovik, A.C.; Huang, T.S.; Munson, D.C., Jr. A generalization of median filtering using linear combinations of order statistics. *IEEE Trans. Acoust. Speech Signal Process.* **1983**, *31*, 1342–1350. [CrossRef]
6. Gabbouj, M.; Coyle, E.J.; Gallagher, N.C., Jr. An overview of median and stack filtering. *Circuits Syst. Signal Process.* **1992**, *11*, 7–45. [CrossRef]
7. Wen, Y.; Zeng, B. A simple nonlinear filter for economic time series analysis. *Econ. Lett.* **1999**, *64*, 151–160. [CrossRef]
8. Hassani, H. Singular spectrum analysis: Methodology and comparison. *J. Data Sci.* **2007**, *5*, 239–257. [CrossRef]
9. Schoellhamer, D.H. Singular spectrum analysis for time series with missing data. *Geophys. Res. Lett.* **2001**, *28*, 3187–3190. [CrossRef]
10. Moskvina, V.; Zhigljavsky, A. An Algorithm Based on Singular Spectrum Analysis for Change-Point Detection. *Commun. Stat. Simul. Comput.* **2003**, *32*, 319–352. [CrossRef]
11. Kondrashov, D.; Ghil, M. Spatio-temporal filling of missing points in geophysical data sets. *Nonlinear Process. Geophys.* **2006**, *13*, 151–159. [CrossRef]
12. Hassani, H.; Zhigljavsky, A.; Patterson, K.; Soofi, A.S. A comprehensive causality test based on the singular spectrum analysis. In *Causality in the Sciences*; Oxford University Press: Oxford, UK, 2011; pp. 379–404. [CrossRef]
13. Mohammad, Y.; Nishida, T. On comparing SSA-based change point discovery algorithms. In Proceedings of the 2011 IEEE/SICE International Symposium on System Integration (SII), Kyoto, Japan, 20–22 December 2011. [CrossRef]
14. Shen, Y.; Peng, F.; Li, B. Improved singular spectrum analysis for time series with missing data. *Nonlinear Process. Geophys.* **2015**, *22*, 371–376. [CrossRef]
15. Ghil, M.; Vautard, R. Interdecadal oscillations and the warming trend in global temperature time series. *Nature* **1991**, *350*, 324–327. [CrossRef]
16. Ghil, M. Advanced spectral methods for climatic time series. *Rev. Geophys.* **2002**, *40*, 3-1–3-41. [CrossRef]
17. Hassani, H.; Thomakos, D. A review on singular spectrum analysis for economic and financial time series. *Stat. Its Interface* **2010**, *3*, 377–397. [CrossRef]
18. Ghodsi, M.; Yarmohammadi, M. Exchange rate forecasting with optimum singular spectrum analysis. *J. Syst. Sci. Complex.* **2014**, *27*, 47–55. [CrossRef]
19. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
20. Flandrin, P.; Rilling, G.; Goncalves, P. Empirical Mode Decomposition as a Filter Bank. *IEEE Signal Process. Lett.* **2004**, *11*, 112–114. [CrossRef]
21. Echeverría, J.C.; Crowe, J.A.; Woolfson, M.S.; Hayes-Gill, B.R. Application of empirical mode decomposition to heart rate variability analysis. *Med. Biol. Eng. Comput.* **2001**, *39*, 471–479. [CrossRef] [PubMed]
22. Battista, B.M.; Knapp, C.; McGee, T.; Goebel, V. Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data. *Geophysics* **2007**, *72*, H29–H37. [CrossRef]
23. Zhang, X.; Lai, K.; Wang, S.Y. A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Econ.* **2008**, *30*, 905–918. [CrossRef]
24. Alexandrov, T.; Bianconcini, S.; Dagum, E.B.; Maass, P.; McElroy, T.S. A Review of Some Modern Approaches to the Problem of Trend Extraction. *Econom. Rev.* **2012**, *31*, 593–624. [CrossRef]
25. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [CrossRef]
26. Rubin, D.B. *Multiple Imputation for Nonresponse in Surveys*; Wiley: New York, NY, USA, 1987.
27. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*; Wiley: New York, NY, USA, 1987.
28. Rubin, D.B. Multiple Imputation after 18+ Years. *J. Am. Stat. Assoc.* **1996**, *91*, 473–489. [CrossRef]
29. Horton, N.J.; Kleinman, K.P. Much Ado About Nothing. *Am. Stat.* **2007**, *61*, 79–90. [CrossRef]
30. Honaker, J.; King, G. What to Do about Missing Values in Time-Series Cross-Section Data. *Am. J. Political Sci.* **2010**, *54*, 561–581. [CrossRef]
31. Horton, N.J.; Lipsitz, S.R. Multiple Imputation in Practice. *Am. Stat.* **2001**, *55*, 244–254. [CrossRef]
32. Graham, J.W.; Olchowski, A.E.; Gilreath, T.D. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prev. Sci.* **2007**, *8*, 206–213. [CrossRef]
33. Van Buuren, S.; Oudshoorn, K. *Flexible Multivariate Imputation by MICE*; TNO Prevention Center: Leiden, The Netherlands, 1999.
34. Azur, M.J.; Stuart, E.A.; Frangakis, C.; Leaf, P.J. Multiple imputation by chained equations: What is it and how does it work? *Int. J. Methods Psychiatr. Res.* **2011**, *20*, 40–49. [CrossRef]
35. Grover, G.; Gupta, V.K. Multiple imputation of censored survival data in the presence of missing covariates using restricted mean survival time. *J. Appl. Stat.* **2014**, *42*, 817–827. [CrossRef]

36. Cleveland, W.S.; Devlin, S.J.; Grosse, E. Regression by local fitting. *J. Econom.* **1988**, *37*, 87–114. [CrossRef]
37. Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; Chapman & Hall/CRC: New York, NY, USA, 1996.
38. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829. [CrossRef]
39. Huber, P.J. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Stat.* **1973**, *1*, 799–821. [CrossRef]
40. Lütkepohl, H. *Handbook of Matrices*; Wiley: New York, NY, USA, 1996.
41. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
42. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; CRC Press: New York, NY, USA, 1990; Volume 43.
43. Turner, J.; Lachlan-Cope, T.; Colwell, S.; Marshall, G.; Connolley, W. Significant warming of the Antarctic winter troposphere. *Science* **2006**, *311*, 1914–1917. [CrossRef]
44. Zhang, J. Increasing Antarctic sea ice under warming atmospheric and oceanic conditions. *J. Clim.* **2007**, *20*, 2515–2529. [CrossRef]
45. Steig, E.J.; Schneider, D.P.; Rutherford, S.D.; Mann, M.E.; Comiso, J.C.; Shindell, D.T. Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year. *Nature* **2009**, *457*, 459–462. [CrossRef]
46. IPCC. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Core Writing Team, Pachauri, R.K., Meyer, L.A., Eds.; IPCC: Geneva, Switzerland, 2014; p. 151.