ROBUST DOMINANT PERIODICITY DETECTION FOR TIME SERIES WITH MISSING DATA

Qingsong Wen[1], *Linxiao Yang*[2], *Liang Sun*[1]

[1]DAMO Academy, Alibaba Group, Bellevue, USA
[2]DAMO Academy, Alibaba Group, Hangzhou, China

ABSTRACT

Periodicity detection is an important task in time series analysis, but still a challenging problem due to the diverse characteristics of time series data like abrupt trend change, outlier, noise, and especially block missing data. In this paper, we propose a robust and effective periodicity detection algorithm for time series with block missing data. We first design a robust trend filter to remove the interference of complicated trend patterns under missing data. Then, we propose a robust autocorrelation function (ACF) that can handle missing values and outliers effectively. We rigorously prove that the proposed robust ACF can still work well when the length of the missing block is less than $1/3$ of the period length. Last, by combining the time-frequency information, our algorithm can generate the period length accurately. The experimental results demonstrate that our algorithm outperforms existing periodicity detection algorithms on real-world time series datasets.

Keywords— periodicity detection, seasonality detection, missing data, robust methods, time series

1. INTRODUCTION

Many time series signals are characterized by repeating cycles, or periodicity in modern applications like the Internet of Things (IoT), Artificial Intelligence for IT Operations (AIOps), and cloud computing [1, 2, 3]. Periodicity detection and adjustment are crucial in many real-world time series applications, like anomaly detection [4, 5, 6], forecasting [7, 8, 9], classification and clustering [10, 11], and decomposition [12, 13, 14]. However, due to different and complicated characteristics of real-world time series data in IoT, AIOps, and cloud computing [2], accurate periodicity detection remains a challenging problem. Here we briefly highlight three challenges for practical periodicity detection. Firstly, most time series are non-stationary, and the trend changes even abrupt trend changes commonly occur. Secondly, the time series data generally contains noises and outliers. Thirdly, many real-world time series data contain missing or even block missing values [15].

The existing periodicity detection algorithms can be categorized into two groups: 1) frequency domain methods relying on periodogram after Fourier transform, such as Fisher's test [16, 17]; 2) time domain methods relying on autocorrelation function (ACF), such as methods in [18, 19]. To combine the advantages of both groups, recent joint time-frequency methods are proposed in [20, 21, 7]. However, the aforementioned methods cannot directly handle time series with missing data. In order to deal with missing values, the Lomb-Scargle periodogram based methods are proposed to detect periodicity as in [22, 23, 24]. Unfortunately, these methods cannot robustly address outliers and noises in time series.

In this paper, we propose a novel periodicity detection method to detect the dominant periodicity of time series under missing data robustly and accurately. Firstly, to mitigate the side effects introduced by trend, especially abrupt trend changes, we design a robust filter to remove the trend component under outliers and missing data. Next, to obtain accurate period length, we design a robust ACF module which can effectively deal with impulse random noise with unknown heavy-tailed noise, as well as block missing data. We rigorously prove that our algorithm can still work successfully when the length of the missing block is less than $1/3$ of the whole length of the time series. Lastly, we combine the time-frequency information for the final accurate period length estimation. Compared with various state-of-the-art algorithms, our proposed robust algorithm performs significantly better on real-world datasets, especially for time series with missing data.

2. PROPOSED PERIODICITY DETECTION

2.1. Framework Overview

We consider the following complex time series model with trend and dominant periodicity/seasonality as

$$y_t = \tau_t + s_t + r_t, \quad t = 0, 1, \cdots, N-1 \tag{1}$$

where y_t represents the observed value at time t, τ_t is the trend component, s_t is the dominant periodic/seasonal component with period length T, and $r_t = a_t + n_t$ denotes the remainder part which contains the noise n_t and possible outlier a_t. Note that y_t may contain missing data. For dominant periodicity detection, we aim to identify if the time series is periodic and its major period length.

Our proposed periodicity detection algorithm contains three steps: 1) robust detrending filter; 2) robust ACF; 3) time-frequency combination. The diagram of the proposed algorithm is depicted in Fig. 1, which will be elaborated in the following subsections.

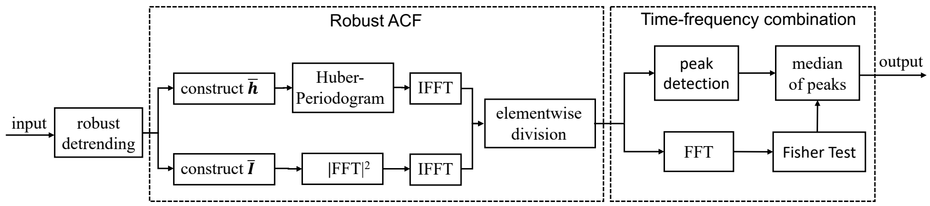


Figure 1: Framework of the proposed robust dominant periodicity detection algorithm for time series with missing data.

2.2. Robust Detrending under Missing Data

Most real-world time series usually have missing data with varying trend and outliers, which makes the ACF hard to capture the period information. To remove the trend component under missing data and motivated by the RobustTrend filter [25], we design a novel robust trend filter under missing data with objective function as

$$||\mathbf{W}(\mathbf{y}-\boldsymbol{\tau})||_1 + \lambda_1||\mathbf{D}^{(1)}\boldsymbol{\tau}||_1 + \lambda_2||\mathbf{D}^{(2)}\boldsymbol{\tau}||_1, \qquad (2)$$

where the $D^{(1)} \in \mathbb{R}^{(N-1)\times N}$ and $D^{(2)} \in \mathbb{R}^{(N-2)\times N}$ are the first- and second-order difference matrix adopted to capture abrupt and slow trend changes, respectively, with the following form:

$$\mathbf{D}^{(1)} = \begin{bmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & \ddots & \\ & & 1 & -1 \end{bmatrix}, \mathbf{D}^{(2)} = \begin{bmatrix} 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix},$$

and $\mathbf{W} = \text{diag}\,(w_t)$ is a diagonal matrix with w_t as 0 if y_t is missing and 1 otherwise. Note that our designed trend filter has two differences compared with the RobustTrend filter [25]. One is the $\mathbf{W}$ matrix incorporated into the loss function, which makes trend extraction possible without missing value imputation. The other is that we update the Huber loss of the original RobustTrend filter to least absolute deviations (LAD) loss for its similar robustness to outliers and easy-to-solve property in alternative direction method of multipliers (ADMM). Specifically, to obtain the trend component from Eq. (2), we can rewrite it in an equivalent form:

$$\begin{array}{ll} \min & ||\mathbf{e}||_1 \\ \text{s.t.} & \mathbf{A}\boldsymbol{\tau} - \mathbf{b} = \mathbf{e} \end{array} \qquad (3)$$

where $\mathbf{A} = [\mathbf{W}^T, \mathbf{D}^{(1)T}, \mathbf{D}^{(2)T}]^T$, $\mathbf{b} = [(\mathbf{W}\mathbf{y})^T, \mathbf{0}^T]^T$. Then we can obtain the augmented Lagrangian as

$$L_\rho(\mathbf{x},\mathbf{y},\mathbf{u}) = ||\mathbf{y}||_1 + \mathbf{u}^T(\mathbf{A}\mathbf{x}-\mathbf{b}-\mathbf{y}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x}-\mathbf{b}-\mathbf{y}||_2^2$$

where $\mathbf{u}$ is the dual variable, and ρ is the penalty parameter. Next, the solution can be obtained through the ADMM [26] as

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}} L_\rho(\mathbf{x},\mathbf{y},\mathbf{u}) = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{b}+\mathbf{y}_t-\mathbf{u}_t/\rho)$$
$$\mathbf{y}_{t+1} = \arg\min_{\mathbf{y}} L_\rho(\mathbf{x},\mathbf{y},\mathbf{u}) = S_{1/\rho}(\mathbf{A}\mathbf{x}_{t+1}-\mathbf{b}+\mathbf{u}_t/\rho)$$
$$\mathbf{u}_{t+1} = \arg\min_{\mathbf{u}} L_\rho(\mathbf{x},\mathbf{y},\mathbf{u}) = \mathbf{u}_t + \rho(\mathbf{A}\mathbf{x}_{t+1}-\mathbf{y}_{t+1}-\mathbf{b})$$

where $S_\kappa(x)$ is the soft thresholding as $S_\kappa(x) = (1-\kappa/|x|)_+ x$.

2.3. Roubust ACF with M-Periodogram

2.3.1. Structure of the Proposed Roubust ACF

Let $\{x_t\}$ denote the detrended time series, i.e. $x_t = y_t - \tau_t$. Then its period can be estimated by finding the position of the largest value of the ACF $\{r_k\}$ which is defined as $r_k = \mathbb{E}(x_{t+k}x_t)$. The ACF is usually estimated using $r_k \approx \frac{1}{|\mathbb{Q}_k|}\sum_{t\in\mathbb{Q}_k} x_{t+k}x_t$, where set $\mathbb{Q}_k$ is defined as $\{t|0 \le t \le N-1, 0 \le n+k \le N-1\}$, and $|\mathbb{Q}_k|$ denote the size of $\mathbb{Q}_k$. However, this estimator can not be applied to time series with missing values directly. To address this problem, we propose a new unbias estimator. To better illustrate our method, let $\{I_t\}$ be a binary sequence that indicates whether $\{x_t\}$ is observed. Specifically, let $I_t = 1$ when x_t is observed, and $I_t = 0$ when x_t is missing. Our proposed ACF estimator is $r_k \approx \frac{1}{|\hat{\mathbb{Q}}_k|}\sum_{t\in\hat{\mathbb{Q}}_k} x_{t+k}x_t$ where the set $\hat{\mathbb{Q}}_k = \{t|I_t = 1, I_{t+k} = 1\}$. It is ready to see that this estimator is an unbiased estimator of the autocorrelation, but directly computing it has a computational complexity of order $O(N^2)$. To reduce the complexity, we define a sequence $\{h_t\}$ by padding $\{x_t\}$ with zeros, i.e. $h_t = x_t$ if $I_t = 1$ and $h_t = 0$ if $I_t = 0$. It is easy to see that $h_t = x_t I_t$. Then our proposed ACF estimator can be rewritten as

$$\begin{aligned} r_k \approx & \frac{1}{|\hat{\mathbb{Q}}_k|}\sum_{t\in\hat{\mathbb{Q}}_k} x_{t+k}x_t = \frac{1}{|\hat{\mathbb{Q}}_k|}\sum_{t\in\hat{\mathbb{Q}}_k} I_{t+k}x_{t+k}I_t x_t \\ = & \frac{1}{|\hat{\mathbb{Q}}_k|}\sum h_{t+k}\hat{h}_{k-(t+k)} = \frac{1}{|\hat{\mathbb{Q}}_k|}(\mathbf{h}*\hat{\mathbf{h}})_k \end{aligned} \qquad (4)$$

where $\hat{h}_t = h_{-t}$, and the operation "$*$" denotes the linear convolution. Recall that $\hat{\mathbb{Q}}_k = \{t|I_t = 1, I_{t+k} = 1\} = \{t|I_t I_{t+k} = 1\}$, we have $|\hat{\mathbb{Q}}_k| = \sum I_n I_{t+k} = \sum \hat{I}_{-t} I_{t+k} = (\hat{\mathbf{I}} * \mathbf{I})_k$, where $\hat{I}_t = I_{-t}$. Furthermore, we can utilize FFT/IFFT based on circular convolution theorem to reduce the complexity of $(\mathbf{h}*\hat{\mathbf{h}})_k$ and $(\hat{\mathbf{I}}*\mathbf{I})_k$ from $O(N^2)$ to $O(N\log N)$. Therefore, our efficient and robust ACF is

$$r_k \approx \frac{(\mathbf{h}*\hat{\mathbf{h}})_k}{(\hat{\mathbf{I}}*\mathbf{I})_k} = \frac{\text{IFFT}(|\text{FFT}(\bar{\mathbf{h}})|^2)_k}{\text{IFFT}(|\text{FFT}(\bar{\mathbf{I}})|^2)_k}, \qquad (5)$$

where the length of $\mathbf{h}$ and $\mathbf{I}$ are doubled by padding N zeros and denoted as $\bar{\mathbf{h}} = [\mathbf{h}^T, 0, \cdots, 0]^T$ and $\bar{\mathbf{I}} = [\mathbf{I}^T, 0, \cdots, 0]^T$.

Besides dealing with missing data, we also consider to mitigate the effect of outliers in $\bar{\mathbf{h}}$ of Eq. (5). Note that $|\text{FFT}(\bar{\mathbf{h}})|^2$ is the conventional periodogram of $\bar{\mathbf{h}}$. To make it robust to outliers, we introduce the Huber-periodogram [7] as

$$(|\text{FFT}(\bar{\mathbf{h}})|^2)_k \approx \frac{N'}{4}\left|\hat{\boldsymbol{\beta}}_M(k)\right|^2 = \frac{N'}{4}\left|\arg\min_{\boldsymbol{\beta}\in\boldsymbol{R}^2}\gamma(\boldsymbol{\phi}\boldsymbol{\beta}-\bar{\mathbf{h}})\right|^2, \qquad (6)$$

where $N' = 2N$, $\hat{\boldsymbol{\beta}}_M(k)$ is the robust estimation of harmonic regressor $\boldsymbol{\phi}_t = [\cos(2\pi kt/N'), \sin(2\pi kt/N')]$, $\boldsymbol{\phi}$ has the matrix forms of $[\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_{N'-1}]^T$, and $\gamma(\mathbf{x})$ is the robustifying objective function defined as the summation of Huber loss as $\gamma(\mathbf{x}) = \sum_{t=0}^{N'-1}\gamma^{hub}(x_t)$.

Based on the aforementioned processing for missing data and outliers, the final structure of our robust ACF is depicted in the middle part of Fig. 1.

2.3.2. Theoretical Constraint under Missing Data

Next, we investigate the maximum missing block length that our robust ACF estimator can bear. Clearly, utilizing Eq. (4) to compute the ACF requires $Q_k = |\hat{\mathbb{Q}}_k| > 0$. While the value of Q_k depends on the positions and the volume of the missing entries, which makes it difficult for analysis. In practice, the worst scenarios for missing entries occur in cluster. In the following, we provide theoretical analysis for the scenario that all the missing entries concentrate in one block, i.e., $I_k = 0$ when $m \leq n \leq m + l - 1$ and $I_k = 1$ otherwise, where m denotes the start position of the block and l denotes the length of the block. Here we assume that the block does not include the beginning and end point of the sequence, i.e., $m > 0$ and $m + l - 1 < N - 1$. Note that the proposed method can work if and only if $Q_k > 0$. Recall that Q_k is $\sum_{t=k}^{N-1} I_t I_{t-k}$ when $k \geq 0$ and $\sum_{t=0}^{N+k} I_t I_{t-k}$ otherwise. Formally, we have the following proposition.

Proposition: For any k and m that satisfy $m > 0$ and $m+l-1 < N-1$, if $l < \frac{N}{3}$, we always have $Q_k > 0$.

Proof: From the definition of Q_k, we have $Q_k \geq 0$, then we only need to prove that it does not exist a k such that $Q_k = 0$. Without loss of generality, we only consider the case that $k > 0$. Let $g_t = I_{t-k}$, then $g_t = 0$ for $m + k \leq t \leq m + l + k - 1$. Suppose there is a k and m such that $Q_k = 0$, then $I_n \hat{I}_n = 0$ for all $n \in [k, N-1]$, then we have

$$[k, N-1] \subset [m, m+l-1] \cup [m+k, m+l+k-1] \quad (7)$$

Note that Eq. (7) implies the following three possibilities:

$$[k, N-1] \subset [m, m+l-1] \quad (8)$$

$$[k, N-1] \subset [m+k, m+l+k-1] \quad (9)$$

$$\begin{aligned} &[k, N-1] \subset [m, m+l-1] \cup [m+k, m+l+k-1] \\ &[m, m+l-1] \cup [m+k, m+l+k-1] = [m, m+l+k-1] \end{aligned} \quad (10)$$

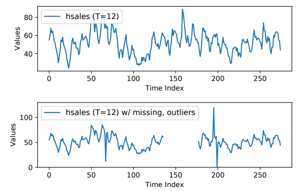
Obviously, Eqs. (8) and (9) requires $N-1 \leq m+l-1$ or $k+1 \geq m+k$ which is contradicted to $m < N-l$ and $m > 0$. Note that Eq. (10) implies $m \leq k$, $m+k \leq m+l$, and $N-1 \leq m+l+k-1$, which can be rewritten as

$$\max(m, N-m-l) \leq k \leq l \quad (11)$$

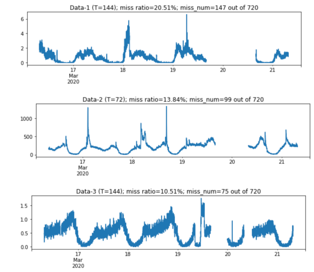
We show that if $l < \frac{N}{3}$, one cannot find a m and k satisfying Eq. (11). It is easy to see that for all $0 < m < N-l$, the minimum of $\max(m, N-m-l)$ is archived at $m = \frac{N-l}{2}$ and equals to $\frac{N-l}{2}$. As $l < \frac{N}{3}$, we can see that l is always less than $\frac{N-l}{2}$, i.e., $l < \frac{N-l}{2} \leq \max(m, N-m-l)$ which makes there is no k satisfying Eq. (11). The proof is completed here.

2.4. Final Time-Frequency Combination

In frequency domain, to detect dominant periodicity, Fisher's test [16] defines g-statistic as $g = \max_k P_k / \sum_{j=1}^{N} P_j, k = 1, 2, \cdots, N$, where $P_k = \text{FFT}\{r_t\}$ is the calculated periodogram via Wiener-Khinchin theorem [27] based on robust ACF r_t. Therefore, this g-statistic is also robust to outliers and missing data. The distribution of g-statistic gives a p-value to determine if a time series is periodic [16]. If this value is less than the predefined threshold α, we reject the null hypothesis H_0 and conclude the time series is periodic with period length candidate as N/k where $k = \arg\max_k P_k$. If the Fisher's test is passed, we further refine the candidate of period length in the time domain similar to [20, 7].



(a) One CRAN periodic time series and its variant with outliers and missing values.



(b) Three typical periodic time series from cloud computing industry with outliers and missing values.

Figure 2: Representative challenging periodic time series.

Specifically, we first summarize the peaks of robust ACF through peak detection [28]. Then, we calculate the median distance of those peaks whose heights exceed predefined threshold. Furthermore, based on the resolution of periodogram, i.e., the peak value of P_k at index k corresponds to period length in the range $[\frac{N}{k}, \frac{N}{k-1})$, the median distance of ACF peaks is the final period length only if it locates in $R_k = \left[\frac{1}{2}\left(\frac{N}{k+1} + \frac{N}{k}\right) - 1, \cdots, \frac{1}{2}\left(\frac{N}{k} + \frac{N}{k-1}\right) + 1\right]$. Note that this combination is necessary. On the one hand, the periodogram has limited resolution and spectral leakage may exist [20], which makes the candidate from Fisher's test not accurate. On the other hand, only relying on ACF may result in false positive results since ACF cannot provide if there is dominant periodicity.

3. EXPERIMENTS AND DISCUSSIONS

3.1. Periodicity Detection Comparisons

Baseline Algorithms and Datasets:

We compare our algorithm with baseline algorithms: ACF-Med (median distance of ACF peaks), Fisher's Test, Lomb–Scargle periodogram [22, 23, 24], and the state-of-the-art time-frequency method RobustPeriod [7]. Note that there exist other types of time-frequency algorithms [20, 21], but their performances are inferior to RobustPeriod [7]. Except Lomb–Scargle and our proposed algorithm, we add linear interpolation in case of missing data for other algorithms since their performances are degraded severely if directly working on time series with missing values.

For datasets, we consider the public single-period time series from CRAN [29, 19], which contains 82 real-world time series in various scenarios (like IoT, sales, sunspots, etc.) with lengths from 16 to 3024 and period lengths from 2 to 52. We also add outliers and block missing data in the datasets to evaluate the robustness of all algorithms. The outlier ratio (OR) is set to 0.01 or 0.05 with outlier amplitude as 5 times standard deviation of original time series, and the block missing ratio (MR) is set to 0.05 or 0.30. Since the outlier ratio is usually small in practice, we set it as 0.05 at most; while the missing ratio could be relatively large due to block missing, we set it as 0.3 at most. One example is shown in Fig. 2(a). Besides, We utilize 3 typical real-world datasets from a leading cloud computing company as shown in Fig. 2(b). The length of the dataset is 5 days with sampling resolution of 10 minutes, where their true period lengths are 144 (one-day), 72 (half-day), and 144 (one-day), respectively. For the evaluation, the precision is calculated by the ratio of the number of time series with correctly estimated period length to the total number of time series.

Table 1: Precision comparisons of different periodicity detection algorithms on public CRAN data. MR and OR indicate missing ratio and outlier ratio, respectively. The best results are highlighted in bold.

Algorithms	CRAN dataset with or without missing values & outliers								
	MR=0			MR=0.05			MR=0.30		
	OR=0	OR=0.01	OR=0.05	OR=0	OR=0.01	OR=0.05	OR=0	OR=0.01	OR=0.05
ACF-Med	0.57	0.55	0.24	0.56	0.50	0.23	0.52	0.50	0.27
Fisher's Test	0.52	0.44	0.32	0.49	0.44	0.30	0.42	0.33	0.24
Lomb–Scargle	0.45	0.13	0.09	0.45	0.13	0.10	0.37	0.11	0.10
RobustPeriod	0.62	0.60	0.59	0.59	0.57	0.50	0.47	0.45	0.43
Proposed	**0.63**	**0.62**	**0.60**	**0.60**	**0.60**	**0.59**	**0.56**	**0.52**	**0.51**

Table 2: Comparisons of different periodicity detection algorithms on 3 typical real-world datasets with missing data.

Algorithms	Data-1 (T=144)	Data-2 (T=72)	Data-3 (T=144)
ACF-Med	144	72	143
Fisher's Test	144	73	143
Lomb–Scargle	140	71	140
RobustPeriod	142	(72,144)	144
Proposed	**144**	**72**	**144**

Performance Comparisons of Different Algorithms: We summarize the detection precision results on public CRAN datasets with or without missing data and outliers in Table 1. It can be seen that, when there are no missing data and outliers, both RobustPeriod and our proposed algorithm exhibit similar results and have better performance than others. When there are no outliers but with increasing missing data (i.e., OR=0, MR=0, 0.05, 0.3), the performance degradation of Lomb–Scargle is relatively smaller than ACF-Med, Fisher's Test, and RobustPeriod. This is due that Lomb–Scargle can directly deal with missing data without imputation. However, the Lomb–Scargle is not robust to outliers, and its performance becomes much worse than others when there are outliers. When we fix the ratio of missing data but with increasing outliers (i.e., OR=0, 0.05, 0.3), the performance degradation of RobustPeriod is much smaller than ACF-Med, Fisher's Test, and Lomb–Scargle, which demonstrates its robustness to outliers. When there are both missing data and outliers, the performances of existing algorithms drop significantly, especially for ACF-Med, Fisher's Test, and Lomb–Scargle methods. The reason is that these methods heavily depend on conventional ACF or FFT while the results of ACF or FFT would be severely distorted by block missing data and outliers. Overall, our proposed algorithm performs best in all test cases, especially exhibits better performance than others when the missing data and outliers ratio are severe.

Table 2 summarizes the detection results of the 3 typical real-world datasets as in Fig. 2(b). It can be seen that the RobustPeriod may generate some false positives, which is due to the fact that it is designed for general multiple-period detection. It is also interesting to find that Lomb–Scargle is not robust to large outliers and noises even though it can directly deal with missing data. Overall, our proposed algorithm achieves the best results.

3.2. Component Investigation

Firstly, we carry out experiments to demonstrate the advantages of the designed robust trend filter under missing data. Fig. 3 depicts the proposed robust trend filter without imputation versus common Hodrick–Prescott (HP) trend filter [30] with linear interpolation for time series under missing data, trend changes, and outliers. As HP filter is unable to deal with time series with missing data, we inpaint the time series using linear interpolation and input it into

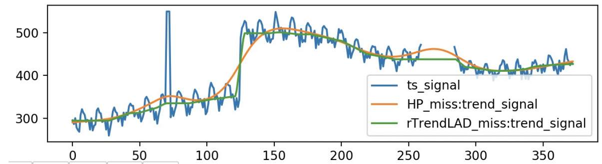
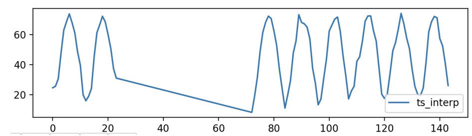
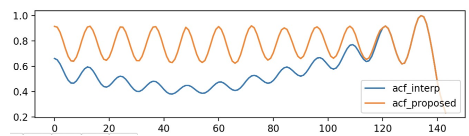


Figure 3: Proposed robust trend filter without imputation vs. HP trend filter with linear interpolation under missing data, trend changes, and outliers (better in colorful version).



(a) Linear interpolated time series



(b) Proposed robust ACF under missing data vs. ACF by linear interpolation (better in colorful version).

Figure 4: Proposed robust ACF without imputation vs. ACF with linear interpolation under missing data.

HP filter. After imputation, the HP filter is still biased by the outlier and abrupt changes of trend. In contrast, our method (denoted as "rTrendLAD_miss") can deal with missing data without imputation, and meanwhile is robust to outliers and trend changes.

Next, we demonstrate the superiority of the proposed robust ACF estimator under missing data. We conduct experiments on a time series of length 144 with period 12. We set the 24th to 71th points as missing, so the missing ratio is $1/3$. Fig. 4(a) shows the incomplete time series with linear interpolation. Fig. 4(b) shows the results of the proposed ACF under incomplete time series and the standard ACF under linear interpolation. We can see that the ACF from linear interpolated time series is severely affected, which makes it hard to find a threshold to locate the peaks of the ACF. In contrast, our proposed robust ACF estimator can provide a promising ACF estimation which brings accurate periodicity detection.

4. CONCLUSION AND FUTURE WORK

In this paper we propose a novel periodicity detection algorithm to detect the dominant periodicity. Our algorithm can robustly deal with trend changes and outliers for time series under missing data scenarios. In the future we plan to extend our framework to handle time series with multiple periodicities, as well as apply it in more real-world applications.

5. REFERENCES

[1] Heonho Kim, Unil Yun, Bay Vo, Jerry Chun-Wei Lin, and Witold Pedrycz, "Periodicity-oriented data analytics on time-series data for intelligence system," *IEEE Systems Journal*, vol. 15, no. 4, pp. 4958–4969, 2020.

[2] Qingsong Wen, Linxiao Yang, Tian Zhou, and Liang Sun, "Robust time series analysis and applications: An industrial perspective," in *KDD 2022*, 2022, pp. 4836–4837.

[3] Shuo Zhang, XiaoFei Chen, JiaYuan Chen, Qiao Jiang, and Hejiao Huang, "Anomaly detection of periodic multivariate time series under high acquisition frequency scene in IoT," in *Int. Conf. on Data Mining Workshops*, 2020, pp. 543–552.

[4] Jingkun Gao, Xiaomin Song, Qingsong Wen, Pichao Wang, Liang Sun, and Huan Xu, "RobustTAD: Robust time series anomaly detection via decomposition and convolutional neural networks," *KDD Workshop MileTS*, 2020.

[5] Ramona Tolas, Raluca Portase, Andrei Iosif, and Rodica Potolea, "Periodicity detection algorithm and applications on IoT data," in *20th International Symposium on Parallel and Distributed Computing*, 2021, pp. 81–88.

[6] Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun, "TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis," in *CIKM*, 2022.

[7] Qingsong Wen, Kai He, Liang Sun, Yingying Zhang, Min Ke, and Huan Xu, "RobustPeriod: Time-frequency mining for robust multiple periodicity detection," in *ACM Int. Conf. on Management of Data (SIGMOD)*, 2021, pp. 2328–2337.

[8] Qingyang Xu, Qingsong Wen, and Liang Sun, "Two-stage framework for seasonal time series forecasting," in *ICASSP*, 2021, pp. 3530–3534.

[9] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International Conference on Machine Learning*, 2022, pp. 27268–27286.

[10] X. Wang, K. Smith-Miles, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining and Knowledge Discovery*, vol. 13, pp. 335–364, 09 2006.

[11] Michail Vlachos, Christopher Meek, Zografoula Vagena, and Dimitrios Gunopulos, "Identifying similarities, periodicities and bursts for online search queries," in *ACM SIGMOD Int. Conf. on Management of data*, 2004, pp. 131–142.

[12] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–73, 1990.

[13] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, Huan Xu, and Shenghuo Zhu, "RobustSTL: A robust seasonal-trend decomposition algorithm for long time series," in *AAAI Conference on Artificial Intelligence*, 2019, pp. 5409–5416.

[14] Linxiao Yang, Qingsong Wen, Bo Yang, and Liang Sun, "A robust and efficient multi-scale seasonal-trend decomposition," in *ICASSP*, 2021.

[15] Xiuwen Yi, Yu Zheng, Junbo Zhang, and Tianrui Li, "ST-MVL: filling missing values in geo-sensory time series data," in *Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2016.

[16] Ronald Aylmer Fisher, "Tests of significance in harmonic analysis," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 125, no. 796, pp. 54–59, 1929.

[17] S. Wichert, K. Fokianos, and K. Strimmer, "Identifying periodically expressed transcripts in microarray time series data," *Bioinformatics*, vol. 20, no. 1, pp. 5–20, 2004.

[18] Wang, J., Chen, T., and Huang, B., "Cyclo-period estimation for discrete-time cyclo-stationary signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 83–94, 2006.

[19] Maximilian Toller, Tiago Santos, and Roman Kern, "SAZED: parameter-free domain-agnostic season length estimation in time series data," *Data Mining and Knowledge Discovery*, 2019.

[20] Michail Vlachos, Philip Yu, and Vittorio Castelli, "On periodicity detection and structural periodic similarity," in *SIAM Int. Conf. on Data Mining*, 2005, pp. 449–460.

[21] Abdullah Almasri, "A new approach for testing periodicity," *Communications in Statistics—Theory and Methods*, vol. 40, no. 7, pp. 1196–1217, 2011.

[22] Feiyan Hu, Alan F Smeaton, and Eamonn Newman, "Periodicity detection in lifelog data with missing and irregularly sampled data," in *IEEE BIBM*, 2014, pp. 16–23.

[23] Earl F Glynn, Jie Chen, and Arcady R Mushegian, "Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms," *Bioinformatics*, vol. 22, no. 3, pp. 310–316, 2006.

[24] Nicholas R Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and space science*, vol. 39, no. 2, pp. 447–462, 1976.

[25] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, and Jian Tan, "RobustTrend: a Huber loss with a combined first and second order difference regularization for time series trend filtering," in *IJCAI*, 2019, pp. 3856–3862.

[26] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[27] Norbert Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258, 1930.

[28] Felix Scholkmann, Jens Boss, and Martin Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012.

[29] Kurt Hornik, "The comprehensive R archive network," *Wiley interdisciplinary reviews: Computational statistics*, vol. 4, no. 4, pp. 394–398, 2012.

[30] Robert J Hodrick and Edward C Prescott, "Postwar US business cycles: an empirical investigation," *Journal of Money, Credit, and Banking*, pp. 1–16, 1997.