Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/159875>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

MuSDRI: Multi-Seasonal Decomposition Based Recurrent Imputation for Time Series

Yujue Zhou, Jie Jiang, Shuang-Hua Yang, Ligang He and Yulong Ding

Abstract—Missing values are a common problem of time series such as sensor-generated data due to issues of data collection, transmission, and storage. To deal with this problem, various data imputation methods have been proposed. Among these methods, variants of recurrent neural networks (RNNs) have attracted a lot of attention as they provide a natural representation of temporal dynamics. Though such methods have been shown to achieve higher imputation accuracy, they mostly focus on the modeling of temporal dynamics from a short-term perspective while ignoring the long-term features such as trend and seasonal information embedded in time series data. In this paper, we investigate different paradigms of time series imputation methods to capture temporal dynamics from both long-term and short-term perspectives. In particular, we propose a new learning paradigm called MuSDRI which combines the time series decomposition method Seasonal Trend decomposition using Loess (STL) with RNNs. Specifically, STL is used to decompose long sequences of time series into trend, seasonal, and remainder components such that long-term dynamics can be revealed; RNNs are used to capture local patterns from the remainder component extracted by STL such that short-term dynamics can be refined. Moreover, MuSDRI is able to handle time series with multiple seasonal patterns by learning their contributions with respect to the imputation task. We evaluate the proposed imputation methods on three real-world datasets. Experimental results show that the imputation methods considering both the long-term and short-term dynamics achieve higher imputation accuracy, and among them MuSDRI in general achieves the best performance.

Index Terms—Data imputation, multi-seasonal decomposition, recurrent neural networks, time series

I. INTRODUCTION

Multivariate time series like sensor-generated data have been widely collected and used for various applications, e.g., weather forecasting [1], [2], health-care [3], water supply system management [4], and traffic engineering [5]. Due to issues of data collection, transmission and storage, missing

values are commonly found in time series data, which not only drags down the data quality, but also seriously affects data processing and analysis, making it difficult for practical applications. [6], [7].

To deal with the problem of missing values in time series data, various imputation methods have been proposed in the literature (e.g. [8]–[10]). Simple statistical methods include imputing the missing values with the last observed value, with the mean value, and with the most common value. The issue with these methods is that they ignore the temporal dynamics in the time series data. There are also works applying machine learning techniques for time series imputation such as KNN [8], ARIMA [9], MICE [10], etc. Recently, with the popularity and success of deep learning, variants of recurrent neural networks (RNNs) [11] have attracted a lot of attention as they provide a natural representation of temporal dynamics. These RNN based imputation methods have been shown to achieve higher imputation accuracy than the traditional machine learning based imputation methods. However, they mostly focus on the modeling of temporal dynamics from a short-term perspective without taking into account the long-term features such as the trend and seasonal information embedded in the time series data. To extract such long-term features, it is often required to model long sequences (a few months of data or even years of data) as a whole. This poses challenges for RNN

This research is supported by the National Natural Science Foundation of China under Grant Nos. 61873119 and 92067109, National Key RD Program of China under Grant Nos. 2019YFC0810705 and the Science and Technology Innovation Commission of Shenzhen under Grant Nos. KQJSCX20180322151418232. (Corresponding author: Jie Jiang and Shuang-Hua Yang.)

Yujue Zhou is with the Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK (e-mail: y.zhou.34@warwick.ac.uk).

Jie Jiang is with Department of Computer Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, P.R. China (e-mail: jiangj@sustech.edu.cn).

Shuang-Hua Yang is with Department of Computer Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, P.R. China, and also with PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen 518055, P.R. China (e-mail: yangsh@sustech.edu.cn).

Ligang He is with the Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK (e-mail: ligang.he@warwick.ac.uk).

Yulong Ding is with Department of Computer Science and Engineering, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, P.R. China, and also with Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, P.R. China (e-mail: dingyl@sustech.edu.cn).

based imputation methods which usually take small sequences of data as inputs for model training (e.g. 36 data points in the case of [12]) due to their high computational cost.

To this end, we investigate different paradigms of time series imputation methods to capture temporal dynamics from both a long-term and a short-term perspective. Specifically, various time series decomposition methods such as Seasonal Trend decomposition using Loess (STL) [13] and Fourier transformation [14] are applied to capture long-term dynamics, while RNNs are applied to capture short-term dynamics. Moreover, we propose a new learning paradigm called MuSDRI, a Multi-Seasonal Decomposition Based Recurrent Imputation method for time series, in which STL is used to decompose long sequences of time series into trend, seasonal and remainder components such that long-term dynamics can be revealed; RNNs are used to capture local patterns from the remainder component extracted by STL such that short-term dynamics can be refined. That is, STL is responsible for modeling time series data as a whole to reveal their long-term features such that RNN can use such information as a prior to refine the modeling of local patterns. In particular, MuSDRI is able to handle time series exhibiting multiple seasonal patterns by providing a learning mechanism to adjust their contributions with respect to the imputation task. We evaluate all the imputation methods on three real-world datasets, including an air quality dataset [15], a water pressure dataset [16], and a gas consumption dataset. Experimental results show that the imputation methods considering both the long-term and short-term dynamics achieve higher imputation accuracy in terms of three evaluation metrics including mean absolute error, mean square error, and mean relative error, and among them MuSDRI achieves the overall best performance.

In summary, the contributions of this paper are as follows:

- We investigate different paradigms of time series imputation methods based on various time series decomposition methods (e.g. STL, Fourier transformation) and RNNs. Experimental results show that the joint modeling of the long-term and short-term features improves the imputation accuracy.
- We propose a new learning paradigm called MuSDRI for time series imputation, which takes into account the situations where time series data may have multiple seasonal patterns and provides a mechanism to learn their contributions with respect to the imputation task. Experimental results show that MuSDRI outperforms the state-of-the-art methods for the imputation task.
- We release to the public another time series dataset, i.e. a gas consumption dataset collected from five canteen users over three years as well as the source code¹ of MuSDRI to facilitate the reproducibility of our work.

The rest of the paper is organized as follows: Section II discusses the related work. Section III gives a formal description of the time series imputation problem. Section IV presents the proposed time series imputation method MuSDRI. Section V describes the datasets used in the experiment. Section VI introduces the baseline methods and a set of

adapted imputation methods based on variants of time series decomposition methods. Sections VII and VIII respectively illustrate the experiment preparation and the evaluation metrics, and analyse the experiment results under the settings of point missing and consecutive missing. Section IX discusses the experiment results, the time complexity as well as the limitations of MuSDRI. Finally, in Section X, we conclude the paper with the possibilities of future work.

II. RELATED WORK

There are many methods proposed for time series imputation in the literature. In general, the existing imputation methods can be divided into three major categories, i.e., statistical imputation methods, machine learning based imputation methods, and deep learning based imputation methods. The simplest method to process missing values is deletion [17]. Other statistical imputation methods include imputing missing values by the last observed value [18], the mean value [19] and the mode value [20]. Machine learning based imputation methods have been discussed in many previous works. For example, K-Nearest Neighbor (KNN) based imputation method [8] uses the mean value of the k nearest non-missing neighbors to impute the missing values. Also based on KNN, [21] proposed a method called IIM (Imputation via Individual Models) to deal with the challenge of data sparsity and heterogeneity for the imputation task. The autoregressive imputation methods use parameterization to fit the stationary part of the time series data, including ARIMA [9], ARFIMA [22], and SARIMA [23]. In addition, [24] and [25] combine ARIMA and Kalman Filter, which achieve better performance. MICE (Multivariate Imputation by Chained Equations) [10] imputes missing values based on the iterative chain equations. Moreover, [26] and [27] utilise Matrix Factorization to fill the missing values in time series. Focusing on time series data with large gaps of missing values, Liu et al. [28] proposed an iterative framework using STL decomposition and multiple segmented gap iteration to provide estimations for the missing values.

Recently, researchers have been attempting to impute the missing values of time series with deep learning based methods. For example, Che et al. [29] proposed GRU-D, which imputes missing values of a clinical dataset with a smooth fashion. It considers that the combination of the last observed value and the global mean can represent the patterns of a missing value. M-RNN [30] adopts bi-directional RNN to impute missing values. BRITS [12] is also a bi-directional RNN model, which considers the correlation between different variables in time series data. End-to-End Generative Adversarial Network (E2GAN) [31], [32] adapts GAN (Generative Adversarial Network) to impute the incomplete time series by the nearest complete time series generated by the pre-trained GAN. Also based on GAN, [33] proposed a framework for traffic data imputation, which uses a new time-related coding method to convert traffic time series data generation problems into image generation problems. NAOMI [34] is a deep generative model which uses multi-resolution modeling and adversarial training to recursively find and predict missing values from coarse resolution to fine-grained resolution.

¹<https://github.com/yujue-zhou/MuSDRI>

Combining ideas from variational autoencoders and Gaussian processes, Fortuin *et al.* [35] proposed a deep probabilistic model for multivariate time series imputation, in which the variational autoencoders map the missing data from the input space into a latent space while the temporal dynamics are modeled by the Gaussian processes.

Although a number of previous works have proven the effectiveness of RNN-based methods for time series imputation, these methods mostly focus on the modeling of temporal dynamics from a short-term perspective without taking into account the long-term features of the time series data such as the trend and seasonal information. In this paper, we propose MuSDRI, a multi-seasonal decomposition based recurrent imputation method for time series data, which explores both the long-term and short-term patterns for time series imputation.

Another line of related work focuses on time series forecasting, which also takes into account both short-term and long-term patterns. For example, [36] proposes a hierarchical prediction method that combines the Exponential Smoothing (ES) model with the Long Short Term Memory (LSTM) neural network to extract local as well as global information from time series data. [37] proposes a model called Long Short-Term Memory Multi-Seasonal Net (LSTM-MSNet), which is able to predict the changes of time series with multiple seasonal patterns. [38] proposes a network traffic prediction model which combines STL and LSTM. Though these works take into account both short-term and long-term patterns of time series, their focus is on time series forecasting rather than imputation. Moreover, our work differs from these works as we not only consider the situations where multiple seasonal patterns exist but also provide a learning mechanism to adjust their contributions with respect to the imputation task.

III. PROBLEM STATEMENT

We denote a d -dimensional multivariate time series X observed at $T = (t_1, t_2, \dots, t_n)$, as $X = (x_{t_1}, x_{t_2}, \dots, x_{t_n}) \in \mathbb{R}^{d \times n}$. We introduce a masking vector $m \in \mathbb{R}^{d \times n}$ to indicate whether the values of X exist or not, defined as follows.

$$m_{t_i}^j = \begin{cases} 0 & \text{if } x_{t_i}^j \text{ is not observed} \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

For example, if $x_{t_i}^j$, $1 \leq i \leq n$, $1 \leq j \leq d$ is not a missing value, $m_{t_i}^j = 1$, otherwise $m_{t_i}^j = 0$.

To represent the time lag between the last observed value to the current value, we introduce a matrix $\delta \in \mathbb{R}^{d \times n}$, defined as follows

$$\delta_{t_i}^j = \begin{cases} t_i - t_{i-1} + \delta_{t_{i-1}}^j & \text{if } i > 1, m_{t_{i-1}}^j = 0 \\ t_i - t_{i-1} & \text{if } i > 1, m_{t_{i-1}}^j = 1 \\ 0 & \text{if } t = 1 \end{cases} \quad (2)$$

In the following sections, we investigate the problem of time series imputation. Specifically, given a time series with missing values denoted as X' and the corresponding masking vector m and time lag matrix δ , we aim at training a model that can impute the missing values of X' and generate a completed time series X .

IV. MUSDRI

In this section, we describe the proposed imputation method for time series called MuSDRI (Multi-Seasonal Decomposition based Recurrent Imputation). The proposal of MuSDRI is based on the idea that time series usually exhibit autocorrelation in both long-term dynamics and short-term dynamics. For example, in the air quality data [15], there is a trend that PM2.5 decreases during summer while increases during winter, which can be considered as long-term dynamics. On the other hand, from a short-term perspective, when there is a high PM2.5 at the previous timestamp it is likely that the following timestamp also features a relatively high PM2.5.

The framework of MuSDRI is shown in Figure 1, consisting of two modules. The first module is called Multi-Seasonal Decomposition which utilizes the time series decomposition method STL to capture the long-term features of time series by decomposing the time series into seasonal component (S), trend component (E) and remainder component (R). The second module is called Recurrent Imputation which adapts the RNN-based imputation method BRITS (Bidirectional Recurrent Imputation for Time Series) [12] to capture the short-term dynamics of the time series and impute the remainder component R obtained from the first module. In particular, MuSDRI takes into account the situations where there are multiple seasonal cycles (e.g., daily and weekly cycles) and automatically learns their contributions with respect to the imputation task. In the rest of this section, we give a detailed illustration of MuSDRI.

A. Multi-Seasonal Decomposition

The aim of the Multi-Seasonal Decomposition module is to extract the long-term features of time series data by revealing their trend and seasonal changes. In particular, we also take into account the situations where time series data exhibit multiple seasonal patterns, e.g. daily, weekly, etc. To achieve this, we adopt the time series decomposition method STL, which has been widely used for time series analysis.

Given a time series $X = (x_{t_1}, \dots, x_{t_n})$, the goal of STL is to decompose x_{t_i} , $1 \leq i \leq n$ into seasonal, trend, and remainder components, indicated respectively as s_{t_i} , e_{t_i} and r_{t_i} such that

$$x_{t_i} = s_{t_i} + e_{t_i} + r_{t_i} \quad (3)$$

The process of applying STL is iterating through the smoothing of the seasonal and trend components. The seasonal component is determined by the cycle of a time series, and the length of the cycle, i.e., the number of data points contained in each cycle. The length of the cycle is denoted as μ and it needs to be manually set in STL. All the smoothing in STL is done using LOESS (locally estimated scatterplot smoothing) regression. For the smoothing of the seasonal component, data points are organized into cycle-subseries, each of which is the set of data points with the same position in each cycle. For example, if the sampling rate of a time series is 1 day and μ is set to 7 days (1 week), one cycle-subseries of this time series would be a subseries of all the data points collected on Mondays, another subseries of all the data points collected

Fig. 1. The framework of MuSDRI

on Tuesdays, etc. The smoothing of the trend component is then performed after removing the seasonal component. Aside from the number of observations per cycle determined by the cycle length μ , there are two other important parameters relating to the seasonal and trend smoothing, i.e. the loess window for seasonal extraction $n_{(s)}$ and the loess window for trend extraction $n_{(t)}$. These two smoothing operations are encapsulated in an inner loop and repeated until convergence. There is also an outer loop which computes robustness weights after each iteration of the inner loop to reduce the influence of outliers.

As the original STL algorithm does not have a sophisticated mechanism to handle time series with missing values, we use an improved version of STL called STL+ [39] to accommodate the need of the Multi-Seasonal Decomposition module. STL+ not only inherits the decomposition capability of STL but is also able to provide estimations of the missing values for the trend and seasonal components via the Approximate Nearest Neighbor Search Routines algorithm and LOESS regression. Given a time series with missing values and a specified cycle length μ , STL+ returns a complete seasonal component $\vec{S}$, a complete trend component $\vec{E}$ and an incomplete remainder component $\vec{R}'$.

Moreover, to cater for the situations where a time series may contain multiple seasonal patterns (e.g., daily and weekly cycles), the Multi-Seasonal Decomposition module applies STL+ with multiple cycle lengths $\{\mu_1, \mu_2, \dots, \mu_q\}$ to the time series, where q is the number of seasonal patterns. As a result, given a time series X' with missing values, the outputs of the Multi-Seasonal Decomposition module are a set of complete seasonal components $\vec{S} = [S_1, S_2, \dots, S_q]$, a set of complete trend components $\vec{E} = [E_1, E_2, \dots, E_q]$ and a set of incomplete remainder components $\vec{R}' = [R'_1, R'_2, \dots, R'_q]$.

The outputs of the Multi-seasonal Decomposition module, i.e., STL+ decompositions of the time series with selected cycle lengths, will be integrated into the other module of MuS-

DRI, namely Recurrent Imputation, to guide the imputation of the incomplete remainder component of the time series. The details will be illustrated in the following section.

B. Recurrent Imputation

The aim of the Recurrent Imputation module is to capture the short-term dynamics of time series data and impute the missing values of the remainder $\vec{R}'$ component obtained from the Multi-Seasonal Decomposition module. To achieve this, we adapt the state-of-the-art RNN-based imputation method BRITS (Bidirectional Recurrent Imputation for Time Series) [12].

Suppose we have a time series X' with missing values. Through the Multi-Seasonal Decomposition module, we obtain a set of seasonal components $\vec{S} = [S_1, S_2, \dots, S_q]$, a set of trend components $\vec{E} = [E_1, E_2, \dots, E_q]$ and a set of remainder components with missing values $\vec{R}' = [R'_1, R'_2, \dots, R'_q]$ with respect to a set of cycle lengths $\{\mu_1, \mu_2, \dots, \mu_q\}$. Given these three sets of components, the Recurrent Imputation module is then applied to impute the missing values of the remainder components and the detailed process is shown on the right side of Figure 1.

First, we adapt RNN-based imputation method BRITS to fill the missing values of $\vec{R}'$. The main idea of BRITS is to use the history and future data sequence of a time series as inputs to predict the data sample at the current timestamp. When the history data sequence or the future data sequence has missing values, the predictions of these missing values are used as inputs. Moreover, BRITS also considers the correlation between the features of a time series for the imputation task.

Given an incomplete time series, BRITS provides three estimations of the missing values, including a historical-based estimation, a feature-based estimation, and a combination of the historical-based and feature-based estimation. The sum of the mean absolute error between the true value and each of the three estimations is used as the loss function. Notice that

when there are missing values the error cannot be calculated. In such cases, when later a true value is observed a delayed error is obtained.

Taking BRITS as a basis, MuSDRI builds a learning process to determine the contribution of each seasonal decomposition as to the imputation of the remainder component $\vec{R}' = [R'_1, R'_2, \dots, R'_q]$. Following BRITS, the loss function of MuSDRI also considers three types of estimations of r'_{k,t_i} which indicates the remainder components at timestamp t_i with respect to the k th seasonal decompositions. Assuming the imputation model is denoted as g_θ , where θ represent the learnable weights of the model. The three types of estimations of r'_{k,t_i} can be obtained:

$$\begin{bmatrix} \hat{r}_{k,t_i}^h & \hat{r}_{k,t_i}^f & \hat{r}_{k,t_i}^s \end{bmatrix} = g_\theta(r'_{k,t_i}, m_{t_i}, \delta_{t_i}) \quad (4)$$

where m_{t_i} and δ_{t_i} are the masking vector and time lag matrix, defined in Equations (1) and (2). $\hat{r}_{k,t_i}^h, \hat{r}_{k,t_i}^f, \hat{r}_{k,t_i}^s$ respectively represent the historical-based estimation, the feature-based estimation, and the weighted sum of the previous two estimations as that in BRITS.

To optimize the estimations of the missing values in $\vec{R}'$ while taking into account the multi-seasonal patterns in time series data, we adapt the loss function of BRITS, and as a result, the loss function of MuSDRI is built with two components.

The first component is an aggregation of the estimation errors, taking into account the contribution of different seasonal decompositions. To this end, we introduce a weight vector $\vec{W} = [w_1, w_2, \dots, w_q]$, $w_1 + w_2 + \dots + w_q = 1$ to combine the estimation errors from the q seasonal decompositions. As such, the first part of the loss function is defined as follows:

$$\begin{aligned} l_1 = & \mathcal{L}_\varepsilon \left(r'_{t_i}, \hat{r}_{t_i}^h; \vec{W}^T \right) \\ & + \mathcal{L}_\varepsilon \left(r'_{t_i}, \hat{r}_{t_i}^f; \vec{W}^T \right) \\ & + \mathcal{L}_\varepsilon \left(r'_{t_i}, \hat{r}_{t_i}^s; \vec{W}^T \right) \end{aligned} \quad (5)$$

where $\mathcal{L}_\varepsilon$ represents the function of weighted mean absolute error. As described in [12], in order to speed up the calculation, all three estimations of r'_{t_i} are used in loss calculation. Different from BRITS, MuSDRI focuses on the estimation of the remainder components and more importantly it also takes into account the influence of multi-seasonal decompositions via the adjustments of $\vec{W}$ to reduce the estimation error.

When a seasonal decomposition has a smaller cycle length μ , the value range of its corresponding remainder component is also smaller, resulting in estimation errors with smaller values. Therefore, without any other constraints, the imputation model tends to assign bigger weights to those with smaller μ . To alleviate such effect, we add a regularization term to the loss function to balance the weights of different seasonal decompositions such that the imputation model will not always assign higher weights to the decompositions whose remainder component has a smaller value range, which is defined as follows:

$$l_2 = \frac{1}{\mu_1} w_1 + \frac{1}{\mu_2} w_2 + \dots + \frac{1}{\mu_q} w_q \quad (6)$$

This regularization term uses a function of μ_k to adjust its corresponding weight w_k and in this paper $\frac{1}{\mu_k}$ is used. When μ_k is larger, $\frac{1}{\mu_k}$ is smaller, and the corresponding w_k can be assigned a larger value to minimise the loss l_1 . Combining l_1 with l_2 , the loss function of MuSDRI is defined as follows:

$$l_{t_i} = l_1 + \lambda l_2 \quad (7)$$

where λ is used to adjust the influence of the regularization term.

During model training, all the parameters in BRITS as well as the weights of the seasonal decompositions $\{w_1, w_2, \dots, w_q\}$ will be updated to optimize the estimations of the missing values. The final imputed time series are obtained as follows:

$$X = (\vec{S} + \vec{E} + \hat{R}^s) \cdot \vec{W}^T \quad (8)$$

where $\vec{S}$ and $\vec{E}$ are respectively the set of seasonal and trend components with respect to the q STL+ decompositions, $\hat{R}^s$ is the combined estimation of the missing values in the corresponding remainder components.

V. DATASETS

A. Air Quality Dataset

The air quality dataset used in this paper comes from the KDD CUP Challenge 2018 [15], which contains air quality measurements from 11 environmental monitoring stations in the urban area of Beijing. The measurements of each monitoring station cover 6 variables including PM2.5, PM10, NO₂, CO, O₃, and SO₂. The measurements are recorded every hour from February 1, 2017 to January 30, 2018. In total, there are 15.1% of missing values. For the experiments, we use the first ten months of data from February 1, 2017 to November 30, 2017 as the training set, and the last two months of data from December 1, 2017 to January 30, 2018 as the test set.

B. Water Pressure Dataset

The second dataset used in this paper comes from the Shenzhen Open Data Innovation Contest (SODIC) 2020 [16], which contains water pressure measurements from 27 water pressure monitoring stations of a water distribution system in Shenzhen. The measurements are recorded around every half an hour from January 1, 2019 to August 31, 2019. To keep the measurements from different monitoring stations aligned, we resample the measurements to one hour. In total, there are 4.1% of missing values. For the experiments, we use the first six months of data from January 1, 2019 to June 30, 2019 as the training set, and the last two months of data from July 1, 2019 to August 31, 2019 as the test set.

C. Gas Consumption Dataset

The third dataset² used in this paper is provided by a Chinese gas company, and is made publicly available along with this paper. This dataset contains three years of gas

²<https://github.com/yujue-zhou/MuSDRI>

consumption data with respect to five canteen users. The measurements are recorded every two hours from January 1, 2018 to January 1, 2021. In total, there are 5.1% of missing values. For the experiments, we use the first two years of data from January 1, 2018 to December 31, 2019 as the training set, and the last year of data from January 1, 2020 to January 1, 2021 as the test set.

VI. IMPUTATION METHODS

In this paper, we compare MuSDRI with a set of baseline methods and a set of adapted imputation methods based on variants of time series decomposition methods shown as follows.

First, we implement three traditional imputation methods as baselines:

- **Last** [18]: The missing values are imputed with the last observed value from the same variable.
- **Mean** [19]: The missing values are imputed with the global mean from the same variable.
- **KNN** [8]: The missing values are imputed with the weighted average of the k nearest neighbors identified by the k -Nearest Neighbor (KNN) algorithm.

To show the effectiveness of using long-term features for data imputation, we combine the traditional methods above with various seasonal trend decomposition methods, and obtain another eight imputation methods shown as follows:

- **STL⁺-Last**: The time series are first decomposed into seasonal, trend, and remainder components with STL+ which also imputes the missing values in the seasonal and trend components. Thereafter, the missing values in the remainder component are imputed using Last. The imputed seasonal, trend and remainder components are summed to generate the final imputed time series.
- **STL⁺-Mean**: Similar to STL⁺-Last except that the missing values in the remainder component are imputed using Mean.
- **STL⁺-KNN**: Similar to STL⁺-Last except that the missing values in the remainder component are imputed using KNN.
- **MSTL⁺-Last**: MSTL⁺ is a multi-seasonal decomposition method which combines Multiple STL (MSTL) [40] with STL+. In specific, we replace STL with STL+ as a basis in MSTL for extracting the seasonal components, the trend component and the remainder component such that the missing values in the seasonal and trend components can be imputed along with the decomposition process. Thereafter, the missing values in the remainder component are imputed using Last. The imputed seasonal, trend and remainder components are summed to generate the final imputed time series.
- **MSTL⁺-Mean**: Similar to MSTL⁺-Last except that the missing values in the remainder component are imputed using Mean.
- **MSTL⁺-KNN**: Similar to MSTL⁺-Last except that the missing values in the remainder component are imputed using KNN.

- **MuSDRI-Last**: Similar to MuSDRI except that the missing values in the remainder component are imputed using Last instead of BRITS.
- **MuSDRI-Mean**: Similar to MuSDRI except that the missing values in the remainder component are imputed using Mean instead of BRITS.

Another baseline implemented in this paper is the state-of-the-art RNN-based imputation method BRITS:

- **BRITS** [12]: BRITS is an adaptation of bi-directional RNNs which imputes the missing values according to the hidden states in both directions in RNN.

Similarly, we combine BRITS with various seasonal trend decomposition methods to show the effectiveness of using long-term features for data imputation and obtain another five imputation methods shown as follows:

- **STL⁺-BRITS**: Similar to STL⁺-Last except that the missing values in the remainder component are imputed using BRITS.
- **MSTL⁺-BRITS**: Similar to MSTL⁺-Last except that the missing values in the remainder component are imputed using BRITS.
- **Fourier-Exogenous-BRITS**: Fourier transformation [14] is used to obtain the Fourier terms corresponding to the multiple seasonal patterns of the time series. Thereafter, the original time series together with the Fourier terms are used as inputs to BRITS to impute the missing values of the original time series.
- **STL⁺-Exogenous-BRITS**: The time series are first decomposed into seasonal, trend and remainder components with STL⁺. Thereafter, the original time series together with the seasonal component and the trend component are used as inputs to BRITS to impute the missing values of the original time series.
- **MSTL⁺-Exogenous-BRITS**: Similar to STL⁺-Exogenous-BRITS except that in MSTL⁺-Exogenous-BRITS multiple seasonal components are extracted with MSTL⁺ and fed into BRITS.

VII. EXPERIMENTS FOR POINT MISSING

This part of the experiments is mainly to evaluate the performance of MuSDRI and the imputation methods presented in VI under the setting of point missing. Under this setting, data points are dropped out at random positions independently.

A. Experiment Setting

Considering the continuity of time series, we divide the complete time series dataset into two segments proportionally. The first part is used as the training set, and the second part is used as the test set. The division of the training set and the test set is described in detail in Section V. Following the previous studies [31], [32], for each of the three datasets as described in Section V, we randomly drop out $p\%$ of the data from the training set and test set ($p = 10, 20, 30$). Note that these randomly dropped data are from the set of data without any missing values in the original dataset such that the ground truth can be obtained. Three evaluation metrics, i.e. Mean

Absolute Error (MAE), Mean Relative Error (MRE), and Mean Square Error (MSE), are used to evaluate the performance of the imputation methods.

B. Parameter Selection

For the imputation methods with seasonal decomposition, there is mainly one hyper-parameter that needs to be tuned, i.e., cycle length μ . Due to limitations of STL+, the minimum value that can be assigned to μ is 4 in which case the range of values of the seasonal component is close to zero. In this paper, we experimented with a set of values for μ . For the data sets with a sampling rate of one hour (air quality data and water pressure data), we set μ to 4, 24, 168. Here, $\mu = 4$ represents the situation when there is no explicit seasonal pattern, $\mu = 24$ represents daily pattern, and $\mu = 168$ represents weekly pattern. For the data set with a sampling rate of two hours (gas consumption data), we set μ to 4, 12, 84. Here, $\mu = 12$ represents daily pattern, and $\mu = 84$ represents weekly pattern.

For the imputation methods that can only deal with a single seasonality such as STL+-Last, STL+-Mean, STL+-KNN, STL+-BRITS and STL+-Exogenous-BRITS, we experiment with different μ and report the best result. For the imputation methods that can deal with multiple seasonalities such as MSTL+-Last, MSTL+-Mean, MSTL+-KNN, MuSDRI-Last, MuSDRI-Mean, MSTL+-BRITS, Fourier-Exogenous-BRITS, MSTL+-Exogenous-BRITS and MuSDRI, we experiment with different combinations of μ . In particular, since MuSDRI has the capability to learn the contributions of different seasonalities for the imputation task, we use a combination of all the three cycle lengths (μ), i.e. 4 (4 hours), 24 (1 day), and 168 (1 week) for the air quality and water pressure datasets, and 4 (8 hours), 12 (1 day), and 84 (1 week) for the gas consumption dataset. Except the cycle length μ , the loess window for seasonal extraction $n_{(s)}$ and the loess window for trend extraction $n_{(t)}$ are chosen empirically: $n_{(s)}$ is set to “periodic” and the corresponding value is $10n + 1$ where n indicates the total length of the time series, in which case seasonal smoothing is effectively replaced by taking the mean; $n_{(t)}$ is set to the default which is defined as follows [13]:

$$nextodd(\lceil \frac{1.5\mu}{1 - \frac{1.5}{n_{(s)}}} \rceil) \quad (9)$$

For all the methods that adopt BRITS, following [12], we set the length of input sequences to 36, the number of RNN layers to 1, and the number of hidden units to 128. Adam is used as the optimiser to train BRITS with a learning rate of 0.001. For MuSDRI, there is another hyper-parameter λ that needs to be tuned, which is used to balance the weights of different seasonal decompositions. In this paper, we experimented with a set of values for λ ranging from 0 to 3.

C. Experiments Results

• Air Quality Data

Table I shows the performance of all the imputation methods with respect to the air quality dataset under the setting of point missing. The best performance achieved by the single-period

imputation methods (denoted as STL+-) is under the setting of $\mu = 4$. This is due to the fact that there is no strong seasonal pattern in the air quality data and the improvement mainly comes from the trend information. Among the set of traditional methods, STL+-KNN achieves the best performance.

Among the set of RNN based imputation methods integrated with seasonal trend decomposition, MuSDRI achieves the best performance under the parameter setting of three cycles $\mu_1 = 4, \mu_2 = 24, \mu_3 = 168$, and $\lambda = 0.0$. In particular, MuSDRI obtains an overall improvement of 23.88% for MAE, 23.88% for MRE, and 33.15% for MSE compared to the state-of-the-art imputation method BRITS.

In general, for the air quality dataset, the single-period imputation methods prefixed by STL+ and STL+-Exogenous outperform those corresponding methods without applying seasonal trend decomposition. Except for the methods based on MuSDRI which have the capability of learning the contribution of different seasonal decompositions, the other multi-seasonal imputation methods prefixed by MSTL+, MSTL+-Exogenous, and Fourier-Exogenous perform poorly on the air quality dataset.

• Water Pressure Data

Table II shows the performance of all the imputation methods with respect to the water pressure data under the setting of point missing. The best performance achieved by the single-period imputation methods is under the setting of $\mu = 24$. This is due to the fact that the water pressure data has a strong cycle of 24 hours. The method that achieves the best performance among the set of traditional methods is also STL+-KNN.

Among the set of RNN based imputation methods integrated with seasonal trend decomposition, MuSDRI achieves the best performance under the parameter setting of three cycles $\mu_1 = 4, \mu_2 = 24, \mu_3 = 168$, and $\lambda = 1.0$. Specifically, MuSDRI obtains an overall improvement of 52.48% for MAE, 52.48% for MRE, and 71.85% for MSE compared to the state-of-the-art imputation method BRITS.

Compared with BRITS, both STL+-BRITS and MSTL+-BRITS achieve better performance while the performance of STL+-Exogenous-BRITS, MSTL+-Exogenous-BRITS, and Fourier-Exogenous-BRITS gets worse. This indicates that it is not beneficial to feed to BRITS the original time series together with the exogenous information obtained from seasonal decomposition. Moreover, the traditional imputation methods integrated with seasonal trend decomposition outperform BRITS in most cases, which shows the importance of long-term features in the case of the water pressure dataset. The inferior of MSTL+ based methods compared with the STL+ based methods is due to the fact that the water pressure data exhibits a strong single seasonal pattern.

• Gas Consumption Data

Table III shows the performance of all the imputation methods with respect to the gas consumption dataset under the setting of point missing. It can be seen that except Fourier-Exogenous-BRITS, all the imputation methods integrated with seasonal trend decomposition outperform their correspondences (Last, Mean, KNN, BRITS), which shows the importance of long-term features for time series imputation.

TABLE I

PERFORMANCE OF ALL THE IMPUTATION METHODS WITH RESPECT TO THE AIR QUALITY DATA UNDER THE SETTING OF POINT MISSING

Missing Rate			10%			20%			30%		
Method		μ	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)
Traditional	Last	\	15.91	22.58	8.52	17.05	24.14	9.88	18.57	26.47	11.36
	Mean		65.37	92.77	75.55	65.33	92.51	75.37	64.94	92.56	73.39
	KNN		13.50	19.16	5.23	14.28	20.22	5.97	14.89	21.23	6.18
Seasonal Trend Decomposition + Traditional	STL ⁺ -Last	4	13.10	18.59	5.29	13.85	19.62	5.93	15.13	21.57	6.88
	STL ⁺ -Mean		13.34	18.94	4.99	14.26	20.19	5.79	15.53	22.13	6.54
	STL ⁺ -KNN		12.20	17.31	4.27	13.06	18.50	4.99	14.32	20.41	5.65
	MSTL ⁺ -Last	{4,24,168}	21.30	30.22	10.69	24.96	35.35	13.90	28.95	41.26	17.49
	MSTL ⁺ -Mean		54.77	77.72	55.50	55.87	79.11	57.52	57.58	82.07	60.53
	MSTL ⁺ -KNN		20.80	29.52	9.00	23.78	33.67	11.42	27.58	39.31	14.91
	MuSDRI-Last	{4,24,168}	13.13	18.63	5.36	13.88	19.65	6.00	15.15	21.59	6.90
MuSDRI-Mean	13.38		18.99	5.01	14.30	20.24	5.80	15.56	22.18	6.55	
RNN	BRITS	\	14.23	20.19	5.27	14.97	21.19	5.89	15.53	22.14	6.16
	STL ⁺ -BRITS	4	10.88	15.44	3.58	11.56	16.37	4.12	12.64	18.02	4.59
Seasonal Trend Decomposition + RNN	MSTL ⁺ -BRITS	{4,24,168}	19.02	27.00	7.77	21.56	30.53	9.59	24.95	35.55	12.28
		{24,168}	19.11	27.11	7.85	21.48	30.42	9.45	24.93	35.54	12.23
	Fourier-Exogenous-BRITS	{4,24,168}	14.67	20.82	5.56	15.25	21.60	6.12	15.85	22.59	6.28
		{24,168}	14.57	20.67	5.48	15.38	21.78	6.22	15.91	22.67	6.45
	STL ⁺ -Exogenous-BRITS	4	12.31	17.47	4.12	13.04	18.46	4.66	13.96	19.90	4.94
	MSTL ⁺ -Exogenous-BRITS	{4,24,168}	15.71	22.29	6.24	16.50	23.36	7.09	17.08	24.35	7.24
		{24,168}	15.72	22.31	6.31	16.25	23.02	6.93	17.01	24.24	7.30
MuSDRI		{4,24,168}	10.51	14.91	3.36	11.25	15.92	3.91	12.33	17.57	4.34

TABLE II

PERFORMANCE OF ALL THE IMPUTATION METHODS WITH RESPECT TO THE WATER PRESSURE DATA UNDER THE SETTING OF POINT MISSING

Missing Rate			10%			20%			30%			
Method		μ	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)	
Traditional	Last	\	40.21	51.92	38.23	41.65	54.09	42.65	43.88	56.68	47.25	
	Mean		65.57	84.68	82.58	65.22	84.70	79.44	65.21	84.23	82.84	
	KNN		28.26	36.49	22.66	29.56	38.39	25.43	31.62	40.84	31.34	
Seasonal Trend Decomposition + Traditional	STL ⁺ -Last	24	29.75	38.42	21.69	30.27	39.31	23.04	30.92	39.93	23.96	
	STL ⁺ -Mean		29.57	38.18	21.07	29.82	38.73	21.17	30.54	39.44	24.31	
	STL ⁺ -KNN		20.80	26.86	11.76	22.56	29.30	13.57	24.59	31.76	17.66	
	MSTL ⁺ -Last	{24,168}	32.09	41.44	23.56	33.34	43.30	25.74	34.58	44.66	27.58	
	MSTL ⁺ -Mean		38.26	49.41	40.86	38.76	50.34	40.12	39.81	51.42	45.33	
	MSTL ⁺ -KNN		27.57	35.60	20.84	30.59	39.72	25.55	33.82	43.68	33.03	
	MuSDRI-Last	{4,24,168}	29.91	38.63	20.66	30.38	39.45	21.56	31.03	40.08	22.68	
MuSDRI-Mean	29.03		37.49	17.11	28.58	38.42	17.68	30.63	39.56	20.95		
RNN		\	35.97	46.44	30.49	35.91	46.64	27.79	39.52	51.04	36.58	
Seasonal Trend Decomposition + RNN	STL ⁺ -BRITS	24	16.74	21.61	7.38	18.05	23.44	8.57	19.92	25.73	12.04	
	MSTL ⁺ -BRITS	{4,24,168}	23.84	30.79	15.99	26.64	34.60	18.01	28.90	37.32	23.42	
		{24,168}	23.88	30.84	15.35	26.56	34.49	17.84	29.03	37.50	23.63	
	Fourier-Exogenous-BRITS	{4,24,168}	35.33	45.63	29.52	37.91	49.23	31.09	39.93	51.57	37.44	
	STL ⁺ -Exogenous-BRITS	{24,168}	35.40	45.71	29.91	36.81	47.80	30.04	40.00	51.66	37.88	
		4	35.89	46.34	29.33	36.69	47.65	28.52	39.22	50.65	34.60	
	MSTL ⁺ -Exogenous-BRITS	{4,24,168}	38.05	49.13	31.29	38.76	50.33	31.48	41.00	52.95	37.91	
		{24,168}	39.53	51.05	33.19	40.20	52.21	33.08	42.24	54.56	39.45	
		MuSDRI	{4,24,168}	16.17	20.89	7.01	17.42	22.62	8.17	19.40	25.05	11.73

The best performance achieved by the single-period imputation methods is under the setting of $\mu = 84$. When the setting of cycle length is under $\mu = 84$, the remainder component extracted by STL⁺ is similar to the remainder component extracted by MSTL⁺ under the setting of $\mu_1 = 4, \mu_2 = 24, \mu_3 = 168$ or the setting of $\mu_1 = 24, \mu_2 = 168$. The reason is that the seasonal component extracted by MSTL⁺ under $\mu_1 = 4$ has a rather small range of values, and the seasonal component extracted by STL⁺ under $\mu = 84$ is almost the same as the sum of the two seasonal components extracted by MSTL⁺ with the two cycle lengths of 12 and 84. Therefore, under this setting, the results of STL⁺ based methods and MSTL⁺ based methods are similar. In general, the methods that achieve the best performance among the set of traditional methods are STL⁺-KNN and MSTL⁺-KNN.

Among the set of RNN based imputation methods integrated with seasonal trend decomposition, MSTL⁺-BRITS achieves the best performance when the missing rate $p = 10$, and MuSDRI achieves the best performance when the missing rate $p = 20, 30$. Moreover, under the parameter setting of three cycles $\mu_1 = 4, \mu_2 = 24, \mu_3 = 168$, and $\lambda = 0.0$, MuSDRI obtains an overall improvement of 14.83% for MAE, 14.83%

for MRE, and 39.24% for MSE compared to the state-of-the-art imputation method BRITS.

VIII. EXPERIMENTS FOR CONSECUTIVE MISSING

This part of the experiments is mainly to evaluate the performance of MuSDRI and the imputation methods presented in VI under the setting of consecutive missing. Under this setting, data points are dropped out starting from a random position with a random running length.

A. Experiment Setting

The experiments under the setting of consecutive missing are based on the gas consumption dataset. The division of the training set and the test set is described in detail in Section V-C. Following the previous study which investigates the problem of missing value imputation for time series data with large gaps [28], we generate consecutive missing values with an upper bound of three days which can be considered as large gaps. That is, the number of data points that will be dropped out ranges from 1 (2 hours) to 36 (3 days) in the case of the gas consumption dataset, the sampling rate of which is two hours. Moreover, we control the total number of

TABLE III

PERFORMANCE OF ALL THE IMPUTATION METHODS WITH RESPECT TO THE GAS CONSUMPTION DATA UNDER THE SETTING OF POINT MISSING

Missing Rate			10%			20%			30%		
Method		μ	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)
Traditional	Last	\	44.94	65.34	106.98	49.50	71.99	121.83	52.97	77.92	131.55
	Mean		67.12	97.59	96.38	67.20	97.72	97.75	66.57	97.91	94.20
	KNN		25.77	37.48	40.13	29.30	42.61	45.88	30.52	44.90	48.78
Seasonal Trend Decomposition + Traditional	STL ⁺ -Last	\	23.58	34.28	24.84	25.50	37.08	29.78	26.37	38.78	30.86
	STL ⁺ -Mean		22.09	32.12	17.74	22.50	32.72	19.12	22.76	33.48	20.78
	STL ⁺ -KNN		21.18	30.79	16.09	22.35	32.51	16.31	23.23	34.17	18.92
	MSTL ⁺ -Last	{12,84}	23.58	34.28	24.84	25.50	37.08	29.78	26.37	38.78	30.86
	MSTL ⁺ -Mean		22.09	32.12	17.74	22.50	32.72	19.12	22.76	33.48	20.78
	MSTL ⁺ -KNN		21.18	30.79	16.09	22.35	32.51	16.31	23.23	34.17	18.92
	MuSDRI-Last	{4,12,84}	23.60	34.32	24.85	25.53	37.13	29.78	26.39	38.81	30.85
MuSDRI-Mean	22.11		32.14	17.73	22.52	32.75	19.11	22.78	33.50	20.76	
RNN	BRITS	\	19.60	28.49	19.71	20.32	29.55	20.78	21.70	31.92	23.34
	STL ⁺ -BRITS	84	16.94	24.63	12.30	17.92	26.05	13.20	18.43	27.11	14.42
Seasonal Trend Decomposition + RNN	MSTL ⁺ -BRITS	{4,12,84}	16.75	24.36	12.05	17.86	25.97	13.41	18.17	26.73	14.05
		{12,84}	16.92	24.60	12.43	17.87	25.99	13.18	18.51	27.23	14.57
	Fourier-Exogenous-BRITS	{4,12,84}	19.21	27.93	19.88	23.21	33.75	21.72	23.66	34.80	23.97
		{12,84}	22.62	32.90	20.05	20.97	30.49	23.74	22.34	32.86	25.01
	STL ⁺ -Exogenous-BRITS	84	18.46	26.85	17.96	19.80	28.80	20.78	20.10	29.56	22.31
	MSTL ⁺ -Exogenous-BRITS	{4,12,84}	17.51	25.46	17.67	18.49	26.89	18.16	20.11	29.57	22.86
		{12,84}	18.70	27.19	18.43	19.46	28.30	20.22	20.71	30.46	21.21
MuSDRI		{4,12,84}	16.85	24.51	12.15	17.50	25.44	12.85	18.10	26.63	13.71

TABLE IV

PERFORMANCE OF ALL THE IMPUTATION METHODS WITH RESPECT TO THE GAS CONSUMPTION DATA UNDER THE SETTING OF CONSECUTIVE MISSING

Missing Rate			10%			20%			30%			
Method			μ	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)	MAE(%)	MRE(%)	MSE(%)
Traditional	Last	\	75.08	116.01	195.84	68.76	97.93	175.95	72.53	102.11	188.78	
	Mean		63.81	98.59	83.69	68.18	97.11	99.55	68.51	96.45	103.48	
	KNN		56.77	87.72	71.82	62.49	89.00	92.63	61.57	86.67	88.96	
Seasonal Trend Decomposition + Traditional	STL ⁺ -Last	84	30.57	47.24	40.70	29.48	41.99	35.63	29.24	41.17	29.58	
	STL ⁺ -Mean		18.70	28.90	13.02	22.34	31.82	19.99	21.28	29.96	16.45	
	STL ⁺ -KNN		19.14	29.57	13.64	22.50	32.04	19.18	21.43	30.16	16.29	
	MSTL ⁺ -Last	{12,84}	30.57	47.24	40.70	29.48	41.99	35.63	29.24	41.17	29.58	
	MSTL ⁺ -Mean		18.70	28.90	13.02	22.34	31.82	19.99	21.28	29.96	16.45	
	MSTL ⁺ -KNN		19.14	29.57	13.64	22.50	32.04	19.18	21.43	30.16	16.29	
	MuSDRI-Last	{4,12,84}	31.40	48.53	40.04	30.04	42.79	35.54	29.91	42.10	29.83	
	MuSDRI-Mean		19.37	29.93	13.19	22.87	32.58	20.01	21.91	30.84	16.73	
RNN			BRITS			\			\			
Seasonal Trend Decomposition + RNN	STL ⁺ -BRITS	84	15.94	24.63	10.61	19.08	27.17	15.00	18.62	26.21	14.30	
	MSTL ⁺ -BRITS	{4,12,84}	16.11	24.90	10.53	18.68	26.60	14.82	18.76	26.40	14.15	
		{12,84}	16.06	24.82	10.42	18.97	27.02	15.21	18.82	26.49	14.40	
	Fourier-Exogenous-BRITS	{4,12,84}	25.94	40.08	30.60	30.15	42.94	35.08	32.96	46.40	41.90	
	STL ⁺ -Exogenous-BRITS	{12,84}	23.75	36.71	32.25	30.68	43.69	45.10	33.04	46.50	53.99	
		84	16.46	25.43	15.95	19.77	28.15	21.30	20.52	28.88	20.57	
	MSTL ⁺ -Exogenous-BRITS	{4,12,84}	16.71	25.82	16.49	19.81	28.22	20.45	18.58	26.16	19.16	
		{12,84}	15.96	24.66	14.72	19.93	28.38	20.36	18.85	26.53	18.56	
MuSDRI			{4,12,84}	15.67	24.21	10.09	18.97	27.02	14.69	18.55	26.11	14.12

missing values with a ratio of $p\%$ ($p = 10, 20, 30$). Note that these randomly dropped consecutive data points are from the set of data without any missing values in the original dataset such that the ground truth can be obtained. Similarly, three evaluation metrics, i.e. MAE, MRE, and MSE, are used to evaluate the performance of the imputation methods.

B. Parameter Selection

The selection of the parameters for the experiments under the setting of consecutive missing is similar to that under the setting of point missing as described in Section VII. The only difference is that for training BRITS, the length of the input sequences is set to 72 (6 days) instead of 36 (3 days) such that in the cases of 36 consecutive missing values BRITS may still extract useful information from the remaining data points of the sequences.

C. Experiments Results

Table IV shows the performance of all the imputation methods with respect to the gas consumption dataset under the setting of consecutive missing. Similar to the results under the

case of point missing, except for Fourier-Exogenous-BRITS, all the imputation methods integrated with seasonal trend decomposition outperform their correspondences (Last, Mean, KNN, BRITS), which again confirms the importance of long-term features for time series imputation. As aforementioned, the remainder component extracted by STL⁺ under the setting $\mu = 84$ is extremely close to the remainder component extracted by MSTL⁺. Therefore, the results of STL⁺ based methods and MSTL⁺ based methods are similar. In general, the methods that achieve the best performance among the set of traditional methods are STL⁺-Mean and MSTL⁺-Mean.

Under the parameter setting of ($\mu_1 = 4, \mu_2 = 12, \mu_3 = 84$) and $\lambda = 0.0$, MuSDRI obtains an overall improvement of 18.40% for MAE, 18.41% for MRE, and 45.68% for MSE compared to the state-of-the-art imputation method BRITS. In general, MuSDRI outperforms the other five imputation methods in the group of RNN based imputation methods integrated with seasonal trend decomposition.

IX. DISCUSSION AND LIMITATIONS

By investigating the experimental results presented in this paper and other experiments implemented on simulated data,

we have the following observations.

The superiority of the STL⁺ based methods against their correspondences (Last, Mean, KNN, BRITS) on these three datasets shows the importance of long-term features for time series imputation. The comparable performance of MuSDRI-Last and MuSDRI-Mean as to that of STL⁺-Last and STL⁺-Mean on the three datasets demonstrates the learning capability provided by MuSDRI. Moreover, MuSDRI in general achieved the best results, which confirms the effectiveness of its mechanism of adjusting the contribution of various seasonal decompositions along with the training of the imputation model.

Overall, STL⁺-BRITS achieved the second-best performance after MuSDRI. For the methods based on MSTL⁺, when the time series exhibits multiple seasonal patterns such as the gas consumption data, the performance is comparable to the methods based on STL⁺; but when the time series does not have an explicit seasonal pattern or only has a strong single periodicity such as the air quality data and the water pressure data, the performance is inferior to the methods based on STL⁺. This shows that the direct use of the MSTL⁺ heavily relies on the accurate selection of the cycle lengths of the time series data, otherwise it will interfere with the imputation process.

Moreover, the methods with the exogenous training mechanism, such as STL⁺-exogenous-BRITS, MSTL⁺-exogenous-BRITS, and Fourier-exogenous-BRITS, generally do not perform well, and sometimes their performances are even worse than that of BRITS. At the same time, STL⁺-exogenous-BRITS and MSTL⁺-exogenous-BRITS are both inferior to STL⁺-BRITS and MSTL⁺-BRITS. These imply that the exogenous training mechanism may not suit the task of time series imputation.

When a time series exhibits multiple seasonalities with different cycle lengths and the longer cycle length can be divisible by the shorter cycle length such as the gas consumption dataset, the decomposition results of STL⁺ under the setting of the longest cycle length is similar to that of MSTL⁺ under the setting of a combination of all the cycle lengths. In such cases, STL⁺-BRITS can achieve comparable performance as that of the imputation methods integrated with multi-seasonal decomposition such as MSTL⁺-BRITS and MuSDRI. However, the application of STL⁺-BRITS or MSTL⁺-BRITS relies on the accurate selection of the cycle lengths of a time series. The advantage of MuSDRI is that it does not require an accurate prior knowledge of the seasonalities of the time series. By including some well-known cycles of human activities, e.g., short cycle, daily cycle, weekly cycle, etc., MuSDRI can then automatically learn their contributions with respect to the imputation task.

Time complexity is also an important factor to be considered. The main operations in STL⁺ include loess smoothing and moving average. Each loess operation is only performed at a fixed subset of time points and interpolated elsewhere. Since STL⁺ only linearly interpolates between data points, it keeps the computation time linear in μn where μ is the cycle length and n is the length of the time series. Moreover, in most cases the cycle length is much smaller than the length

of the time series, i.e. $\mu \ll n$. Therefore, compared to RNN-based imputation methods such as BRITS, the computational cost added by STL⁺ is neglectable.

As shown in Section VII and VIII, the application of MuSDRI is limited to the time series with missing values less than or equal to 30% as it is difficult for STL⁺ to effectively extract the trend and seasonal information from data containing more missing values. Except for the energy consumption dataset, in the rest of the three real-world datasets used in this paper, the amount of missing values is relatively small (15.1%, 4.1%, 5.1%) and we argue this usually holds for time series data collected from real-world applications. That is, MuSDRI can be applied to a broad range of real-world datasets.

X. CONCLUSION

In this paper, we carry out an extensive investigation of different paradigms of time series imputation methods to capture temporal dynamics from both a long-term and a short-term perspective. In particular, we propose a new learning paradigm called MuSDRI to impute the missing values in multivariate time series.

- On the one hand, MuSDRI makes use of the time series decomposition method STL to extract the long-term dynamics with multiple seasonal patterns. On the other hand, MuSDRI takes advantage of the multi-seasonal decomposition results and adapts the state-of-the-art RNN based imputation method to refine the learning of short-term dynamics.
- MuSDRI provides a mechanism to learn the contributions of multiple seasonal decompositions with respect to the imputation accuracy, which is able to deal with both situations when there is a strong single seasonal pattern as well as when no single seasonal pattern stands out.
- Through a set of experiments with three real-world datasets, we show that MuSDRI achieves more accurate results for the imputation task under both settings of point missing and consecutive missing.

For future work, we intend to investigate ways of setting up interactions between the Multi-seasonal Decomposition module and the Recurrent Imputation module such that the imputation results can also be used to refine the time series decomposition.

REFERENCES

- [1] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.
- [2] J. Bright, C. Smith, P. Taylor, and R. Crook, "Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data," *Solar Energy*, vol. 115, pp. 229–242, 2015.
- [3] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," *arXiv preprint arXiv:1706.02633*, 2017.
- [4] S. Zhou, T. McMahon, A. Walton, and J. Lewis, "Forecasting operational demand for an urban water supply zone," *Journal of hydrology*, vol. 259, no. 1–4, pp. 189–202, 2002.
- [5] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, ser. AAAI'17. AAAI Press, 2017, p. 1655–1661.

- [6] B. Fekade, T. Maksymyuk, M. Kyryk, and M. Jo, "Probabilistic recovery of incomplete sensed data in iot," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2282–2292, 2018.
- [7] J. Du, M. Hu, and W. Zhang, "Missing data problem in the monitoring system: A review," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 13 984–13 998, 2020.
- [8] A. T. Hudak, N. L. Crookston, J. S. Evans, D. E. Hall, and M. J. Falkowski, "Nearest neighbor imputation of species-level, plot-scale forest structure attributes from lidar data," *Remote Sensing of Environment*, vol. 112, no. 5, pp. 2232–2245, 2008.
- [9] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control*. John Wiley & Sons, 2011, vol. 734.
- [10] I. R. White, P. Royston, and A. M. Wood, "Multiple imputation using chained equations: issues and guidance for practice," *Statistics in medicine*, vol. 30, no. 4, pp. 377–399, 2011.
- [11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [12] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "Brits: Bidirectional recurrent imputation for time series," in *Advances in Neural Information Processing Systems*, 2018, pp. 6775–6785.
- [13] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "Stl: A seasonal-trend decomposition," *Journal of official statistics*, vol. 6, no. 1, pp. 3–73, 1990.
- [14] A. C. Harvey and N. Shephard, "10 structural time series models," 1993.
- [15] "KDD-CUP," <http://www.kdd.org/kdd2018/>, 2020.
- [16] "Slawek2020hybrid," <https://www.sodic.com.cn/>, 2020.
- [17] J. W. Graham, "Missing data analysis: Making it work in the real world," *Annual review of psychology*, vol. 60, pp. 549–576, 2009.
- [18] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011.
- [19] M. Amiri and R. Jensen, "Missing data imputation using fuzzy-rough methods," *Neurocomputing*, vol. 205, pp. 152–164, 2016.
- [20] A. Purwar and S. K. Singh, "Hybrid prediction model with missing value imputation for medical data," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5621–5631, 2015.
- [21] A. Zhang, S. Song, Y. Sun, and J. Wang, "Learning individual models for imputation," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 160–171.
- [22] J. W. Galbraith, V. Zinde-Walsh *et al.*, "Autoregression-based estimators for arfima models," CIRANO, Tech. Rep., 2001.
- [23] C. Hamzaçebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Information Sciences*, vol. 178, no. 23, pp. 4550–4559, 2008.
- [24] D. S. Fung, "Methods for the estimation of missing values in time series," 2006.
- [25] E. Afrifa-Yamoah, U. A. Mueller, S. Taylor, and A. Fisher, "Missing data imputation of high-resolution temporal climate time series data," *Meteorological Applications*, vol. 27, no. 1, p. e1873, 2020.
- [26] J. Mei, Y. De Castro, Y. Goude, and G. Hébrail, "Nonnegative matrix factorization for time series recovery from a few temporal aggregates," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2382–2390.
- [27] X. Song, Y. Guo, N. Li, and S. Yang, "A novel approach based on matrix factorization for recovering missing time series sensor data," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 491–13 500, 2020.
- [28] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Missing value imputation for industrial iot sensor data with large gaps," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6855–6867, 2020.
- [29] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.
- [30] J. Yoon, W. R. Zame, and M. van der Schaar, "Multi-directional recurrent neural networks: A novel method for estimating missing data," in *Time Series Workshop in International Conference on Machine Learning*, 2017.
- [31] Y. Luo, X. Cai, Y. Zhang, J. Xu *et al.*, "Multivariate time series imputation with generative adversarial networks," in *Advances in Neural Information Processing Systems*, 2018, pp. 1596–1607.
- [32] Y. Luo, Y. Zhang, X. Cai, and X. Yuan, "E2gan: End-to-end generative adversarial network for multivariate time series imputation," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, 2019, pp. 3094–3100.
- [33] T. Huang, P. Chakraborty, and A. Sharma, "Deep convolutional generative adversarial networks for traffic data imputation encoding time series as images," *arXiv preprint arXiv:2005.04188*, 2020.
- [34] Y. Liu, R. Yu, S. Zheng, E. Zhan, and Y. Yue, "Naomi: Non-autoregressive multiresolution sequence imputation," *arXiv preprint arXiv:1901.10946*, 2019.
- [35] V. Fortuin, D. Baranchuk, G. Rätsch, and S. Mandt, "Gp-vae: Deep probabilistic time series imputation," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 1651–1661.
- [36] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020, m4 Competition. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207019301153>
- [37] K. Bandara, C. Bergmeir, and H. Hewamalage, "Lstm-msnet: Leveraging forecasts on sets of related time series with multiple seasonal patterns," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.
- [38] Y. Huo, Y. Yan, D. Du, Z. Wang, Y. Zhang, and Y. Yang, "Long-term span traffic prediction model based on stl decomposition and lstm," in *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2019, pp. 1–4.
- [39] R. P. Hafen, *Local Regression Models: Advancements, applications, and new methods*. Purdue University, 2010.
- [40] R. J. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeeen, "forecast: Forecasting functions for time series and linear models," 2018.

Yujue Zhou received the B.S. degree in physics from Zhejiang University, China, in 2015, and M.S. degree in data analytics (with distinction) from University of Warwick, UK, in 2018. She is currently a PhD student in the Department of Computer Science at the University of Warwick under a joint PhD program between University of Warwick and Southern University of Science and Technology (SUSTech). She is a member of Marie Curie Alumni Association.

Jie Jiang received her B.S. in Automation from Chang'an University, China in 2007, M.S. degree in Control Theory and Engineering from Xi'an Jiaotong University, China in 2010, and Ph.D. degree in Artificial Intelligence from Delft University of Technology, the Netherlands in 2015. After finishing her PhD, she joined University of Surrey as a Research Fellow. She is currently a Research Assistant Professor in the department of computer science and engineering at Southern University of Science and Technology. Her

research interests primarily lie in IoT, time series analysis, and machine learning.

Shuang-Hua Yang received the B.S. degree in instrument and automation and the M.S. degree in process control from the China University of Petroleum (Huadong), Beijing, China, in 1983 and 1986, respectively, and the Ph.D. degree in intelligent systems from Zhejiang University, Hangzhou, China, in 1991. He was awarded DSc from Loughborough University in 2014 to recognize his academic contribution to wireless monitoring research. He is currently a chair professor of computer science and Executive Deputy

Dean of Graduate School with Southern University of Science and Technology (SUSTech), Shenzhen, China. Before joined SUSTech in 2016 he had spent over two decades in Loughborough University, as a professor in computer science and head of department. His current research interests include cyber-physical system safety and security, Internet of Things, wireless network-based monitoring and control. He is a Fellow of IET and a Fellow of InstMC, U.K. He is an Associate Editor of the IET Journal Cyber-Physical Systems Theory and applications, the InstMC Journal Measurement and Control, and the International Journal of Computing and Automation.

Ligang He received the B.S. and M.S. degree from Huazhong University of Science and Technology, China, and received the Ph.D. degree in Computer Science at the University of Warwick, United Kingdom. After PhD studies, he worked as a post-doctoral researcher at the University of Cambridge, UK, and then joined in the Department of Computer Science at the University of Warwick as Assistant Professor. He is now the Reader in the department. His research interests focus on parallel and distributed processing

and Cloud computing. He has published more than 100 papers in international conferences and journals. He has been a co-chair or a member of the program committee for a number of international conferences, and been the reviewers for many international journals. He is a member of the IEEE.

Yulong Ding received the B.S. and M.S. degrees in Chemical Engineering from Tsinghua University, China, in 2005 and 2008, respectively and the Ph.D. in Chemical Engineering from University of British Columbia, Canada, in 2012. He is currently a Research Assistant Professor with the Academy for Advanced Interdisciplinary Studies of Southern University of Science and Technology. His main interests are Industrial Internet of Things and Low-Power Wide Area Networks (LPWAN).