Statistics and Computing manuscript No.
(will be inserted by the editor)

A wavelet-based approach for imputation in nonstationary multivariate time series

Rebecca E. Wilson · Idris A. Eckley · Matthew A. Nunes · Timothy Park

Received: date / Accepted: date

Abstract Many multivariate time series observed in practice are second order nonstationary, i.e. their covariance properties vary over time. In addition, missing observations in such data are encountered in many applications of interest, due to recording failures or sensor dropout, hindering successful analysis. This article introduces a novel method for data imputation in multivariate nonstationary time series, based on the so-called locally stationary wavelet modelling paradigm. Our methodology is shown to perform well across a range of simulation scenarios, with a variety of missingness structures, as well as being competitive in the stationary time series setting. We also demonstrate our technique on data arising in an environmental application.

Keywords imputation; local stationarity; missing data; multivariate time series.

1 Introduction

Time series data arise in a variety of different areas including finance (Taylor, 2007), biology (Bar-Joseph et al, 2003) and energy (Alvarez et al, 2011; Doucoure et al, 2016). The collection and recording of time series can be interrupted due to various reasons, including human error or technical faults with the recording equipment, inducing missingness within the time series. Little and Rubin (2002) describe the occurrence of missing values in data through a number of "missingness mechanisms":

1. Missing completely at random (MCAR) - The probability of missingness is the same for all units, i.e. the missing value is not dependent on other variables.
2. Missing at random (MAR) - The probability of missingness depends only on available information, i.e. the missing value depends on other variables.
3. Not missing at random (NMAR) - The missingness probability depends on the variable itself, i.e. the missing observation depends on other missing values.

Regardless of the type of missingness present, further analysis of the time series such as autocovariance or spectral estimation can be difficult without first replacing the missing data with appropriate estimates. This estimation process is called imputation.

There exists a rich literature dedicated to the imputation of missing values within *stationary* time series; see Pratama et al (2016) for an recent review of this literature. For univariate time series, explicit consideration of the temporal dependence (autocorrelation properties) is key for successful imputation. Moritz and Bartz-Beielstein (2017) discuss various univariate time series imputation approaches, ranging from simple methods that replace missing values with the mean to more advanced approaches that involve spline interpolation

Rebecca E. Wilson
STOR-i Centre for Doctoral Training, Lancaster University, LA1 4YF, United Kingdom

Idris A. Eckley
Department of Mathematics and Statistics, Lancaster University, LA1 4YF, United Kingdom

Matthew A. Nunes
Department of Mathematics, University of Bath, BA2 7AY, United Kingdom
E-mail: m.a.nunes@bath.ac.uk

Timothy Park
Statistics and Data Science, Shell Global Solutions International BV, Amsterdam, The Netherlands

or model fitting combined with the use of a Kalman filter.

Popular techniques suitable for multivariate data such as Multiple Imputation (see e.g. Rubin (1987) or Audigier et al (2016)), Hot-Deck (Ford, 1983) and Expectation-Maximisation (EM) (Dempster et al, 1977) make use of inter-variable correlations to estimate missing data. Most variants of the EM approach within the literature are based on assumptions of Gaussianity (see e.g. Junger and de Leon (2015), Honaker and King (2010) or Honaker et al (2011)). Alternative methods have also been developed, combining EM with other modelling procedures such as PCA fixed-effects models (Caussinus, 1986) and Gaussian Mixture Models (Ghahramani and Jordan, 1994). Other model-based methods for imputation within time series make use of restrictive classes of statistical models to infer missing values, such as those based on purely autoregressive processes (Sridevi et al, 2011).

Other approaches to imputation include those based on heuristics such as genetic algorithms (Lobato et al, 2015; Tang et al, 2015), or those incorporating machine learning methodology, ranging from support vector machines (Wu et al, 2015) and random forests (Stekhoven and Bühlmann, 2011), to more advanced techniques such as recurrent neural networks (Cao et al, 2018; Che et al, 2018) or adversarial networks (Luo et al, 2018; Yoon et al, 2018). The drawback of many of these approaches is that they often require training on previously-seen complete data. However, in many contexts, we only have access to a single observed multivariate series.

An alternative strategy for coping with missingness is to estimate the spectral information of the series in some way. A range of methods have been developed for spectral estimation in stationary time series with missing values or irregularly sampled observations within the time series and signal processing literature. The Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982) estimates the Fourier spectrum from the irregularly sampled data but can be subject to strong bias which hinders its ability to describe slopes within the spectrum. Variants of this approach have been applied in a range of different fields including astronomy (Wen et al, 1999), biology (Van Dongen et al, 1999) and biomedical engineering (Laguna et al, 1998). Other widely used techniques involve fitting time series models directly to the unequally spaced data and using this to estimate spectral information for stationary processes (Jones, 1980; Bos et al, 2002; Broersen, 2006).

In practice however, the assumptions imposed by modelling an observed time series as stationary can be restrictive and unrealistic. Nonstationary time series, i.e. series with time-varying second order structure, are observed in various fields including finance (Stărică and Granger, 2005; Fryzlewicz et al, 2006), medicine (Cranstoun et al, 2002) and oceanography (Killick et al, 2013). Hence the ability to impute missing values in multivariate, nonstationary time series has potential widespread benefits. Many techniques have been developed for modelling and analysing complete multivariate, nonstationary data including the locally stationary Fourier model (Dahlhaus, 2000), the smooth localized complex exponential (SLEX) model (Ombao et al, 2005) and the multivariate locally stationary wavelet (mvLSW) framework (Park et al, 2014) but the literature on how to deal with missingness within such data is sparse. In the univariate setting, Knight et al (2012) propose a method for estimating spectral information of a LSW process containing missing values where information is estimated at the observed time locations. However, the problem of spectral estimation within multivariate, nonstationary time series has not been widely studied.

In this article, we address the challenging problem of imputation in the multivariate locally stationary time series setting and, more specifically, where additional training data is unavailable. Our approach involves first estimating the local wavelet spectral matrix of a mvLSW process with missing observations before forecasting and backcasting the missing values of the time series using a multivariate extension of the wavelet forecasting approach of Fryzlewicz et al (2003). As a final step, we average the estimates obtained from the forward and backward pass to obtain an overall estimate of the time series. Through the use of simulated examples and a case study, we demonstrate that our method performs well for a range of realistic missingness scenarios in both the stationary and nonstationary setting.

This work is organised as follows. Within Section 2, we review existing methods for modelling locally stationary time series and forecasting within this context. In Section 3, we introduce the proposed imputation method. Section 4 contains a simulation study evaluating the performance of the proposed imputation method using synthetic examples. We also describe a case study using a dataset arising from a Carbon Capture and Storage facility. Finally, Section 6 includes some concluding remarks.

2 Background

In this section we provide a brief overview of recent work modelling locally stationary time series. For a comprehensive review of nonstationary time series more generally, see Dahlhaus (2012). The section is organised as follows; we first review existing methods for

modelling locally stationary time series using the LSW framework both in a univariate and multivariate context in Section 2.1 before looking at one-step ahead forecasting in this setting in Section 2.2.

2.1 Modelling locally stationary wavelet processes

Within the univariate setting, the locally stationary wavelet (LSW) framework introduced by Nason et al (2000) provides a flexible model for nonstationary time series that captures the changing second-order structure of such series. Nason et al (2000) define a LSW process $\{X_{t,T}\}_{t=0,\ldots,T-1}$, $T = 2^J \geq 1$ to be a sequence of (doubly-indexed) stochastic processes that can be represented as

$$X_{t,T} = \sum_{j=1}^{\infty} \sum_{k} W_j(k/T)\psi_{j,t-k}\xi_{j,k}. \tag{1}$$

The vectors $\psi_j = \left\{\psi_{j,0}, \psi_{j,1}, \ldots, \psi_{j,N_j-1}\right\}$ are discrete non-decimated wavelets associated with a low- / high-pass filter pair, $\{\mathcal{H}, \mathcal{G}\}$. The elements of ψ_j can be calculated using the following expression:

$$\psi_{1,n} = \sum_{k} g_{n-2k}\delta_{0,k} = g_n \quad \text{for } n = 0, 1, \ldots, N_1 - 1,$$
$$\psi_{j+1,n} = \sum_{k} h_{n-2k}\psi_{j,k} \quad \text{for } n = 0, 1, \ldots, N_{j+1} - 1.$$

In the equations above, $\delta_{0,k}$ is the Kronecker delta function and $N_j = \left(2^j - 1\right)(N_h - 1) + 1$ where N_h is the number of non-zero elements of the filter $\mathcal{H} = \{h_k\}_{k\in\mathbb{Z}}$. $\{\xi_{j,k}\}_{j,k}$ is a sequence of zero-mean, orthonormal random variables and $\{W_j(k/T)\}$ is a set of amplitudes on which a number of assumptions are imposed to control the behaviour of the LSW process (see Nason et al (2000) for further details).

Park et al (2014) introduced the multivariate locally stationary wavelet (mvLSW) processes as a multivariate extension to the LSW framework.Following Park et al (2014), a P-variate locally stationary wavelet process $\{\mathbf{X}_{t,T}\}_{t=0,1,\ldots,T-1}$, $T = 2^J \geq 1$ has the following representation

$$\mathbf{X}_{t,T} = \sum_{j=1}^{\infty} \sum_{k} \mathbf{V}_j(k/T)\psi_{j,t-k}\mathbf{z}_{j,k} \tag{2}$$

where $T = 2^J \geq 1\{\psi_{j,t-k}\}_{j,k}$ is a set of discrete non-decimated wavelets constructed according to Nason et al (2000) and $\mathbf{V}_j(k/T)$ is the transfer function matrix. The random vectors $\mathbf{z}_{j,k}$ are uncorrelated, have mean vector $\mathbf{0}$ and variance-covariance matrix equal to the $P\times P$ identity matrix. The transfer function matrix consists of Lipschitz continuous functions with Lipschitz constants, L_j, that satisfy $\sum_{j=1}^{\infty} 2^j L_j^{(p,q)} < \infty$ for each pair of channels (p, q). The conditions imposed upon the transfer function matrix control the time-varying contribution to the variance made by each channel at a particular scale.

The local wavelet spectral (LWS) matrix is an important quantity within the mvLSW framework as it provides a scale-dependent decomposition of the variance and cross-covariance between channels at a particular (rescaled) time z. Given a mvLSW signal $\mathbf{X}_t$ with transfer function matrix $\mathbf{V}_j(k/T)$, the LWS matrix is given by

$$\mathbf{S}_j(z) = \mathbf{V}_j(z)\mathbf{V}_j(z)^\top \tag{3}$$

for scale j and rescaled time z. Following Park et al (2014), the local auto- and cross-covariance between channels p and q are defined as

$$c^{(p,p)}(u, \tau) = \sum_{j=1}^{\infty} S_j^{(p,p)}(u)\Psi_j(\tau), \tag{4}$$

$$c^{(p,q)}(u, \tau) = \sum_{j=1}^{\infty} S_j^{(p,q)}(u)\Psi_j(\tau) \tag{5}$$

where $S_j^{(p,p)}(u)$ and $S_j^{(p,q)}(u)$ denote the spectra and cross-spectra respectively of the series, and $\Psi_j(\tau)$ is the discrete autocorrelation wavelet defined by $\Psi_j(\tau) = \sum_k \psi_{j,k}\psi_{j,k-\tau}$ for $j \in \mathbb{N}$ and $\tau \in \mathbb{Z}$ (Eckley and Nason, 2005).

In practice, the LWS matrix and local auto- and cross-covariance are unknown for an observed multivariate series and need to be estimated. In the complete data case, where there are no observations missing, the LWS matrix of a multivariate signal can be estimated by first calculating the empirical wavelet coefficient vector $\mathbf{d}_{j,k} = \sum_t \mathbf{X}_t\psi_{j,k}$. The raw wavelet periodogram matrix is then defined as $\mathbf{I}_{j,k} = \mathbf{d}_{j,k}\mathbf{d}_{j,k}^\top$.

Park et al (2014) show that the raw wavelet periodogram is a biased and inconsistent estimator of the true LWS matrix, $\mathbf{S}_j(z)$. The bias within the raw wavelet periodogram can be removed using the inverse of the inner product matrix of discrete autocorrelation wavelets, $\mathbf{A}$. The elements of $\mathbf{A}$ are given by $\mathbf{A}_{j,\ell} = \sum_\tau \Psi_j(\tau)\Psi_\ell(\tau)$ where $\Psi_j(\tau) = \sum_k \psi_{jk}(0)\psi_{jk}(\tau)$ (see Eckley and Nason (2005) or Nason et al (2000) for further details). To obtain consistency, the resulting estimate must be smoothed in some way, for example using a rectangular kernel smoother (Park et al, 2014). The local auto- and cross-covariance can then be estimated by substituting the estimated LWS matrix $\hat{\mathbf{S}}$ into Equations (4) and (5) respectively. The LWS matrix along with the local auto- and cross-covariance structure are important quantities within the imputation method we propose in Section 3

as they are used within the prediction step to estimate missing values. See Taylor et al (2019) for further details and an introduction to software that implements the mvLSW estimation scheme.

In the missing data setting, there exist some time points for which we do not have an observed value, $\mathbf{X}$, for one or more channels of the P-variate series. Due to this, we cannot directly apply the above approach for estimating the wavelet periodogram and have to modify the method to allow us to obtain estimates of the wavelet coefficients at all scales and locations. This will be discussed further in Section 3.

2.2 Forecasting multivariate locally stationary wavelet processes

As outlined in Section 1, a key element of the imputation scheme that we introduce in Section 3 is the ability to perform one-step ahead forecasts in a locally stationary setting. Fryzlewicz et al (2003) introduce a one-step ahead predictor in the univariate LSW setting which uses the autocovariance structure of an LSW process to form generalised Yule-Walker equations. Our approach is a straightforward extension of the foregoing work to the multivariate setting that makes use of the mvLSW model outlined in Section 2.1.

Due to the separable structure of the mvLSW model, we can form one-step ahead prediction equations for each channel combination (p, q) using the local auto- and cross-covariance defined in equations (4) and (5) respectively. The multivariate prediction equations are defined by

$$\sum_{s=0}^{t-1} b_{t-1-s,T}^{(p,q)} \; c^{(p,q)}\left(\frac{s+n}{2T}, s-n\right) = c^{(p,q)}\left(\frac{n+t}{2T}, t-n\right). \tag{6}$$

As Fryzlewicz et al (2003) describe in the univariate setting, the coefficients $\mathbf{b}_t^{(p,q)}$ that solve the prediction equations can be shown to minimise the mean square prediction error

$$\begin{aligned}\mathrm{MSPE}\big(\hat{X}_{t,T}^{(p,q)}, X_{t,T}^{(p,q)}\big) &= \mathrm{E}\big(\hat{X}_{t,T}^{(p,q)} - X_{t,T}^{(p,q)}\big)^2 \\ &= \mathbf{b}_t^{(p,q)\top} \boldsymbol{\Sigma}_{t,T}^{(p,q)} \mathbf{b}_t^{(p,q)},\end{aligned}$$

where $\mathbf{b}_t^{(p,q)} = (b_{t-1,T}^{(p,q)}, \ldots, b_{0,T}^{(p,q)}, -1)$ and $\boldsymbol{\Sigma}_{t,T}^{(p,q)}$ is the covariance matrix of $X_{0,T}^{(p,q)}, \ldots, X_{t,T}^{(p,q)}$.

The one-step ahead predictor of $\mathbf{X}_{t,T}$ given previous multivariate observations $\mathbf{X}_{0,T}, \ldots, \mathbf{X}_{t-1,T}$ is then given by the *m-observation clipped predictor*

$$\hat{X}_{t,T}^{(p)} = \sum_{q \in \{1,\ldots,P\}} \sum_{s=t-m}^{t-1} b_{t-1-s;T}^{(p,q)} X_{s,T}^{(q)} \tag{7}$$

for $p \in \{1, \ldots, P\}$, where m is the number of recent observations used in prediction.

Our proposed imputation approach (introduced in the next section) uses the one-step ahead prediction in the mvLSW setting outlined above to replace missing values in a multivariate locally stationary time series.

3 Imputation for multivariate locally stationary wavelet processes

In this section we introduce our multivariate imputation method which uses the local auto- and cross-covariance structure of a nonstationary time series to estimate missing observations. The key challenge in this context is that the usual mvLSW spectral estimation process cannot be used due to the presence of missingness. For this reason, the first step of the algorithm involves estimating the wavelet periodogram of a mvLSW process with missing observations, this will be discussed in Section 3.1. Using the estimate of the LWS matrix, we then form the local auto- and cross-covariance structure and carry out a forward pass of the data where we forecast missing values. To obtain more accurate estimates of the time series at missing locations, we also implement a backward pass of the data where we backcast the missing values. We then average the series obtained from the forward and backward pass in order to get an overall estimate. The forecasting and backcasting steps will be described in Section 3.2 and 3.3 respectively. A complete overview of the steps carried out in one iteration of the method can be found in Appendix A.

3.1 Spectral Estimation

Suppose we have a P-variate time series of length $T = 2^J$ containing missing values which we wish to impute. The first step of the `mvLSWimpute` algorithm involves estimating the LWS matrix of the time series. The presence of missing values means that we can not use the usual estimation procedure and have to modify our approach.

First, we calculate the empirical wavelet coefficient vector, $\mathbf{d}_{j,k}$, for the time series ensuring that any wavelet coefficients at scales and locations affected by the initial missing data will also be missing. The Haar wavelet is used within the calculation of the empirical wavelet coefficient vector as this will ensure that more levels of the wavelet transform will contain information. From this, we can form the raw wavelet periodogram matrix as described in Section 2.1. Since the raw wavelet periodogram will also have entries missing, we need to

perform an intermediate step to fill in these missing values before smoothing and correcting as described in Section 2.1 to obtain an estimate of the LWS matrix. We note here that other wavelets could be used in the calculation of the coefficients; for further discussion of the choice of wavelet in the LSW context, we refer the reader to Gott and Eckley (2013).

In order to fill in the missing values, for each spectra and cross-spectra we linearly interpolate the periodogram values at the levels in which we have some information present. For the coarsest levels of the periodogram where the coefficients are all missing, we replace them in a different manner. To do this, we take the coarsest level of the periodogram that contains information and recursively apply the wavelet filter equations. This process generates coefficients that allow us to replace the values in the coarsest levels of the periodogram.

To obtain an estimate of the LWS matrix, we correct the periodogram by multiplying by $\mathbf{A}^{-1}$ and then smooth the result using a running mean smoother with window length $\lfloor\sqrt{T}\rfloor$, implemented in the `mvLSW` *R* package (Taylor et al (2017)). This estimate can then be substituted into equations (4) and (5) to form the local auto and cross-covariance which are used in the forecasting and backcasting steps of the algorithm.

3.2 Forecasting

In order to replace missing data in the time series, we first carry out a forward pass of the series where we use the one-step ahead multivariate wavelet forecasting approach outlined in Section 2.2. A missing index is defined to be a timepoint at which one or more channels of the P-variate time series has missing values present.

For each missing index i, we forecast the missing values sequentially in the following way. First, calculate the local auto- and cross-covariance using the estimated spectra from time 1 to time $i-1$ and equations (4) and (5).

For each channel combination (p, q) where $p, q \in \{1, \ldots, P\}$, form the prediction equations using the local auto- and cross-covariance at certain locations and lags, as in equation (6). Solving the prediction equations allows us to obtain $\mathbf{b}^{(p,q)}$ vectors used to predict the values of the series at time i using the one-step ahead predictor defined in equation (7). The channels of the multivariate time series that contain missing data at time i are then replaced by the corresponding predicted values from the forecasting step.

It is important to note that, for efficiency, we use a clipped predictor in the forecasting step in which only the most recent m observations are used in the prediction, similar to Fryzlewicz et al (2003) in the univariate setting.

3.3 Backcasting

After carrying out the forward pass of the data, the next step is to backcast the missing values sequentially. This backcasting step is included in order to improve the accuracy of the imputation method since this allows us to incorporate information from both sides of the missing observation in our estimation. Similar to the approach of Trindade (2003), we can form backward Yule-Walker equations in the mvLSW setting by beginning at time T and again using the multivariate wavelet forecasting approach from Section 2.2, but ensuring to order the spectral values in this case. Note that the backward pass is carried out independently to the forward pass and does not depend on the imputed time series obtained in the previous step.

For each missing index i (considered in descending order), we proceed as in the forecasting case and form the local auto- and cross-covariance using the estimated spectra from time T to $i+1$. As in the forward pass, for each channel combination (p, q), we can solve the prediction equations using the local auto- and cross-covariance to obtain the $\mathbf{b}^{(p,q)}$ vectors. However, the one-step ahead predictor has a slightly different form in the backcasting step:

$$\hat{X}^{(p)}_{t,T} = \sum_{q \in \{1,\ldots,P\}} \sum_{s=t+m}^{t+1} b^{(p,q)}_{t+1-s;T} X^{(q)}_{s,T} \tag{8}$$

for $p \in \{1, \ldots, P\}$. The one-step ahead predictor in equation (8) is then used to backcast the value of the time series at index i and then any missing entries within channels are replaced using their corresponding predicted values.

After carrying out the forward and backward pass independently, we obtain two imputed time series which are then averaged to get an overall estimate of the time series. The process can then be iterated but from the second iteration onwards the spectral estimation step no longer requires linear interpolation and the LWS matrix can now be estimated using equation (3). The forecasting and backcasting steps remain the same and we again average to obtain an updated estimate of the time series.

4 Simulated performance of `mvLSWimpute`

We now assess the performance of our proposed multivariate imputation method through a range of simu-

lated data examples. The generating series used within the simulation study exhibit varying degrees of nonstationarity and dependence. These have been chosen to test the ability of our method to impute missing values in multivariate, nonstationary time series. A number of different scenarios have been chosen for the missingness structure in order to mimic situations arising in practice.

For datasets containing missing entries, traditional analysis based on complete cases has proven to be reasonably accurate provided that the amount of missing values is small (Graham, 2009). However, such methods yield poor results when the proportion of missing entries increases. To evaluate the performance of the proposed method as the amount of missingness increases, we remove 10%, 20%, 30% and 40% of values from the generating series at random, either from all channels simultaneously or from one channel independently. In practice, time series obtained from industrial applications can contain gaps that may extend over hours or even days due to faults in the recording equipment or human error. In order to reflect this, we also consider the case where information is missing from one or more variables of the time series for a period of 20 consecutive time points. As a third scenario, we also include the situation where the missingness occurs in bursts up to length 20 before the signal returns to normal for a set period of time.

For all missingness scenarios, the coefficients of the generating series randomly switch at set times in order to test the ability of the imputation methods to deal with slowly and rapidly evolving dependence within a signal. The time series used in each case have the following forms:

- Slowly changing structure: Trivariate signal of length $T = 512$, two changes in the generating coefficients of the series, occurring at time 150 and 300.
- Rapidly changing structure: Trivariate signal of length $T = 512$, four changes in the generating coefficients of the series, occurring at time 100, 200, 300 and 400.

The first example we consider is a mvLSW process with changing spectral structure, chosen in such a way that there is strong coherence between channels of the signal. In this case, the example consists of two underlying classes with differing LWS matrices as defined below

$$\text{Class 1}: \quad S_j^{(1,2)}(z) = \begin{cases} 5 & \text{for } j = 1, \\ -6 & \text{for } j = 2, \\ -4 & \text{for } j = 5. \end{cases}$$

$$S_j^{(1,3)}(z) = \begin{cases} 2 & \text{for } j = 3, \\ -4 & \text{for } j = 4. \end{cases}$$

$$S_j^{(2,3)}(z) = -6 \text{ for } j = 2.$$

$$\text{Class 2}: \quad S_j^{(1,2)}(z) = \begin{cases} -5 & \text{for } j = 1, \\ 6 & \text{for } j = 2, \\ 4 & \text{for } j = 5. \end{cases}$$

$$S_j^{(1,3)}(z) = \begin{cases} 8 & \text{for } j = 3, \\ 4 & \text{for } j = 4. \end{cases}$$

$$S_j^{(2,3)}(z) = 6 \text{ for } j = 2.$$

Figure 1(a) displays a dataset simulated from this process using the different dependence structures describe above. The second example we examine is a time-varying vector autoregressive moving average process with three different classes defined by the following coefficient matrices

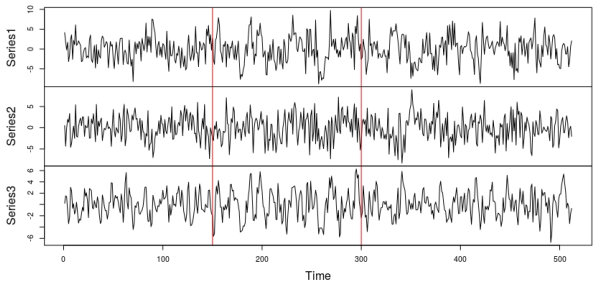
$$\text{Class 1}: \ \mathbf{X}_t = \begin{pmatrix} 0.4 & 0.1 & -0.2 \\ 0.1 & 0.3 & -0.3 \\ -0.2 & -0.3 & -0.2 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t + \begin{pmatrix} 1 & 0.8 & 0.4 \\ 0.8 & 1 & 0.1 \\ 0.4 & 0.1 & 1 \end{pmatrix} \mathbf{Z}_{t-1}$$

$$\text{Class 2}: \ \mathbf{X}_t = \begin{pmatrix} -0.3 & -0.2 & 0.3 \\ -0.2 & -0.3 & 0.1 \\ 0.3 & 0.1 & 0.2 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t + \begin{pmatrix} 1 & -0.6 & 0.3 \\ -0.6 & 1 & -0.3 \\ 0.3 & -0.3 & 1 \end{pmatrix} \mathbf{Z}_{t-1}$$

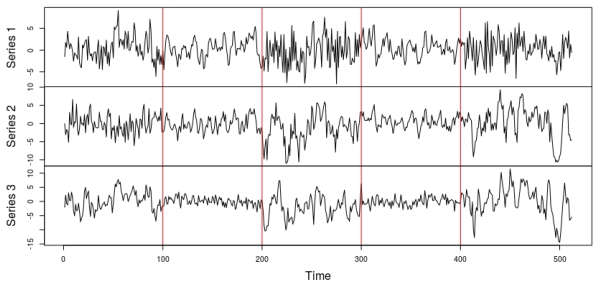
$$\text{Class 3}: \ \mathbf{X}_t = \begin{pmatrix} -0.6 & 0.4 & 0.1 \\ 0.4 & 0.2 & 0.3 \\ 0.1 & 0.3 & 0.5 \end{pmatrix} \mathbf{X}_{t-1} + \mathbf{Z}_t + \begin{pmatrix} 1 & 0.2 & -0.7 \\ 0.2 & 1 & 0.6 \\ -0.7 & 0.6 & 1 \end{pmatrix} \mathbf{Z}_{t-1}$$

where $\mathbf{Z}_t$ and $\mathbf{Z}_{t-1}$ are zero-mean multivariate normal realisations, distributed with class-dependent covariances

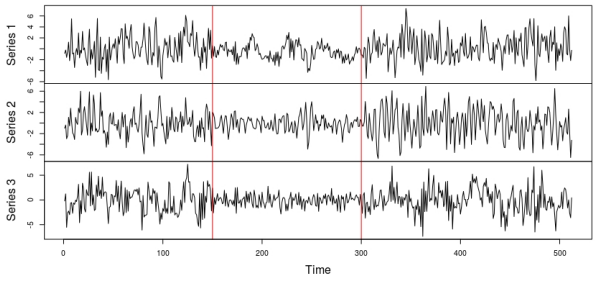
$$\Sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \Sigma_2 = \Sigma_3 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$



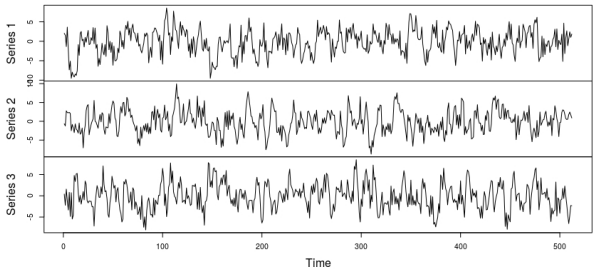
(a) Slow changes, mvLSW process



(b) Rapid changes, vector autoregressive moving average series



(c) Slow changes, vector autoregressive series



(d) Stationary vector moving average series

Fig. 1: Example realisations of generating processes for the different scenarios used in the simulation study. (a), (c) Slowly evolving dependence, class changes at time 150 and 300; (b) Rapidly changing dependence structure, class changes at time 100, 200, 300 and 400; (d) Stationary signal, no changes in the generating coefficient matrices of the process.

for classes 1-3 respectively. The signal switches randomly between each of the three classes at different times depending on whether we are considering slowly or rapidly evolving dependence. This ensures that the intra- and cross-channel dependence of the process changes over time (see Figure 1(b)).

The third example we consider is a time-varying vector autoregressive process with two classes defined by the following coefficient matrices

$$\text{Class 1}: \ \mathbf{X}_t = \begin{pmatrix} 0.3 & 0.2 & -0.2 \\ 0.2 & 0.4 & -0.2 \\ -0.2 & -0.2 & -0.1 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} 0.4 & -0.2 & 0.3 \\ -0.2 & -0.4 & 0.1 \\ 0.3 & 0.1 & -0.2 \end{pmatrix} \mathbf{X}_{t-2} + \epsilon_1$$

$$\text{Class 2}: \ \mathbf{X}_t = \begin{pmatrix} 0.2 & -0.2 & 0 \\ -0.2 & 0.4 & -0.2 \\ 0 & -0.2 & 0.2 \end{pmatrix} \mathbf{X}_{t-1} + \begin{pmatrix} -0.1 & 0 & 0 \\ 0 & -0.4 & 0.3 \\ 0 & 0.3 & 0.3 \end{pmatrix} \mathbf{X}_{t-2} + \epsilon_2$$

where the innovation vectors ϵ_i are zero-mean multivariate normal realisations, distributed with per-class covariances

$$\Sigma_{\epsilon_1} = \begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0.1 \\ 0 & 0.1 & 1 \end{pmatrix}, \qquad \Sigma_{\epsilon_2} = \begin{pmatrix} 5 & 1.2 & 2 \\ 1.2 & 5 & 1.5 \\ 2 & 1.5 & 5 \end{pmatrix}.$$

A realisation of such a process can be seen in Figure 1(c).

Within the simulation study, we compare our method to a range of multivariate imputation approaches, some of which assume that the data follow a (time-constant) multivariate normal distribution. For this reason, we also include a stationary example where the coefficients of the moving average process do not change over time. The coefficient matrices for this process are defined as follows

$$\mathbf{X}_t = \mathbf{Z}_t + \begin{pmatrix} 1 & 0.5 & -0.2 \\ 0.5 & 1 & 0.3 \\ -0.2 & 0.3 & 1 \end{pmatrix} \mathbf{Z}_{t-1} + \begin{pmatrix} 1 & -0.4 & 0.2 \\ -0.4 & 1 & -0.6 \\ 0.2 & -0.6 & 1 \end{pmatrix} \mathbf{Z}_{t-2} + \begin{pmatrix} 1 & 0.1 & -0.5 \\ 0.1 & 1 & -0.3 \\ -0.5 & -0.3 & 1 \end{pmatrix} \mathbf{Z}_{t-3}$$

where $\mathbf{Z}_{t-1}, \mathbf{Z}_{t-2}$ and $\mathbf{Z}_{t-3}$ are zero-mean, multivariate normal realisations, with covariances given by

$$\Sigma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

An example of this stationary process can be found in Figure 1(d).

4.1 Competitor methods

In the simulation study, we compare our method with a number of alternative multivariate imputation approaches, suitable for one single realisation of a multivariate process. Firstly we consider the modified Expectation-Maximization (EM) approach of Junger and de Leon (2015) implemented in the *R* package `mtsdi` (Junger and de Leon, 2018). Within this method, cross-channel correlations are taken into account within the multivariate normal modelling structure and inter-time behaviour is accounted for using a level estimation step in which temporal behaviour of each of the univariate time series is estimated. In the `mtsdi` package, a number of different methods are implemented for estimating the level of the univariate time series. For all simulated examples, we fit a cubic spline to each univariate component where the number of degrees of freedom of each spline is chosen by cross-validation.

Secondly, we compare against the multiple imputation method that combines Expectation-Maximization with bootstrapping proposed by Honaker and King (2010), available in the `Amelia II` *R* package (Honaker et al, 2015), see also Honaker et al (2011) for implementation details. As this is a multiple imputation approach, the method produces m completed datasets which are then averaged to obtain a final imputed dataset. Within the simulations, we choose $m = 5$ as this is suggested to be suitable unless the rate of missingness is very high (Schafer and Olsen, 1998; Honaker et al, 2011). As both of these methods assume that the data can be modelled using a multivariate normal distribution, we would expect them to perform poorly in cases where the underlying time series is highly nonstationary.

As we wish to impute missing values in a multivariate, nonstationary time series, it is important to compare our method to a range of other model-based approaches available in the literature. Specifically, we apply the iterative PCA multiple imputation method of Audigier et al (2016), which is available in the `missMDA` *R* package; see Josse and Husson (2016) and Husson and Josse (2018) for more details. Within the simulations, we apply the regularised iterative PCA algorithm with the number of random initializations set to 10 and the default parameters. In addition, we compare to the non-parametric random forest imputation method (Stekhoven and Bühlmann, 2011) implemented in the *R* package `missForest` (Stekhoven, 2013) where again we use the default parameters.

Since our method involves using a one-step ahead forecasting and backcasting step within the mvLSW framework, as a direct comparison to this we also apply the vector autoregressive prediction approach from the *R* package `MTS` (Tsay, 2015), described in the text Tsay (2013). For each missing index, the approach fits a vector autoregressive process to the available observations and then produces one-step ahead forecasts to predict the missing values. In an attempt to ensure fair comparison with our proposed method, we implement the vector autoregression prediction performing a backward pass of the data and combine the estimates from the forecasting and backcasting steps by averaging, similar to the `mvLSWimpute` method. This is denoted VAR-fb. For completeness, we also include the results from applying one-step ahead forecasting only within this setting (denoted VAR-f). Our proposed `mvLSWimpute` method was implemented using modifications to the code in the `wavethresh` (Nason, 2016) and `mvLSW` (Taylor et al, 2017) *R* packages which perform estimation of multivariate LSW process quantities.

4.2 Evaluation measures

For each of the missingness scenarios and dependence structures described in Section 4, $K = 100$ realizations of the test signals are simulated and four different evaluation measures are considered. In order to assess the performance of the imputation methods, we consider a modified version of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The majority of the simulated examples we consider contain changes in variability over time; this volatility affects the standard RMSE and MAE and makes it difficult to directly compare results for slowly and rapidly evolving dependence. For this reason, we scale the results over time using the true standard deviation.

Let $\sigma_{t,P}$ denote the true standard deviation of the signal at time t for channel P. The calculations of the modified RMSE and MAE only include the predicted values at the missing time points and not the full time series. Let N denote the total number of missing values across all channels and timepoints, $t_{\text{mis}} = \{t_1, t_2, \ldots\}$ contain the timepoints where observations are missing and $P_{\text{mis}} = \{P_{t_1}, P_{t_2}, \ldots\}$ denote the corresponding channels which are affected. Then the modified RMSE

can be defined by

$$\text{RMSE} = \frac{1}{N} \sum_{b \in P_{\text{mis}}} \sum_{s \in t_{\text{mis}}} \frac{(X^b_{s,T} - \hat{X}^b_{s,T})^2}{\sigma^2_{s,b}}, \tag{9}$$

and the modified MAE as

$$\text{MAE} = \frac{1}{N} \sum_{b \in P_{\text{mis}}} \sum_{s \in t_{\text{mis}}} \frac{|X^b_{s,T} - \hat{X}^b_{s,T}|}{\sigma_{s,b}}. \tag{10}$$

In addition to this, we also rank the imputation methods based on the modified RMSE and MAE results. For each of the $K = 100$ realizations carried out, we track which of the imputation methods gives the lowest scaled RMSE and MAE and sum the results.

The imputation results for each of the examples described above (as in Figure 1) can be seen in Tables 1-5. For each example, we consider 10%, 20%, 30% and 40% missingness at random as well as chunks of 20 consecutive time points missing and bursts of missingness up to length 20. A description of how the bursts of missingness are generated can be found in Appendix B. Note that we include the results for the situation where the missingness occurs in all channels simultaneously. In each case, we record the modified RMSE, modified MAE and rankings over the $K = 100$ realizations based on these errors (as described above); the numbers within the brackets represent the standard deviation of these quantities.

When we consider rapidly evolving dependence within the time varying vector autoregressive moving average setting (Table 1), it can be seen that the `mvLSWimpute` method performs well both in terms of the modified error measures and the rankings when percentages of the data are missing at random. On the other hand, the competitor methods which rely on the assumption of an underlying stationary model cannot cope with the changing dependence structure. Note that the addition of a backcasting step into the vector autoregressive prediction approach (VAR-fb) provides an improvement in performance. However, it can be seen that the results weaken when we look at more extreme missingness scenarios such as bursts or chunks missing. When imputing missing values in areas of a signal where consecutive time points are missing, perhaps unsurprisingly all methods =struggle to accurately reconstruct the dependence behaviour within these areas.

We next turn to consider the stationary moving average process (Table 2). It can be seen that the `mvLSWimpute` approach again produces more accurate results followed by both VAR-fb and `mtsdi`. Despite the competitor methods being designed for imputation within stationary time series, the `mvLSWimpute` method outperforms them both in terms of the modified error measures and the rankings.

As expected, for the mvLSW process exhibiting slowly varying spectral structure the `mvLSWimpute` method performs strongly across all evaluation measures. We report the full results for this scenario in Table 3 of Appendix B. For the slowly evolving vector autoregressive series, the `mvLSWimpute` method consistently performs better than the competitors. However, the VAR-fb approach also performs well in this setting due to the underlying model used for the one step ahead predictions being designed for this scenario. In this case, the results produced are comparable to `mvLSWimpute` in terms of the modified RMSE and MAE. Again, we report the full results for this example in Table 4 of Appendix B.

Note that in nearly all cases across the examples, `mvLSWimpute` performs consistently well in terms of the error measures considered, despite some of the competitor methods being designed specifically for imputation within those scenarios. We also observe that the use of a backcasting step within both the mvLSW and VAR imputation methods improves their performance, justifying its inclusion. However, it should be noted here that the VAR-f and VAR-fb methods struggled with fitting models when missingness occurs near the start (respectively end) of the series, due to the number of observations available for parameter estimation; this limits practical use of these techniques in many contexts and reflects similar experiences remarked by other authors, see for example Knight et al (2020). For the scenarios discussed above, missingness in one channel only was also considered and similar performance was observed. The results for the vector autoregressive moving average setting are included in Table 5 of Appendix B.

5 Case study

In the previous section, we evaluated the performance of our multivariate imputation method against a range of alternatives for a variety of simulated scenarios. We now consider an application related to health-related motion analysis.

The reduction in cost and size of accelerometers over the past decade has led to the use of these devices in many areas of scientific research, e.g. sport science (Troiano et al, 2014), engineering component calibration (Yin and Huang, 2014) and computer security (Mayrhofer and Gellersen, 2009). Accelerometer data are particularly useful in cases when the measurement of other meaningful physiological signals is difficult or obtrusive. As such, the have gained popularity in disease-related assessment, for example analysis of sleep disorders (Van Hees et al, 2015), obesity and other cardiometabolic biomarkers (Brocklebank et al,

Table 1: Performance of the imputation methods over $K = 100$ realizations of vector moving average, autoregressive series with rapidly changing dependence structure for different missingness scenarios occurring simultaneously across all channels, using the evaluation measures described in the text. Numbers in brackets represent the standard deviation of estimation errors. Bold numbers indicate best result.

	10%	20%	30%	40%	Chunks	Bursts
Method	Scaled by true standard deviation - RMSE					
mvLSWimpute-f	1.67 (0.13)	1.76 (0.10)	1.80 (0.10)	1.88 (0.10)	2.02 (0.52)	1.81 (0.16)
mvLSWimpute-fb	**1.45 (0.11)**	**1.54 (0.09)**	**1.60 (0.09)**	**1.70 (0.09)**	**2.00 (0.52)**	**1.62 (0.14)**
mtsdi	1.69 (0.12)	1.74 (0.11)	1.78 (0.10)	1.89 (0.11)	2.33 (0.64)	1.82 (0.18)
Amelia	2.40 (0.19)	2.42 (0.14)	2.41 (0.14)	2.44 (0.14)	2.36 (0.51)	2.37 (0.22)
VAR-f	1.75 (0.15)	1.84 (0.11)	1.90 (0.12)	1.98 (0.11)	2.05 (0.53)	1.84 (0.16)
VAR-fb	1.54 (0.12)	1.65 (0.09)	1.71 (0.10)	1.82 (0.09)	2.02 (0.52)	1.68 (0.16)
PCA	2.10 (0.18)	2.12 (0.13)	2.11 (0.13)	2.15 (0.13)	2.07 (0.53)	2.07 (0.20)
Random forest	2.19 (0.21)	2.19 (0.14)	2.21 (0.15)	2.26 (0.18)	2.15 (0.54)	2.16 (0.22)
Method	Scaled by true standard deviation - MAE					
mvLSWimpute-f	1.32 (0.10)	1.39 (0.08)	1.42 (0.08)	1.48 (0.07)	1.61 (0.43)	1.42 (0.13)
mvLSWimpute-fb	**1.14 (0.09)**	**1.22 (0.07)**	**1.26 (0.07)**	**1.34 (0.07)**	**1.59 (0.43)**	**1.27 (0.11)**
mtsdi	1.34 (0.10)	1.37 (0.09)	1.41 (0.07)	1.48 (0.08)	1.86 (0.52)	1.44 (0.14)
Amelia	1.89 (0.15)	1.91 (0.11)	1.91 (0.11)	1.92 (0.11)	1.90 (0.42)	1.87 (0.17)
VAR-f	1.38 (0.11)	1.45 (0.09)	1.50 (0.09)	1.56 (0.08)	1.64 (0.44)	1.46 (0.13)
VAR-fb	1.21 (0.09)	1.30 (0.07)	1.35 (0.08)	1.43 (0.07)	1.62 (0.43)	1.33 (0.13)
PCA	1.65 (0.14)	1.67 (0.10)	1.66 (0.10)	1.68 (0.10)	1.66 (0.44)	1.64 (0.16)
Random forest	1.72 (0.16)	1.72 (0.11)	1.74 (0.12)	1.78 (0.15)	1.73 (0.44)	1.71 (0.17)
Method	Ranking - RMSE					
mvLSWimpute-f	0	0	0	0	19	0
mvLSWimpute-fb	**92**	**95**	**99**	**97**	**27**	**80**
mtsdi	0	0	0	1	3	2
Amelia	0	0	0	0	2	0
VAR-f	0	0	0	0	11	0
VAR-fb	8	5	1	2	17	18
PCA	0	0	0	0	7	0
Random forest	0	0	0	0	14	0
Method	Ranking - MAE					
mvLSWimpute-f	1	0	0	0	20	1
mvLSWimpute-fb	**90**	**98**	**98**	**99**	**28**	**77**
mtsdi	0	0	0	1	5	3
Amelia	0	0	0	0	3	0
VAR-f	0	0	0	0	9	0
VAR-fb	9	2	2	0	15	19
PCA	0	0	0	0	7	0
Random forest	0	0	0	0	13	0

2015; Augustin et al, 2017) and post-diagnosis changes in physiology (Sekine et al, 2004; McDonald et al, 2019).

The benefits of wavelet-based accelerometry analysis has been well-established in the scientific literature, see Bidargaddi et al (2007) or Godfrey et al (2008); since accelerometry data often represents movements at different frequencies over time. It is also acknowledged that accelerometry signals are nonstationary in nature (Preece et al, 2009a,b).

It has been suggested in many studies that attempts to draw conclusions from accelerometer signals can be hindered by the presence of missingness within the data, and missing entries should be replaced with suitable estimates before further analysis can take place, particularly in the case of such data collected from smartphones (Barnett et al, 2018). Within this section, we focus on the problem of imputing missing values in accelerometer data arising from an experiment in which

Table 2: Performance of the imputation methods over $K = 100$ realizations of stationary vector moving average series for different missingness scenarios occurring simultaneously across all channels, using the evaluation measures described in the text. Numbers in brackets represent the standard deviation of estimation errors. Bold numbers indicate best result.

	10%	20%	30%	40%	Chunks	Bursts
Method	Scaled by true standard deviation - RMSE					
mvLSWimpute-f	1.81 (0.12)	1.89 (0.09)	1.94 (0.07)	2.00 (0.07)	2.14 (0.26)	1.97 (0.15)
mvLSWimpute-fb	**1.63 (0.11)**	**1.71 (0.09)**	**1.77 (0.06)**	**1.83 (0.07)**	**2.13 (0.25)**	**1.81 (0.14)**
mtsdi	1.71 (0.11)	1.79 (0.10)	1.89 (0.10)	2.02 (0.12)	2.33 (0.31)	1.94 (0.19)
Amelia	2.36 (0.15)	2.41 (0.11)	2.42 (0.10)	2.42 (0.10)	2.41 (0.26)	2.43 (0.18)
VAR-f	1.83 (0.12)	1.91 (0.10)	1.96 (0.08)	2.03 (0.08)	2.19 (0.26)	2.01 (0.16)
VAR-fb	1.66 (0.11)	1.75 (0.09)	1.81 (0.07)	1.89 (0.08)	2.16 (0.26)	1.85 (0.14)
PCA	2.15 (0.14)	2.21 (0.10)	2.21 (0.08)	2.22 (0.10)	2.20 (0.25)	2.20 (0.16)
Random forest	2.30 (0.19)	2.37 (0.18)	2.35 (0.17)	2.37 (0.16)	2.39 (0.35)	2.35 (0.23)
Method	Scaled by true standard deviation - MAE					
mvLSWimpute-f	1.44 (0.10)	1.51 (0.08)	1.54 (0.06)	1.59 (0.06)	1.73 (0.23)	1.56 (0.11)
mvLSWimpute-fb	**1.29 (0.09)**	**1.36 (0.07)**	**1.40 (0.05)**	**1.45 (0.06)**	**1.72 (0.22)**	**1.43 (0.11)**
mtsdi	1.36 (0.09)	1.43 (0.08)	1.49 (0.07)	1.58 (0.08)	1.86 (0.25)	1.52 (0.14)
Amelia	1.89 (0.13)	1.93 (0.09)	1.93 (0.07)	1.93 (0.09)	1.94 (0.24)	1.94 (0.16)
VAR-f	1.45 (0.10)	1.52 (0.08)	1.56 (0.06)	1.61 (0.07)	1.78 (0.23)	1.59 (0.12)
VAR-fb	1.32 (0.09)	1.39 (0.07)	1.44 (0.05)	1.50 (0.07)	1.76 (0.23)	1.47 (0.11)
PCA	1.72 (0.12)	1.76 (0.09)	1.76 (0.06)	1.77 (0.09)	1.79 (0.23)	1.76 (0.14)
Random forest	1.84 (0.16)	1.90 (0.15)	1.88 (0.14)	1.90 (0.13)	1.95 (0.31)	1.88 (0.20)
Method	Ranking - RMSE					
mvLSWimpute-f	1	0	0	0	26	1
mvLSWimpute-fb	**67**	**79**	**84**	**95**	**31**	**71**
mtsdi	12	8	6	0	11	14
Amelia	0	0	0	0	0	0
VAR-f	0	0	0	0	5	1
VAR-fb	20	13	10	5	13	13
PCA	0	0	0	0	3	0
Random forest	0	0	0	0	11	0
Method	Ranking - MAE					
mvLSWimpute-f	1	0	0	0	25	1
mvLSWimpute	**63**	**78**	**90**	**92**	**30**	**68**
mtsdi	12	9	5	4	10	17
Amelia	0	0	0	0	0	0
VAR-f	0	0	0	0	10	0
VAR-fb	24	13	5	4	10	14
PCA	0	0	0	0	4	0
Random forest	0	0	0	0	11	0

subjects were asked to perform a sequence of predetermined activities or postures. Such gait and postural transition analyses are important in assessing patients' balance after neurological episodes such as strokes (Janssen et al, 2008). In particular, we consider recordings in a period of high activity for the first subject in the HAPT smartphone dataset (Reyes-Ortiz et al, 2016), obtained from the UCI data repository[1](Dua and Graff, 2017). In what follows we first difference the data to remove any trend, as is commonplace prior to secondary analysis (Ahrabian et al, 2017).

The resulting data we analyze is a trivariate signal of length $T = 2048$. Accelerometer data is often blighted by missingness, which usually occurs in bursts or chunks

[1] `https://archive.ics.uci.edu/ml/datasets.php`

across all axes (see e.g. Ae Lee and Gill (2018) for a discussion of missingness patterns and imputation of accelerometry in a different modelling setting). To this end, we induce bursts of missingness of length $l = 30$ with spacings of $d = 350$ according to the procedure in Appendix B. Figure 2(a) shows the accelerometer measurements over time for three sensors, with locations of missingness shown with red triangles (Figure 2(b)).

We apply the mvLSW-based imputation approach with $m = 2$ points considered in the clipped predictor for both the forecasting and backcasting steps. For comparison, we apply the `mtsdi` method and the VAR-fb approach (with $p = 2$) as, of the existing methods, these performed better in the simulation study.

The imputation results for each of the methods are shown in Figure 3 for the first burst of missingness (between time indices $t = 50$ and $t = 80$). Missing values are denoted with back dots, whilst imputed values for `mtsdi` are denoted by red crosses, those for the VAR-fb method are given in blue 'plusses' and those for the `mvLSWimpute` method are denoted by green triangles.

It can be seen that, whilst the imputation results for all three methods are quite similar, the `mvLSWimpute` method produces the most visually reliable estimate for the missing data. In particular, the `mtsdi` method does not track any variation in the volatility of the series, the estimates being essentially constant over time. The VAR-fb method, whilst better, tends to produce some significantly over-/underestimated values for the series. Our `mvLSWimpute` technique is able to strike a balance between accurate imputation and the changing dynamics of the data. We observed similar behaviour of the methods for different axes and gaps.

Next we consider the VAR-fb and `mvLSWimpute` methods with $p = 2$ and $m = 2$ respectively. However, in some settings, accelerometer data will feature longer temporal dependence, see for example the study in Khan et al (2013). When modelling the HAR dataset $m > 2$ for our `mvLSWimpute` approach, we observed similarly accurate results; on the other hand, the VAR-fb method suffered from drastically variable data estimates for differing p, with the error often increasing 100-fold.

The overall aim of an analysis with accelerometer data is to, typically, perform activity recognition and analysis that could be used to provide health interventions. It is therefore important to be able to replace missing values with reasonable estimates which will then allow further analysis to be carried out. Our mvLSW imputation approach can be used as a first step to infill any missing values before attempting to predict activity levels or other secondary analysis tasks of interest.

6 Concluding remarks

In this work, we have introduced a wavelet-based imputation method that can be used to replace missing values within a multivariate, nonstationary time series. We compared the performance of our method against existing imputation approaches using simulated data examples and a smartphone accelerometer dataset. The simulated data examples demonstrate that the use of a backcasting step within imputation can improve the performance of the prediction methods. The case study shows that our method can be used to successfully impute missing values within time series containing both nonstationarity and seasonality, resulting in a more reliable imputation estimate compared to existing approaches.

In practice, we have found that, as with other competitor methods, the performance of our approach suffers when we have extreme scenarios such as chunks of consecutive time points missing or bursts of missingness. An avenue for future research could be to look at ways in which we could improve the imputation results for these cases.

Acknowledgements Wilson gratefully acknowledges financial support from EPSRC and Shell via the STOR-i Centre for Doctoral Training. Eckley also gratefully acknowledges the financial support of EPSRC grant EP/N031938/1.

An R package implementing the imputation method in this article will be publicly available in due course.

A Overview of the `mvLSWimpute` algorithm

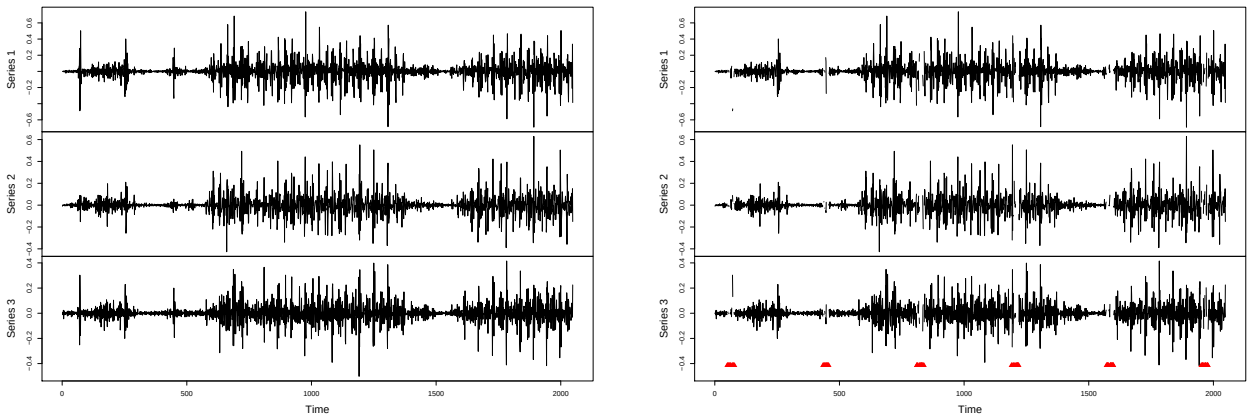
In this section we provide an algorithmic overview of our proposed imputation technique for multivariate nonstationary time series. More specifically, Algorithm 1 outlines the steps involved in one iteration of our `mvLSWimpute` methodology.

B Additional simulation results

In this appendix, we present the results for some of the simulated examples in Section 4. Specifically, we include the imputation results for slowly evolving mvLSW and vector autoregressive series with missingness occurring simultaneously across all channels; these can be found in Tables 3 and 4 respectively. Finally, Table 5 contains the results for vector moving average, autoregressive series with rapidly changing dependence structure, where missingness occurs across one channel only.

Bursts of missingness

Here we describe how we generated bursts of missingness for the examples in the main text. For all the simulations described in Section 4, we generate 5 bursts of missingness that



(a) (b)

Fig. 2: Time series plots of the accelerometer readings for the three sensors (axes) over the same time period: (a) original series; (b) series with induced missingness – red triangles show locations of missing datapoints.

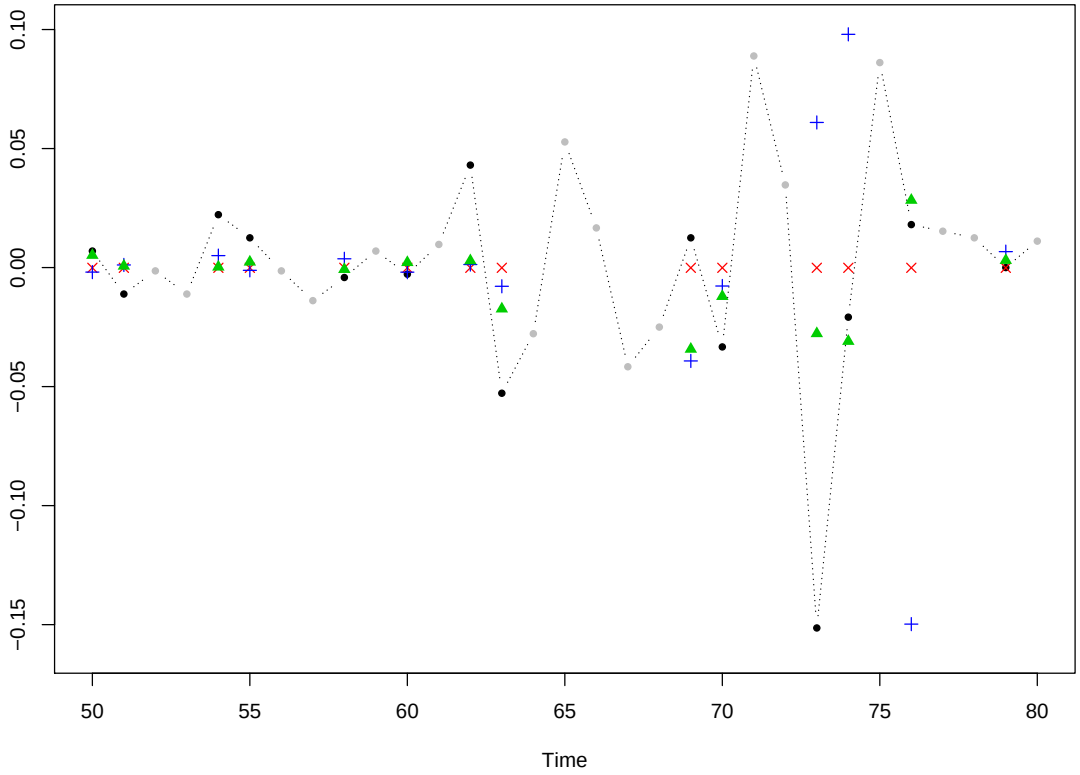


Fig. 3: Imputation results for the first period of missingness in the trivariate accelerometry signal from the case study: missing values denoted by black dots; imputed values for the `mtsdi`, VAR-fb and `mvLSWimpute` methods shown by red crosses (×), blue plusses (+) and green triangles (▲).

have maximum length $l = 20$ and are a minimum of $d = 70$ timepoints apart. First, we randomly sample 50% of the indices without replacement which form the set of candidate locations for missing data. Next, we select the start location

1: **Spectral estimation step**
Estimate the LWS matrix of the signal containing missing data in the following way:
   (a) Estimate the raw wavelet periodogram keeping any missing values intact; any wavelet coefficients affected by the initial missing values will also be missing.
   (b) For each spectra and cross-spectra, linearly interpolate the missing wavelet coefficients by level of the periodogram.
   (c) Recursively apply the wavelet filter equations to the coarsest level of the periodogram that contains information to replace any levels that consist solely of missing values.
   (d) Smooth the periodogram using a running mean smoother and correct using the inverse of the inner product matrix of discrete autocorrelation wavelets $\mathbf{A}$.

2: **Forecasting step**
For each missing index i, forecast the missing value in the following way:
   (a) Consider the spectra obtained in Step 1 from time 1 to time $i-1$.
   (b) Form the local auto- and cross-covariance by substituting estimated spectra from time 1 to $i-1$ into Equations (4) and (5).
   (c) For each channel combination (p, q), solve the prediction equations given in Equation (6) to obtain $\mathbf{b}^{(p,q)}$.
   (d) Use $\mathbf{b}^{(p,q)}$ vectors along with the clipped predictor in Equation (7) to estimate the value of the time series at missing index i.

3: **Backcasting step**
For each missing index i, backcast the missing value in the following way:
   (a) Consider the spectra obtained in Step 1 from time T to time $i+1$.
   (b) Form the local auto and cross-covariance by substituting estimated spectra from time T to $i+1$ into Equations (4) and (5).
   (c) For each channel combination (p, q), solve the prediction equations given in Equation (6) to obtain $\mathbf{b}^{(p,q)}$.
   (d) Use $\mathbf{b}^{(p,q)}$ vectors along with the clipped predictor in Equation (8) to estimate the value of the time series at missing index i.

4: **Averaging step**
Average the estimates of the time series obtained from the forward pass described in Step 2 and the backward pass described in Step 3.

Alg. 1: One iteration of the `mvLSWimpute` algorithm.

for the first burst by sampling an index k between 30 and 70. The start locations s of the bursts are then defined by $k + l(n-1) + d(n-1)$ where $n \in \{1, \ldots, 5\}$ and the end locations are determined by $s + l$. For each of the bursts, the missing indices are chosen by selecting the candidate locations (determined in Step 1) that are between the start and end point of the burst. An example time series containing bursts of missingness can be seen in Figure 4, the filled red triangles represent the locations of the missing values.

References

Ae Lee J, Gill J (2018) Missing value imputation for physical activity data measured by accelerometer. Stat Meth Med Res 27(2):490–506

Ahrabian A, Elsaleh T, Fathy Y, Barnaghi P (2017) Detecting changes in the variance of multi-sensory accelerometer data using MCMC. In: IEEE Sensors, IEEE, pp 1–3

Alvarez FM, Troncoso A, Riquelme JC, Ruiz J (2011) Energy time series forecasting based on pattern sequence similarity. IEEE Trans on Knowl Data Eng 23(8):1230–1243

Audigier V, Husson F, Josse J (2016) Multiple imputation for continuous variables using a Bayesian principal component analysis. J Stat Comput Simul 86(11):2140–2156

Augustin NH, Mattocks C, Faraway JJ, Greven S, Ness AR (2017) Modelling a response as a function of high-frequency count data: The association between physical activity and fat mass. Stat Meth Med Res 26(5):2210–2226

Bar-Joseph Z, Gerber GK, Gifford DK, Jaakkola TS, Simon I (2003) Continuous representations of time-series gene expression data. J of Comput Biol 10(3-4):341–356

Barnett I, Torous J, Staples P, Keshavan M, Onnela JP (2018) Beyond smartphones and sensors: choosing appropriate statistical methods for the analysis of longitudinal data. J Am Med Inform Assoc 25(12):1669–1674

Bidargaddi N, Sarela A, Boyle J, Cheung V, Karunanithi M, Klingbei L, Yelland C, Gray L (2007) Wavelet based approach for posture transition estimation using a waist worn accelerometer. In: 29th Annual International Conference of the IEEE Eng. Med. Biol. Soc., 2007., pp 1884–1887

Bos R, De Waele S, Broersen PM (2002) Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data. IEEE Trans Instrum Meas 51(6):1289–1294

Brocklebank LA, Falconer CL, Page AS, Perry R, Cooper AR (2015) Accelerometer-measured sedentary time and cardiometabolic biomarkers: a systematic review. Prevent Med 76:92–102

Broersen PMT (2006) Automatic spectral analysis with missing data. Digital Signal Process 16(6):754–766

Cao W, Wang D, Li J, Zhou H, Li L, Li Y (2018) Brits: bidirectional recurrent imputation for time series. In: Adv. Neural Info. Process. Syst., pp 6775–6785

Caussinus H (1986) Models and uses of principal component analysis. Multidimension Data Anal 86:149–170

Che Z, Purushotham S, Cho K, Sontag D, Liu Y (2018) Recurrent neural networks for multivariate time series with missing values. Sci Rep 8(1):6085

Cranstoun S, Ombao H, Von Sachs R, Guo W, Litt B (2002) Time-frequency spectral estimation of multichannel EEG using the auto-SLEX method. IEEE Trans Biomed Eng 49(9):988–996

Dahlhaus R (2000) A likelihood approximation for locally stationary processes. Ann Stat 28(6):1762–1794

Dahlhaus R (2012) Locally stationary processes. In: Handbook of Statistics, vol 30, Elsevier, pp 351–413

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B pp 1–38

Doucoure B, Agbossou K, Cardenas A (2016) Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. Renewable Energy 92:202–211

Dua D, Graff C (2017) UCI machine learning repository. URL `http://archive.ics.uci.edu/ml`

Table 3: Performance of the imputation methods over $K = 100$ realizations of mvLSW process with slowly evolving dependence for different missingness scenarios occurring simultaneously across all channels, using the evaluation measures described in the text. Numbers in brackets represent the standard deviation of estimation errors. Bold numbers indicate best result.

	10%	20%	30%	40%	Chunks	Bursts
Method	Scaled by true standard deviation - RMSE					
mvLSWimpute-f	0.88 (0.06)	0.90 (0.05)	0.92 (0.04)	0.94 (0.04)	1.00 (0.14)	0.92 (0.08)
mvLSWimpute-fb	**0.81 (0.06)**	**0.85 (0.04)**	**0.87 (0.04)**	**0.89 (0.04)**	**1.00 (0.14)**	**0.88 (0.08)**
mtsdi	0.93 (0.06)	0.95 (0.05)	0.97 (0.04)	1.01 (0.07)	1.15 (0.22)	0.97 (0.09)
Amelia	1.10 (0.07)	1.10 (0.06)	1.11 (0.05)	1.09 (0.04)	1.11 (0.15)	1.09 (0.09)
VAR-f	0.92 (0.06)	0.94 (0.05)	0.95 (0.04)	0.97 (0.04)	1.03 (0.15)	0.95 (0.08)
VAR-fb	0.86 (0.06)	0.88 (0.05)	0.90 (0.04)	0.92 (0.04)	1.02 (0.15)	0.89 (0.08)
PCA	1.00 (0.07)	1.00 (0.05)	1.00 (0.04)	0.99 (0.04)	1.02 (0.14)	0.99 (0.08)
Random forest	1.06 (0.09)	1.06 (0.09)	1.06 (0.07)	1.04 (0.05)	1.08 (0.15)	1.04 (0.09)
Method	Scaled by true standard deviation - MAE					
mvLSWimpute-f	0.70 (0.05)	0.72 (0.04)	0.73 (0.03)	0.75 (0.03)	0.81 (0.12)	0.73 (0.06)
mvLSWimpute-fb	**0.65 (0.05)**	**0.67 (0.04)**	**0.69 (0.03)**	**0.71 (0.03)**	**0.80 (0.12)**	**0.70 (0.06)**
mtsdi	0.75 (0.06)	0.76 (0.04)	0.77 (0.03)	0.80 (0.04)	0.93 (0.18)	0.77 (0.07)
Amelia	0.88 (0.06)	0.88 (0.05)	0.88 (0.04)	0.87 (0.03)	0.89 (0.12)	0.87 (0.07)
VAR-f	0.73 (0.05)	0.75 (0.04)	0.76 (0.03)	0.77 (0.03)	0.83 (0.13)	0.75 (0.06)
VAR-fb	0.68 (0.05)	0.70 (0.04)	0.71 (0.03)	0.73 (0.03)	0.82 (0.13)	0.71 (0.06)
PCA	0.80 (0.06)	0.80 (0.04)	0.81 (0.04)	0.79 (0.03)	0.82 (0.12)	0.79 (0.06)
Random forest	0.85 (0.07)	0.85 (0.07)	0.85 (0.05)	0.83 (0.04)	0.87 (0.13)	0.84 (0.07)
Method	Ranking - RMSE					
mvLSWimpute-f	3	1	0	0	29	10
mvLSWimpute-fb	**90**	**91**	**97**	**98**	**32**	**67**
mtsdi	1	0	0	0	7	4
Amelia	0	0	0	0	2	0
VAR-f	0	0	0	0	11	0
VAR-fb	6	8	3	2	12	19
PCA	0	0	0	0	3	0
Random forest	0	0	0	0	4	0
Method	Ranking - MAE					
mvLSWimpute-f	5	2	1	1	24	10
mvLSWimpute-fb	**88**	**90**	**92**	**89**	**31**	**66**
mtsdi	0	0	0	0	10	4
Amelia	0	0	0	0	3	0
VAR-f	1	0	0	0	11	0
VAR-fb	6	8	7	10	11	20
PCA	0	0	0	0	2	0
Random forest	0	0	0	0	8	0

Eckley IA, Nason GP (2005) Efficient computation of the discrete autocorrelation wavelet inner product matrix. Stat Comput 15(2):83–92

Ford BL (1983) An overview of hot-deck procedures. Incompl Data Sample Surv 2 (Part IV):185–207

Fryzlewicz P, Van Bellegem S, Von Sachs R (2003) Forecasting non-stationary time series by wavelet process modelling. Ann Inst Stat Math 55(4):737–764

Fryzlewicz P, Sapatinas T, Subba Rao S (2006) A Haar–Fisz technique for locally stationary volatility estimation. Biometrika 93(3):687–704

Ghahramani Z, Jordan MI (1994) Supervised learning from incomplete data via an EM approach. In: Adv. Neural Info. Process. Syst., pp 120–127

Godfrey A, Conway R, Leonard M, Meagher D, OLaighin GM (2008) Motion analysis in delirium: A wavelet based approach for sub classification. In: 30th Ann. Intern. Conf. IEEE Eng. Med. Biol. Soc., 2008., pp 3574–3577

Gott AN, Eckley IA (2013) A note on the effect of wavelet choice on the estimation of the evolutionary wavelet spectrum. Commun Stat – Simul Computat 42(2):393–406

Table 4: Performance of the imputation methods over $K = 100$ realizations of vector autoregressive series with slowly evolving dependence structure for different missingness scenarios occurring simultaneously across all channels, using the evaluation measures described in the text. Numbers in brackets represent the standard deviation of estimation errors. Bold numbers indicate best result.

	10%	20%	30%	40%	Chunks	Bursts
Method	Scaled by true standard deviation - RMSE					
mvLSWimpute-f	1.10 (0.08)	1.13 (0.05)	1.14 (0.05)	1.16 (0.04)	1.14 (0.13)	1.14 (0.08)
mvLSWimpute-fb	**1.04 (0.07)**	**1.07 (0.05)**	**1.08 (0.05)**	**1.11 (0.04)**	**1.13 (0.13)**	**1.09 (0.07)**
mtsdi	1.14 (0.08)	1.17 (0.05)	1.18 (0.05)	1.19 (0.05)	1.21 (0.18)	1.17 (0.09)
Amelia	1.34 (0.09)	1.36 (0.07)	1.36 (0.07)	1.37 (0.05)	1.35 (0.25)	1.34 (0.09)
VAR-f	1.11 (0.09)	1.15 (0.05)	1.15 (0.06)	1.17 (0.04)	1.15 (0.14)	1.14 (0.08)
VAR-fb	1.06 (0.08)	1.09 (0.05)	1.10 (0.05)	1.13 (0.04)	1.15 (0.20)	1.10 (0.08)
PCA	1.16 (0.08)	1.17 (0.05)	1.16 (0.06)	1.17 (0.04)	1.17 (0.18)	1.17 (0.08)
Random forest	1.24 (0.12)	1.26 (0.12)	1.27 (0.11)	1.27 (0.10)	1.21 (0.17)	1.25 (0.12)
Method	Scaled by true standard deviation - MAE					
mvLSWimpute-f	0.88 (0.07)	0.90 (0.04)	0.91 (0.04)	0.92 (0.03)	0.94 (0.14)	0.90 (0.06)
mvLSWimpute-fb	**0.82 (0.06)**	**0.85 (0.04)**	**0.86 (0.04)**	**0.88 (0.03)**	**0.93 (0.14)**	**0.86 (0.07)**
mtsdi	0.91 (0.06)	0.93 (0.04)	0.94 (0.04)	0.95 (0.04)	0.97 (0.14)	0.93 (0.07)
Amelia	1.06 (0.07)	1.08 (0.05)	1.08 (0.04)	1.08 (0.04)	1.08 (0.20)	1.06 (0.08)
VAR-f	0.88 (0.07)	0.91 (0.05)	0.92 (0.05)	0.93 (0.03)	0.95 (0.15)	0.91 (0.06)
VAR-fb	0.84 (0.06)	0.87 (0.04)	0.87 (0.04)	0.89 (0.03)	0.94 (0.15)	0.87 (0.07)
PCA	0.92 (0.06)	0.94 (0.04)	0.93 (0.04)	0.94 (0.03)	0.95 (0.15)	0.93 (0.07)
Random forest	0.99 (0.10)	1.00 (0.09)	1.01 (0.08)	1.01 (0.08)	0.99 (0.20)	0.99 (0.10)
Method	Ranking - RMSE					
mvLSWimpute-f	0	0	0	0	**22**	2
mvLSWimpute-fb	**68**	**66**	**72**	**73**	17	**57**
mtsdi	0	0	0	0	12	1
Amelia	0	0	0	0	2	0
VAR-f	0	0	0	0	14	1
VAR-fb	32	34	28	27	21	38
PCA	0	0	0	0	3	1
Random forest	0	0	0	0	9	0
Method	Ranking - MAE					
mvLSWimpute-f	1	0	0	1	**22**	3
mvLSWimpute-fb	**63**	**64**	**67**	**67**	17	**55**
mtsdi	0	0	0	0	11	0
Amelia	0	0	0	0	2	0
VAR-f	1	0	0	0	14	2
VAR-fb	35	36	33	32	20	39
PCA	0	0	0	0	3	1
Random forest	0	0	0	0	11	0

Graham J (2009) Missing data analysis: Making it work in the real world. Ann Rev Psychol 60:549–576

Honaker J, King G (2010) What to do about missing values in time–series cross–section data. Am J Polit Sci 54(2):561–581

Honaker J, King G, Blackwell M (2011) Amelia II: A Program for Missing Data. J Stat Softw 45(7):1–47

Honaker J, King G, Blackwell M (2015) Amelia: A Program for Missing Data. URL `https://CRAN.R-project.org/package=Amelia`

Husson F, Josse J (2018) missMDA: Handling Missing Values with Multivariate Data Analysis. URL `https://CRAN.R-project.org/package=missMDA`

Janssen WGM, Külchü DG, Horemans HLD, Stam HJ, Bussmann JBJ (2008) Sensitivity of accelerometry to assess balance control during sit-to-stand movement. IEEE Trans Neur Syst Rehab Eng 16(5):479–484

Jones RH (1980) Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics 22(3):389–395

Table 5: Performance of the imputation methods over $K = 100$ realizations of vector moving average, autoregressive series with rapidly changing dependence structure for different missingness scenarios occurring across one channel, using the evaluation measures described in the text. Numbers in brackets represent the standard deviation of estimation errors. Bold numbers indicate best result.

	10%	20%	30%	40%	Chunks	Bursts
Method	Scaled by true standard deviation - RMSE					
mvLSWimpute-f	1.63 (0.20)	1.64 (0.14)	1.67 (0.13)	1.70 (0.11)	1.91 (0.56)	1.72 (0.21)
mvLSWimpute-fb	**1.40 (0.18)**	**1.44 (0.11)**	**1.45 (0.11)**	**1.49 (0.09)**	**1.79 (0.48)**	**1.55 (0.18)**
mtsdi	1.70 (0.21)	1.72 (0.13)	1.73 (0.12)	1.75 (0.11)	2.06 (0.58)	1.87 (0.23)
Amelia	2.29 (0.28)	2.24 (0.19)	2.25 (0.18)	2.29 (0.16)	2.24 (0.55)	2.21 (0.31)
VAR-f	1.73 (0.17)	1.72 (0.13)	1.74 (0.13)	1.79 (0.12)	1.96 (0.52)	1.80 (0.22)
VAR-fb	1.50 (0.16)	1.52 (0.11)	1.55 (0.11)	1.59 (0.10)	1.82 (0.49)	1.60 (0.17)
PCA	2.02 (0.22)	1.99 (0.16)	2.03 (0.16)	2.05 (0.16)	2.02 (0.50)	2.00 (0.29)
Random forest	2.07 (0.23)	2.07 (0.17)	2.09 (0.16)	2.14 (0.15)	2.13 (0.47)	2.09 (0.30)
Method	Scaled by true standard deviation - MAE					
mvLSWimpute-f	1.30 (0.16)	1.30 (0.11)	1.32 (0.10)	1.36 (0.09)	1.49 (0.45)	1.36 (0.17)
mvLSWimpute-fb	**1.12 (0.15)**	**1.14 (0.09)**	**1.15 (0.09)**	**1.18 (0.07)**	**1.45 (0.41)**	**1.23 (0.14)**
mtsdi	1.36 (0.18)	1.36 (0.12)	1.37 (0.10)	1.39 (0.09)	1.64 (0.50)	1.49 (0.18)
Amelia	1.81 (0.22)	1.79 (0.17)	1.79 (0.15)	1.82 (0.12)	1.85 (0.46)	1.76 (0.25)
VAR-f	1.39 (0.15)	1.39 (0.13)	1.40 (0.11)	1.41 (0.09)	1.61 (0.45)	1.42 (0.16)
VAR-fb	1.19 (0.14)	1.21 (0.10)	1.22 (0.10)	1.24 (0.09)	1.47 (0.43)	1.28 (0.13)
PCA	1.60 (0.18)	1.59 (0.16)	1.62 (0.12)	1.65 (0.12)	1.58 (0.41)	1.60 (0.23)
Random forest	1.64 (0.19)	1.64 (0.13)	1.66 (0.13)	1.69 (0.12)	1.67 (0.39)	1.65 (0.24)
Method	Ranking - RMSE					
mvLSWimpute-f	3	1	0	0	16	4
mvLSWimpute-fb	**76**	**86**	**87**	**93**	22	**65**
mtsdi	3	2	1	0	6	2
Amelia	0	0	0	0	3	0
VAR-f	0	0	0	0	9	1
VAR-fb	18	11	12	7	**33**	28
PCA	0	0	0	0	8	0
Random forest	0	0	0	0	3	0
Method	Ranking - MAE					
mvLSWimpute-f	5	1	0	0	19	8
mvLSWimpute-fb	**74**	**84**	**86**	**93**	21	**62**
mtsdi	0	1	1	0	6	2
Amelia	0	0	0	0	4	0
VAR-f	0	0	0	0	13	2
VAR-fb	21	14	13	7	**26**	26
PCA	0	0	0	0	8	0
Random forest	0	0	0	0	3	0

Josse J, Husson F (2016) missMDA: A package for handling missing values in multivariate data analysis. J Stat Softw 70(1):1–31, DOI 10.18637/jss.v070.i01, URL `https://www.jstatsoft.org/v070/i01`

Junger W, de Leon AP (2018) mtsdi: Multivariate Time Series Data Imputation. URL `https://CRAN.R-project.org/package=mtsdi`

Junger WL, de Leon AP (2015) Imputation of missing data in time series for air pollutants. Atmos Environ 102:96–104

Khan AM, Siddiqi MH, Lee SW (2013) Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones. Sensors 13(10):13,099–13,122

Killick R, Eckley IA, Jonathan P (2013) A wavelet-based approach for detecting changes in second order structure within nonstationary time series. Electron J Stat 7:1167–1183

Knight MI, Nunes MA, Nason G (2012) Spectral estimation for locally stationary time series with missing observations. Stat Comput 22(4):877–895

Knight MI, Leeming KA, Nason GP, Nunes MA (2020) Generalised network autoregressive processes and the GNAR

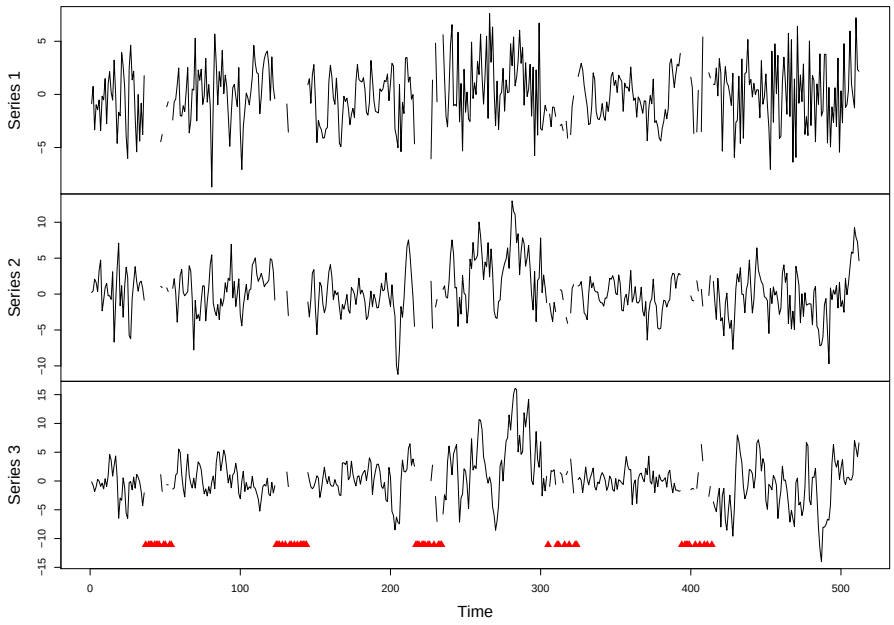


Fig. 4: Example time series containing bursts of missingness; missing time locations are represented by the filled red triangles.

package. J Stat Soft 96(5):1–36
Laguna P, Moody GB, Mark RG (1998) Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. IEEE Trans Biomed Eng 45(6):698–715
Little R, Rubin DB (2002) Statistical Analysis with Missing Data (Second Edition). Wiley
Lobato F, Sales C, Araujo I, Tadaiesky V, Dias L, Ramos L, Santana A (2015) Multi-objective genetic algorithm for missing data imputation. Pattern Recognit Lett 68:126–131
Lomb NR (1976) Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci 39(2):447–462
Luo Y, Cai X, Zhang Y, Xu J, et al (2018) Multivariate time series imputation with generative adversarial networks. In: Adv. Neural Info. Process. Syst., pp 1596–1607
Mayrhofer R, Gellersen H (2009) Shake well before use: Intuitive and secure pairing of mobile devices. IEEE Trans Mob Comput 8(6):792–806
McDonald L, Oguz M, Carroll R, Thakkar P, Yang F, Dhalwani N, Cox A, Merinopoulou E, Malcolm B, Mehmud F, et al (2019) Comparison of accelerometer-derived physical activity levels between individuals with and without cancer: a UK Biobank study. Fut Oncol 15(33):3763–3774
Moritz S, Bartz-Beielstein T (2017) imputeTS: Time Series Missing Value Imputation in R. R J 9(1):207–218
Nason GP (2016) wavethresh: Wavelets Statistics and Transforms. URL `https://CRAN.R-project.org/package=wavethresh`
Nason GP, Von Sachs R, Kroisandt G (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. J R Stat Soc B 62(2):271–292
Ombao H, Von Sachs R, Guo W (2005) SLEX analysis of multivariate nonstationary time series. J Am Stat Assoc 100(470):519–531
Park T, Eckley IA, Ombao HC (2014) Estimating time-evolving partial coherence between signals via multivariate locally stationary wavelet processes. IEEE Trans Signal Process 62(20):5240–5250
Pratama I, Permanasari AE, Ardiyanto I, Indrayani R (2016) A review of missing values handling methods on time-series data. In: 2016 Int. Conf. Info. Technol. Syst. Innovation (ICITSI), IEEE, pp 1–6
Preece S, Goulermas J, Kenney L, Howard D (2009a) A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng 53(3):871–879
Preece SJ, Goulermas JY, Kenney LPJ, Howard D, Meijer K, Crompton R (2009b) Activity identification using body-mounted sensors—a review of classification techniques. Physiol Meas 30(4):R1
Reyes-Ortiz JL, Oneto L, Samà A, Parra X, Anguita D (2016) Transition-aware human activity recognition using smartphones. Neurocomput 171:754–767
Rubin DB (1987) Multiple Imputation for Nonresponse in Surveys (Wiley Series in Probability and Statistics). Wiley
Scargle JD (1982) Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. Astrophys J 263:835–853
Schafer JL, Olsen MK (1998) Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Multivar Behav Res 33(4):545–571
Sekine M, Akay M, Tamura T, Higashi Y, Fujimoto T (2004) Investigating body motion patterns in patients with Parkinson's disease using matching pursuit algorithm. Med Biol Eng Comput 42(1):30–36
Sridevi S, Rajaram S, Parthiban C, SibiArasan S, Swadhikar C (2011) Imputation for the analysis of missing values and

prediction of time series data. In: 2011 Int. Conf. Recent Trends Info. Technol. (ICRTIT), IEEE, pp 1158–1163

Stekhoven DJ (2013) missForest: Nonparametric Missing Value Imputation using Random Forest. URL `https://CRAN.R-project.org/package=missForest`

Stekhoven DJ, Bühlmann P (2011) MissForest–nonparametric missing value imputation for mixed-type data. Bioinf 28(1):112–118

Stărică C, Granger C (2005) Nonstationarities in stock returns. Rev Econ Stat 87(3):503–522

Tang J, Zhang G, Wang Y, Wang H, Liu F (2015) A hybrid approach to integrate fuzzy c-means based imputation method with genetic algorithm for missing traffic volume data estimation. Transp Res Part C: Emerging Technol 51:29–40

Taylor S, Park T, Eckley IA, Killick R (2017) mvLSW: Multivariate Locally Stationary Wavelet Process Estimation. URL `https://CRAN.R-project.org/package=mvLSW`

Taylor S, Park T, Eckley IA (2019) Multivariate locally stationary wavelet analysis with the mvLSW R package. J Stat Softw 90(11):1–19, DOI 10.18637/jss.v090.i11, URL `https://www.jstatsoft.org/v090/i11`

Taylor SJ (2007) Modelling Financial Time Series (Second Edition). World Scientific Publishing

Trindade AA (2003) Implementing modified Burg algorithms in multivariate subset autoregressive modeling. J Stat Softw 8(1):1–68

Troiano RP, McClain JJ, Brychta RJ, Chen KY (2014) Evolution of accelerometer methods for physical activity research. Brit J Sports Med 48(13):1019–1023

Tsay RS (2013) Multivariate Time Series Analysis: with R and Financial Applications. John Wiley & Sons

Tsay RS (2015) MTS: All-Purpose Toolkit for Analysing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models. URL `https://CRAN.R-project.org/package=MTS`

Van Dongen HPA, Olofsen E, Van Hartevelt JH, Kruyt EW (1999) A procedure of multiple period searching in unequally spaced time-series with the Lomb–Scargle method. Biol Rhythm Res 30(2):149–177

Van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, Abell JG, Kivimäki M, Trenell MI, Singh-Manoux A (2015) A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PloS one 10(11)

Wen L, Cui W, Levine AM, Bradt HV (1999) Orbital modulation of X-rays from Cygnus X-1 in its hard and soft states. Astrophys J 525(2):968–977

Wu SF, Chang CY, Lee SJ (2015) Time series forecasting with missing values. In: 2015 1st Int. Conf. Ind. Networks Intell. Syst. (INISCom), IEEE, pp 151–156

Yin S, Huang Z (2014) Performance monitoring for vehicle suspension system via fuzzy positivistic c-means clustering based on accelerometer measurements. IEEE Trans Mechatron 20(5):2613–2620

Yoon J, Jordon J, Van Der Schaar M (2018) Gain: Missing data imputation using generative adversarial nets, arXiv preprint arXiv:1806.02920