



HAL
open science

Dynamic Mode Decomposition for Univariate Time Series: Analysing Trends and Forecasting

Santosh Tirunagari, Samaneh Kouchaki, Norman Poh, Mirosław Bober, David Windridge

► **To cite this version:**

Santosh Tirunagari, Samaneh Kouchaki, Norman Poh, Mirosław Bober, David Windridge. Dynamic Mode Decomposition for Univariate Time Series: Analysing Trends and Forecasting. 2017. hal-01463744

HAL Id: hal-01463744

<https://hal.science/hal-01463744>

Preprint submitted on 9 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dynamic Mode Decomposition for Univariate Time Series: Analysing Trends and Forecasting

Santosh Tirunagari^{*‡}, Samaneh Kouchaki[†], Norman Poh^{*}, Miroslaw Bober[‡], and David Windridge[§]

^{*} Department of Computer Science.

[‡] Center for Vision, Speech and Signal Processing.

^{*‡} University of Surrey, Guildford, Surrey, United Kingdom GU2 7XH.

[†] Department of Evolutionary and Genomic Sciences, The University of Manchester, UK.

[§] Department of Computer Science, Middlesex University, The Burroughs, Hendon, London NW4 4BT.

{santosh.tirunagari, n.poh, m.bober}@surrey.ac.uk, samaneh.kouchaki@manchester.ac.uk, and d.windridge@mdx.ac.uk.

Abstract—This study introduces the method of Dynamic Mode Decomposition (DMD) for analysing univariate time series by forecasting as well as extracting trends and frequencies. The key advantage of DMD is its data-driven nature which does not rely on any prior assumptions (like Singular Spectrum Analysis (SSA)) except the inherent dynamics which are captured over time. Indeed, this study will show that the DMD eigenvalues with frequencies that are closer to the origin in the complex plane capture the trends in the time series. Moreover, the temporal evolution of the DMD modes, which is preserved via the Vandermonde matrix, can be used to reconstruct the desired components and perform forecasting at the same time. The results at various noise levels on simulated data suggests that DMD is a promising approach to modelling a time series with a noisy structure. Although these properties are not new in the DMD literature, the novel contributions of this paper are in making the method of DMD work for univariate time series through a four staged pipeline. Thus, this is the first work that shows DMD can be used for modelling, predicting, and forecasting a univariate time series.

Keywords—Time series modelling, dynamic mode decomposition, DMD, forecast, prediction, trends, noise.

I. INTRODUCTION

Time series analysis includes extracting meaningful information like filtering trends, and forecasting future values [1], [2]. It has applications in many areas, such as quantitative finance and biomedical engineering [3], [4], [5]. This signal processing problem is often tackled by decomposing the data into several components in the frequency domain, thus decoupling signals or trends from noise. To this end, matrix decomposition techniques such as singular spectrum analysis (SSA) [3], [6] is often used. Analogous to SSA, this paper presents a data-driven approach, Dynamic Mode Decomposition (DMD).

The method of DMD was introduced originally in the area of Computational Fluid Dynamics (CFD) for extracting coherent structures from spatio-temporal complex fluid flow data [7]. DMD takes in time series data and computes a set of modes, each of which is associated with a complex eigenvalue. DMD analysis is closely associated with spectral analysis of the Koopman operator, which provides linear but infinite-dimensional representation of nonlinear dynamical systems [8]. Therefore, by using DMD a nonlinear system could be described by a superposition of modes whose dynamics are governed by the eigenvalues. The key advantage of DMD

is its data-driven nature which does not rely on any prior assumptions except the inherent dynamics which are observed over time. Its capability for extracting relevant modes from complex fluid flows [9], [10], [11], [12] has seen significant application across multiple fields, including computer vision, for robustly separating video frames into a background model and multiple foreground objects [13], [14], [15]. In biometrics for detecting spoofed samples [16], [17] from the live ones. In robotics and neuroscience, DMD has been used to estimate perturbation in human robot interactions [18] and to extract coherent patterns in large-scale neural recordings [19].

All the aforementioned DMD applications consider a data matrix as an input. For instance, to analyse an image sequence (a video) using DMD, the input data is converted into column vectors, each of which representing a single frame of image. However, being a fundamentally multidimensional method, DMD has never been applied to a univariate time series. Therefore, our contribution is to introduce the method of DMD through a four staged approach for decomposing a univariate time series into a number of interpretable elements in terms of noise, trend and harmonics. Moreover, time series forecasting using DMD is also presented in this study.

The methodological pipeline consists of four stages similar to that of SSA. However, it is different from SSA in the formulation. The beauty of DMD is that it captures frequency components that describe the growth/decay and oscillations in the data whilst at the same time it can forecast time series during the reconstruction stage.

In the remainder of this paper, the DMD methodology is described in Section II and SSA is described in Section III. Section IV presents the datasets used in this study. Experiments and results are presented in Section V. Finally, conclusions with discussions are drawn in Section VI.

II. METHODOLOGY

This section formulates DMD for univariate time series analysis. Our algorithm consists of four stages: (1) embedding, (2) dynamic mode decomposition, (3) reconstruction and forecasting, and (4) diagonal averaging. Each step is explained in more detail in the following subsections:

A. Embedding

Embedding can be considered as a mapping that converts a univariate time series data $\mathbf{x} = [x_1, x_2, \dots, x_T]$ of length

T into a multidimensional series $\mathbf{X} = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_L]$, of size $K \times L$. The vector $\bar{\mathbf{x}}_n = [x_n, x_{n+1}, \dots, x_{n+K-1}]^T \in \mathbb{R}^K$ is called a K -lagged vector, and $L \in [2 \leq L \leq \frac{T}{2}]$ is the length of window whereas $K = T - L + 1$, is the number of overlapping segments. Thus, the time series is decomposed into K overlapping segments of length L . This produces a trajectory matrix $\mathbf{X} \in \mathbb{R}^{K \times L}$. Unfolding the matrix \mathbf{X} with elements $K \times L$, we have:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_L \\ x_2 & x_3 & \dots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_K & x_{K+1} & \dots & x_{K+L-1} \end{pmatrix}. \quad (1)$$

Note that \mathbf{X} is a Hankel matrix, because its anti-diagonal elements $(m+n) = \text{const}$, are equal. Where, m, n are the row and column indices of the Hankel matrix¹

B. Dynamic Mode Decomposition

The K -lagged vectors $\bar{\mathbf{x}}_n \in \mathbb{R}^K$ are evenly spaced by $\delta t = K$. Therefore, it can be justified that a mapping \mathbf{A} exists between consecutive lagged vectors. Formally, we can say that these vectors span the krylov subspace [20], [21], [22], i.e.,

$$\mathbf{P} = [\bar{\mathbf{x}}_1, \mathbf{A}\bar{\mathbf{x}}_1, \mathbf{A}^2\bar{\mathbf{x}}_1, \mathbf{A}^3\bar{\mathbf{x}}_1, \dots, \mathbf{A}^{N-1}\bar{\mathbf{x}}_1] \quad (2)$$

for an unknown matrix \mathbf{A} , this formulation (Eq. (2)), essentially reconstructs the data $\tilde{\mathbf{X}}$ at any instance $l \in \{1, \dots, L-1, \dots, L-1+F\}$ after the initial vector $\bar{\mathbf{x}}_1$. Here L is the number of columns in the trajectory matrix \mathbf{X} and F is the number of future points. This will be discussed in Section II-C. Eq. (2) can be rewritten by introducing two data matrices $\mathbf{P}_2 \equiv [\bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \dots, \bar{\mathbf{x}}_L]$ and $\mathbf{P}_1 \equiv [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_{L-1}]$ in the following way:

$$\mathbf{P}_2 = \mathbf{A}\mathbf{P}_1. \quad (3)$$

The mapping matrix (also known as Koopman operator) \mathbf{A} is responsible for capturing the dynamics within these vectors. The sizes of the matrices \mathbf{P}_2 and \mathbf{P}_1 are $K \times (L-1)$ each. Therefore, the size of the unknown matrix \mathbf{A} would be $K \times K$. For $(K \gg L)$, unfortunately, solving for \mathbf{A} will be computationally very expensive. Moreover, our problem of interest is to find the temporal evolution which are essentially captured in terms of column vectors. Therefore, it is necessary to represent these column vectors as a linear combination. Our assumption that these vectors form a Krylov span, allows us to introduce \mathbf{H} (as a starting point of the standard Arnoldi iteration),

$$\mathbf{A}\mathbf{P}_1 \approx \mathbf{P}_1\mathbf{H}. \quad (4)$$

Here, \mathbf{H} is a companion matrix, also known as a shifting matrix, that simply shifts vectors 1 through $L-2$. In essence, \mathbf{H} approximates the last vector $L-1$ by linearly combining the previous $L-2$ vectors, i.e., $\bar{\mathbf{x}}_{L-1} = c_0\bar{\mathbf{x}}_1 + \dots + c_{L-2}\bar{\mathbf{x}}_{L-2}$. \mathbf{H} requires the storage of $(L-1) \times (L-1)$ data matrix and is significantly fewer in dimension than \mathbf{A} .

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & \dots & 0 & c_0 \\ 1 & 0 & \dots & 0 & c_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & c_{L-2} \\ 0 & 0 & 0 & 1 & c_{L-1} \end{pmatrix}. \quad (5)$$

Thus, for the last K lagged vector $\bar{\mathbf{x}}_{L-1}$, where L is significantly fewer in dimensionality of \mathbf{A} ($K \times K$), one can write \mathbf{P}_2 as a linear combination of the previous vectors. Consistent with Eqs. (3) and (4), we then have:

$$\mathbf{P}_2 = \mathbf{P}_1\mathbf{H} + \mathbf{r}.e_{L-1}^*, \quad (6)$$

where, \mathbf{r} is the residual error, $(.)^*$ is the conjugate transpose operator, e_{L-1} is the $(L-1)$ unit vector. From Eqs. (4) and (6), we have $\mathbf{A}\mathbf{P}_1 \approx \mathbf{P}_2 \approx \mathbf{P}_1\mathbf{H}$. Although we can solve \mathbf{H} using the LU [16] or the QR decomposition [11], [23], a more robust solution can be achieved using SVD [11] on \mathbf{P}_1 in Eq. (3), to obtain \mathbf{U} , Σ and \mathbf{V}^* matrices.

$$\mathbf{P}_2 = \mathbf{A}\mathbf{U}\Sigma\mathbf{V}^*. \quad (7)$$

$\therefore \mathbf{P}_1 = \sum_{i=1}^d \Sigma_i \mathbf{u}_i \mathbf{v}_i^*$, here, Σ_i is the i^{th} singular value of \mathbf{P}_1 , \mathbf{u}_i and \mathbf{v}_i^* are the corresponding left and right singular vectors respectively; and d is the total number of singular values. Rearranging Eq. (7), we obtain the the full-rank matrix \mathbf{A} ,

$$\mathbf{A} = \mathbf{P}_2\mathbf{V}\Sigma^{-1}\mathbf{U}^*. \quad (8)$$

Since the eigenvalue analysis is agnostic to any linear projection; so solving the eigen problem of $\tilde{\mathbf{H}}$ is easier than that of solving for \mathbf{A} directly. Moreover, the associated eigenvectors of $\tilde{\mathbf{H}}$ provide the coefficients for the linear combination that is necessary to express the dynamics within the time series basis.

$$\tilde{\mathbf{H}}\omega = \sigma\omega, \quad (9)$$

where, ω are the eigenvectors and σ a diagonal matrix containing the corresponding eigenvalues of $\tilde{\mathbf{H}}$ matrix. The eigenvalues of $\tilde{\mathbf{H}}$ approximate some of the eigenvalues of the full system \mathbf{A} [13], we then have:

$$\begin{aligned} \mathbf{A}\mathbf{U} &\approx \mathbf{U}\tilde{\mathbf{H}}, \\ \mathbf{A}\mathbf{U} &\approx \mathbf{U}\omega\sigma\omega^{-1}, \\ \mathbf{A}(\mathbf{U}\omega) &\approx (\mathbf{U}\omega)\sigma. \end{aligned} \quad (10)$$

Therefore $\tilde{\mathbf{H}}$ is determined on the subspace spanned by the orthogonal singular basis vectors \mathbf{U} obtained via \mathbf{P}_1 ,

$$\begin{aligned} \tilde{\mathbf{H}} &= \mathbf{U}^*(\mathbf{A})\mathbf{U}, \\ \tilde{\mathbf{H}} &= \mathbf{U}^*(\mathbf{P}_2\mathbf{V}\Sigma^{-1}\mathbf{U}^*)\mathbf{U}, \end{aligned} \quad (11)$$

which can be rewritten as:

$$\tilde{\mathbf{H}} = \mathbf{U}^*\mathbf{P}_2\mathbf{V}\Sigma^{-1}. \quad (12)$$

Here $\mathbf{U}^* \in \mathbb{C}^{(L-1) \times K}$ and $\mathbf{V} \in \mathbb{C}^{(L-1) \times (L-1)}$ are the conjugate transpose of \mathbf{U} and \mathbf{V}^* , respectively; and $\Sigma^{-1} \in \mathbb{C}^{(L-1) \times (L-1)}$ denotes the inverse of the singular values Σ .

By replacing $\Psi = \mathbf{U}\omega$ in Eq. (10) i.e., $\mathbf{A}(\Psi) \approx (\Psi)\sigma$, we obtain the dynamic modes Ψ . $\therefore \mathbf{U} = \mathbf{P}_2\mathbf{V}\Sigma^{-1}$, therefore, we have:

$$\Psi = \mathbf{P}_2\mathbf{V}\Sigma^{-1}\omega \quad (13)$$

The complex eigenvalues σ contain growth/decay rates and frequencies of the corresponding DMD modes [10], [9]. If σ_j are the diagonal elements of σ from Eq. (9), the temporal behaviour of the DMD modes is then formed via Vandermonde matrix \mathbf{V}_{and} , which raises its column vector to the appropriate

¹The authors in [19] use block-Hankelisation approach on the multi-dimensional data. Our proposed algorithm is especially for univariate data.

power. $\mathbf{V}_{\text{and}}(f)$ with $(L-1) \times (f+1)$ elements will be defined by the following:

$$\mathbf{V}_{\text{and}}(f) = \begin{pmatrix} 1 & \sigma_1^1 & \sigma_1^2 & \dots & \sigma_1^f \\ 1 & \sigma_2^1 & \sigma_2^2 & \dots & \sigma_2^f \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \sigma_{L-1}^1 & \sigma_{L-1}^2 & \dots & \sigma_{L-1}^f \end{pmatrix}, \quad (14)$$

$\mathbf{V}_{\text{and}}(L)$ is a standard Vandermonde matrix for reconstruction but if $f > L$, this is used for forecasting, which will be discussed in section II-C.

DMD modes with frequencies μ_j is defined by:

$$\mu_j = \frac{\ln(\sigma_j)}{\delta t}, \quad (15)$$

where δt is the lag between the vectors and may be considered to consist of K time steps. The real part of μ_j regulates the growth or decay of the DMD modes, while the imaginary part of μ_j drives oscillations in the DMD modes. The frequencies near the origin can be interpreted as the trend of the time series, and the frequencies bounded away from the origin contain harmonics of the time series. Specifically, the trend in the time series has an associated frequency at the origin of the complex plane with $\|\mu_j\| \approx 0$.

C. Reconstruction and Forecasting

The DMD reconstruction of the data $\hat{\mathbf{X}}$ at any instance f after the initial vector $\bar{\mathbf{x}}_1$ is given by,

$$\hat{\mathbf{X}} = \Psi(\mathbf{V}_{\text{and}}(f) \circ \mathbf{b}), \quad \forall f \in \{1, 2, \dots, F\}. \quad (16)$$

Where $\mathbf{V}_{\text{and}}(f) = \sigma^{f-1}$ and F is the number of future points that are to be forecast. The vector $\mathbf{b} \approx \Psi^{-1} \bar{\mathbf{x}}_1$ contains the initial amplitudes for the dynamic modes. Operator (\circ) is the point-wise (element-wise) multiplication of \mathbf{b} with every column of $\mathbf{V}_{\text{and}}(f)$. The first lagged vector $\bar{\mathbf{x}}_1$ of the data \mathbf{X} reduces to

$$\bar{\mathbf{x}}_1 = \Psi \mathbf{b}, \quad (17)$$

and \mathbf{b} can be obtained through the Moore-Penrose pseudo inverse,

$$\mathbf{b} = (\Psi^* \Psi)^{-1} \Psi^* \bar{\mathbf{x}}_1. \quad (18)$$

The significance of the DMD modes Ψ are determined by their scaling factors \mathbf{b} when it comes to approximating the entire data sequence \mathbf{P}_2 during reconstruction and forecasting.

Since the slow varying frequencies μ_j capture trends in the time series, it is desirable to reconstruct the trajectory matrix with selected DMD triples. A DMD triple consists of a DMD eigenvalue, its corresponding DMD mode and an associated scaling factor, i.e., amplitude.

$$\hat{\mathbf{X}} = \Psi_{\{S\}}(\mathbf{V}_{\text{and}\{S\}}(f) \circ \mathbf{b}_{\{S\}}), \quad S \subset \{1, \dots, L-1\}. \quad (19)$$

$\mathbf{V}_{\text{and}\{S\}}(f)$ is a subset of $\mathbf{V}_{\text{and}}(f)$ formed by selecting only the appropriate rows which correspond to the selected DMD eigenvalues. $\mathbf{V}_{\text{and}\{S\}}(f)$ will have the dimension of $S \times (f+1)$ and $\mathbf{b}_{\{S\}}$ has the dimension of $(S \times 1)$ for the selected S eigenvalues.

D. Hankelisation and Diagonal Averaging

Finally, the reconstructed trajectory matrix $\hat{\mathbf{X}}$ is Hankelised, in order to ensure that the anti-diagonal elements are equalised. The Hankelisation operator \mathbb{H} for an $K \times F$ matrix $\tilde{\mathbf{X}}$ is defined as

$$\mathbb{H}\hat{\mathbf{X}} = \tilde{\mathbf{X}} = \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 & \dots & \tilde{x}_F \\ \tilde{x}_2 & \tilde{x}_3 & \dots & \tilde{x}_{F+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_K & \tilde{x}_{K+1} & \dots & \tilde{x}_{K+F-1} \end{pmatrix}. \quad (20)$$

The desired or reconstructed time-series is given by $\Re(\tilde{\mathbf{x}}_k) = 1/\text{num}(D_k) \sum_{m,n \in D_k} \hat{\mathbf{X}}_{m,n}$. Here k is the index of the time series, $\Re(\cdot)$ is the real part of $\tilde{\mathbf{x}}_k$, $\text{num}(\cdot)$ is the number of combinations of (m, n) , $m+n = k+1$ and D_k is given by $\{(m, n) : 1 \leq m \leq K, 1 \leq n \leq F, m+n = k+1\}$.

The methodology, as summarised via utilising Eq. (19), can be used to decompose data into a number of desired subcomponents including trend, noise and future components, but at the expense of introducing a time stamp delay. This is because, \mathbf{P}_2 has been used for projection from Eq. (13), which does not contain information about the first data point. Therefore, this study recommends that the first column vector of the trajectory should be repeated.

III. SINGULAR SPECTRUM ANALYSIS

SSA is another powerful method for analysing real-valued time series [24]. The initial idea of SSA is associated with the work of Broomhead [25], [26]. It combines the multivariate statistics, classical time series analysis, dynamical systems, and signal processing.

SSA is becoming an effective method in various areas such as economics [24], [27] and biomedical engineering [28], [29], [30]. Basically, SSA decomposes a data into a number of interpretable elements with different subspaces, such as noise and trend [31], and can be used for any time series with complex structure [27]. For instance, for decades, SSA has been used for both trend detection and prediction in financial data [24], [27]. SSA can be used in several applications including trends fitting, extraction of cycles with various periods and amplitudes, smoothing, and finding some structures in short time series.

Ordinary SSA is a subspace decomposition algorithm with four stages similar to that of the proposed method but with different formulation in the second and third stages. SSA decomposes the resulted matrix of embedding step using SVD:

$$\mathbf{C} = \sum_{i=1}^d \mathbf{C}_i = \sum_{i=1}^d \sqrt{\sigma_i} \mathbf{u}_i \mathbf{v}_i^T, \quad (21)$$

where σ_i is the i th eigenvalue of covariance matrix $\mathbf{C} = \mathbf{X}\mathbf{X}^T$, \mathbf{u}_i is the corresponding eigenvector, d is the total number of eigenvalues, and $\mathbf{v}_i = \mathbf{C}^T \mathbf{u}_i / \sqrt{\sigma_i}$.

In the reconstruction step, first the elementary matrices from the previous stage are grouped into several sub-matrices

$$\mathbf{C} = \sum_{q=1}^Q \hat{\mathbf{C}}_q, \quad (22)$$

where Q determines the total number of groups, index q refers to q th subgroup of eigenvalues, and $\hat{\mathbf{X}}_q$ indicates the sum of \mathbf{X}_i within group q . Finally, the desired subgroup is selected and transformed into a Hankel matrix. Later, the average of the anti diagonal elements form the reconstructed desired time series.

SSA has recurrent and vector based forecasting techniques:

A. SSA Recurrent Forecasting

If the desired group of eigentriples is considered as I and the corresponding i th eigenvector as \mathbf{u}_{I_i} , then

$$r = \frac{1}{1 - v^2} \sum_{i \in I} \pi_i \mathbf{u}_{I_i} \quad (23)$$

where $r = (a_{l_1-1}, \dots, a_1)^T$, \mathbf{u}_{I_i} and π_i respectively show all except for the last and the last column of \mathbf{u}_{I_i} , and $v^2 = \sum_{i \in I} \pi_i^2$. Consequently, the recurrent forecasting can be formulated as

$$x_i = \begin{cases} \tilde{x}_i & i = 1, \dots, T \\ \sum_{j=1}^{l_1-1} a_j x_{i-j} & i = T + 1, \dots, T + F \end{cases} \quad (24)$$

where F is the number of forecasting points. The reconstructed terms and Eq. (23) are used to forecast the value of new points. The linear operator φ can be defined as

$$\varphi \mathbf{x} = \begin{pmatrix} \mathbf{x} \\ \mathbf{r}^T \mathbf{x} \end{pmatrix}, \quad (25)$$

where \mathbf{x} shows $K - 1$ components of vector \mathbf{x} . Therefore, trajectory matrix of the time series \mathbf{x} of length $T + F$ can be written as the vector representation of Eq. (24)

$$\mathbf{z}_i = \begin{cases} \mathbf{x}_i & i = 1, 1, \dots, L \\ \varphi \mathbf{z}_{i-1} & i = L + 1, \dots, T + F \end{cases} \quad (26)$$

IV. DATASET

A. Real Data

To demonstrate and validate the methodology, publicly available Box & Jenkins airline data [2] have been used. This consists of the monthly total of international airline passengers from the year 1949 to 1960. For comparing with SSA, BJ sales and UK gas data are also considered. All of the aforementioned datasets have been obtained from Github repository for R time series datasets². These datasets then have been transformed into the logscale (y-axis) and the time (x-axis) is considered in increments of 1.

B. Simulated Data

The simulated data are a mixture of two signals

$$x(t) = \mathbf{a}s(t) + \beta e(t), \quad (27)$$

where $s(t)$ shows the desired signal and \mathbf{a} is the mixing vector. $x(t)$ is the input or mixture being examined, $e(t)$ indicates the unwanted signal or noise, and β is the noise level.

Signal-to-noise ratio (SNR) in terms of mean square (MS) is used as a measure for noise level which is adjusted by changing β

$$\text{SNR} = \text{MS}(s(t)) / \beta^2. \quad (28)$$

To evaluate the simulation performance, the evaluation criteria in terms of mean square error (MSE) is considered as

$$\text{MSE} = \frac{\text{MS}(s(t) - \hat{s}(t))}{\text{MS}(s(t))}, \quad (29)$$

where $\hat{s}(t)$ is the estimated source.

V. EXPERIMENTS AND RESULTS

The real time series containing airline data is transformed to a trajectory matrix with K -lagged vectors. The parameter that needs to be predetermined is the window length L because our conjecture is that an appropriate choice of this can produce a better result.

A. Effect of Window Length L

Fig. 1 (a) shows our experimental results with window lengths L for reconstructing the original time series. Our result shows that the reconstruction error for $L \in \{12, \dots, 66\}$ gave a stable performance with relatively low reconstruction error. If L is relatively small, it can cause the trajectory matrix \mathbf{X} to miss the trends in the data especially when dealing with seasonal components. Furthermore, if L is closer to $\frac{T}{2}$, where T is the length of the time series, the decomposition of the data becomes more detailed. Therefore, the selection of L can neither be close to 2 nor $\frac{T}{2}$. Therefore, this study, recommends choosing L between 8% to 45% of the length of the time series. In our experiments we arbitrarily considered L to be 25% of the length of the time series, i.e., $L = 35$.

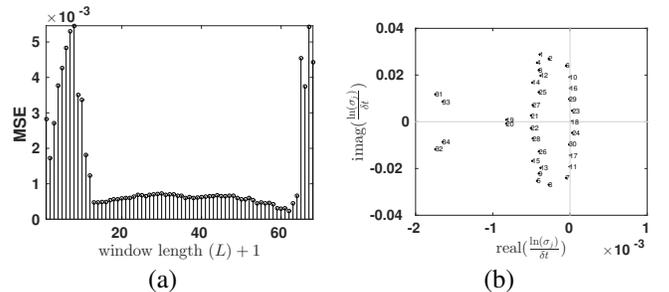


Fig. 1. (a) Effect of window length L over the reconstruction performance. (b) DMD eigenvalue plot, showing eigenvalues $\{18, 19, 20, 21, 22\}$ that are vertically closer to the X-axis

²<https://github.com/vincentarelbundock/Rdatasets>

B. DMD for Extracting Trends

The logarithmic values of DMD eigenvalues give us the frequencies, the real part captures the growth/decay and imaginary part, the oscillations within the time series. The frequencies that are closer to origin are the slow varying components (for example, eigenvalues $\{18, 19, 20, 21, 22\}$ in Fig. 1 (b) which are vertically closer to the X-axis) that essentially capture the trends.

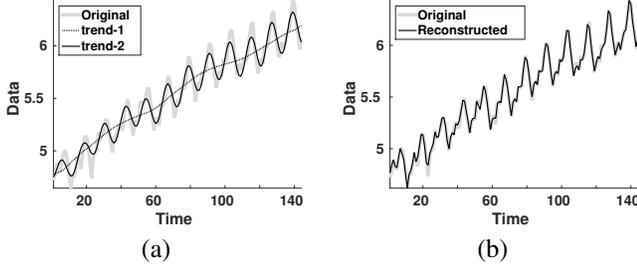


Fig. 2. (a) Trend analysis of the timeseries, revealing two trends (i) trend1: linear trend and (ii) periodic trend. (b) A complete reconstruction.

The ones which are further away from the origin capture various frequency growth/decay rates. In order to establish the order in the DMD, the absolute frequencies of the DMD eigenvalues are sorted in the ascending order. The absolute value of the frequencies associated with each of the eigenvalues are shown in Fig. 3 (b).

Reconstruction of the original time series with the aforementioned slow varying DMD triples, i.e., with $\{18, 19, 20\}$ in Fig. 1 (b) (or the first three DMD triples) gave us the trend (trend-1) as shown in Fig. 2 (a).

Reconstruction with other DMD triples that are closer to the origin, i.e., $\{18, 19, 20, \dots, 26\}$ gave a smoother version (trend-2) of the time series (Fig. 2 (a)).

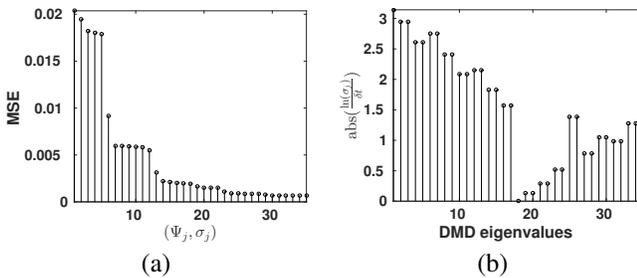


Fig. 3. (a) DMD triples vs reconstruction error. (b) Corresponding absolute frequencies.

As we reconstruct the data with increasing DMD triples, the reconstruction mean square error (MSE) decreases (Fig. 3 (a)). A complete reconstruction of the original time series with all of the DMD triples produces the original time series as shown in Fig. 2 (b) with a reconstruction MSE of 0.0006.

C. DMD for Time Series Forecasting

To demonstrate the capability of DMD for forecasting, 124 out of the total 144 datapoints are considered as the training set, leaving the remaining 20 points as the test set. The window length L considered comprises of 25% of the data i.e., $L = 31$. The forecast trends in the data are shown in Fig. 4 (a).

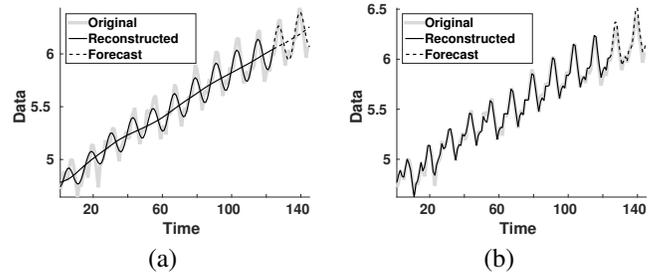


Fig. 4. (a) Trend forecast. (b) A complete forecast of the timeseries.

The MSE recorded for the forecast trend-1 was 0.0007 and for the smoother trend-2 was 0.0080 when evaluated against the reconstructed trend. The result of the forecast is shown in Fig. 4 (b) with a MSE of 0.0090 when compared to the original data.

D. Comparison to SSA

1) *On Simulated Data with Effect of Noise:* In order to compare with SSA, the performance of both the methods should be investigated in the presence of noise. For this purpose, simulated signal consisting of two sinusoids with different frequencies and powers are considered, which were then mixed with a Gaussian noise at various signal to noise ratio (SNR) levels $\{-10, -9, \dots, 19, 20\}dB$.

2) *Selection of the triples:* The proposed method using the first 9 DMD triples and SSA with first 4 singular triples were applied as they those parameters gave the least reconstruction MSE across all the noise levels on this simulated data as shown in Figure5.

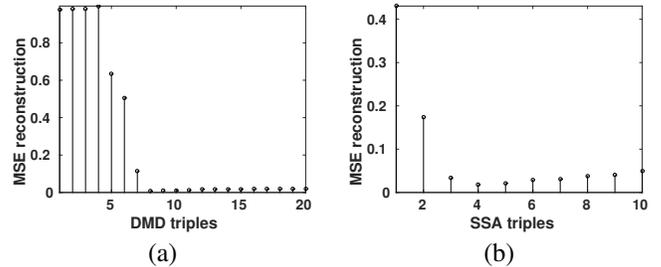


Fig. 5. MSE while reconstruction of the original time series in the presence of noise at $-2dB$, using first t (a) DMD triples, (b) SVD triples.

3) *Robustness to Noise:* It can be noticed from Fig. 6 that, as the SNR increases, the MSE in reconstruction (with respect to the clean signal) decreases in both of the methods. The average time taken for DMD across all the noise levels was 3.76, while for SSA was 46.94 seconds³. This result shows that DMD is computationally more efficient than SSA and is consequently a promising approach for analysing a univariate time series in the presence of noise.

Fig. 7 shows the original simulated signal and the noisy signal at $SNR = -2dB$. The reconstruction along with forecasting of 750 points, using the first 5 DMD triples and 4 singular triples is shown. The MSE recorded after reconstruction and forecast was 0.0205 and for SSA was 0.0241. The time taken for DMD and SSA was 0.17 and 23.23 seconds respectively.

³On a Dell PowerEdge R715 AMD - 2-CPU/24-Core - 128GB RAM Running Precise (Ubuntu 12.04.5).

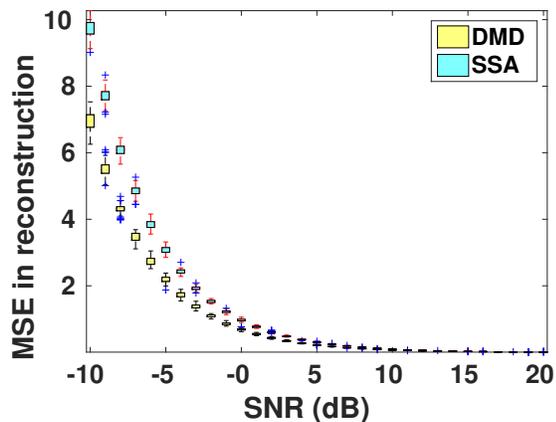


Fig. 6. Effect of noise in reconstruction at various SNRs across 500 iterations.

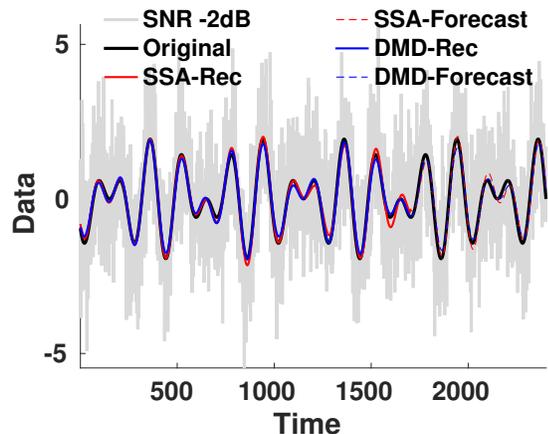


Fig. 7. Reconstruction and forecasting at $-2dB$

4) *On Real Data:* The results on the real world datasets are shown in Fig. 8, (a) Airline passengers data, (b) BJ sales data) and (c) UK gas data. The MSE error in reconstructing and forecasting these data, including the parameters chosen are given in Table I. These results show that DMD based univariate time series analysis is superior to SSA particularly when forecasting the time series.

VI. CONCLUSION AND DISCUSSIONS

In this study we proposed the method of DMD in a four-stage pipeline to decompose a univariate time series into a number of interpretable elements with different components, such as noise and underlying data trend. In addition, this four-stage DMD pipeline was also shown to be used for forecasting a time series.

DMD eigenvalues with frequencies that are closer to the origin in the complex plane capture the trends in the time series. Moreover, the temporal evolution of the DMD modes, which is preserved via the Vandermonde matrix, can be used to reconstruct the desired components and perform forecasting at the same time. Window length L needs to be predetermined and we recommend choosing L between 8% to 45% of the length of the time series. In reconstruction, to compensate one time stamp delay, we recommend that the first column vector

of the trajectory matrix be repeated once. The experimental results at various noise levels on simulated data suggests that DMD is a promising approach to modelling a time series with a noisy structure. DMD was compared to SSA, a popular method employed extensively in biomedical signal processing applications for separating artefacts and extracting features from electromyography (EMG), electrocardiogram (ECG) and electroencephalogram (EEG) signals [28].

Thus DMD can potentially be applied to analyse biomedical signals, especially EEG, where brain rhythms manifest themselves as narrow frequency band components. As a good example, sleep is a dynamic process which consists of different stages with different neural activity levels. Each stage is characterised by a distinct set of physiological and neurological features and dominant frequency band. SSA, however, does not exploit the narrowband property of these kind of cyclic data since it requires manual grouping or optimisation [30] to detect the desired features. DMD on the other hand due to its intrinsic nature of complex eigenvalues that capture frequency bands, can help find cycling components automatically. Therefore, in such studies, DMD might capture clues for chaotic subcomponents/trends and noise. These subcomponents later can be used as feature vectors for e.g. in sleep stage classification, epilepsy detection, and seizure predictions (for e.g., using SSA in [28], [3], [4], [5]). Moreover, our experiments showed that DMD is computationally more efficient than SSA as well as robust in the presence of added white Gaussian noise.

ACKNOWLEDGEMENT

The funding for this work has been provided by the Department of Computer Science and the Centre for Vision, Speech and Signal Processing (CVSSP) - University of Surrey. ‘S.T’ and ‘N.P’ have benefited from the Medical Research Council (MRC) funded project “Modelling the Progression of Chronic Kidney Disease” under the grant number R/M023281/1. The details of the project are available at www.modellingCKD.org.

REFERENCES

- [1] N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for time series*. Springer Science & Business Media, 2013.
- [2] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [3] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals*. CRC Press, 2015.
- [4] S. M. Mohammadi, S. Kouchaki, M. Ghavami, and S. Sanei, “Improving time-frequency domain sleep {EEG} classification via singular spectrum analysis,” *Journal of Neuroscience Methods*, pp. –, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016502701630187X>
- [5] S. Kouchaki, S. Sanei, E. L. Arbon, and D. J. Dijk, “Tensor based singular spectrum analysis for automatic scoring of sleep eeg,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, pp. 1–9, Jan 2015.
- [6] J. B. Elsner and A. A. Tsonis, *Singular spectrum analysis: a new tool in time series analysis*. Springer Science & Business Media, 2013.
- [7] P. J. Schmid and J. L. Sesterhenn, “Dynamic mode decomposition of numerical and experimental data,” *In Bull. Amer. Phys. Soc., 61st APS meeting, San Antonio*, p. 208, 2008.
- [8] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, “On dynamic mode decomposition: theory and applications,” *arXiv preprint arXiv:1312.0041*, 2013.
- [9] P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.

TABLE I. THE PARAMETERS THAT HAVE BEEN USED FOR ANALYSING THE REAL DATA USING DMD AND SSA. HERE PARAMETERS, T AND L REPRESENT THE TOTAL LENGTH OF THE TIME SERIES AND LENGTH OF THE WINDOW CONSIDERED FOR THE ANALYSIS RESPECTIVELY. ‘REC’ DENOTES RECONSTRUCTION AND ‘FOR’ DENOTES FORECASTING. MSE DENOTES THE MEAN SQUARE ERROR BETWEEN THE RECONSTRUCTED/FORECAST TIME SERIES AND THE ORIGINAL TIMESERIES.

Dataset	Parameters				DMD Triples				SSA Triples				MSE in Rec ,For	
	T		L		Trend		Period		Trend		Period		DMD	SSA
	Rec	For	Rec	For	Rec	For	Rec	For	Rec	For	Rec	For		
Airline Passengers	144	130	36	33	2	2	7	9	1	2	3	3	0.0010	0.0004
BJSales	150	135	38	34	2	2	7	7	1	1	7	7	0.00006	0.0001
UKgas	97	87	24	22	2	2	12	12	1	2	2	2	0.0088	0.0142

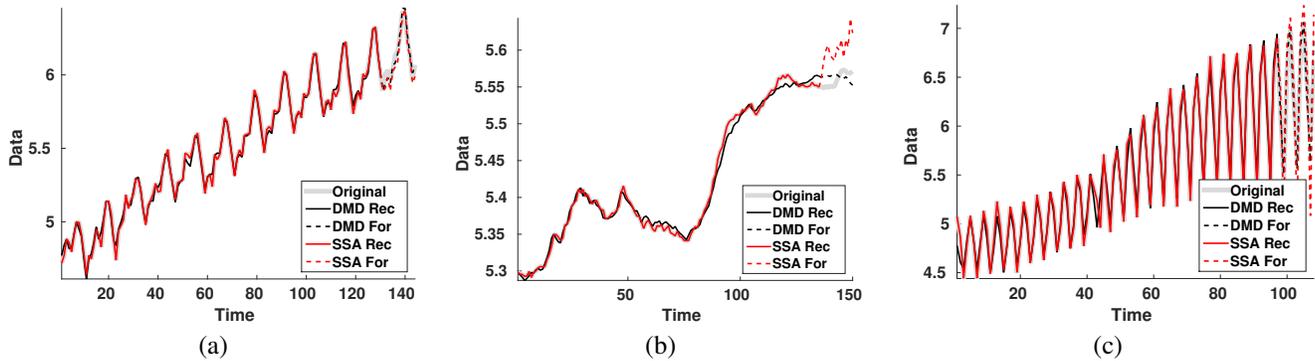


Fig. 8. Comparison of DMD and SSA on (a) Airline passengers data, (b) BJ sales data) and (c) UK gas data.

- [10] P. J. Schmid, K. E. Meyer, and O. Pust, “Dynamic mode decomposition and proper orthogonal decomposition of flow in a lid-driven cylindrical cavity,” in *8th International Symposium on Particle Image Velocimetry*, 2009, pp. 25–28.
- [11] P. Schmid, L. Li, M. Juniper, and O. Pust, “Applications of the dynamic mode decomposition,” *Theoretical and Computational Fluid Dynamics*, vol. 25, no. 1-4, pp. 249–259, 2011.
- [12] S. Tirunagari, V. Vuorinen, O. Kaario, and M. Larmi, “Analysis of proper orthogonal decomposition and dynamic mode decomposition on les of subsonic jets,” *CSI Journal of Computing*, vol. 1, pp. 20–26, 2012.
- [13] J. Grosek and J. N. Kutz, “Dynamic mode decomposition for real-time background/foreground separation in video,” *CoRR*, vol. abs/1404.7592, 2014. [Online]. Available: <http://arxiv.org/abs/1404.7592>
- [14] S. Tirunagari, N. Poh, M. Bober, and D. Windridge, “Can dmd obtain a scene background in color?” *arXiv preprint arXiv:1607.06783, International Conference on Image, Vision and Computing (ICIVC), August 3-5, 2016, Portsmouth, UK*, 2016.
- [15] N. Erichson and C. Donovan, “Randomized low-rank dynamic mode decomposition for motion detection,” *Computer Vision and Image Understanding*, vol. In Press, 2 2016.
- [16] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. Ho, “Detection of face spoofing using visual dynamics,” *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 762–777, 2015.
- [17] S. Tirunagari, N. Poh, M. Bober, and D. Windridge, “Windowed dmd as a microtexture descriptor for finger vein counter-spoofing in biometrics,” in *Information Forensics and Security (WIFS), 2015 IEEE International Workshop on*. IEEE, 2015.
- [18] E. Berger, M. Sastuba, D. Vogt, B. Jung, and H. B. Amor, “Dynamic mode decomposition for perturbation estimation in human robot interaction,” in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Aug 2014, pp. 593–600.
- [19] B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz, “Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition,” *Journal of Neuroscience Methods*, vol. 258, pp. 1–15, 2016.
- [20] A. Krylov, “On the numerical solution of the equation by which in technical questions frequencies of small oscillations of material systems are determined,” *Izvestija AN SSSR (News of Academy of Sciences of the USSR), Otdel. mat. i estest. nauk*, vol. 7, no. 4, pp. 491–539, 1931.
- [21] Y. Saad, “Krylov subspace methods for solving large unsymmetric linear systems,” *Mathematics of Computation*, vol. 37, no. 155, pp. 105–126, 1981.
- [22] A. Ruhe, “Rational krylov sequence methods for eigenvalue computation,” *Linear Algebra and its Applications*, vol. 58, pp. 391–405, 1984.
- [23] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. SIAM, 1974, vol. 161.
- [24] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of time series structure: SSA and related techniques*. CRC Press, 2010.
- [25] D. Broomhead and G. P. King, “Extracting qualitative dynamics from experimental data,” *Physica D: Nonlinear Phenomena*, vol. 20, no. 2, pp. 217–236, 1986.
- [26] D. S. Broomhead and G. P. King, “On the qualitative analysis of experimental dynamical systems;” in *Nonlinear Phenomena and Chaos*, S. Sarkar, Ed. Bristol, England: Adam Hilger, 1986, pp. 113–144.
- [27] T. Alexandrov and N. Golyandina, “The automatic extraction of time series trend and periodical components with the help of the Caterpillar-SSA approach,” *Exponenta Pro* 3, vol. 4, pp. 54–61, 2004.
- [28] S. Sanei, T. K.-M. Lee, and V. Abolghasemi, “A new adaptive line enhancer based on singular spectrum analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 428–434, 2012.
- [29] F. Ghaderi, H. R. Mohseni, and S. Sanei, “Localizing heart sounds in respiratory signals using singular spectrum analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 12, pp. 3360–3367, 2011.
- [30] S. Sanei, M. Ghodsi, and H. Hassani, “An adaptive singular spectrum analysis approach to murmur detection from heart sounds,” *Medical Engineering & Physics*, vol. 33, no. 3, pp. 362–367, 2011.
- [31] S. Kouchaki and S. Sanei, “Supervised single channel source separation of eeg signals,” in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2013, pp. 1–5.