CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Conference Paper

A Dynamic Mode Decomposition approach with Hankel blocks to forecast multi-channel temporal series

Journal to Conference paper (CISTER-TR-190502).

Ênio Filho*

Paulo Lopes dos Santos

*CISTER Research Centre

CISTER-TR-190805

2019/12/11

A Dynamic Mode Decomposition approach with Hankel blocks to forecast multi-channel temporal series

Ênio Filho*, Paulo Lopes dos Santos

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: enpvf@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Forecasting is a task with many concerns, such asthe size, quality, and behavior of the data, the computing power todo it, etc. This paper proposes the Dynamic Mode Decompositionas a tool to predict the annual air temperature and the sales of astores 19 chain. The Dynamic Mode Decomposition decomposes thedata into its principal modes, which are estimated from a trainingdata set. It is assumed that the data is generated by a lineartime-invariant high order autonomous system. These modes areuseful to find the way the system behaves and to predict itsfuture states, without using all the available data, even in a noisyenvironment. The Hankel block allows the estimation of hiddenoscillatory modes, by increasing the order of the underlyingdynamical system. The proposed method was tested in a casestudy consisting of the long term prediction of the weekly salesof a chain of stores. The performance assessment was based onthe Best Fit Percentage Index. The proposed method is comparedwith three Neural Network Based predictors.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2019.2917811, IEEE Control Systems Letters

A Dynamic Mode Decomposition Approach With Hankel Blocks to Forecast Multi-Channel Temporal Series

Enio Vasconcelos Filho, Paulo Lopes dos Santos, *Member, IEEE*

***Abstract*—Forecasting is a task with many concerns, such as the size, quality, and behavior of the data, the computing power to do it, etc. This paper proposes the Dynamic Mode Decomposition as a tool to predict the annual air temperature and the sales of a stores' chain. The Dynamic Mode Decomposition decomposes the data into its principal modes, which are estimated from a training data set. It is assumed that the data is generated by a linear time-invariant high order autonomous system. These modes are useful to find the way the system behaves and to predict its future states, without using all the available data, even in a noisy environment. The Hankel block allows the estimation of hidden oscillatory modes, by increasing the order of the underlying dynamical system. The proposed method was tested in a case study consisting of the long term prediction of the weekly sales of a chain of stores. The performance assessment was based on the Best Fit Percentage Index. The proposed method is compared with three Neural Network Based predictors.**

***Index Terms*—Dynamic Mode Decomposition, Hankel Matrix, Prediction, System Identification.**

I. Introduction

FORECASTING is a task with many concerns such as the size, quality and how to find the behavior of the data, the computational power, etc. Since time-series can be of distinct nature and have different behaviors, there are several approaches to model and predict them. For example, temperature, and sales can have completely different modes and behaviors, even if they are measured in the same place at the same time interval.

Many different approaches can be used to predict time-series, such as fuzzy Evolving Methods [1], regression [2], state space and Auto-regressive Integrated Moving Average (ARIMA) [3], moving average [4] and Neural Networks [5]. Yet, most of these approaches require a large amount of training data [6] and are subjected to many optimization problems, like the difficulty to define the best parameters to achieve the best results [7].

This paper addresses the problem of long term forecasting time-series with a periodical or quasi-periodical behaviour. The Dynamic Mode Decomposition (DMD) is proposed to estimate the system principal modes and to predict its future behaviour. The DMD is a powerful analysis tool that reveals spatial coherent structures associated with temporal spectral components [8]. Although it was originally proposed to analyse fluid flow data [9] it can also be used to model and predict temporal series [10]. It is a data-driven method that decomposes a set of signals into principal modes. As it exploits the low-dimensional structure of the data it spends low computational resources [11]. The DMD assumes multi-channel time-series being generated by autonomous linear time-invariant systems where each channel is a state variable. If these state-variables do not span the time-series principal modes, DMD fails. In this paper, signal observations are arranged in a block Hankel matrix before performing DMD, creating new shifted in time state-variables for the underlying autonomous system. This strategy increases the system's order enabling the DMD to capture more signal modes. The DMD with Hankel matrix was also used in [12] to develop an alternative view of Koopman analysis. This approach was called the Hankel alternative view of Koopman (HAVOK) and enables the estimation of linear models that capture the periodic or almost periodic dynamics of nonlinear systems in a nearly perfect way. Some other authors also proposed the DMD with the Hankel matrix [13], [14], [15], [16].In [13], the Hankel matrix is referred but not tested. In [14] the DMD is used with the well know subspace identification algorithm N4SID. In [15], the authors propose a learning method where the "lags" of the Hankel Matrix are defined by an Artificial Neural Network. But, one of the most important contributions is in [16], where the Birkhoff's ergodic theorem is used to show that the DMD with an Hankel matrix yields the true Koopman eigenfunctions and eigenvalues. One of the contributions of this work is to use an autonomous LTI system modal analysis approach to prove the benefits of the use of the Hankel matrix in the DMD. These benefits are illustrated in a real case study presented in the next section where the use of the Hankel matrix significantly improved the DMD based long term predictions.

The performance of the proposed method is illustrated in a case study consisting in the weekly sales data of a chain of stores. Each store is a time-series channel. Despite the large number of channels (2659), the block Hankel matrix almost doubled the number of estimated modes, significantly improving the prediction accuracy. The DMD was also compared with three artificial neural network (ANN) predictors.

The paper is organized as follows: After Section I, this introductory section, Section II describes the DMD and introduces the use of the Hankel block matrix. Section III presents the

Enio Vasconcelos Filho is with Cister Research Centre, ISEP, Rua Alfredo Allen 535, 4200-135 Porto, Portugal, e-mail: enpvf@isep.ipp.pt.

Paulo Lopes dos Santos is with INESC TEC - Institute for Systems and Computer Engineering Technology and Departmento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 PORTO, Portugal

Manuscript received March 01, 2019; revised April 25, 2019.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2019.2917811, IEEE Control Systems Letters

case study, discusses the results of the DMD with and without a block Hankel matrix and compares the DMD with the ANN predictors. Finally, the main conclusions are drawn and some future work directions are indicated in Section IV.

II. Dynamic Mode Decomposition with a Hankel Block Matrix

A. The Dynamic Mode Decomposition Algorithm

In this section, the DMD algorithm is briefly reviewed. More details can be found in [17], [9] and [18]. It is assumed that data is generated by an autonomous dynamic system described by the set of differential equations $\dot{x} = f(x,t)$, where $x(t) \in \mathbb{R}^n$ is the state at time t. The discrete-time representation is $x_{k+1} = \mathbf{F}(x_k)$ where $x_k = x(k\Delta t)$. The DMD algorithm works with temporal series, and, in its earlier versions, assumed a constant sampling period $T_s = \Delta_t$, such that the measured state-values are snapshots in a Krylov subspace. However, in [11] a modern definition is proposed, where the algorithm can work with sparse spatial and temporal collections of data. The initial condition can be defined as $x(0) = x_0$. The function $\mathbf{F}(\cdot)$ defines the governing equations of the system. It is non-linear and unknown. The initial conditions and the state measurements are the only available information. The state is represented by x. So, x is an n-dimensional vector with $(n_x >> 1)$. The DMD allows the construction of an approximate locally linear dynamical system $\dot{x} = Ax$. A solution with $x(0) \neq 0$ is [19]:

$$x(t) = \sum_{j=1}^{n} b_j \phi_j \exp(\omega_j t), \tag{1}$$

where Φ_j and ω_j, are the *eigenvectors* and *eigenvalues* of A, b_j, $j = 1, \ldots, n$ the coordinates of $x(0)$ in the base defined by the eigenvectors, and $\exp(\omega_j t)$, the system modes, all for $j = 1, \ldots, n,$. An eigenvalue with positive real part generates a growing mode. The opposite happens if the real part is negative. If the real part is zero, i.e., if the eigenvalue is in the imaginary axis, the mode is oscillatory. If $x(t)$ is uniformly sampled, the solution (1) at the sampling instants, $t_k = kT_s$ is

$$x(k) = \sum_{j=1}^{n} \phi_j \lambda_j^k b_j = \Phi \Lambda^k b, \tag{2}$$

where λ_j, $j = 1, \ldots, n$, are the *eigenvalues* of $A_d = \exp(AT_s)$, the state-matrix of the discrete-time map $x_{k+1} = A_d x_{k+1}$, $\Phi = \begin{bmatrix} \phi_1 & \cdots & \phi_n \end{bmatrix}$, $\Lambda = \text{diag}(\lambda_j)$, and $b = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T$. Using the DMD, it is possible to carry out a spatio-temporal decomposition of the data into a set of dynamic modes coming from measurements (or snapshots) equally spaced in time. Those modes have temporal frequencies associated with the imaginary parts of the eigenvalues ω_j [20]. The data collection is arranged in a $n \times m$ matrix where n is the number of state-variables and m is the number of observations (snapshots). The matrix takes the form $X = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}$. Splitting the data in two matrices delayed in time gives:

$$X_1 = \begin{bmatrix} x_1 & \cdots & x_{m-1} \end{bmatrix} \tag{3}$$

$$X_2 = \begin{bmatrix} x_2 & \cdots & x_m \end{bmatrix} \tag{4}$$

and X_2 can be expressed as

$$X_2 = A_d X_1 \tag{5}$$

because the operator A_d maps x_i into x_{i+1}. The DMD uses the *eigenvalues* λ_j of the best-fit linear operator A_d relating X_2 and X_1. The *eigenvalues* of A are

$$\omega_j = \ln(\lambda_j)/T_s, \quad j = 1, ..., n. \tag{6}$$

The eigenvectors ϕ_j and the system modes, $\exp(\omega_j t)$, are extracted from a low-rank structure. To find them, x_m can be expressed as a linear combination of the x previous values, $x_m = \sum_{i=1}^{m-1} a_i x_i + v$, where v is the residual vector. X_2 is related with X_1 by

$$X_2 = X_1 S + v e_{m-1}, \tag{7}$$

where S is the companion matrix [21]

$$S = \begin{bmatrix} 0 & \cdots & 0 & a_1 \\ 1 & \cdots & 0 & a_2 \\ \vdots & \ddots & \vdots & \ldots \\ 0 & \ldots & 1 & a_{m-1} \end{bmatrix},$$

and $e_{(m-1)}$ is the $(m-1)^{th}$ column of the m-dimensional identity matrix. The coefficients a_i, $i = 1, \ldots, m-1$, can be computed by solving a least squares problem, which minimizes the overall residual. Using (5) and (7), it is possible to see that $X_2 = A_d X_1 \cong X_1 S$. So, the eigenvalues of A_d are approximately those of S. The estimation of S can be replaced by a Singular Value Decomposition (SVD) [22] approach that yields a projection $\tilde{A}$ of A_d into the subspace spanned by the eigenvectors of its dominant modes [23]. The advantage of this approach is the power of the reduced SVD to attenuate the noise in the data and to compensate numerical truncation issues. $\tilde{A}$ is found from the SVD of X_1, $X_1 = U\Sigma V^*$, where $U \in C^{n \times r}$, $\Sigma \in C^{r \times r}$ and $V \in C^{m \times r}$. If the rank of X_1 is approximated by r, then $X_1 \approx U_r \Sigma_r V_r^*$ is the reduced SVD approximation of X_1, where Σ_r is a diagonal matrix containing the first r singular values of X_1, and U_r and V_r are orthonormal matrices with its first r left and right singular vectors. A_d is estimated in the least squares sense by $A_d = X_2 X_1^\dagger = X_2 V_r \Sigma_r^{-1} U_r^*$, where $X_1^\dagger = V_r \Sigma_r^{-1} U_r^*$ is the pseudoinverse of X_1. $\tilde{A}$ is an $r \times r$ matrix given by $\tilde{A} = U_r^* A_d U_r = U_r^* X_2 V_r \Sigma_r^{-1}$. It defines a low dimensional system $\tilde{x}_{k+1} = \tilde{A}\tilde{x}_k$ and the high dimensional state can be reconstructed from $x_k = U_r \tilde{x}_k$. The *eigendecomposition* of $\tilde{A}$ is $\tilde{A}W = W\Lambda$, where Λ is a diagonal matrix with the eigenvalues and the columns of W are the *eigenvectors*. The *eigendecomposition* of A_d is reconstructed from W and Λ. Λ contains the dominant eigenvalues of A_d and the columns of $\Phi = X_2 V_r \Sigma_r^{-1} W$ contain its dominant (DMD) *eigenvectors*, which are also dominant *eigenvectors* of A. Using (6) to convert the DMD discrete-time eigenvalues to continuous-time, the future behavior of the system can be predicted by:

$$x(t) \approx \sum_{j=1}^{r} \boldsymbol{\phi}_j \exp(\omega_j t) b_j = \boldsymbol{\Phi} \exp(\boldsymbol{\Omega} t) \boldsymbol{b} \tag{8}$$

where $\boldsymbol{\Phi}$ is the matrix with the DMD *eigenvectors*, b_j is coordinate of the initial value in the direction of ϕ_j, and $\Omega = diag(\omega)$ is a diagonal matrix with the *eigenvalues* ω_j. To compute the initial values b_j take the initial sate, $x(0)$, equal to the initial sample, x_1. Equation (8) becomes $x_1 = \boldsymbol{\Phi b}$. Then, it is possible to find b from $b = \Phi^\dagger x_1$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2019.2917811, IEEE Control Systems Letters

B. Adjusting the DMD with an Hankel matrix

The DMD extracts the dominant modes present in the measured signals to predict its future behaviour. However, this is only possible if the sets of linearly independent signals containing the dominant modes have a cardinality greater or equal than the number of these modes. i.e, if the measured signals span the dominant modes. Otherwise the DMD fails and gives inaccurate results. For instance, the signal

$$x(t) = A\cos(\omega_0 t) = \frac{A}{2}\exp(j\omega_0 t) + \frac{A}{2}\exp(-j\omega_0 t)$$

contains two modes: $\exp(j\omega_0 t)$ and $\exp(-j\omega_0 t)$. Hence, a DMD using $x(t)$ alone cannot extract them because it assumes a first order underlying autonomous system with only one mode. To extract both modes it needs at least a second order system and this is only possible if the DMD

Let $x(t) \in \mathbb{R}^n$ be the output of an r^{th} order autonomous linear time invariant system with distinct eigenvalues. Then, $x(t)$ is a linear combination of the distinct modes $d_{c,j}(t) = \exp(\omega_j t)$, $j = 1, \ldots, r$ of the system, i.e, $x(t) = Td_c(t)$, where $T \in \mathbb{R}^{n\times r}$ and $d_c(t) = \begin{bmatrix} d_{c,1}(t) & \cdots & d_{c,n}(t) \end{bmatrix}^T$. If $x(t)$ is uniformly sampled with a sampling period of T_s, then at the sampling instants $t_k = kT_s$, $x(t)$ is given by $x_k = Td_k$ where

$$d_k = \begin{bmatrix} \lambda_1^k & \cdots & \lambda_n^k \end{bmatrix}^T \tag{9}$$

with $\lambda_j = \exp(\omega_j T_s)$, $j = 1, \ldots, r$. It is well known that d_k is the state of the discrete-time autonomous linear time-invariant (LTI) system $d_{k+1} = \Lambda d_k$, with $\Lambda = \mathrm{diag}\{\lambda_j\}$, $j = 1, \ldots, r$. The state-variables d_k, and consequently, the modes $d_{c,j}(t) = \exp(\omega_j t)$, can be recovered from x_k if and only if rank$(T) = r$. Under this condition, $d_k = T^\dagger x_k$, where $T^\dagger \in \mathbb{R}^{r\times n}$ is the pseudo inverse of T, also known as the Moore-Penrose inverse. Hence $x_{k+1} = A_d x_k$ with $A_d = T\Lambda T^\dagger \in \mathbb{R}^{n\times n}$ and both d_k and $d_c(t)$ may be found by the DMD. But if rank$(T) < r$ it is not possible to recover d_k from x_k and the DMD fails. Suppose now that $x(t) \in \mathbb{R}$ is a linear combination of the modes $d_{c,j} = \exp(\omega_j t)$, $j = 1, \ldots, r$ and is uniformly sampled with a period T_s. From the discussion above it is not possible to extract these modes from this signal alone. But, x_k and its n lagged copies $x_{k+1}, \ldots, x_{k+n-1}$ are given by

$$x_{k+i} = a_1\lambda_1^{k+i} + \cdots + a_r\lambda_r^{k+i}, \quad i = 0, \ldots, n-1,$$

where $\lambda_j = \exp(\omega_j T_s)$. Rewriting this equation as

$$x_{k+1} = a_1\lambda_1^i\lambda_1^k + \cdots + a_r\lambda_r^i\lambda_r^k, \quad i = 0, \ldots, n-1$$

then, from (9), $\begin{bmatrix} x_k^T & x_{k+1}^T & \cdots & x_{k+n-1}^T \end{bmatrix} = Td(k)$, with the entries i, j of $T \in \mathbb{R}^{n\times r}$ given by $t_{i,j} = a_j\lambda^{i-1}$. As λ_j are distinct, then, if $n > r$, rank$(T) = r$ and the signal modes can be recovered with the DMD applied to this set of lagged signals.

The DMD is supposed to work with low-rank matrices. If $n >> m$, the matrix has a low rank and the DMD is easily applied. In this case, the measured variables span all dominant modes and the rank of T is equal to the number these modes. However, in some data sets, there are linear dependencies among the measured signals preventing them to span all dominant modes.Hence, the rank of T is too low and the DMD fails to correctly identify the system characteristics. To solve this problem, the the rank of T is increased by adding time lagged coordinates before applying the DMD. This is done by rearranging the data set in the Hankel matrix:

$$X_1 = \begin{bmatrix} x_1 & x_2 & \cdots & x_j \\ x_2 & x_3 & \cdots & x_{j+1} \\ \vdots & \vdots & \vdots & \vdots \\ x_i & x_{i+1} & \cdots & x_{j+i-1} \end{bmatrix} \tag{10}$$

This approach redesigns the input data of the system, creating new state variables. Those variables allow to find the best low-rank matrix, keeping the almost linear structure [24]. However, the introduction of the new variables is made at the expense of reducing the number of samples of the training data set. Hence, the number of these new variables (number of rows the Hankel matrix) a has to be a balance between the ability to detect dominant modes and the accuracy of the estimated model.

III. Case Study

A. Data Set

The data set consists of the weekly sales of a company with 2659 stores collected in a period of two and a half years (143 weeks). The DMD is used to predict the weekly sales in a period of 50 weeks ahead (about one year) in each store based on the sales data of the previous 93 weeks.

In order to guarantee that the desired data is suitable to the DMD, it was necessary to certify some periodicity. Thus, the Power Spectral Density (PSD) of the total sales volume time-series was compared for different time windows. Hence the data was splitted into two halfs and the PSD was estimated for each half and for the full data. The PSD estimates are compared in Figure 1 where it can be seen that they do not differ significantly. All estimates have a peak at a frequency corresponding to a period of approximately 25 weeks. The yearly periodicity corresponding to a frequency of 0.12 is also noticeable in the PSD estimates but with smaller power, due to the short widths of the time-domain data windows.

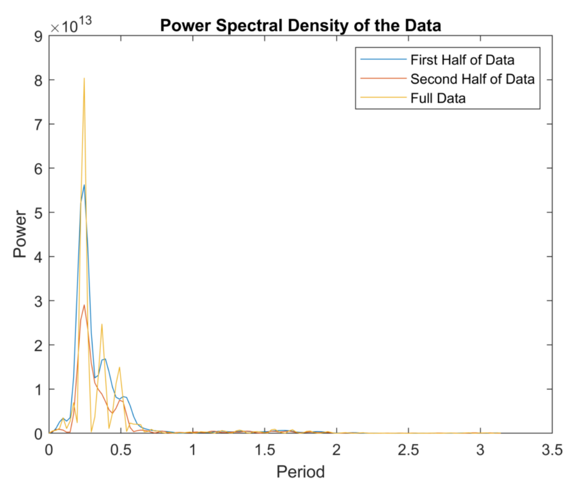

Fig. 1. Power Spectral Density of the Data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2019.2917811, IEEE Control Systems Letters

B. Forecast Quality Evaluation

The quality of the forecasts was assessed with the Best Fit Percentage (BFT) index [25], that uses the Normalised Root Mean Square Error (NRMSE), defined as

$$BFT = \left(1 - \frac{\| A_t - F_f \|}{\| A_t - \bar{A}_t \|}\right) 100\% \quad (11)$$

where F_t is the forecast, A_t is the actual value at instant t, $\bar{A}_t$ is the mean value of A_t and $\|$ indicates the 2-norm of a vector. This index was calculated for each store and measures how well the estimated model predicts the alternated component of the signal. It is not biased by the sales mean values which is very easy to predict because it remains almost stationary.

C. Sales forecasting

The DMD was firstly used with the raw data and then with an Hankel block matrix. Table I summarizes the results with both approaches. With raw data, the best BFT was achieved with a rank of 16. Figure 2 shows the actual, the estimated and the predicted total sales volume (sum of the sales of all stores). In this scenario, the DMD is as a good predictor, but it fails to follow some peaks and the last predictions converge to the mean. The BFT of the total predicted sales was of 39% as displayed in table I

TABLE I
SUMMARY OF FORECAST SALES WITH DMD

Data Set	Number of *measured signals*	Length of the training set	Reduced rank	BFT
Raw data	2659	93	16	39%
Hankel Matrix	63816	70	47	58%

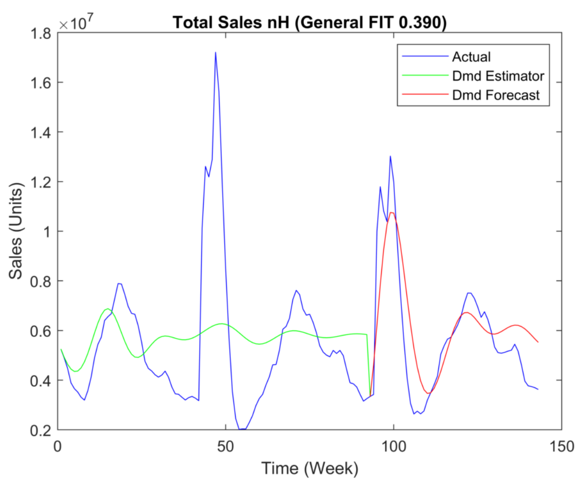

Fig. 2. Estimated and predicted total sales volume of the DMD with raw-data vs actual values.

These results were significantly improved by the use of a block Hankel matrix. More dominant modes became spanned by the augmented measured signals enabling the DMD to estimate them. This is denoted in Figure 3 where the actual values of the total sales volume is depicted in conjunction with

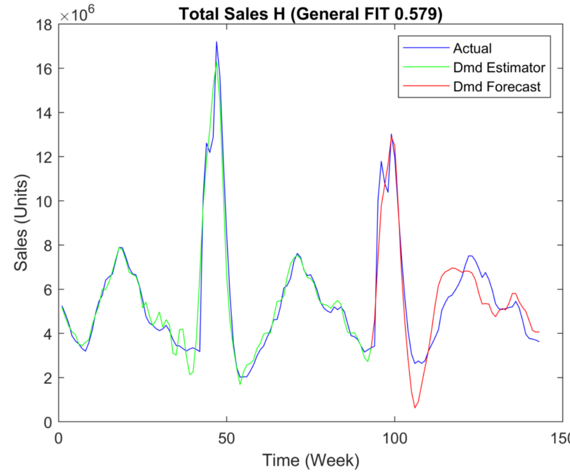

Fig. 3. Estimated and predicted total sales volume of the DMD with an Hankel Block vs actual values

the estimated and predicted values of the DMD with an Hankel block. It can be seen that both the estimates and the predictions follow the original signal. The BFT of the predicted total sales volume was increased to 58% (see Table I). Figure 4 shows the prediction errors of the DMD with and without the Hankel Matrix Block.

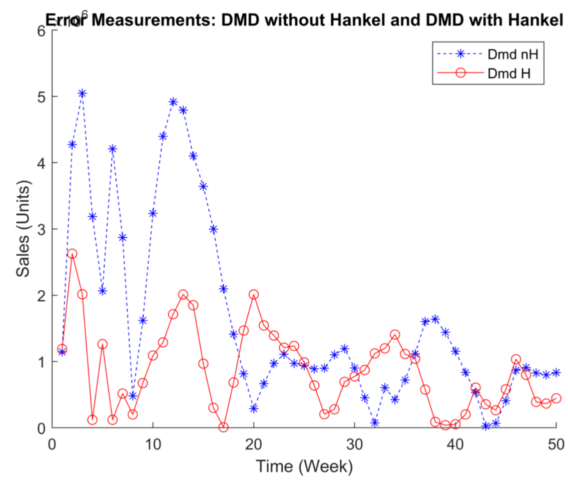

Fig. 4. Comparison between the Errors in the Data set with or without Hankel

Figure 5 displays the eigenvalues estimated by the DMD without and with an Hankel block matrix. It is possible to see

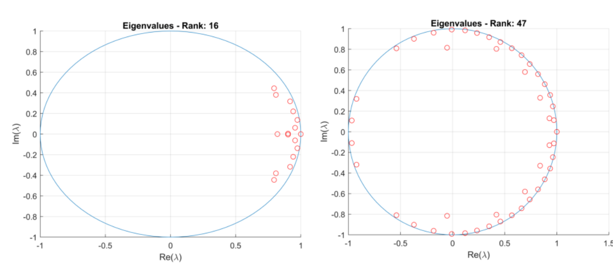

Fig. 5. The discrete-time eigenvalues estimated by the DMD with raw-data (LEFT) and using Hankel block matrix (RIGHT).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2019.2917811, IEEE Control Systems Letters

that the eigenvalues estimated with the block Hankel matrix (Figure 5 - RIGHT) are virtually on the unit circumference. There is also a pair of eigenvalues close to $1 + j0$, denoting the existence of a double integrator. This indicates a ramp profile in the data with a growing or descending tendency of the sales. Due to their large number it is impossible to perform detailed analysis of the sales of each store. Instead, histograms of the prediction BFTs in all stores are shown in Figure 6. This

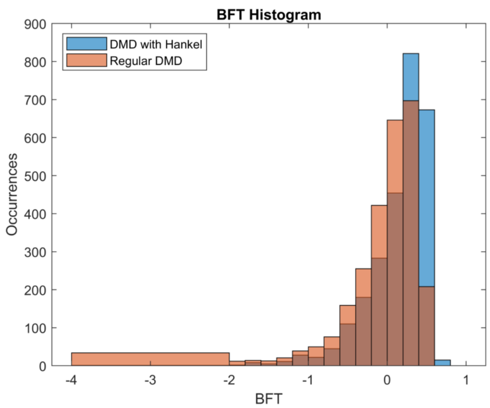

Fig. 6. Histograms of the BFTs of the Kernel predictor with raw-data and with a Block Hankel Matrix.

figure confirms the significant improvements of the Hankel Block matrix. In 43% of the stores, the predictions of the DMD with an Hankel matrix had a BFT greater than 30% while, without this matrix, only in 20% of the stores, the BFT surpassed this limit. Moreover, also without the Hankel matrix, the same BFTs were negative in 42% of the stores. With the Hankel block it was negative only in 26%. Another important conclusion drawn from the histograms is that, even with the Hankel block, the DMD did not work well in all stores. But a more detailed analysis revealed that those stores were the ones with the lowest sales volume. Hence, their contribution to the total sales is marginal. If the objective is only to estimate and predict the total sales one can think that it is sufficient to perform the DMD with the time-series of the total sales volume. In fact, such a DMD, with only one signal, spends much lower computational resources than a DMD with the 2659 signals from all stores. For this reason the DMD was performed just considering the total volume of sales. The best BFT was of 38% and was achieved with a 50 rows Hankel matrix. It is significantly lower than the 58% obtained by the full DMD (with all stores). Figure 7 compares estimated and predicted sales with the actual values. Both follow the measured signal reasonably but they are worse than full DMD estimates and predictions depicted in figure 3.

D. Comparison with RBF Neural Network

ANN predictors are widely used to forecast temporal series [5]. Hence, several ANN were also designed to predict the sales. But, due to the large number of stores, an ANN to predict all sales demands too many computational resources. Therefore the ANN was only designed for the total volume of sales. As the DMD prediction horizon is of 50 weeks, a 50 steps ahead ANN (NN50) predictor was first set up. It

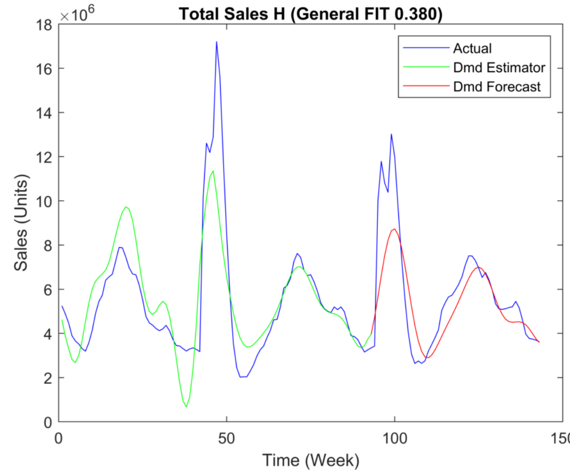

Fig. 7. Estimated and predicted total sales volume of the DMD with only one signal

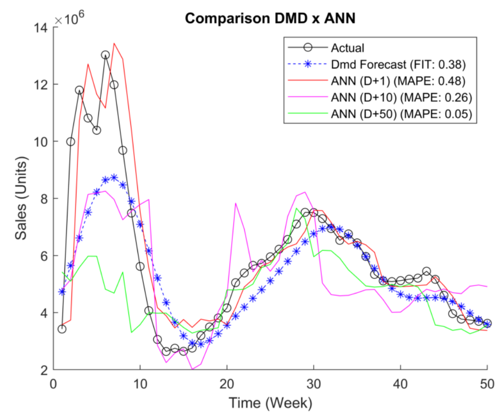

Fig. 8. ANN and DMD Estimator and Predictor.

had 1 input, and a hidden layer. It was tuned to optimize the BFT index of the prediction error over the test data. The others predictors have prediction horizons of 1 (NN1) to 10 (NN10) steps ahead. Figure 8 compares the forecasts of the ANN predictors the DMD with the total volume of sales. The DMD clearly outperforms the NN10 and NN50 and. This is confirmed by the BFT indexes: $BFT(DMD) = 38\%$, $BFT(ANN50) = 5\%$ and $BFT(ANN10) = 26\%$. The ANN1 performed slightly better than the DMD. However, it is not fair to compare these estimators because the AAN1 is fed back by all past values while the DMD does not have any feedback. Yet, the global DMD predictor, also without feedback, achieved a better BFT than the ANN1. Perhaps the 1 step forecasts of the DMD would be better, or at least, won't be worst than the ANN1 predictions. A comparison between these predictors is not shown because the reported results so far illustrate the superiority of the DMD in this case study.

IV. Conclusions and Future Work

The DMD is a recognized tool to capture spatial coherent structures associated with temporal spectral components. It can also be used to predict almost periodic temporal series from the estimation of its principal modes. But it can only estimate modes spanned by the measured signals. If their number is insufficient the DMD fails to estimate the temporal series principal modes and the predictions are very inaccurate. The number of measured signals can be artificially increased by arranging them in a Block Hankel matrix. This increases the dimension of the subspace spanned by these signal until it contains all principal modes. Hence, the DMD becomes an effective tool to forecast temporal series even for a small number of channels. In this paper the DMD predictor is used to forecast sales of big chain of stores. Although the number of stores is very large, the DMD failed to estimate some principal modes. The predictor accuracy increased significantly with an Hankel block matrix, enabling The DMD to forecast the total volume of sales for one year period with a BFT of 58% which is a very good figure for such a long period. Even with the Hankel matrix the DMD did not performed well in all stores. But those stores were the ones with the lowest sales volume. A DMD with only the total sales volume was also performed. It spends much lower computational resources but its estimates and predictions were less accurate than those of the full DMD. This DMD predictor was also compared with ANN1, ANN10 and ANN50, with predictions horizons of 1, 10 and 50 weeks. The DMD outperformed the ANN10 and ANN50. It was slightly worse than the ANN1. But the comparison of the ANN1 and DMD predictions were not fair because the ANN1 was fed back by all past values and the DMD worked without any feedback. Despite this the full DMD, also without feedback, achieved a better BFT.

The accuracy of the principal modes of the DMD with an Hankel block will addressed in a future work. In fact sometimes the Hankel matrices are badly conditioned and thus the DMD produces poor estimates. It will be studied if orthogonal or almost orthogonal matrices extracted form the Hankel block improve the DMD accuracy. It will be also considered other approaches to time-series forecasting such as Kernel based regressions.

ACKNOWLEDGMENTS

Enio Vasconcelos Filho was supported by National Funds, through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit, (UID/CEC/04234); Paulo Lopes dos Santos was and also supported by FCT within project: UID/EEA/50014/2019.

REFERENCES

[1] D. Dovžan, V. Logar, and I. Škrjanc, "Solving the sales prediction problem with fuzzy evolving methods," in 2012 IEEE Congress on Evolutionary Computation, Jun. 2012, pp. 1–8.

[2] C.-Y. Chen, W.-I. Lee, H.-M. Kuo, C.-W. Chen, and K.-H. Chen, "The study of a forecasting sales model for fresh food," Expert Systems with Applications, vol. 37, no. 12, pp. 7696–7702, Dec. 2010.

[3] P. Ramos, N. Santos, and R. Rebelo, "Performance of state space and ARIMA models for consumer retail sales forecasting," Robotics and Computer-Integrated Manufacturing, vol. 34, pp. 151–163, Aug. 2015.

[4] S. L. Lin, "A new two-stage hybrid approach of credit risk in banking industry," Expert Systems with Applications, vol. 36, no. 4, pp. 8333–8341, May 2009.

[5] M. Tkáč and R. Verner, "Artificial neural networks in business: Two decades of research," Applied Soft Computing, vol. 38, pp. 788–804, jan 2016.

[6] A. V. Devadoss and T. A. Alphonnse Ligori, "Stock Prediction Using Artificial Neural Networks," International Journal of Web Technology, vol. 002, no. 002, pp. 45–51, Dec. 2013.

[7] C. H. Aladag, M. A. Basaran, E. Egrioglu, U. Yolcu, and V. R. Uslu, "Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations," Expert Systems with Applications, vol. 36, no. 3, pp. 4228–4231, apr 2009.

[8] F. Guéniat, L. Mathelin, and L. R. Pastur, "A dynamic mode decomposition approach for large and arbitrarily sampled systems," Physics of Fluids, vol. 27, no. 2, p. 025113, Feb. 2015.

[9] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," Journal of Fluid Mechanics, vol. 656, pp. 5–28, Aug. 2010.

[10] S. Tirunagari, S. Kouchaki, N. Poh, M. Bober, and D. Windridge, "Dynamic Mode Decomposition for Univariate Time Series: Analysing Trends and Forecasting," HAL archives-ouvertes.fr, vol. hal-01463744, p. 8, 2017.

[11] J. N. Kutz, S. L. Brunton, D. M. Luchtenburg, C. W. Rowley, and J. H. Tu, "On dynamic mode decomposition: Theory and applications," Journal of Computational Dynamics, vol. 1, no. 2, pp. 391–421, Dec. 2014.

[12] K. Champion, S. L. Brunton, and J. N. Kutz, "Discovery of nonlinear multiscale systems: Sampling strategies and embeddings," arXiv preprint arXiv:1805.07411, 2018.

[13] N. Takeishi, Y. Kawahara, and T. Yairi, "Subspace dynamic mode decomposition for stochastic koopman analysis," Physical Review E, vol. 96, no. 3, Sep. 2017.

[14] T. Runolfsson, "Output dynamic mode decomposition: An extension of dynamic mode decomposition based on output functional expansions," in 2018 IEEE Conference on Decision and Control (CDC). IEEE, Dec. 2018.

[15] T. Y. Naoya Takeishi, Yoshinobu Kawahara, "Learning koopman invariant subspaces for dynamic mode decomposition," Advances in Neural Information Processing System, vol. 30, 2017.

[16] H. Arbabi and I. Mezić, "Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator," SIAM Journal on Applied Dynamical Systems, vol. 16, no. 4, pp. 2096–2126, Jan. 2017.

[17] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," Journal of Fluid Mechanics, vol. 641, p. 115, Dec. 2009.

[18] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, DMD: Data-Driven Modeling of Complex Systems, 1st ed. siam, 2016.

[19] W. E. Boyce and R. C. DiPrima, Elementary differential equations and boundary value problems, 7th ed. New York: Wiley, 2001.

[20] J. Mann and J. N. Kutz, "Dynamic mode decomposition for financial trading strategies," Quantitative Finance, vol. 16, no. 11, pp. 1643–1655, Nov. 2016.

[21] A. Edelman and H. Murakami, "Polynomial Roots From Companion Matrix Eigenvalues," Mathematics of Computation, vol. 64, no. 210, pp. 763–776, Apr. 1995.

[22] L. N. Trefethen and D. Bau III, Numerical Linear Algebra. SIAM, Jun. 1997, google-Books-ID: 4Mou5YpRD_kC.

[23] C. W. Rowley, T. Colonius, and R. M. Murray, "Model reduction for compressible flows using POD and Galerkin projection," Physica D: Nonlinear Phenomena, vol. 189, no. 1-2, pp. 115–129, Feb. 2004.

[24] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, "Hankel Matrix Rank Minimization with Applications to System Identification and Realization," SIAM Journal on Matrix Analysis and Applications, vol. 34, no. 3, pp. 946–977, Jan. 2013.

[25] L. Ljung, "System identification toolbox for use with {MATLAB}," The MathWorks, 2007.