An. Ştiinţ. Univ. Al. I. Cuza Iaşi. Mat. (N.S.)
Tomul LXVI, 2020, f. 2

Application of deterministic and randomized dynamic mode decomposition in epidemiology and fluid dynamics

Diana A. Bistrian · Gabriel Dimitriu ·
Ionel M. Navon

Abstract The present work represents a survey of modern reduced order modelling (ROM) methods developed by us with applications in epidemiology and fluid dynamics. The paper outlines the Koopman decomposition technique and the principles governing dynamic mode decomposition (DMD). We provide the description of two DMD based algorithms employed in our applications, a deterministic DMD algorithm endowed with a mode's selection criterion and a randomized DMD algorithm, respectively. We present a framework for reduced order modelling and forecasting of non-intrusive data with application to epidemiology, employing randomized DMD and ARIMA models. We provide a comparison between the two aforementioned algorithms and we perform the model order reduction of non-intrusive data originating from Saint–Venant model or shallow water equations model.

Keywords Dynamic mode decomposition · Randomized DMD · ARIMA models · Reduced order modelling · Model forecasting

Mathematics Subject Classification (2010) 81T80 · 93A30 · 60G25

1 Introduction

Over the last decades, mathematical modeling have been helpful to gain insights into the transmission dynamics of infectious diseases. The use of disease transmission models to generate short-term and long-term epidemic forecasts has increased with the rising number of emerging and re-emerging infectious disease outbreaks. Modelling of epidemiological data is even more difficult because we handle with discrete or so called *non-intrusive data*.

Identification of a mathematical model from raw data is a challenging topic in epidemiology. Many efforts were directed towards non-intrusive techniques, that do not require the projection of the governing equations onto the reduced basis modes, assuming that these governing equations or their numerical code are not known. The challenge among practitioners is obtaining a reliable approximation of disease dynamics by models of low complexity, i.e. reduced order models (ROMs).

Although most investigators studying epidemiological infections are using the existing SEIR model [24] (Susceptible-Exposed-Infectious-Removed), difficulty to calibrate this model to real data still exists. The existing models are not immediately applicable as predictive models for serious quantitative work. Additional features could be integrated into more quantitative epidemiology models that append geographical location, imperfect immunity, age and race, gender etc.

An attempt to data-driven model in epidemiology represents the use of Agent-based models (ABMs), which are a type of computer simulation composed of agents that can interact with each other and with an environment [50]. To realistically model an outbreak and to be useful in real world scenario, an ABM needs to model characteristics of a disease, as well as characteristics of the agents and their environment, all at an appropriate level of detail which requires an extensive computational effort [51]. Thus existing modeling tools have limitations both in calibration to real data and predictive features.

Researchers' interests were directed in recent years to identification of a reliable approximation of the complex behavior of discrete (also called *non-intrusive*) data by models of low complexity, i.e. reduced order models (ROM) [86], [93], [56]. This is possible given the fact that any complex system often consists of a superposition of a small number of underlying coherent structures [49]. Because these patterns have different weights in the reconstitution of the data, once the most influential coherent structures are identified, it is assumed that these contain the essential behavior of the system and the ROM is constructed by overlapping them. Reduced order modelling represents a state of the art data-driven tool [89], [76], [77], [70] among different model discovery approaches developed in recent years, including linear methods [66], [57], Koopman theory [61], [87], neural networks [31], [59] and genetic programming [22], [78]. Application of any of these tools requires various levels of mathematical expertise and computational effort.

1.1 Powerful ROM tools and related work

Among several model order reduction techniques that perform well with non-intrusive data, Proper Orthogonal Decomposition (POD) and Dynamic Mode Decomposition (DMD) are widely applied to create surrogate models of dynamics of complex systems in different applications.

The well known method of POD has been illustrated on a variety of examples ranging from turbulent flows (Wang et al. [86], Xiao et al. [90]), convection-dominated flows (Gunzburger et al. [47]), fluid mechanics (Luchtenburg and Rowley [58], Liberge and Hamdouni [56], Xiao et al. [93]), oceanography (San et al. [77], [76], Dumon et al. [39], Östh et al. [70]) to engineering applications (Bistrian and Resiga [16], Buljak and Maier [26], Mariani and Dessi [60]) or aerodynamics (Xiao et al. [91]). POD approach has been incorporated for reduced order modeling purposes by many practitioners, for exemplification see Du et al. [38], Fang et al. [42], Ştefănescu et al. [82].

POD proved to be an effective technique also in inverse problems (Winton et al. [88], Chen et al. [29], [30] and Cao et al. [27], [28]) and also in association with data assimilation techniques (Zerfas et al. [94], Dimitriu et al. [33], [34]). Recently POD was involved in harnessing the power of Machine Learning tools, especially artificial neural networks (ANNs) for reduced order modeling in the research work of Xiao et al. [92] and San et al. [1].

Koopman Mode Decomposition (KMD) [53] introduced by the French-born American mathematician B. O. Koopman, provides a theoretical background for modal decomposition, global modes analysis and hydrodynamic stability in problems describing oscillating phenomena. Once it was proved that normal modes of linear oscillations (or *shape modes*) have their natural analogs Koopman modes in the context of nonlinear dynamics [62], the Koopman theory was intensively applied for the purposes of reduced order modelling [73], [74], [30], [12], [13], [61]. Schmid and Sesterhenn [79] were the first to introduce a numerical algorithm to compute this type of modal decomposition that was called Dynamic Mode Decomposition (DMD). In a bold manner, Rowley and his coworkers [73] proved that the linearity assumption of the Koopman operator is not necessary in order to derive the DMD. Schmid [80] explored the similarities between POD and DMD and recommended a better conditioned DMD algorithm.

Multiple applications of DMD originated in the fluid dynamics community [43], [81], [14], [3], [54], [55] and recently this technique emerged into the niche fields like human robot interaction (Berger et al. [15]) and neuroscience (Brunton et al. [25]).

A considerable amount of work has focused on understanding and improving the method of dynamic mode decomposition and several variants of DMD have been released: optimized DMD (Chen et al. [30]), exact DMD (Tu et al. [84]), sparsity promoting DMD (Jovanović et al. [52]), multi-resolution DMD (Kutz et al. [55]), extended DMD (Williams et al. [87]), recursive dynamic mode decomposition (Noack et al. [68]), DMD with control (Proctor et al. [71]), randomized low-rank DMD (Erichson and Donovan [40], Erichson et al. [41]), adaptive randomized DMD (Bistrian and Navon [17]).

A comparison of DMD vs. POD in reduced order modelling was illustrated in our previous paper [18], for the study of shallow water equations model. Although spatial basis functions for DMD and POD respectively, offer an insight of the coherent structures in the flow field, there are several major differences between these two ROM techniques. POD computes a set of orthonormal basis vectors associated with the most influential modes by which the given data is characterized, which are energetically ranked. The most energetic POD modes are selected to generate the ROM. The advantage of the DMD modes compared to POD modes is that each shape mode is associated with a pulsation, a growth rate and each mode has a single distinct frequency. Selection of Koopman modes used for the flow reconstruction constitutes the source of many discussions among modal decomposition practitioners. It was reported that neither the selection of the modes based on their amplitude, nor the selection based on the frequency are certain to lead to the finding of the dominant modes [67], [83].

One of the most challenging problems is represented by the modeling of biological phenomena, due to the great number of parameters involved in the dynamics. The control theory combined with modelling tools has been increasingly used to describe some of the most important features of age-dependent population dynamics. New trends in the control theory are based on shape optimization problems using regional controls in spatially explicit models (e.g. the work provided by the efforts of Aniţa et al. [5], [6], [7], [8], [4], [9]).

1.2 Contribution to POD-DMD-ROM models

Several procedures for selecting the most influential modes in dynamic mode decomposition can be found in our previous papers and we'll briefly mention them. In [18] we proposed a framework for modal decomposition of $2D$ flows, when numerical data are captured with large time steps. We arranged the modes in descending order of their energy weighted by the inverse of the Strouhal number. The modes that contribute weakly to the data sequence were eliminated based on the conservation of quadratic integral invariants [65] by the reduced order flow. In [2] we proposed a new framework for dynamic mode decomposition based on the reduced Schmid operator. We investigated a variant of DMD algorithm and we explored the selection of the modes based on sorting them in decreasing order of their amplitudes. This procedure works well for models without modes that are very rapidly damped, having very high amplitudes. Therefore the selection of modes based on their amplitude is effective only in certain situations.

In [19] we focused on the effects of modes selection in dynamic mode decomposition. We proposed a dynamic filtering criterion for which the amplitude of any mode is weighted by its growth rate, resulting in identification of dynamically relevant flow features of time-resolved numerical data.

Although the reported results were satisfactory, defining a DMD modes'selection criterion requires additional attention in the offline stage of the algorithm and increases the CPU time. It is more desirable to reduce the problem dimension to avoid a computationally expensive singular value decomposition of data (SVD) that plays a central role in computing the DMD. We have introduced in [17] the procedure of randomization of data prior to singular value decomposition. Thus, we endowed the DMD algorithm with a randomized SVD function, aiming to improve the accuracy of the reduced order linear model and to reduce the CPU time. A reduced DMD basis is obtained in the offline stage, that does not require a dominant mode selection algorithm. We proved the efficiency of randomized dynamic mode decomposition for reduced order modelling in Bistrian and Navon [20].

We pioneered in [21] the application of an enhanced adaptive predictive model for malaria disease. We proved the efficiency of this research avenue for epidemiological data, with several major advantages: the prediction is performed in time for a short-medium period, over any period of time, with small errors; the prediction is performed for all geographical locations simultaneously; the tendency of epidemiologic dynamics is correctly captured.

High dimensionality continues to represent a challenge in computational systems biology. The model reduction method has been analyzed in mathematical modelling of biological systems.

An application of DEIM combined with POD to provide dimension reduction of a system of two nonlinear partial differential equations describing the spatio–temporal dynamics of a predator-prey community was presented by Dimitriu et al. [35]. The same DEIM algorithm combined with POD has been implemented in [36] to provide dimension reduction of a model describing the aggregative response of parasitoids to hosts in a coupled multi-species system.

A comparative analysis using three reduced-order strategies, Missing Point Estimation (MPE), Gappy POD method and Discrete Empirical Interpolation Method (DEIM), applied to a biological model describing the spatio–temporal dynamics of a predator–prey community was performed in [37]. An optimal control problem was studied in [10] for a reaction–diffusion system that models an ecosystem composed by one predator and two prey populations. A prey–predator model defined by an initial–boundary value problem whose dynamics is described by a Holling type III functional response was introduced in [11]. A numerical approximation of the system was carried out with a spectral method coupled with the fourth-order Runge–Kutta time solver and the biological relevance of the comparative numerical results were also presented.

The present work represents a survey of modern methods of ROM developed by us with applications in epidemiology and fluid dynamics. The remainder of the manuscript is organized as follows: Section 2 outlines the Koopman decomposition technique and the principles governing the DMD. We provide the description of two DMD based algorithms employed in our applications. A deterministic DMD algorithm and a randomized DMD algorithm, respectively. Section 3 harnesses the application of DMD-ROM to epidemiological data. In Section 4 we provide a comparison between the two aforementioned algorithms and we perform the model order reduction of non-intrusive data originating from Saint-Venant model or shallow water equations model. Summary and conclusions are drawn in the final section.

2 Reduced order modelling based on Dynamic Mode Decomposition

2.1 The key steps of Dynamic Mode Decomposition

Dynamic mode decomposition (DMD) is a data processing tool which is applied to numerical or experimental data, in order to identify the coherent structures of dynamics or for the purpose of surrogate modelling. In the present paper we apply the method of DMD for an efficient reduced order modelling of numerical data. We outline in the following the key steps of DMD.

We proceed by collecting data $r_i\left(t,\mathbf{x}\right)=r\left(t_i,\mathbf{x}\right),\ t_i=i\Delta t,\ i=0,...,N,$ at a constant sampling time Δt, $\mathbf{x}$ representing the spatial coordinates whether Cartesian or cylindrical. The two integer parameters N and M involved in the process of data acquisition represent, respectively:
$N+1=$ total number of snapshots taken in time,
$M=$ number of spatial measurements per time snapshot.

We form a data matrix whose columns represent the individual data samples, called *the snapshot matrix*

$$V=\left[\,r_0\;r_1\;...\;r_N\,\right]\in\mathbb{R}^{M\times(N+1)}. \tag{2.1}$$

Each column r_i is a vector with M components, representing the numerical measurements. For simplicity of description, we consider here real data $r_i\in\mathbb{R}^M$.

The Koopman decomposition assumed that a propagator matrix $\mathcal{A}$ exists, that maps every vector column onto the next one

$$\left\{r_0,\ r_1=\mathcal{A}r_0,\ r_2=\mathcal{A}r_1=\mathcal{A}^2r_0,.\,..,\ r_N=\mathcal{A}r_{N-1}=\mathcal{A}^Nr_0\right\}. \tag{2.2}$$

The DMD algorithm constructs the best approximation of the propagator matrix $\mathcal{A}$. The next computational step consists in forming two data matrices from the snapshot sequence. A matrix V_0^{N-1} is formed with the first N columns and the matrix V_1^N contains the last N columns of V:

$$V_0^{N-1}=\left[\,r_0\;r_1\;...\;r_{N-1}\,\right]\in\mathbb{R}^{M\times N},\ V_1^N=\left[\,r_1\;r_2\;...\;r_N\,\right]\in\mathbb{R}^{M\times N}. \tag{2.3}$$

For a sufficiently long sequence of the snapshots, we suppose that the last snapshot v_N can be written as a linear combination of previous N vectors, such that

$$r_N=c_0r_0+c_1r_1+...+c_{N-1}r_{N-1}+\mathcal{R}, \tag{2.4}$$

in which $\mathrm{c}_i\in\mathbb{R},\mathrm{i}=0,...,\mathrm{N}-1$ and $\mathcal{R}$ is the residual vector. We assemble the following relations

$$\left\{r_1,r_2,...r_N\right\}=\mathcal{A}\left\{r_0,r_1,...r_{N-1}\right\}=\left\{r_1,r_2,...,V_0^{N-1}c\right\}+\mathcal{R}, \tag{2.5}$$

where $c=\left(\,c_0\;c_1\;...\;c_{N-1}\,\right)^T$ is the unknown column vector.

In matrix notation form, Eq. (2.5) reads

$$\mathcal{A}V_0^{N-1} = V_0^{N-1}\mathcal{S} + \mathcal{R}, \quad \mathcal{S} = \begin{pmatrix} 0 & \dots & 0 & c_0 \\ 1 & & 0 & c_1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & c_{N-1} \end{pmatrix}, \tag{2.6}$$

where $\mathcal{S}$ is the companion matrix.

Relation (2.6) is true when the residual

$$\mathcal{R} = r_N - V_0^{N-1}c, \tag{2.7}$$

is minimized when c is chosen such that $\mathcal{R}$ is orthogonal to

$$span\left\{r_0, ..., r_{N-1}\right\}.$$

.

The goal of DMD algorithm is to solve the eigenvalue problem of the companion matrix $\mathcal{S}$

$$V_1^N = \mathcal{A}V_0^{N-1} = V_0^{N-1}\mathcal{S} + \mathcal{R}, \tag{2.8}$$

where $\mathcal{S}$ approximates the eigenvalues of $\mathcal{A}$ when $\|\mathcal{R}\|_2 \to 0$.

The objective at this step is to solve the minimization problem

$$\underset{\mathcal{S}}{Minimize}\ \mathcal{R} = \left\|V_1^N - V_0^{N-1}\mathcal{S}\right\|_2. \tag{2.9}$$

An estimate can be computed by multiplying V_1^N by the Moore-Penrose pseudoinverse of V_0^{N-1}:

$$\mathcal{S} = \left(V_0^{N-1}\right)^+ V_1^N, \tag{2.10}$$

where $\left(V_0^{N-1}\right)^+$ is computed according to Moore-Penrose pseudoinverse definition [45].

Following [80], we developed an alternate algorithm based on Singular Value Decomposition (SVD) of snapshot matrix V_0^{N-1}. This approach is helpful when the matrix V_0^{N-1} is rank deficient ($M > N$). The minimization problem (2.9) has the following solution. We first identify a singular value decomposition of V_0^{N-1}:

$$V_0^{N-1} = U\Sigma W^H, \tag{2.11}$$

where U contains the proper orthogonal modes of V_0^{N-1}, Σ is a square diagonal matrix containing the singular values of V_0^{N-1} and W^H is the conjugate transpose of W.

Relations $\mathcal{A}V_0^{N-1} = V_1^N = V_0^{N-1}S + \mathcal{R}, \|\mathcal{R}\|_2 \to 0$ and $V_0^{N-1} = U\Sigma W^H$ yield:

$$\begin{aligned} \mathcal{A}U\Sigma W^H = V_1^N = U\Sigma W^H S \\ \Rightarrow \quad U^H \mathcal{A}U\Sigma W^H = U^H U\Sigma W^H S \\ \Rightarrow \quad S = U^H \mathcal{A}U. \end{aligned} \tag{2.12}$$

From $\mathcal{A}U\Sigma W^H = V_1^N$ it follows that $\mathcal{A}U = V_1^N W\Sigma^{-1}$ and hence

$$S = U^H \left(V_1^N W\Sigma^{-1}\right). \tag{2.13}$$

A direct consequence of solving the minimization problem (2.9) is that decreasing the residual increases overall convergence and therefore the eigenvalues λ_j and the eigenvectors $\phi_j, j = 1, ..., N$ of $\mathcal{S}$ will converge toward the eigenvalues and the eigenvectors of the Koopman operator $\mathcal{A}$, respectively. Koopman operator has infinite number of eigenvalues while DMD is linear and has finite no of terms. More specifically, every column vector $r_i, i = 1, ..., N$ can be written as a linear combination of its predecessor:

$$r_i = \mathcal{A}v_{i-1} = \ldots = \mathcal{A}^{i-1}r_1, \quad i = 1, ..., N. \tag{2.14}$$

The eigenvectors of $\mathcal{S}$ form a basis for the span of $\mathcal{A}$, therefore, we can write every column vector as a linear combination of the eigenvectors

$$\begin{aligned} r_1 &= \sum_{j=1}^{N} a_j \phi_j, \\ r_i &= \sum_{j=1}^{N} \mathcal{A}^{i-1} a_j \phi_j, \quad i = 1, ..., N, \\ r_i &= \sum_{j=1}^{N} a_j \lambda_j^{i-1} \phi_j, \quad i = 1, ..., N. \end{aligned} \tag{2.15}$$

A straightforward interpretation of relations (2.15) brings the data snapshots at every time step $\{\mathrm{t}_1,...,\mathrm{t}_N\}$ as a linear combination of DMD modes according to

$$r(t_i, \mathbf{x}) = \sum_{j=1}^{N} a_j \phi_j(\mathbf{x}) \lambda_j^{i-1}, \quad i \in \{1, ..., N\}, \quad t_i \in \{\mathrm{t}_1, ..., \mathrm{t}_N\}, \tag{2.16}$$

or in matrix formulation:

$$\begin{aligned} V_1^N &= [\, r_1\ r_2\ ...\ r_N \,] = \\ &= [\, \phi_1\ \phi_2\ ...\ \phi_N \,] \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_N \end{pmatrix} \begin{pmatrix} 1 & \lambda_1^1 & \lambda_1^2 & \ldots & \lambda_1^{N-1} \\ 1 & \lambda_2^1 & \lambda_2^2 & \ldots & \lambda_2^{N-1} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_N^1 & \lambda_N^2 & \ldots & \lambda_N^{N-1} \end{pmatrix}, \end{aligned} \tag{2.17}$$

where the right eigenvectors of $\mathcal{S}$, $\phi_j \in \mathbb{C}$ are dynamic *shape* (or Koopman) modes, the eigenvalues of $\mathcal{S}$, λ_j are called Ritz values [32] and coefficients $a_j \in \mathbb{C}$ are denoted as amplitudes or Koopman eigenfunctions. Each Ritz value $\lambda_j = e^{(\sigma_j + i\omega_j)\Delta t}$ is associated with the growth rate $\sigma_j = \frac{\log(|\lambda_j|)}{\Delta t}$ and the frequency $\omega_j = \frac{\arg(|\lambda_j|)}{\Delta t}$.

The modes' selection plays a central role in model reduction. The superposition of all Koopman modes, weighted by their amplitudes and complex frequencies, approximates the entire data sequence, but there are also modes that have a weak contribution. Our goal is to produce a reduced order model (ROM) of the data involving only the most significant modes, having a strong contribution to the data representation, which we are calling *leading modes*.

Thus, the data snapshots at every time step $\{\mathrm{t}_1,...,\mathrm{t}_N\}$ will be represented as a linear combination of the leading DMD modes according to

$$r_{DMD}\left(t_i, \mathbf{x}\right) = \sum_{j=1}^{k} a_j \phi_j\left(\mathbf{x}\right) \lambda_j^{i-1}, \quad i \in \{1, ..., N\}, \quad t_i \in \{\mathrm{t}_1, ..., \mathrm{t}_N\}, \tag{2.18}$$

where k represents the number of DMD modes involved in reconstruction of data snapshots, or in matrix formulation

$$\begin{aligned} V_1^N &= \left[\, r_1 \; r_2 \; ... \; r_N \,\right] = \\ &= \left[\, \phi_1 \; \phi_2 \; ... \; \phi_k \,\right] \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \vdots & \\ & & & a_k \end{pmatrix} \begin{pmatrix} 1 & \lambda_1^1 & \lambda_1^2 & \dots & \lambda_1^{N-1} \\ 1 & \lambda_2^1 & \lambda_2^2 & \dots & \lambda_2^{N-1} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_k^1 & \lambda_k^2 & \dots & \lambda_k^{N-1} \end{pmatrix}. \end{aligned} \tag{2.19}$$

Here we point out that the k leading modes involved in ROM representation of data (2.19) are not the first k modes from representation (2.17). The leading modes represent a subset of DMD modes that will be selected from all computed DMD modes via several criteria, which will be the subject of discussion in the next section.

It can be seen that the Singular Value Decomposition plays a central role in computing the DMD. We developed two dynamic mode decomposition algorithms based on SVD. The first consists in a deterministic DMD method endowed with a vector filtering criterion to select the most influential modes. The second algorithm utilizes an adaptive randomized dynamic mode decomposition to obtain the leading modes in the offline stage, that does not require an additional selection algorithm of the DMD modes. The two algorithms are presented in the following subsections.

2.2 Deterministic Dynamic Mode Decomposition with modes selection criterion (CrDMD)

We recently investigated different techniques of modes selection in DMD. Ref. [18] aimed to present a preliminary survey on DMD modes selection. We proposed a framework for modal decomposition of $2D$ flows, when numerical data are captured with large time steps. Key innovations for the DMD-based ROM introduced in [18] are the use of the Moore-Penrose pseudoinverse in the DMD computation that produced an accurate result and a novel selection method for the DMD modes. We eliminate the modes that contribute weakly to the data sequence based on the conservation of quadratic integral invariants [63] by the reduced order flow.

In [2] we proposed a new framework for dynamic mode decomposition based on the reduced Schmid operator. We investigated a variant of DMD algorithm and we explored the selection of the modes based on sorting them in decreasing order of their amplitudes. This procedure works well for models without modes that are very rapidly damped, having very high amplitudes. Therefore the selection of modes based on their amplitude is effective only in certain situations, as reported also by Noack et al. [67].

In [19] we focused on the effects of modes selection in dynamic mode decomposition. The algorithm related in [19] proposed a dynamic filtering criterion for which the amplitude of any mode is weighted by its growth rate. This method proved to be perfectly adapted to the flow dynamics, resulting in identification of the most influential modes for the swirling flow investigated problems.

The first DMD algorithm we address in this survey is a deterministic (or classic) DMD together with a vector filtering criterion for selection of DMD modes involved in the ROM.

We define the amplification of any DMD mode as

$$A_j = \frac{1}{T}\int_0^T a_j\left(\lambda_j^{t/\Delta t} + \lambda_j^{-t/\Delta t}\right)dt =$$

$$= \frac{a_j}{\sigma_j T}\left(e^{\sigma_j T} + e^{-\sigma_j T} - 2\right), \quad j = 1,\ldots,N, \quad T = (N-1)\Delta t, \tag{2.20}$$

where λ_j are the Ritz values, $a_j \in \mathbb{C}$ are the modal amplitudes and $\sigma_j = \frac{\log(|\lambda_j|)}{\Delta t}$ represent the growth rates.

We define the relative error of the low-rank model as the L_2-norm of the difference between the flow variables and approximate DMD solutions over the exact one, that is,

$$Er_{DMD} = \frac{\left\|r(\mathbf{x}) - r_{DMD}(\mathbf{x})\right\|_2}{\left\|r(\mathbf{x})\right\|_2}, \tag{2.21}$$

where $r(\mathbf{x})$ represents the numerical data and $r_{DMD}(\mathbf{x})$ is the low-rank DMD approximation.

We retain dynamic modes and associated frequencies in descending order of their amplification defined by (2.20) until a minimum relative error of reconstruction is achieved. To produce the reduced order model amounts to finding the solution to the following minimization problem:

$$\begin{cases} \underset{k\in\mathbb{N},k\geq 2}{Find} \quad r_{DMD}\left(t_i,\mathbf{x}\right)=\sum\limits_{j=1}^{k} a_j\phi_j\left(\mathbf{x}\right)\lambda_j^{i-1}, i\in\left\{1,...,N\right\}, t_i\in\left\{\mathrm{t}_1,...,\mathrm{t}_N\right\}, \\ Subject\ to \quad \arg\min\limits_{k}\left\{A_1>A_2>...>A_k,\ Er_{DMD}\leq\varepsilon\right\}, \varepsilon=10^{-5}. \end{cases} \tag{2.22}$$

Consequently, the modes and frequencies that have strongest influence on the quality of approximation are selected to be included in the reduced order model. The CrDMD algorithm proceeds as follows:

Algorithm 1 (CrDMD): Deterministic DMD algorithm with modes selection criterion

---

Initial data: $V_0^{N-1}\in\mathbb{R}^{M\times N}$, $V_1^N\in\mathbb{R}^{M\times N}$.

1. Produce the economy-size singular value decomposition, where U contains the proper orthogonal modes of V_0^{N-1} and Σ contains the singular values:
$$[U,\Sigma,W]=SVD\left(V_0^{N-1}\right).$$
2. Solve the minimization problem (2.9): $S=U^H\left(V_1^N W\Sigma^{-1}\right)$.
3. Compute dynamic modes solving the eigenvalue problem $SX=X\Lambda$ and obtain dynamic modes as $\Phi=UX$. The diagonal entries of Λ represent the eigenvalues λ.
4. Project dynamic modes onto the first snapshot to calculate the vector containing dynamic modes amplitudes $Ampl=\left(a_j\right)_{j=1}^{rank(\Lambda)}$.
5. Solve the minimization problem (2.22) and obtain the rank k. Retain dynamic modes and associated frequencies in descending order of their amplification defined by (2.20).
6. The reconstructed data at every time step $\{\mathrm{t}_1,...,\mathrm{t}_N\}$ involving the selected DMD modes is given by the product
$$\begin{aligned} V^{DMD} &= \Phi\cdot diag\left(Ampl\right)\cdot Van = \\ &= \left[\,\phi_1\ \phi_2\ ...\ \phi_k\,\right]\begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \vdots & \\ & & & a_k \end{pmatrix}\begin{pmatrix} 1 & \lambda_1^1 & \lambda_1^2 & \dots & \lambda_1^{N-1} \\ 1 & \lambda_2^1 & \lambda_2^2 & \dots & \lambda_2^{N-1} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_k^1 & \lambda_k^2 & \dots & \lambda_k^{N-1} \end{pmatrix} \end{aligned} \tag{2.23}$$

Output: k, V_{DMD}.

Consequently, the modes and frequencies that have strongest influence on the quality of approximation are selected to be included in the reduced order model.

2.3 Adaptive randomized Dynamic Mode Decomposition (ARDMD)

The Singular Value Decomposition (SVD) plays a central role in computing the DMD. Therefore, the Moore-Penrose pseudoinverse approach we previously employed in [18] might not be feasible when dealing with high dimensional non-intrusive data. It is more desirable to reduce the problem dimension to avoid a computationally expensive SVD. We have introduced in [17] the procedure of randomization of data prior to singular value decomposition. Thus, we endow the DMD algorithm with a randomized SVD function adapted after [48], aiming to improve the accuracy of the reduced order linear model and to reduce the CPU time.

We developed a randomized dynamic mode decomposition as a fast and accurate option in model order reduction of non-intrusive data. To the best of our knowledge, the paper [17] was the first to introduce the randomized dynamic mode decomposition algorithm with application to fluid dynamics, after the randomized SVD algorithm recently introduced in [40] for processing of high resolution videos.

A straightforward interpretation of this technique brings the data snapshots at every time step $\{\mathrm{t}_1,...,\mathrm{t}_{N_t}\}$ as a linear combination of DMD modes, such that the ROM for each data sequence can be written according to

$$r_{DMD}\left(t_i,x\right)=\sum_{j=1}^{N_{DMD}}\underbrace{a_j\left(t_i\right)}_{Amplitude}\underbrace{\phi_j\left(x\right)}_{Spatial\ modes}\quad,\quad t_i\in\left\{\mathrm{t}_1,...,\mathrm{t}_{N_t}\right\},\tag{2.24}$$

where $\phi_j \in \mathbb{C}$ are dynamic Koopman) modes,
$a_j\left(t_i\right)=\widetilde{a}_j\lambda_j^{i-1}$, $i\in\left\{1,...,N_t\right\}$, $j\in\left\{1,...,N_{DMD}\right\}$ are modal amplitudes and N_{DMD} represents the number of the DMD modes that we include in the reduced order model.

The objective of the DMD based reduced order model (denoted in the following as DMD-ROM) is to represent with high fidelity the non-intrusive epidemiological data using dynamics given by the DMD modes. Our goal is to produce a reduced order model involving only the most significant modes, providing a strong contribution to the data representation, which we are calling the *leading modes*.

The main task amounts to finding the optimal rank N_{DMD} of the randomized SVD function, by formulating restriction optimisation issues. This is equivalent to finding the solution to one of the following constrained optimization problems

$$\begin{cases} \underset{N_{DMD}\in\mathbb{N},N_{DMD}\geq 2}{Find} \quad r_{DMD}\left(t_i,x\right)=\sum\limits_{j=1}^{N_{DMD}} a_j\left(t_i\right)\phi_j\left(x\right), \quad t_i\in\left\{\mathrm{t}_1,...,\mathrm{t}_{N_t}\right\}, \\ Subject\ to\ \ N_{DMD}=\arg\min\left\{C_{DMD}\geq 0.999\right\}, \end{cases} \tag{2.25}$$

$$\begin{cases} \underset{N_{DMD}\in\mathbb{N},N_{DMD}\geq 2}{Find} \quad r_{DMD}\left(t_i,x\right)=\sum\limits_{j=1}^{N_{DMD}} a_j\left(t_i\right)\phi_j\left(x\right), \quad t_i\in\left\{\mathrm{t}_1,...,\mathrm{t}_{N_t}\right\}, \\ Subject\ to\ \ N_{DMD}=\arg\min\left\{E_{DMD}\leq \varepsilon\right\}, \end{cases} \tag{2.26}$$

where C_{DMD} is the correlation coefficient of the low-rank DMD model

$$C_{DMD}=\left\langle \frac{\left(\left\|r\left(x,t\right)\cdot r_{DMD}\left(x,t\right)\right\|_2\right)^2}{\left\|r(x,t)^H\cdot r\left(x,t\right)\right\|_2\left\|r_{DMD}(x,t)^H\cdot r_{DMD}\left(x,t\right)\right\|_2}\right\rangle_T, \tag{2.27}$$

where $r\left(x,t\right)$ means the numerical data, $r_{DMD}\left(x,t\right)$ represent the computed solution by means of the reduced order DMD model, $(\cdot)$ represents the Hermitian inner product and H denotes the conjugate transpose, $\langle\cdot\rangle_T$ is the time average operator defined by Eq. (2.28).

$$\left\langle f\left(t\right)\right\rangle_T=\frac{1}{N_t}\sum_{i=1}^{N_t} f\left(t_i\right), \quad t_i\in\left\{t_1,t_2,...,t_{N_t}=T\right\}. \tag{2.28}$$

E_{DMD} is the error given by the DMD-ROM defined by

$$E_{DMD}=\langle\left\|r\left(x,t\right)-r_{DMD}\left(x,t\right)\right\|_2\rangle_T, \tag{2.29}$$

$\|\cdot\|_2$ is the L_2-norm of $\mathbb{R}^{N_x}$, Parameter ε in Eq. (2.26) represents a constant that sets the maximum admissible limit to the error of data reconstruction. The accuracy of the numerical procedure can be adjusted according to the value chosen for this parameter. Usually, we set this value to $\varepsilon=10^{-5}$.

The adaptive randomized DMD algorithm (**Algorithm 2**) proceeds as follows:

Algorithm 2 (ARDMD): Adaptive randomized DMD algorithm

Initial data: $V_0^{N-1}\in\mathbb{R}^{M\times N}$, $V_1^N\in\mathbb{R}^{M\times N}$, $M\geq N$, integer target rank $k\geq 2$ and $k<N$.

1. For $k=2$ to $N-1$.

2. Produce the randomized singular value decomposition:
$$[U, \Sigma, W] = \mathbf{RSVD}\left(V_0^{N-1}, k\right),$$
where U contains the proper orthogonal modes of V_0^{N-1} and Σ contains the singular values. The RSVD function is described in continuation of this algorithm.
3. Solve the minimization problem (2.9): $S = U^H\left(V_1^N W \Sigma^{-1}\right)$.
4. Solving the eigenvalue problem $[X, \Lambda] = eig\,(S)$, obtain dynamic modes as $\Phi = UX$. The diagonal entries of Λ represent the eigenvalues λ.
5. Project dynamic modes onto the first snapshot to calculate the vector containing dynamic modes amplitudes $Ampl = (a_j)_{j=1}^{rank(\Lambda)}$.
6. The DMD model of rank k is given by the product
$$\begin{aligned} V^{DMD} &= \Phi \cdot diag\,(Ampl) \cdot Van = \\ &= [\,\phi_1\ \phi_2\ ...\ \phi_k\,] \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \vdots & \\ & & & a_k \end{pmatrix} \begin{pmatrix} 1 & \lambda_1^1 & \lambda_1^2 & \dots & \lambda_1^{N-1} \\ 1 & \lambda_2^1 & \lambda_2^2 & \dots & \lambda_2^{N-1} \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda_k^1 & \lambda_k^2 & \dots & \lambda_k^{N-1} \end{pmatrix} \end{aligned} \tag{2.30}$$
7. Solve the optimization problem (2.25) or (2.26) and obtain the lowest rank k and the N_{DMD} leading DMD modes.
8. Finalize the reduced order model at every time step $\{\mathrm{t}_1, ..., \mathrm{t}_N\}$, involving the leading k DMD modes, with the product given by Eq. (2.30).
Output: N_{DMD}, V_{DMD}.

Randomized SVD function (RSVD):

Initial data: $V_0^{N-1} \in \mathbb{R}^{M\times N}$, $M \geq N$, integer target rank $k \geq 2$ and $k < N$.

1. Generate random test matrix $M = rand\,(N, r)$, $r = \min\,(N, 2k)$.
2. Compute sampling matrix by multiplication of snapshot matrix with random matrix $Q = V_0^{N-1}M$.
3. Orthonormalization of sampling matrix via Gram–Schmidt orthonormal method $Q \leftarrow GramSchmidt\,(Q)$.
4. Projection of snapshot matrix to smaller space $V = Q^H V_0^{N-1}$, where H denotes the conjugate transpose.
5. Produce the economy-size singular value decomposition of low-dimensional snapshot matrix $[Q_1, \Sigma, W] = SVD\,(V)$.
6. Compute the right singular vectors $U = QQ_1$.
Output: Procedure returns $U \in \mathbb{R}^{M\times k}$, $\Sigma \in \mathbb{R}^{k\times k}$, $W \in \mathbb{R}^{N\times k}$.

2.4 Short-medium term prediction of epidemiologic dynamics using DMD-ROM and ARIMA models

As we have seen so far, the reduced order model (2.24) is built-up of two components. The first one is represented by stationary DMD modes and the second one is represented by time-dependent modal amplitudes. Our strategy in prediction of epidemiologic dynamics consists of two stages. First, we will use the ARIMA (AutoRegressive Integrated Moving Average) models to a high-fidelity fitting of the amplitudes of DMD-ROM model. In the second step, we will forecast in time the fitted ARIMA models of amplitudes and we will assemble the predicted models of epidemiological data.

ARIMA models represent an adaptation of discrete-time filtering methods [23] that are used for fitted time series data, to better understand the data or to predict future points in the series, i.e. forecasting. In general form, the equation for the predicted value of amplitude a in period t, based on data observed up to period $t-1$ is

$$a_t^{fitted} = \underbrace{\mu}_{const.} + \underbrace{\varphi_1 a_{t-1} + ... + \varphi_p a_{t-p}}_{AR\ terms} \underbrace{-\theta_1 e_{t-1} - ... - \theta_q e_{t-q}}_{MA\ terms}. \tag{2.31}$$

Usually, an ARIMA model is denoted by ARIMA(p, d, q). The autoregressive (AR) part of ARIMA indicates that the evolving variable is regressed on its own lagged values, so parameter p is the order (or number of time lags) of the autoregressive model. The moving average (MA) part indicates that the regression error is expressed as a linear combination of actual and previous terms, so parameter q is the order of moving average model. The I stands for "integrated" and indicates that the data have been differenced once or more times. Its role is to eliminate the dependence on the time at which the series is observed. Parameter d represents the degree of differencing.

In order to forecast in time the model amplitudes and to assemble the predicted dynamics of epidemiological data, we have created the following algorithm using a *forecasting sliding window technique*, that consists of the following main steps:

1. We consider an input data vector denoted *sliding vector*, consisting in a small part of epidemiological data. Using the *sliding vector* we compute the reduced order DMD-ROM model.
2. We consider an input data vector denoted *forecast vector*, consisting in DMD-ROM amplitudes.
3. Optimal ARIMA(p, d, q) model is found to fit the *forecast vector*, when we look for the indices $p \in \{1, 2, 3\}$, $d \in \{1, 2\}$ and $q = 1$. The optimal ARIMA model is the one minimizing the square errors $\min \sum_j \left(a_j - a_j^{fitted} \right)^2$, where a_j represents the DMD amplitudes and a_j^{fitted} are the corresponding ARIMA fitted values.

4. The amplitudes are forecasted one step ahead using the corresponding ARIMA model.
5. The predicted epidemiological data are computed using the DMD modes according to

$$r_{forecast}(t,x) = \sum_{j=1}^{N_{DMD}} \phi_j\left(x\right) a_j^{forecast}\left(t\right), \tag{2.32}$$

   where $\phi_j\left(x\right)$ are DMD modes and $a_j^{forecast}$ represent the forecasted modal amplitudes.
6. The first point is dropped, the forecasted point is added to current *sliding vector*.
7. Go to **Step 1**.

3 Application of DMD-ROM models in epidemiology

3.1 Processing malaria disease data

In order to apply and test the proposed DMD-ROM method and the prediction technique to epidemiological data, we consider the following initial data:

$N_x = 137$, representing the number of geo-spatial locations from which the epidemiological data were extracted;

$N_t = 27$, i.e. each epidemiological vector has 27 recordings, concerning malaria infection rates;

$\Delta t = 1$ i.e. time step is one year.

The length of *sliding vector* is 10, i.e. 10 raw data in each data set are used to obtain the reduced order DMD-ROM model;

The length of *forecast vector* is 7, i.e. last 7 amplitudes are used to obtain the ARIMA model to forecast the DMD-ROM.

The raw epidemiological data collected yearly from 1990 to 2016 along 137 geo-spatial locations is illustrated in Figure 1.

Adaptive randomized dynamic mode decomposition algorithm (ARDMD) that we applied, produces a significantly reduced size spectrum which incorporates the dominant DMD modes. The optimal rank of the reduced DMD model is the unique solution of the optimization problem (2.25). By employing simulated annealing technique [69] we obtained the solution to the optimization problem (2.25).

This leads to the optimal low rank N_{DMD} and associated DMD subspace where the most influential DMD bases live. The rank of the reduced DMD-ROM model is automatically found such that the error given by the DMD-ROM defined by Eq. (2.29) becomes sufficiently small. Figure 2 presents an insight of how the proposed randomized DMD algorithm finds the number of leading DMD modes $N_{DMD} = 9$.

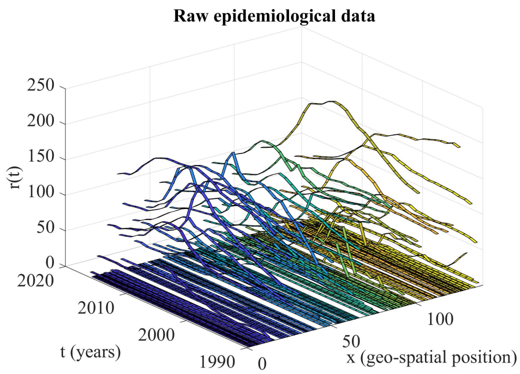


Fig. 1 The raw epidemiological data collected yearly from 1990 to 2016 along 137 geo-spatial locations

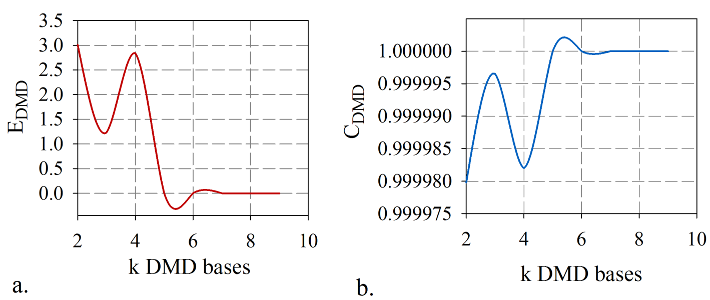


Fig. 2 a) The DMD-ROM error computed as a function of retained number of dynamic modes, b) The correlation coefficient of DMD-ROM computed as a function of retained number of dynamic modes. The randomized DMD algorithm selects a number of $N_{DMD} = 9$ leading modes

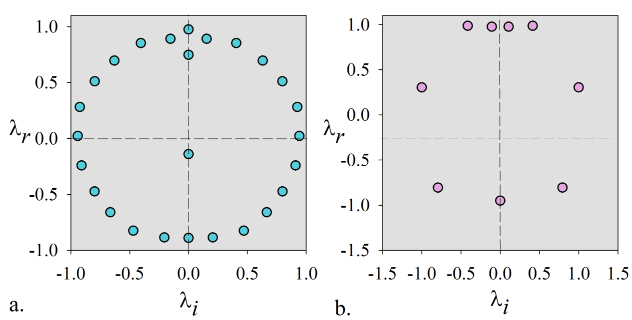
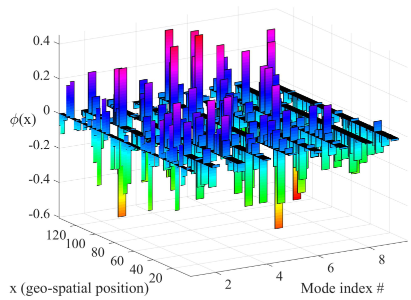
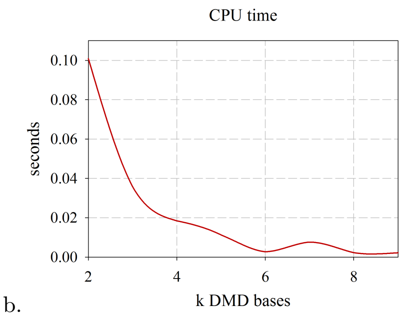


Fig. 3 a) The spectrum of dynamic decomposition of the epidemiological data; b) The spectrum of the reduced order model (2.24)



a.



b.

Fig. 4 a) The leading DMD modes used to generate the reduced order model; b) The CPU time required in the offline stage (i.e. obtaining DMD-ROM) by randomized dynamic mode decomposition

Table 1 The DMD-ROM error E_{DMD}, the correlation coefficient C_{DMD} and the reduced order modelling rank N_{DMD} obtained from randomized modal decomposition.

Absolute error	Correlation coefficient	DMD-ROM Rank
$E_{DMD} = 1.3872 \times 10^{-10}$	$C_{DMD} = 0.99999$	$N_{DMD} = 9$

The spectrum of dynamic decomposition of the epidemiological data is illustrated in Figure 3a. The positions of the Ritz values indicate purely oscillatory modes since they lie on the unit circle, whereas the Ritz values that lie within the unit circle have a decaying dynamic characteristic. In Figure 3b we illustrate the spectrum of the reduced order model (2.24), when only a small part of raw data were used to obtain an accurate DMD-ROM. A straightforward analysis of spatial-temporal disease data subject to dynamic mode decomposition is provided by Proctor and Eckhoff [72].

The absolute error E_{DMD} of the reduced order model, the correlation coefficient C_{DMD} and the rank N_{DMD} of reduced order model of the raw data, provided by the randomized DMD algorithm are presented in Table 1.

The dynamic modes can also indicate the epidemiological connectedness of spatial locations. Figure 4a presents the leading DMD modes used to generate the reduced order model. The presence of low magnitude to high-magnitude modes may indicate the spread of the epidemic from the boundaries to the interior of the investigated area.

Another major advantage offered by the proposed randomized DMD can be found in the significant reduction of CPU time for computation of huge epidemiological data. The CPU time required in the offline stage (i.e. obtaining DMD-ROM) by applying randomized dynamic mode decomposition is illustrated in Figure 4b.

In this survey we focus on prediction of disease spreading. Using the *forecasting sliding window technique* described in the previous section, we obtained a yearly prediction of the spreading. The results are validated against true data. The accuracy of the predicted data is verified by means of root

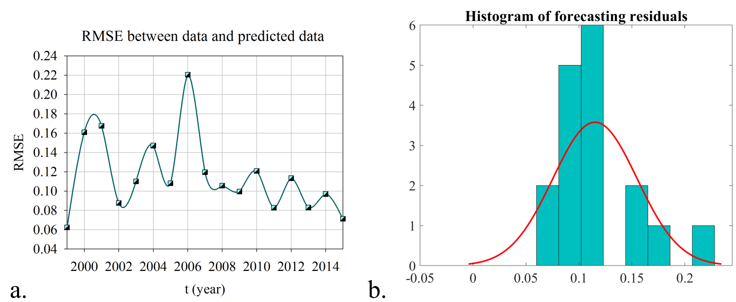


Fig. 5 a) RMSE between data and predicted data; b) Histogram of forecasting residuals

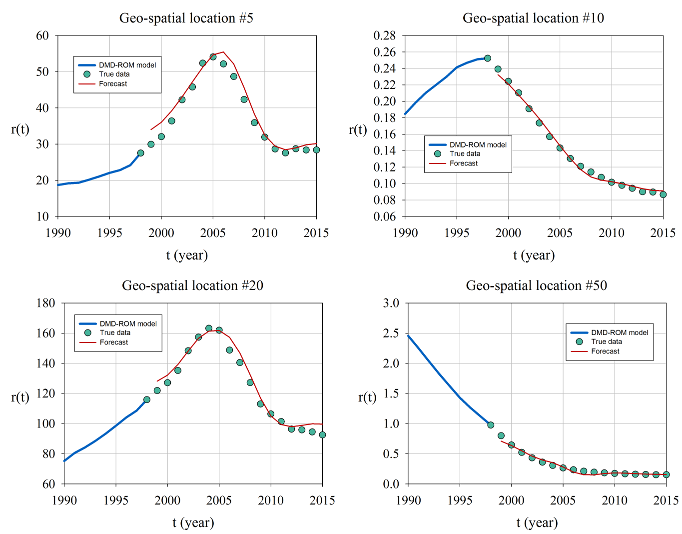


Fig. 6 The reduced order DMD model that generates the forecasting and the predicted data against true epidemiological data, for several geo-spatial locations identified by their indexes.

mean square error (RMSE) indicator

$$RMSE(t_i) = \sqrt{\frac{1}{N_x}\sum_{j=1}^{N_x}\left(r_{forecast}\left(t_i, x_j\right) - r\left(t_i, x_j\right)\right)^2}, \quad t_i \in \{t_{11}, ..., t_{N_t}\}. \tag{3.1}$$

Figure 5 plots the RMSE between data and predicted data, together with the normal distribution of forecasting residuals. Figures 6-7 present the reduced order DMD model that generates the forecasting and the predicted data against true epidemiological data, for several geo-spatial locations identified by their indexes.

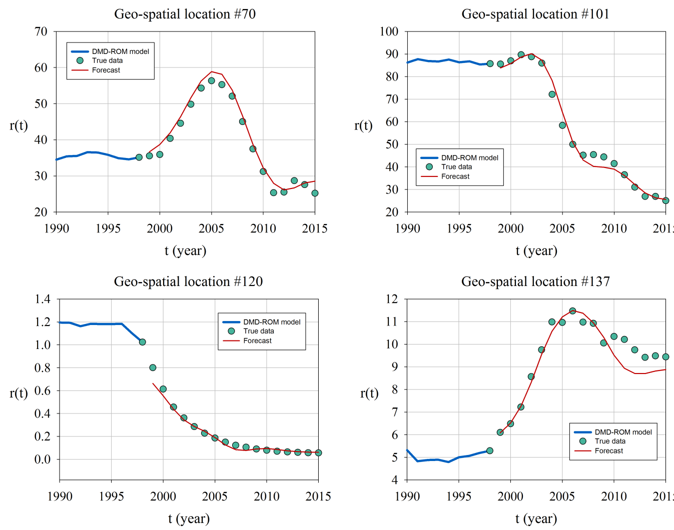


Fig. 7 The reduced order DMD model that generates the forecasting and the predicted data against true epidemiological data, for several geo-spatial locations identified by their indexes.

3.2 Modelling dynamic patterns from COVID-19 data

We have acquired the preliminary data provided by World Health Organization (WHO), consisting in number of COVID-19 infections, registered daily in the period 28/02/2020 – 18/05/2020 (i.e. 80 days) for 12 countries including the American continent and Europe. We consider the following initial data:

$N_x = 12$, representing the number of geo-spatial locations from which the epidemiological data were extracted;

$N_t = 80$, i.e. each epidemiological vector has 807 recordings;

$\Delta t = 1$ i.e. time step is one day.

The length of *sliding vector* is 12, i.e. 12 raw data in each data set are used to obtain the reduced order DMD-ROM model;

The length of *forecast vector* is 10, i.e. last 10 amplitudes are used to obtain the ARIMA model to forecast the DMD-ROM.

Adaptive randomized dynamic mode decomposition algorithm (ARDMD) is applied to identify the optimal rank of mathematical model of the infection spread. Figure 8 presents an insight of how the proposed randomized DMD algorithm finds the optimal number of leading DMD modes $N_{DMD} = 12$. The RMSE of the reduced order model as a function of retained number of dynamic modes is illustated in Figure 9. It is confirmed that the lower value is obtained for the optimal number of leading DMD modes $N_{DMD} = 12$.

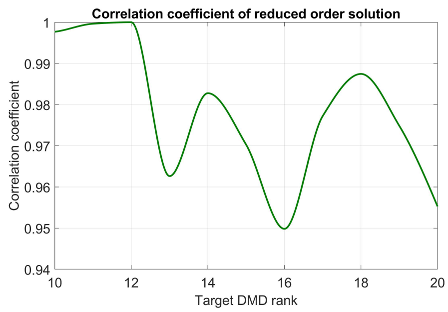


Fig. 8 The correlation coefficient of DMD-ROM computed as a function of retained number of dynamic modes. The randomized DMD algorithm selects a number of $N_{DMD} = 12$ leading modes.

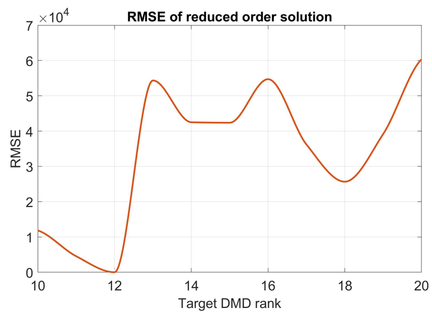


Fig. 9 RMSE of the reduced order model as a function of retained number of dynamic modes. It is confirmed that the lower value is obtained for the optimal number of leading DMD modes $N_{DMD} = 12$.

The correlation coefficient C_{DMD} of the reduced order model, the RMSE of the reduced order model and the rank N_{DMD} of reduced order model provided by the randomized DMD algorithm are presented in Table 2.

Table 2 The DMD-ROM correlation coefficient C_{DMD}, RMSE and the reduced order modelling rank v obtained from randomized modal decomposition.

Correlation coefficient	RMSE	DMD-ROM Rank
$C_{DMD} = 1$	$E_{DMD} = 1.4361 \times 10^{-6}$	$N_{DMD} = 12$

The spectrum of the reduced order model (2.24) is illustrated in Figure 10a. The spectrum of DMD decomposition consists in 12 eigenvalues associated to dominant modes. The application of randomized DMD leads to a robust, stable and super fast algorithm. Figure 10b illustrates the CPU time required in the offline stage, computed as a function of retained number of dynamic modes.

Figure 11 illustrates the raw data of Covid-19 infections vs. the high fidelity reduced order mathematical model.

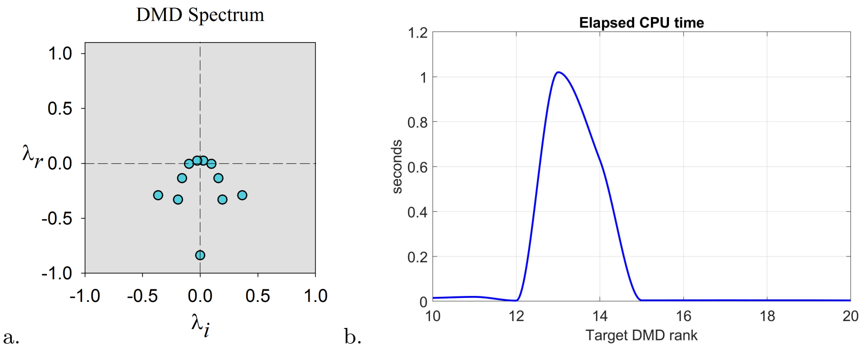


Fig. 10 a. The spectrum of the reduced order model (2.24); b. CPU time required in the offline stage, computed as a function of retained number of dynamic modes.

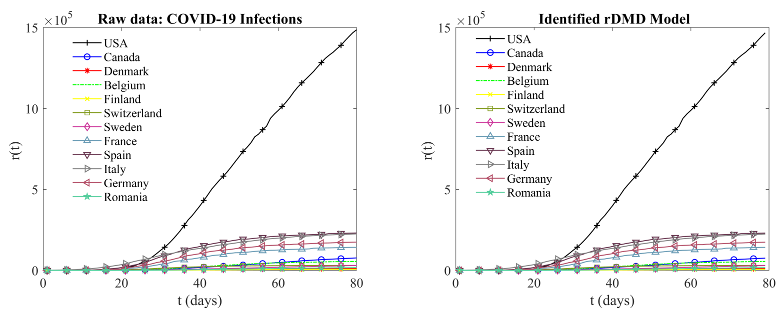


Fig. 11 The raw data of COVID-19 infections vs. the high fidelity identified rDMD mathematical model. COVID-19 infection rates are registered daily in the period 28/02/2020 – 18/05/2020 (i.e. 80 days) for 12 countries including the American continent and Europe.

After obtaining the high fidelity mathematical model we use a small part of the raw data to initiate the predictive model and the rest of the data to validate the prediction results.

The optimal ARIMA model parameters have been identified during the sliding window prediction algorithm by using the Akaike criterion (AIC), by investigating autocorrelation function (ACF) and partial autocorrelation function (PACF) and by testing the normal distribution assumption about residuals [23]. The next figures illustrate an exploratory analysis of the sliding window predictive algorithm. The sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) are used to identify ARIMA lag order (p), degree of differencing (d) and the order of moving average (q) (Figure 12).

The accuracy of the predicted data is verified by means of root mean square error indicator (RMSE) defined by Eq. 3.1 Figure 13 plots the root mean square error between data and predicted data, together with the normal distribution of forecasting residuals.

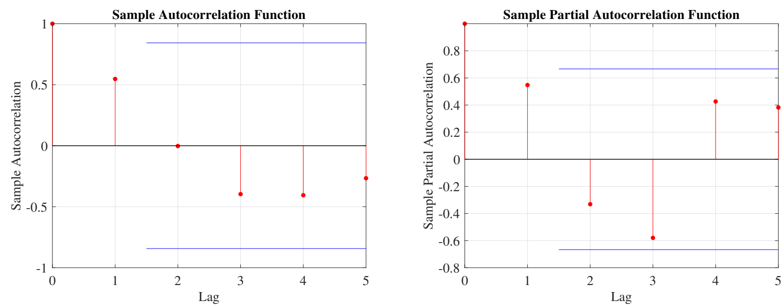


Fig. 12 The sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) are used to identify ARIMA lags.

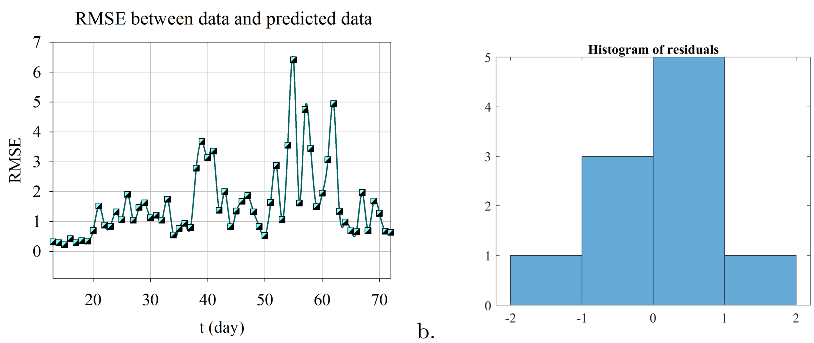


Fig. 13 a) Root mean square error between data and predicted data; b) The normal distribution of forecasting residuals.

Figure 14-15 present a daily prediction of COVID-19 infection spreading in different countries. The results are validated against true data.

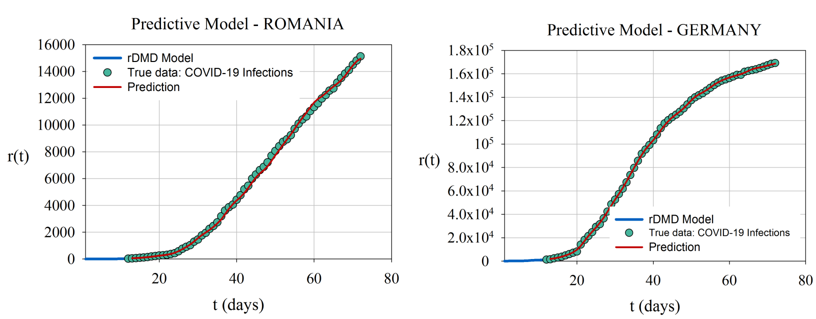


Fig. 14 Daily predicted infection data against true epidemiological data for Romania and Germany, respectively.

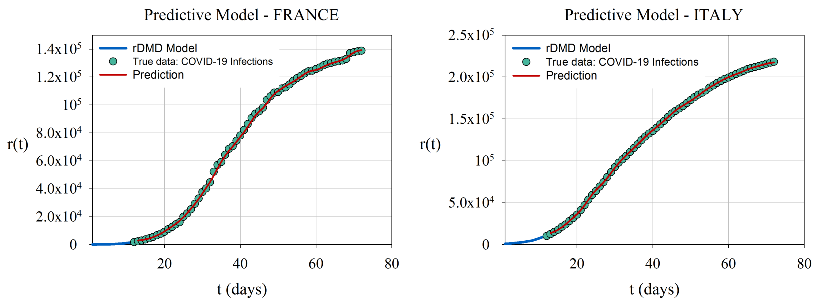


Fig. 15 Daily predicted infection data against true epidemiological data for France and Italy, respectively.

The major advantages of the adaptive randomized dynamic mode decomposition (**ARDMD**) are:

- This method provides an efficient tool in developing the linear model of epidemiological dynamics described by non-intrusive data.
- This method does not require an additional selection algorithm of the DMD modes. ARDMD produces a reduced order subspace of Ritz values, having the same dimension as the rank of randomized SVD function, where the most significant DMD modes live.
- We gain a significant reduction of the offline CPU time in computation of the ROM compared with the classic dynamic mode decomposition associated with different modes selection criteria.

4 Application of DMD-ROM models in fluid dynamics

In order to compare the performances of the proposed algorithms **CrDMD** and **ARDMD**, we performed the model order reduction of non-intrusive data originating from Saint-Venant model or shallow water equations model.

Unlike the classic selection based on modes amplitude, dynamic vector filtering criterion **CrDMD** offers two major advantages: it provides an automatic selection of the most representative modes, even when they exhibit rapid growth with lower amplitudes or they are high amplitudes fast damped modes. The **CrDMD** algorithm proved its efficiency in application to shallow water equations model, preserving a very good approximation of the full solution by the ROM model.

To overcome the inconveniences of developing and implementing a mode selection criterion associated with dynamic mode decomposition, the adaptive randomized DMD algorithm (**ARDMD**) is endowed with a randomized singular value decomposition (RSVD) function. The rank of the ROM is given as the unique solution of an optimization problem whose constraints consist of a sufficiently small relative error of data reconstruction and a sufficiently high correlation coefficient between the numerical data and the DMD solution. Solving the optimization problem (2.25) using a collaborative optimization technique involving hybrid simulated annealing and sequential quadratic programming (SA-SQP-CO) [69] we gain a fast and accurate adaptive randomized DMD algorithm, with a significantly lower rank for the new ROM, compared to the case of deterministic **CrDMD** algorithm with modes selection criterion (2.22).

The Saint-Venant equations, named after the French mathematician Adhémar Jean Claude Barré de Saint-Venant (1797 − 1886), represent a system of conservation laws that describe the flow below a pressure surface in a fluid. The test problem used in this paper consists of the nonlinear Saint-Venant equations (also called the shallow water equations [75]) in a channel on the rotating earth:

$$\frac{\partial\left(\tilde{u}\widetilde{h}\right)}{\partial t}+\frac{\partial\left(\tilde{u}^2\widetilde{h}+g\widetilde{h}^2/2\right)}{\partial x}+\frac{\partial\left(\tilde{u}\tilde{v}\widetilde{h}\right)}{\partial y}=\widetilde{h}\left(f\tilde{v}-g\frac{\partial H}{\partial x}\right), \tag{4.1}$$

$$\frac{\partial\left(\tilde{v}\widetilde{h}\right)}{\partial t}+\frac{\partial\left(\tilde{u}\tilde{v}\widetilde{h}\right)}{\partial x}+\frac{\partial\left(\tilde{v}^2\widetilde{h}+g\widetilde{h}^2/2\right)}{\partial y}=\widetilde{h}\left(-f\tilde{u}-g\frac{\partial H}{\partial y}\right), \tag{4.2}$$

$$\frac{\partial\widetilde{h}}{\partial t}+\frac{\partial\left(\tilde{u}\widetilde{h}\right)}{\partial x}+\frac{\partial\left(\tilde{v}\widetilde{h}\right)}{\partial y}=0, \tag{4.3}$$

where $\tilde{u}$ and $\tilde{v}$ are the velocity components in the $\tilde{x}$ and $\tilde{y}$ axis directions respectively, $\tilde{h}$ represents the depth of the fluid, $H\left(x,y\right)$ is the orography field, $\tilde{f}$ is the Coriolis factor and g is the acceleration of gravity.

A description of the Saint-Venant system as a result of depth-integration of the Navier-Stokes equations is posed by Saint-Venant in his 1871 paper [75]. Details can be found in [64], [85] and [44].

We consider that the reference computational configuration is the rectangular $2D$ domain $\Omega=[0,L_{\max}]\times[0,D_{\max}]$. Subscripts represent the derivatives with respect to time and the streamwise and spanwise coordinates.

In this model, the Coriolis parameter is modelled as varying linearly in the spanwise direction, such that

$$\widetilde{f}=f_0+\beta(\widetilde{y}-D_{\max}), \tag{4.4}$$

where f_0,β are constants, $L_{\max},D_{\max}$ are the dimensions of the rectangular domain of integration. The height of the orography is given by the fixed two-dimensional field

$$H\left(x,y\right)=\alpha e^{y^2-x^2}. \tag{4.5}$$

The model (4.1)-(4.3) is associated with periodic boundary conditions in the $\tilde{x}$-direction and solid wall boundary condition in the $\tilde{y}$-direction:

$$\tilde{u}\left(0,\tilde{y},\tilde{t}\right)=\tilde{u}\left(L_{\max},\tilde{y},\tilde{t}\right),\ \tilde{v}\left(\tilde{x},0,\tilde{t}\right)=\tilde{v}\left(\tilde{x},D_{\max},\tilde{t}\right)=0, \tag{4.6}$$

and also with the initial condition of [46] as the initial height field, which propagates the energy in wave number one, in the streamwise direction:

$$\begin{aligned} & h_0(\tilde{x},\tilde{y}) = H_0 + H_1 \tanh\left(\tfrac{9(D_{\max}/2-\tilde{y})}{2D_{\max}}\right) + \\ & H_2 \sin\left(\tfrac{2\pi\tilde{x}}{L_{\max}}\right) \cosh^{-2}\left(\tfrac{9(D_{\max}/2-\tilde{y})}{D_{\max}}\right) \end{aligned} \tag{4.7}$$

Using the geostrophic relationship $\tilde{u} = -\tilde{h}_{\tilde{y}}\left(g/\tilde{f}\right)$, $\tilde{v} = \tilde{h}_{\tilde{x}}\left(g/\tilde{f}\right)$, the initial velocity fields are derived as:

$$\begin{aligned} u_0(\tilde{x},\tilde{y}) = & -\frac{g}{\tilde{f}}\frac{9H_1}{2D_{\max}}\left(\tanh^2\left(\frac{9D_{\max}/2-9\tilde{y}}{2D_{\max}}\right)-1\right) - \\ & \frac{18g}{\tilde{f}}H_2 \sinh\left(\frac{9D_{\max}/2-9\tilde{y}}{D_{\max}}\right)\frac{\sin\left(\frac{2\pi\tilde{x}}{L_{\max}}\right)}{D_{\max}\cosh^3\left(\frac{9D_{\max}/2-9\tilde{y}}{D_{\max}}\right)}, \end{aligned} \tag{4.8}$$

$$v_0(\tilde{x},\tilde{y}) = 2\pi H_2 \frac{g}{\tilde{f}L_{\max}} \cos\left(\frac{2\pi\tilde{x}}{L_{\max}}\right)\cosh^{-2}\left(\frac{9(D_{\max}/2-\tilde{y})}{D_{\max}}\right). \tag{4.9}$$

The constants used for the test model are

$$f_0 = 10^{-4}s^{-1}, \quad \alpha = 4000, \quad \beta = 1.5\times 10^{-11}s^{-1}m^{-1}, \quad g = 9.81ms^{-1},$$

$$\mathrm{D}_{\max}= 50\times 10^3\mathrm{m}, \quad \mathrm{L}_{\max}= 254\times 10^3\mathrm{m},$$

$$H_0 = 10\times 10^3 m, \quad H_1 = -400m, \quad H_2 = -300m.$$

We set the error of the numerical algorithms to be less than $\varepsilon = 10^{-5}$. To measure the accuracy of the reduced shallow water model, we undertake a non-dimensional analysis of the shallow water model. Reference quantities of the dependent and independent variables in the shallow water model are considered, i.e. the length scale $L_{ref} = L_{\max}$ and the reference units for the height and velocities, respectively, are given by the initial conditions $h_{ref} = h_0$, $u_{ref} = u_0$. A typical time scale is also considered, assuming the form $t_{ref} = L_{ref}/u_{ref}$. In order to make the system of equations (4.1)-(4.3) non-dimensional, we define the non-dimensional variables

$$(t,x,y) = \left(\tilde{t}/t_{ref}, \tilde{x}/L_{ref}, \tilde{y}/L_{ref}\right), \quad (h,u,v) = \left(\tilde{h}/h_{ref}, \tilde{u}/u_{ref}, \tilde{v}/u_{ref}\right).$$

The numerical results are obtained employing a Lax-Wendroff finite difference discretization scheme and used in further numerical experiments in dimensionless form.

The data comprises a number of 145 unsteady solutions of the two-dimensional shallow water equations model (4.1)-(4.3), at regularly spaced time intervals of $\Delta t = 3600s$ for each solution variable.

To perform the comparison between the two DMD algorithms presented herein, we illustrate in Figure 16 the spectra of DMD decompositions of geopotential height field h, in case of DMD with modes selection criterion (**CrDMD**) and adaptive randomized DMD (**ARDMD**). The figures highlight the effect of the presented algorithms to select different modes, having the strongest impact on the reconstruction of the flow dynamics.

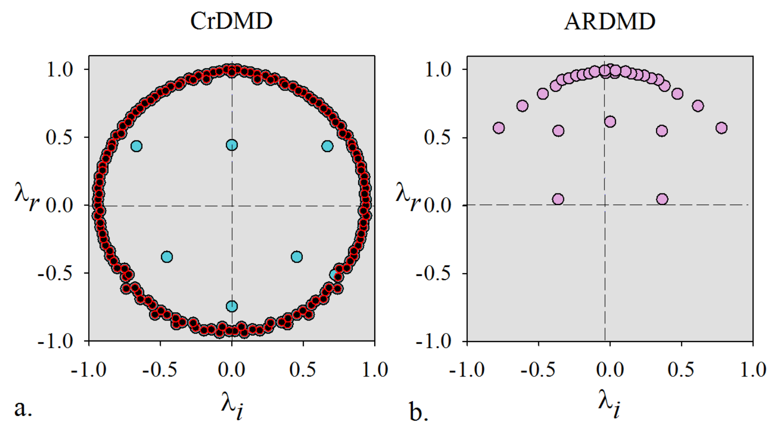


Fig. 16 The spectrum of DMD decomposition of height field h in case of: a) DMD with modes selection criterion - CrDMD: 144 modes (lighter colored dots), 137 modes selected for ROM (darker colored dots), b) Adaptive randomized DMD - ARDMD: 30 modes selected for ROM.

In the case of deterministic algorithm **CrDMD**, the number of selected modes for representation of h, u, v fields by the reduced DMD model are presented in Table 3 and illustrated in Figure 16a, where lighter circles represent total computed modes (144 modes), while darker circles represent the retained modes after DMD optimization (137 modes). In case of application of dynamic vector filtering criterion, the dominant modes are included in a smaller subspace, preserving a very good approximation of the full solution by the ROM model. The benefit of the proposed filtering criterion consists in eliminating the DMD modes that contribute weakly to the data sequence. It provides an automatic selection of the most representative modes, even when they exhibit rapid growth with lower amplitudes or they consist of high amplitudes fast damped modes.

Table 3 presents the number of DMD modes computed and involved in the ROM, provided by the two algorithms, respectively and the relative error Er_{DMD} of reconstruction of the flow field. Compared to **CrDMD** algorithm, the randomized DMD algorithm **ARDMD** produces a significantly reduced size spectrum.

Another benefit of the randomized dynamic mode decomposition is that the low order solution (2.18) is guaranteed to satisfy the boundary conditions (4.6) of the full model, because the DMD modes provided by the

Table 3 Comparison of the efficiency of the two algorithms, $\Delta t = 3600s$, 145 data snapshots have been processed.

Algorithm	Computed DMD modes	The ROM rank	The relative error
CrDMD[a]	144	$k = 137$	$Er^h{}_{DMD} = 2.8362 \times 10^{-6}$
ARDMD[b]	30	$k = 30$	$Er^h{}_{DMD} = 2.7189 \times 10^{-6}$

a–DMD with modes selection criterion.
b–Adaptive randomized dynamic mode decomposition.

ARDMD algorithm satisfy the relations

$$\phi_j^u(0, y) = \phi_j^u(L_{\max}, y), \; j = 1, ..., k \tag{4.10}$$

$$\phi_j^v(x, 0) = \phi_j^v(x, D_{\max}) = 0, \; j = 1, ..., k \tag{4.11}$$

where ϕ_j^u, ϕ_j^v are dynamic modes of the u and v fields, respectively, and k represents the number of the retained modes in the ROM.

The representation of the height field, based on the selected modes in case of algorithms **CrDMD** and **ARDMD** is displayed, respectively, in Figure 17. The vorticity field is illustrated in Figure 18, computed with the two algorithms. The ROM exhibits a relative error of order $\mathcal{O}\left(10^{-6}\right)$ in both cases. Data presented in Figure 19 confirms the efficiency of the **ARDMD** algorithm. Although the previous techniques detailed in [18] , [2], [20] lead to a reduced number of retained modes, there are still missing modes that would contribute to data approximation. Hence the relative error of flow reconstruction by the reduced order model is the best in the case of randomized dynamic mode decomposition. The great advantage of adaptive randomized DMD algorithm **ARDMD** is that it avoids the computational efforts required for implementing an additional criterion of dominant modes'selection, since they are automatically selected. Thus a significant reduction in computational time is also achieved compared with deterministic dynamic mode decomposition associated with different modes selection criteria. The CPU time required in the offline stage is presented in Figure 20. By employing the randomized DMD algorithm **ARDMD** in comparison with deterministic DMD associated with the energetic criterion for modes selection **CrDMD**, the computational complexity of the low order model is significantly reduced from the very beginning, as illustrated in Figure 20. Figure 21 and Figure 22 present examples of DMD modes involved in the reduced order model, in case of application of the two aforementioned algorithms.

5 Summary and conclusions

The present survey has focused on a subject of great interest in epidemiology and fluid dynamics, namely: identification of a reduced order model (ROM) from numerical code output by non-intrusive techniques (i.e. not requiring projecting of the governing equations onto the reduced basis modes,

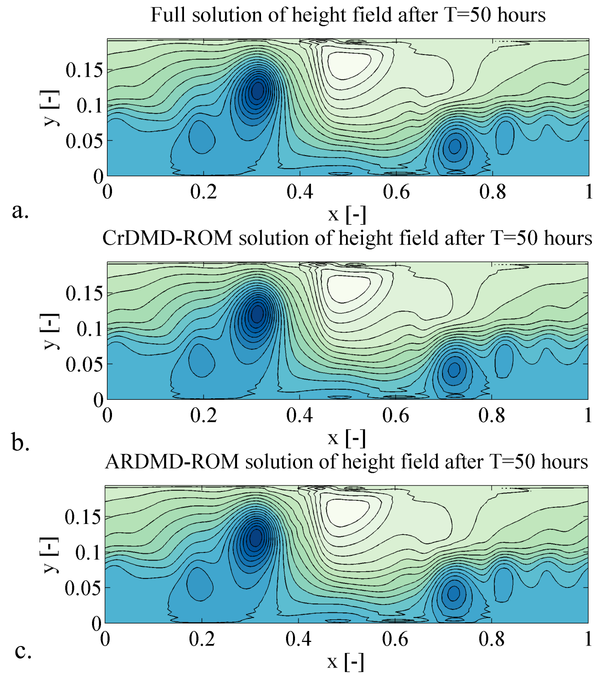


Fig. 17 a. Full solution of height field after $T = 50h$; b. ROM solution of height field obtained with CrDMD algorithm ($k = 137$ modes); c. ROM solution of height field obtained with ARDMD algorithm ($k = 30$ modes). The relative error is of order $\mathcal{O}\left(10^{-6}\right)$ in both cases.

assuming that these governing equations or their numerical code are not available).

We performed a comparison between two methods of model order reduction based on dynamic mode decomposition (DMD). The first method is a deterministic (classic) DMD technique endowed with a dynamic filtering criterion of selection of modes used in the ROM model (**CrDMD**). The second method is an adaptive randomized DMD algorithm (**ARDMD**) based on a randomized singular value decomposition.

In order to compare the performances of the proposed algorithms, we performed the model order reduction of non-intrusive data originating from epidemiology and for Saint-Venant model (or shallow water equation model).

Dynamic vector filtering criterion CrDMD offers two major advantages: it provides an automatic selection of the most representative modes, even when they exhibit rapid growth with lower amplitudes or they are high amplitudes fast damped modes. The CrDMD algorithm proved its efficiency in application to shallow water equations model, preserving a very good approximation of the full solution by the ROM model.

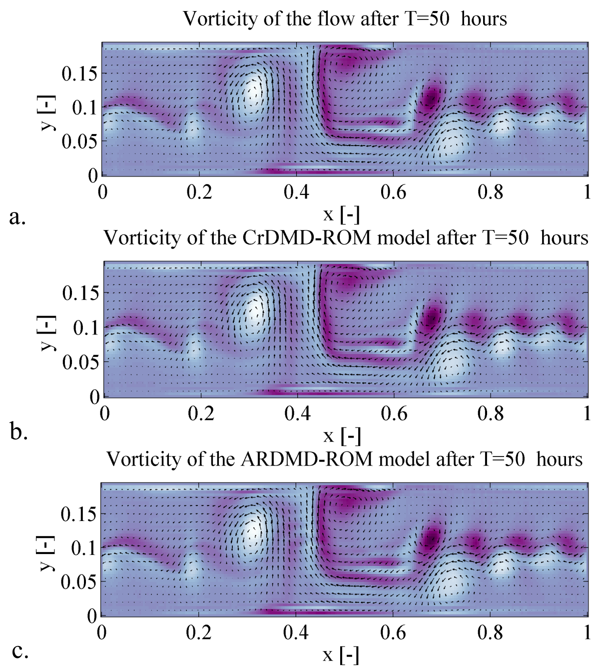


Fig. 18 a. Vorticity field after $T = 50h$; b. ROM solution of vorticity field obtained with CrDMD algorithm ($k = 137$ modes); c. ROM solution of vorticity field obtained with ARDMD algorithm ($k = 30$ modes). The relative error is of order $\mathcal{O}\left(10^{-6}\right)$ in both cases.

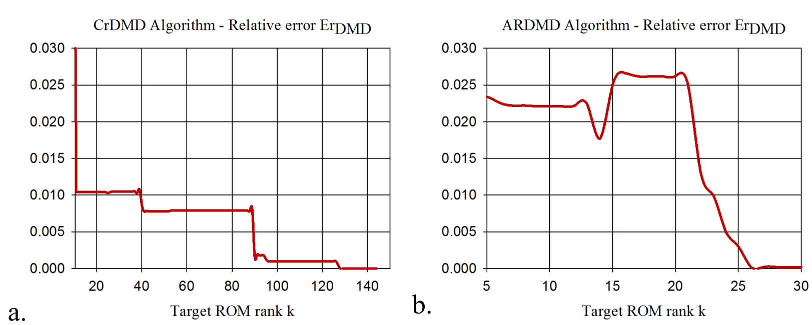


Fig. 19 The relative errors computed with respect to the ROM rank, in case of application of: a. deterministic DMD with modes selection criterion **CrDMD** and b. adaptive randomized DMD algorithm **ARDMD**, $\Delta t = 3600s$, 145 data snapshots have been processed.

The efficiency of the adaptive randomized DMD method is especially evident in the case of large amount of snapshots data. The rank of the ROM produced by the adaptive randomized **ARDMD** algorithm is significantly

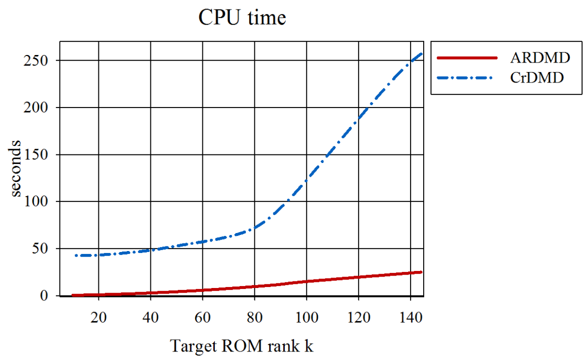


Fig. 20 The CPU time required in the offline stage by applying adaptive randomized dynamic mode decomposition **ARDMD** and deterministic DMD associated with the energetic criterion for modes selection **CrDMD**, $\Delta t = 3600s$, 145 data snapshots have been processed.

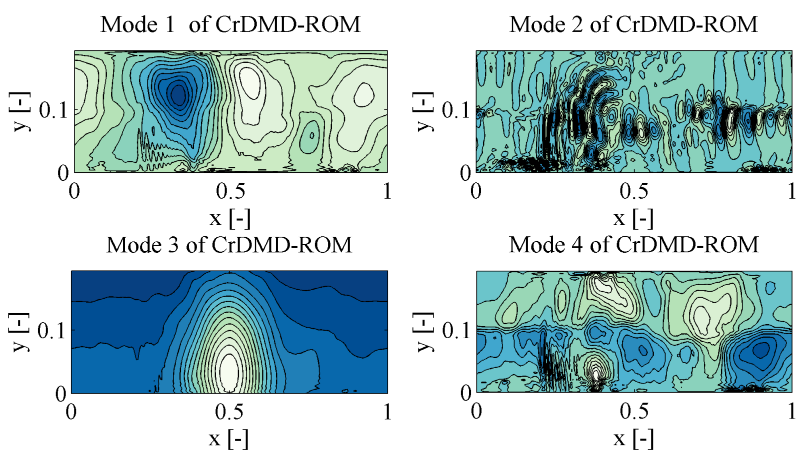


Fig. 21 Examples of DMD modes computed with **CrDMD** algorithm, which are involved in the ROM model.

smaller than in the case of deterministic algorithm with a selection criterion **CrDMD**.

Application of the adaptive randomized dynamic mode decomposition algorithm to epidemiological data offers the main advantage of deriving a reduced order model capable to provide an accurate short-medium term forecasting of epidemiologic dynamics.

We addressed the investigation of epidemiologic dynamics of the new Corona virus (Covid-19) through the perspective of reduced order modelling (ROM) using Randomized Dynamic Mode Decomposition (rDMD). The objective of rDMD is to build a model of smaller dimension from raw data, to decrease the computational burden while leading to a controlled

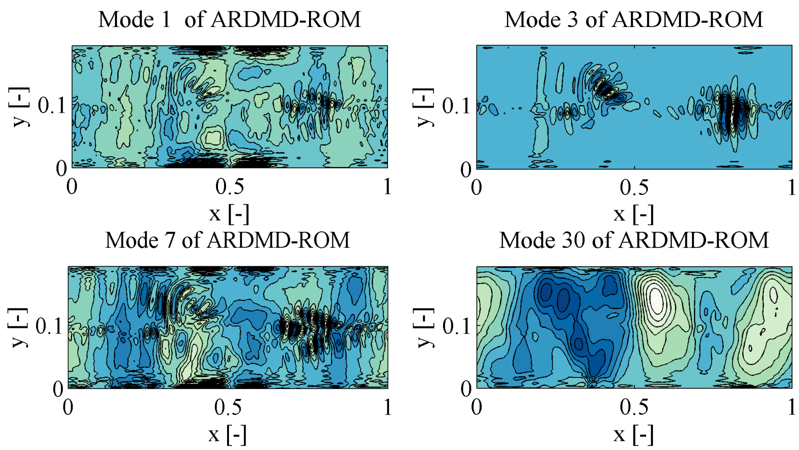


Fig. 22 Examples of DMD modes computed with **ARDMD** algorithm, which are involved in the ROM model.

loss of accuracy. To asses predictions from the non-intrusive data of disease spreading an additional key innovation in the paper consists in forecasting the DMD-ROM model in time using autoregressive integrated moving average (ARIMA) models, using a sliding time window technique also referred to as adaptive DMD. The proposed mathematical model was applied for the first time on Covid-19 infectious disease data.

The key benefits of the proposed methods are as follows: dynamic patterns of infectious disease spread are obtained, the information provided by the mathematical model can be useful for allocating vaccine resources per year, sending surveillance teams to monitor the disease and timing the interventions to leverage natural disease dynamics. Quantitative modeling and analysis plays a key role in understanding disease spread and optimally applying intervention resources such as viral new medications to maximize the probability of success for eradication.

References

1. Ahmed, S.E.; San, O.; Bistrian, D.A.; Navon, I.M. – *Sampling and resolution characteristics in reduced order models of shallow water equations: intrusive vs non-intrusive*, Internat. J. Numer. Methods Fluids, **92** (2020), no. 8, 992-1036.
2. Alekseev, A.K.; Bistrian, D.A.; Bondarev, A.E.; Navon, I.M. – *On linear and nonlinear aspects of dynamic mode decomposition*, Internat. J. Numer. Methods Fluids, **82** (2016), no. 6, 348-371.
3. Alla, A.; Kutz, J.N. – *Nonlinear model order reduction via dynamic mode decomposition*, SIAM J. Sci. Comput., **39** (2017), B778-B796.
4. Aniţa, L.I.; Aniţa, S.; Arnăutu, V. – *Internal null stabilization for some diffusive models of population dynamics*, Applied Mathematics and Computation, **219** (2013), no. 20, 10231-10244.
5. Aniţa, S.; Capasso, V.; Dimitriu, G. – *Regional control for a spatially structured malaria model*, Math. Methods Appl. Sci., **42** (2019), no. 8, 2909-2933.

6. Aniţa, S.; Capasso, V.; Dimitriu, G. – *Controlling an alien predator population by regional controls*, Nonlinear Anal. Real World Appl., **46** (2019), 82-97.
7. Aniţa, S. – *Analysis and control of age-dependent population dynamics*, Mathematical Modelling: Theory and Applications, **11**, Kluwer Academic Publishers, Dordrecht, (2000).
8. Aniţa, S.; Arnăutu, V.; Capasso, V. – *An introduction to optimal control problems in life sciences and economics*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser/Springer, New York, (2011).
9. Aniţa, S.; Capasso, V.; Lipnicka, M.; Nowakowski, A. – *A Shape Optimization Problem Concerning the Regional Control of a Class of Spatially Structured Epidemics: Sufficiency Conditions*, SEMA SIMAI Springer Series, (2020), ch. Aguiar M., Braumann C., Kooi B.W., Pugliese A., Stollenwerk N., Venturino E. (eds), Current Trends in Dynamical Systems in Biology and Natural Sciences, 165-183.
10. Apreutesei, N.; Dimitriu, G.; Strugariu, R. – *An optimal control problem for a two–prey and one–predator model with diffusion*, Comput. Math. Appl., **67** (2014), no. 12, 2127-2143.
11. Apreutesei, N.; Dimitriu, G. – *On a prey-predator reaction-diffusion system with Holling type III functional response*, J. Comput. Appl. Math., **235** (2010), no. 2, 366-379.
12. Bagheri, S. – *Computational hydrodynamic stability and flow control based on spectral analysis of linear operators*, Arch. Comput. Methods Eng., **19** (2012), no. 3, 341-379.
13. Bagheri, S. – *Koopman–mode decomposition of the cylinder wake*, J. Fluid Mech., **726** (2013), 596-623.
14. Balajewicz, M.J.; Dowell, E.H.; Noack, B.R. – *Low–dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier-Stokes equation*, J. Fluid Mech., **729** (2013), 285-308.
15. Berger, E.; Sastuba, M.; Vogt, D.; Jung, B.; Amor, H.B. – *Estimation of perturbations in robotic behavior using dynamic mode decomposition*, Advanced Robotics, **29** (2015), no. 5, 331-343.
16. Bistrian, D.A.; Susan-Resiga, R.F. – *Weighted proper orthogonal decomposition of the swirling flow exiting the hydraulic turbine runner*, Appl. Math. Modell., **40** (2016), no. 5-6, 4057-4078.
17. Bistrian, D.A.; Navon, I.M. – *Randomized dynamic mode decomposition for non-intrusive reduced order modelling*, Internat. J. Numer. Methods Engrg., **112** (2017), no. 1, 3-25.
18. Bistrian, D.A.; Navon, I.M. – *An improved algorithm for the shallow water equations model reduction: dynamic mode decomposition vs POD*, Internat. J. Numer. Methods Fluids, **78** (2015), no. 9, 552-580.
19. Bistrian, D.A.; Navon, I.M. – *The method of dynamic mode decomposition in shallow water and a swirling flow problem*, Internat. J. Numer. Methods Fluids, **83** (2017), no. 1, 73-89.
20. Bistrian, D.A.; Navon, I.M. – *Efficiency of randomised dynamic mode decomposition for reduced order modelling*, Int. J. Comput. Fluid Dyn., **32** (2018), no. 2-3, 88-103.
21. Bistrian, D.A.; Dimitriu, G.; Navon, I.M. – *Processing epidemiological data using dynamic mode decomposition method*, AIP Conference Proceedings, **2164** (2019), no. 1, 080002.
22. Bongard, J.; Lipson, H. – *Automated reverse engineering of nonlinear dynamical systems*, Proceedings of the National Academy of Sciences of the Unitated States of America, **104** (2007), no. 24, 9943-9948.
23. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. – *Time Series Analysis: Forecasting and Control*, Wiley Series in Probability and Statistics, 5th Edition., Wiley, (2015).
24. Brauer, F.; Castillo-Chavez, C. – *Mathematical models in population biology and epidemiology*, Second edition, Texts in Applied Mathematics, 40, Springer, New York, (2012).
25. Brunton, S.L.; Brunton, B.W.; Proctor, J.L.; Kutz, J.N. – *Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control*, PLoS ONE, **11** (2016), 0150171.
26. Buljak, V.; Maier, G. – *Proper orthogonal decomposition and radial basis functions in material characterization based on instrumented indentation*, Engineering Structures, **33** (2011), no. 2, 492-501.

27. Cao, Y.; Zhu, J.; Luo, Z.; Navon, I.M. – *Reduced-order modeling of the upper tropical Pacific Ocean model using proper orthogonal decomposition*, Comput. Math. Appl., **52** (2006), no. 8-9, 1373-1386.
28. Cao, Y.; Zhu, J.; Navon, I.M.; Luo, Z. – *A-reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition*, International Journal for Numerical Methods in Fluids, **53** (2007), 1571-1583.
29. Chen, W.-S. – *Use of recurrence plot and recurrence quantification analysis in Taiwan unemployment rate time series*, Physica A: Statistical Mechanics and its Applications, **390** (2011), no. 7, 1332-1342.
30. Chen, K.K.; Tu, J.H.; Rowley, C.W. – *Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses*, J. Nonlinear Sci., **22** (2012), no. 6, 887-915.
31. Cheng, M.; Fang, F.; Pain, C.C.; Navon, I.M. – *Data–driven modelling of nonlinear spatio–temporal fluid flows using a deep convolutional generative adversarial network*, Comput. Methods Appl. Mech. Engrg., **365** (2020), 113000, 1-18.
32. Chopra, A.K. – *Dynamics of Structures*, 4th edition edn., Prentice-Hall International Series in Civil Engineering and Engineering Mechanics, (2000).
33. Dimitriu, G.; Apreutesei, N. – *Comparative study with data assimilation experiments using proper orthogonal decomposition method*, Lecture Notes in Computer Science, **4818** (2008), 393-400.
34. Dimitriu, G.; Apreutesei, N.; Ştefănescu, R. – *Numerical simulations with data assimilation using an adaptive POD procedure*, Lecture Notes in Computer Science, **5910** (2010), 165-172.
35. Dimitriu, G.; Navon, I.M.; Ştefănescu, R. – *Application of POD-DEIM approach for dimension reduction of a diffusive predator–prey system with Allee effect. Large-scale scientific computing*, Lecture Notes in Comput. Sci., Springer, Heidelberg, **8353** (2014), 373-381.
36. Dimitriu, G.; Ştefănescu, R.; Navon, I.M. – *POD-DEIM approach on dimension reduction of a multi–species host–parasitoid system*, Ann. Acad. Rom., Sci. Ser. Math. Appl., **7** (2015), no. 1, 173-188.
37. Dimitriu, G.; Ştefănescu, R.; Navon, I.M. – *Comparative numerical analysis using reduced–order modeling strategies for nonlinear large-scale systems*, J. Comput. Appl. Math., **310** (2017), 32-43.
38. Du, J.; Fang, F.; Pain, C.C.; Navon, I.M.; Zhu, J.; Ham, D.A. – *POD reduced order unstructured mesh modeling applied to 2D and 3D fluid flow*, Comput. Math. Appl., **65** (2013), no. 3, 362-379.
39. Dumon, A.; Allery, C.; Ammar, A. – *Proper generalized decomposition method for incompressible Navier–Stokes equations with a spectral discretization*, Appl. Math. Comput., **219** (2013), no. 15, 8145-8162.
40. Erichson, N.B.; Donovan, C. – *Randomized low-rank dynamic mode decomposition for motion detection*, Computer Vision and Image Understanding, **146** (2016), 40-50.
41. Erichson, N.B.; Mathelin, L.; Kutz, J.N.; Brunton, S.L. – *Randomized dynamic mode decomposition*, SIAM J. Appl. Dyn. Syst., **18** (2019), no. 4, 1867-1891.
42. Fang, F.; Pain, C.C.; Navon, I.M.; Cacuci, D.G.; Chen, X. – *The independent set perturbation method for efficient computation of sensitivities with applications to data assimilation and a finite element shallow water model*, Comput. & Fluids, **76** (2013), 33-49.
43. Frederich, O.; Luchtenburg, D.M. – *Modal analysis of complex turbulent flow*, The 7th International Symposium on Turbulence and Shear Flow Phenomena (TSFP-7), Ottawa, Canada (2011).
44. Galdi, G.P. – *An introduction to the mathematical theory of the Navier-Stokes equation*, Springer Tracts in Natural Philosophy, Springer-Verlag, New York, (1994).
45. Golub, G.H.; Van Loan, C.F. – *Matrix computations, Third edition*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, (1996).
46. Grammeltvedt, A. – *A survey of finite-diference schemes for the primitive equations for a barotropic fluid*, Monthly Weather Review, **97** (1969), no. 5, 384-404.
47. Gunzburger, M.; Iliescu, T.; Mohebujjaman, M.; Schneier, M. – *Nonintrusive stabilization of reduced order models for uncertainty quantification of time–dependent convection–dominated flows*, Preprint arXiv, **1810.08746** (2018).

48. Halko, N.; Martinsson, P.G.; Tropp, J.A. – *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., **53** (2011), no. 2, 217-288.
49. Holmes, P.; Lumley, J.L.; Berkooz. G. – *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge Monographs on Mechanics, Cambridge University Press, Cambridge (1996).
50. Hunter, E.; Namee, B.M.; Kelleher, J.D. – *A taxonomy for agent–based models in human infectious disease epidemiology*, Journal of Artificial Societies and Social Simulation, **20** (2017), no. 3, 1-17.
51. Hunter, E.; Namee, B.M.; Kelleher, J.D. – *An open–data–driven agent–based model to simulate infectious disease outbreaks*, PLOS ONE, **14** (2019), no. 1, 1-35.
52. Jovanović, M.R.; Schmid, P.J.; Nichols, J.W. – *Sparsity–promoting dynamic mode decomposition*, Physics of Fluids, **26** (2014), no. 2, 024103.
53. Koopman, B.O. – *Hamiltonian systems and transformations in Hilbert space*, Proc. Nat. Acad. Sci., **17** (1931), 315-318.
54. Kutz, J.N.; Brunton, S.L.; Brunton, B.W.; Proctor, J.L. – *Dynamic mode decomposition. Data-driven modeling of complex systems*, SIAM, Philadelphia, (2016).
55. Kutz, J.N.; Fu, X.; Brunton, S.L. – *Multiresolution dynamic mode decomposition*, SIAM J. Appl. Dyn. Syst., **15** (2016), no. 2, 713-735.
56. Liberge, E.; Hamdouni, A. – *Reduced order modelling method via proper orthogonal decomposition (POD) for flow around an oscillating cylinder*, Journal of Fluids and Structures, **26** (2010), no. 2, 292-311.
57. Ljung, L. – *Perspectives on system identification*, Annual Reviews in Control, **34** (2010), no. 1, 1-12.
58. Luchtenburg, D.M.; Rowley, C.W. – *Model reduction using snapshot-based realizations*, American Physical Society, 64th Annual Meeting of the APS Division of Fluid Dynamics, (2011).
59. Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. – *VAMPnets for deep learning of molecular kinetics*, Nature Communications, **9** (2018), no. 5.
60. Mariani, R.; Dessi, D. – *Analysis of the global bending modes of a floating structure using the proper orthogonal decomposition*, Journal of Fluids and Structures, **28** (2012), 115-134.
61. Mezić, I. – *Analysis of fluid flows via spectral properties of the Koopman operator*, Annual review of fluid mechanics, **45** (2013), 357-378.
62. Mezić, I. – *Spectral properties of dynamical systems, model reduction and decompositions*, Nonlinear Dynamics, **41** (2005), no. 1-3, 309-325.
63. Navon, I.M.; DeVilliers, R. – *Gustaf: A quasi-Newton nonlinear Adi Fortran IV program for solving the shallow-water equations with augmented lagrangians*, Computers & Geosciences, **12** (1986), no. 2, 151-173.
64. Navon, I.M. – *Numerical Methods for the Solution of the Shallow-Equations in Meteorology*, Thesis (PH.D.) - University of the Witwatersrand, Johannesburg, South Africa, **41** (1979), no. 3.
65. Navon, I.M. – *A Numerov-Galerkin technique applied to a finite-element shallow-water equations model with enforced conservation of integral invariants and selective lumping*, J. Comput. Phys., **52** (1983), no. 2, 313-339.
66. Nelles, O. – *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*, Springer (2001).
67. Noack, B.R.; Morzynski, M.; Tadmor, G. – *Reduced-Order Modelling for Flow Control*, Springer (2011).
68. Noack, B.R.; Stankiewicz, W.; Morzyński, M.; Schmid, P.J. – *Recursive dynamic mode decomposition of transient and post-transient wake flows*, J. Fluid Mech., **809** (2016), 843-872.
69. Nocedal, J.; Wright, S.J. – *Numerical optimization*, Second edition, Springer Series in Operations Research and Financial Engineering, Springer, New York, (2006).
70. Östh, J.; Noack, B.R.; Krajnović, S.; Barros, D.; Borée, J. – *On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high–Reynolds-number flow over an Ahmed body*, J. Fluid Mech., **747** (2014), 518-544.
71. Proctor, J.L.; Brunton, S.L.; Kutz, J.N. – *Dynamic mode decomposition with control*, SIAM J. Appl. Dyn. Syst., **15** (2016), no. 1, 142-161.
72. Proctor, J.L.; Eckhoff, P.A. – *Discovering dynamic patterns from infectious disease data using dynamic mode decomposition*, International Health, **7** (2015), no. 2, 139-145.

73. Rowley, C.W.; Mezić, I.; Bagheri, S.; Schlatter, P.; Henningson, D.S. – *Spectral analysis of nonlinear flows*, J. Fluid Mech., **641** (2009), 115-127.
74. Rowley, C.W.; Mezić, I.; Bagheri, S.; Schlatter, P.; Henningson, D.S. – *Reduced-order models for flow control: balanced models and Koopman modes*, Seventh IUTAM Symposium on Laminar-Turbulent Transition, IUTAM Bookseries, 2010.
75. Saint-Venant, A.J.C. Barré de – *Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et a l'introduction de marées dans leurs lits*, Comptes Rendus de l'Académie des Sciences, **73** (1871), 237-240.
76. San, O.; Iliescu, T. – *A stabilized proper orthogonal decomposition reduced–order model for large scale quasigeostrophic ocean circulation*, Advances in Computational Mathematics, **41** (2015), 1289-1319.
77. San, O.; Staples, A.E.; Wang, Z.; Iliescu, T. – *Approximate deconvolution large eddy simulation of a barotropic ocean circulation model*, Ocean Modelling, **40** (2011), no. 2, 120-132.
78. Schmidt, M.; Lipson, H. – *Distilling free–form natural laws from experimental data*, Science, **324** (2009), no. 5293, 81-85.
79. Schmid, P.; Sesterhenn, J. – *Dynamic mode decomposition of numerical and experimental data*, 61st Annual Meeting of the APS Division of Fluid Dynamics, San Antonio, Texas (2008).
80. Schmid, P. – *Dynamic mode decomposition of numerical and experimental data*, Journal of Fluid Mechanics, **656** (2010), 5-28.
81. Schmid, P.; Violato, D.; Scarano, F. – *Decomposition of time-resolved tomographic PIV*, Exp. Fluids, **52** (2012), no. 6, 1567-1579.
82. Ştefănescu, R.; Sandu, A.; Navon, I.M. – *Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations*, Internat. J. Numer. Methods Fluids, **76** (2014), no. 8, 497-521.
83. Tissot, G.; Cordier, L.; Benard, N.; Noack, B.R. – *Model reduction using dynamic mode decomposition*, Comptes Rendus Mécanique, **342** (2014), no. 6-7, 410-416.
84. Tu, J.H.; Rowley, C.W.; Luchtenburg, D.M.; Brunton, S.L.; Kutz, J.N. – *On dynamic mode decomposition: theory and applications*, J. Comput. Dyn., **1** (2014), no. 2, 391-421.
85. Vreugdenhil, C.B. – *Numerical Methods for Shallow-Water Flow*, Water Science and Technology Library, Kluwer Academic Publishers, **13** (1994).
86. Wang, Z.; Akhtar, I.; Borggaard, J.; Iliescu, T. – *Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison*, Comput. Methods Appl. Mech. Engrg. **237/240** (2012), 10-26.
87. Williams, M.O.; Kevrekidis, I.G.; Rowley, C.W. – *A data-driven approximation of the Koopman operator: extending dynamic mode decomposition*, J. Nonlinear Sci., **25** (2015), no. 6, 1307-1346.
88. Winton, C.; Pettway, J.; Kelley, C.T.; Howington, S.; Eslinger, O.J. – *Application of Proper Orthogonal Decomposition (POD) to inverse problems in saturated groundwater flow*, Advances in Water Resources, **34** (2011), no. 12, 1519-1526.
89. Xiao, D.; Fang, F.; Buchan, A.G.; Pain, C.C.; Navon, I.M.; Muggeridge, A. – *Non–intrusive reduced order modelling of the Navier-Stokes equations*, Comput. Methods Appl. Mech. Engrg, **293** (2015), 522-541.
90. Xiao, D.; Heaney, C.E.; Fang, F.; Mottet, L.; Hu, R.; Bistrian, D.A.; Aristodemou, E.; Navon, I.M.; Pain, C.C. – *A domain decomposition non–intrusive reduced order model for turbulent flows*, Comput. & Fluids, **182** (2019), 15-27.
91. Xiao, M.; Breitkopf, P.; Filomeno Coelho, R.; Knopf-Lenoir, C.; Villon, P.; Zhang, W. – *Constrained proper orthogonal decomposition based on QR-factorization for aerodynamical shape optimization*, Appl. Math. Comput., **223** (2013), 254-263.
92. Xiao, D.; Heaney, C.E.; Mottet, L.; Fang, F.; Lin, W.; Navon, I.M.; Guo, Y.; Matar, O.K.; Robins, A.G.; Pain, C.C. – *A reduced order model for turbulent flows in the urban environment using machine learning*, Building and Environment, **148** (2019), 323-337.
93. Xiao, D.; Fang, F.; Heaney, C.E.; Navon, I.M.; Pain, C.C. – *A domain decomposition method for the non–intrusive reduced order modelling of fluid flow*, Comput. Methods Appl. Mech. Engrg., **354** (2019), 307-330.
94. Zerfas, C.; Rebholz, L.G.; Schneier, M.; Iliescu, T. – *Continuous data assimilation reduced order models of fluid flow*, Comput. Methods Appl. Mech. Engrg. **357** (2019), 1-18.

Received: 9.XI.2020 / Accepted: 4.XII.2020

Authors

Diana A. Bistrian,
Politehnica University of Timişoara,
Department of Electrical Engineering and Industrial Informatics,
Revoluţiei Nr. 5,
Hunedoara 331128, România,
E-mail: diana.bistrian@fih.upt.ro

Gabriel Dimitriu,
University of Medicine and Pharmacy "Grigore T. Popa",
Department of Mathematics and Informatics,
Universităţii 16,
Iaşi 700115, România,
E-mail: gabriel.dimitriu@umfiasi.ro

Ionel M. Navon,
Florida State University,
Department of Scientific Computing,
Dirac Science Library Building, Tallahassee,
FL 32306-4120, USA,
E-mail: inavon@fsu.edu