Physics Informed Modeling of Ecosystem Respiration via Dynamic Mode Decomposition with Control Input

Maha Shadaydeh
Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Germany
Email: maha.shadaydeh@uni-jena.de

Joachim Denzler
Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Germany

Mirco Migliavacca
European Commission Joint Research Centre, 1050 Brussels, Belgium

Abstract

Ecosystem respiration (Reco) represents a significant component of the global carbon cycle, and accurate characterization of its dynamics is essential for a comprehensive understanding of ecosystem-climate interactions and the impacts of climate extremes on the ecosystem. In this paper, we present a novel data-driven and physics-aware method for estimating Reco dynamics using the dynamic mode decomposition with control input (DMDc), an emerging tool for analyzing nonlinear dynamical systems. The proposed model represents Reco as a state space model with an autonomous component and an exogenous control input. The control input can be any ecosystem driver(s), such as air temperature, soil temperature, or soil water content. This unique modeling approach allows controlled intervention to study the effects of different inputs on the system. We further discuss using time delay embedding of both components to characterize Reco dynamics. Experimental results using Fluxnet2015 data show that the prediction accuracy of Reco dynamics achieved with DMDc is comparable to state-of-the-art methods, making it a promising tool for analyzing the dynamic behavior of different vegetation ecosystems on multi-temporal scales in response to different climatic drivers.

1 Introduction

Carbon losses from ecosystems affect climate change (Melillo et al., 2017). Ecosystem respiration (Reco), the sum of autotrophs (plants) and heterotrophs (bacteria, fungi, and animals) respiration, represents a significant component of the global carbon cycle. Accurate estimation of Reco dynamics is necessary for a better understanding of ecosystem-climate interactions and climate extremes' impact on ecosystems. This paper presents a data-driven yet physics-aware method for estimating Reco dynamics using the dynamic mode decomposition with control input (Proctor et al., 2016), an emerging tool for analyzing nonlinear dynamical systems.

Ecosystem respiration is typically described as an exponential function of temperature based on the law of thermodynamics (Lloyd & Taylor, 1994). However, this exponential relationship is defined based on some parameters, such as temperature sensitivity, which are assumed to remain constant. Several studies have pointed to the dependence of these parameters on other drivers of ecosystems (Mahecha et al., 2010), which could lead to either over- or under-estimation of Reco depending on the temperature range. Regression models partially compensate for such problems by using temporally moving windows to estimate the model's parameters (Reichstein et al., 2012).

The eddy covariance (EC) technique is widely used to measure the net ecosystem exchange (NEE), which is the difference between the total CO_2 release due to all respiration processes and the gross

carbon uptake by photosynthesis (GPP) (Baldocchi et al., 2001). The two CO2 fluxes, Reco and GPP, are derived using what is known as NEE partitioning methods (Lloyd & Taylor, 1994; Reichstein et al., 2005; Pastorello, 2020). These approaches use NEE measurements to fit physiologically based nonlinear relationships to estimate GPP and Reco using meteorological drivers. These functional relationships are either a light-response-based function linking global incoming radiation and GPP (Gilmanov et al., 2003) or respiration functions based on temperature (Reichstein et al., 2012). These methods are adopted as standard processing tools in the Fluxnet community (Pastorello, 2020) and are referred to as daytime (DT) and nighttime (NT) methods, respectively.

Recently, the capabilities of deep neural networks have been used for modeling Reco dynamics (Tramontana et al., 2020; Trifunov et al., 2021). These approaches provide data-driven equation-free estimates of Reco with the flexibility to include other meteorological and biological drivers affecting Reco. Despite their improved performance, these methods require sufficient training data and experimental tuning of the hyper-parameters of the deep networks used. Hence, the trained model can not consider short-term variations in ecosystem respiration.

Koopman operator (Koopman, 1931) enables the transformation of finite-dimensional nonlinear system dynamics to an infinite-dimensional linear dynamical system. However, finding the Koopman operator's eigenfunctions remains a significant obstacle to its implementation. Dynamic mode decomposition (DMD) is a simple numerical algorithm approximating the Koopman operator with a best-fit linear model that advances measurements from one time step to the next (Mezić, 2005; Schmid & Peters, 2010; Rowley et al., 2009; Brunton et al., 2017). It is an equation-free system identification method where the underlying dynamics of the system are learned from snap-shots in time of measurement data. DMD decomposes system dynamics into temporal modes whereby each mode represents a fixed oscillation frequency and decay/growth rate. The extension of DMD to dynamical systems with control inputs is the dynamic mode decomposition with control (DMDc) (Proctor et al., 2016). DMDc provides a robust framework for analyzing the behavior of complex dynamical systems under the influence of external inputs and has been successfully applied in diverse nonlinear dynamical systems, e.g., fluids dynamics and neuroscience.

In this paper, we extend our initial work (Shadaydeh et al., 2022) and present a novel data-driven Reco estimation approach that can also serve as an NEE partitioning method using DMDc. The system state Reco is presented as a state space dynamical model with an autonomous component and an exogenous component that serves as a control input function. The control input to the system could be air temperature (Tair) per the thermodynamics-based exponential function and/or other observed drivers such as soil temperature, soil water contents (SWC), etc. Such modeling of Reco allows disentangling the exogenous effect of the control input, e.g., Tair, from the autonomous dynamics of Reco. Subsequently, it facilitates intervention studies with different control inputs to study their effect on the ecosystem. To deal with unobserved drivers of Reco other than temperature or any used control input, we use time delay embedding (TDE) of the history of the system's state and control input. According to Takens Theory (Takens, 1981), embedding the state history guarantees under certain conditions that the system will learn the trajectories of the original system. The model training is performed in a sliding temporal window on night-time NEE (GPP ≈ 0) of Fluxnet2015 data (Pastorello, 2020). Reco day- and nighttime values, within forecast intervals that span one to several days, are then predicted using the trained model with the future values of the control input(s). The performance of the DMDc with different control inputs, namely Tair and SWC, is compared to DT and NT NEE partitioning methods reported in the Fluxnet2015 data. The impact of TDE on the performance of the proposed DMDc approach is also investigated for Reco's long-term forecasting.

2 METHODOLOGY

2.1 DYNAMIC MODE DECOMPOSITION WITH CONTROL (DMDC)

DMD (Schmid & Peters, 2010) analyzes the relationship between pairs of measurements from a dynamical system. The measurements, $\mathbf{x}(k)$ and $\mathbf{x}(k+1)$, where k indicates the temporal iteration from a discrete dynamical system, are assumed to be approximately related by a linear operator $\mathbf{A}$ such that the relation $\mathbf{x}(k+1) \approx \mathbf{A}\mathbf{x}(k)$ holds for all pairs of measurements k, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$. Let $\mathbf{X} = [\ \mathbf{x}(1) \quad \mathbf{x}(2) \quad \cdots \quad \mathbf{x}(M-1)\ ]$ and the time-shifted matrix $\mathbf{X}' = [\ \mathbf{x}(2) \quad \mathbf{x}(3) \quad \ldots \quad \mathbf{x}(M)\ ]$ where M is the total number of snapshots. This relationship can be described in a matrix form as $\mathbf{X}' \approx \mathbf{A}\mathbf{X}$. The best-fit linear operator is computed as $\mathbf{A} =$

$\mathbf{X}'\mathbf{X}^\dagger$ where † is the Moore-Penrose pseudoinverse which is computed using the singular value decomposition (SVD) of $\mathbf{X}$. The eigendecomposition of $\mathbf{A}$ yields the eigenvalues and eigenvectors (modes) that can be used for future state prediction. Specifically, let ϕ_j and λ_j be the jth eigenvector-eigenvalue pair of $\mathbf{A}$, $\mathbf{x}(k)$ at any time k is defined as

$$\mathbf{x}(k) = \sum_{j=1}^{N} b_j \phi_j \lambda_j^k \tag{1}$$

where b_j denotes the weights of the modes in the initial state $\mathbf{x}(0)$.

DMDc (Proctor et al., 2016) extends the DMD to dynamical systems with control inputs. The goal of DMDc is to analyze the relationship between a future system measurement $\mathbf{x}(k+1)$ with the current measurement $\mathbf{x}(k)$ and the current control input $\mathbf{u}(k)$ such that $\mathbf{x}(k+1) \approx \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k)$, where $\mathbf{x}(k) \in \mathbb{R}^n, \mathbf{u}(k) \in \mathbb{R}^l, \mathbf{A} \in \mathbb{R}^{n \times n}$, and $\mathbf{B} \in \mathbb{R}^{n \times l}$. Let $\mathbf{\Upsilon} = [\ \mathbf{u}(1)\ \ \mathbf{u}(2)\ \ \cdots\ \ \mathbf{u}(m-1)\ ]$. Utilizing the three data matrices, DMDc is focused on finding best-fit approximations to the mappings $\mathbf{A}$ and $\mathbf{B}$ such that

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{\Upsilon} \tag{2}$$

holds for all trios of data. The augmented snapshot matrix $[\mathbf{X}; \mathbf{\Upsilon}]$ is decomposed using SVD to extract the dynamic modes of the system, including the effects of the control inputs. Once the dynamic modes are identified, they can be used to reconstruct the system's evolution under the influence of control inputs. The DMDc steps are summarised in A.1.

2.2 Modeling the dynamics of Reco using DMDc

In Eq. 2, let the state of the system $\mathbf{X}$ and the control input $\mathbf{\Upsilon}$ be two vectors which constitute of M nights measurement of nighttime NEE (NEEnight) and the control input, e.g., Tair, respectively. Specifically, for Tair as a control input, we define

$$\begin{aligned} \mathbf{X} &= [\ \text{NEEnight}(t_1)\ \ \text{NEEnight}(t_2)\ \ \cdots\ \ \text{NEEnight}(t_{M-1})\ ] \\ \mathbf{\Upsilon} &= [\ \text{Tair}(t_1)\ \ \text{Tair}(t_2)\ \ \cdots\ \ \text{Tair}(t_{M-1})\ ]. \end{aligned} \tag{3}$$

Here, NEEnight(t_i), $i = 1, \cdots, M$ refers to a vector of nighttime measurement of the day t_i. The matrices $\mathbf{A}$ and $\mathbf{B}$ in Eq. 2 are identified using the steps of the DMDc algorithm summarized in A.1.

The dynamical model of Reco is learned using the DMDc method applied to the values of NEEnight (when photosynthesis, and therefore GPP, is assumed to be 0) and the selected control input, e.g., Tair, using a temporal sliding window of size M. The trained state space model is then used to forecast day and night time values of Reco using the day and night time values of the control input within the forecast horizon, which might span one to several days.

2.3 Time delay embedding (TDE) for characterization of Reco dynamics

Time delay embedding is a classical approach for dealing with partial state information. Most often, we can observe only partially the drivers affecting the state of the observed system, i.e., the ecosystem respiration system in this study. This method augments a state $\mathbf{x}$ with its history. According to Takens theory (Takens, 1981), under mild conditions on the observables, the dynamics of the augmented state are guaranteed to be diffeomorphic to the dynamics of the original state $\mathbf{x}$, provided that the embedding dimension N satisfies $N \geq 2n + 1$, where n is the dimension of the state. One method to obtain the TDE dimension N is by computing the SVD of a Hankel matrix (Brunton et al., 2017). The Hankel matrix is formed by stacking the columns of the snapshot matrices $\mathbf{X}$ with delay embedding, e.g., the Hankel matrix of NEEnight in A.2. Another advantage of using TDE in the proposed method is that it allows for learning Reco dynamics from short data as it compensates for using advanced measurement in time. This is relevant as it enables ecosystem forecasts to consider very transient factors affecting short-term variations in the system dynamics, such as perspiration.

The singular values of the Hankel matrix capture the energy or variance associated with each of the DMDc modes (Brunton et al., 2017). Thus, the number of dominant modes necessary for Reco forecast can be estimated using the rank of the Hankel matrix of NEEnight. Analyzing the spread of singular values of the Hankel matrix also provides valuable insights into the underlying system's dynamics. A well-structured spread of singular values indicates a well-observable or predictable
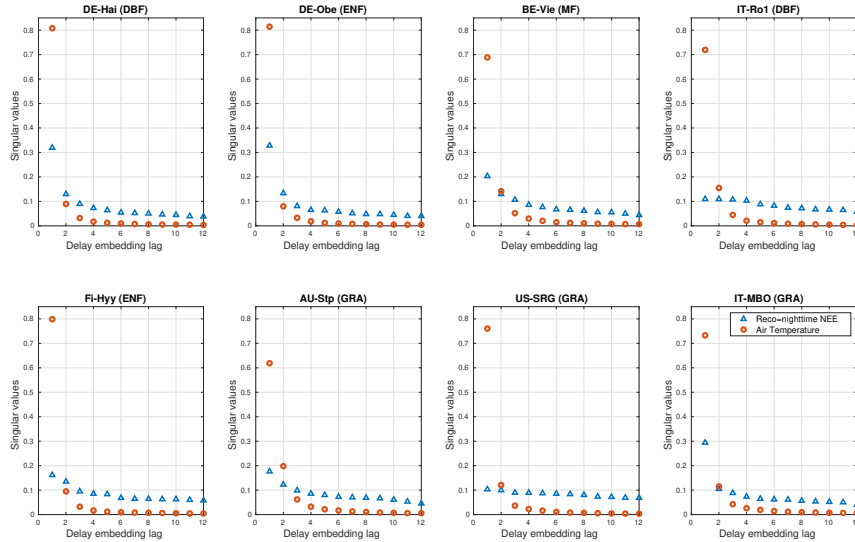
Figure 1: Singular values of the Hankel matrices H of NEEnight and Tair in different Fluxnet sites. The spread of the singular values is an indicator of the predictability of the system. The number of distinct singular values is the number of distinct dynamic modes. We can observe four to six dominant modes depending on the vegetation type and climate of the site location.

system. Smaller singular values indicate significant noise in the system. Comparing the spread of singular values of different sites can help characterize the differences in Reco dynamics in different vegetation and average temperature sites, as shown in the following section.

3 EXPERIMENTAL RESULTS AND DISCUSSION

We used the half-hourly EC Fluxnet2015 data (Pastorello, 2020) measured at Fluxnet sites with different vegetation types and average temperature to investigate the performance of our method and the impact of different control inputs, e.g., Tair and SWC, on Reco dynamics. In particular, we selected five different forests of various types: Deciduous Broadleaf Forests (DBF), Evergreen Needleleaf Forests (ENF), Mixed Forests (MF), as well as three Grassland (GRA) sites.

Since the direct measurement of daytime Reco is not available, the root mean square error (RMSE) between the Reco forecast values and NEEnight (ground-truth Reco assuming GPP=0 at night) during the nights of the forecast interval is used as a validation measure. We used the variable 'Night' of the Fluxnet data as a reference for nighttime. The method is applied on a temporal sliding window over the last two years of several sites of Fluxnet 2015 data from May to September (January to April for the southern hemisphere). Only good quality data with at least 80% of quality flag $Q < 2$ is used. All variables were min-max normalized and then set to the range $[-1, 1]$. We compare the performance of the proposed method using DMDc with and without control input (DMD) to NT and DT Reco values of the Fluxnet data.

Figure 1 shows the singular values of the Hankel matrices of NEEnight and Tair for different vegetation sites. In most sites, we can observe four to six dominant modes. Results of RMSE of one day forecast using five nights for training are given in Table 1. These results indicate a comparable performance of the proposed method to DT and NT methods. The spread of the singular values of the NEEnight Hankel matrix of the different sites shown in Figure 1 can explain the prediction accuracy of our method. We can notice that sites with well-spread singular values, such as DE-Hai (DBF) and BE-Vie(MF), are more predictable, and hence our proposed DMDc-based method outperformed DT and NT methods. However, for sites with less structured singular values, such as IT-Ro1(GRA) and AU-Stp (GRA), we can observe a slightly lower accuracy of DMDc. This is expected since the

Table 1: Comparison of different methods' nighttime Reco forecasts to the ground truth NEEnight. We show the NEEnight estimation RMSE (μmol CO_2 m^{-2} s^{-1}) metric for all methods and vegetation types. The performance of DT and NT NEE partitioning methods are compared to those of DMD and DMDc when using five nights of NEEnight for training and one-day Reco forecast, i.e., validation on the next night. The best results are in bold.

Method (Control Input)	DE-Hai (DBF)	DE-Obe (ENF)	BE-Vie (MF)	IT-Ro1 (DBF)	FY-Hyy (ENF)	AU-Stp (GRA)	US-SRG (GRA)	IT-MBO (GRA)
Fluxnet NT Reco	0.192	**0.190**	0.143	0.147	0.204	0.096	0.163	0.127
Fluxnet DT Reco	0.251	0.209	0.173	**0.126**	0.221	**0.091**	**0.122**	0.165
DMD	0.195	0.222	0.140	0.144	0.191	0.098	0.174	0.130
DMDc(Tair)	**0.182**	0.207	**0.137**	0.145	**0.188**	0.099	0.175	**0.126**
DMDc(SWC)	0.191	0.213	0.141	0.145	0.197	0.098	0.171	0.127

Table 2: Comparison of different methods' nighttime Reco forecasts to the ground truth NEEnight. We show the NEEnight estimation RMSE (μmol CO_2 m^{-2} s^{-1}) metric for all methods and vegetation types. The performance of DMD and DMDc with Tair and SWC as control inputs and different embedding dimension N are compared using two weeks of training data and two weeks of Reco forecast, i.e., validation on the nights of the next two weeks. The best results are in bold.

Method (Control input, embedding dimension)	DE-Hai (DBF)	DE-Obe (ENF)	BE-Vie (MF)	IT-Ro1 (DBF)	FY-Hyy (ENF)	AU-Stp (GRA)	US-SRG (GRA)	IT-MBO (GRA)
DMD	0.216	0.323	0.199	0.193	0.279	**0.114**	0.242	0.171
DMDc (Tair)	0.235	0.301	**0.189**	**0.186**	0.274	0.115	0.243	**0.161**
DMDc-TDE (Tair, N=4)	0.230	0.297	0.191	0.188	0.271	0.115	0.240	0.162
DMDc-TDE (Tair, N=6)	0.231	**0.294**	0.190	0.190	**0.260**	0.117	0.236	0.163
DMDc(SWC)	0.219	0.326	0.198	0.195	0.296	0.115	0.230	0.180
DMDc-TDE (SWC, N=4)	0.216	0.316	0.194	0.204	0.281	0.116	0.225	0.180
DMDc-TDE (SWC, N=6)	0.219	0.315	0.198	0.192	0.279	0.122	**0.219**	0.181

low spread of singular values indicates the absence of distinct modes, resulting in less predictability. The low spread of singular values could result from significant noise in the system or slow climate dynamics in the site location. It should be noted, however, that the DMD method on its own, without any control input, can still produce reasonable accuracy, which is comparable to the DT and NT methods in some sites. Our results also show a promising performance in capturing the influence of different ecosystem drivers. We can notice that SWC is a better predictor of Reco dynamics for grasslands or higher average temperature sites, while Tair is a better predictor for forest sites.

In a second set of experiments, we aim to investigate the impact of using time delay embedding to improve the accuracy of longer Reco forecast intervals. To this end, we train the DMDc state-space model with two weeks of NEEnight measurement data and forecast the next two weeks. Table 2 shows the results of DMD and DMDc for two different control inputs, Tair and SWC, and two TDE dimensions, N=4 and 6. These two embedding values are selected based on the number of distinct singular values we can observe in Figure 1. Results indicate that, on the one hand, Tair is a better long-term predictor in most studied sites. On the other hand, we can observe that TDE enhanced the prediction accuracy with SWC as the control input for all studied sites.

4 CONCLUSIONS

We proposed a data-driven model for analyzing ecosystem respiration dynamics using DMDc. Such modeling of Reco allows disentangling the exogenous effect of the control input, e.g., Tair, from the autonomous dynamics of Reco. Subsequently, it facilitates interventional studies on the control input to study its effect on the ecosystem. The performance of the DMDc with different control inputs, namely Tair and SWC, is compared to DT and NT NEE partitioning methods reported in the Fluxnet2015 data for different vegetation sites. Our results show a promising performance of our approach in the characterization of Reco dynamics and in capturing the influence of different ecosystem drivers, making it a promising tool for analyzing the dynamic behavior of different vegetation ecosystems in response to climate change. Experimental results also showed that for most sites, time delay embedding of different ecosystem drivers could improve long-term prediction accuracy. Current research efforts are focused on implementing the proposed approach on multitemporal scales to improve its long-term forecast accuracy.

5 ACKNOWLEDGMENT

REFERENCES

D. Baldocchi, E. Falge, L. H. Gu, et al. Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, 82(11):2415–2434, 2001.

Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, Eurika Kaiser, and J. Nathan Kutz. Chaos as an intermittently forced linear system. *Nature Communications*, 8:19, May 2017. doi: 10.1038/s41467-017-00030-8.

Tagir G. Gilmanov, Shashi B. Verma, Phillip L. Sims, Tilden P. Meyers, James A. Bradford, George G. Burba, and Andrew E. Suyker. Gross primary production and light response parameters of four southern plains ecosystems estimated using long-term co2-flux tower measurements. *Global Biogeochemical Cycles*, 17(2), 2003. doi: https://doi.org/10.1029/2002GB002023.

B. O. Koopman. Hamiltonian systems and transformation in hilbert space. 17(5):315–318, 1931. ISSN 0027-8424. doi: 10.1073/pnas.17.5.315.

J. Lloyd and J. A. Taylor. On the temperature dependence of soil respiration. *Functional Ecology*, 8 (3):315–323, 1994. ISSN 02698463, 13652435.

Miguel D Mahecha, Markus Reichstein, Nuno Carvalhais, Gitta Lasslop, Holger Lange, Sonia I Seneviratne, Rodrigo Vargas, Christof Ammann, M Altaf Arain, Alessandro Cescatti, et al. Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science*, 329(5993): 838–840, 2010.

J. M. Melillo, S. D. Frey, K. M. DeAngelis, W. J. Werner, M. J. Bernard, F. P. Bowles, G. Pold, M. A. Knorr, and A. S. Grandy. Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, 358(6359):101–105, 2017. doi: 10.1126/science.aan2874.

Igor Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dynamics*, 41:309–325, 2005.

Gilberto et al Pastorello. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, 7(1):225, 2020. ISSN 2052-4463. doi: 10.1038/s41597-020-0534-3.

Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016. doi: 10.1137/15M1013857.

Markus Reichstein, Eva Falge, Dennis Baldocchi, et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology*, 11(9):1424–1439, 2005.

Markus Reichstein, Paul C. Stoy, Ankur R. Desal, et al. Partitioning of net fluxes. In Marc Aubinet, Timo Vesala, and Dario Papale (eds.), *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, pp. 263–289. Springer Netherlands, 2012. ISBN 978-94-007-2350-4.

Clarence W. Rowley, Igor Mezic, Shervin Bagheri, Philipp Schlatter, and Dan S. Henningson. Spectral analysis of nonlinear flows. *Journal of Fluid Mechanics*, 641:115–127, 2009. doi: 10.1017/S0022112009992059.

Schmid and Peters. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. doi: 10.1017/S0022112010001217.

Maha Shadaydeh, Joachim Denzler, and Mirco Migliavacca. Partitioning of net ecosystem exchange using dynamic mode decomposition and time delay embedding. *Engineering Proceedings*, 18(1), 2022. doi: 10.3390/engproc2022018013.

Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young (eds.), *Dynamical Systems and Turbulence, Warwick 1980*, pp. 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg. ISBN 978-3-540-38945-3.

Gianluca Tramontana, Mirco Migliavacca, Martin Jung, et al. Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. *Global Change Biology*, 26(9):52355–5253, 2020.

Violeta Teodora Trifunov, Maha Shadaydeh, Jakob Runge, Markus Reichstein, and Joachim Denzler. A data-driven approach to partitioning net ecosystem exchange using a deep state space model. *IEEE Access*, 9:107873–107883, 2021. doi: 10.1109/ACCESS.2021.3101129.

A APPENDIX

A.1 MAIN STEPS OF THE DMDC METHOD PROCTOR ET AL. (2016)

1. Collect the system measurement and control snapshots
2. Form the matrices $\mathbf{X}, \mathbf{X}'$, and $\mathbf{\Upsilon}$
3. Stack the data matrices $\mathbf{X}$ and $\mathbf{\Upsilon}$ to construct the matrix $\mathbf{\Omega}$
4. Compute the SVD of $\mathbf{\Omega}$ thereby obtaining the decomposition
$\mathbf{\Omega} \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^*$ with truncation value p.
5. Compute the SVD of $\mathbf{X}'$, thereby obtaining the decomposition
$\mathbf{X}' \approx \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}^*$ with truncation value r
6. Compute the following:

$$\tilde{\mathbf{A}} = \hat{\mathbf{U}}^*\mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{U}}_1^*\hat{\mathbf{U}} \tag{4}$$

$$\tilde{\mathbf{B}} = \hat{\mathbf{U}}^*\mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{U}}_2^* \tag{5}$$

7. Perform the eigenvalue decomposition

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda} \tag{6}$$

8. Compute the dynamic modes of the operator $\mathbf{A}$.

$$\mathbf{\Phi} = \mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{U}}_1^*\hat{\mathbf{U}}\mathbf{W} \tag{7}$$

A.2 HANKEL MATRIX OF NEENIGHT

The Hankel matrix of NNnight is defined as

$$\mathbf{H} = \begin{bmatrix} \text{NEEnight}(t_1) & \text{NEEnight}(t_2) & \cdots & \text{NEEnight}(t_{M-N}) \\ \text{NEEnight}(t_2) & \text{NEEnight}(t_3) & \cdots & \text{NEEnight}(t_{M-N+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{NEEnight}(t_N) & \text{NEEnight}(t_{N+1}) & \cdots & \text{NEEnight}(t_M) \end{bmatrix} \tag{8}$$