Geoscientific
Model Development

Methods for assessment of models

Data-driven discovery and model reduction methods for the atmospheric effects of high altitude emissions

Jurriaan A. van 't Hoff[1], **Tom S. van Cranenburgh**[1], **Urban Fasel**[2], and **Irene C. Dedoussi**[1,3]

[1]Operations & Environment, Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, the Netherlands
[2]Department of Aeronautics, Faculty of Engineering, Imperial College London, Exhibition Road, SW7 2AZ, London, UK
[3]Department of Engineering, University of Cambridge, 1 JJ Thomson Avenue, Cambridge, CB3 0DY, Cambridge, UK

Correspondence: Irene C. Dedoussi (icd23@cam.ac.uk)

Abstract. Chemistry transport models play a crucial role in the evaluation of the effect of anthropogenic emissions on the atmosphere and climate, but they come with high computational costs and require specialized know-how. This renders them impractical for applications in multidisciplinary optimisation, or regulatory and operational decision-making processes where environmental effects are to be considered. Such applications require computationally efficient surrogate models of the complex chemistry transport models. Here we investigate the use of data-driven discovery and reduced-order modelling methods for this purpose. Specifically, we examine the dynamic mode decomposition (DMD) and proper orthogonal decomposition coupled with the sparse identification of non-linear dynamics (POD-SINDy). We evaluate their ability to reconstruct and forecast changes in the distribution of ozone in response to the introduction of supersonic aircraft as modelled by the GEOS-Chem chemistry transport model. Of the tested methods, we find that optimized DMD and bagging optimized DMD with constrained eigenvalues perform best. These methods can reconstruct and forecast full-atmospheric ozone responses for up to several years without losing stability, at smaller errors than estimates using the spatio-temporal mean of the data. On average, the constrained optimized DMD method reduces the reconstruction error by 63.5 % and that of forecasting by 25.8 % compared to the spatio-temporal mean. For the constrained bagging optimized DMD these reductions are 45.0 % and 23.1 %, respectively. The resulting change in global ozone column, calculated from the reconstructed atmospheres, has an error smaller than 10 %. This is achieved while reducing the computational and storage requirements by several orders of magnitude, which may be a worthwhile tradeoff for some applications.

1 Introduction

Chemistry transport models (CTMs) and climate chemistry models (CCMs) are critical to our understanding of how anthropogenic emissions affect the environment and climate. These models simulate the chemistry, transport, and deposition of hundreds of chemical species in the atmosphere under evolving meteorological conditions, allowing them to capture complex chemical feedback mechanisms and responses. This level of complexity requires specialised expertise and comes with a considerable computational cost, which is why these models usually require access to dedicated high-performance computer systems. Despite this barrier, CTMs and CCMs are widely used to study a variety of problems in atmospheric composition and chemistry, although their use is often limited to the evaluation of a handful of case-studies.

In the consideration of new technologies, such as new aircraft concepts, multiple parties have interest in the evaluation of CTMs and CCMs. Engineers may want to integrate environmental assessment into multidisciplinary optimisation (MDO) approaches to minimise environmental effects in their design process. Regulatory bodies may be interested in evaluating the effectiveness of certain regulations, and operators may want to know if environmental effects can be reduced through operations. These parties often take iterative

approaches, that may require up to hundreds or thousands of cycles over their applications. Considering that CTMs and CCMs often require days, if not weeks, to perform their evaluations, they are impractical to integrate in these approaches, even if the technical know-how to use them is already present. Therefore, to support the integration of environmental evaluations into MDO approaches or regulatory and operational decision-making processes, we need easy-to-use alternatives to replicate or estimate the output of these models at a fraction of their computational cost.

The recent developments in the field of machine learning have led to new methods that may be suitable for low-cost surrogate models, but many of them (e.g. neural networks) require vast amounts of data, which is impractical to generate with CTMs and CCMs. Data-driven model discovery and reduced-order modelling methods, that require less data to operate and are more interpretable than "black-box" neural network models, can be suitable alternatives. Two methods of interest are Dynamic Mode Decomposition (DMD) and Proper Orthogonal Decomposition (POD) coupled with the Sparse Identification of Non-Linear Dynamics (SINDy). These methods extract low-dimensional features from data that can be used for analysis and to construct reduced-order models. This application is similar to linear inverse modelling methods, which have been used before to model coupled atmospheric-oceanic systems (Perkins and Hakim, 2020), atmospheric flows (Kwasniok, 2022) and surface temperatures (Newman, 2013), but DMD and POD-SINDy are more widely used in the field of fluid mechanics (e.g. Taira et al., 2017, 2020; Champion et al., 2019; Khodkar and Hassanzadeh, 2021; Callaham et al., 2022). Atmospheric processes share some similarities with fluid flows, and earlier work has shown that these methods may be transferable. For example, Yang et al. (2024) have shown that dimensionality reduction techniques can be combined with SINDy to reproduce atmospheric data from CTMs, and Velegar et al. (2024) have shown that DMD methods can be used to produce efficient reduced-order models to forecast the concentration of several tropospheric chemicals over the course of multiple months. These initial explorations show promising results, but the application of the methods is still limited to small subsets of CTM data.

We expand on this exploration by assessing the suitability of the DMD and POD-SINDy methods on large-scale and comprehensive datasets that describe chemistry and transport processes across the entire atmosphere. Where earlier work assessed the capability of these methods to reproduce parts of unperturbed atmospheres, we evaluate their use on data describing perturbations over the entire atmosphere. Specifically, we apply the methods to datasets that describe the difference of the ozone (O_3) distribution in response to different supersonic aircraft scenarios (van 't Hoff et al., 2025c, 2024b). The aircraft attributable ozone effect is identified as the difference between two independent CTM evaluations (one with supersonic aircraft emissions, and one

without). If data-driven methods can reconstruct and predict these differences, they may valuable to support the analysis of adoption scenarios and more efficient exchange and embedding of the spatiotemporal data. If stable models can be found this may also open up avenues for future research to directly use these methods to support MDO applications by combining multiple dynamical models. In this work, we aim to investigate the potential of these methods through a stepwise approach, starting from their interpretability for the analysis of atmospheric processes, followed by their reproductions of data in specific locations and the entire atmosphere. We also assess the possibility of calculating other metrics from the reproduced atmosphere, a capability that may be valuable for several applications.

2 Case studies

We use seven datasets that describe the changes in global ozone distribution in response to the adoption of supersonic aircraft. These datasets were generated using the Dutch national supercomputer Snellius using an estimate of around 493 920 CPU-hours. From van 't Hoff et al. (2024b) we use four datasets which describe the response to annual emissions equivalent to 8 Tg of annual fuel consumption by supersonic aircraft, originating from different flight corridors. These are the transatlantic flight corridor (TAC) and the corridor over the southern Arabian sea (SAS). In both cases we use datasets with emission altitudes of 16.2 and 20.4 km. These are therefore denoted as SAS204, TAC204, SAS162, and TAC162 respectively. These datasets are generated with version 13.3.1 of the GEOS-Chem CTM and represent 10-year of daily changes in global ozone mass at a horizontal resolution of $4° \times 5°$ (latitude $\times$ longitude) with 72 vertical pressure levels.

We also use three other datasets from van 't Hoff et al. (2025c). These also describe changes in the distribution of the ozone mass over 10 years, but in response to emissions from a global supersonic aircraft fleet rather than from a regional source. Due to the global distribution of these emissions we expect that these datasets may exhibit more challenging dynamical behaviour compared to the regional emission datasets. These datasets are generated with version 14.3.0 of the high-performance GEOS-Chem model, with a horizontal resolution of $2° \times 2.5°$ and 72 vertical pressure levels. Following the notation of van 't Hoff et al. (2025c) we denote these datasets as S1, S2, and S3. These represent the change in ozone in response to a theoretical supersonic fleet operating at Mach 2.0 with cruise altitudes of 16.5 to 19.5 km, one with increased NO_x emissions, and one with a lower cruise altitude and speed (Mach 1.6, 14–16.5 km), respectively. Detailed descriptions of the emission scenarios and chemistry transport model setups are provided in the related works (van 't Hoff et al., 2024b, 2025c).

3 Methodology

The methodology for this work is outlined in Fig. 1. Subsequent subsections will discuss the preprocessing of data and the model discovery methods.

3.1 Data description and processing

The datasets that we use contain four-dimensional (time, longitude, altitude, latitude) descriptions of the daily-averaged changes in ozone mass over the course of 10 years. We take several pre-processing steps to reduce the complexity of our data. All datasets exhibit an initial transient response to the introduction of the supersonic aircraft emissions, followed by a stabilized response where the new atmospheric quasi-equilibrium is reached. We first isolate the stable response from the data by removing the transient part, which spans the first five years. Secondly, we reduce the dimensionality of the dataset by calculating the longitudinally-averaged ozone response. Longitudinal averaging is standard practice in studies of the environmental effects of supersonic aircraft, because in the multi-annual timespans considered, the atmosphere may be considered as well-mixed over the longitude (Zhang et al., 2021a, 2023; Eastham et al., 2022). Thirdly, the GEOS-Chem model uses a non-uniform vertical grid that is denser near the surface. This is not uncommon for CTMs, but in this case study the majority of the ozone response occurs in the coarser stratospheric grid. To avoid over-representation of the near-surface conditions in the data, we therefore interpolate the data to a normalized vertical grid with a resolution of 1 km. Finally, we discard the grid cells above 51 km altitude, as this is the vertical limit of GEOS-Chem's extensive chemical solver for ozone (Eastham et al., 2014).

The data may be considered as combination of a dominant steady-state change in the ozone distribution and a dynamic component that is driven by seasonal and day-to-day changes in meteorology. Across all datasets, the steady-state response exhibits similar characteristics, with increases of ozone in the lower stratosphere and depletion of ozone in the upper stratosphere, yielding a net-loss of the global ozone column. For detailed explanations of the mechanisms behind these responses, we refer to the related articles (van 't Hoff et al., 2024b, 2025c). Figure 2 shows the average steady state for one of the datasets (TAC204) alongside four seasonal averages, highlighting that the ozone response shifts seasonally. This shift is part of the dynamic component, that we isolate by subtracting the temporal mean of the data similar to Velegar et al. (2024) and Yang et al. (2024). This mean is re-added after the dynamics are predicted to recompose the full signal. The data is considered as a matrix of snapshots $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ x(t_1) & x(t_2) & \cdots & x(t_m) \\ | & | & & | \end{bmatrix}. \tag{1}$$

The matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ represents the ozone response, with n representing the product of the latitude and altitude dimensions and m the number of daily snapshots. In this representation, the latitude and altitude are flattened into a single column $x(t_y)$. To evaluate the modelling of the entire atmosphere, we also calculate the mean global change in the ozone columns in Dobson Units (DU). This metric is also commonly used in studies of the environmental effects of supersonic aircraft (Zhang et al., 2023, 2021a, b; Eastham et al., 2022; van 't Hoff et al., 2025c).

3.2 Proper orthogonal decomposition

The POD originates from applications in the field of fluid mechanics and it uses the singular value decomposition (SVD) in the context of spatiotemporal data (Brunton and Kutz, 2021; Weiss, 2019). Applied to spatiotemporal data, the POD returns spatial patterns based on the variance in the data, also referred to as modes, and an associated temporal evolution of these modes. The spatial modes are hierarchically ordered, making truncated POD approximations of the high-dimensional data straightforward by isolating the first r spatial modes. The POD seeks to find an approximation of the mass distribution of ozone $\mathbf{X}$ at spatial dimension x (both latitude and altitude) and time t:

$$\mathbf{X}(x, t) = \overline{\mathbf{X}}(x) + \sum_{k=1}^{r} a_k(t) \boldsymbol{\psi}_k(x), \tag{2}$$

where $\overline{\mathbf{X}}(x)$ is the temporal mean of ozone response, $\boldsymbol{\psi}_k(x)$ the kth POD spatial mode, and $a_k(t)$ the POD temporal coefficient of the kth mode at time t. By combining the product of the spatial modes and temporal coefficients up to a certain rank r, reduced-order representations of the original data can be reconstructed.

3.3 Sparse identification of non-linear dynamics (SINDy)

To forecast future behaviour, the future state of the temporal coefficient needs to be forecasted, and for this we use sparse identification of non-linear dynamics (SINDy) to find a dynamical system model for the temporal coefficients. The SINDy method can be applied to data to discover ordinary differential equations (ODEs) and partial differential equations (PDEs) which describe its dynamics (Brunton et al., 2016; Rudy et al., 2017). SINDy has been applied to a range of model discovery problems, including problems in fluid dynamics (Loiseau and Brunton, 2018), computational chemistry (Boninsegna et al., 2018; Yang et al., 2024), and climate models (Zanna and Bolton, 2020). It is often assumed that the equations that govern these systems are sparse and have only a few active terms in the dynamics, therefore SINDy uses sparse regression to identify the active terms out of a library of candidate model terms. We use SINDy to find analytical models for the dynamics of the temporal coefficients of the
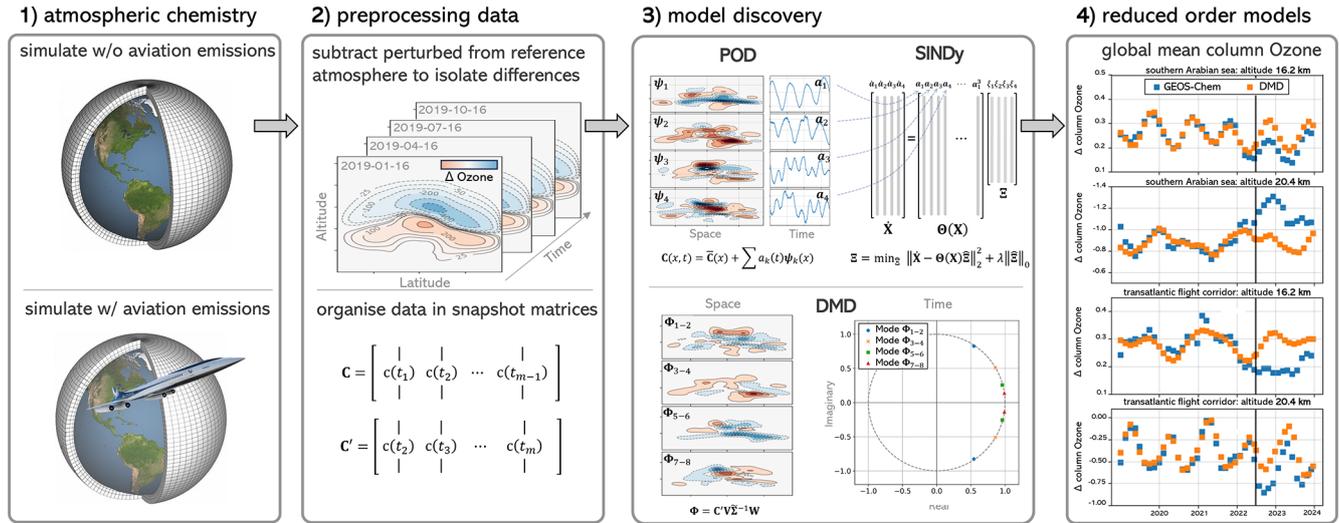
Figure 1. Methodology overview. In step 1 the effects of emissions are isolated by combining two GEOS-Chem atmosphere simulations, of which one is affected by the additional emissions. The resulting spatiotemporal datasets describe the daily average changes in ozone mass distribution for a period of 10 years. In step 2 the data is preprocessed and organised as a snapshot of matrices prior to applying data-driven methods (step 3). In step 4 the data-driven methods are used to forecast future behaviour of the ozone response, based on which the data-driven methods are assessed.
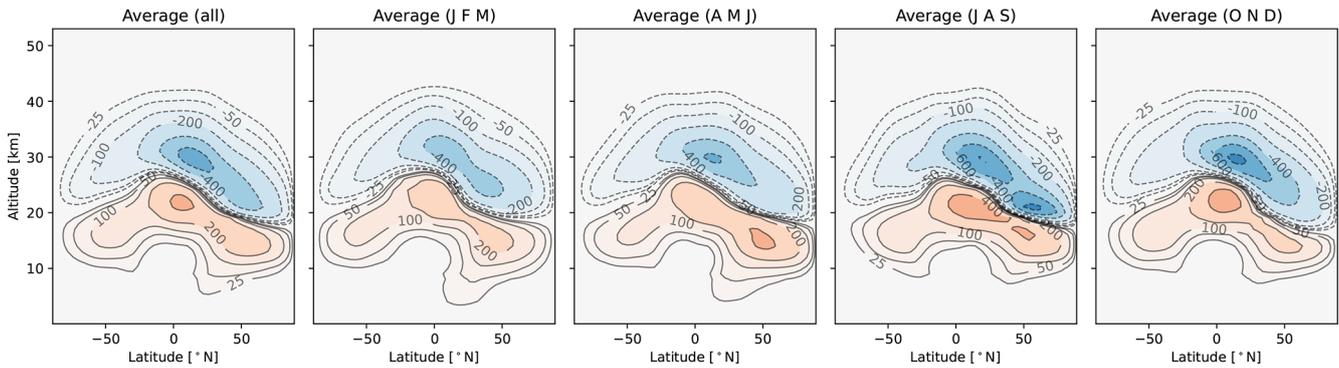

Figure 2. Average ozone response of the TAC204 dataset over different timespans in tons of ozone. The leftmost plot shows the average response over all 5 years of data, and the other four show averages taken over 3-monthly snapshots from the entire dataset. For example, the rightmost figure (Average (OND)) shows the average ozone response for the months of October, November, and December across the 5 years of data.

POD ($a_k(t)$ in Eq. 2). SINDy attempts to identify a system of first-order ODEs of the form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{a} = \dot{\mathbf{a}} = \mathbf{f}(\mathbf{a}) \approx \boldsymbol{\Theta}(\mathbf{a})\,\boldsymbol{\Xi}, \tag{3}$$

where the temporal coefficients of the POD ($\mathbf{a} = [a_1, a_2, \cdots, a_r]^T$) are the states of the dynamical system, and the derivatives ($\dot{\mathbf{a}}$) are approximated as a linear combination of nonlinear functions ($\boldsymbol{\Theta}(\mathbf{a})$) in a library with coefficients ($\boldsymbol{\Xi}$). The dynamics ($\mathbf{f}$) of the coefficients can be represented with a library of components. This library can be made up of any set of functions, with low-order polynomial terms commonly being used as a starting point. To identify an accurate and sparse model, the number of nonzero terms in $\boldsymbol{\Xi}$ should be as small as possible. To

find $\boldsymbol{\Xi}$, a penalized least-squares problem is solved using the sequentially thresholded least-squares algorithm (STLSQ) (Brunton et al., 2016):

$$\boldsymbol{\Xi} = \min_{\hat{\boldsymbol{\Xi}}} \left\| \dot{\mathbf{P}} - \boldsymbol{\Theta}(\mathbf{P})\hat{\boldsymbol{\Xi}} \right\|_2^2 + \lambda \| \hat{\boldsymbol{\Xi}} \|_0. \tag{4}$$

Here $\mathbf{P} = [\mathbf{a}(t_1), \mathbf{a}(t_2), \cdots, \mathbf{a}(t_m)]$ is the POD coefficient matrix, and $\dot{\mathbf{P}}$ is its derivative. STLSQ sequentially solves a least squares problem and includes a sparsity-promoting threshold λ, setting terms below the threshold to zero, approximating the L_0 norm in Eq. (4). To find suitable SINDy models, we therefore perform sequential grid searches over a range of threshold values, using the corrected Aikaike information criteria (AIC-c) (Mangan et al., 2017) to optimize the model

selection for sparsity. Specifically, we use up to three grid searches in sequence with 60 to 100 grid points over a range of thresholds. Thresholds with the lowest AIC-c score are then used to narrow the range of threshold values for the subsequent searches.

The SINDy method uses the derivative of the data, which in this case describes both longer term trends in the ozone response as well as day-to-day variations. The latter are of sufficient magnitude to affect the derivative, resulting in a noise-like layer on top of long-term trends. We therefore take additional pre-processing steps to improve the robustness of the SINDy method when applied to the temporal coefficients. We first reduce noise in the coefficients using a Savitzky–Golay filter (Savitzky and Golay, 1964) with a 60 d temporal window. Following this, the values of the time coefficients are normalized, and finally the time coefficients are temporally interpolated by a factor 10. These steps reduce the magnitude and variability of the derivatives, making them easier to model using the SINDy method. In addition to this pre-processing, we also apply the weak and ensembling forms of the SINDy algorithm that are more suitable for dealing with noisy data (Reinbold et al., 2020; Messenger and Bortz, 2021; Fasel et al., 2022). We use a sequential grid search over a range of threshold values to identify optimal sparse formulations based on the AIC-c. We do this for the standard SINDy, the ensemble, and the weak formulations. These methods are incorporated in the PySINDy Python package (de Silva et al., 2017; Kaptanoglu et al., 2021b).

3.4 Dynamic mode decomposition

The dynamic mode decomposition decomposes multidimensional spatiotemporal data into spatially coherent structures with linear temporal behaviour, each with an associated frequency of oscillation and rate of growth or decay. For the DMD method, two matrices of snapshots are arranged ($\mathbf{X}$ and $\mathbf{X}'$). These are similar to the $\mathbf{X}$ matrix of the POD method, but these matrices omit the final and first snapshots of the data respectively:

$$\mathbf{X} = \begin{bmatrix} | & | & & | \\ \mathbf{x}(t_1) & \mathbf{x}(t_2) & \cdots & \mathbf{x}(t_{m-1}) \\ | & | & & | \end{bmatrix} \tag{5}$$

$$\mathbf{X}' = \begin{bmatrix} | & | & & | \\ \mathbf{x}(t_2) & \mathbf{x}(t_3) & \cdots & \mathbf{x}(t_m) \\ | & | & & | \end{bmatrix}. \tag{6}$$

The DMD algorithm finds the best-fit linear operator $\mathbf{A}$ that relates these snapshot matrices.

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} \tag{7}$$

The following steps are followed when applying DMD:

1. The truncated singular value decomposition of matrix $\mathbf{X}$ is calculated, where $\tilde{\mathbf{U}} \in \mathbb{C}^{n \times r}$ and $\tilde{\mathbf{V}} \in \mathbb{C}^{m \times r}$ and $\tilde{\mathbf{\Sigma}} \in \mathbb{R}^{r \times r}$ depend on the chosen rank r.

$$\mathbf{X} \approx \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^* \tag{8}$$

2. Matrix $\mathbf{A}$ is then obtained by computing the pseudo inverse of $\mathbf{X}$. By projecting $\mathbf{A}$ on the SVD modes $\mathbf{U}$, the full matrix $\mathbf{A}$ does not have to be computed:

$$\tilde{\mathbf{A}} = \tilde{\mathbf{U}}^*\mathbf{A}\tilde{\mathbf{U}} = \tilde{\mathbf{U}}^*\mathbf{X}'\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}^{-1}. \tag{9}$$

3. The eigenvalues $\mathbf{\Lambda}$ of the reduced matrix $\tilde{\mathbf{A}}$ are the same as the eigenvalues of $\mathbf{A}$. The eigendecomposition of $\tilde{\mathbf{A}}$ is performed:

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}. \tag{10}$$

4. The DMD modes $\mathbf{\Phi}$ are reconstructed with the eigenvectors $\mathbf{W}$ and the time-shifted snapshot matrix of the original data $\mathbf{X}'$.

$$\mathbf{\Phi} = \mathbf{X}'\mathbf{V}\tilde{\mathbf{\Sigma}}^{-1}\mathbf{W} \tag{11}$$

To increase the noise robustness of the DMD algorithm, an optimised DMD (OptDMD) was introduced by Askham and Kutz (2018). This was further extended to bagging optimised DMD (BOPDMD) by Sashidhar and Kutz (2022). Both methods are incorporated in the open source Python package PyDMD (Demo et al., 2018; Ichinaga et al., 2024). In this work we assess the applicability of the standard DMD, standard and constrained OptDMD (OptDMD, C-OptDMD), and standard and constrained BOPDMD (BOPDMD, C-BOPDMD).

We split the steady-state component of the datasets (the last five years) into two separate sets. We use the first three and a half years to fit the methods and to test their ability to reproduce the original data. The latter one and a half years of data is used to test the methods' ability to forecast the future behaviour of the dynamical system. We evaluate the methods' forecast and reproduction of the ozone response in terms of the root mean squared error (RMSE). For comparison, we consider the RMSE that would be achieved if the temporal mean of the fitting data were used instead for reconstruction and forecasting. We use this for reference as the temporal mean is also commonly used to represent ozone response data in publications (van 't Hoff et al., 2024b; Eastham et al., 2022; Zhang et al., 2023; Speth et al., 2021), and it is the simplest method through which future behaviour can be estimated lacking more comprehensive modelling methodologies.

4 Results

We evaluate four different aspects of these reduced-order modelling and model discovery methods. First, we discuss their general application and use for analysis in Sect. 4.1.

We then evaluate the applicability of these methods in modelling tasks of different dimensionality and complexity. We start by modelling the ozone response in individual grid cells (Sect. 4.2), expanding to the modelling of the 2-dimensional zonally average ozone response (Sect. 4.3), and finally to the modelling of relevant metrics derived from the global ozone response, specifically the globally averaged ozone column response (Sect. 4.4). We note that the results shown in this section do not represent the best models that we found, but they represent the median model performance from 29,400 variations evaluated in a hyperparameter search (Appendix A, discussed further in Sect. 5).

4.1   Analysis of spatial modes

The POD and DMD methods produce spatial modes that represent patterns of dominant spatial features with coupled time coefficient series that describe how the spatial modes evolve over time. For the DMD method, the spatial modes represent patterns with similar temporal frequencies, whereas for the POD method the spatial modes represent features with high spatial correlations. Based on an extensive hyperparameter search (Appendix A), we present results for DMD methods with a rank eight and POD-SINDy methods with rank four. Considering that the DMD modes are coupled conjugate pairs, this results in DMD and POD-SINDy models with similar levels of complexity. Further increases in rank did not yield considerable improvements to the model discovery methods in terms of mean average errors. Examples of the first four spatial and temporal modes from the C-OptDMD and POD methods applied to the TAC204 dataset are shown in Fig. 3. Both of these methods identify different spatial modes and time dynamics, with the C-OptDMD method resulting in oscillating time coefficients of different frequencies, whereas POD results in more irregular time coefficients.

The spatial patterns identified by both methods are driven by changes in the ozone response over time. These are a result of the interactions of chemistry and transport processes, that are also affected by meteorological conditions and the distribution of the perturbed emissions in the atmosphere. This overlap of mechanisms makes it challenging to attribute spatial modes, or patterns therein, to any single process. The time coefficients may however inform on some of the drivers behind these modes. For example: the UV radiation that affects the ozone chemistry changes seasonally, driving seasonal fluctuations in the ozone perturbation (also shown in Fig. 2). In the northern hemisphere, this causes the average altitude of the positive lower-stratospheric ozone perturbation to oscillate yearly by up to 2 km (Fig. C1). This oscillation is likely captured by the annual modes of the POD and DMD methods, such as modes $\Phi_{3\text{--}4}$ of the C-OptDMD method and modes ψ_1 and ψ_2 of the POD method shown in Fig. 3. Another example is the change in seasonal strato- to tropospheric transport of ozone in the northern hemisphere, visible in mode ψ_1 of the POD. While indications of some

processes might be inferred, all spatial modes capture dynamic behaviour across both hemispheres, suggesting that they may capture multiple phenomena that have similar temporal frequencies. Other mechanisms that may be captured here are the seasonal variation of the tropopause altitude and atmospheric circulations that change the ozone distribution. The tendency to capture multiple processes at once limits the usability of POD and DMD modes for analysis on full atmospheric systems, at least for a species as complicated as ozone. It may be possible to identify the role of different processes in more restricted case studies, such as the formation of the ozone hole or the surface concentrations of ozone.

4.2   Individual grid cells

The modelling of individual grid cells may be of interest to several case studies, such as forecasting air quality metrics over specific areas. To assess this capability we consider the performance of the model discovery methods in replicating and forecasting the ozone perturbations for several individual grid cells over time. Figure 4 shows the reconstruction and forecast of 3 cells from the TAC204 dataset, located at the same latitude of 22° N at altitudes of 1.1, 18.3, and 26 km. We select these grid cells because they are located in the troposphere, tropopause, and stratosphere, respectively, in an active part of the Brewer-Dobson circulation. The TAC204 dataset is used as example because it has no perturbing emissions in any of these grid cells, and the dynamics in these cells are therefore related only to atmospheric transport of emission species and meteorology.

We find that the methods have different performance in the different atmospheric regimes. The reconstruction and forecast errors are shown in Table B1. In the lower troposphere (1.1 km altitude, 22° N), the ozone response exhibits a stable seasonal oscillation, as the response in this domain is primarily driven by the availability of residual UV radiation. This behaviour is generally reconstructed well by most methods, with RMSEs ranging from 0.42 t (C-OptDMD) to 1.59 t (POD-SINDy) of ozone. Using the data-mean, the reconstruction RMSE is 1.9 t of ozone at this altitude, meaning that in this case all data-driven methods provide an improvement over the propagation of the data-mean. Despite this, not all methods are numerically stable. Figure C2 shows the eigenvalues of the DMD variants applied in this work over all datasets. It shows that the DMD, OptDMD, and BOPDMD methods result in stable models when applied to most datasets (eigenvalues within the unit circle) while being unstable for some. The models with constrained eigenvalues (C-OptDMD, C-BOPDMD) are stable in all cases. The POD-SINDy models are generally less stable, even compared to the unstable DMDs. For example, Fig. 4 shows that the normal and ensemble variants POD-SINDy models stagnate after less than a year of propagation, whereas the weak POD-SINDy performs considerably better but has larger deviations compared to the other methods.
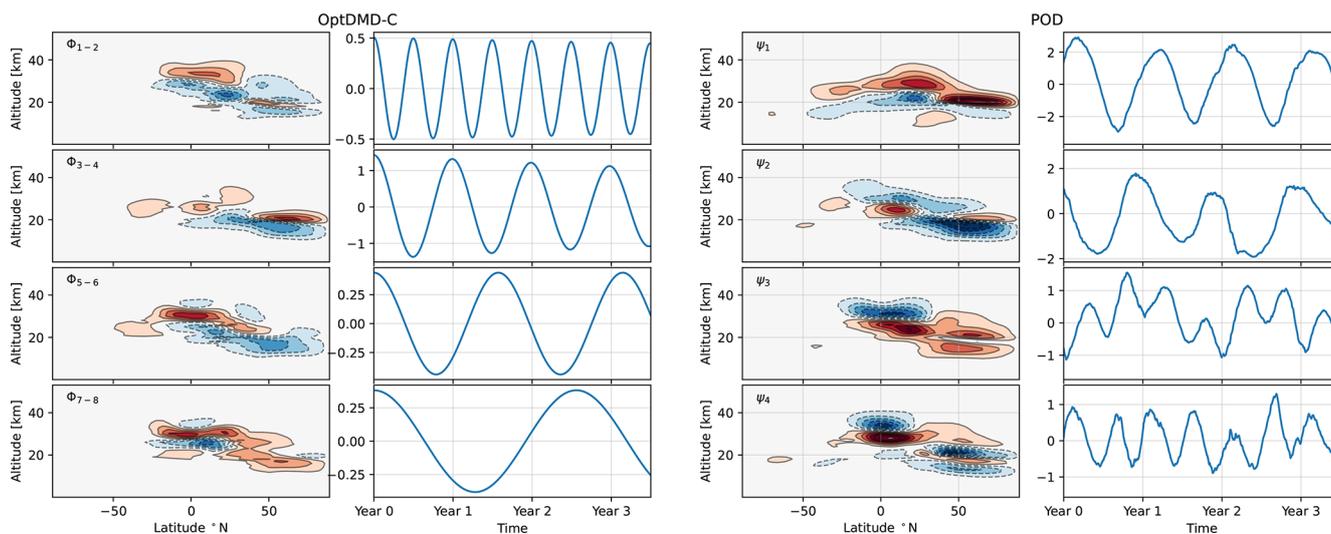
Figure 3. Spatial modes 1 to 4 and associated time coefficients of the C-OptDMD methods (left panels) and the POD method (right panels) extracted from the 3.5 years of fitting data from the TAC204 dataset. Modes of the C-OptDMD method are shown as pairs due to their nature as conjugate-couples. In both cases, values of the time coefficients are scaled down by a factor of 1 million kg.
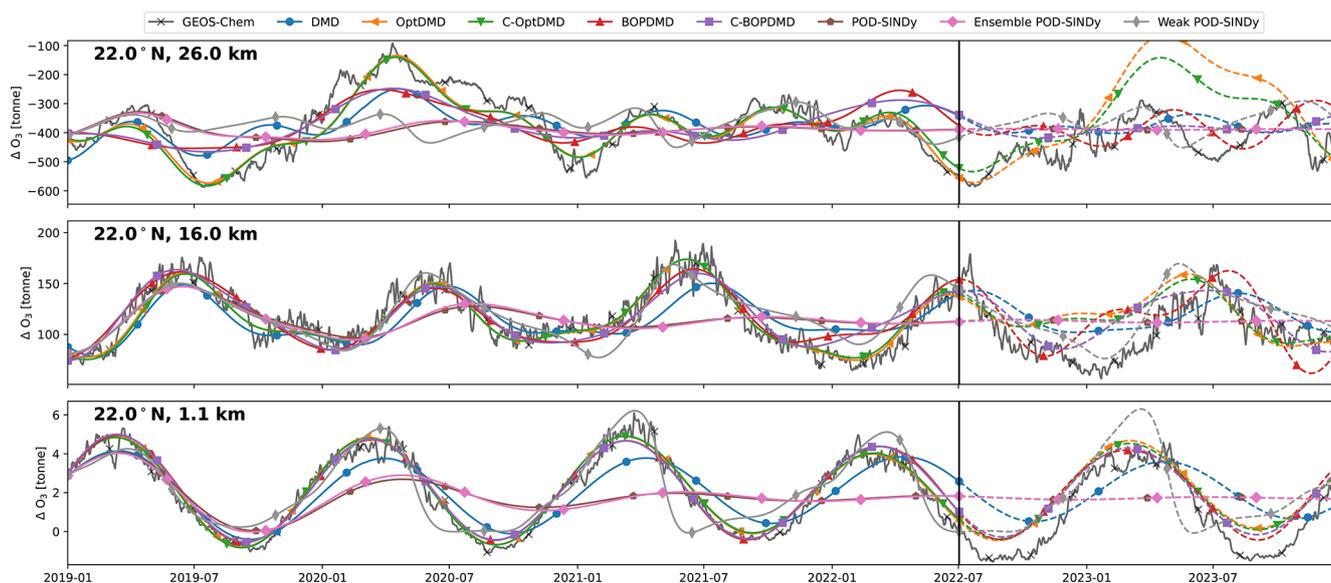

Figure 4. Ozone response in tonnes for 3 grid cells of the TAC204 dataset, with reconstructions and short-term forecasts from the DMD and POD-SINDy methods. Cells shown are located at 22° N latitude and altitudes of 26.0, 18.3, and 1.4 km from top to bottom respectively. Vertical black lines mark the difference between the reconstruction (solid lines) and forecast (dashed lines).

For the surface grid cell, the best results are achieved by the BOPDMD, C-BOPDMD, and C-OptDMD models. These methods, respectively, achieve forecast RMSEs of 0.98, 1.00, and 1.04 t, while that of the other methods ranges from 1.57 t (weak POD-SINDy) to 2.16 t (ensemble POD-SINDy, Table B1). Despite the instability of the POD-SINDy methods, the model discovery methods provide an improvement over forecasting for 1.5 years with the data-mean in all cases (RMSE of 2.32 t). As the altitude increases, the intricacy of the ozone response increases due to the higher complexity of local ozone chemistry and meteorological conditions, as well as the closer proximity to the perturbed emissions. Figure 4 shows that the seasonal behaviour seen near the surface is less influential at 16 km altitude, and almost non-existent at 26.0 km altitude. Lacking clear seasonal patterns, it also becomes more difficult for all methods to reproduce and forecast the ozone response in these domains. BOPDMD, C-BOPDMD, and C-OptDMD again produce the best reproductions of the original data, although they do not capture all features at these altitudes. These meth-

ods replicate the TAC204 data with RMSE of 13.88, 14.41, and 10.17 t for the 16 km grid cell, and 82.23, 83.29, and 43.70 t for the 26 km grid cell (BOPDMD, C-BOPDMD, and C-OptDMD, respectively). When forecasting, these respective errors increase to 33.34, 33.36 and 27.09 t at 16 km, with 97.61, 90.30, and 123.27 t at 26 km.

These results highlight that accommodating all of these regimes within a single dynamical model is a significant challenge. For the near-surface grid cell, the ozone response is most stable and exhibits clear seasonal patterns, assisting its predictability. In this cell, forecasts from the data-driven methods are an improvement over propagating the mean in 78.6 % of the tested cases, with the DMD-driven methods representing improvement in almost all cases. For the 16 and 26 km cells these fractions change to 69.6 % and 60.7 % respectively (Table B1) as forecasting future states becomes more difficult. At these altitudes, the DMD-driven methods still offer the most consistent improvements over the propagation of the temporal mean, but ultimately all methods have difficulties with the ozone response in the stratospheric grid cell. For all other datasets we find similar patterns in the results (Table B1, Figs. C3–C8). It may be that the data is just too complex at this altitude for these methods or that the size of the dataset may be insufficient to identify the appropriate large-scale features. Given the importance of this domain for this case study, this will negatively affect the ability to forecast in aggregated metrics such as the perturbation of ozone columns.

4.3 Zonal average

The differences in the ability to forecast and reconstruct different atmospheric domains are also apparent in the estimates of zonal average perturbations. The RMSE of the zonal average perturbation reconstructions and forecasts is presented in Table 1. We find that the DMD and POD-SINDy methods provide an improvement over the use of the temporal mean in most cases. When used for reconstruction the methods produce a lower RMSE than the propagation of the temporal mean in 82.1 % of the tested combinations, and when forecasting this is the case for 75.0 %. Of all methods tested, the C-OptDMD, C-BOPDMD, and BOPDMD variants produce the smallest overall RMSE in both reconstruction and forecasting of the zonal average. Figure 5 shows that C-OptDMD consistently reduces the reconstruction RMSE by up to −70 % compared to using the temporal mean, with reductions of −40 % for forecasting. For BOPDMD these reductions are up to −60 % and −30 % for reproducing and forecasting, respectively. The C-BOPDMD has similar performance as BOPDMD in most cases, except for the S3 dataset. In this case the constrained model prevents instability in the forecasting, but the reproduction is slightly worse. Compared to the DMD methods, the POD-SINDy methods are less successful. They only provide notable (< −10 %) reductions over the use of the temporal

mean in a few datasets, in some cases even yielding less accurate estimations. There are several factors that may affect the reduced performance of the POD-SINDy methods, which will be discussed in subsequent sections.

The variability and the reconstruction and forecast errors of the methods applied to the TAC204 dataset are shown in Fig. 6. The variability of the ozone response is localised around two areas of the stratosphere: the middle stratosphere and the area around the source of emissions. These areas coincide with the boundary between areas of net increases and depletion of stratospheric ozone, that are subject to seasonal variations. This variability pattern is generally captured by the DMD and POD-SINDy models, although their peak variability is lower because only low-ranking features are used in the reduced-order models. The effect of the lower variability is also visible in Fig. 4, as the reconstructed signals do not exhibit the day-to-day variations of the original signal.

Figure 6 also shows that the methods have similar reconstruction errors despite their algorithmic differences. This suggests that the errors may predominantly be driven by sources outside of the models, such as the lack of higher-order features. We also find that all methods exhibit similar error patterns when forecasting. In all forecasts, the methods underestimate the loss of the middle stratospheric ozone while overestimating the increases of the lower stratospheric ozone in the northern hemisphere. They also overestimate ozone depletion in the tropic upper-stratosphere. The consistency in these errors across the methods suggests that they may be caused by a common external factor, that cannot be forecast from the given data. This results in larger overall forecast errors, although in this application the data-driven methods are still better than propagating the temporal mean in 42 out of 56 cases (75.0 %, Table 1).

Comparing the datasets with the local and global distribution of emissions, we do not find notable differences in the forecasting or reconstruction performance of the methods, despite the higher complexity of the latter datasets. Relative to reconstructing or forecasting with the temporal mean, there is no directly differentiable difference in method performance for the global emission datasets (Fig. 5), although this may be because the accuracy of the temporal-mean estimate also reduces against more complex dynamics. Comparing the coefficient of determination, calculated by comparing all grid cells in the zonal average, we see that the coefficients are lower for the global emission datasets (Fig. C9). This shows that applying model discovery methods to these datasets is more challenging, but the difference is not big enough to affect the aggregated metrics.

4.4 Derivative metric: global ozone columns

In this section, we evaluate the ability to calculate derivative metrics, in this case the mean perturbation of the global ozone column in Dobson units (DU), from the zonal-average perturbations provided by the reduced-order models. This
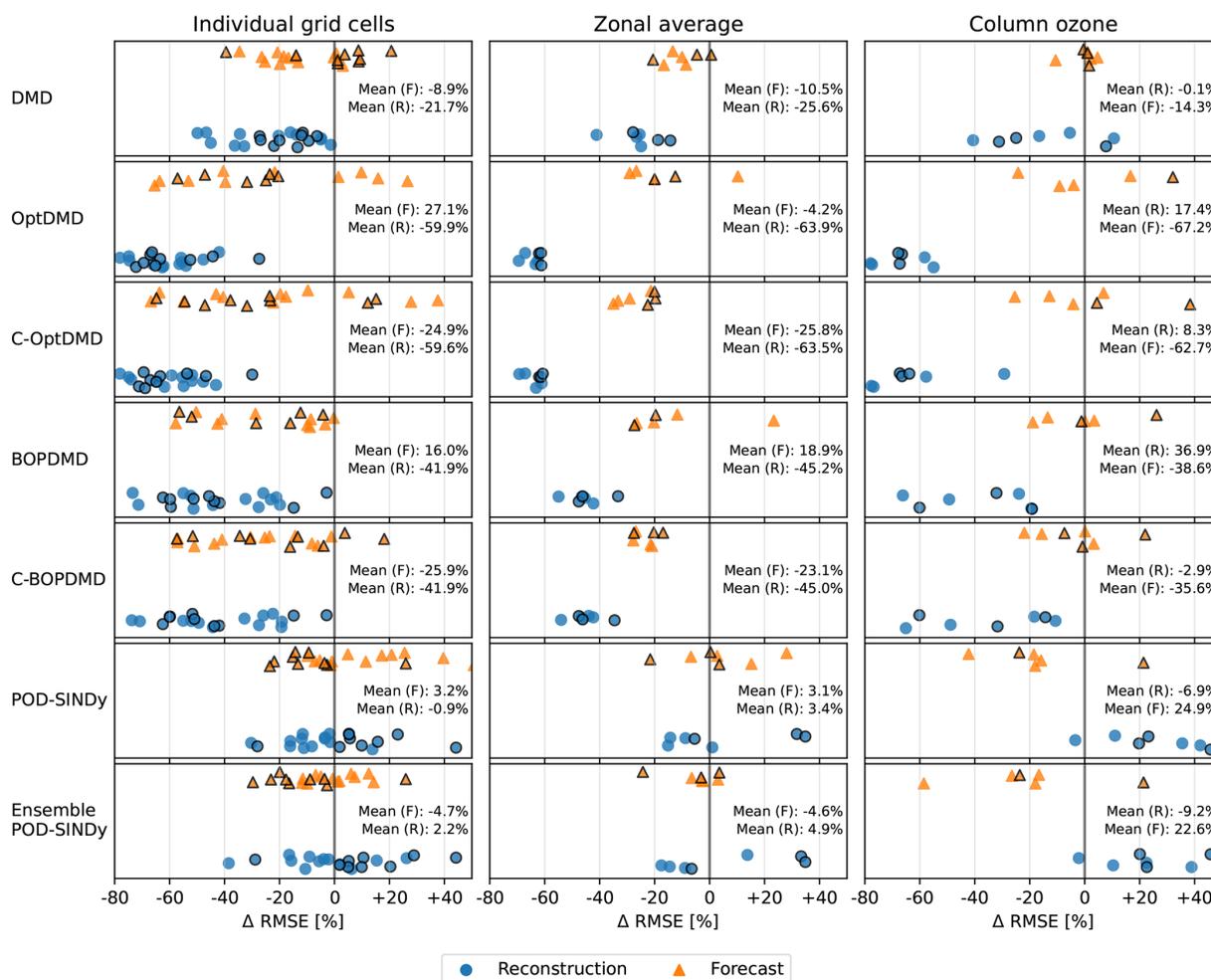
Figure 5. Change in the reconstruction (blue) and forecast (orange) RMSE with respect to reconstruction and forecasts using the temporal mean over all datasets for the methods used in this work. The left column shows the reconstruction of individual grid cells at altitudes of 1, 16, and 26 km (Table B1), the middle column the zonal average (Table 1), and the rightmost column the estimation of column ozone (Table 2). Markers with a black border denote the S1, S2, or S3 datasets.

metric is an aggregate of the global ozone distribution, which is more sensitive to latitudinal misplacement of the ozone response because it depends on the surface area of the grid cells. The global mean ozone column response over time is shown in Fig. 7 for the datasets, together with reconstructions and forecasts from the data-driven methods and their variants. The RMSEs of these reconstructions and forecasts are shown in Table 2.

We find that the ability of the data-driven methods to reconstruct and forecast the perturbation of the mean global ozone column varies considerably between the scenarios. Compared to propagating the temporal mean, the DMD and POD-SINDy methods provide better reconstructions in 64.3 % of the applications. This is a reduction of 17.9 % compared to the improvement of propagating the zonal average perturbation (Sect. 4.3). Figure 5 also shows that the RMSE reduction from the application of the methods also worsens upon calculation of the ozone columns. We expect that the re-

duction in accuracy is related to the higher sensitivity to latitudinal misplacement in the ozone column calculation. The POD-SINDy method and variants are most affected, producing higher RMSEs than the temporal mean in most cases. Only the C-OptDMD and both BOPDMD methods still provide consistent improvements over the temporal mean for reconstruction, with C-OptDMD resulting in the largest reductions in error. Compared to the original data, these methods have relative errors smaller than 10 % in their reconstructions of the ozone column effect across all datasets.

Being an extension of the zonal averages, the ozone column estimates are also less accurate when forecasting future responses. These forecasts are better than propagating the mean in 55.3 % of the applications. We also find that methods that reconstruct the data well do not necessarily provide the best forecasts (Fig. 5). This suggests that, in some cases, such as the weak POD-SINDy method applied to the SAS204 data, the low forecast error may be attributable to

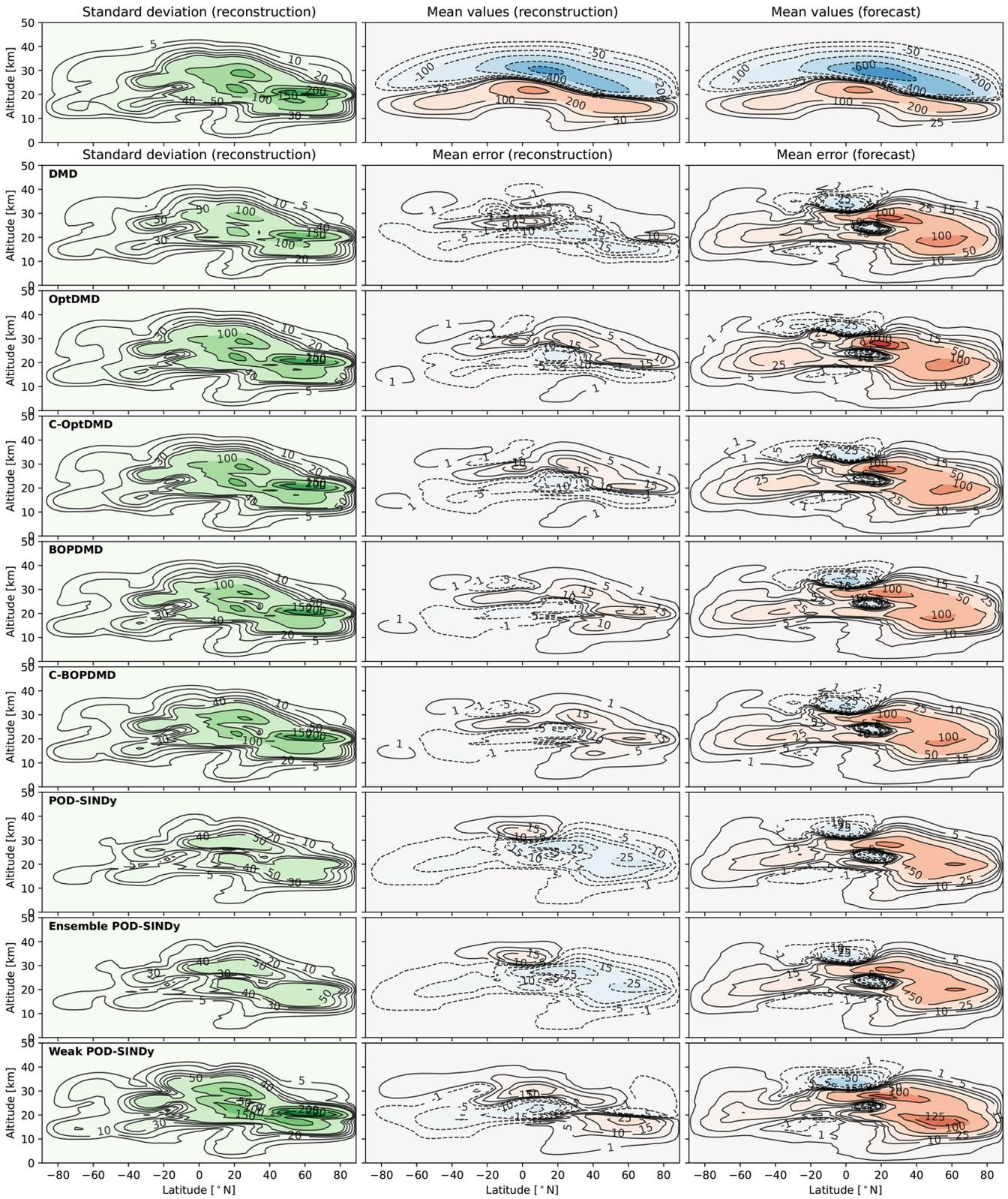**Figure 6.** Standard deviation, reconstruction errors, and forecast errors of the zonal average ozone response. The top row shows the reference standard deviation, and mean values of the TAC204 dataset for the reconstruction and forecast period. The subsequent rows show the standard deviation and mean error for the reconstruction and forecast for the different methods. The standard deviation and errors are shown in tonnes of ozone.

Table 1. RMSE (root mean squared error) of the reconstruction (R) and forecast (F) of longitudinally-averaged ozone response in tonnes for the DMD and POD-SINDy methods over the datasets. The data mean column shows the zonal average reconstruction and forecast RMSE if the perturbation is estimated using the temporal mean from the fitting data. When method's RMSE is lower than the data mean approach, its values are printed in bold.

		DMD	OptDMD	C-OptDMD	BOPDMD	C-BOPDMD	POD-SINDy	POD-SINDy (ensemble)	POD-SINDy (weak form)	Data mean
SAS204	R	**42.48**	**21.68**	**22.50**	**33.42**	**33.39**	58.45	65.84	103.63	57.88
	F	**64.41**	85.22	**51.54**	**61.63**	**61.08**	98.94	**75.17**	87.19	77.26
TAC204	R	**27.93**	**14.51**	**14.59**	**21.37**	**21.81**	**40.65**	**40.55**	**35.37**	47.43
	F	**51.12**	**43.28**	**38.35**	**52.10**	**46.35**	**55.02**	**55.19**	**48.55**	59.02
SAS162	R	**13.05**	**5.76**	**5.76**	**9.36**	**9.36**	**15.97**	**15.94**	17.70	17.51
	F	**16.49**	**12.79**	**12.79**	**13.25**	**13.20**	18.53	18.58	**15.74**	18.03
TAC162	R	**15.26**	**7.45**	**7.49**	**11.09**	**11.39**	**17.25**	**16.76**	**17.58**	20.33
	F	**19.81**	**17.61**	**17.31**	27.16	**15.90**	25.36	**21.47**	**20.60**	22.01
S1	R	**7.07**	**3.72**	**3.72**	**5.13**	**5.13**	**9.26**	**9.15**	**9.65**	9.79
	F	**9.10**	**10.02**	**9.20**	**8.32**	**8.30**	**8.98**	**8.67**	**7.65**	11.45
S2	R	**23.27**	**11.07**	**11.07**	**15.41**	**15.40**	37.72	38.17	29.15	28.63
	F	32.20	**25.61**	**25.61**	**25.74**	**25.51**	32.12	**31.04**	39.39	32.01
S3	R	**3.04**	**1.37**	**1.39**	**2.36**	**2.32**	4.78	4.78	3.74	3.54
	F	**4.71**	8.31	**3.83**	15.54	**4.10**	5.12	5.12	**4.64**	4.94

Table 2. RMSE (root mean squared error) of the reconstruction (R) and forecast (F) of mean changes in global column ozone for the DMD and POD-SINDy methods over the datasets in Dobson Units. The methods are fitted and tested for reconstruction on 3.5 years of data, whereas the forecast is tested on the subsequent 1.5 years of data. The data mean column shows the reconstruction and forecast errors if these were estimated using mean values from the fitting data. When method's RMSE is lower than the data mean approach, its values are printed in bold.

		DMD	OptDMD	C-OptDMD	BOPDMD	C-BOPDMD	POD-SINDy	POD-SINDy (ensemble)	POD-SINDy (weak form)	Data mean
SAS204	R	**0.08**	**0.04**	**0.06**	**0.07**	**0.07**	0.12	0.12	0.28	0.09
	F	0.28	0.32	**0.24**	**0.28**	0.28	**0.23**	**0.20**	**0.09**	0.28
TAC204	R	**0.10**	**0.04**	**0.04**	**0.06**	**0.06**	**0.17**	**0.17**	**0.09**	0.17
	F	**0.27**	**0.23**	**0.22**	**0.24**	**0.23**	**0.24**	**0.25**	**0.25**	0.30
SAS162	R	**0.04**	**0.01**	**0.01**	**0.02**	**0.02**	0.05	0.05	0.06	0.05
	F	0.07	0.07	**0.07**	**0.07**	0.07	**0.06**	**0.06**	**0.05**	0.07
TAC162	R	0.05	**0.02**	**0.02**	**0.03**	**0.04**	0.06	0.05	0.05	0.04
	F	0.09	**0.08**	0.09	**0.07**	**0.07**	**0.05**	**0.04**	**0.05**	0.09
S1	R	**0.05**	**0.02**	**0.02**	**0.05**	**0.05**	0.09	0.09	0.08	0.07
	F	0.09	0.14	0.12	**0.08**	**0.08**	**0.07**	**0.07**	**0.05**	0.09
S2	R	**0.15**	**0.07**	**0.07**	**0.08**	**0.08**	0.25	0.24	0.20	0.20
	F	**0.15**	0.23	0.23	0.19	0.18	0.22	0.22	0.20	0.15
S3	R	0.03	**0.01**	**0.01**	**0.02**	**0.03**	0.04	0.04	0.03	0.03
	F	0.05	0.06	0.05	0.15	**0.04**	0.06	0.06	**0.05**	0.05

coincidence more than the method. Still, Fig. 7 shows that for some datasets the DMD variants predict seasonal trends in the ozone response, although the values may drift from the GEOS-Chem values. We expect that, like the drifts in zonal-average predictions, these shifts are also related to shifts in background conditions that are not foreseen by the methods.

5 Discussion

In this work, we present the evaluation of seven variations of model discovery methods applied to seven different datasets. The results that we present in the previous section do not however represent the best results that we obtained. We have
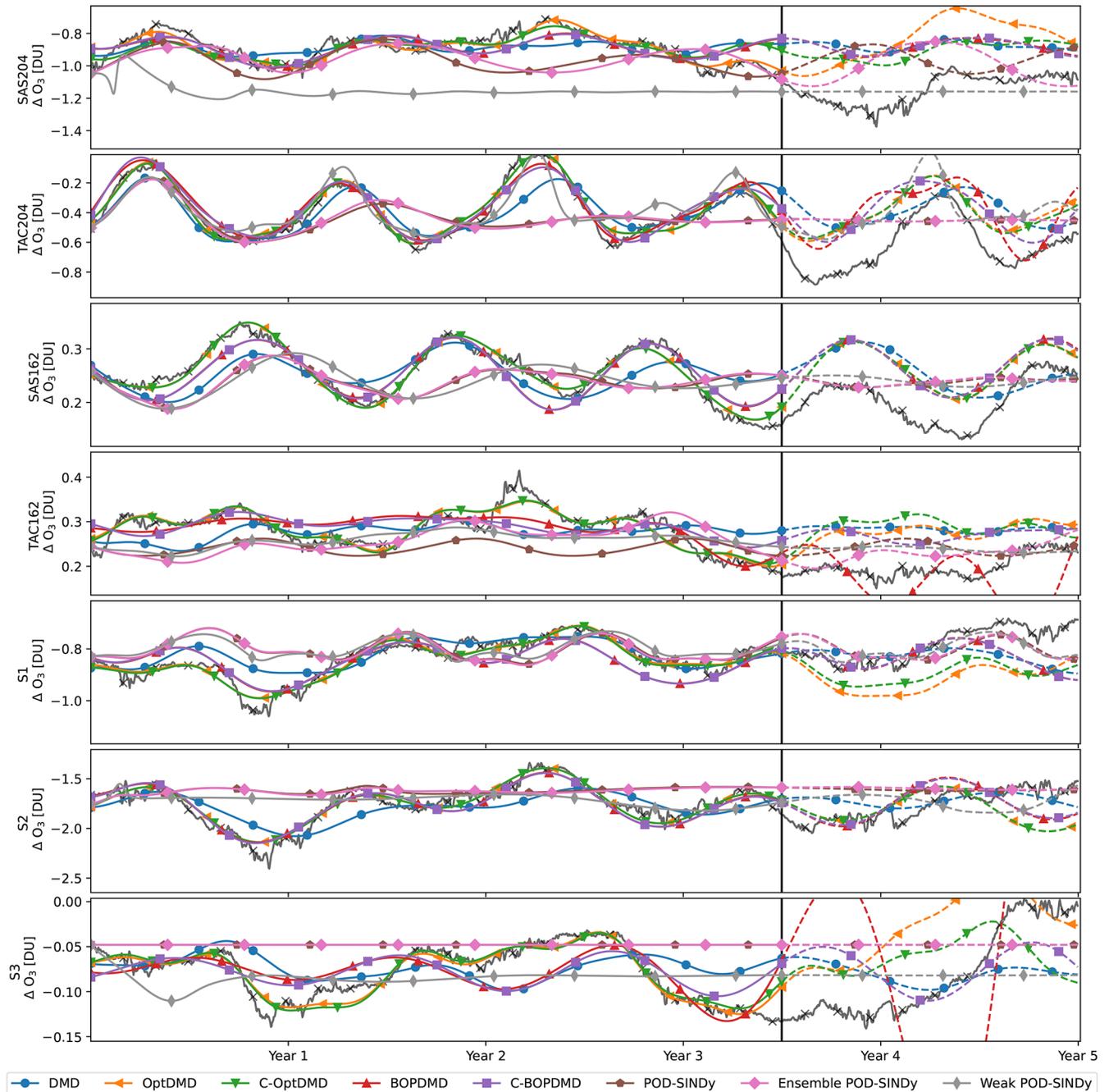
Figure 7. Reconstruction and forecast of the change in global column ozone in DU over time for all datasets and methods. The vertical black line denotes the change from the reconstruction to forecasting domains, solid lines are reconstructions and dashed lines mark forecasts of future behaviour.

performed extensive hyperparameter searches in which more than 29 000 configurations of the methods were fitted to different parts of the data (Appendix A). Through this approach, we found configurations with considerably smaller reconstruction and forecast errors for the ozone response than those we show in Sect. 4 (e.g., see Fig. A3). We do not present these as our main results, as cherry-picking the

best results does not necessarily paint an accurate picture of the performance that could be expected from these methods. Most prospective users of these methods may not have access to the future system behaviour to test the quality of forecasting, and they may also not have the resources to test tens of thousands of configurations. The results that we share in this work are representative of the median performance

of the hyperparameter search, because we consider that this level of performance is closest to what can be expected when these methods are applied to a dataset without extensive exploration. For example, from 71 different configurations of the C-OptDMD method with 3.5 years of fitting data, we find a median forecast error of 16.7 % for the S2 scenario, with a standard deviation 13.0 %, and a range 6 % to 75.4 %. The C-OptDMD configuration presented in Sect. 4.4 has a relative error of 9.4 %. With some exploration we consider it likely that configurations can be found with better performance than the median we present, but we note that model selection for forecasting future behaviour remains challenging. For example, we find that selecting models for the lowest reconstruction error tends to increase forecasting errors as this strategy promotes high-ranking solutions which overfit to the training data. This suggests that it may be better to reserve part of the data to evaluate short-term forecasting to guide model selection, but this needs to be explored further in future work.

Overall, we find that the OptDMD or BOPDMD methods have the most robust performance in terms of forecasting and reconstruction of the data, especially their constrained variants (C-OptDMD, C-BOPDMD). The constraints on the eigenvalues prevent numerical instability of the models, resulting in better forecasts at a small cost to reconstruction accuracy (Fig. 5). The benefit of these constraints can clearly be seen in Fig. 7, where the BOPDMD ensemble is unstable for the TAC162 and S3 datasets (Fig. C2) resulting in large errors in the forecast that are mitigated in the constrained model. This pattern is reflected in our hyperparameter searches, where these methods generally perform best, although there are instances where DMD or unconstrained OptDMD perform better. The hyperparameter search has shown that there is no clear optimum for the rank of the DMD methods, with the exception of BOPDMD which becomes unstable at higher ranks (see Appendix A). Therefore, we do not expect that our results change significantly when the rank of the DMD methods is varied. For the application of DMD methods to other atmospheric chemistry datasets, we recommend the constrained OptDMD or BOPDMD methods first, because they most consistently yield low reconstruction errors while maintaining dynamical stability. The other methods we assess are able to yield similar results, but they are less consistent in numerical stability. In this work we have however not assessed the effect of time delays in the DMD variants, which could affect the model performance and relative stability.

Compared to the DMD methods, we obtained worse performing reconstructions and forecasts of the reference data from the POD-SINDy based methods across all datasets. This has continually also been the case for the extensive searches that were run, and it is predominantly related to the difficulty of finding stable SINDy models for the time period that we assess. We expect this could be improved by the application of different SINDy extensions that are designed to promote model stability, such as the "Trapping SINDy" method by Kaptanoglu et al. (2021a), as our results show these stability-promoting approaches are very effective for DMD methods. In addition to the use of these extensions, it may also be possible that the SINDy methods may improve from further exploration of library model terms, as also discussed by Yang et al. (2024). Similar to Yang et al. (2024) and recommendations from Fasel et al. (2022), we use a library with linear and quadratic terms, but this may not be best suited to this application. The use of different function libraries may improve results of this method, but this requires a more thorough exploration of terms that is outside the scope of this work. It could also be considered to integrate other data in the SINDy method than the POD time coefficients, such as the temperature or the distribution of other chemical species, as these might also directly affect the temporal coefficients.

Regardless of the method, we find that it is difficult to forecast future behaviour of the ozone response. We expect that this is partially related to the the reduced-order models not being bound by physical constraints such as stationary processes or the conservation of mass, in addition to the difficulty of forecasting background conditions. Ozone is influenced by meteorological conditions, and shifts in these conditions may have considerable effects on its distribution. In our approach we attempt to predict ozone's behaviour from existing dynamics without integrating any meteorological information, and we show that this is a considerable challenge. We expect that better forecasts could be obtained if meteorological variables, such as global irradiance or temperature fields, are incorporated into the data provided to the methods. Alternatively, stochastic approaches, such as those proposed by Rao et al. (2024) and Boninsegna et al. (2018), may also allow for better consideration of uncertainties due to changes in future conditions.

While our results show that all methods have some degree of error relative to the GEOS-Chem data, these methods all require but a fraction of the computational cost and storage space footprint. The datasets that we use all require several gigabytes of storage each, and were created using hundreds of thousands of CPU-hours. Using the C-OptDMD method as an example, it can reproduce an estimate of the GEOS-Chem data in a matter of seconds while only needing 20 MB of storage space for the fitted model. This represents a reduction of several orders of magnitude in computational time and storage requirements. Furthermore, we have also shown that these reproductions can be accurate enough to calculate derivative metrics such as the change in global column ozone. It is likely that better ozone column models can be found by applying the methods to ozone column data directly, but the ability to do this conversion highlights the versatility of these methods. Having shown that some of these methods can create stable models of coupled tropo- and stratospheric systems, we consider that future work may look into combining multiple dynamical models to predict dynam-

ical responses not comprehensively simulated, such as the consideration of new aircraft cruise altitudes. Should this be possible these methods could be used as very efficient but detailed surrogate models to support decision-making or optimisation processes.

6 Conclusions and recommendations

Chemistry transport models (CTM) are critical in the evaluation of the environmental effects of emissions on our atmosphere, but their high computational cost and difficulty to use makes them impractical for implementation in multidisciplinary optimisation or operational and regulatory decision-making processes. Model discovery methods such as dynamic mode decomposition (DMD) and proper orthogonal decomposition (POD) paired with the sparse identification of non-linear dynamics (SINDy) can help to develop easy-to-use reduced-order models from chemistry transport model data. In this work, we evaluate the application of these methods on seven datasets from case studies that assess changes in the global ozone distribution in response to supersonic aviation emissions over up to 10 years. Compared to earlier works, this represents a considerable increase in problem scope and complexity.

We show that these model discovery methods can be used to reconstruct and forecast datasets spanning entire atmospheres, with the constrained variants of optimised DMD (C-OptDMD) and bagging optimised DMD (C-BOPDMD) methods being most effective. These methods can reconstruct and forecast the ozone response with smaller errors than the temporal data mean for several years without losing stability, resulting in substantial reductions in reconstruction ($> 50\%$) errors compared to estimations using the temporal mean. The reconstructed atmospheres of these methods can be accurate enough to calculate changes in the ozone column with relative errors smaller than 10% compared to the real data. Approaches based on DMD, unconstrained OptDMD, and POD-SINDy can also return reduced-order models with similar performance in some cases, but these methods are less reliable because they do not promote stable solutions, affecting their dynamical stability over the multi-year timeframes. Ultimately, we find that stable reduced-order models with relative errors smaller than 5% can be found with all methods, although the difficulty of finding stable configurations differs greatly between methods. Such models can reduce the computational and data storage requirements by several orders of magnitude and without requiring the same level of technical expertise as the CTM models. Forecasting future behaviour of ozone responses from only historic data remains challenging however, and this may require embedding more data such as meteorological variables. The ability to accurately represent temporal atmospheric responses with easy-to-distribute and resource-efficient models such as those from C-OptDMD may be valuable to facilitate the ex-

change of data and embedding complex temporal behaviour in applications such as those supporting decision-making processes. This may already improve the level of detail used to support decision-making or multidisciplinary optimisation processes, but to unlock the full value of these methods further research is needed to improve the forecasting of complex responses or to combine several fitted models to predict responses to new situations.

Appendix A: Hyperparameter search

The usability of methods like POD and DMD can vary considerably depending on the hyperparameters that are used (e.g. rank, constraints). While there are some precedents from literature to similar data (Velegar et al., 2024; Yang et al., 2024), this is not enough to conclude on best-practice approaches. Which is why as part of this work we ran extensive hyperparameter sweeps where thousands of variations of the methods were applied to different parts of the dataset. In this appendix, we shortly summarize the methodology and main results.

We explored the effect of model ranks, different DMD variants, and the selection of fitting and forecast datasets. Specifically, we performed a grid search over the length of input data (1 to 7.5 years, steps of 0.5), the length of the forecast time (1 to 3.5 years, steps of 0.5), and the rank of the models (1 to 26, steps of 1) for each of the five DMD variants, resulting in 4200 points per dataset. We find that there is no clear best practice with regard to the rank and variant of the DMD method, but some trends can be identified (Fig. A1). For the conventional DMD method and unconstrained Opt-DMD, the median performance and the variance between the methods is not affected by the rank. For the constrained Opt-DMD and BOPDMD the median forecast accuracy and its spread increase with higher ranks. The latter in particular exhibiting increasing errors above the rank of 8.

Concerning the size of the dataset, we find that DMD and OptDMD generally work better with small datasets ($<$ 500 datapoints) whereas BOPDMD methods require at least several years of data (> 1000 datapoints) to achieve similar performance (Fig. A2). We expect that this larger need for data is related with the bagging process in the BOPDMD method. We use a trial size of 20% in this work, meaning that the models in the ensemble have access to only 20% of the available data. With short datasets this may not be sufficient data to properly capture annual trends, which is not a problem with the non-bagging methods. Across all methods the median forecast error and its spread decrease when the methods have access to more data.

Regardless of the rank of the method and the length of the fitting data, DMD models can be found that have forecast errors smaller than 10%, but the odds of finding such models improve under the conditions mentioned above. Regardless of the settings, no methods have better relative forecast

errors than 5 %, which may represent a limit for what can be expected of DMD methods for data as variable as daily changes in ozone concentrations.

In the end, we do not find clear suggestions for the best-practice application of DMD methods to ozone perturbation data. Our results suggest that the most suitable DMD method and its configuration depend primarily on the data on which it is applied, and therefore we recommend that prospective users of these methods explore different configurations as well for their own data.
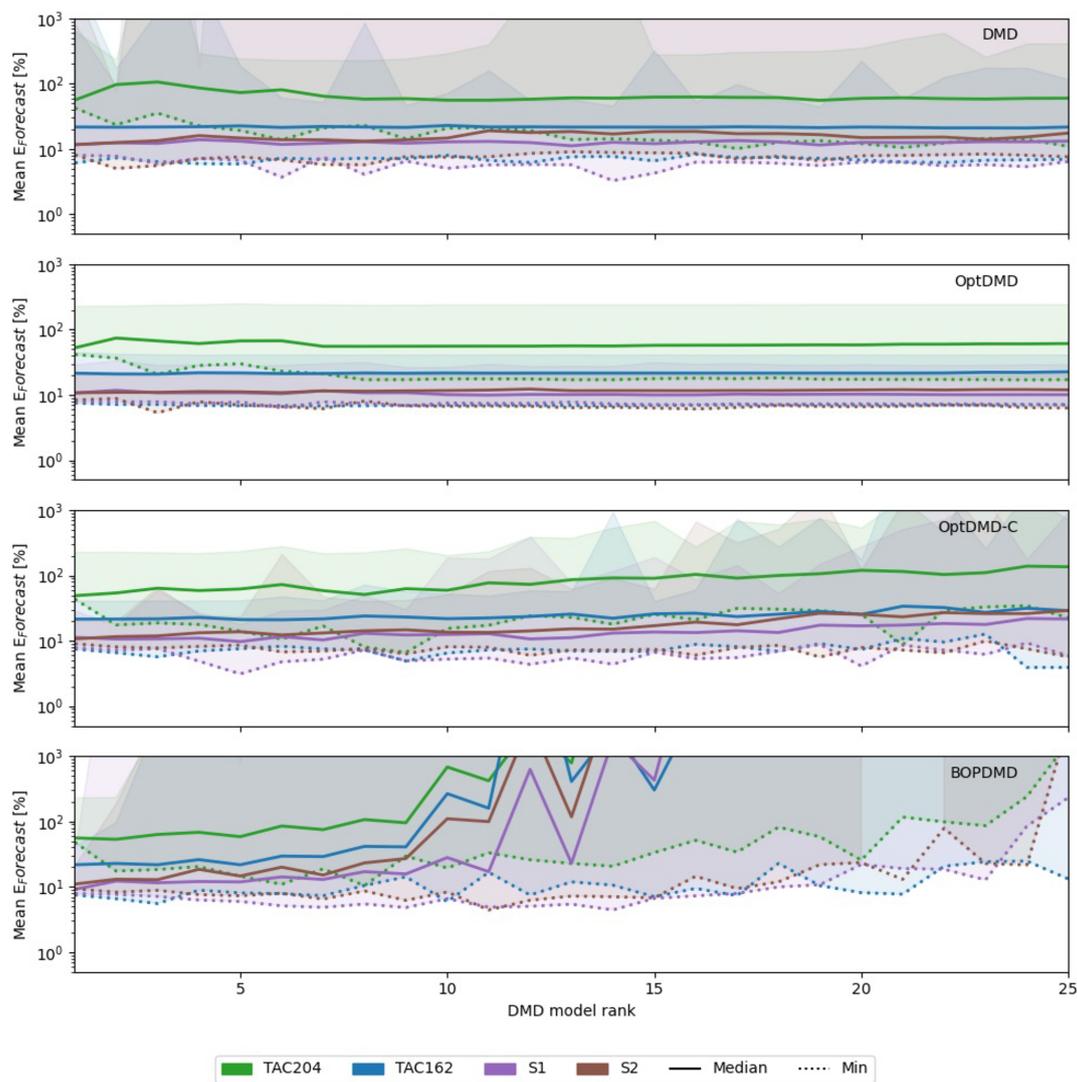
Figure A1. Results of hyperparameter sweep of DMD methods over the DMD rank, for the TAC204, TAC162, S1, and S2 emission scenario datasets. The solid lines indicate the median O_3 column change forecast error in percent. The shaded areas mark the full range of the error distribution, and the dotted lines mark the lowest forecast error achieved by any of the tried configurations.
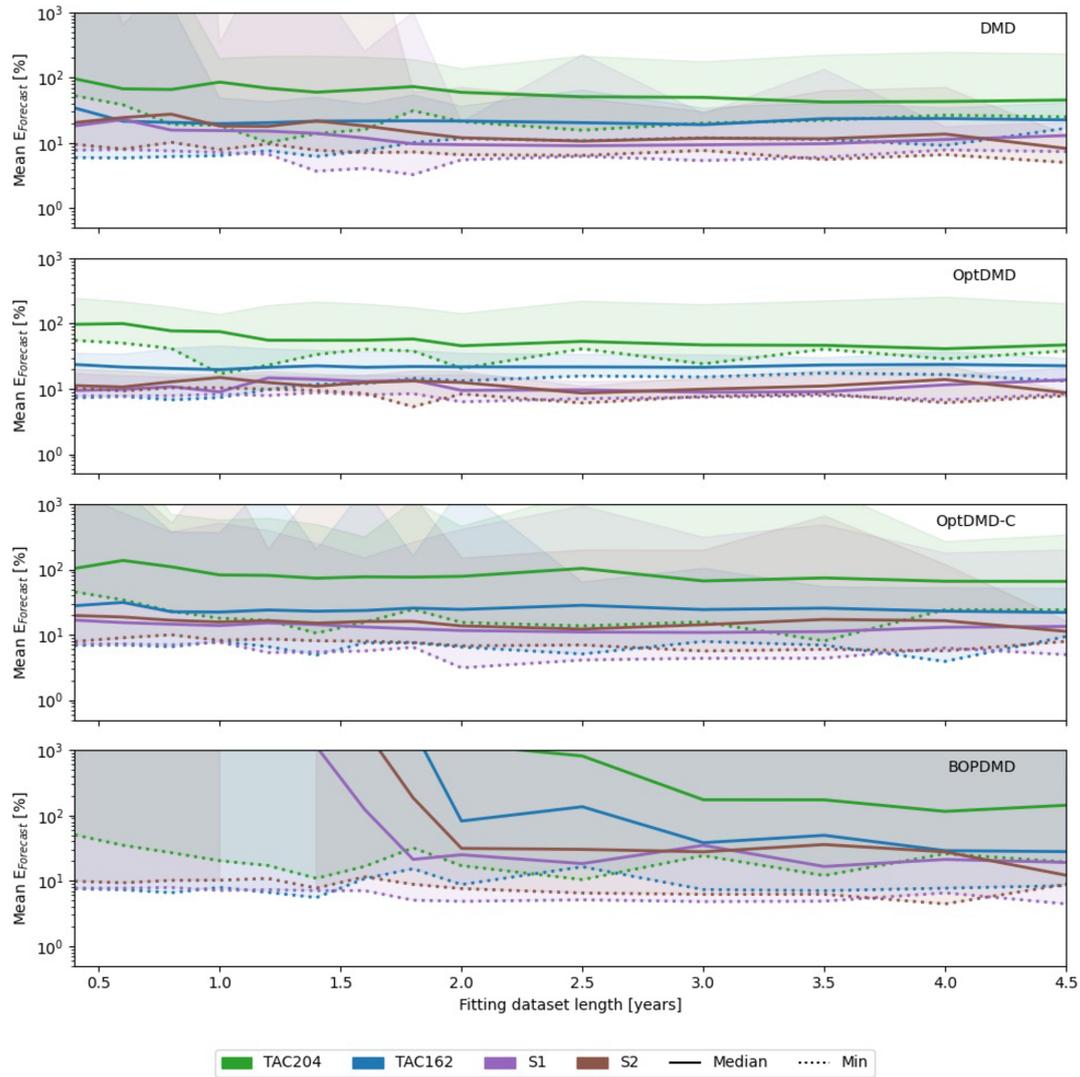
Figure A2. Results of hyperparameter sweep of DMD methods over the fitting dataset length, for the TAC204, TAC162, S1, and S2 emission scenario datasets. The solid lines indicate the median O_3 column change forecast error in percent. The shaded areas mark the full range of results, and the dotted lines mark the lowest forecast errors achieved by any of the tried configurations.
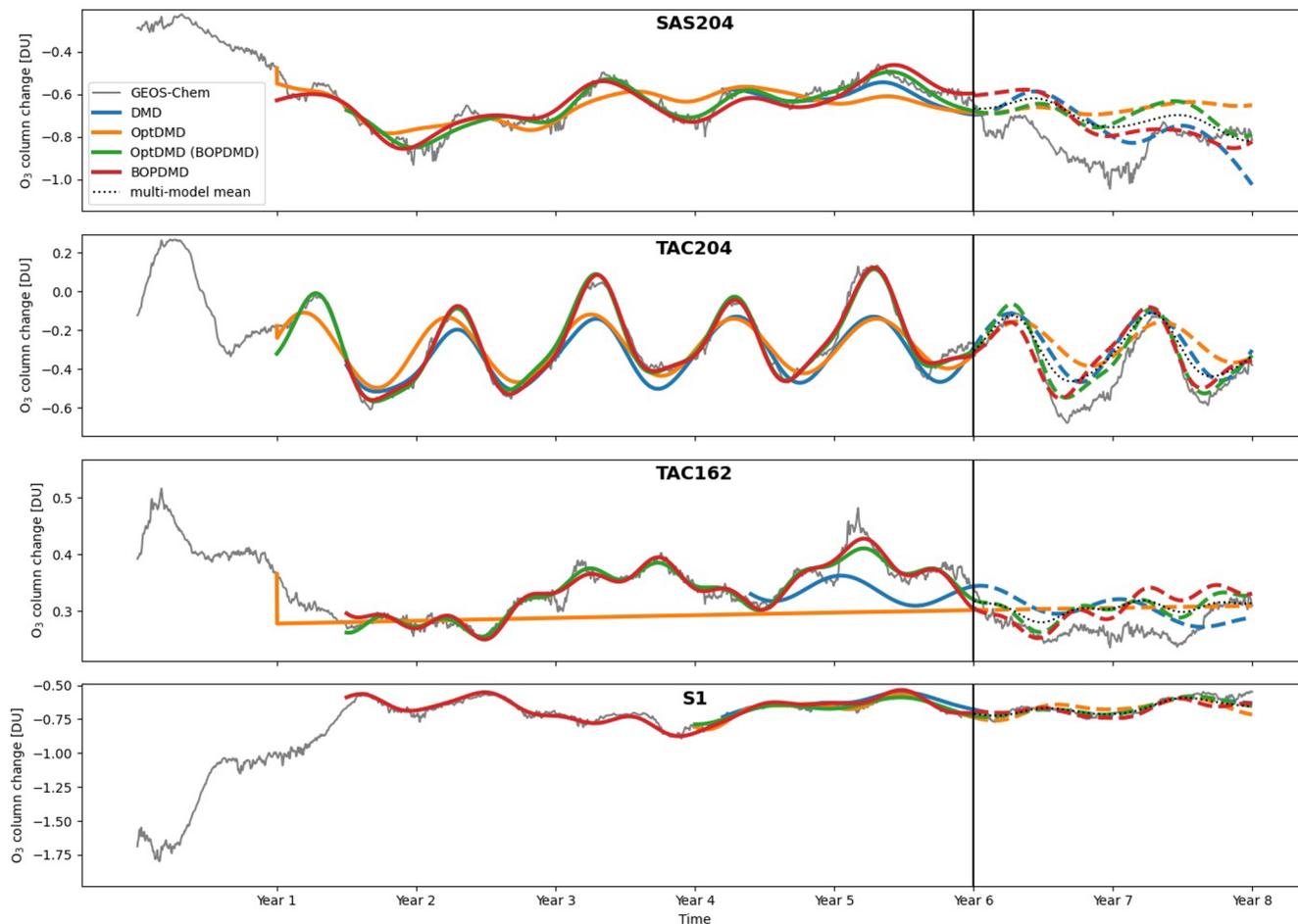
Figure A3. Selection of DMD models identified by the hyperparameter sweep, selected for the best 2-year forecast performance for the SAS204, TAC204, TAC162, and S1 scenarios. The rank and the length of fitting data is allowed to vary between the configurations, therefore the selected best results have different ranks and ranges of fitting data. In some cases however, such as for OptDMD with the TAC162 scenario, suitable configurations could still not be found.

Appendix B: Additional tables

Table B1. RMSE (root mean squared error) of the reconstruction (R) and forecast (F) of mean changes near-surface ozone mass in tonnes for the DMD and POD-SINDy methods over the datasets. The Data mean column shows the reconstruction and forecast errors if these were estimated using mean values from the fitting data. Values printed in bold are lower than the data-mean estimate.

|		DMD	OptDMD	C-OptDMD	BOPDMD	C-BOPDMD	POD-SINDy	POD-SINDy (weak form)	POD-SINDy (ensemble)	Data mean
1 km altitude										
SAS204	R	**1.13**	**0.94**	**0.96**	**1.01**	**1.02**	**1.88**	2.26	3.59	2.13
	F	**1.75**	**1.62**	**1.53**	**1.55**	**1.50**	3.74	**2.37**	2.68	2.68
TAC204	R	**0.95**	**0.42**	**0.42**	**0.51**	**0.50**	**1.59**	**1.59**	**0.86**	1.90
	F	**1.73**	**1.09**	**1.04**	**0.98**	**1.00**	**2.15**	**2.16**	**1.57**	2.32
SAS162	R	**1.12**	**0.91**	**0.91**	**0.97**	**0.98**	**1.47**	**1.47**	**1.73**	1.75
	F	**1.28**	**1.04**	**1.04**	**1.03**	**1.03**	**1.73**	1.77	**1.54**	1.75
TAC162	R	**1.05**	**0.48**	**0.50**	**0.54**	**0.56**	**1.33**	**1.17**	**1.44**	1.91
	F	**1.59**	**0.69**	**0.67**	**1.42**	**0.85**	2.52	**2.00**	**1.95**	2.01
S1	R	**0.16**	**0.13**	**0.12**	**0.17**	**0.17**	0.18	0.18	0.19	0.17
	F	0.28	**0.19**	**0.20**	**0.25**	**0.25**	**0.22**	**0.22**	**0.21**	0.26
S2	R	**0.68**	**0.25**	**0.25**	**0.28**	**0.28**	0.87	0.90	0.75	0.75
	F	0.89	**0.50**	**0.50**	**0.53**	**0.51**	**0.64**	**0.61**	0.88	0.74
S3	R	**0.15**	**0.09**	**0.09**	**0.13**	0.23	0.27	0.27	**0.18**	0.19
	F	0.24	0.78	0.26	0.58	0.39	**0.23**	**0.23**	**0.23**	0.23
16 km altitude										
SAS204	R	**20.61**	**11.28**	**11.84**	**17.76**	**17.81**	27.95	30.99	27.40	24.55
	F	**21.08**	32.74	**20.76**	**23.70**	**22.03**	38.99	29.10	**23.85**	25.87
TAC204	R	**18.70**	**9.66**	**10.17**	**13.87**	**14.41**	**25.12**	**25.48**	**18.28**	28.50
	F	**24.96**	30.42	**27.09**	**7.24**	**22.88**	**28.40**	**28.52**	**23.29**	29.99
SAS162	R	**49.90**	**18.71**	**18.69**	**33.34**	**33.36**	**68.19**	**67.55**	75.32	74.29
	F	**57.41**	**26.04**	**26.02**	**35.45**	**35.62**	75.10	76.88	**63.52**	71.60
TAC162	R	**21.23**	**12.50**	**12.24**	**16.98**	**17.41**	**20.77**	24.82	21.72	21.51
	F	**21.69**	23.87	22.88	36.29	**16.25**	24.22	**19.18**	**20.85**	21.74
S1	R	**1.52**	**1.16**	**1.11**	**1.78**	**1.78**	2.20	2.19	2.18	2.09
	F	2.69	**1.11**	**1.18**	**2.17**	**2.17**	**2.02**	**1.99**	**1.54**	2.59
S2	R	**6.87**	**2.84**	**2.84**	**3.80**	**3.76**	9.56	10.01	7.98	7.77
	F	9.60	**6.71**	**6.71**	**7.73**	**7.61**	**7.53**	**7.03**	11.06	8.78
S3	R	**2.48**	**1.04**	**0.97**	**1.88**	**2.98**	3.16	3.16	3.24	3.10
	F	**3.39**	12.02	4.54	11.36	5.52	**3.80**	**3.80**	**3.85**	3.94
26 km altitude										
SAS204	R	**160.58**	**63.08**	**64.5**	**135.01**	**136.43**	**150.16**	172.44	312.50	168.98
	F	**122.34**	490.47	194.02	**135.8**	**139.37**	**139.00**	149.48	205.00	140.97
TAC204	R	**85.39**	**40.58**	**43.70**	**82.23**	**83.29**	**105.56**	**103.05**	**106.44**	107.21
	F	96.97	172.80	123.27	97.61	**90.30**	**86.99**	**86.85**	101.20	96.40
SAS162	R	**26.94**	**13.92**	**13.96**	**21.13**	**21.18**	**30.70**	**30.61**	32.75	31.26
	F	25.23	**19.16**	**19.03**	**22.04**	**22.47**	28.70	27.97	**22.04**	24.48
TAC162	R	**39.60**	**18.13**	**20.00**	**30.97**	**30.93**	**40.14**	**39.27**	**40.92**	41.59
	F	**28.69**	38.77	**27.54**	50.44	**23.28**	40.40	34.04	**33.12**	33.45
A1	R	**14.84**	**5.64**	**5.86**	**8.17**	**8.18**	**14.61**	**14.44**	**16.36**	20.30
	F	**12.01**	15.78	**12.33**	**8.64**	**8.62**	**15.19**	**13.95**	**14.53**	19.84
A2	R	**70.30**	**24.42**	**24.42**	**32.14**	**31.98**	84.08	88.28	**79.31**	79.85
	F	84.21	**43.96**	**43.96**	**39.84**	**40.16**	**75.41**	**75.81**	100.80	83.19
A3	R	**7.07**	**2.84**	**2.87**	**4.76**	**7.50**	8.98	8.98	9.10	8.17
	F	10.27	31.88	**3.33**	25.28	**6.39**	11.89	11.89	**8.50**	9.44

Appendix C: Additional figures
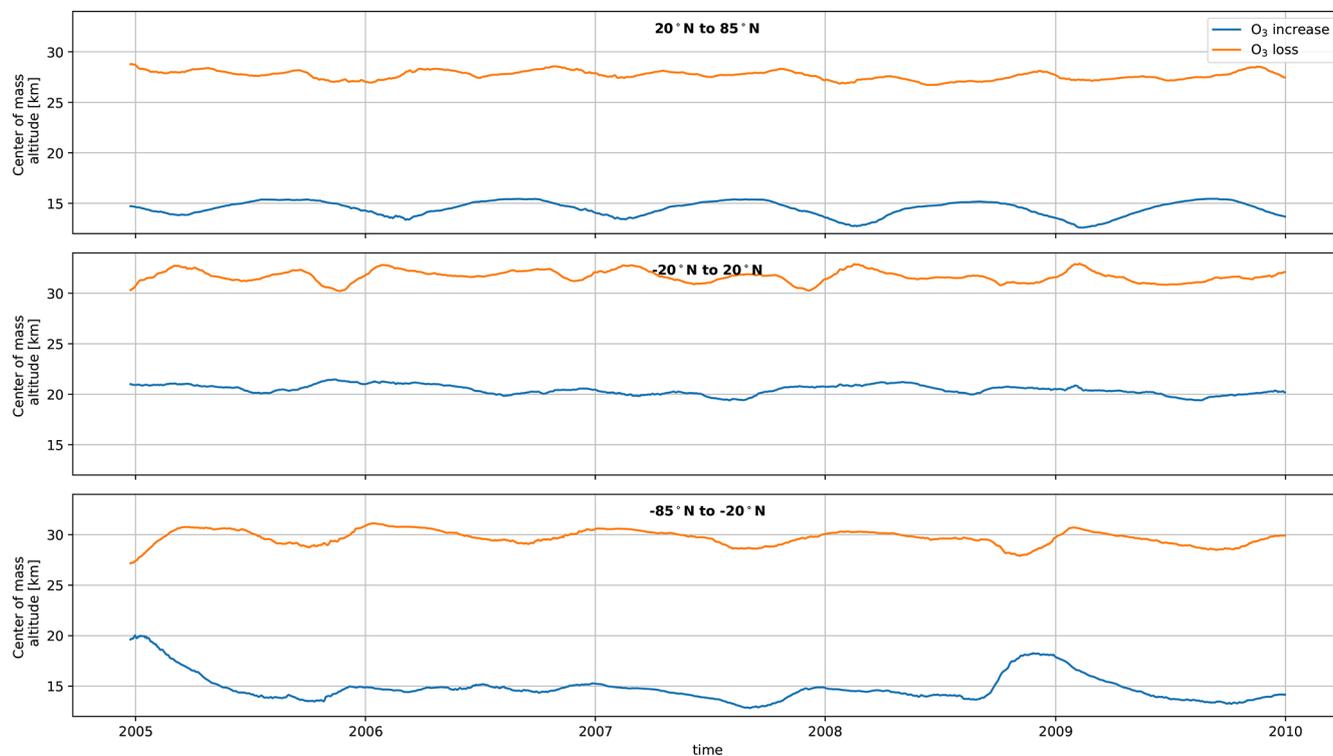
Figure C1. Altitude of the center of mass of the ozone response in the TAC204 dataset over time, calculated through a weighted average of the grid cells. Top row shows the northern hemisphere (20 to 85° N), the middle shows the tropics (−20 to 20° N), and the bottom row shows the southern hemisphere (−85 to −20° N).
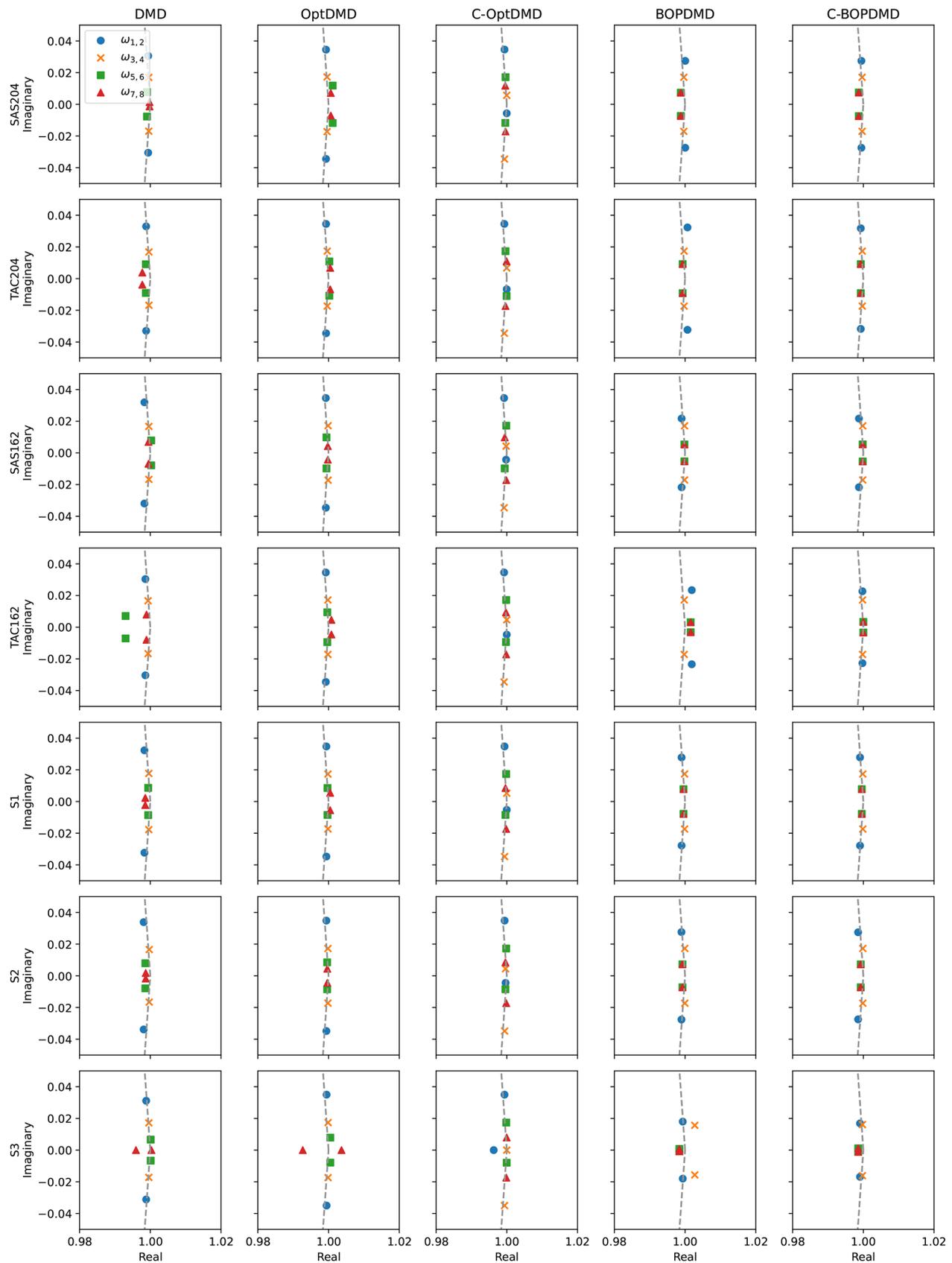
Figure C2. Eigenvalues of the conjugate pairs for the first four modes for each of the DMD variants applied to the different datasets. Markers show the conjugate pairs of the modes, and the dashed line shows the unit circle.
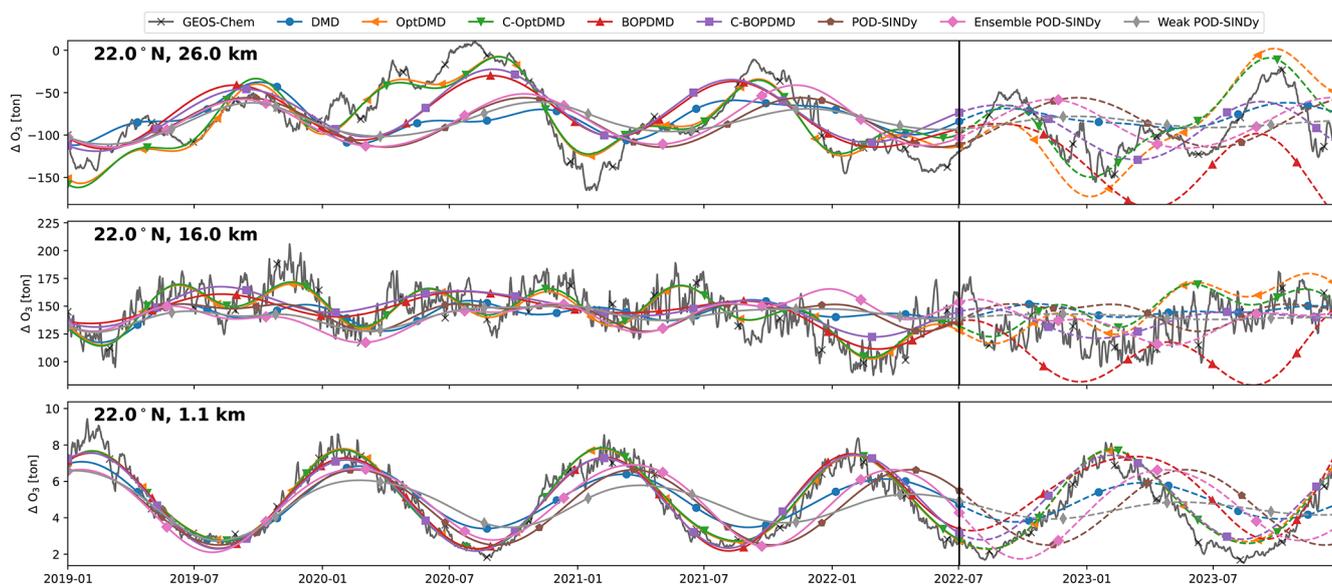
Figure C3. Ozone response in tonnes for 3 grid cells of the TAC162 dataset, with reconstructions and short-term forecasts from the DMD and POD-SINDy methods. Cells shown are located at 22° N latitude and altitudes of 26.0, 18.3, and 1.4 km from top to bottom respectively. Vertical black lines mark the difference between the reconstruction (solid lines) and forecast (dashed lines).
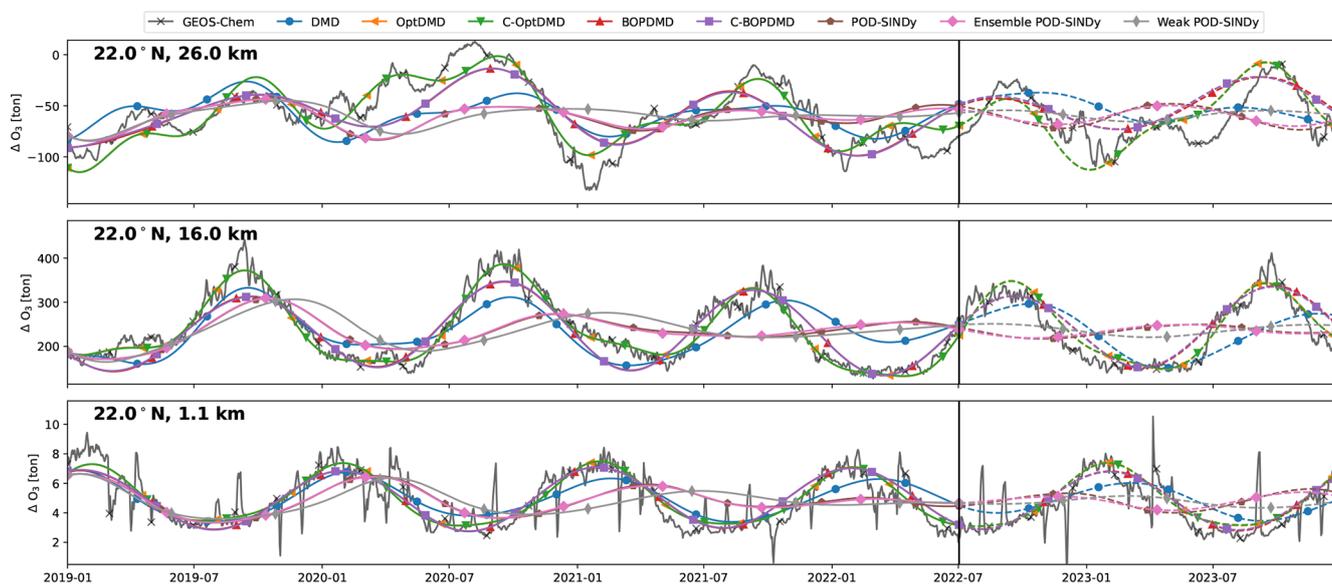
Figure C4. Same as Fig. C3, but for the SAS162 dataset.
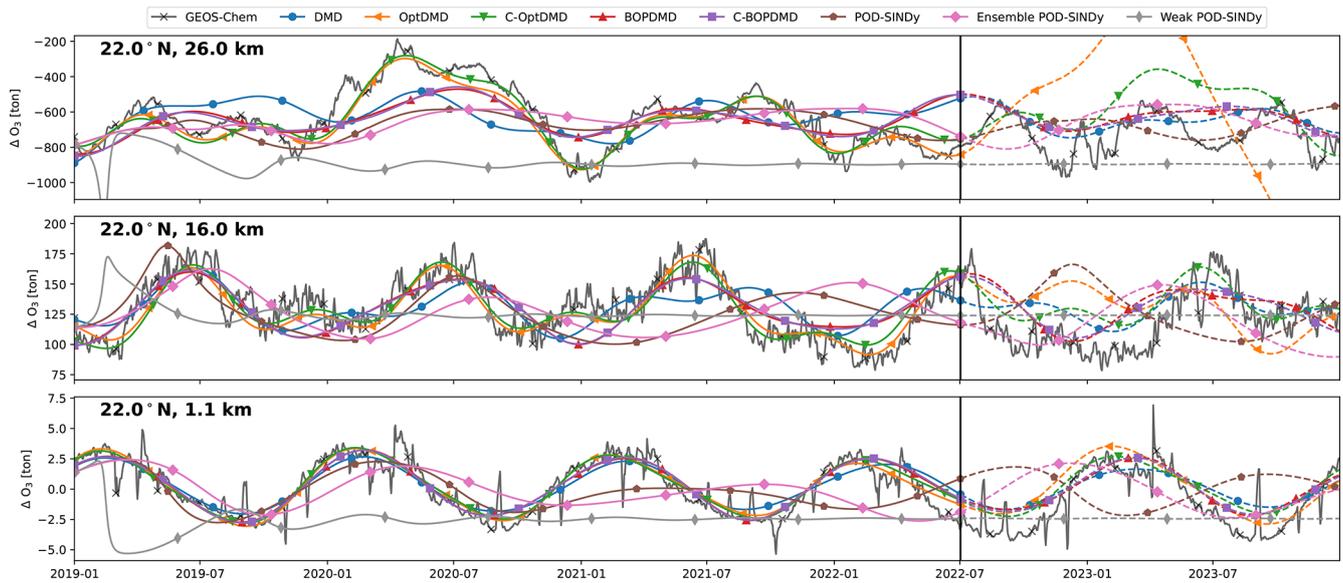
Figure C5. Same as Fig. C3, but for the SAS204 dataset.
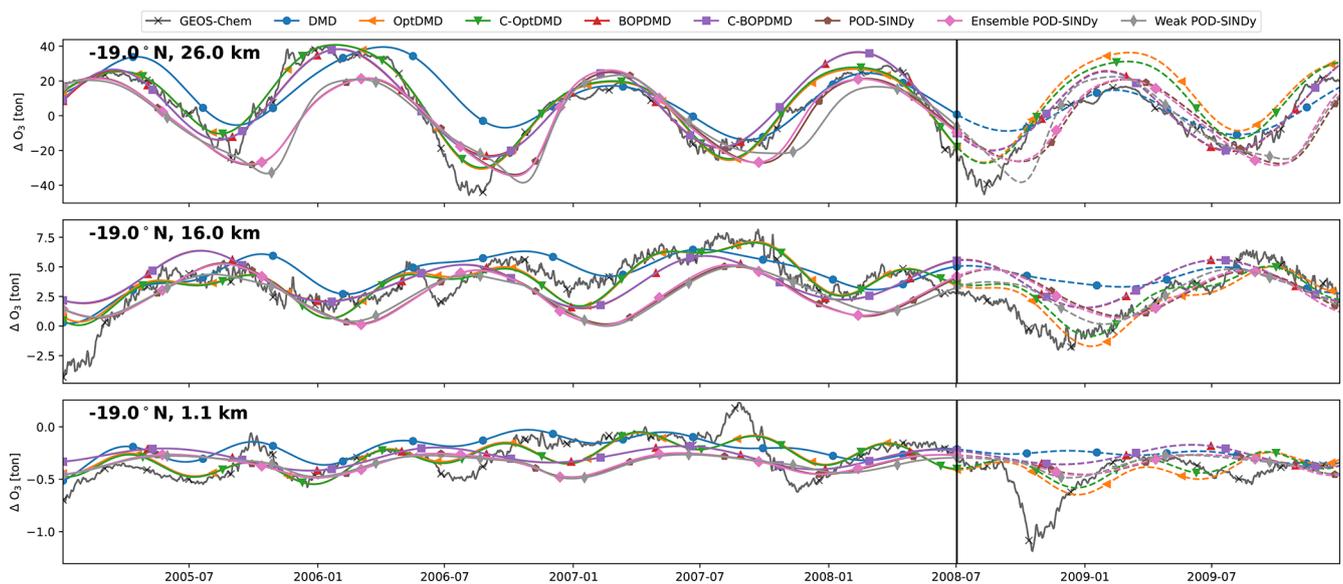
Figure C6. Same as Fig. C3, but for the S1 dataset.
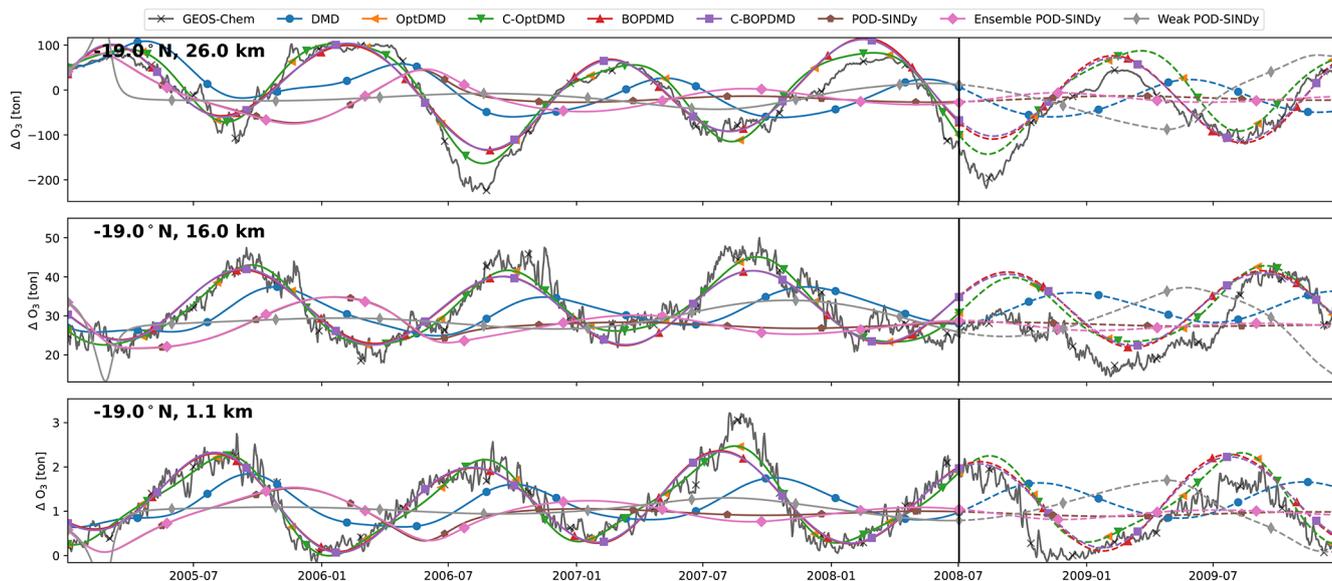
Figure C7. Same as Fig. C3, but for the S2 dataset.
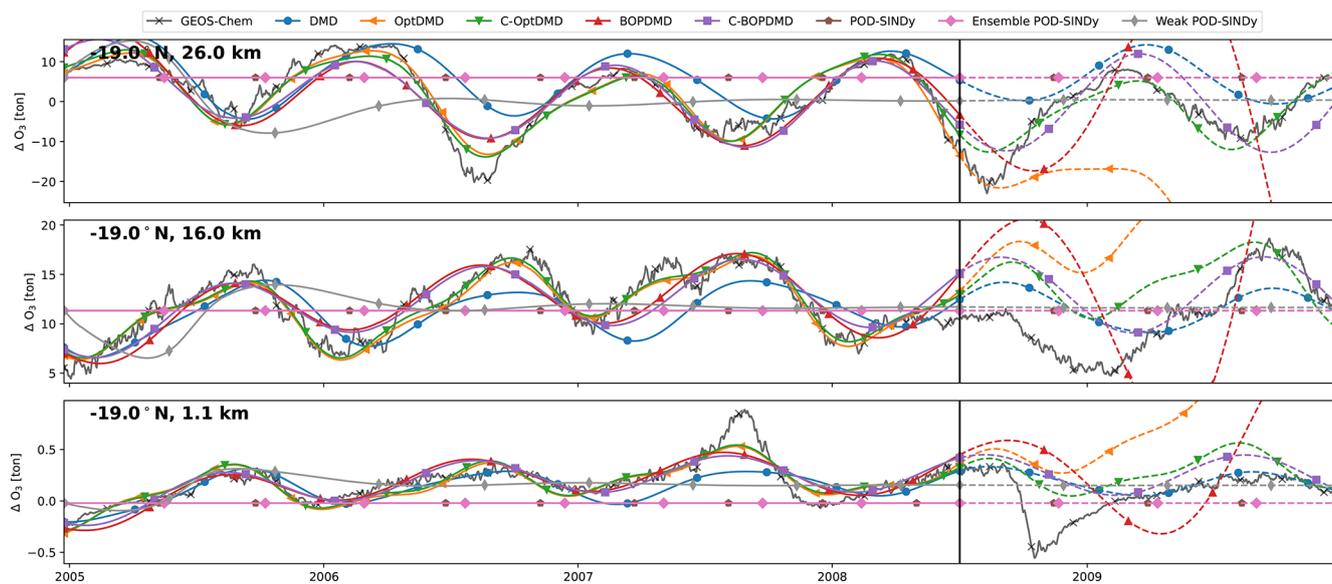
Figure C8. Same as Fig. C3, but for the S3 dataset.
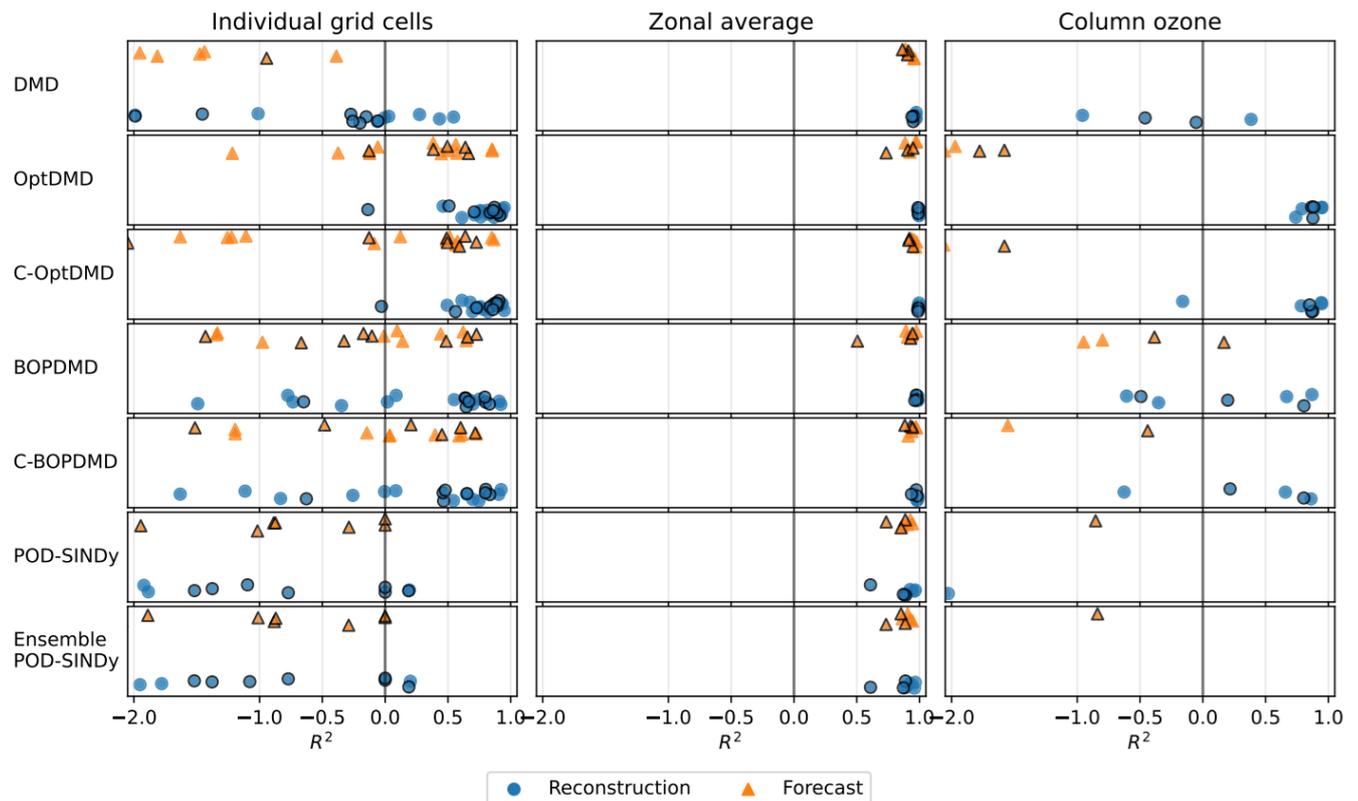
Figure C9. Comparison of the coefficient of determination for the single-cell (left panels), zonal-average (middle panels), and column ozone response (right panels). Reconstructions (blue) and forecasts (orange) are shown separately for the individual methods across all datasets.

Author contributions. Conceptualization: ID, UF; Data curation: JH; Investigation: All; Methodology: All; Software: TC, UF, JH; Writing – original draft: TC, JH; Writing – review and editing: all.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

References

Askham, T. and Kutz, J. N.: Variable projection methods for an optimized dynamic mode decomposition, SIAM J. Appl. Dynam. Syst., 17, 380–416, https://doi.org/10.1137/M1124176, 2018.

Boninsegna, L., Nüske, F., and Clementi, C.: Sparse Learning of Stochastic Dynamical Equations, J. Chem. Phys., 148, 241723, https://doi.org/10.1063/1.5018409, 2018.

Brunton, S. L. and Kutz, J. N.: Data-Driven Science and Engineering Machine Learning, Dynamical Systems, and Control, https://doi.org/10.1017/9781009089517, 2021.

Brunton, S. L., Proctor, J. L., Kutz, J. N., and Bialek, W.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems, P. Natl. Acad. Sci. USA, 113, 3932–3937, https://doi.org/10.1073/PNAS.1517384113, 2016.

Callaham, J., Brunton, S., Kutz, J. N., and Storti, D.: Multiscale model reduction for unsteady fluid flow, PhD thesis, https://digital.lib.washington.edu/researchworks/items/b4329d93-710b-4abe-ba60-fb3e2554d8cf (last access: 2 February 2026), 2022.

Champion, K., Lusch, B., Nathan Kutz, J., and Brunton, S. L.: Data-driven discovery of coordinates and governing equations, P. Natl. Acad. Sci. USA, 116, 22445–22451, https://doi.org/10.1073/PNAS.1906995116, 2019.

Demo, N., Tezzele, M., and Rozza, G.: PyDMD: Python Dynamic Mode Decomposition Software, J. Open Sour. Softw., 3, 530, https://doi.org/10.21105/joss.00530, 2018.

de Silva, B. M., Champion, K., Quade, M., Loiseau, J.-C., Nathan Kutz, J., and Brunton, S. L.: PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data, J. Open Sour. Softw., 5, 2104, https://doi.org/10.21105/joss.02104, 2017.

Eastham, S. D., Weisenstein, D. K., and Barrett, S. R. H.: Development and Evaluation of the Unified Tropospheric–Stratospheric Chemistry Extension (UCX) for the Global Chemistry-Transport Model GEOS-Chem, Atmos. Environ., 89, 52–63, https://doi.org/10.1016/j.atmosenv.2014.02.001, 2014.

Eastham, S. D., Fritz, T., Sanz-Morère, I., Prashanth, P., Allroggen, F., Prinn, R. G., Speth, R. L., and Barrett, S. R.: Impacts of a near-future supersonic aircraft fleet on atmospheric composition and climate, Environ. Sci.: Atmos., 2, 388–403, https://doi.org/10.1039/d1ea00081k, 2022.

Fasel, U., Kutz, J. N., Brunton, B. W., and Brunton, S. L.: Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control, P. Roy. Soc. A, 478, https://doi.org/10.1098/RSPA.2021.0904, 2022.

Ichinaga, S. M., Andreuzzi, F., Demo, N., Tezzele, M., Lapo, K., Rozza, G., Brunton, S. L., and Kutz, J. N.: PyDMD: A Python package for robust dynamic mode decomposition, Journal of Machine Learning Research, 25, 1–9, 2024.

Kaptanoglu, A. A., Callaham, J. L., Aravkin, A., Hansen, C. J., and Brunton, S. L.: Promoting Global Stability in Data-Driven Models of Quadratic Nonlinear Dynamics, Phy. Rev. Fluids, 6, 094401, https://doi.org/10.1103/PhysRevFluids.6.094401, 2021a.

Kaptanoglu, A. A., de Silva, B. M., Fasel, U., Kaheman, K., Goldschmidt, A. J., Callaham, J. L., Delahunt, C. B., Nicolaou, Z. G., Champion, K., Loiseau, J.-C., Kutz, J. N., and Brunton, S. L.: PySINDy: A comprehensive Python package for robust sparse system identification, J. Open Sour. Softw., 7, 3994, https://doi.org/10.21105/joss.03994, 2021b.

Khodkar, M. A. and Hassanzadeh, P.: A data-driven, physics-informed framework for forecasting the spatiotemporal evolution of chaotic dynamics with nonlinearities modeled as exogenous forcings, J. Comput. Phys., 440, 110412, https://doi.org/10.1016/J.JCP.2021.110412, 2021.

Kwasniok, F.: Linear Inverse Modeling of Large-Scale Atmospheric Flow Using Optimal Mode Decomposition, J. Atmos. Sci., 79, 2181–2204, https://doi.org/10.1175/JAS-D-21-0193.1, 2022.

Loiseau, J.-C. and Brunton, S. L.: Constrained sparse Galerkin regression, J. Fluid Mech., 838, 42–67, 2018.

Mangan, N. M., Kutz, J. N., Brunton, S. L., and Proctor, J. L.: Model selection for dynamical systems via sparse regression and information criteria, P. Roy. Soc. A, 473, https://doi.org/10.1098/rspa.2017.0009, 2017.

Messenger, D. A. and Bortz, D. M.: Weak SINDy for partial differential equations, J. Comput. Phys., 443, 110525, https://doi.org/10.1016/J.JCP.2021.110525, 2021.

Newman, M.: An Empirical Benchmark for Decadal Forecasts of Global Surface Temperature Anomalies, J. Climate, 26, 5260–5269, https://doi.org/10.1175/JCLI-D-12-00590.1, 2013.

Perkins, W. A. and Hakim, G.: Linear Inverse Modeling for Coupled Atmosphere-Ocean Ensemble Climate Prediction, J. Adv. Model. Earth Syst., 12, e2019MS001778, https://doi.org/10.1029/2019MS001778, 2020.

Rao, P., Dwight, R., Singh, D., Maruhashi, J., Dedoussi, I., Grewe, V., and Frömming, C.: The Ozone Radiative Forcing of Nitrogen Oxide Emissions from Aviation Can Be Estimated Using a Probabilistic Approach, Commun. Earth Environ., 5, 1–12, https://doi.org/10.1038/s43247-024-01691-2, 2024.

Reinbold, P. A. K., Gurevich, D. R., and Grigoriev, R. O.: Using noisy or incomplete data to discover models of spatiotemporal dynamics, Phys. Rev. E, 101, https://doi.org/10.1103/PhysRevE.101.010203, 2020.

Rudy, S. H., Brunton, S. L., Proctor, J. L., and Kutz, J. N.: Data-driven discovery of partial differential equations, Sci. Adv., 3, 123, https://doi.org/10.1126/sciadv.1602614, 2017.

Sashidhar, D. and Kutz, J. N.: Bagging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification, Philos. T. Roy. Soc. A, 380, https://doi.org/10.1098/RSTA.2021.0199, 2022.

Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem., 36, 1627–1639, https://doi.org/10.1021/ac60214a047, 1964.

Speth, R. L., Eastham, S. D., Prashanth, P., and Allroggen, F.: Global Environmental Impact of Supersonic Cruise Aircraft in the Stratosphere, Tech. rep., Massachusetts Institute of Technology, Cambridge, Massachusetts, https://ntrs.nasa.gov/api/citations/20205009400/downloads/CR-20205009400.pdf (last access: 2 February 2026), 2021.

Taira, K., Brunton, S. L., Dawson, S. T., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., and Ukeiley, L. S.: Modal analysis of fluid flows: An overview, AIAA J., 55, 4013–4041, https://doi.org/10.2514/1.J056060, 2017.

Taira, K., Hemati, M. S., Brunton, S. L., Sun, Y., Duraisamy, K., Bagheri, S., Dawson, S. T., and Yeh, C. A.: Modal analysis of

fluid flows: Applications and outlook, AIAA J., 58, 998–1022, https://doi.org/10.2514/1.J058462, 2020.

van 't Hoff, J., Grewe, V., and Dedoussi, I. C.: Supporting Dataset for "Sensitivities of Atmospheric Ozone and Radiative Forcing to Supersonic Aircraft Emissions across Two Flight Corridors", 4TU.ResearchData [data set], https://doi.org/10.4121/D5947A0D-F34D-400B-87DE-46EBDA16EC44.V1, 2024a.

van 't Hoff, J., Hauglustaine, D., Pletzer, J., Skowron, A., Grewe, V., Matthes, S., Meuser, M. M., Thor, R. N., and Dedoussi, I. C.: Supporting Dataset for "Multi-model Assessment of the Atmospheric and Radiative Effects of Supersonic Transport Aircraft", 4TU.ResearchData [data set], https://doi.org/10.4121/DD38833D-6C5D-47D8-BB10-7535CE1EECF1.V1, 2025a.

van 't Hoff, J., van Cranenburgh, T., Fasel, U., and Dedoussi, I.: Supporting data and code for "Data-driven discovery and model reduction methods for the atmospheric effects of high altitude emissions", 4TU.ResearchData [code and data set], https://doi.org/10.4121/D7C8091B-FC2B-4C21-A498-D4A01C9A7A40, 2025b.

van 't Hoff, J. A., Grewe, V., and Dedoussi, I. C.: Sensitivities of Ozone and Radiative Forcing to Supersonic Aircraft Emissions Across Two Flight Corridors, J. Geophys. Res.-Atmos., 129, e2023JD040476, https://doi.org/10.1029/2023JD040476, 2024b.

van 't Hoff, J. A., Hauglustaine, D., Pletzer, J., Skowron, A., Grewe, V., Matthes, S., Meuser, M. M., Thor, R. N., and Dedoussi, I. C.: Multi-model assessment of the atmospheric and radiative effects of supersonic transport aircraft, Atmos. Chem. Phys., 25, 2515–2550, https://doi.org/10.5194/acp-25-2515-2025, 2025c.

Velegar, M., Keller, C., and Kutz, J. N.: Optimized Dynamic Mode Decomposition for Reconstruction and Forecasting of Atmospheric Chemistry Data, arXiv [preprint], https://doi.org/10.48550/arXiv.2404.12396, 2024.

Weiss, J.: A Tutorial on the Proper Orthogonal Decomposition, Tech. rep., Technische Universität Berlin, https://doi.org/10.14279/depositonce-8512, 2019.

Yang, X., Guo, L., Zheng, Z., Riemer, N., and Tessum, C. W.: Atmospheric Chemistry Surrogate Modeling With Sparse Identification of Nonlinear Dynamics, J. Geophys. Res.-Mach. Learn. Comput., 1, e2024JH000132, https://doi.org/10.1029/2024JH000132, 2024.

Zanna, L. and Bolton, T.: Data-driven equation discovery of ocean mesoscale closures, Geophys. Res. Lett., 47, e2020GL088376, https://doi.org/10.1029/2020GL088376, 2020.

Zhang, J., Wuebbles, D., Kinnison, D., and Baughcum, S. L.: Potential Impacts of Supersonic Aircraft Emissions on Ozone and Resulting Forcing on Climate: An Update on Historical Analysis, J. Geophys. Res.-Atmos., 126, https://doi.org/10.1029/2020JD034130, 2021a.

Zhang, J., Wuebbles, D., Kinnison, D., and Baughcum, S. L.: Stratospheric Ozone and Climate Forcing Sensitivity to Cruise Altitudes for Fleets of Potential Supersonic Transport Aircraft, J. Geophys. Res.-Atmos., 126, e2021JD034971, https://doi.org/10.1029/2021JD034971, 2021b.

Zhang, J., Wuebbles, D., Pfaender, J. H., Kinnison, D., and Davis, N.: Potential Impacts on Ozone and Climate From a Proposed Fleet of Supersonic Aircraft, Earth's Future, 11, e2022EF003409, https://doi.org/10.1029/2022EF003409, 2023.