

Causal inference for quantifying chemical–dynamical pathways controlling tropical middle stratospheric ozone variability

Evgenia Galytska^{1,2}, Birgit Hassler², Carlo Arosio¹, Martyn P. Chipperfield^{3,4}, Sandip S. Dhomse^{3,4}, Kimberlee Dubé⁵, Wuhu Feng^{3,6}, Fernando Iglesias-Suarez², and Jakob Runge⁷

¹University of Bremen, Institute of Environmental Physics, Bremen, Germany

²Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

³School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK

⁴National Centre for Earth Observation, University of Leeds, Leeds LS2 9JT, UK

⁵Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, Canada

⁶National Centre for Atmospheric Science, University of Leeds, Leeds LS2 9PH, UK

⁷Department of Computer Science, University of Potsdam, Potsdam, Germany

Correspondence: Evgenia Galytska (egalytska@iup.physik.uni-bremen.de)

Abstract.

Understanding the chemical–dynamical interactions controlling ozone (O_3) variability in the tropical middle stratosphere is essential for interpreting short-term trends and their sensitivity to dynamical fluctuations. This study applies a process-oriented causal inference framework that combines causal discovery and causal effect estimation. This approach integrates qualitative physical knowledge through a causal graph applied to satellite observations and a chemistry-transport model (CTM) simulation, using monthly data for the period 2004–2021. Causal inference robustly identifies a dominant chemical–dynamical pathway, in which variability in residual vertical velocity (w^*) modulates nitrous oxide (N_2O), subsequently affecting nitrogen dioxide (NO_2) and ultimately O_3 . Estimates of direct causal effect capture that O_3 variability is dominated by this indirect NO_2 -mediated pathway, while the direct influence of w^* on O_3 is weak. The total causal effect (direct and mediated) peaks at a lag of approximately two-three months, indicating the cumulative impact of persistent, vertically coupled w^* anomalies associated with the QBO. Regime-oriented analysis applied to the observations reveals that the chemical links (N_2O – NO_2 and NO_2 – O_3) strengthen during westerly QBO shear compared to easterly shear.

Our study highlights the pivotal role that causal inference can play in disentangling complex chemical-dynamical influences on O_3 , complementing traditional statistical methods. This approach lays the foundation for broader applications in stratospheric chemistry, where the understanding of various feedback pathways remains uncertain. By discovering and quantifying causal links, this methodology can be adapted to address open questions with environmental and societal relevance. Integrating causal reasoning into data-driven science can enhance process understanding and also strengthen the synergy between machine learning and statistical methods in Earth and environmental sciences.

1 Introduction

20 Stratospheric ozone (O_3) is essential for protecting life on Earth by absorbing most of the harmful solar ultraviolet (UV-B) radiation (280–315 nm). The tropical (10°S – 10°N) middle stratosphere (~ 10 hPa) is a key region for the photochemical formation of O_3 (Chapman, 1930). A balance between photochemical production and loss, together with dynamical transport, determines the overall abundance of stratospheric O_3 (Match et al., 2025). Its global distribution and inter-annual changes are mainly determined by dynamical and chemical processes in conjunction with their superimposed variability of different origins and periodicity, such as Brewer–Dobson circulation (BDC), El Niño–Southern Oscillation (ENSO), Quasi-Biennial Oscillation (QBO), concentrations of greenhouse gases (GHGs) and Ozone Depleting Substances (ODSs). Reduction in stratospheric O_3 concentrations allows more UV radiation to reach the Earth’s surface, resulting in harmful effects, including damage to plant life and crops, disruption of aquatic ecosystems, and increased risks of skin cancer, cataracts, immune suppression, and erythema in humans (WMO, 2022; Zerefos et al., 2023). To mitigate these harmful effects, actions taken under the Montreal Protocol and its Amendments and Adjustments have significantly reduced emissions and atmospheric abundances of ODSs, contributing to the recovery of the stratospheric O_3 layer.

Apparent changes in tropical middle stratospheric O_3 can vary substantially depending on the length and timing of the analyzed period. For example, in the early 2000s, several studies reported statistically significant O_3 decline in this region using a variety of datasets and methodologies (Kyrölä et al., 2013; Eckert et al., 2014; Gebhardt et al., 2014; Nedoluha et al., 2015; Galytska et al., 2019; Arosio et al., 2019; Iglesias-Suarez et al., 2021). Subsequent analyses showed that the slight differences in the tropical O_3 trends across different data sources arise from small shifts in the analyzed time period, largely due to endpoint anomalies influenced by the curvature of long-term O_3 variability and unaccounted fluctuations in the record (Petropavlovskikh et al., 2019; Sofieva et al., 2021). Recent assessments show that the magnitude of O_3 trends in the tropical middle stratosphere since the early 2000s is highly uncertain (–2 to 0% per decade, WMO, 2022) with multiple studies finding no statistically significant or robust long-term trend across different datasets and time periods (Petropavlovskikh et al., 2019; Steinbrecht et al., 2017; Godin-Beekmann et al., 2022; Arosio et al., 2024). In addition, Szeląg et al. (2020) highlighted a strong seasonal dependence in tropical middle stratospheric O_3 trends during 2000–2018, with a significant increase in spring (2–3% per decade) and a non-significant decrease in autumn (–1% to –2% per decade), resulting in non-significant trends, consistent with Galytska et al. (2019); Li et al. (2023). However, the absence of a persistent long-term trend in recent analyses does not rule out future changes of O_3 in the tropical middle stratosphere. Similarly, negative trends observed in the early 2000s could recur if comparable dynamical–chemical conditions arise. In this context, understanding the mechanisms governing interannual O_3 variability becomes particularly important, as such variability can substantially influence trends derived from a limited observational dataset.

While previously discussed O_3 trends motivate this study, our objective is to quantify the mechanisms that control O_3 variability on monthly timescales and thus, that can modulate trends over limited time periods. The sensitivity of O_3 trends in the tropical middle stratosphere to the analyzed period highlights the dominant role of chemical–dynamical variability in this region. In this context, Galytska et al. (2019) showed that the O_3 decline observed in the early 2000s was dynamically

controlled and linked to increases in nitrogen dioxide (NO_2). The elevated NO_2 levels enhanced O_3 loss through the catalytic NO_x ($\text{NO}_x = \text{NO} + \text{NO}_2$) cycle, dominant in the middle stratosphere (Portmann et al., 2012). The increase in NO_2 , also confirmed by Dubé et al. (2020), resulted from the prolonged residence time of its primary source, nitrous oxide (N_2O). Since N_2O is a long-lived species (Portmann et al., 2012; Chipperfield et al., 2014), the changes in its abundance reflect variations in tropical upwelling within the BDC, therefore, they are directly affected by changes in stratospheric transport (Iglesias-Suarez et al., 2021). Prather et al. (2023) showed that during 2005-2021, N_2O increased in the tropical middle stratosphere more than expected from the rate of tropospheric increases. This implies a more vigorous BDC, leading to a shorter lifetime. Minganti et al. (2020) evaluated the climatological impact of the stratospheric BDC on N_2O and reported that, in the tropical middle stratosphere, the interannual variability of vertical residual advection exhibits a significant spread, which reflects the influence of the major source of interannual variability in the equatorial stratosphere, i.e. the QBO on tropical upwelling (Abalos et al., 2015). Chipperfield and Gray (1992), and later Park et al. (2017) highlighted that the source gas N_2O and the reactive nitrogen species (NO_y) display coherent QBO signals within the tropical stratosphere. O_3 also exhibits a strong QBO signal, with the QBO directly influencing O_3 levels by altering the chemical reactions responsible for O_3 depletion, see also Chipperfield et al. (1994) and Ming et al. (2025). This oscillation modulates the rates of these reactions, leading to additional variations in O_3 concentration. It is also important to note that recent work by Hitchcock and Ming (2025) demonstrates that ozone can play a leading role in modulating the QBO, highlighting a two-way chemical–dynamical interaction, which is the subject of ongoing research and clarification (Orbe et al., 2026).

Therefore, to set the stage for our subsequent causal analysis of these chemical-dynamical feedback pathways on O_3 variability in the tropical middle stratosphere, Figure 1 presents scatter plots of detrended monthly mean anomalies of (a) residual vertical velocity (w^*) versus N_2O , (b) N_2O versus NO_2 , (c) NO_2 versus O_3 , and (d) N_2O versus O_3 for observations and the TOMCAT CTM simulation for 2004-2021. Both observations and the TOMCAT CTM simulation exhibit similar correlations and slopes. The relationship between w^* and N_2O is moderately positive (panel a). The strong anti-correlations between N_2O and NO_2 (panel b) result from their opposing response to transport-driven variability, where, e.g. enhanced upwelling increases N_2O while reducing its chemical loss, which is crucial for NO_2 production (see further discussion in Sect. 4.1). The negative relationship between NO_2 and O_3 (panel c) is a result of NO_2 being the primary sink of O_3 in the middle stratosphere (Park et al., 2017). The positive relationship between N_2O and O_3 (panel d) reflects transport-controlled variability in N_2O , while O_3 is mainly controlled by photochemistry (Brasseur and Solomon, 2005); the observed correlation results from both tracers varying systematically with altitude in the tropical middle stratosphere.

While the key chemical reactions and dynamical processes governing tropical middle-stratospheric O_3 have been identified and can be simulated in chemistry-climate models (CCMs), previous research relies mostly on correlation and different types of regression analyses that do not explicitly distinguish between direct and mediated causal effects within a multivariate system. A causal inference framework proposed in this study represents the relationships between the selected variables as a directed acyclic graph (DAG), in which nodes correspond to physical quantities and edges represent causal influences. Therefore, the structure of the DAG is informed by established physical understanding and subsequently evaluated using causal inference methods applied to the data. Causal inference can then be used to estimate the targeted relationships under explicit

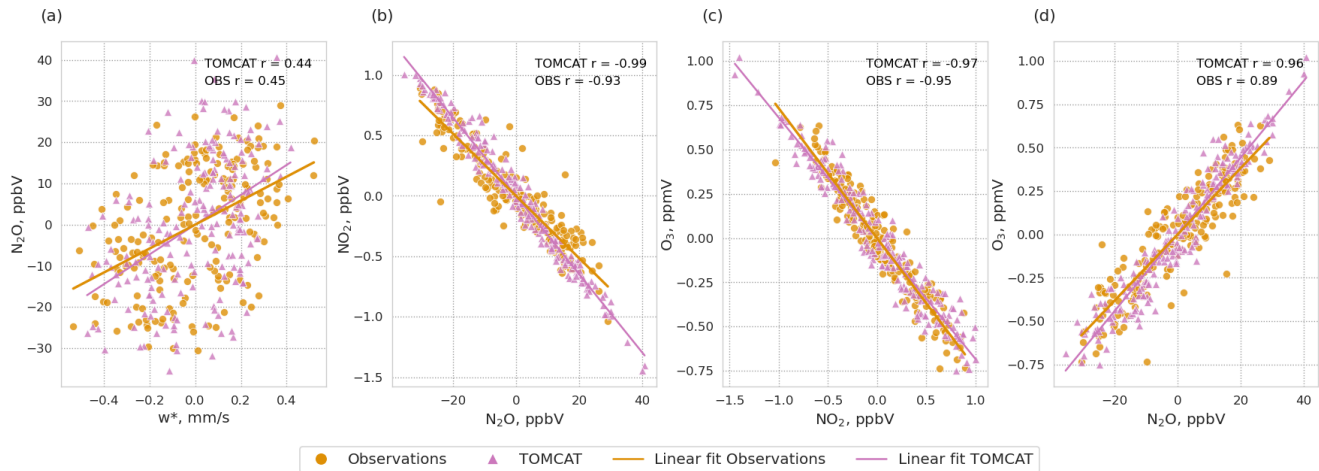


Figure 1. Scatter plots of detrended monthly mean anomalies in the tropical middle stratosphere for 2004–2021, from observations (circles) and the TOMCAT CTM simulation (triangles). Panels show (a) w^* versus N_2O , (b) N_2O versus NO_2 , (c) NO_2 versus O_3 , and (d) N_2O versus O_3 . Solid lines indicate linear regressions for observations (orange) and TOMCAT (pink). The corresponding Pearson correlation coefficients (r) are shown in each panel.

assumptions. Applying such an approach to a well-understood chemical–dynamical system provides a reliable test of whether causal inference can recover known physical pathways and quantify their contributions in the analyzed system. To connect the discussion of stratospheric chemical–dynamical interactions with the causal inference framework used in this study, Appendix A provides a glossary that links key causal terms to their interpretation in a stratospheric context. For a broader overview of the terminology, the reader is referred to Camps-Valls et al. (2023) and Runge et al. (2023). It is important to highlight that to ensure statistical consistency, this study analyzes detrended anomalies rather than long-term trends. Therefore, the results describe variability-driven processes and should not be interpreted as a direct explanation of decadal O_3 trends caused by externally forced long-term changes in CO_2 , N_2O emissions, or ODSs, but rather in the context of dynamical processes that are strongly influenced by the QBO.

2 Data sources

This study uses monthly data from satellite observations, reanalysis products, and a TOMCAT CTM simulation, focusing on four core variables, namely w^* , N_2O , NO_2 , and O_3 . We also use a measure of the QBO as an additional variable to examine how the relationships among the four core variables vary under different dynamical regimes. By selecting these variables, we focus the analysis on the interpretability of key chemical-dynamical processes that play a major role in controlling O_3 in the tropical middle stratosphere. We intentionally limit the number of variables to maintain a high level of interpretability of the causal graphs. We then compare the causal graphs from observations and the TOMCAT CTM simulation.

105 2.1 Observations

Since no single satellite instrument provides all the required variables with sufficient temporal and spatial resolution, we integrate data from multiple sources into this study. The following observational or reanalysis-based datasets were used:

- 110 – w^* : Derived Transformed Eulerian Mean (TEM) momentum terms (v0.1.1, Serva, 2022), based on European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis (Hersbach et al., 2020) using diagnostics from Serva (2023); Serva et al. (2024);
- N_2O : Profiles from the Earth Observing System (EOS) Microwave Limb Sounder (MLS, v5.01) instrument on NASA's Aura satellite, which offer a vertical resolution of 5–8 km and a horizontal along-track resolution of 165–265 km (Lambert et al., 2020);
- 115 – NO_2 : Profiles retrieved from limb-scattered sunlight observations on the OSIRIS instrument aboard the Swedish Odin satellite (Murtagh et al., 2002; Llewellyn et al., 2004). OSIRIS NO_2 v7.3 is retrieved via spectral fitting in the 435–477 nm range from 10.5 to 39.5 km with a 2–3 km vertical resolution in most of the stratosphere (Dubé et al., 2022). Due to the pronounced diurnal cycle of NO_2 (Galytska, 2019; Dubé et al., 2020) the photochemical correction from Dubé et al. (2020) was applied to standardize all measurements to a reference time of 12:00 p.m.;
- 120 – O_3 : OSIRIS O_3 v7.3 (Bognar et al., 2022), an improved version of the v5.10 product with corrected long-term drift by accounting for systematic errors in the instrument limb-pointing (Bourassa et al., 2018);
- QBO : equatorial zonal mean zonal wind taken from the Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, see Kerzenmacher and Braesicke (2026).

In the following, the combination of these datasets is collectively referred to as "observations".

2.2 TOMCAT Chemical Transport Model

125 TOMCAT is a three-dimensional off-line CTM (Chipperfield, 2006), driven here by winds and temperatures from the ERA5 reanalysis (Hersbach et al., 2020). Given prescribed atmospheric transport and temperature fields, TOMCAT calculates the distributions of chemical species in the troposphere and stratosphere. A stratospheric full-chemistry simulation, including all of the NO_y chemistry discussed in this paper, was run at a horizontal resolution of $2.8^\circ \times 2.8^\circ$ with approximately 1.5 km vertical resolution in the stratosphere (Chrysanthou et al., 2025). The model uses time varying sulfate aerosol surface area density Dhomse et al. (2015), solar fluxes (Dhomse et al., 2016), and lower boundary concentrations of GHGs and ODSs (WMO, 2022), recommended for CMIP6 simulations. We used the monthly average output in our analysis. The TOMCAT CTM simulation was chosen for its ability to provide a continuous time series without spatial or temporal gaps, making it ideal for robust comparison with the observational datasets.

2.3 Data preprocessing

135 Monthly anomalies of w^* , N_2O , NO_2 , and O_3 in the tropical ($10^\circ S$ – $10^\circ N$) middle (10 hPa) stratosphere were calculated from observations and the TOMCAT CTM simulation relative to the climatological mean over August 2004–December 2021. The analysis begins in August 2004, due to the availability of MLS N_2O observations. However, for simplicity, we refer to the period as 2004–2021. To maintain a robust comparison against temporal sampling discrepancies, TOMCAT CTM data were masked to align with observations, omitting any dates with missing observational data. For causal inference, all time series
 140 were standardized. Since the application of causality requires the stationarity of the time series (see Sect. 3.1), we remove the linear trends from all analyzed time series. For more details about preprocessed timeseries from observations and the TOMCAT CTM simulation and their further comparison, see Appendix B.

3 Methods

3.1 Causal inference

145 We apply the latent Peter-Clark momentary conditional independence (LPCMCI) algorithm (Gerhardus and Runge, 2020), which is an extension of the PCMCI+ algorithm specifically designed to deal with latent (i.e. unobserved) variables (Runge et al., 2019b; Runge, 2020). LPCMCI employs ideas from the Fast Causal Inference (FCI) algorithm to learn not only directed causal relationships but also infer the presence of latent confounders (Spirtes, 1995). LPCMCI benefits from the same ideas underlying PCMCI+ by increasing the effect size of conditional independence (CI) tests through including causal parents in
 150 conditioning sets. The LPCMCI method seeks to learn a Directed Partially Ancestral Graph (DPAG), which captures the causal relationships among the observed variables. In contrast to Maximal Ancestral Graphs (MAGs) that contain directed arrows (\rightarrow) and bidirected edges (in other words, double-headed arrows \leftrightarrow), PAGs can include additional edge types. These edges, drawn as $\circ \rightarrow$ and/or $\circ - \circ$, indicate the presence of hidden variables or uncertainty about the exact causal direction.

To understand the causal structure of the underlying complex dynamical system, the observed time series $\mathbf{X}_t = (X_t^1, \dots, X_t^N)$,
 155 where N stands for the different variables represented by time series, is assumed to follow the following causal process:

$$X_t^j = f_j \left(\mathcal{P}(X_t^j), \eta_t^j \right),$$

where f_j is a measurable function that depends on all its inputs, η_t^j represents dynamical noise (independent across $t' \neq t$). Here, $\mathcal{P}(X_t^j) \subset \mathbf{X}_{t+1}^- = (\mathbf{X}_t, \mathbf{X}_{t-1}, \dots) \setminus \{X_t^j\}$ denotes the set of parent variables of X_t^j since the value of X_t^j is determined from the variables in $\mathcal{P}(X_t^j)$ and the dynamical noise η_t^j . Bidirected links between X_t^i and X_t^j in this model imply that the
 160 associated noise terms η_t^i and η_t^j are dependent through unobserved η confounding. We assume causal stationarity, meaning $X_{t-\tau}^i \in \mathcal{P}(X_t^j)$ if and only if $X_{t-\tau-\Delta t}^i \in \mathcal{P}(X_{t-\Delta t}^j)$. In practice, causal discovery algorithms, including LPCMCI, require approximately stationary time series, meaning time series whose statistical properties, such as mean and variance, remain approximately constant over time. Nonstationary behavior, such as long-term trends, can introduce spurious statistical dependencies

and bias CI tests, leading to incorrect causal links. Therefore, removing or accounting for such trends (via masking or sliding window) is a methodological necessity to ensure that the algorithm identifies causal relationships associated with the internal dynamics of the system rather than coincidental alignment of long-term shifts in the variables (Runge et al., 2019b, 2023).

Additionally, we also assume the absence of cyclic causal relationships, which, due to the temporal order, limits interactions to contemporaneous cases only when $\tau = 0$. Furthermore, we assume the standard assumptions of constraint-based causal discovery, the Markov condition and the faithfulness condition (Spirtes et al., 2000; Runge et al., 2023), which implies that conditional independence in the observed distribution generated by the structural causal model above implies directional separation (i.e. separation of variables by conditioning on appropriate sets of other variables) in the associated time series graph and vice versa.

We conducted a series of sensitivity tests with various settings of significance level α_{pc} and the maximum time delay τ_{max} , but only used the causal graphs with $\alpha_{pc}=0.02$ and $\tau_{max} = 1$ for the toy model (see Sect. 4.1), and with $\alpha_{pc}=0.05$ and $\tau_{max} = 2$ for the observations and the TOMCAT CTM simulation (see Sect. 4.2). Since some of the analyzed variables have non-gaussian distributions, we use the RobustParCorr conditional independence test, which transforms the data to a normal distribution before the partial correlation test. This usage implies that we assume the functions f_j in the model above to be linear. As a trade-off to its ability to also deal with latent confounding, LPCMCI suffers from lower recall compared to, e.g., PCMCI+ (Runge, 2020). Despite this, LPCMCI successfully identified causal connections in observational data that align closely with expert knowledge and the literature review. However, LPCMCI did not robustly detect anticipated connections in the TOMCAT CTM simulation. To estimate the direct causal effects for the period 2004-2021 in the TOMCAT CTM simulation, we refined the causal graphs by incorporating expert knowledge and insights from the literature review (as further discussed in Sect. 3.3 and depicted in Fig. 2).

3.2 Causal effect estimation

Over a hundred years ago, Wright (1921) suggested a method to estimate causal effects in linear models. This approach estimates the so-called path coefficients for all links in causal paths and then sums the products of these path coefficients over all causal paths. This method applies only to DAGs; therefore, in the case of DPAG, we ensure that all edges are directed before applying causal effect estimation. Causal effect estimation consists of the following steps:

1. For all causal links $i \rightarrow j$ that belong to causal paths from X to Y (where X is a parent and Y is a child), estimate the path coefficient $\beta_{i \rightarrow j}$ by regressing j on all its parents in a multivariate regression and taking the coefficient corresponding to parent i (see e.g. Runge et al., 2023).
2. The causal effect is then computed as:

$$CE = \sum_{\text{causal paths}} \prod_{\text{link } i \rightarrow j \text{ in path}} \beta_{i \rightarrow j}$$

By restricting this estimator to paths that pass through at least one node among a selected set of mediators M^* , it is also possible to compute mediated causal effects (MCE). These are defined as:

$$MCE = \sum_{\text{causal paths through at least one } M \in M^* \text{ link } i \rightarrow j \text{ in path}} \prod \beta_{i \rightarrow j}$$

3.3 Process-oriented causal inference framework

To quantitatively characterize O_3 variability in the tropical middle stratosphere during 2004–2021, we employ a process-oriented causal inference approach depicted in Fig. 2. Although causal analysis has already gained significant application in atmospheric sciences, including the analysis of Arctic processes and their links to middle latitudes (Polkova et al., 2021; Docquier et al., 2022; Galytska et al., 2023; Kretschmer et al., 2020), teleconnections (Karmouche et al., 2023; Tibau et al., 2022; Carvalho-Oliveira et al., 2024), atmosphere-biosphere interactions (Krich et al., 2020), and evaluating climate models (Nowack et al., 2020; Debeire et al., 2025) and their sensitivities (Ricard et al., 2024), it has not yet been applied to the study of stratospheric chemical-dynamical interactions.

Figure 2 outlines the process-oriented causal inference framework, which is built upon three essential components (blue squares): (1) a causal graph that contains information about qualitative cause-and-effect relationships (Runge et al., 2023), (2) observational and/or modeled data, and (3) a method for estimating causal effects. To construct (1) the causal graph, we employ a triangulated approach (Denzin, 2010; Uleman et al., 2024) that integrates (i) expert knowledge, (ii) a comprehensive literature review, and (iii) a data-driven causal discovery algorithm. While each of these components can independently contribute to the creation of the causal graph, we recommend employing the triangulated approach to ensure a more robust and reliable framework. To assess the performance of the selected (iii) causal discovery algorithm before applying it to real-world data, we first construct a (iv) "toy model" using synthetic data. This synthetic dataset is designed to replicate the properties and challenges of the real system while incorporating known underlying ground truth from (i) expert knowledge and (ii) a literature review. The toy model is then used to evaluate the performance of the causal discovery method in realistic, finite sample scenarios (Camps-Valls et al., 2023). To ensure the robustness of the results, it is recommended to further perform sensitivity tests on free algorithm parameters, such as α_{pc} and τ_{max} for (iii) the causal discovery algorithm and (iv) the toy model.

It is important to note that if (iii) the causal discovery algorithm does not robustly detect anticipated relationships in the analyzed (2) real-world or modelled data, the user can integrate physical knowledge into the algorithm. Alternatively, the causal graph may be constrained purely based on (i) expert knowledge and (ii) a comprehensive literature review, including previous successful applications of causal discovery to related research topics. The final (1) causal graph, derived from (2) real-world data, serves as a foundation for estimating (3) direct and total causal effects. Causal effect estimation can be further refined through a process-oriented analysis, such as, for example, masking on different atmospheric regimes. Additionally, (v) total causal effects can be assessed across different time lags. This complex approach outlined in Fig. 2 ensures robust and reliable causal inference, particularly for complex systems such as the stratospheric chemical-dynamical interactions investigated here.

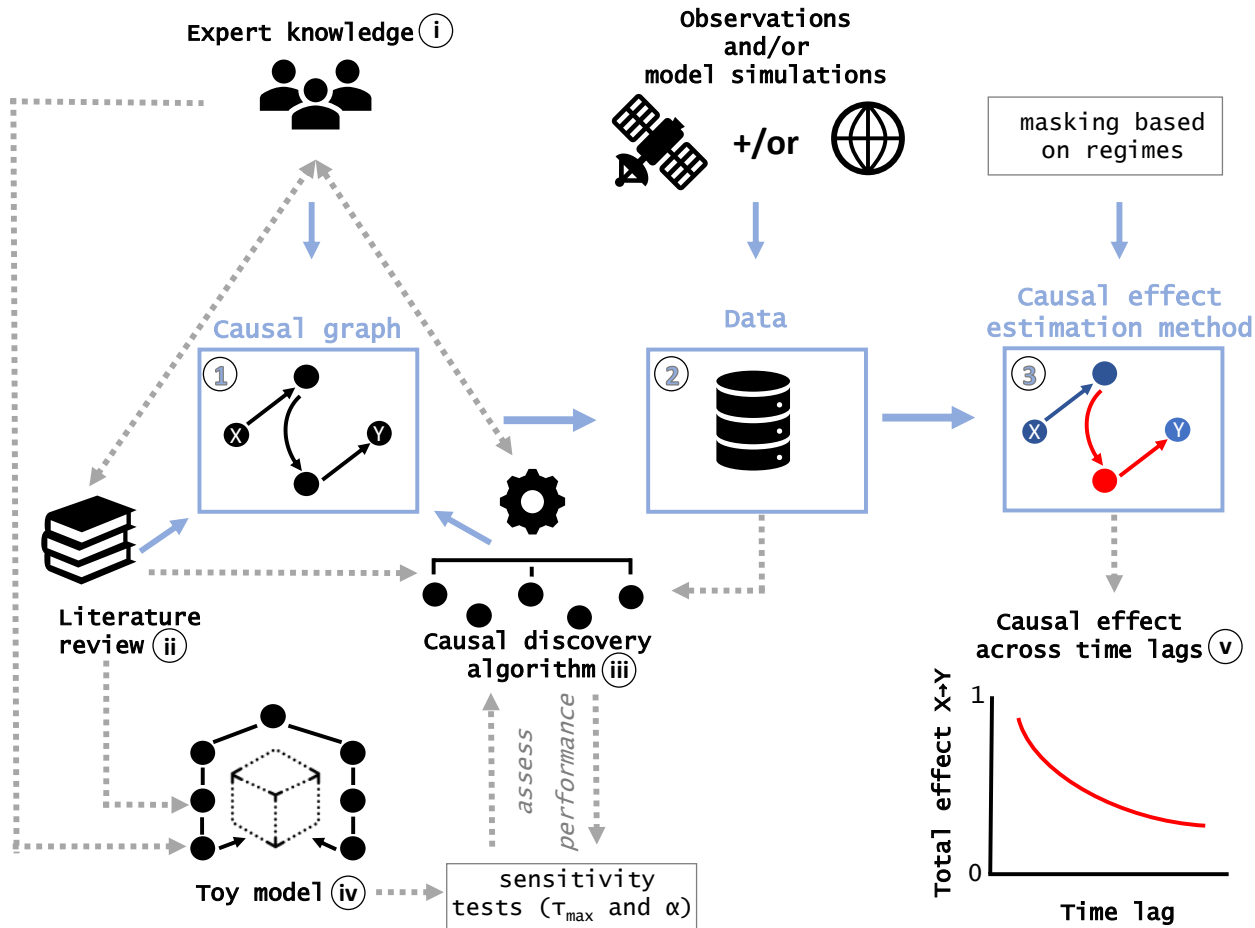


Figure 2. Process-oriented causal inference framework built upon three essential components (blue squares): (1) a causal graph, (2) data, and (3) a method for causal effect estimation. To construct the (1) causal graph for the studied system, a triangulated approach (Uleman et al., 2024) is applied, integrating (i) expert knowledge, (ii) a literature review, and (iii) a data-driven causal discovery algorithm. Before applying the (iii) causal discovery algorithm to real-world data, we construct a (iv) toy model to assess the performance of the selected (iii) causal discovery algorithm. The final (1) causal graph, based on (2) real-world data, serves as a foundation for estimating (3) causal effects, which can be further refined through process-oriented analysis, including masking on atmospheric regimes and additional sensitivity tests.

225 3.4 Confidence intervals and masking for regime-oriented analysis

Confidence intervals for direct causal effect estimates, computed using Wright’s path coefficient (Sect. 4.2), were obtained via bootstrapping with 500 repetitions. Only significant direct causal effects are shown, defined as those for which the bootstrap confidence interval does not include 0. For both direct and total causal effects across different time lags (Sect. 4.4), confidence intervals were also obtained from 500-member bootstrapping. Because the causal effect estimates were conducted on the same
230 data as the prior causal discovery step, the confidence intervals do not cover the uncertainty from this prior model-selection step. In our case, this issue is less pronounced because the causal graph was based on a triangulation and not fully data-driven.

For the regime-oriented analysis of causal effects during different QBO phases (Sect. 4.3), we first calculate the QBO wind shear as the vertical gradient of the zonal mean zonal wind between 10 and 30 hPa. For observations, the shear is derived directly from radiosonde-based zonal wind at the two pressure levels (Kerzenmacher and Braesicke, 2026). For the TOMCAT
235 CTM simulation, the zonal wind is first averaged over 10°S–10°N before computing the vertical gradient between 10 and 30 hPa. The resulting shear time series is subsequently standardized for use in the process-oriented causal analysis. Positive (negative) values correspond to a westerly (easterly) shear zone, which plays a key role in modulating secondary circulation and stratospheric transport. We focus on the 10-30 hPa shear layer since no data is available above 10 hPa in the observational record used here.

240 4 Results and Discussions

4.1 Causal justification and validation with a toy model

Before applying causal discovery to analyze chemical–dynamical interactions using observations and the TOMCAT CTM simulation, we first summarize the relationships of O₃ variability in this region in a shape of DAG-based on expert knowledge and literature review (as discussed and interpreted by e.g. Galytska et al., 2019; Nedoluha et al., 2015), following the procedure
245 discussed in Sect. 3.3. Figure 3a depicts a simple linear chain from the cause w^* to the outcome O₃ (grey nodes), with mediating variables N₂O and NO₂ (magenta nodes). The inferred DAG, therefore, represents an effective causal structure that emerges under the influence of dynamical variability, rather than a representation of isolated chemical relationships. In particular, a positive relationship from w^* to N₂O (labelled A) indicates that an increase in residual vertical velocity leads to enhanced N₂O concentrations. The relationship from N₂O to NO_x (labelled B) is negative, despite N₂O being a source of NO_x. This
250 apparent contradiction is an example of Simpson’s paradox (Blyth, 1972) and arises because tropical residual velocity w^* acts as a confounding dynamical process, leading to an anti-correlation between N₂O and NO₂. Namely, slower (faster) upwelling results in lower (higher) N₂O concentrations and consequently longer (shorter) N₂O residence time in this region, which allows more (less) time for the photochemical production of NO_x from N₂O. Consequently, higher (lower) NO_x levels lead to lower (higher) O₃ concentrations via the NO_x-catalyzed O₃ destruction cycle, resulting in a negative relationship (labelled C, see
255 Crutzen, 1970). Table 1 summarizes the discussed chemical-dynamical relationships in the tropical middle stratosphere as depicted in Fig. 3a.

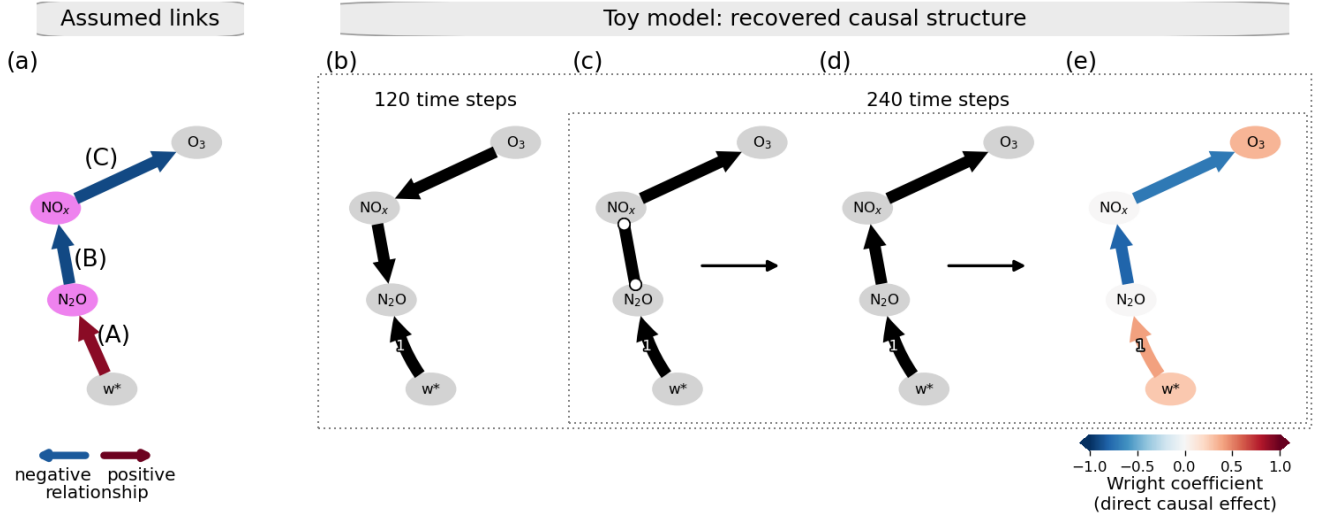


Figure 3. Causal justification and validation. (a) Assumed relationships based on expert knowledge and literature review. Magenta nodes indicate mediators in the total influence of w^* on O_3 . (b)-(e) Causal discovery applied to a generated toy model with contemporaneous (time lag = 0) and lagged (time lag > 0) dependencies based on linear partial correlation as a conditional independence test for 120 time steps (b) and 240 time steps (c), which corresponds to the DPAG. The application of the triangulated approach from Fig. 2 resulted in the DAG (d), which in turn was used as a basis for causal effect estimation, where edge colors indicate the estimated direct causal effects (e).

We further justify the assumed causal DAG in Fig. 3a and validate the reliability of the causal inference method. For this, we require a benchmark dataset with known causal ground truth for validation as depicted in Fig. 3a. We consider the linear structural causal process with four time series as an example that comes from a data-generating process using the following model:

$$\begin{aligned}
 X_t^0 &= 0.3X_{t-1}^0 + \eta_t^0 \\
 X_t^1 &= 0.3X_{t-1}^1 + 0.4X_{t-1}^0 + \eta_t^1 \\
 X_t^2 &= 0.3X_{t-1}^2 - 0.94X_t^1 + \eta_t^2 \\
 X_t^3 &= 0.3X_{t-1}^3 - 0.95X_t^2 + \eta_t^3
 \end{aligned} \tag{1}$$

where η_t stands for the independent Gaussian white noise processes with variances σ^2 , X^0 depicts residual vertical velocity w^* , X^1 - the concentration of N_2O , X^2 - NO_x , X^3 - O_3 . This set of variables and their dependencies define a toy model. Although this toy model is designed to replicate causal dependencies in the tropical middle stratosphere, it is important to emphasize that there are multiple approaches to constructing such a model. While additional variables could be further introduced based on expert knowledge and a thorough literature review, the goal of the analysis here is not to maximize the number of variables but to create an intuitive system that can simply and effectively replicate the processes under study.

Table 1. Summary of chemical-dynamical processes in the tropical middle stratosphere.

Label in Fig. 3a	Connection	Link type	Description
A	$w^* \rightarrow \text{N}_2\text{O}$	positive	Transport is the primary source of stratospheric N_2O . In addition to determining N_2O concentrations (an increase of w^* leads to an increase of N_2O), it also affects its residence time (an increase in w^* leads to a shorter residence time of N_2O).
B	$\text{N}_2\text{O} \rightarrow \text{NO}_x$	negative	N_2O is the primary source of NO_2 via the slow reaction $\text{N}_2\text{O} + \text{O}(^1\text{D}) \longrightarrow 2\text{NO}$ followed by the rapid reaction $\text{NO} + \text{O}_3 \longrightarrow \text{NO}_2 + \text{O}_2$. As N_2O increases due to faster circulation, its residence time decreases, and therefore, NO_2 concentrations decline (there is less time to produce NO_2). The resulting link is therefore negative. In the absence of dynamical variability, the relationship would appear positive, reflecting only the underlying chemical production of NO_2 from N_2O .
C	$\text{NO}_x \rightarrow \text{O}_3$	negative	NO_2 is the primary sink of O_3 in the tropical middle stratosphere via the NO_x ($\text{NO} + \text{NO}_2$) catalytic cycle.

Figure 3b-e illustrates the causal graphs inferred from the causal discovery algorithm (see Sect. 3.1) when applied to the toy model from Eq. (1). To evaluate the performance of the causal discovery algorithm on time series of different lengths, Fig. 3b,c show causal graphs for generated time series of 120 time steps (equivalent to 10 years of monthly data) and 240 time steps (equivalent to 20 years), respectively. When applied to the shorter time series (Fig. 3b), causal discovery fails to detect several expected connections as anticipated from Fig. 3a, likely due to limitations in the toy model, such as weaker causal signals or higher noise. In contrast, despite the plausible limitations of the toy model, most expected connections are recovered in the 20-year time series depicted in Fig. 3c. However, the inferred graph lacks directionality between N_2O and NO_2 (shown as $\circ-\circ$ edge), resulting in a DPAG. Since causal effect estimation requires a fully directed DAG, we applied the triangulation approach to resolve ambiguities, yielding the DAG in Fig. 3d. This final causal graph (Fig. 3d) serves as the basis for causal effect estimation, shown in Fig. 3e. Given approximately linear relationships among analyzed variables (see Fig. 1b-d), we assume linear causal effects and apply Wright’s method (Wright, 1921). This approach is akin to linear regression slopes between two variables X and Y, with the critical distinction that the graph is used to detect and eliminate confounding influences before regression (see Sect. 3.2 and Fons et al., 2023).

4.2 Physical description of observed and modeled causal relations

Figure 4 presents the magnitude and sign of direct causal effects computed using Wright’s approach (Wright, 1921) on causal graphs identified by the triangulation (Uleman et al., 2024) based on observations (a) and the TOMCAT CTM simulation (b) during 2004–2021. The original graphs inferred by the causal discovery algorithm (see Sect. 3.1) without the application of the triangulated approach are shown in Appendix C. The causal discovery algorithm successfully identifies the anticipated

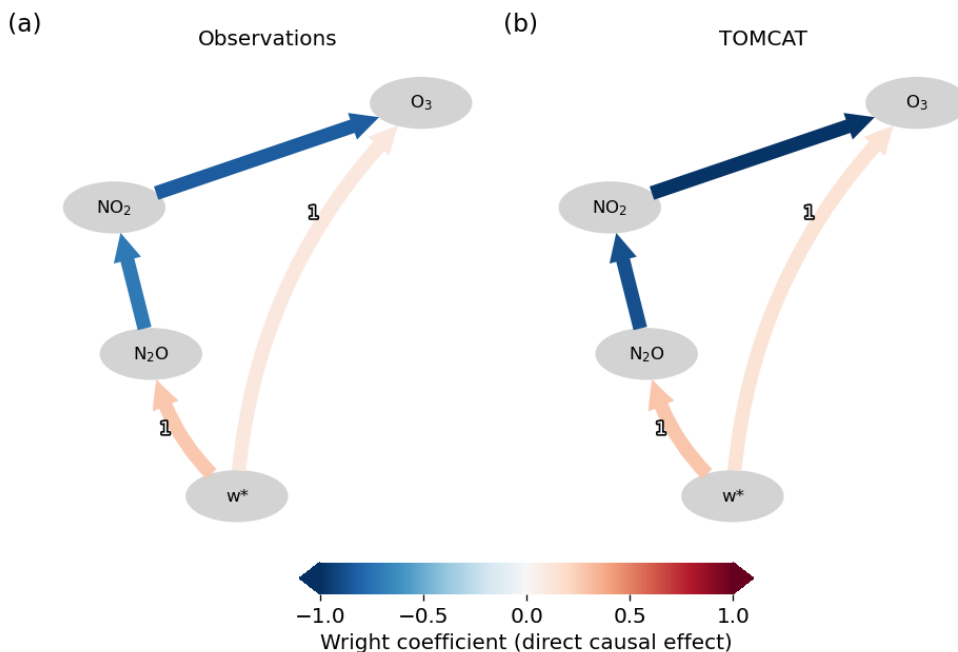


Figure 4. Causal inference: Magnitude and sign of the direct causal effects. The direct causal effects calculated for (a) the observations and (b) the TOMCAT CTM simulation from the detrended monthly anomalies for 2004-2021. Straight arrows show the contemporaneous (time lag = 0) connections; curved arrows indicate lagged links (time lag > 0); edge color stands for the estimated direct causal effects. For TOMCAT, causal graphs identified from the observations were used, and direct causal effects were estimated using the TOMCAT data.

connections in the observations (Fig. 4a) since the signs of the direct causal effects align well with the expected processes outlined in the Introduction and as discussed in Sect. 4.1.

Notably, the positive lagged link in observations from w^* to N_2O indicates that an increase in residual vertical velocity intensifies N_2O transport. This enhanced transport, in turn, reduces the residence time of N_2O , leading to less time for NO_2 production via the reaction $N_2O + O(^1D) \longrightarrow NO + NO$ (also labelled B in Fig. 3a). Consequently, the causal contemporaneous link from N_2O to NO_2 reflects the negative relationship confounded by upwelling. The negative contemporaneous link from NO_2 to O_3 indicates that lower/higher NO_2 levels are associated with higher/lower O_3 concentrations, as O_3 loss in the tropical middle stratosphere is largely driven by catalytic destruction by NO_x . Causal discovery further detects a bidirected connection between w^* and O_3 in the observations, indicating the presence of a latent common driver of w^* and O_3 and that neither variable is an ancestor of the other (see Appendix C, Fig. C1a). As causal effect estimation requires a DAG, we define the direction from w^* to O_3 to quantify the strength of this link. This choice allows us to estimate the direct dynamical influence of w^* on O_3 , which is sometimes identified by a causal discovery algorithm in sensitivity tests. Based on Fig. 4, the direct influence of w^* on O_3 is much weaker, and a mediated pathway via N_2O and NO_2 dominates. However, temperature-mediated effects could amplify the apparent strength of the connection from NO_2 to O_3 , since enhanced upwelling induces both cooling

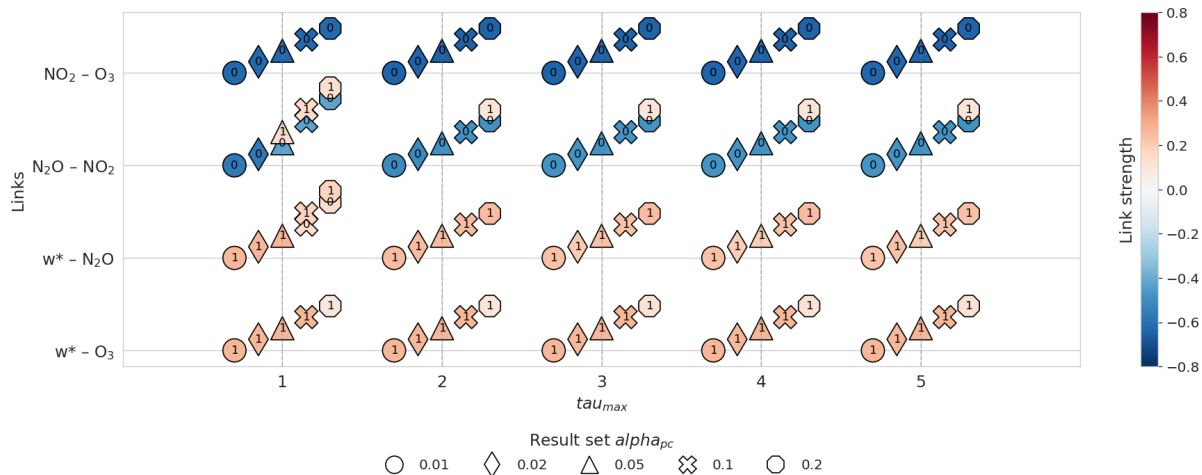


Figure 5. Sensitivity tests of the detected links to the choice of τ_{max} (x-axis) and α_{pc} (depicted with markers) for the observations for the period 2004-2021. For all tests, the minimum time delay $\tau_{min} = 0$. Only link pairs shown in Fig. 4 are considered.

300 (increasing O_3) and reduced NO_x species. To assess this, we performed additional analysis, including temperature anomalies in the tropical middle stratosphere. The inferred graph structure was not robust, likely because w^* is a derived diagnostic that depends on thermodynamic fields. Removing w^* altered the parent structure and prevented a direct comparison of direct causal effects shown in Fig. 4a. We therefore interpret the identified NO_2 -mediated pathway as the dominant mechanism, while acknowledging that temperature-related effects may potentially project onto this link.

305 Unlike the observations, in the TOMCAT CTM simulation, the causal discovery algorithm does not fully reproduce the expected chemical-dynamical coupling as outlined in the Introduction, Sect. 4.1 and as shown in Fig. 4a. In particular, the anticipated one-month lagged link from w^* to N_2O is not robustly detected. We found that this occurs because O_3 exhibits very strong contemporaneous coupling with both N_2O and NO_2 , such that conditioning on O_3 makes the one-month lagged w^* to N_2O link statistically insignificant. As further demonstrated and discussed in Appendix C, this does not indicate a lack
 310 of dynamical coupling in the TOMCAT CTM simulation. Instead, it reflects the strong shared variability among the chemical tracers as the model uses a chemically consistent scheme for all the variables. In order to still assess the strength of the processes represented in TOMCAT, we therefore did not rely on the TOMCAT-derived graph. Instead, by adopting a fixed graph structure derived from observational ground truth and expert knowledge, we can accurately estimate direct causal effects with the TOMCAT data (Fig. 4b), providing a valid and pragmatic solution for quantifying model sensitivities. Similar to
 315 observations, the direct one-month lagged w^* - N_2O connection is estimated as significant in the TOMCAT CTM simulation for the analyzed period. Additionally, the N_2O - NO_2 negative contemporaneous link is slightly stronger in the TOMCAT CTM simulation (Fig. 4b) compared to those in the observations (Fig. 4a), with a similar causal pattern observed for the NO_2 - O_3 link.

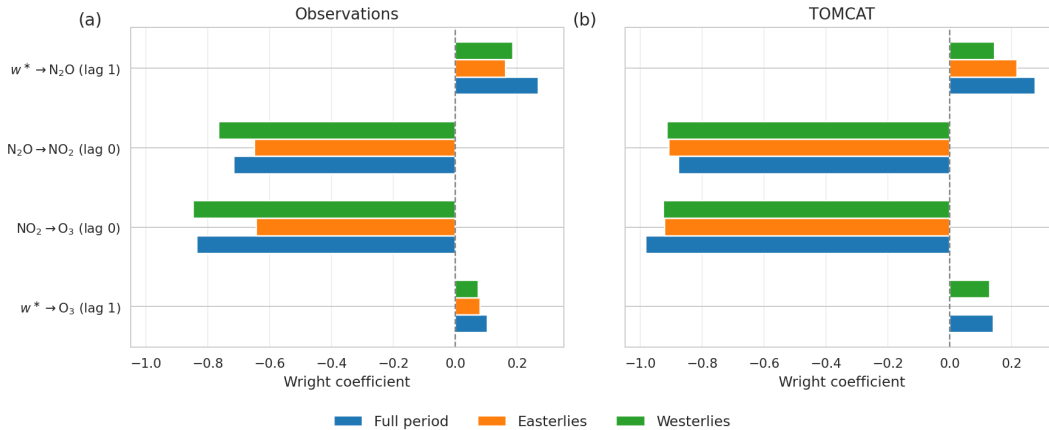


Figure 6. Regime-oriented direct causal effects from (a) observations and (b) the TOMCAT CTM simulation for the full 2004-2021 period (blue), and for easterly (orange) and westerly (green) QBO regimes.

To ensure the robustness of the connections detected in Fig.4a in observations in the tropical middle stratosphere during the
 320 period 2004-2021, Fig. 5 demonstrates the results of the application of the causal discovery algorithm with different setups
 of τ_{max} (depicted on the x-axis) and α_{pc} (depicted with markers). τ_{min} is set to zero to account for the contemporaneous
 connections. It should be noted that choosing a τ_{max} that is too low risks missing causal links with longer delays, which
 violates the assumption of causal sufficiency. However, choosing τ_{max} too high without mitigation can dilute the detection
 power of the causal algorithm. Specifically, a larger τ_{max} expands the search, as the algorithm tests more possible lagged pairs.
 325 This leads to larger conditioning sets, which can further reduce the effect size and detection power. Therefore, it is important
 to condition only on a few relevant variables that actually explain the relationship (Runge et al., 2019b). Sensitivity tests from
 Fig. 5 show very similar results for different configurations of τ_{max} and α_{pc} . The chemical connections related to N_2O - NO_2
 and NO_2 - O_3 pairs are robustly detected as contemporaneous across all experiments. The dynamical connections w^* - N_2O and
 w^* - O_3 are robustly detected with a one-month lag.

330 A further analysis of the sensitivity experiments reveals an additional feature in the N_2O - NO_2 relationship. A positive
 one-month lagged link from N_2O to NO_2 is detected across all tested τ_{max} , but primarily at relaxed significance thresholds
 ($\alpha_{pc}=0.2$). For $\tau_{max}=1$, the link is also identified at $\alpha_{pc}=0.1$ and 0.05. Such a positive lagged connection is physically plausible,
 as NO_2 is produced from N_2O , as discussed in Table 1, and a delayed response may emerge at the monthly timeseries. However,
 given its sensitivity to the choice of α_{pc} , this link cannot be considered a robust pathway.

335 4.3 Process-oriented analysis

We adapt the process-oriented causal analysis (Nowack et al., 2020; Karmouche et al., 2023; Eyring et al., 2024; Debeire et al.,
 2025) to major drivers of tropical middle stratospheric O_3 variability to further understand the robustness of the connections
 during different regimes. Given that N_2O , NO_2 , and O_3 exhibit a strong QBO signal (Chipperfield et al., 1994; Tian et al.,

2006; Park et al., 2017; Ming et al., 2025), we mask the data to easterly and westerly shear zones (see Sect. 3.4). This approach
340 makes it possible to explore the regime-dependent robustness of causal relationships. A robust connection between specific
variables indicates that the particular link is consistently detected across multiple resampled datasets. A robust connection also
suggests that the relationship is less sensitive to variations in the data, providing higher confidence that the detected connection
is not a result of random fluctuations or sampling variability. It is also important to note that the analyzed period includes an
unprecedented QBO disruption in 2016 (Tweedy et al., 2017; Match and Fueglistaler, 2021), which stalled the descent of the
345 easterly shear toward 10 hPa and caused anomalous upwelling and wind patterns, temporarily altering transport and mixing in
the tropical middle stratosphere. To assess how QBO phase affects the strength of the identified relationships, we keep the causal
graph inferred from the full period and re-estimate the link strengths separately for easterly and westerly QBO conditions. This
approach isolates regime-oriented changes in coupling strength without altering the underlying network structure. Figure 6
illustrates the estimated direct causal effects for both observations (panel a) and the TOMCAT CTM simulation (panel b) for
350 the full period 2004-2021 (blue), which also corresponds to direct causal effects shown in Fig. 4a,b, easterly (orange), and
westerly phase of the QBO.

The dynamical positive link from w^* to N_2O at a lag of one month shows slightly reduced magnitude in observations when
separating into QBO phases compared to the full period. In the TOMCAT CTM simulation, the link strength is similar to
observations and also slightly reduces across QBO phases. The negative contemporaneous chemical coupling between N_2O
355 and NO_2 is stronger in the TOMCAT CTM simulation compared to observations and of similar strength across phases. In
observations, the absolute magnitude is slightly larger during the westerly phase than during the easterly phase. The contemporaneous
negative link from NO_2 to O_3 is likewise stronger in the TOMCAT CTM simulation compared to the observations. In the observations,
this link weakens during the easterly QBO phase, whereas in the TOMCAT CTM simulation its magnitude
is comparatively stable across QBO phases.

360 The imposed one-month lagged connection from w^* to O_3 is the weakest among the analyzed pathways and substantially
smaller than the chemically mediated pathway. In the TOMCAT CTM simulation, no link is detected for the easterly QBO
phase. This phase-dependent absence of this connection points to its limited robustness. This also indicates that the direct
influence of the residual circulation on O_3 variability is predominantly mediated through N_2O and NO_2 , rather than through a
direct dynamical effect. Overall, the sign of all connections is robust across datasets and QBO phases. In the observations, the
365 chemical links (from N_2O to NO_2 and from NO_2 to O_3) tend to strengthen during the westerly QBO phase.

4.4 Causal effects across different time lags

While the causal discovery algorithm, along with causal effects estimation of the direct links, offers a general overview of
whether the methodology captures the expected dependencies, one can further study the propagation of causal effects through
the causal graph across different time lags. The total causal effects (for further discussions, see Appendix D), which are not
370 necessarily just depicted by a single arrow in the causal graph, can be derived from direct causal effects using Wright's path
analysis for specific relationships across different time lags. Figure 7 shows total (lines) and direct (labels) causal effects across
different time lags for a selection of different (X, Y) pairs of variables, where a positive connection, similar to Fig. 4, indicates

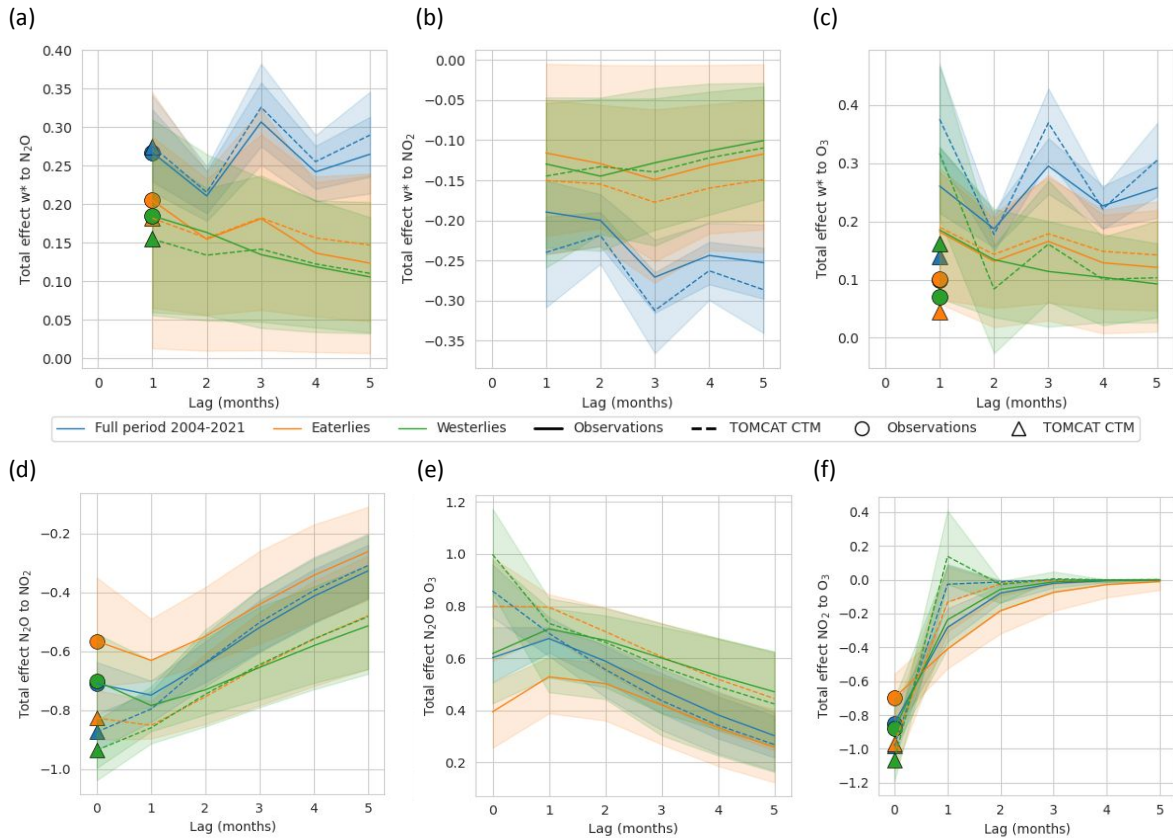


Figure 7. Total causal effects across different time lags in months from observations (solid lines) and the TOMCAT CTM simulation (dashed) for (a) w^* on N_2O , (b) w^* on NO_2 , (c) w^* on O_3 , (d) N_2O on NO_2 , (e) N_2O on O_3 , (f) NO_2 on O_3 . All plots show the results obtained for 2004–2021 (blue), easterly (orange), and westerly QBO phase (green). The markers indicate direct causal impact similar to Fig. 4 and 6. The shading corresponds to the 90% bootstrap confidence interval.

that an increase in X leads to an increase in Y after a time lag (τ). Due to the limited number of variables in the causal graphs, we do not analyze the role of mediators for specific connections, as their influence is straightforward. For example, the total causal effect from w^* to NO_2 is mediated solely by N_2O , as shown in Fig. 4. However, for the analyses that involve more complex causal graphs, similar to Fons et al. (2023), we recommend estimating the contribution of various mediators on a specific set of connections.

Figure 7 demonstrates that the total causal effects across different time lags in the TOMCAT CTM simulation (dashed lines) closely align with observations (solid lines) across analyzed regimes, indicating that the model reproduces not only the sign but also the temporal structure of the effects. However, the amplitudes in the TOMCAT CTM simulation are slightly larger, suggesting a somewhat stronger coupling between analyzed variables in the simulation. The direct causal effect of w^* on N_2O (Fig. 7a) exhibits a clear maximum at a lag of around three months for the full period 2004–2021. The total causal effects of

w^* on NO_2 (Fig. 7b) and on O_3 (Fig. 7c and Appendix D) also show similar lagged maxima for the full period 2004–2021. This lagged behavior likely indicates the spatiotemporal structure of the covariability between w^* anomalies at the analyzed
385 10 hPa level and below. Upwelling anomalies are primarily governed by the QBO and descend over time with the associated shear zones. As a result, an anomaly in w^* at 10 hPa, originating from lower altitudes, produces an instantaneous but modest local effect on composition, while over subsequent months, it persists and increasingly affects lower altitudes. This leads to a cumulative impact on N_2O , NO_2 , and consequently O_3 , with a delayed maximum in the total causal effect. The timing of this maximum is therefore linked to the vertical extent of the w^* anomaly and the rate at which QBO shear zones descend.

390 We can further decompose the total causal effect of w^* on O_3 into direct and indirect contributions. As shown by Fig. 4a and Fig. 6a, the direct one-month lagged w^* on the O_3 pathway is approximately 0.10 in the observations and 0.14 in the TOMCAT CTM simulation. The indirect contribution, mediated via N_2O and subsequently NO_2 , reaches 0.16 in the observations and 0.23 in the simulation (not shown here). Therefore, the total causal effect of w^* on O_3 shown in Fig. 7c at a lag of one month therefore represents the sum of both pathways, yielding approximately 0.26 for the observations and 0.37 for the TOMCAT
395 CTM simulation over the full analyzed period. The same analysis is performed for the easterly and westerly QBO phases, but an analogous description is not discussed further. Conditioning on both N_2O and NO_2 does not alter the magnitude of the mediated effect, indicating that the dominant indirect pathway proceeds sequentially through w^* , N_2O , NO_2 , and O_3 .

The N_2O on NO_2 relationship (Fig. 7d) strengthens gradually with increasing lag, with close agreement between observations and the TOMCAT CTM simulation after two to three months. Differences are largest at short lags, where the observational
400 estimate shows stronger variability. The observations suggest that this link appears weaker during easterly QBO shear, linked to vertical ascent due to the strengthening of tropical upwelling (Baldwin et al., 2001). The total N_2O to O_3 effect (Fig. 7e) increases from lag zero to lag one (in the observations) and then gradually decreases toward longer lags. Here, observations also show a weaker total causal effect during easterly QBO phase. These results are also consistent with direct causal effects analyzed for different QBO regimes (see Fig. 6). The NO_2 on O_3 effect (Fig. 7f) peaks at lag zero and rapidly decreases,
405 approaching zero after about one month. This behavior is consistent with fast NO_x -driven O_3 chemistry. A slightly longer persistence during the earlier subperiod suggests a more sustained influence of NO_2 on O_3 during that time.

5 Conclusions

This study applies causal inference to quantify the contributions of chemical-dynamical drivers that control O_3 variability in the tropical middle stratosphere. Using a causal discovery algorithm applied to observations over the 2004–2021 period, we
410 robustly identify a dominant chemical–dynamical pathway, in which variability in residual vertical velocity w^* modulates N_2O , subsequently affecting NO_2 and ultimately O_3 . The resulting causal graphs are then used within the causal inference framework to quantify the temporal contribution of specific variables on monthly O_3 variability under different QBO phases and to separate the direct and mediated causal effects. In the TOMCAT CTM simulation, strong shared variability among the chemical tracers limited the ability of the discovery algorithm to detect the expected dynamical coupling. Therefore, we applied
415 observational graphs based on triangulation (Uleman et al., 2024) to estimate causal relationships in the TOMCAT monthly

data. This example illustrates a practical workaround for cases where models or datasets do not yield the expected causal structure through discovery. Importantly, estimating direct or total causal effects does not require that the graph be learned by an algorithm since a graph informed by expert knowledge and supported by the literature can serve as a valid alternative (Fons et al., 2023). In our case, the close agreement between the observationally derived graph and the main relationships anticipated from established chemical-dynamical interactions (see Fig. 4 and as discussed in the Introduction) demonstrates that the discovery algorithm can indeed recover the expected structure when applied to suitable data. Additionally, the toy model validation further demonstrates reliability under finite-sample conditions (Sect. 4.1). Therefore, the methodology adopted in this study, which integrates triangulation (Uleman et al., 2024) for the construction of causal graphs and an algorithm for causal effect estimation, demonstrates a comprehensive approach to causal inference. This ensures the robustness of the analyzed system and facilitates the quantification of specific connections in a physically meaningful domain.

Direct causal effects applied to the observations and the TOMCAT CTM simulation reveal that tropical middle-stratospheric O_3 variability is dominated by an indirect NO_2 -mediated pathway, consistent with previous studies of chemical-dynamical coupling (see Fig. 4 and e.g. Portmann et al., 2012; Chipperfield et al., 2014; Galytska et al., 2019; Iglesias-Suarez et al., 2021; Prather et al., 2023; Ming et al., 2025). This mechanism of identified connections captures variability under different QBO phases via regime-oriented analysis and is also supported by sensitivity tests. For example, the direct influence of w^* on O_3 is weak and not robust across QBO phases. The total causal effects that consist of direct and mediated pathways peak at a lag of approximately two-three months (as discussed in Sect. 4.4, Fig. 7), associated with the cumulative impact of persistent, vertically coupled w^* anomalies linked to the QBO. The process-oriented analysis for different QBO phases applied to the observations suggests that the chemical links between N_2O and NO_2 , and NO_2 and O_3 strengthen during westerly shear compared to easterly shear (see Fig. 6). Additional sensitivity tests, including temperature anomalies in the tropical middle stratosphere, did not yield a robust and interpretable graph. Temperature-related effects may therefore partly project onto the identified NO_2 -mediated pathway, but within the present framework their contribution cannot be robustly separated and quantified.

Our study highlights the pivotal role causal inference can play in identifying, disentangling, and quantifying complex physical and chemical-dynamical processes in the stratosphere. This work lays the foundation for extending the application of causal inference to other areas involving complex chemical-dynamical interactions. Given its potential, this approach is particularly valuable for systems with connections that are not well understood. In the scope of the application of causal inference to stratospheric chemical-dynamical research, we emphasize the following limitation to the reader. The present analysis considers variability at a single pressure level of 10 hPa and does not explicitly resolve the vertical structure of the analyzed system. In the tropical stratosphere, where air is continuously ascending, variability at a given level is dynamically linked to conditions at lower altitudes. As a result, within such a single-level framework, the identified total lagged causal effects (discussed in Sect. 4.4) reflect a combination of local processes and effects arising from vertical coupling in the circulation. Some of the further general limitations and challenges are discussed in Runge et al. (2019a); Galytska et al. (2023).

We would like to highlight that an integration of causal reasoning into data-driven science will help us to enhance the understanding of complex processes and will support the development of robust methodologies that combine machine learning

with statistical approaches. This integration is especially relevant for Earth and environmental sciences, as it benefits both observational and modelling studies. Causal inference has already proven to be an effective tool for climate model evaluation, particularly by enabling comparisons between causal graphs derived from models and those based on observations. This emerging methodology is gaining traction in process-oriented climate model evaluation (Debeire et al., 2025; Karmouche et al., 2023; Galytska et al., 2023; Nowack et al., 2020) and offers valuable insights into the physical mechanisms driving the varying performance of different models. Notably, causal model evaluation not only identifies physically based connections but can also reveal the processes that are poorly represented in models. We also identify a key direction for future research focused on cross-evaluating the process-oriented performance of different CTMs and coupled CCMs. Such efforts could provide valuable insights into model accuracy and the reliability of CCMs, ensuring that the distribution of simulated species or processes is not driven by incorrect or unknown factors. Additionally, integrating causal inference with well-established statistical methods could further advance stratospheric studies, particularly in analyzing the variability of chemical species. Testing these approaches across different chemical and dynamical processes in various stratospheric regions represents a promising avenue for improving model evaluation and understanding complex interactions.

Code and data availability. The code used to reproduce results, including figures for this manuscript, will be accessible in Zenodo and in the following GitHub repository <https://github.com/EyringMLClimateGroup/>. The causal discovery algorithm is implemented in the Python package Tigramite is released under GNU General Public License v3.0. Tigramite v5 is publicly available on Zenodo <https://doi.org/10.5281/zenodo.6247837> Runge (2022) or via <https://github.com/jakobrunge/tigramite>, last access: 01.08.2025.

Transformed Eulerian mean data from the ERA5 reanalysis (monthly means) is available on Zenodo (<https://zenodo.org/records/7081721>, Serva, 2022). MLS N₂O v5.01 Lambert et al. (2020) is publicly available <https://disc.gsfc.nasa.gov/> upon registration. OSIRIS O₃ and NO₂ data are available at <https://research-groups.usask.ca/osiris/data-products.php#>. QBO equatorial winds are provided by the Institute of Meteorology and Climate Research at the Karlsruhe Institute of Technology (KIT) and are publicly available on Zenodo (<https://zenodo.org/records/18850668>, Kerzenmacher and Braesicke, 2026). Data from the TOMCAT CTM simulation is available upon request from the authors.

Appendix A: Glossary

Table A1 summarizes commonly used terms in this manuscript and provides brief examples illustrating their application in this study. For further acquaintance with the terminology, we refer the reader to Runge et al. (2023).

Table A1: Glossary linking causal inference terminology to stratospheric chemistry context.

Term	Formal meaning	Stratospheric example
Causal graph	Graphical model, where nodes correspond to physical quantities and edges represent direction and associated time lags of assumed or learned direct causal influences.	Residual vertical velocity w^* is used as a proxy for upwelling in the tropical middle stratosphere, influencing N_2O concentrations and thereby affecting NO_2 and O_3 . The anticipated graph is: $w^* \rightarrow \text{N}_2\text{O} \rightarrow \text{NO}_2 \rightarrow \text{O}_3$.
Causal effect	Change in one variable due to an intervention on another variable. In the linear case and with standardized time series, causal effect estimates are in units of standard deviations (but not confined to the interval $[-1, 1]$). Unlike correlation, causal effects are directional and account for confounding and mediation if the assumed graph is correct.	How will N_2O change if w^* is altered? From Fig. 4, the direct one-month lagged causal effect of w^* on N_2O is around 0.28 for both (a) the observations and (b) the TOMCAT CTM simulation. This means a 1-standard deviation change in w^* leads to a 0.28 standard deviation change in N_2O .
Connection	An open path between two variables in a causal graph according to the directional separation criterion (not necessarily causal).	Relationship between w^* and N_2O (moderate correlation from Fig. 1a) and direct one-month lagged causal connection from w^* to N_2O (e.g. Fig. 4), reflecting the influence of upwelling on tracer transport.
Dependency	A statistical association between two variables that, assuming faithfulness, may arise from a connection in the causal graph due to direct causation, indirect pathways, or common drivers. Can be instantaneous or lagged.	Lagged (partial) correlations show how dependencies decay with time (see further Appendix B) across different variable pairs.
Direct causal effect	Direct causal influence through a directed causal link, not mediated by other variables. In linear models, the direct causal effect corresponds to the regression coefficient of a predictor on a response when conditioning on its parents.	NO_2 directly affecting O_3 . For example, the direct causal effect of NO_2 on O_3 in observations for the analyzed 2004-2021 period is equal to -0.83 (see Fig. 4a).
Edge	An adjacency between two nodes. This may be directed, bidirected, or unoriented.	$w^* \rightarrow \text{N}_2\text{O}$.

Term	Formal meaning	Stratospheric example
Latent variable	Unobserved, hidden variable influencing the system.	Solar forcing is a strong common driver of many processes. If not included in the causal graph, it will act as a latent variable, causing the appearance of spurious links (Gerhardus and Runge, 2020; Runge et al., 2019a). To account for this effect, the analyzed time series should be either anomalized (see Sect. 2.3) or solar forcing can be included in the causal graph as a node.
Mediator	Variable on the causal path between two variables.	N_2O and NO_2 serve as mediators linking w^* to O_3 .
Node	A variable represented in a graph.	w^* , N_2O , NO_2 , O_3 .
Total causal effect	Causal effect across both direct and indirectly mediated causal paths between two variables. In a simple causal structure without mediators, the total causal effect is equal to the direct causal effect.	w^* effect on O_3 via all pathways: direct $w^* \rightarrow O_3$ and mediated via N_2O and NO_2 , $w^* \rightarrow N_2O \rightarrow NO_2 \rightarrow O_3$.

Appendix B: Comparison of observations and the TOMCAT CTM simulation

The detrended monthly mean anomalies in the tropical middle stratosphere from observations and the TOMCAT CTM simulation that were used for the causal inference are shown in Fig. B1. Monthly anomalies from the TOMCAT CTM simulation
480 closely align with observations.

Figure B2 depicts lagged dependencies from observations (blue) and the TOMCAT CTM simulation (orange) in the tropical middle stratosphere during 2004-2021 using the RobustParCorr class, i.e., computing lagged correlations after transforming all marginal distributions to Gaussians. Overall, the model reproduces both the sign and the magnitude of the observed relationships across most variable pairs and lags, with correlations generally decreasing in magnitude with increasing time lag. The
485 autocorrelations of the analyzed variables are mostly similar between observations and the TOMCAT CTM simulation, except for NO_2 , which exhibits slightly stronger autocorrelation in the TOMCAT CTM simulation across all time lags compared to observations. The connections between w^* and O_3 (first row, last column), N_2O and NO_2 (second row, third column), and O_3 and NO_2 (last row, third column) also show stronger lagged dependencies in the TOMCAT CTM simulation in comparison to observations.

Figure B3 depicts kernel density estimates of the joint and marginal (diagonal panels) densities of w^* , N_2O , NO_2 , and O_3 in the tropical middle stratosphere. Color coding corresponds to Fig. B2. The contours illustrate the covariance structure between the variables, and the annotated τ values indicate the lag at which the strongest association is found. The overall orientation of
490

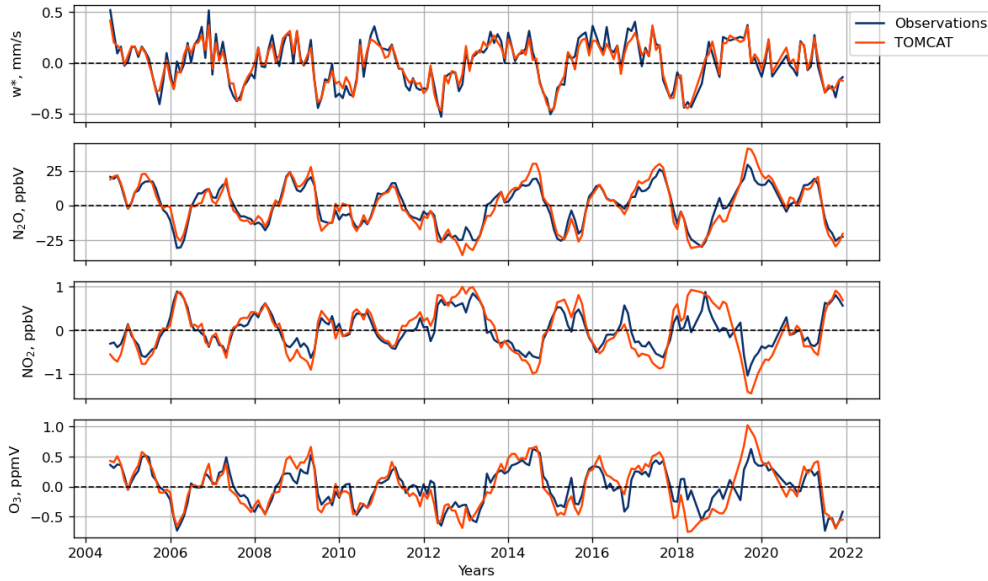


Figure B1. Monthly mean detrended anomalies in the tropical middle stratosphere during 2004-2021 in observations (blue) and the TOMCAT CTM simulation (orange). The TOMCAT CTM data is masked to match the occurrence of the observations.

the contours is similar between observations and the TOMCAT CTM simulation, suggesting that the model captures the main sign of the relationships.

495 It is important to highlight that the lagged dependencies (Fig. B2) and density plots (Fig. B3) only quantify pairwise co-variability. They do not separate direct from indirect causal effects and do not condition on other variables. Therefore, the lag at which the correlation is strongest should not be interpreted as the lag of a direct causal interaction. In a system with common dynamical forcing and persistence, pairwise correlations can reflect mixed pathways operating on different time scales.

Appendix C: Graphs inferred by the causal discovery algorithm and sensitivity testing

500 Figure C1 shows the causal graphs detected by the LPCMCI algorithm for the observations (a) and the TOMCAT CTM simulation (b) for the period 2004-2021 with $\alpha_{pc} = 0.05$ and $\tau_{max} = 2$ in. For the observations, the causal discovery algorithm does not identify the direction of the N_2O-O_3 link. In addition, the w^*-O_3 connection is detected as bidirectional, indicating the presence of a latent common driver of w^* and O_3 rather than a resolved causal direction. For the TOMCAT CTM simulation, in contrast to observations, the causal discovery algorithm does not detect the anticipated negative link from w^* to N_2O , however
 505 captures the negative contemporaneous links (without directions) between N_2O and NO_2 , and NO_2 and O_3 .

Further analysis of the output from the causal discovery from the observations revealed that the algorithm successfully identified a significant lagged causal link from w^* to N_2O , indicating that past values of w^* have a direct influence on N_2O . This relationship remained significant even after conditioning on other variables and their lags, suggesting a robust causal

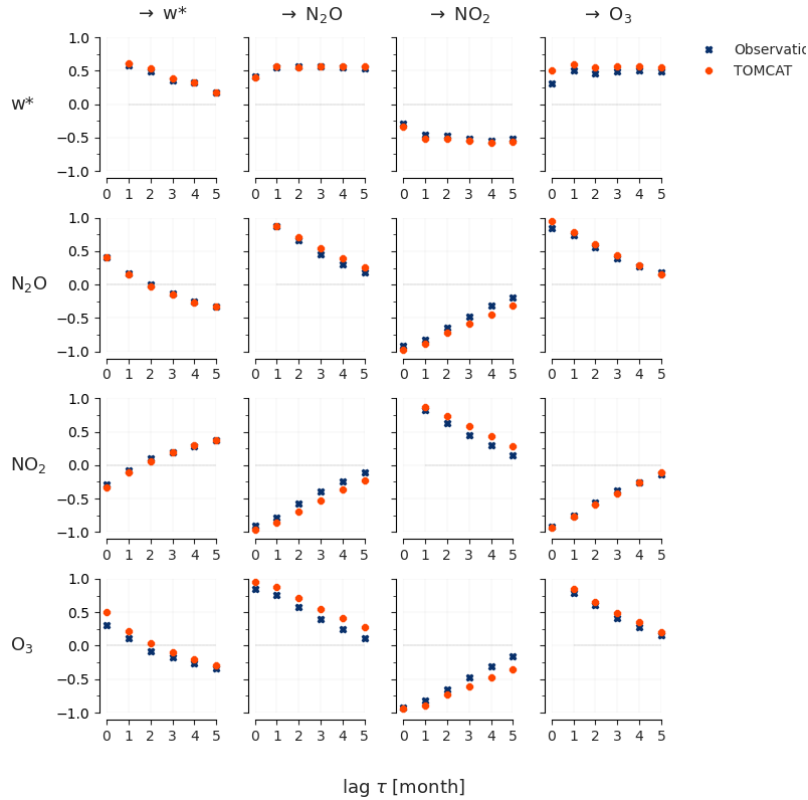


Figure B2. Lagged dependencies in the tropical middle stratosphere during 2004-2021 in observations (blue) and the TOMCAT CTM simulation (orange) based on the RobustParCorr class.

connection. In contrast, in the TOMCAT CTM simulation, the same lagged link was initially detected but became statistically
 510 insignificant once O_3 was included in the conditioning set, indicating that O_3 explains most of the shared variability between
 transport and N_2O . To further investigate the reason for the removal of one-month lagged w^* to N_2O link in the TOMCAT
 CTM simulation for the period 2004-2021, Fig. C2 shows the results of causal discovery after excluding O_3 from the variable
 set. In this reduced setup, the w^* to N_2O one-month lagged link reappeared and was statistically significant across tested
 α_{pc} . This indicates that the previously missing transport signal is not due to a lack of dynamical coupling in the TOMCAT
 515 CTM simulation, but rather to conditional independence induced by the strong covariance among the chemical variables. In
 the TOMCAT CTM simulation, O_3 has strong instantaneous coupling with both N_2O and NO_2 , and therefore accounts for a
 large part of their common variability. Consequently, once conditioning on O_3 , the remaining unique contribution of w^* to
 N_2O variability is no longer statistically detectable. The re-emergence of the w^* to N_2O link in the reduced system (without
 O_3) therefore indicates that the transport influence is present, but in the full multivariate setup it becomes statistically hidden
 520 because the strongly coupled chemical tracers (O_3 and N_2O , and O_3 and NO_2) explain most of the same variability.

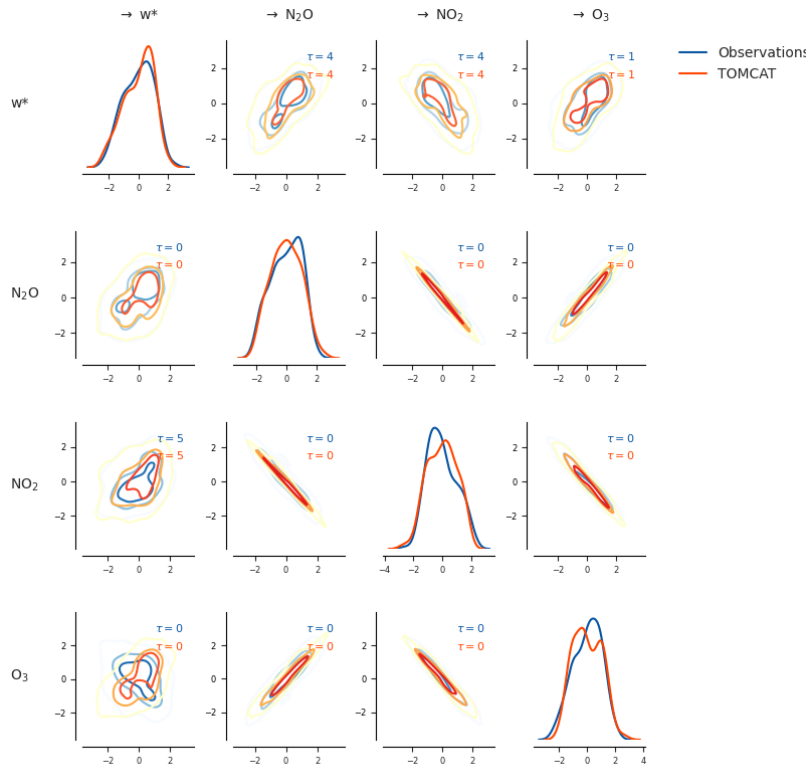


Figure B3. Density estimates of the joint and marginal densities over tropical middle stratosphere during 2004-2021 in observations (blue) and the TOMCAT CTM simulation (orange) based on the RobustParCorr class.

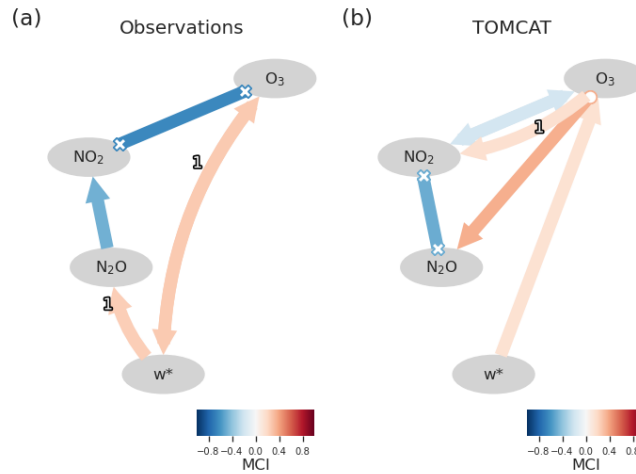


Figure C1. Causal graphs detected by the causal discovery algorithm from (a) observations and (b) the TOMCAT CTM simulation for the period 2004-2021 with significance level $\alpha_{pc}=0.05$ and $\tau_{max}=2$.

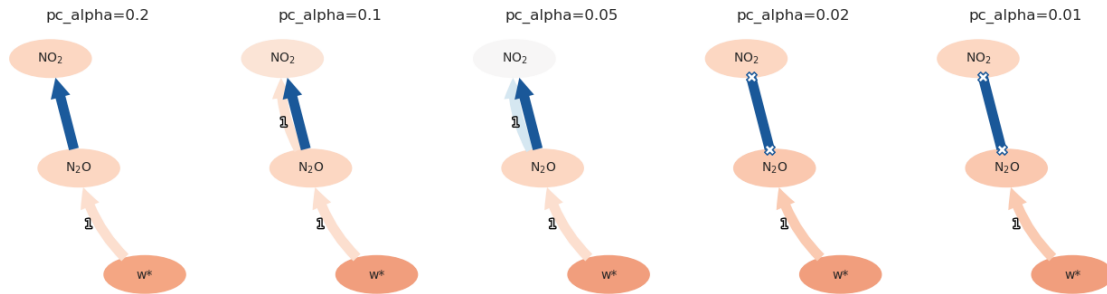


Figure C2. Causal graphs detected by the causal discovery algorithm from the TOMCAT CTM simulation for the period 2004-2021 with different significance levels α_{pc} and $\tau_{max}=2$.

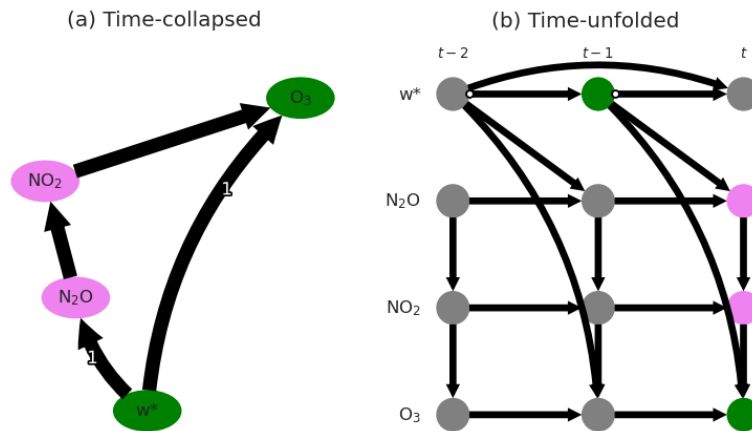


Figure D1. Example of the total causal effect of w^* on O_3 (green) as a sum of a direct lagged link and indirect via N_2O and NO_2 mediators (violet) in the time-collapsed graph (a), which corresponds to the causal graph from observations shown in Fig. 4a and time-unfolded graph (b).

Appendix D: Total causal effects

To better understand the causal effects across different time lags, Fig. D1 depicts (a) time-collapsed and (b) time-unfolded causal graphs based on the observational causal graph detected in Fig.4a in the manuscript. As an example, the total causal effect of w^* on O_3 (green nodes) consists of a direct path from w^* to O_3 and an indirect path, which is mediated by N_2O and NO_2 (violet nodes) as shown in Fig. D1a. The indirect path of the influence of w^* on O_3 is then better represented via a time series causal graph (Fig. D1b). The total causal effect across different time lags of this relationship is shown and discussed in Fig. 7c in the manuscript.

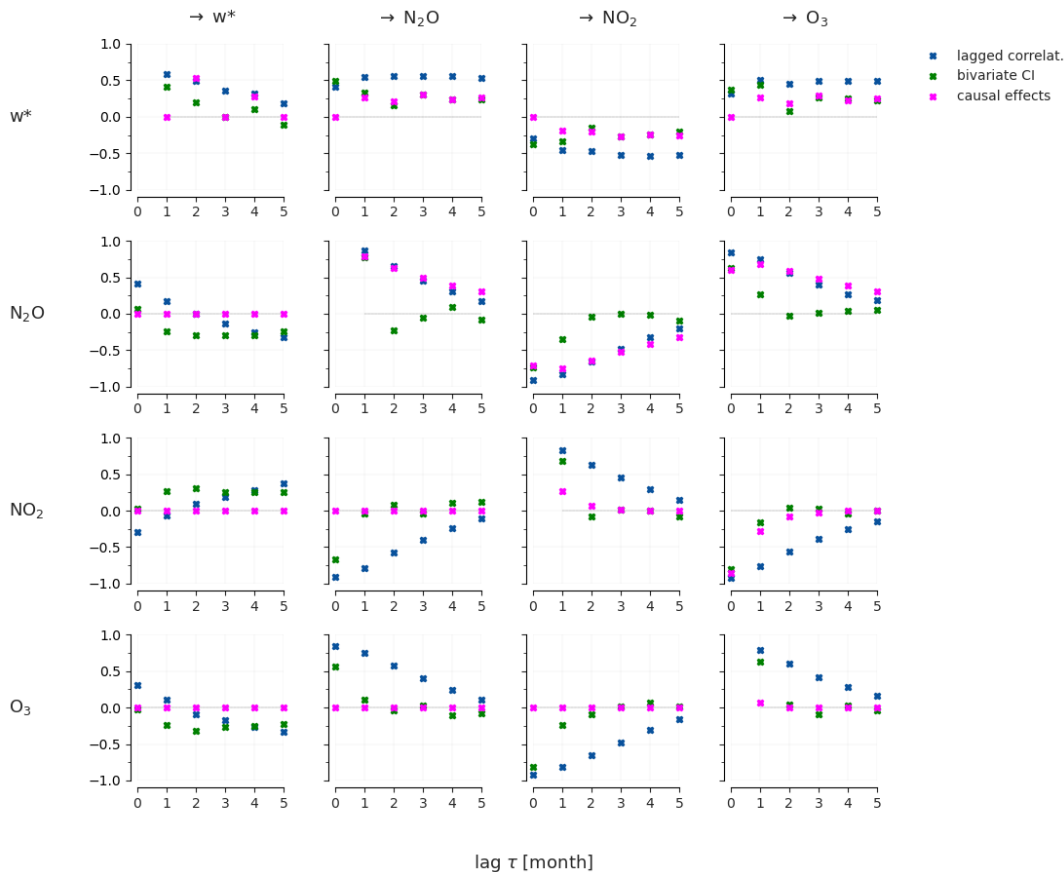


Figure E1. Lagged dependencies across different time lags (in months) from observations for the period 2004-2021 based on partial correlations (blue), bivariate partial correlations (green), and estimated Wright coefficients (magenta).

Appendix E: Dependencies vs. causal effects across different time lags

530 The lagged unconditional dependencies, such as the lagged correlations based on partial correlation, are helpful to identify the maximal time lag τ_{max} to choose in the causal discovery algorithm. However, large autocorrelation might inflate lag peaks (Runge et al., 2014). To condition out some part of the autocorrelation, the bivariate, lagged CI test (partial correlation as an example) can be applied. The comparison between the lagged dependencies from both tests is shown in Fig. E1. We also plot lagged dependencies based on the calculated Wright coefficient (labeled as 'causal effects'), which directly indicates the strength of causal relationships and takes into account the direction of causal influence (see, e.g. Fig. 4 in the manuscript). For 535 example, the causal effects in the last row in Fig. E1 are all zero since O_3 is not a cause of any analyzed variable in the system.

Author contributions. E.G. performed the analysis, prepared all figures, and led the writing of the manuscript. J.R. developed the causal discovery tool that supported this study. E.G. developed the code for data pre- and post-processing and implemented a causal inference workflow using J.R.'s package and code from Fons et al. (2023). M.P.C., S.S.D., and W.F. designed and performed the TOMCAT CTM simulation. All co-authors commented on the initial and revised drafts of the manuscripts and contributed to the interpretation of the results.

540 *Competing interests.* The authors declare no financial or non-financial competing interests

Acknowledgements. This research was funded by the Central Research Development Fund at the University of Bremen, Funding No: ZF04A/2023/FB1/Galytska Evgenia. Part of the funding for this study was provided by the European Research Council (ERC) Synergy Grant “Understanding and Modelling the Earth System with Machine Learning (USMILE)” under the Horizon 2020 research and innovation programme (Grant agreement 855187), the European Union’s Horizon 2020 research and innovation programme under Grant Agreement
545 101003536 (ESM2025—Earth System Models for the Future), and the “Advanced Earth System Model Evaluation for CMIP (EVal4CMIP)” project funded by the Helmholtz Society. J.R. has received funding from the European Research Council (ERC) Starting Grant CausalEarth under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 948112). M.P.C. and S.S.D. are supported by the NCEO TerraFIRMA, NERC LSO3 (NE/V011863/1) and ESA OREGANO (4000137112/22/I-AG) projects. This work used
550 the computational resources of the Deutsches Klimarechenzentrum (DKRZ, Germany) granted by its Scientific Steering Committee (WLA) under project ID bd1083. The TOMCAT simulation was performed on the UK Archer2 HPC system. The authors thank Veronika Eyring for her comments on the study. The authors thank the Swedish National Space Agency and the Canadian Space Agency for the continued operation and support of Odin-OSIRIS.

References

- Abalos, M., Legras, B., Ploeger, F., and Randel, W. J.: Evaluating the advective Brewer-Dobson circulation in three reanalyses for the period
555 1979–2012, *Journal of Geophysical Research: Atmospheres*, 120, 7534–7554, <https://doi.org/10.1002/2015JD023182>, 2015.
- Arosio, C., Rozanov, A., Malinina, E., Weber, M., and Burrows, J. P.: Merging of ozone profiles from SCIAMACHY, OMPS and SAGE II
observations to study stratospheric ozone changes, *Atmospheric Measurement Techniques*, 12, 2423–2444, <https://doi.org/10.5194/amt-12-2423-2019>, 2019.
- Arosio, C., Chipperfield, M. P., Rozanov, A., Weber, M., Dhomse, S., Feng, W., Jaross, G., Zhou, X., and Burrows, J. P.: Investigating Zonal
560 Asymmetries in Stratospheric Ozone Trends From Satellite Limb Observations and a Chemical Transport Model, *Journal of Geophysical Research: Atmospheres*, 129, e2023JD040353, <https://doi.org/10.1029/2023JD040353>, 2024.
- Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnarsley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, *Reviews of Geophysics*, 39, 179–229, <https://doi.org/10.1029/1999RG000073>, 2001.
- 565 Blyth, C. R.: On Simpson’s Paradox and the Sure-Thing Principle, *Journal of the American Statistical Association*, 67, 364–366, <https://doi.org/10.1080/01621459.1972.10482387>, 1972.
- Bognar, K., Tegtmeier, S., Bourassa, A., Roth, C., Warnock, T., Zawada, D., and Degenstein, D.: Stratospheric ozone trends for 1984–2021 in the SAGE II–OSIRIS–SAGE III/ISS composite dataset, *Atmospheric Chemistry and Physics*, 22, 9553–9569, <https://doi.org/10.5194/acp-22-9553-2022>, 2022.
- 570 Bourassa, A. E., Roth, C. Z., Zawada, D. J., Rieger, L. A., McLinden, C. A., and Degenstein, D. A.: Drift-corrected Odin-OSIRIS ozone product: algorithm and updated stratospheric ozone trends, *Atmospheric Measurement Techniques*, 11, 489–498, <https://doi.org/10.5194/amt-11-489-2018>, 2018.
- Brasseur, G. and Solomon, S.: *Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere*, Atmospheric and Oceanographic Sciences Library, Springer Netherlands, ISBN 9781402038242, <https://books.google.de/books?id=Z5OtlDjfxkkC>, 2005.
- 575 Camps-Valls, G., Gerhardus, A., Ninad, U., Varando, G., Martius, G., Balaguer-Ballester, E., Vinuesa, R., Diaz, E., Zanna, L., and Runge, J.: Discovering causal relations and equations from data, *Physics Reports*, 1044, 1–68, <https://doi.org/10.1016/j.physrep.2023.10.005>, discovering causal relations and equations from data, 2023.
- Carvalho-Oliveira, J., Di Capua, G., Borchert, L. F., Donner, R. V., and Baehr, J.: Causal relationships and predictability of the summer East
580 Atlantic teleconnection, *Weather and Climate Dynamics*, 5, 1561–1578, <https://doi.org/10.5194/wcd-5-1561-2024>, 2024.
- Chapman, S.: XXXV. On ozone and atomic oxygen in the upper atmosphere, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10, 369–383, <https://doi.org/10.1080/14786443009461588>, 1930.
- Chipperfield, M. P.: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments, *Quarterly Journal of the Royal Meteorological Society*, 132, 1179–1203, <https://doi.org/10.1256/qj.05.51>, 2006.
- 585 Chipperfield, M. P. and Gray, L. J.: Two-dimensional model studies of the interannual variability of trace gases in the middle atmosphere, *Journal of Geophysical Research: Atmospheres*, 97, 5963–5980, <https://doi.org/10.1029/92JD00029>, 1992.
- Chipperfield, M. P., Gray, L. J., Kinnarsley, J. S., and Zawodny, J.: A Two-Dimensional Model Study of the QBO Signal in SAGE II NO₂ and O₃, *Geophysical Research Letters*, 21, 589–592, <https://doi.org/10.1029/94GL00211>, 1994.

- Chipperfield, M. P., Liang, Q., Strahan, S. E., Morgenstern, O., Dhomse, S. S., Abraham, N. L., Archibald, A. T., Bekki, S., Braesicke, P.,
590 Di Genova, G., Fleming, E. L., Hardiman, S. C., Iachetti, D., Jackman, C. H., Kinnison, D. E., Marchand, M., Pitari, G., Pyle, J. A.,
Rozanov, E., Stenke, A., and Tummon, F.: Multimodel estimates of atmospheric lifetimes of long-lived ozone-depleting substances:
Present and future, *Journal of Geophysical Research: Atmospheres*, 119, 2555–2573, <https://doi.org/10.1002/2013JD021097>, 2014.
- Chrysanthou, A., Dubé, K., Tegtmeier, S., and Chipperfield, M. P.: Hemispheric Asymmetry in Stratospheric Trends of HCl and
Ozone: Impact of Chemical Feedback on Ozone Recovery, *Journal of Geophysical Research: Atmospheres*, 130, e2024JD042161,
595 <https://doi.org/10.1029/2024JD042161>, e2024JD042161 2024JD042161, 2025.
- Crutzen, P. J.: The influence of nitrogen oxides on the atmospheric ozone content, *Quarterly Journal of the Royal Meteorological Society*,
96, 320–325, <https://doi.org/10.1002/qj.49709640815>, 1970.
- Debeire, K., Bock, L., Nowack, P., Runge, J., and Eyring, V.: Constraining uncertainty in projected precipitation over land with causal
discovery, *Earth System Dynamics*, 16, 607–630, <https://doi.org/10.5194/esd-16-607-2025>, 2025.
- 600 Denzin, N. K.: The Fundamentals: Introducing Triangulation, [https://www.shortcutstv.com/wp-content/uploads/2021/01/
Introducing-Triangulation.pdf](https://www.shortcutstv.com/wp-content/uploads/2021/01/Introducing-Triangulation.pdf), 2010.
- Dhomse, S. S., Chipperfield, M. P., Feng, W., Hossaini, R., Mann, G. W., and Santee, M. L.: Revisiting the hemispheric asymmetry in
midlatitude ozone changes following the Mount Pinatubo eruption: A 3-D model study, *Geophysical Research Letters*, 42, 3038–3047,
<https://doi.org/https://doi.org/10.1002/2015GL063052>, 2015.
- 605 Dhomse, S. S., Chipperfield, M. P., Damadeo, R. P., Zawodny, J. M., Ball, W. T., Feng, W., Hossaini, R., Mann, G. W., and Haigh, J. D.:
On the ambiguous nature of the 11 year solar cycle signal in upper stratospheric ozone, *Geophysical Research Letters*, 43, 7241–7249,
<https://doi.org/https://doi.org/10.1002/2016GL069958>, 2016.
- Docquier, D., Vannitsem, S., Ragone, F., Wyser, K., and Liang, X. S.: Causal Links Between Arctic Sea Ice and Its Potential Drivers Based
on the Rate of Information Transfer, *Geophysical Research Letters*, 49, e2021GL095892, <https://doi.org/10.1029/2021GL095892>, 2022.
- 610 Dubé, K., Zawada, D., Bourassa, A., Degenstein, D., Randel, W., Flittner, D., Sheese, P., and Walker, K.: An improved OSIRIS NO₂ profile
retrieval in the upper troposphere–lower stratosphere and intercomparison with ACE-FTS and SAGE III/ISS, *Atmospheric Measurement
Techniques*, 15, 6163–6180, <https://doi.org/10.5194/amt-15-6163-2022>, 2022.
- Dubé, K., Randel, W., Bourassa, A., Zawada, D., McLinden, C., and Degenstein, D.: Trends and Variability in Stratospheric NO_x De-
rived From Merged SAGE II and OSIRIS Satellite Observations, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031798,
615 <https://doi.org/10.1029/2019JD031798>, 2020.
- Eckert, E., von Clarman, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S.,
Leblanc, T., McDermid, S., Pastel, M., Steinbrecht, W., Swart, D. P. J., Walker, K. A., and Bernath, P. F.: Drift-corrected trends and periodic
variations in MIPAS IMK/IAA ozone measurements, *Atmospheric Chemistry and Physics*, 14, 2571–2589, [https://doi.org/10.5194/acp-
14-2571-2014](https://doi.org/10.5194/acp-14-2571-2014), 2014.
- 620 Eyring, V., Collins, W. D., Gentine, P., Barnes, E. A., Barreiro, M., Beucler, T., Bocquet, M., Bretherton, C. S., Christensen, H. M.,
Dagon, K., et al.: Pushing the frontiers in climate modelling and analysis with machine learning, *Nature Climate Change*, 14, 916–928,
<https://doi.org/10.1038/s41558-024-02095-y>, 2024.
- Fons, E., Runge, J., Neubauer, D., and Lohmann, U.: Stratocumulus adjustments to aerosol perturbations disentangled with a causal approach,
npj Climate and Atmospheric Science, 6, 130, <https://doi.org/10.1038/s41612-023-00452-w>, 2023.
- 625 Galytska, E.: Spatio-temporal variations of observed and modelled stratospheric trace gases, Ph.D. thesis, Universität Bremen, <https://nbn-resolving.de/urn:nbn:de:gbv:46-00107599-10>, 2019.

- Galytska, E., Rozanov, A., Chipperfield, M. P., Dhomse, Weber, M., Arosio, C., Feng, W., and Burrows, J. P.: Dynamically controlled ozone decline in the tropical mid-stratosphere observed by SCIAMACHY, *Atmospheric Chemistry and Physics*, 19, 767–783, <https://doi.org/10.5194/acp-19-767-2019>, 2019.
- 630 Galytska, E., Weigel, K., Handorf, D., Jaiser, R., Köhler, R., Runge, J., and Eyring, V.: Evaluating Causal Arctic-Midlatitude Teleconnections in CMIP6, *Journal of Geophysical Research: Atmospheres*, 128, e2022JD037978, <https://doi.org/10.1029/2022JD037978>, 2023.
- Gebhardt, C., Rozanov, A., Hommel, R., Weber, M., Bovensmann, H., Burrows, J. P., Degenstein, D., Froidevaux, L., and Thompson, A. M.: Stratospheric ozone trends and variability as seen by SCIAMACHY from 2002 to 2012, *Atmospheric Chemistry and Physics*, 14, 831–846, <https://doi.org/10.5194/acp-14-831-2014>, 2014.
- 635 Gerhardus, A. and Runge, J.: High-recall causal discovery for autocorrelated time series with latent confounders, in: *Advances in Neural Information Processing Systems*, edited by Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., vol. 33, pp. 12 615–12 625, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2020/file/94e70705efae423efda1088614128d0b-Paper.pdf, 2020.
- Godin-Beekmann, S., Azouz, N., Sofieva, V. F., Hubert, D., Petropavlovskikh, I., Effertz, P., Ancellet, G., Degenstein, D. A., Zawada, D., Froidevaux, L., Frith, S., Wild, J., Davis, S., Steinbrecht, W., Leblanc, T., Querel, R., Tourpali, K., Damadeo, R., Maillard Barras, E., Stübi, R., Vigouroux, C., Arosio, C., Nedoluha, G., Boyd, I., Van Malderen, R., Mahieu, E., Smale, D., and Sussmann, R.: Updated trends of the stratospheric ozone vertical distribution in the 60° S–60° N latitude range based on the LOTUS regression model, *Atmospheric Chemistry and Physics*, 22, 11 657–11 673, <https://doi.org/10.5194/acp-22-11657-2022>, 2022.
- 640 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- 645 Hitchcock, P. and Ming, A.: The Role of Ozone in the Secondary Circulation of the QBO: Linear Theory, *Journal of Geophysical Research: Atmospheres*, 130, e2025JD044766, <https://doi.org/https://doi.org/10.1029/2025JD044766>, e2025JD044766 2025JD044766, 2025.
- Iglesias-Suarez, F., Wild, O., Kinnison, D. E., Garcia, R. R., Marsh, D. R., Lamarque, J.-F., Ryan, E. M., Davis, S. M., Eichinger, R., Saiz-Lopez, A., and Young, P. J.: Tropical Stratospheric Circulation and Ozone Coupled to Pacific Multi-Decadal Variability, *Geophysical Research Letters*, 48, e2020GL092162, <https://doi.org/10.1029/2020GL092162>, 2021.
- 655 Karmouche, S., Galytska, E., Runge, J., Meehl, G. A., Phillips, A. S., Weigel, K., and Eyring, V.: Regime-oriented causal model evaluation of Atlantic–Pacific teleconnections in CMIP6, *Earth System Dynamics*, 14, 309–344, <https://doi.org/10.5194/esd-14-309-2023>, 2023.
- Kerzenmacher, T. and Braesicke, P.: QBO: monthly zonal stratospheric winds from tropical radiosonde data (mainly Singapore), <https://doi.org/10.5281/zenodo.18472673>, 2026.
- Kretschmer, M., Zappa, G., and Shepherd, T. G.: The role of Barents–Kara sea ice loss in projected polar vortex changes, *Weather and Climate Dynamics*, 1, 715–730, <https://doi.org/10.5194/wcd-1-715-2020>, 2020.
- 660 Krich, C., Runge, J., Miralles, D. G., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D.: Estimating causal networks in biosphere–atmosphere interaction with the PCMCi approach, *Biogeosciences*, 17, 1033–1061, <https://doi.org/10.5194/bg-17-1033-2020>, 2020.

- Kyrölä, E., Laine, M., Sofieva, V., Tamminen, J., Päivärinta, S.-M., Tukiainen, S., Zawodny, J., and Thomason, L.: Combined SAGE
 665 II–GOMOS ozone profile data set for 1984–2011 and trend analysis of the vertical distribution of ozone, *Atmospheric Chemistry and
 Physics*, 13, 10 645–10 658, <https://doi.org/10.5194/acp-13-10645-2013>, 2013.
- Lambert, A., Livesey, N., Read, W., and Fuller, R.: MLS/Aura Level 3 Monthly Binned Nitrous Oxide (N₂O) Mixing Ratio on
 Assorted Grids V005, https://disc.gsfc.nasa.gov/datasets/ML2N2O_003/summary, <https://doi.org/10.5067/Aura/MLS/DATA/3545>, ac-
 cessed: [05.03.2025], 2020.
- 670 Li, Y., Dhomse, S. S., Chipperfield, M. P., Feng, W., Bian, J., Xia, Y., and Guo, D.: Quantifying stratospheric ozone trends over 1984–
 2020: a comparison of ordinary and regularized multivariate regression models, *Atmospheric Chemistry and Physics*, 23, 13 029–13 047,
<https://doi.org/10.5194/acp-23-13029-2023>, 2023.
- Llewellyn, E. J., Lloyd, N. D., Degenstein, D. A., Gattinger, R. L., Petelina, S. V., Bourassa, A. E., Wiensz, J. T., Ivanov, E. V., McDade,
 I. C., Solheim, B. H., McConnell, J. C., Haley, C. S., von Savigny, C., Sioris, C. E., McLinden, C. A., Griffioen, E., Kaminski, J., Evans,
 675 W. F., Puckrin, E., Strong, K., Wehrle, V., Hum, R. H., Kendall, D. J., Matsushita, J., Murtagh, D. P., Brohede, S., Stegman, J., Witt, G.,
 Barnes, G., Payne, W. F., Piché, L., Smith, K., Warshaw, G., Deslauniers, D. L., Marchand, P., Richardson, E. H., King, R. A., Wevers,
 I., McCreath, W., Kyrölä, E., Oikarinen, L., Leppelmeier, G. W., Auvinen, H., Mégie, G., Hauchecorne, A., Lefèvre, F., de La Nöe, J.,
 Ricaud, P., Frisk, U., Sjöberg, F., von Schéele, F., and Nordh, L.: The OSIRIS instrument on the Odin spacecraft, *Canadian Journal of
 Physics*, 82, 411–422, <https://doi.org/10.1139/p04-005>, 2004.
- 680 Match, A. and Fueglistaler, S.: Anomalous Dynamics of QBO Disruptions Explained by 1D Theory with External Triggering, *Journal of the
 Atmospheric Sciences*, 78, 373 – 383, <https://doi.org/10.1175/JAS-D-20-0172.1>, 2021.
- Match, A., Gerber, E. P., and Fueglistaler, S.: Protection without poison: why tropical ozone maximizes in the interior of the atmosphere,
Atmospheric Chemistry and Physics, 25, 4349–4366, <https://doi.org/10.5194/acp-25-4349-2025>, 2025.
- Ming, A., Hitchcock, P., Orbe, C., and Dubé, K.: Phase and Amplitude Relationships Between Ozone, Temperature,
 685 and Circulation in the Quasi-Biennial Oscillation, *Journal of Geophysical Research: Atmospheres*, 130, e2024JD042 469,
<https://doi.org/https://doi.org/10.1029/2024JD042469>, e2024JD042469 2024JD042469, 2025.
- Minganti, D., Chabrilat, S., Christophe, Y., Errera, Q., Abalos, M., Prignon, M., Kinnison, D. E., and Mahieu, E.: Climatological impact of
 the Brewer–Dobson circulation on the N₂O budget in WACCM, a chemical reanalysis and a CTM driven by four dynamical reanalyses,
Atmospheric Chemistry and Physics, 20, 12 609–12 631, <https://doi.org/10.5194/acp-20-12609-2020>, 2020.
- 690 Murtagh, D., Frisk, U., Merino, F., Ridal, M., Jonsson, A., Stegman, J., Witt, G., Eriksson, P., Jiménez, C., Megie, G., Noë, J. d. I., Ricaud,
 P., Baron, P., Pardo, J. R., Hauchcorne, A., Llewellyn, E. J., Degenstein, D. A., Gattinger, R. L., Lloyd, N. D., Evans, W. F., McDade,
 I. C., Haley, C. S., Sioris, C., Savigny, C. v., Solheim, B. H., McConnell, J. C., Strong, K., Richardson, E. H., Leppelmeier, G. W.,
 Kyrölä, E., Auvinen, H., and Oikarinen, L.: An overview of the Odin atmospheric mission, *Canadian Journal of Physics*, 80, 309–319,
<https://doi.org/10.1139/p01-157>, 2002.
- 695 Nedoluha, G. E., Siskind, D. E., Lambert, A., and Boone, C.: The decrease in mid-stratospheric tropical ozone since 1991, *Atmospheric
 Chemistry and Physics*, 15, 4215–4224, <https://doi.org/10.5194/acp-15-4215-2015>, 2015.
- Nowack, P., Runge, J., Eyring, V., and Haigh, J. D.: Causal networks for climate model evaluation and constrained projections, *Nature
 communications*, 11, 1–11, <https://doi.org/10.1038/s41467-020-15195-y>, 2020.
- Orbe, C., Ming, A., Chiodo, G., Prather, M., Diallo, M., Tang, Q., Chrysanthou, A., Naoe, H., Zhou, X., Thaler, I., Elsbury, D., Bednarz,
 700 E., Wright, J. S., Match, A., Watanabe, S., Anstey, J., Kerzenmacher, T., Versick, S., Marchand, M., Li, F., and Keeble, J.: Experimental

- protocol for phase 1 of the APARC QUOCA (QUasi-biennial oscillation and Ozone Chemistry interactions in the Atmosphere) working group, Geoscientific Model Development, 19, 773–794, <https://doi.org/10.5194/gmd-19-773-2026>, 2026.
- 705 Park, M., Randel, W. J., Kinnison, D. E., Bourassa, A. E., Degenstein, D. A., Roth, C. Z., McLinden, C. A., Stioris, C. E., Livesey, N. J., and Santee, M. L.: Variability of Stratospheric Reactive Nitrogen and Ozone Related to the QBO, *Journal of Geophysical Research: Atmospheres*, 122, 10,103–10,118, <https://doi.org/10.1002/2017JD027061>, 2017.
- Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B., and Sofieva, V.: SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere, Tech. rep., <https://doi.org/10.17874/f899e57a20b>, 9th assessment report of the SPARC project, published by the International Project Office at DLR-IPA. also: GAW Report No. 241; WCRP Report 17/2018, 2019.
- 710 Polkova, I., Afargan-Gerstman, H., Domeisen, D. I. V., King, M. P., Ruggieri, P., Athanasiadis, P., Dobrynin, M., Aarnes, , Kretschmer, M., and Baehr, J.: Predictors and prediction skill for marine cold-air outbreaks over the Barents Sea, *Quarterly Journal of the Royal Meteorological Society*, 147, 2638–2656, <https://doi.org/10.1002/qj.4038>, 2021.
- Portmann, R. W., Daniel, J. S., and Ravishankara, A. R.: Stratospheric ozone depletion due to nitrous oxide: influences of other gases, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1256–1264, <https://doi.org/10.1098/rstb.2011.0377>, 2012.
- 715 Prather, M. J., Froidevaux, L., and Livesey, N. J.: Observed changes in stratospheric circulation: decreasing lifetime of N₂O, 2005–2021, *Atmospheric Chemistry and Physics*, 23, 843–849, <https://doi.org/10.5194/acp-23-843-2023>, 2023.
- Ricard, L., Falasca, F., Runge, J., and Nenes, A.: network-based constraint to evaluate climate sensitivity, *Nature Communications*, 15, 6942, <https://doi.org/10.1038/s41467-024-50813-z>, 2024.
- Runge, J.: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets, in: Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), edited by Peters, J. and Sontag, D., vol. 124 of *Proceedings of Machine*
- 720 *Learning Research*, pp. 1388–1397, PMLR, <https://doi.org/10.48550/arXiv.2003.03685>, 2020.
- Runge, J.: jakobrunge/tigramite: Tigramite 5.0, <https://doi.org/10.5281/zenodo.6247837>, 2022.
- Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models, *Journal of climate*, 27, 720–739, <https://doi.org/10.1175/JCLI-D-13-00159.1>, 2014.
- 725 Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., et al.: Inferring causation from time series in Earth system sciences, *Nature communications*, 10, 2553, 2019a.
- Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D.: Detecting and quantifying causal associations in large nonlinear time series datasets, *Science Advances*, 5, eaau4996, <https://doi.org/10.1126/sciadv.aau4996>, 2019b.
- Runge, J., Gerhardus, A., Varando, G., Eyring, V., and Camps-Valls, G.: Causal inference for time series, *Nature Reviews Earth & Environment*, 4, 487–505, <https://doi.org/10.1038/s43017-023-00431-y>, 2023.
- 730 Serva, F.: Transformed Eulerian mean data from the ERA5 reanalysis (monthly means), <https://doi.org/10.5281/zenodo.7081721>, 2022.
- Serva, F.: Transformed Eulerian mean diagnostics (tem-diag), <https://doi.org/10.5281/zenodo.10180386>, 2023.
- Serva, F., Christiansen, B., Davini, P., von Hardenberg, J., van den Oord, G., Reerink, T. J., Wyser, K., and Yang, S.: Changes in Stratospheric Dynamics Simulated by the EC-Earth Model From CMIP5 to CMIP6, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003756, <https://doi.org/10.1029/2023MS003756>, e2023MS003756 2023MS003756, 2024.
- 735 Sofieva, V. F., Szelağ, M., Tamminen, J., Kyrölä, E., Degenstein, D., Roth, C., Zawada, D., Rozanov, A., Arosio, C., Burrows, J. P., Weber, M., Laeng, A., Stiller, G. P., von Clarmann, T., Froidevaux, L., Livesey, N., van Roozendaal, M., and Retscher, C.: Measurement report: regional trends of stratospheric ozone evaluated using the MErged GRIdded Dataset of Ozone Profiles (MEGRIDOP), *Atmospheric Chemistry and Physics*, 21, 6707–6720, <https://doi.org/10.5194/acp-21-6707-2021>, 2021.

- Spirtes, P.: Directed cyclic graphical representations of feedback models, in: Proceedings of the Eleventh Conference on Uncertainty in
740 Artificial Intelligence, UAI'95, p. 491–498, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, ISBN 1558603859, 1995.
- Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D.: Causation, prediction, and search, MIT press, 2000.
- Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J.,
Frith, S., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber,
745 M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke,
K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N.,
Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F.: An update
on ozone profile trends for the period 2000 to 2016, *Atmospheric Chemistry and Physics*, 17, 10 675–10 690, <https://doi.org/10.5194/acp-17-10675-2017>, 2017.
- Szeląg, M. E., Sofieva, V. F., Degenstein, D., Roth, C., Davis, S., and Froidevaux, L.: Seasonal stratospheric ozone trends over 2000–2018
750 derived from several merged data sets, *Atmospheric Chemistry and Physics*, 20, 7035–7047, <https://doi.org/10.5194/acp-20-7035-2020>,
2020.
- Tian, W., Chipperfield, M. P., Gray, L. J., and Zawodny, J. M.: Quasi-biennial oscillation and tracer distributions in a coupled chemistry-
climate model, *Journal of Geophysical Research: Atmospheres*, 111, 2006.
- Tibau, X.-A., Reimers, C., Gerhardus, A., Denzler, J., Eyring, V., and Runge, J.: A spatiotemporal stochastic climate model for benchmarking
755 causal discovery methods for teleconnections, *Environmental Data Science*, 1, e12, <https://doi.org/10.1017/eds.2022.11>, 2022.
- Tweedy, O. V., Kramarova, N. A., Strahan, S. E., Newman, P. A., Coy, L., Randel, W. J., Park, M., Waugh, D. W., and Frith, S. M.:
Response of trace gases to the disrupted 2015–2016 quasi-biennial oscillation, *Atmospheric Chemistry and Physics*, 17, 6813–6823,
<https://doi.org/10.5194/acp-17-6813-2017>, 2017.
- Uleman, J. F., Luijten, M., Abdo, W. F., Vyrastekova, J., Gerhardus, A., Runge, J., Rod, N. H., and Verhagen, M.: Triangulation for causal
760 loop diagrams: constructing biopsychosocial models using group model building, literature review, and causal discovery, *npj Complexity*,
1, 19, 2024.
- WMO: Scientific Assessment of Ozone Depletion: 2022, Tech. Rep. 278, World Meteorological Organization, Geneva, 2022.
- Wright, S.: Correlation and causation, *Journal of agricultural research*, 20, 557, 1921.
- Zerefos, C., Fountoulakis, I., Eleftheratos, K., and Kazantzidis, A.: Long-term variability of human health-related solar ultraviolet-B radiation
765 doses from the 1980s to the end of the 21st century, *Physiological Reviews*, 103, 1789–1826, <https://doi.org/10.1152/physrev.00031.2022>,
pMID: 36787480, 2023.