https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


# Spatio-temporal pattern analysis of MOPITT total column CO using varimax rotation and singular spectrum analysis

John McKinnon[1], Chayan Roychoudhury[1], and Avelino Arellano, Jr.[1]

[1]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

**Correspondence:** John McKinnon (jmckinnon1@arizona.edu)

**Abstract.** In this paper, we apply varimax empirical orthogonal function (EOF) analysis to MOPITT total column CO retrievals to answer the question of whether or not it is possible to disentangle the dominant CO sources associated with inferred modes of variability at regional to global scales. Additionally, we write this manuscript with the intention to highlight the strengths and weaknesses of EOF analysis, specifically their usage in the field of atmospheric chemistry. We analyzed daily MOPITT Version 8 Level 3 joint (TIR-NIR) products from 2005 to 2018, aggregated every 8 days on a 1° by 1° grid. Our findings show that EOF patterns of MOPITT CO are consistent across various regional subdomains, demonstrating that these spatial patterns are independent of the chosen domain. A comparison of the eigenvalue spectrum reveals that unrotated EOF analysis yields three distinct modes, while varimax rotation reduces these to two. The power spectra of the principal components indicate that the first two unrotated modes are primarily driven by annual and semi-annual cycles, while the third mode reflects seasonal variations occurring over roughly three months. To further isolate these modes, we employed singular spectrum analysis (SSA) at each grid point to generate long-term, seasonal, and residual EOF patterns. The power spectrum analysis of the principal components shows that the long-term EOFs replicate the original two dominant modes, while the seasonal EOFs reveal significant variations over 2 to 3 months, and the residual modes exhibit time scales of 2 months or shorter. By plotting the mean skewness field we show the dataset is non-Gaussian, leading us to conclude each principal component is time-dependent despite being uncorrelated. The periodic decay observed in the temporal auto-correlation function for each time series suggests a classification of wide-sense cyclostationary or wide-sense polycyclostationary behavior. We find the non-stationarity of each time series together with the temporal dependence of modes leads us to conclude that EOF analysis alone cannot fully disentangle individual CO sources. Consequently, we recommend exploring cyclostationary EOF analysis in future studies. These limitations must be carefully considered when interpreting EOF patterns of other composition datasets with similar characteristics.

## 1 Introduction

### 1.1 The Role of Carbon Monoxide in Atmospheric Composition

Carbon monoxide (CO) plays an important role in tropospheric chemistry and composition. Firstly, its reaction with the hydroxyl radical (OH) has a significant impact on the atmospheric oxidizing capacity. CO as well as non-methane volatile organic

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


compounds (NMVOCs) are both responsible for positive climate forcing because they raise tropospheric ozone ($O_3$) and $CH_4$ by creating a global sink for the hydroxyl radical OH (Fiore et al. (2012)). The spatiotemporal variations in CO (and hence OH) can lead to changes in $O_3$ abundance depending on whether the concentration of nitrogen oxides ($NO_x$) is low or high. Regions of high $NO_x$ experience ozone production while clean areas with low $NO_x$ experience the destruction of ozone (Holloway et al. (2000)). Because OH is a photochemical sink of CO through oxidation, its global distribution determines the lifetime of CO and therefore values of CO are lowest with the shortest lifetime in regions with the highest OH. Secondly, CO can also serve as a tracer of pollution transport given its medium length lifetime (weeks to months) (Cicerone, 1988). Mainly a product of incomplete combustion of wood and fossil fuels, CO is co-emitted with nitrogen dioxide ($NO_2$), black carbon, and carbon dioxide ($CO_2$). It is more concentrated in the atmospheric boundary layer (with less in the stratosphere) so observing CO can be useful as well to study the exchange of these gases and aerosols across the troposphere. Thirdly, as discussed by Gaubert et al. (2020), because CO is the main sink of OH it is important for quantifying losses of methane in the troposphere as well (Myhre and Shindell (2013), Gaubert et al. (2016) Gaubert et al. (2017), Nguyen et al. (2020)). The lack of constraints on OH spatiotemporal variability as well as the interannual variability of $CH_4$ have both made it difficult to accurately determine the global $CH_4$ budget (Saunois et al. (2016), Prather and Holmes (2017), Turner et al. (2019). This means there is a need to reduce the uncertainty of primary drivers of OH including CO, ozone, water vapor, nitrogen oxides, and non-methane volatile organic compounds (Gaubert et al. (2020)).

Determining the spatial and temporal distribution of CO, along with the distribution of co-emitted species including $NO_2$, $CO_2$, $CH_4$ and aerosols helps us to understand trends of combustion, biomass burning, and other sources of air pollution. For example Lin et al. (2020b) uses a combined mean and standard deviation approach for total column CO to classify local regions of Southeast and East Asia based on combustion activity into areas of intense urbanization, large-scale biomass burning, regions undergoing significant change from one type to another, and clean regions. Raman and Arellano (2017) utilized variations in enhancement ratios of elemental carbon (EC), CO, and $NO_2$ to investigate processes driving EC levels across the United States. Tang and Arellano Jr. (2017) demonstrated that local enhancement ratios of $\Delta CO/\Delta NO_2$ can distinguish different combustion types, identifying whether or not a fire was in a flaming phase or smoldering phase while Silva et al. (2013) used $\Delta CO/\Delta CO_2$ to identify combustion efficiency patterns in urban agglomeration. Building on this, Silva and Arellano (2017) combined $\Delta CO/\Delta CO_2$ and $\Delta CO/\Delta NO_2$ to better characterize regional-scale combustion process while Tang et al. (2019) used $\Delta CO/\Delta NO_2$ and $\Delta CO/\Delta SO_2$ to track the common combustion emission pathways across China. Recently, Mottungan et al. (2024) demonstrated the value of multi-species enhancement ratios (CO, $CH_4$, $CO_2$) along with multivariate regression methods to differentiate between regional and local influences, and identify the dominant processes driving these patterns.

However, determining accurate trends in air pollution is difficult in practice for many different reasons. While there has been an increasing availability of satellite measurements of trace gases and aerosols in the recent decade, most of these remote sensing retrievals cannot in general tell us where their original source or sink was due to the diffusive nature of atmospheric transport (Murayama et al., 2004). Another challenge is that space-based CO measurements often lack accuracy in areas with persistent cloud cover and near the surface, where thermal infrared sensors struggle due to interference from surface radiation and limited retrieval sensitivity. These significant sources of uncertainty require us to cross-validate satellites across different

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


instruments as well as with ground and aircraft measurements during field campaigns and is especially problematic over areas such as Sub-Saharan Africa where there are very few air quality monitoring stations. Determining the background value of constituents is especially difficult on a global scale because it changes across time and space based on the sources and sinks for a given region and chemical regime. Because satellites measure the total amount of a constituent rather than the source of emission for a given region, it is difficult to tell whether local changes in the overall trend are due to anthropogenic or biogenic sources or the background or residual burden. As discussed in Buchholz et al. (2021) there are no overall trends in either CO or aerosol optical depth (AOD) in Central/Southern Africa despite the increasing trend in anthropogenic combustion in Central Africa as determined by Zheng et al. (2019) which may be counteracting the global decrease in CO. Buchholz et al. (2021) have also noted that the increase in burning in the region as shown in Andela et al. (2017) may be potentially counteracting transport.

There are also a number of complex challenges when determining emission inventories. As discussed by Gaubert et al. (2020), emission inventories in China are low in comparison to in situ observations in the case of both forward and inverse models, and in particular there is an underestimation of CO despite having revised inventories (Kong et al. (2020), Feng et al. (2020), Li et al. (2017)) as explained by an underestimation of residential coal combustion from either heating or cooling (Chen et al. (2017), Cheng et al. (2017), Zhi et al. (2017)). Even though this underestimation of fossil fuel burning does explain the underestimation of Northern Hemisphere (NH) CO in global models (Shindell et al. (2006)) this is confounded by the large difference in modeled variability of the regional distributions of OH in the NH which could offer an alternative explanation (Naik et al. (2013) Young (2013)). Gaubert et al. (2020) used observations from the Korea-United States Air Quality (KORUS-AQ) field campaign to investigate this underestimate over East Asia and found that assimilation of CO improves the representation of OH in global chemical transport models by correcting $OH/HO_2$ partitioning. This also implies that assimilation of CO should be helpful in studying coupled $CH_4$-CO-OH reactions which allow for chemical feedback. Effectively distinguishing the signature of secondary CO from the influence of surface emissions—whether anthropogenic, biogenic, or from fires— is crucial for enhancing the reliability of source attribution studies, including inverse modeling for air quality assessments and chemistry-climate simulations. This separation also aids in the physical interpretation of observed air pollution patterns, with important implications for understanding the distributions of both OH and $CH_4$. The primary questions that we are interested in this study include the following:

1. How can we use the spatial and temporal distribution of CO to detect the difference between anthropogenic and natural variability of CO?

2. Can the global background of CO be disentangled from the local-to-regional influences to observed CO abundance?

To attempt an answer to these questions, we will use an empirical orthogonal function (EOF) analysis, which is a method in pattern analysis to determine the dominant spatial and temporal patterns of a dataset and is described in full detail in the next section on Methods. It has been extensively utilized in climatology, a historical use being from Wallace and Gutzler (1981) who used EOF analysis of geopotential height fields to produce EOF patterns that we now refer to as teleconnections. It has had fewer applications to atmospheric chemistry, but for instance Li et al. (2009) use varimax rotated EOF analysis of

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


de-seasonalized aerosol index in order to isolate major dust and biomass burning regions and dust transport. Li et al. (2013) performs a similar analysis on AOD data. Lin et al. (2020a) verifies the result of their classification of urban and biomass burning regions using EOF analysis. Yin et al. (2019) used EOF analysis to analyze the dominant patterns in summer ozone pollution over Eastern China and their relationship to atmospheric circulations. The focus of this current paper will be on the use of EOF analysis on total column CO data to attempt the separation of anthropogenic and biogenic modes of variability from the background. While EOF have proven useful in climatological studies, this approach has significant limitations that require careful consideration, as detailed in this study.

## 2 Methods

### 2.1 Data Description

We use total column CO data from the Measurements of Pollution in the Troposphere (MOPITT) instrument which uses a correlation radiometer from the NASA Terra instrument. Terra flies at a normal altitude of 705 km in a Sun-synchronous polar orbit that passes over the equator at approximately 10:30 and 22:30 local time. The temporal resolution is daily which has been averaged into 8-day periods, and the spatial pixel resolution is 22 km x 22 km at nadir, resampled to 1° from L3 products. The cross-track scanning angle is ±26 degrees and yields a swath of approximately 640 km with global coverage every 3 days.

The radiometer performs gas correlation spectrometry for broadband measurements in thermal infrared (TIR) at approximately 2140 $cm^{-1}$ and near-infrared (NIR) around 4725 $cm^{-1}$ (Buchholz et al. (2021), Drummond et al. (2010)). This is done by modulating either pressure or length of a correlation cell filled with gas of the target to determine spectral line differences. The TIR is measured using terrestrial thermal radiation while the overtone band from NIR is measured from reflected solar radiation and enhances retrievals, especially near the surface. Products are available in thermal infrared only, near-infrared only, or joint TIR-NIR, although NIR signals are only available during the day and over land. (Buchholz et al. (2017)) For this study, we use Level 3 data using the V8 retrieval algorithm for TIR-NIR. Level 2 and Level 3 products for MOPITT for the V8 retrieval algorithms (and its most recent version, V9) are publicly available through repositories located at https://terra.nasa.gov/data/mopitt-data.

The V8 product includes updates to MODIS cloud cover Collection version 6.1, spectral data for $H_2O$ and $N_2$ as well as aircraft data from NOAA Deeter et al. (2019). This enables a radiance-based bias correction and accounts for both temporal bias drift and bias as a result of water vapor. Daytime retrievals of total column CO for V8 are stable and have a nominal drift of $-0.015 \pm 0.061\%$ relative to NOAA airborne flask sampling during the MOPITT mission. MOPITT measurements have been verified extensively using cross-validation with other satellites, in-situ measurements at the ground, as well as via aircraft, and more recently through ground-based solar Fourier transform infrared spectrometer measurements (Buchholz et al. (2017); Martínez-Alonso et al. (2020); George et al. (2015)).

We note that satellite retrievals of tropospheric CO are also available using a multitude of other instruments including the Atmospheric InfraRed Sounder (AIRS) onboard Aqua, the Tropospheric Emission Spectrometer (TES) on Aura, the Infrared Atmospheric Sounder Interferometer (IASI) on the MetOp platform, the Tropospheric Monitoring Instrument (TROPOMI),

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


and the Cross-track Infrared Sounder (CrIS). These instruments are all sun-synchronous, capable of measuring CO in the infrared region, and have been shown to have consistent hemispheric CO variability when compared to MOPITT CO records from 2002 to 2018 (in the case of CrIS, from 2015 onward) (Buchholz et al. (2021)). Because CrIS was launched in October 2011, it does not have a long or consistent enough period of measurement to be suitable for the determination of long-term trends of CO. The same applies for TROPOMI which was launched in 2017. While CO measurements for IASI-A are long enough to determine trends, they have potential discontinuities due to the fact that CO retrievals are not done using consistent humidity, temperature, and cloud information. Measurements from TES are 2 orders of magnitude fewer than other instruments and are more difficult to verify due to not being collocated. Measurements from AIRS have a long and stable enough record for trend determination and have been shown to reproduce trends from MOPITT, however, AIRS uses a cloud-clearing algorithm that increases global coverage at the expense of increasing ground pixel sizes to 45 km x 45 km resulting in coarser resolution.

Our choice of using MOPITT retrievals is because it has the longest available CO record with a verified history of reliable and stable observations and relatively fine spatial resolution. We use daily data averaged to 8-day periods to give us a sufficient resolution to observe CO during its lifetime while also reducing computation time and giving a more consistent overpass time. We use measurements for the 14 year duration of 2005-2018 so that we have sufficient data to be able to study long-term patterns. Plots are shown below of the spatial (Fig. 1) and temporal (Fig. 2) mean of de-trended CO for global data over our stated observation period. We will discuss the significance of the variability observed by MOPITT in Sect. 3, focusing on its relationship to the sources and sinks of CO.

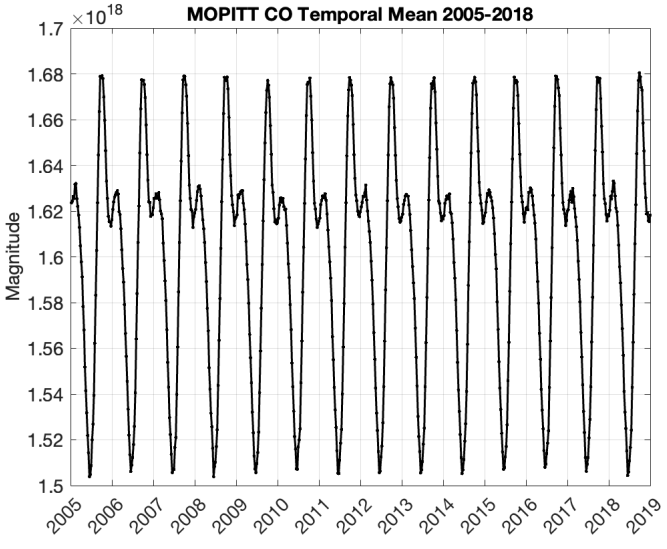


**Figure 1.** Spatially averaged time series of de-trended MOPITT CO over the period of 2005-2018 across the globe from 60°S to 60°N (in molecules $cm^{-2}$).

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


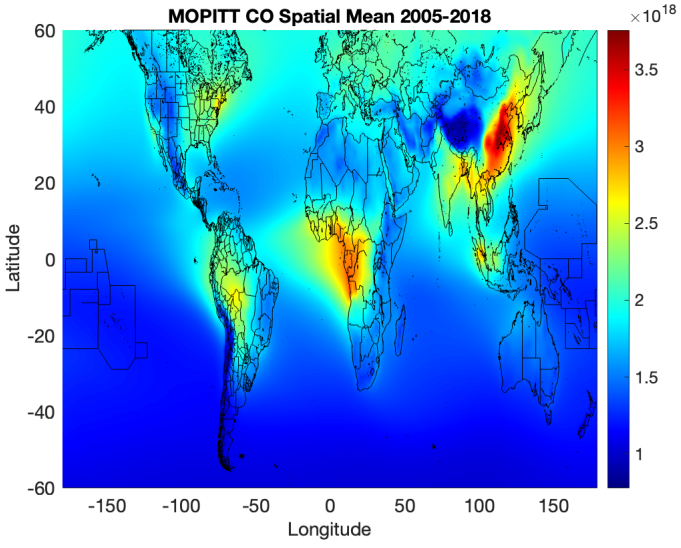


**Figure 2.** Temporally averaged global map of de-trended MOPITT CO over the period of 2005-2018 across the globe from 60°S to 60°N (in molecules $cm^{-2}$).

### 2.2 Source Signatures of CO

Here, we briefly describe the source categories of CO, which include fossil fuel/biofuel burning, biomass burning, biogenic, non-methane hydrocarbon oxidation, and methane oxidation as summarized in Holloway et al. (2000), Edwards et al. (2004) and Edwards et al. (2006). Describing these categories is useful because surface emissions of CO can display unique signatures that may be linked to specific spatial and temporal patterns in observed CO abundance.

CO from anthropogenic combustion (fossil fuel/biofuel burning) is much larger in the NH compared to the SH which is attributed to significantly more fossil fuel consumption in urban and industrial regions. We consider these areas of high emission to be mostly associated with eastern China, India, the eastern United States, and Western Europe. Historically, eastern China has contributed significantly to anthropogenic emissions because of a distinct reliance on coal and biofuel burning, however, recent regulations have shown to make progress toward slowing down emissions Buchholz et al. (2021). The lifetime of CO is much longer during the wintertime and high pollution events may persist in the atmosphere for weeks at a time. It is found in general that CO loading tends to reach maximum values in April and minimum values in September (Edwards et al. (2004)).

Biomass burning is largest in the Southern Hemisphere and a primary source of CO. Biomass burning consists of burning of savanna, forests, agricultural residue, fuelwood, and animal waste and is most concentrated in tropical and sub-tropical forests and grasslands Holloway et al. (2000). The largest contributions of fire emissions come from Africa and South America, although fires also occur in Indonesia, the Boreal forests of Canada, Alaska, and Russia as well as Australia. African NH fires occur in the savanna south of the Sahara Desert and in tropical rainforests north of the equator, while SH fires begin burning in the western portion of the continent near Angola and then spread southeast down the eastern coast. Fires in South America

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025




happen on the southern edge of the Amazon rain forest, the cerrado grasslands, and drought-deciduous forests in the south (Edwards et al., 2006). Another distinct feature of CO fire emissions is their timing (Wiedinmyer et al., 2023). While long-term drivers such as land-use and land cover changes, as well as drought conditions, heavily influence fire occurrences, seasonal variations are primarily driven by agricultural burning practices and prevailing weather conditions. The anomalously high CO levels from these fires, coupled with their distinct spatial and temporal variability inherent to biomass burning, provide a more reliable signature for distinguishing fire emissions from other CO sources.

The contribution of methane oxidation as well as oxidation from non-methane hydrocarbons (NMHC) represent secondary sources of CO that are most apparent in the Southern Hemisphere. The oxidation of NMHCs includes a wide variety of constituents which includes isoprene, terpenes, acetone, as well as industrial and biomass NMHCs. Temperature-dependent biogenic emissions of isoprenes account for about a third of hydrocarbon emissions. They are volatile organic compounds that are emitted by trees, shrubs, and animals including humans (measurable in breath). Terpenes are produced primarily from coniferous forests and are released in higher amounts with warmer weather. The largest emissions of hydrocarbons occur in the tropics which is primarily from isoprene and occurs because of large density of biomass and high temperature conditions. During the NH winter, there is a strong gradient resulting from biogenic emissions that are south of the equator where the intertropical convergence zone is located, and includes emissions from South America, Southern Africa, and Australia. During the NH Summer biogenic sources contribute more in the NH resulting in a weaker but still observable gradient. This gradient is significantly stronger for surface observations but is observable at all levels of the atmosphere (Holloway et al., 2000). CO production from methane oxidation can be observed across the troposphere and tends to be spatially homogeneous given the longer lifetime of $CH_4$ ($\approx$ 10 years). It does however exhibit some seasonality due to chemical loss with its reaction to OH.

### 2.3 Mean Spatial and Temporal Distribution of CO

In this section, we describe the spatial and temporal variations in CO abundance, illustrated in Figs. 1 and 2 using MOPITT retrievals of total CO columns.

In addition to source signatures, the variability in CO, particularly its seasonality, is significantly influenced by its chemical loss with the OH radical. Notably, CO serves as the primary sink for OH, while $CH_4$ serving only as a secondary sink. A recent study by Bryant and Stevenson (2024) has shown that increased hydrogen can also increase the mixing ratio of CO in the troposphere through photolysis of formaldehyde (HCHO) and reactions with OH and nitrate ($NO_3$).

Reports of the inter-annual variability of CO can be identified as fluctuations in the observed gradient of its abundance between the Northern and Southern Hemispheres. This gradient is driven by the significant differences in surface emissions between the hemispheres, along with the associated air mass exchange that occurs over a year. respectively. To a first approximation, CO concentrations are lowest where OH abundance is highest, as CO's chemical lifetime is strongly dependent on reactions with OH. These areas are characterized by having high photochemical activity (solar radiation), as well as high concentrations of $O_3$ and $NO_x$, and correspond to continental regions during the summer. On the other hand, oceanic (remote) regions during the winter correspond to areas with the lowest OH and therefore the highest CO. However, the transport and mixing of surface emissions complicates this first order relationship.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Seasonal differences in the distribution of OH and the imbalance of emission in each hemisphere lead to a North-South gradient in CO that is clearly visible during the NH winter but is absent in the NH Summer. During the NH winter, low OH in the NH lead to high CO while high OH in the SH lead to low CO. This in turn reinforces large NH anthropogenic emissions from Europe, North America, and Asia that peak during early spring time which leads to a steep gradient during the NH winter. During the NH summer, we instead have low CO in the NH, but high CO in the SH which balances a much lower budget of emissions of CO in the SH.

Significant changes in the annual cycle have been connected to fire emissions due to biomass burning, and limited studies have connected variability in the SH to climate indices where biomass burning is the primary driver. In the NH, fires from Boreal forests in North America and Eurasia have been shown to contribute to significant interannual variability. In the SH, while fires in southern Africa and South America generate overall high CO loadings, a primary driver of variability is due to fires in the Maritime Continent and Northern Australia. Buchholz et al. (2018) were able to use multilinear regression with four climate indices (ENSO, IOD, TSA, and AAO) to explain the variability of CO across the Southern Hemisphere and tropics including Maritime SEA, two subregions in each of Australasia, Southern Africa, and South America. They found that more than $50\%$ variability is explained over all regions and more than $70\%$ variability was explained across maritime SEA and North Australasia. The fire intensity and burned area are connected to the amount, type, and dryness of fuel availability which depend on climate-sensitive water availability and ecosystem responses. This indicates that SH interannual variability of CO loading has a clear but complex dependence on climate.

### 2.4 EOF Analysis

#### 2.4.1 Overview of EOF Analysis

The following is a simplified description of EOF analysis. A more complete description can be found in Hannachi et al. (2007), Björnsson and Venegas (1997), Wilks (2011), Hannachi (2021), and von Storch ad F. Zwiers (2003). The exposition presented here is fairly involved because we wish to provide readers both practical and technical foundations that are tailored for the use of EOF analysis in atmospheric chemistry datasets, especially because most references approach the subject from the perspective of meteorological applications. We do this to simultaneously provide a tutorial for future atmospheric chemistry students, as well as to highlight both the strengths and the limitations of EOF analysis in our findings.

Assume that we have a set of $n$ observations in time of a gridded field with a total of $m$ gridpoints. Let $f_{ij}$ where $i = 1, 2, ...n$ $j = 1, 2, ...m$ represent the value of the field observed at time $i$ and location $j$, then the data is represented by the $n \times m$ matrix $F$:

$$F = (\mathbf{f}_1, \mathbf{f}_2, ...\mathbf{f}_n)^T = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1m} \\ f_{21} & f_{22} & \cdots & f_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nm} \end{pmatrix}$$

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


so every row of $F$ represents a $1 \times m$ vector corresponding to a single observation in time across all grid points.

We now construct the anomaly matrix $F'$ that represents the departure from normal climatology by removing from each grid point its corresponding mean across time. This is done by taking $F$ and individually subtracting the mean for each of its $m$ columns. Instead of constantly referring to $F'$ as the anomaly matrix, we will simply refer to it as $F$ from now on.

Our goal for EOF Analysis will be to decompose the anomaly matrix $F$ into a matrix of orthogonal spatial patterns called the EOF modes $C$, and a set of uncorrelated time series $\Phi$ such that

$$F(x,t) = \lambda_1\phi_1(t)c_1(x)^T + \lambda_2\phi_2(t)c_2(x)^T + \cdots \lambda_p\phi_p(t)c_p(x)^T = \sum_{k=1}^{p} \lambda_k\phi_k(t)c_k(x)^T$$

where $p \leq r$, with $r = \text{rank}(F) \leq min(n,m)$. The number of modes $p$ to retain in the truncation is always chosen such that the maximum percent variance of the original dataset can be explained using the new set of basis functions. The data is compressed more when fewer modes of variability are kept in the expansion.

This decomposition can be accomplished by phrasing the question in the form of an optimization problem. We will find uncorrelated linear combinations of the $p$ spatial patterns $C = (c_1, ..., c_p)^T$ such that the resulting projected time series $\Phi = FC$ have maximum variance and are uncorrelated. In terms of the $m \times m$ spatial covariance matrix $R = \frac{1}{n}F^TF$:

The maximization problem to be solved for $C$ such that :

$$c_k^T\, R\, c_k = \frac{1}{n}||Fc_k||^2 \text{ is maximized,} \quad \text{with } c_k^T c = 1 \text{ for } k = 1, ..., p.$$

If we define $\lambda_k = \frac{1}{n}||Fc_k||^2 \geq 0$ The maximization problem is equivalent to solving the eigenvalue problem given by:

$$RC = C\Lambda$$

where $\Lambda = diag(\lambda_1, \lambda_2, ...\lambda_p)$.

So we see that the EOF spatial patterns $C$ correspond to the eigenvectors of the spatial covariance matrix $R$, and the PC time series $\Phi$ correspond to the projections $\Phi = FC$. It is important to note that because the covariance matrix $R$ is symmetric and positive semi-definite it has non-negative eigenvalues with well-defined square roots as well as orthogonal eigenvectors. An equivalent formulation is to use singular value decomposition on $F$, so we expand it into unitary matrices $\Phi$ and $C$ and a matrix $\Sigma = \Lambda^{1/2}$ of singular values such that:

$$F = \Phi\Sigma C^T.$$

Using this expression for $F$, we can verify using direct substitution that the EOF spatial patterns $C$ are indeed the eigenvectors of the spatial covariance $R = \frac{1}{n}F^TF$, and additionally that the PC time series $\Phi$ are the eigenvectors of the temporal covariance

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


$L = \frac{1}{n} F F^T$. Note that in the case that each column of $F$ represents a different observed field, it should also be divided by its standard deviation so that the covariance matrix then becomes a correlation matrix.

In Appendix A, we detail exactly how the EOF patterns and PC time series can be computed in Matlab either using SVD or by using eigendecomposition on $R$ or $L$ and then projecting the resulting eigenvectors onto the original matrix $F$. While using SVD is stable computationally, it may require too much computing power in the case that $F$ is very large because it requires us to compute eigenvectors of both $R$ and $L$. When the spatial grid is large, then $\dim(R) >> \dim(L)$ and we decompose $L$. On the other hand, if the number of observations is large, then $\dim(L) >> \dim(R)$ and we instead will decompose $R$. An even better alternative that is considerably faster with no loss of stability is to use Lanczos iteration to solve the eigenvalue problem as detailed in Toumazou and Cretaux (2001).

### 2.4.2 Separation of Modes Using the North Test

Assuming that the $p$ eigenvalues have been sorted in descending order of magnitude (this is done automatically in SVD algorithms but not necessarily in eigendecompositions), then the percent variance that is accounted for by the $m$th mode is given by Hannachi et al. (2007):

$$\% \, var = \frac{\lambda_m}{\sum_{k=1}^{p} \lambda_k} \times 100\%.$$

Because of correlations between variables (spatial autocorrelation in $F$), it is not possible for $F$ to have truly independent samples. The sampling uncertainty for the eigenvalues can be estimated by using a variety of methods including Asymptotic Approximation, Probabilistic PCA, Monte-Carlo Resampling Methods, or Bootstrapping (Hannachi (2021)). In this paper, we emphasize the asymptotic method as described by North et al. (1982) who established an estimate of the typical error $\Delta\lambda_k$ between two neighboring eigenvalues to be:

$$\Delta\lambda_k \approx \sqrt{\frac{2}{N^*}} \lambda_k$$

so we see that the error in each eigenvalue is proportional to its magnitude. Because this approximation relies on the central limit theorem and the law of large numbers, it is valid as long as the number of independent samples $N^*$ is large.

According to Hartmann (2016), $N^*$ is estimated in terms of a red noise (AR1) time series defined by exponentially decaying autocorrelation. If $a = r_1(\Delta t)$ is the autocorrelation at one time step $\Delta t$, then $N^*$ for first order processes is approximated in terms of the true sample size $N$ as:

$$N^* = N \left( \frac{1-a}{1+a} \right).$$

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


This expression is not accurate for the case of non-Gaussian red noise. North also gives an estimate for the error between eigenmodes ($c_k$ is the EOF for $\lambda_k$) as:

$$\Delta c_k = \left( \frac{\Delta \lambda_k}{\lambda_j - \lambda_k} \right) c_k$$

which shows that large errors in eigenvalues will also cause large errors in a neighboring EOF mode. When there is a significant overlap between the error bars of adjacent eigenvalues, it is a good indicator that these modes are either degenerate due to being contaminated by error, or potentially represent white noise.

### 2.4.3 Weaknesses of EOF Analysis

While EOF analysis can be used as a very powerful tool to compress data and determine dominant modes of climate variability, the method suffers from numerous potential limitations. EOF modes have been extensively found to suffer from the following major drawbacks in comparison to Rotated EOFs (REOFs) (Hannachi (2021), Richman (1986)):

1. Domain Dependence
2. Subdomain Instability (Non-locality)
3. Sampling Error
4. Inaccurate Representation of Physical Phenomena

Domain dependence is a problem that was first reported by Kaiser (1958). Buell (1978) determined that the primary factor that determines the shape of an EOF was not its covariance structure, but rather the shape and size of the domain (also Lehr and Hohenbrink (2024)). This is very immediately a problem, as the covariance structure contains the features of our data and allows us to separate it into different modes of variability. Buell (1978) also showed that it is possible to find several different correlation functions that have a very similar structure over the same domain.

Sub-domain instability is closely related to the issue of domain dependence and references the lack of invariance when EOFs from a given domain are compared to those generated from localized subdomains (Richman (1986)). In order for an EOF to be physically insightful, it should for instance be the case that EOF patterns over Africa and Asia should reproduce those generated globally. If this was not the case, the modes generated would be impossible to reproduce with any semblance of consistency. This difficulty is caused by the orthogonality constraint of EOFs which causes the variance criterion to become maximized globally rather than locally, and leads to the inability of EOF analysis to separate independent and localized features that share similar covariance structure with each other (Horel (1981), Monahan et al. (2009)). This non-locality issue is also consistent with the mixing property for EOFs, which refers to their tendency to take a signal which begins as a linear superposition of several different independent and potentially uncorrelated signals, and then combine them together in order to maximize variance (Hannachi (2021)). Sampling error between neighboring eigenvalues is an issue that we have discussed in Sect. 2.4.2 of this document, so we will not examine it further here.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


The difficulty EOF analysis suffers in expressing physical phenomena is a complex issue, and here we summarize the main points on this topic as discussed in Monahan et al. (2009). The first major issue is that statistical modes generated in EOF analysis do not in general represent dynamical modes that correspond to true physical modes of variability, and in fact hold only under very strict circumstances. This is not unexpected, as the forced orthogonality of modes for EOFs is a mathematical constraint rather than a physical one. In particular, linearized climate models have conditionally orthogonal modes. The linearized barotropic model given in Simmons et al. (1983) does not have orthogonal processes. A more general climate model with conditional orthogonality is given in Monahan et al. (2009).

We should also not expect EOF analysis to produce principal component time series that are truly independent, because in the absence of autocorrelation, we could still have uncorrelated variables that may be related in a nonlinear way. Only in the case that $X$ has a joint Gaussian distribution will it have modes that are uncorrelated and independent. To see this fact, we note that in the case of 2 random variables if $X_1$ and $X_2$ are uncorrelated and jointly Gaussian then the associated covariance matrix is diagonal and the joint PDF takes the form

$$f_X(x_1,x_2) = \frac{1}{\sqrt{(2\pi)^2\sigma_1^2\sigma_2^2}}\exp\left(-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right)$$

which can be written as:

$$f_X(x_1,x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right)\right)\times\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{1}{2}\left(\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right)$$

$$= f_X(x_1)\times f_X(x_2)$$

so therefore $X_1$ and $X_2$ are independent. Without this joint Gaussian assumption, it does not follow in general that uncorrelated random variables are independent. A simple counterexample is given in Hannachi 2021 where if we take $X$ to be standard normal, then $X$ and $Y = X^2$ are uncorrelated, but are not independent. The degree to which a field $F$ deviates from Gaussian behavior can be estimated by calculating the skewness:

$$\text{skew}(F) = \left\langle\left(\frac{F-\langle F\rangle}{\text{std}(F)}\right)^3\right\rangle$$

Climate data which has a relatively low degree of skewness can still be thought of as being Gaussian and therefore having independent principal components.

Some, but not all of these weaknesses in EOF analysis can be resolved using REOFs, but even they come with their own drawbacks, which will be discussed in the next section.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


#### 2.4.4 Varimax Rotation

We describe rotated EOFs (REOFs) using the varimax criterion, which are EOFs that have been rotated according to a maximization condition. REOFs are motivated by the need to make EOF patterns more localized, easier to interpret in physical space, and less dependent on the domain. This is accomplished by relaxing orthogonality constraints for the EOFs and/or the uncorrelated constraints of the PCs subject to a specified criterion. Rotations can either be oblique or orthogonal and include more than ten different possible criteria according to Richman (1986). We will only discuss the varimax criterion in detail, which has popularity due to the fact that orthogonal rotation is less sensitive to changes in the number of variables, and it avoids some convergence issues that can happen during matrix inversion in oblique rotation.

Suppose we begin with an $m \times p$ matrix $U_p$ of the first $p$ leading EOF spatial modes:

$$U_p = \{u_1, ...u_p\}$$

Then the new $m \times p$ matrix of REOF modes $B$ is defined as:

$$B = U_p Q$$

where $Q$ is a $p \times p$ rotation matrix. In varimax rotation, the matrix $Q$ is chosen to be orthogonal so that it satisfies

$$Q^T Q = Q Q^T = I_p$$

where $I_p$ is the $p \times p$ identity matrix. The varimax criterion requires maximizing the function $f(B)$ where:

$$f(B) = \sum_{k=1}^{p} \left[ m \sum_{j=1}^{m} b_{jk}^4 - \left( \sum_{j=1}^{m} b_{jk}^2 \right)^2 \right]$$

In plain English, the varimax criterion is named because it attempts to maximize the sum of the spatial variances of the squared and rotated EOF patterns $b_{ij}$. Orthogonality is preserved while achieving simplicity by forcing patterns to have magnitude near either 0 or 1 so that each pattern can be easily represented by the linear combination of only a few basis vectors.

One potential drawback from varimax rotation is that leading patterns produced by globally maximizing variance could potentially be destroyed. As $p$ increases and we rotate a higher number of EOF patterns, the rotated EOFs become more localized at the expense of the leading patterns, which may no longer be invariant under the orthogonal transformation.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


## 2.5 Seasonal Decomposition using Singular Spectrum Analysis

Singular spectrum analysis (SSA) is a very powerful tool in time series analysis with a wide variety of applications, some of the more notable ones including forecasting, change point detection, filtering, and detection of spatio-temporal patterns. SSA also has applications to meteorology and air quality monitoring, for instance, Macias et al. (2014) uses SSA on the global surface temperature to separate a multi-decadal oscillation from a secular trend, and Gruszczynska et al. (2019) uses multichannel SSA to perform spatiotemporal analysis for atmospheric, continental hydrology, and non-tidal ocean changes to determine an annual signal for 16 sections depending on the climate zone. Another application is from Espinosa et al. (2022) which uses the time derivative of a resulting SSA signal to detect air quality anomalies.

Broadly, our goal is to take a time series $x = [x_1, x_2, ..., x_N]$ of length $N$ and then decompose it into a sum of trends, periodic (or quasi-periodic) components, or noise, the sum of which will reconstruct the original series. SSA is accomplished in two main stages: decomposition and reconstruction each of which has two different steps. A detailed description of the method can be found in Golyandina and Zhigljavsky (2013). Here we summarize the main points as discussed in Hassani (2007).

In stage 1 the decomposition is done by beginning with an embedding where the original 1-D time series is mapped into $K = N - L + 1$ different lagged vectors of length $L$ where $L$ is called the window length, a number that satisfies $2 <= L < N$ and should be large enough to capture a key feature of the data. The $L$ trajectory matrix $X$ is then a Hankel matrix of size $(L, K)$ where each column is one of the lagged vectors and each element is constant along an anti-diagonal $i + j = const$:

$$X = [X_1, X_2, ...X_K] = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N. \end{bmatrix} \quad (1)$$

After completing the embedding, SVD is performed on the trajectory matrix by computing the eigenvalues $\lambda_1 \geq \lambda_2, ... \geq \lambda_L$ and eigenvectors $U_1, U_2, ...U_L$ of $S = XX^T$ (since $L < N$). The eigenvectors of $X^T X$ are then computed as $V_i = X^T U_i / \sqrt{\lambda_i}$ $(i = 1, 2, ...d)$ with $d = \text{rank}(A)$ and the decomposition for $X$ is completed by writing it as a sum of $d$ rank 1 elementary matrices:

$$X = U\Sigma V^T = X_1 + \ldots + X_d$$

where $\Sigma = \Lambda^{1/2}$ is the matrix of singular values.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Suppose that $I = 1, 2, \ldots, ...d$ is the set of indices corresponding to our listing of elementary matrices. The reconstruction stage is begun by splitting the set $I$ into $m$ disjoint subsets so that:

$$I = I_1 \cup I_2 \ldots I_m.$$

The partitioning gives us $d_1$ elements in $I_1$, $d_2$ elements in $I_2$ etc. and finally we have $d_m$ elements in $I_m$ so that $d = d_1 + d_2 + \ldots d_m$. This procedure of choosing the $m$ subsets is called eigentriple grouping and allows us to write the reconstruction as $m$ distinct sums:

$$X = X_{I_1} + X_{I_2} \ldots + X_{I_m}$$

$$= \sum_{i \in I_1} X_i + \sum_{i \in I_2} X_i + \ldots + \sum_{i \in I_m} X_i.$$

In the case that $m = d$ we reproduce the elementary grouping of rank 1 matrices. The percent variance that a specific group $I_p$ contributes to the total variance is found by taking a ratio of their eigenvalues:

$$\%var = \frac{\sum_{i \in I_p} \lambda_i}{\sum_{i=1}^{d} \lambda_i}.$$

After the eigentriple grouping is finished a diagonal averaging procedure called "Hankelization" is done to each matrix in the sum by averaging across the antidiagonal elements in the matrix where $i + j = k + 1$ for $k = 1, \ldots N$ in order to create $m$ different time series of length $N$. We therefore obtain a reconstruction of the original time series:

$$x = \tilde{x_1} + \tilde{x_2} + \ldots + \tilde{x_m}$$

where $\tilde{x_j}$ corresponds to the resulting time series after $X_{I_j}$ has been diagonally averaged.

In practice, there is a large amount of subtlety to the eigentriple grouping procedure, however, one important comment is the fact that the ability to extract cycles depends upon the chosen window length $L$. In general, structure in short-term time series is detected by choosing a smaller window length, and structure in longer-term series is detected by choosing a longer window length.

## 3 Results and Discussion

### 3.1 Domain Independence and Separation of Modes for MOPITT CO

As discussed in Methods (Sect. 2), one of the first potential issues that needs to be addressed in order to produce physical EOF modes is whether or not the modes capture the local structure and if they are reproducible across differing subdomains. Prior to generating EOF modes, our data matrix of size $(longitude, latitude, time)$ is reshaped into size $(time, grid)$, and every

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


grid point is mean centered across time and then standardized. In Fig. 3, we show the first leading EOF mode for the dataset across the entire globe (top left) and then recompute the first mode after subdividing the domain into North American (top right), African (bottom left), and Asian (bottom right). The spatial structure of the original pattern has been preserved in each of the three subdomains, showing us that the local and global structure for the pattern is the same, and therefore our dataset is independent of its domain.

In some problems, it is often a good idea to attempt varimax rotation in order to ensure that the resulting patterns are simple and easy to interpret, but in our case, there is not a real need to do so because our problem does not suffer from domain instability. To demonstrate that varimax does not improve our resulting modes, we show the eigenvalue spectrum of the first 6 unrotated EOF modes together with the spectrum of our varimax modes in Fig. 3. Each mode is plotted as a fraction of the percent variance with the exact value shown at the top. We compute the error bars using the expression for $\Delta\lambda_k$ in Sect. 2.2.2 and plot the results in red. For our unrotated modes, the variance together with error bars (the length above/below each data value) are given by: $61.55 \pm 3.43\%, 20.52 \pm 1.143\%, 8.24 \pm 0.459\%, 3.92 \pm 0.218\%, 1.10 \pm 0.061\%$ and $0.69 \pm 0.038\%$. After using varimax rotation, the percent variance of each mode was: $86.02 \pm 4.793\%, 11.56 \pm 0.644\%, 1.87 \pm 0.104\%, 0.49 \pm 0.027\%, 0.03 \pm 0.002\%$ and $0.01 \pm 0.001\%$. None of these eigenvalues represent degenerate modes because their error bars do not overlap with adjacent neighbors. We note however that varimax modes 4,5, and 6 have such a small percent variance it is unlikely for them to rise above the level of noise and represent independent and important features. On the other hand, our unrotated EOFs have higher modes with a much larger percent variance and are more likely to represent independent features, with modes 5 and 6 most likely representing noise.

### 3.2 Seasonal Decomposition for MOPITT CO

We perform a seasonal decomposition on the mean-centered and standardized time series $A$ for each grid point of our dataset by using the *trenddecomp* function in Matlab with one mode of seasonal variability. This gives us $A = LT + S + R$ where $LT$ is the long-term variation, $S$ is the seasonal variation, and $R$ is the remainder. To visualize this decomposition, rather than choosing a single grid point as an example we plot a decomposition for the temporal mean of the dataset in Fig. 5 to show the average overall effect. The result is that the long-term component (blue) has a yearly periodicity that reproduces the peaks of the original MOPITT series (black), the seasonal component (red) has a shorter temporal periodicity (6 months) while the remainder (dashed, black) shows the shortest periodicity, being on the order of 3 months. Note that in each case the decomposition is phase shifted with reduced magnitude in comparison to MOPITT.

By implementing SSA on each grid point, we produce 3 matrices each with the same size as our original dataset which are again mean-centered and standardized. We then conduct EOF analysis on the resulting matrices, displaying the first 3 global modes for MOPITT in Fig. 6 together with the first 3 modes of the long-term, seasonal, and remainder datasets. While there are some differences in magnitude and percent variance, we note that the first two long-term EOF modes reproduce the spatial pattern of the first two modes for the original MOPITT CO patterns. Examination of the third long-term EOF mode shows that the pattern appears to be noise, as indicated by the low percent variance and the complete lack of any spatial structure. We also note that there is a very similar structure between the third MOPITT EOF mode and the first Seasonal EOF mode,

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


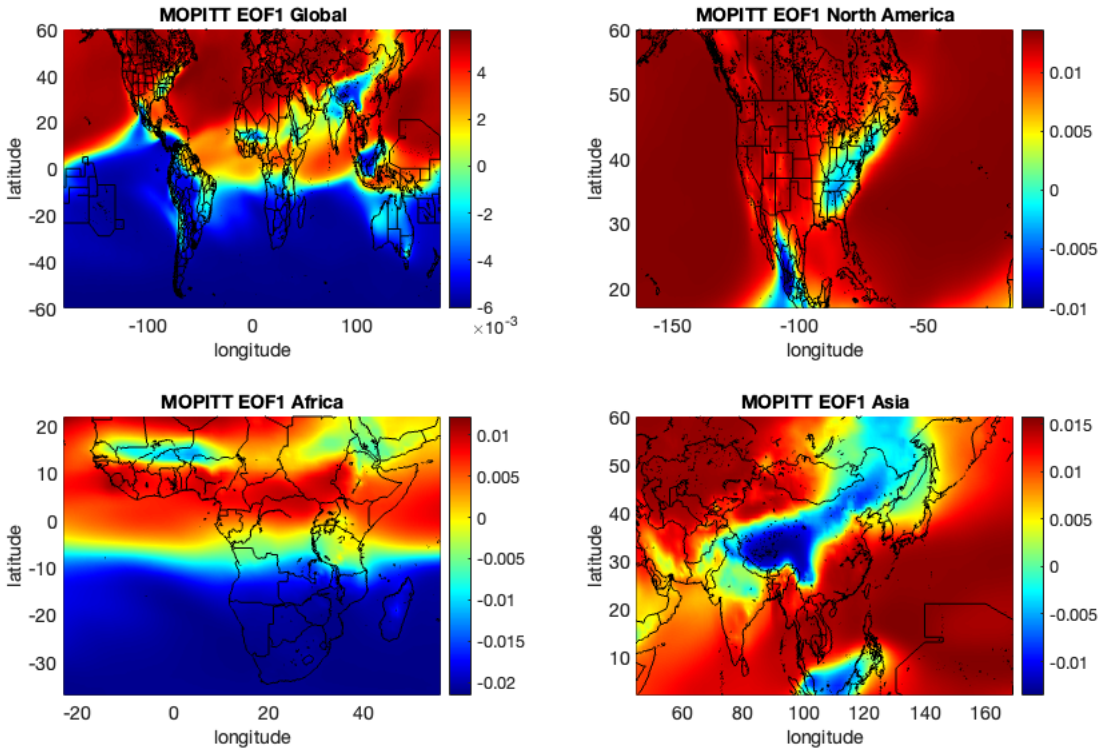


**Figure 3.** Spatial map showing the first EOF mode of MOPITT CO for the globe (Top Left), North America (Top Right), Africa (Bottom Left), and Asia (Bottom Right).

however, the first seasonal mode has a larger contrast between the positive and negative patterns which is especially noticeable in Northeastern Asia and the east coast of the United States. The structure for seasonal modes 2 and 3 as well as all three residual modes represent patterns distinct from our MOPITT EOF modes before we performed SSA on the data.

Our intention in performing EOF analysis was that each EOF mode would represent the variability from either a source or sink in carbon monoxide, however, the patterns depicted in Fig. 6 do not correlate with physically separable sources. If each mode represented a source of CO then we would expect to see modes that correlated with variability due to fossil fuel burning, biomass burning, biogenic oxidation, or methane oxidation. However, none of the modes depicted in Fig. 6 show patterns of variability that can be directly connected with source distributions.

As an example, let us attempt to connect MOPITT EOF 2 with fossil fuel burning. If we attempt the interpretation that blue areas represent areas with high contributions of fuel burning, this would be inconsistent with our knowledge of fossil fuel burning as a source of CO. While we do see significant areas of blue highlighted on the east coast of the United States which is consistent with combustion, the pattern shows too little activity over Eastern China and India and higher activity over Russia than expected. Similarly, when we examine EOF 3 we see a pattern that has some consistency with biomass burning in Africa

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


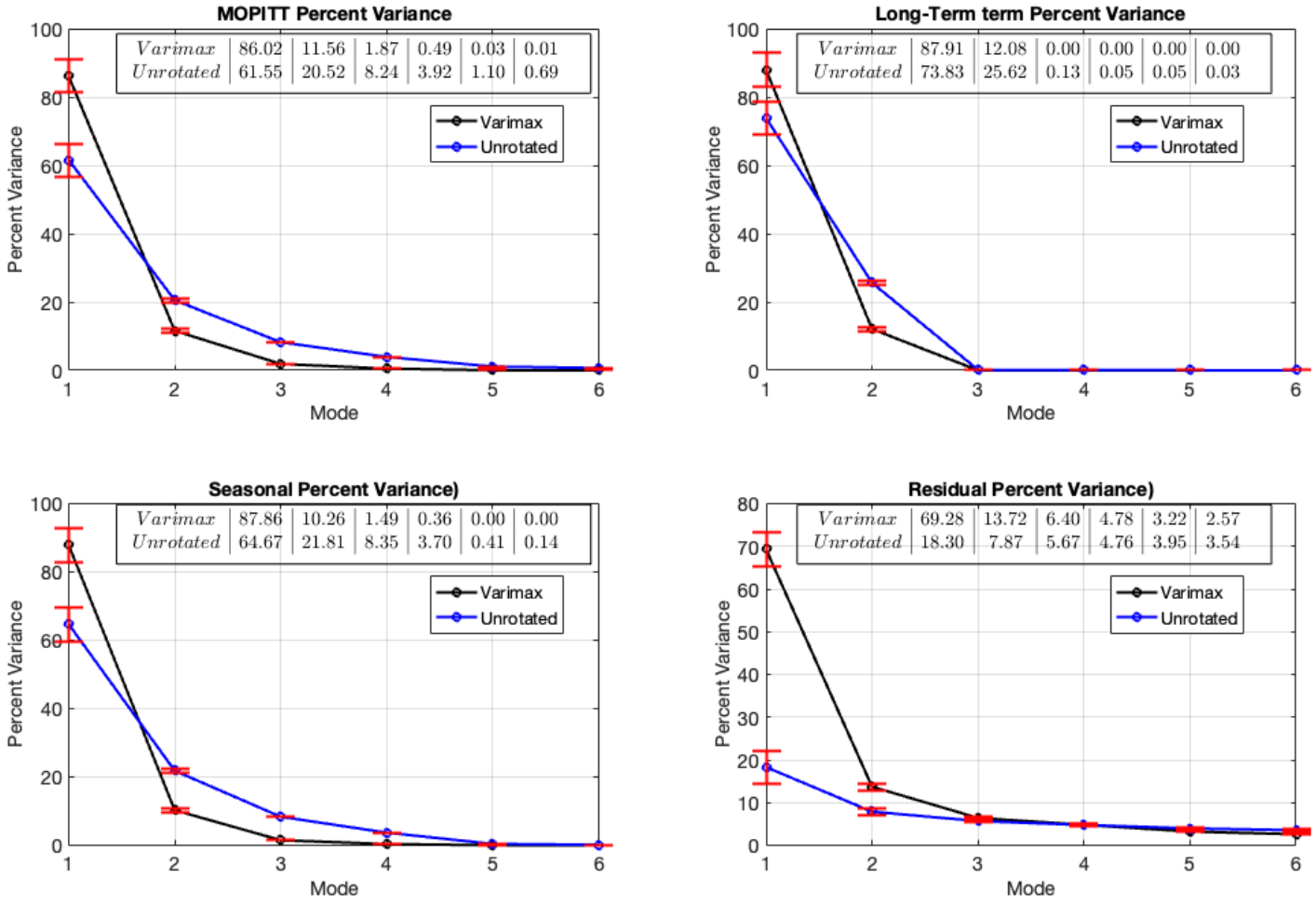


**Figure 4.** The percent variance that is explained by the first six EOF modes for the original MOPITT Dataset and the long term, seasonal, and residual components after using an SSA decomposition. Results are shown globally for both for varimax EOFs (black) and unrotated EOFs.

and South America, however, the features are far too smeared to be a potential interpretation. For example, the pattern stretches too far across oceanic regions as well as the Sahara and Saudia Arabia to be classified as potential fire activity. For similar reasoning, it would not make sense for us to interpret this as biogenic emission.

When examining EOF mode 1 we see a very clear boundary in the Northern and Southern Hemisphere as indicative of the annual cycle in CO caused by the combination of larger combustion activity and lower hydroxide radicals in the Northern Hemisphere during the winter season. This interpretation is consistent with an observation we state in Sect. 3.3 on spectral analysis and time scales which is that the time scale for EOF 1 is dominated by annual/semiannual variation.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


The majority of the areas with the highest contrast in Residual EOF mode 1 are seen to be over oceanic regions, which leads us to believe there may be a connection between Residual mode 1 and oceanic variability. However, these patterns do not match the variation of CO due to oceanic cycles, such as patterns as shown in Conte et al. (2019). This could be because we would expect oceanic variability to be strongest near the surface and patterns due to surface variations would be obscured by the changes in CO throughout other areas of the atmosphere which we are accounting for by using total column data.

Examining the remaining seasonal and residual patterns we see patterns that could be indicative of regional-scale transport, however, verification and further discussion of these modes is beyond the scope of our analysis at this time.

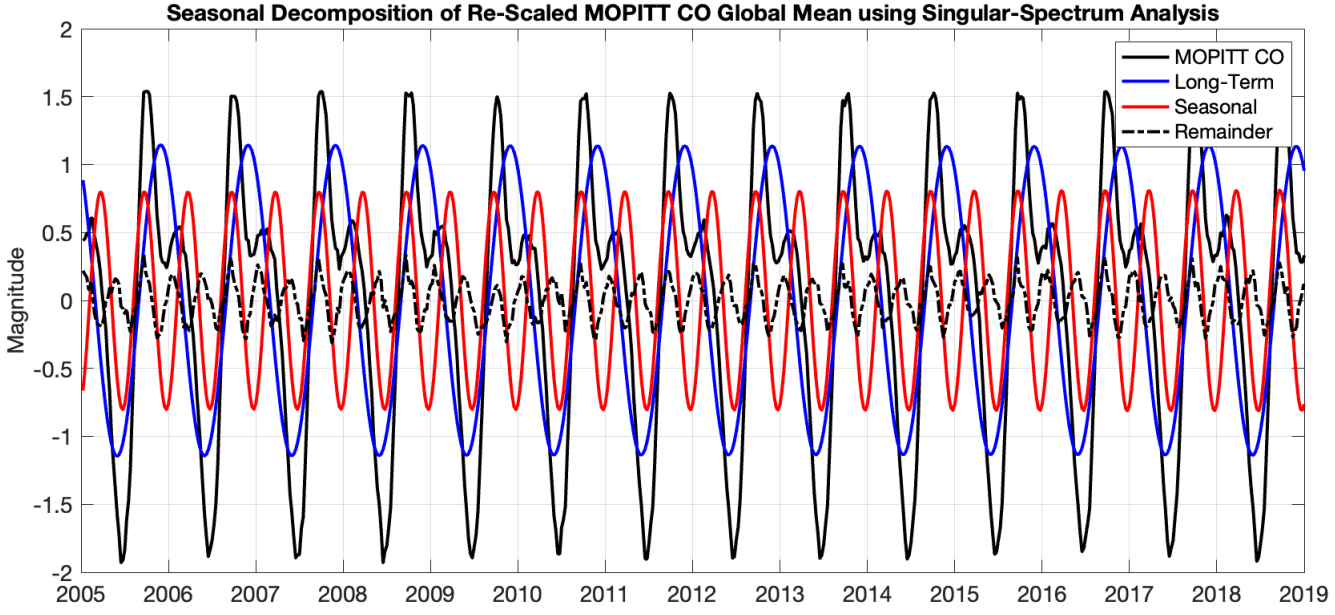


**Figure 5.** A time series plot of the spatially averaged MOPITT CO global mean (black) and the long-term (blue), seasonal (red) and remainder (black dotted) components of the SSA decomposition.

### 3.3 Spectral Analysis and Time Scales

In Fig. 7 we display the principal component time series for the original MOPITT dataset as well as the long-term, seasonal, and residual components. We first note that the first two long-term principal components have very similar periodic patterns as the original 2 MOPITT modes, while the third mode is of course noise because as we already seen the third long-term EOF pattern had no spatial structure. We also see the first two seasonal components have a similar periodic pattern to the third MOPITT component even though we saw from Fig. 6 that the second seasonal EOF has a very different spatial structure. The long-term modes have the largest period while the seasonal components have shorter periodicity and the residual modes the shortest.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


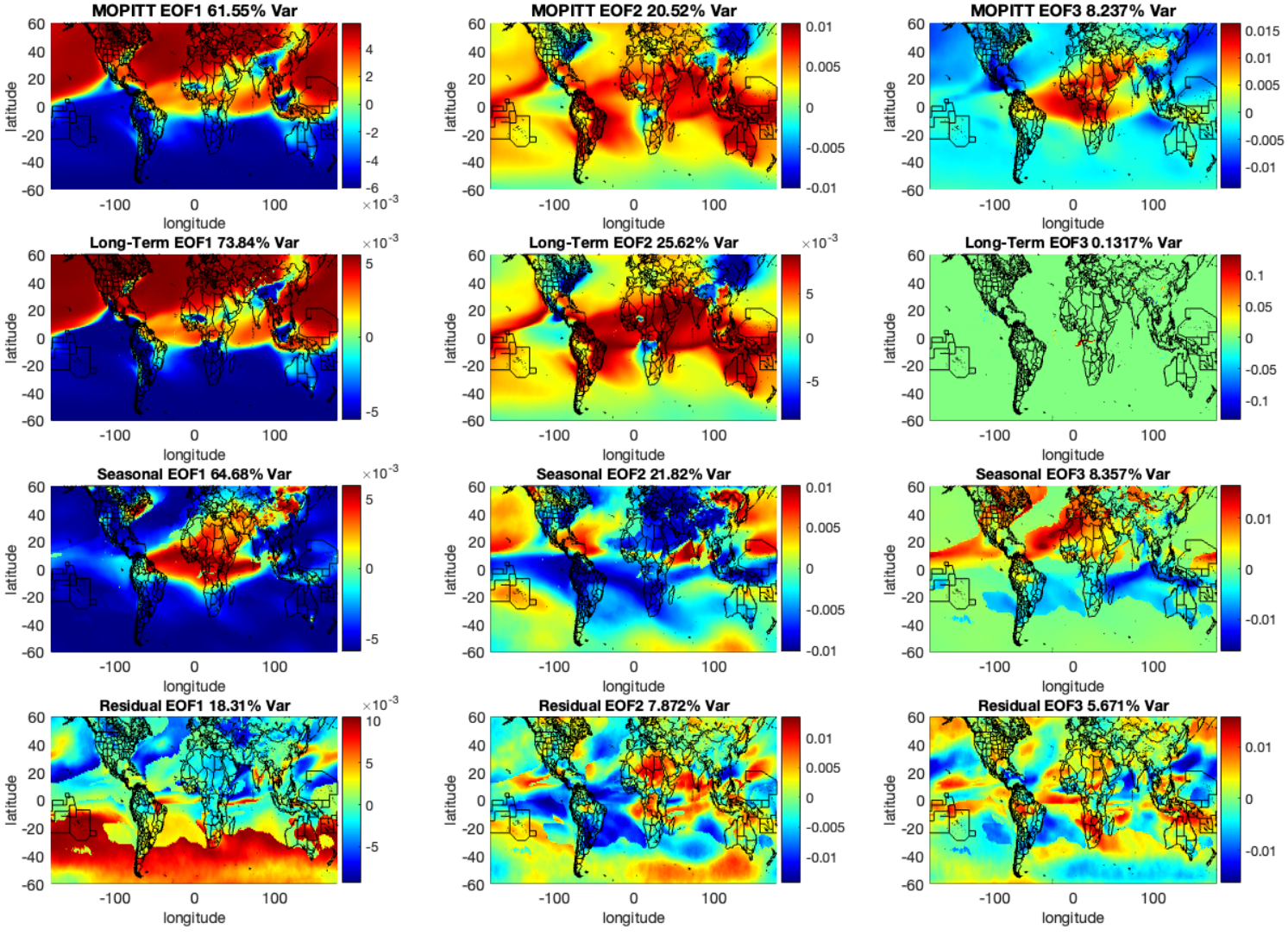


**Figure 6.** Spatial maps of the first 3 unrotated EOF modes of the MOPITT dataset and the long-term, seasonal, and residual components after using an SSA decomposition. The percent variance that is explained for each mode is displayed in the title.

In order to determine the most dominant time scale for a given mode, we used Fourier transforms to compute the power spectral density of each principal component and then used a numerical trapezoid rule to determine the total cumulative power as a percentage which is then plotted in Fig. 8 on a logarithmic scale. The percent of total power is labeled for a period of oscillation for 100 days, 1 year, and 1000 days. A dotted line and the value $t_{max}$ indicates the period of oscillation which has

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


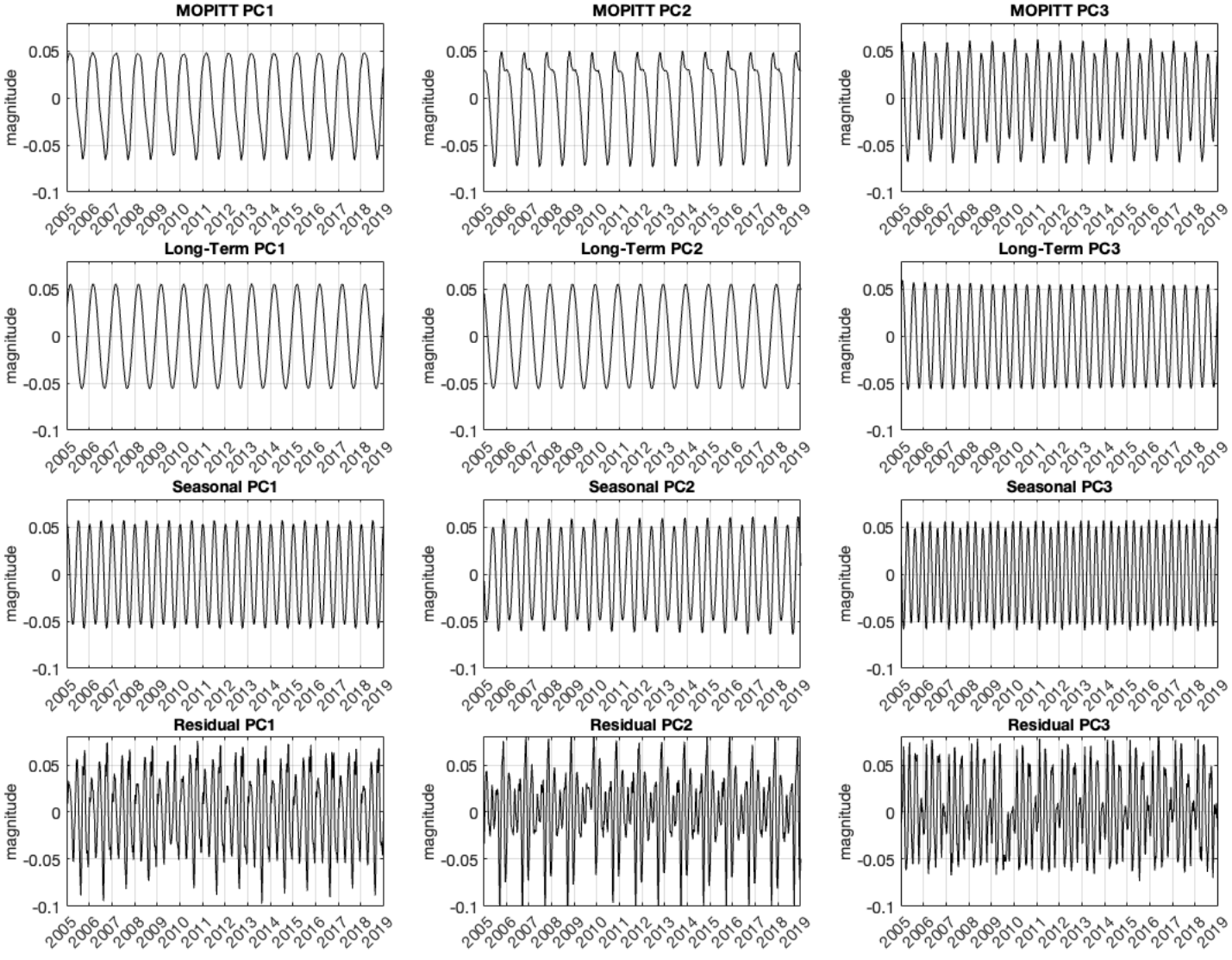


**Figure 7.** Time series for the first 3 principle components of the MOPITT dataset and the long-term, seasonal, and residual components after using an SSA decomposition.

contributed the largest amount to the total power. For example, examining Mopitt Spectrum Mode 1 we see that 1.3% of power is from periods of 100 days or smaller, 26.9% of power is from periods of 1 year or smaller, and 43.2% of power is from periods of 1000 days or smaller. We see that the first two Mopitt patterns correspond to variation dominated by an annual/semi-annual

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


480 cycle while the third mode is dominated by seasonal changes on the order of 3 months. Our long-term EOFs reproduce our original 2 modes, while our seasonal EOFs show dominant time scales on the order of 3 or 2 months, and our residual modes show dominant time scales on the order of 2 months or less.

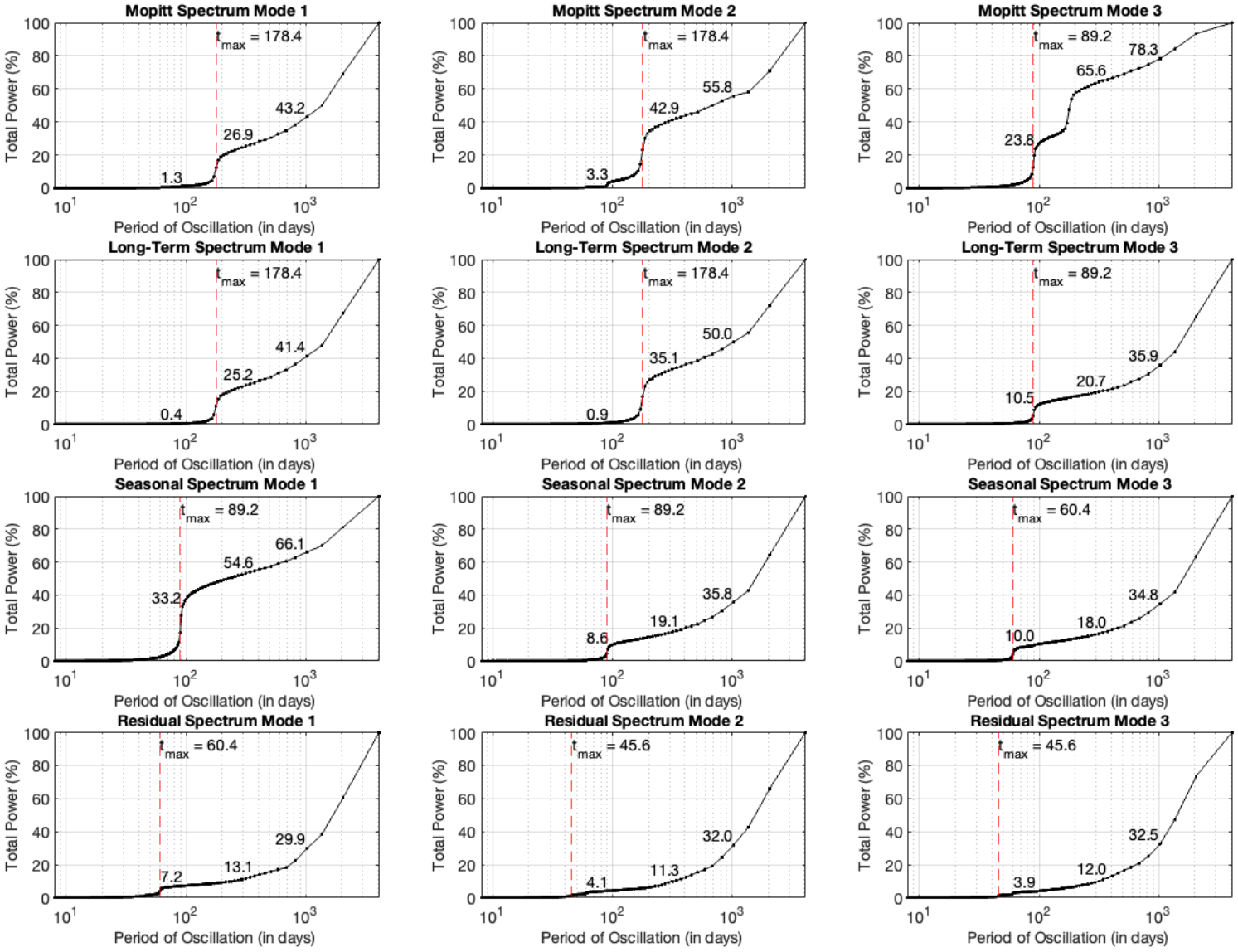


**Figure 8.** The cumulative percent total power for each principal component. Power values are labeled at 100 days, 1 year, and 1000 days. The dotted line and the value $t_{max}$ indicates the period of oscillation which has contributed the largest amount to the total power.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


### 3.4 Time Dependence and Non-Stationarity of Modes

As we have said in Sect. 2.2.3, one weakness of EOF analysis is that even though the principal component time series are constructed to be uncorrelated, it is not necessarily appropriate to assume that they represent independent processes which complicates the interpretation of assigning physical meaning to the modes. Only in the special case that the data is Gaussian does it follow that uncorrelated random variables are then deemed independent. In Fig. 9, we plot the mean skewness field for MOPITT CO over the period of interest from 2005-2018. It is standard to consider values of absolute skewness less than 0.5 to have negligible skewness corresponding to a relatively symmetric distribution, while values between 0.5 and 1.0 are moderately skewed, and values larger than 1.0 are highly skewed. Contour lines are shown for skewness values of -1, -0.5, +0.5, and +1.0 to highlight areas that have a notable degree of skewness. We see that our dataset has significant skewness in the Southern Hemisphere, especially in South America as well as over China, Indonesia, and East Asia. This indicates that we cannot consider our dataset to be approximately Gaussian, and is a potential explanation for why our principal components show significant dependence in the distributions of their spectral responses.

EOF analysis assumes that the dataset is stationary so that the mean and temporal covariance remains constant in time, which is done in order to sure that the first and second moments (the mean and covariance) of random variables can be determined using a limit of simple summations (or integrals). In practice, this assumption is rarely realistic with physical data. An equivalent statement is that the temporal autorrelation is constant as a function of lag time $\tau$. If we consider the mean and covariance of carbon monoxide it would be realistic to assume that these statistics *should* be a function of time because their sources and sinks are correlated temporally. If we imagine a smokestack emitting CO at a fixed location, the CO measured in the future is highly correlated with the amount of CO in the present. This brings us to our next main point which is that another reason why our modes cannot be interpreted in terms of individual sources and sinks lies in the non-stationarity of the MOPITT dataset which remains unaccounted for in standard EOF analysis and is a fact we will now show.

As shown in Fig. 10, the autocorrelation for the original MOPITT time series as well as each component from the seasonal decomposition do not decay to zero but instead show periodic variation that persists across time. The MOPITT series has a positive correlation that peaks for every year exactly and shows two negative peaks, both of which also reoccur at intervals of 1 year. We could therefore consider the MOPITT autocorrelation as being a sum of three periodic components. The long-term and seasonal autocorrelation are simpler functions and both have a single periodic component that peak exactly once per year, and 6 months respectively. The residual autocorrelation like MOPITT is a sum of periodic processes. It is important to note that we *cannot* force the dataset to become stationary by removing an annual cycle, because the autocorrelation represents a superposition of multiple periodic processes. We may therefore consider both the long-term and seasonal series to be wide-sense cyclostationary processes since they have periodic temporal mean and a periodic temporal autocorrelation. Because the MOPITT series and residual series are composed of multiple periodic processes, we may consider them to be wide-sense polycyclostationary. We conclude this discussion by recommending the use of cyclostationary EOF (CSEOF) analysis on either the original time series or on the seasonal modes which is a method that is well suited for datasets that possess autocorrelation functions of this type.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


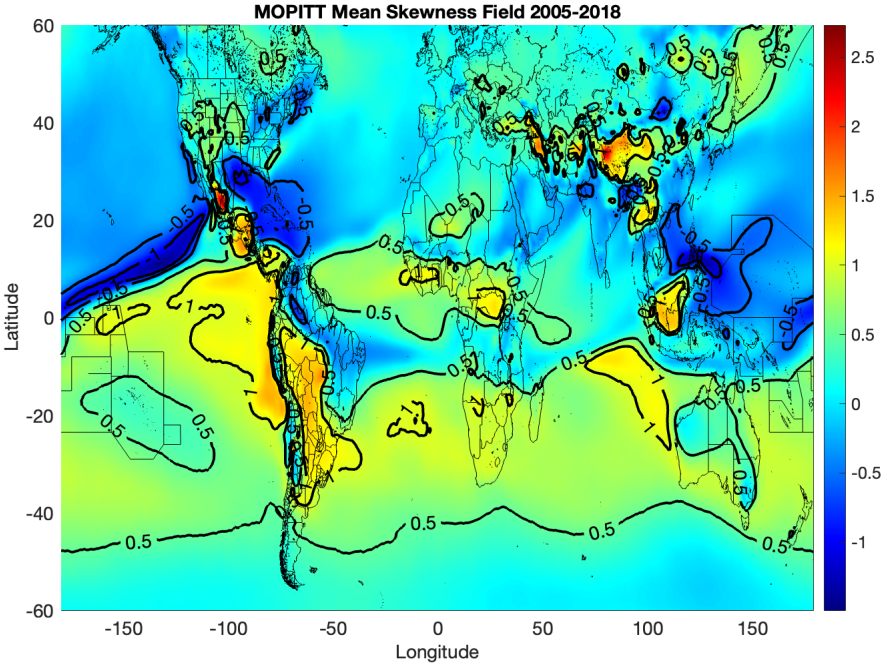


**Figure 9.** A spatial plot showing the mean skewness field of MOPITT CO from 2005-2018. Contour lines are shown to highlight skewness values of -1.0, -0.5, 0.5, and 1.0 indicating regions that are moderately or heavily skewed.

## 4 Conclusions

In this study, we investigated the use of EOF analysis and its variants to assess whether source signatures of CO can be identified from observed CO abundance. We hypothesize that the spatial and temporal differences are linked to surface emission patterns, which are intertwined with CO from methane and non-methane hydrocarbons, atmospheric transport, and reactions with OH. Our results highlight the dominant modes of variability in MOPITT CO columns from 2005 to 2018, with a focus on examining the methodological rationale (limitations) underlying these findings rather than provide physical interpretations on these patterns.

First, our results show that the EOF modes of MOPITT CO columns are not domain-dependent as generating regional domains produced the same spatial patterns as those produced using a global domain. We have also shown that the resulting modes of variability are highly dependent on each other in time. By examining the seasonal skewness, we verify that even though the principal component time series are constructed to be uncorrelated, they are not independent because the dataset is non-Gaussian and therefore they do not correspond to potential sources and sinks.

Spectral analysis for the principal components was used to compute the total accumulated power for each of the modes and to show the dominant time scale for each mode. This verifies that lower EOF modes correspond to the effects of long-term variability in the data while higher modes correspond to the effects of shorter term variability. Because each spectrum is

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


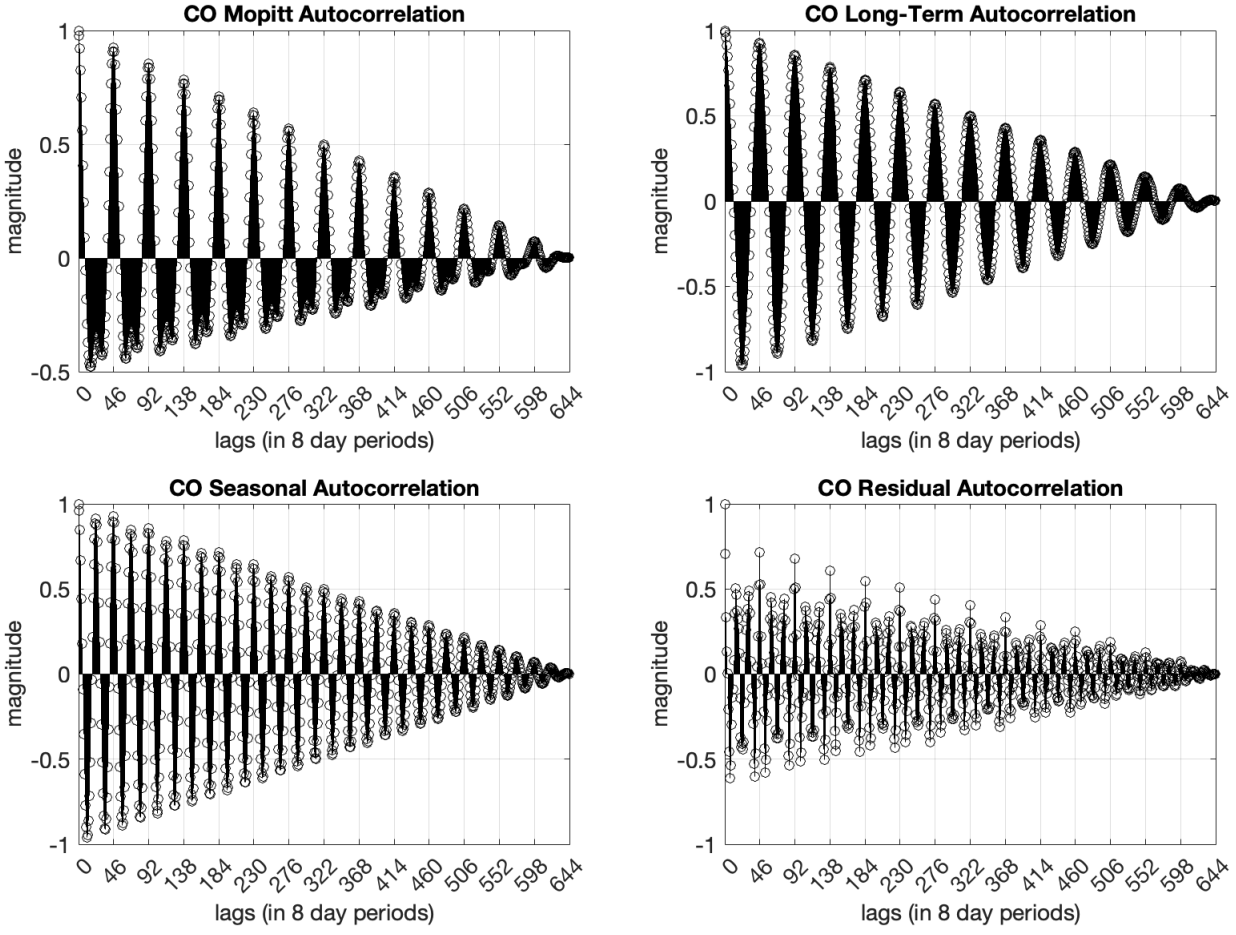


**Figure 10.** The autocorrelation for MOPITT CO as well as the autocorrelation of the long-term, seasonal, and residual modes after using an SSA decomposition plotted as a function of lag time (measured in 8-day periods).

continuous with a significant spread of varying frequencies, we see that each mode is produced using a combination of many different time scales rather than an individual.

The first unrotated MOPITT pattern as well as the first long-term pattern are indicative of the annual/semiannual cycle for CO. Because each mode is highly dependent on time, it is very difficult for us to attribute the variability of higher modes for CO with the individual effects of a source or sink and we therefore cannot offer a complete physical interpretations beyond the understanding of what time scale contributed the most to its variability. Part of the reason for having dependent modes in time could be due to the fact that we chose to use total column CO data which may be averaging the modes of variability together. In future studies, it seems worthwhile to examine the statistics for CO throughout multiple levels of the atmosphere to see whether it is Gaussian or not and whether the spectrum shows behavior that is more time-independent.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


As we showed in the autocorrelation functions in Fig. 10, the MOPITT CO time series can be considered to be wide-sense polycyclostationary and as such we recommend using CSEOF analysis which is a method that is able to handle cyclostationary time series data by assuming that the covariance is a sum of periodic functions of time. This is unlike standard EOF analysis, which uses the major simplification to assume the covariance is stationary (constant in time). In practice, physical modes often have statistics that evolve in time and are an important feature that should be accounted for in the variability of time series data. The use of CSEOF analysis is therefore a more appropriate method to apply to the MOPITT time series data in future analysis and may be able to separate modes of variability based on the individual sources and sinks of CO.

If this was indeed possible, then this could have future implications about the ability of EOF methods to identify biases in chemical transport models. For example, Covariance Discrimination Analysis (CDA) is a method that is able to tell statistical differences between datasets based on the information content of their covariance matrices (or equivalently their principal components). If we were able to separate physical modes of variability for CO using CSEOFs, we may then be able to use CDA to identify statistical differences between model and satellite variability of CO and then connect differences in each mode to biases in the attribution of individual sources including fossil fuel and biomass burning, as well as methane and hydrocarbon oxidation. This would be very useful to identify which sources contribute the highest bias when attempting to model the chemical transport of CO.

Finally, we recognize that these limitations result from the convolution of atmospheric transport and mixing's highly diffusive nature with the reactive chemistry of CO and OH, compounded by the limited information content of columnar CO data. These factors make it challenging to distinguish shorter-scale regional source signatures from the longer-scale background CO. Using a multi-species and multi-scale approach to leverage synergies between various data sources, combined with advanced algorithms—such as the previously mentioned pattern recognition techniques, dimensionality reduction, and machine learning—can greatly improve our ability to distinguish these signatures and attribute them to physical processes. This is especially true with the recent availability of higher-resolution atmospheric composition datasets.

*Data availability.* The MOPITT V8 joint TIR-NIR L3 datasets can be accessed at the following URLs : https://asdc.larc.nasa.gov/project/MOPITT/MOP03J_8.

## Appendix A

### A1 Eigenvectors of Temporal and Spatial Covariance Matrices

Suppose $F$ is an $n$ x $m$ matrix where measurements are taken over $n$ points in time and over a spatial grid with $m$ points. We define the $m$ x $m$ spatial covariance matrix as $R = F^T F$ and the $n$ x $n$ temporal covariance matrix as $L = FF^T$. In the case that the number of gridpoints is very large and $n >> m$ it becomes computationally infeasible to perform an eigendecomposition on $R$. It is instead desirable to decompose the much smaller covariance matrix $L$ and project the resulting PC time series in order to compute the EOF spatial patterns which are eigenvectors of R.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


We begin by performing an eigendecomposition on the spatial covariance matrix $R$ to yield the eigenvector matrix $C$ which corresponds to the EOF spatial patterns and satisfies:

$$RC = C\Lambda$$

Multiplying on the LHS yields:

$$FRC = FF^T FC = LFC \tag{A1}$$

and multiplication by F on the RHS then gives us:

$$L\,(FC) = (FC)\,\Lambda \tag{A2}$$

And defining $A = FC$ we have:

$$LA = A\Lambda \tag{A3}$$

which shows that $A$ is the matrix of eigenvectors for $L$, while the eigenvalues $\Lambda$ are the same as for $R$. We note that while the eigenvectors $C$ of $R$ are size $m$ x $m$ and represent a spatial pattern, the eigenvector matrix $A$ of $L$ is size $n$ x $n$ and represents the principal component (PC) time series constructed by projecting the original matrix $F$ onto the EOF patterns $C$.

We are able to reconstruct the original matrix $F$ using the EOF spatial patterns and the un-normalized PC time series as:

$$F = AC^T \tag{A4}$$

The fact that $A$ is not normalized is apparent from the PCA property of $R$:

$$A^T A = (FC)^T FC = C^T F^T FC = C^T RC = C^T C\Lambda CC^T = \Lambda \tag{A5}$$

which shows us that $||A^T A|| = ||\Lambda|| \neq I$. We can therefore define a normalized PC time series $\Phi$ by dividing each term by $\lambda^{1/2}$:

$$\Phi = FC\Lambda^{-1/2} = A\Lambda^{-1/2} \tag{A6}$$

then it follows that $\Phi$ is orthogonal:

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


$$\Phi^T\Phi = \left(A\Lambda^{-1/2}\right)^T A\Lambda^{-1/2} = \left(A^T A\right)\Lambda^{-1} = \Lambda\Lambda^{-1} = I \tag{A7}$$

This expression of $\Phi$ is important, because it allows us to express the singular value decomposition of the matrix $F$ as:

$$F = \Phi\Lambda^{1/2}C^T = \Phi\Sigma C^T \tag{A8}$$

If we use the singular value decomposition for $F$ in the covariance matrices then we recover the eigendecompositions for both $R$ and $L$:

$$R = F^T F = \left(\Phi\Sigma C^T\right)^T \Phi\Sigma C^T = C\Lambda C^T \tag{A9}$$

$$L = FF^T = \left(\Phi\Sigma C^T\right)^T \Phi\Sigma C^T = \Phi\Lambda\Phi^T \tag{A10}$$

so we see that orthogonal eigenvector matrices for $R$ and $L$ are the normalized EOF spatial patterns $C$ and the PC time series $\Phi$, respectively.

Therefore, when $n >> m$ and we cannot easily compute $C$, our strategy is to perform eigendecomposition on $L$ to find $\Phi$, and then determine $C$ by rearranging the SVD expression:

$$C = \left(F^T\Phi\right)\Lambda^{-1/2} = G\Lambda^{-1/2} \tag{A11}$$

where $G = F^T\Phi$. The original matrix $F$ can then be reconstructed using the full rank expression for the SVD. If we are using correlation matrices so that $F$ has zero mean and has been scaled by its standard deviation, the original matrix in physical space $F^*$ is given by:

$$F^* = F\,\text{std(F)} + \text{mean}(F) = \left(\Phi\Sigma C^T\right)\text{std}(F) + \text{mean}(F). \tag{A12}$$

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


### A2 Equivalence of Determining EOFs and PCs Using Eigenvector Decomposition and SVD

In the case of eigenvector decomposition, we will describe cases. When we have $n >> m$ so the spatial grid is large then $L = FF^T$ has size $m$ x $m$ and as described before we compute the eigendecomposition of $L$ to determine the PC time series which are the column vectors of $\Phi$. In Matlab we would simply use $[\Phi, \Lambda] = \text{eig}(L)$ to compute the normalized eigenvectors and the eigenvalues. We can then compute the matrix $G$ and find the EOF spatial patterns which are the column vectors of $C$ (or the row vectors of $C^T$) as shown in (11).

In the case that $m >> n$ so the time points are large, then $R = F^T F$ has size $n$ x $n$ and we first compute the EOF spatial patterns $C$ which are the normalized eigenvectors of $R$, and are found in Matlab as $[C, \Lambda] = \text{eig}(R)$. We then determine the PC time series $\Phi$ using equation (6).

When $m \approx n$ then either method from above would be appropriate to compute the EOFs and PCs. It would also be appropriate (and equivalent) to compute the singular value decomposition of $F$ using equation (8) which is determined in matlab as $[\Phi, \Sigma, C] = \text{svd}(F)$, where the singular values $\Sigma$ are the square roots of the eigenvalues $\Lambda$. In this case, the left singular vectors are the normalized PC time series, and the right singular vectors are the normalized EOF spatial patterns.

| | $R = F'F$ | $L = FF$' | SVD |
|---|---|---|---|
| Matlab | $[C, \Lambda] = eig(R)$ | $[\Phi, \Lambda] = eig(L)$ | $[\Phi, \Sigma, C] = svd(F)$ |
| PC (un-normalized) | $FC$ | $\Phi\Lambda^{1/2}$ | $\Phi\Sigma$ |
| PC (normalized) | $FC\Lambda^{-1/2}$ | $\Phi$ | $\Phi$ |
| EOF | $C$ | $F^T\Phi\Lambda^{-1/2}$ | $C$ |
| $F$ (un-normalized) | $(\text{PC})\,(\text{EOF})^T$ | $(\text{PC})\,(EOF)^T$ | $(\text{PC})\,(\text{EOF})^T$ |
| $F$ (normalized) | $(\text{PC})\,\Lambda^{1/2}\,(\text{EOF})^T$ | $(\text{PC})\,\Lambda^{1/2}\,(\text{EOF})^T$ | $(\text{PC})\,\Sigma\,(\text{EOF})^T$ |

*Author contributions.* Conceptualization: JMM,AFA Jr.; investigation: JMM; methodology: JMM, CR, AFA Jr.; formal analysis: JMM; data curation: JMM, CR; validation: JMM; visualization: JMM; supervision: AFA Jr.; writing (original draft preparation): JMM; writing (review and editing): JMM, BG, CR, AFA Jr.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This research work is supported by NASA ACMAP (grant no. 80NSSC19K0947). We also acknowledge the insights and expertise from the National Center for Atmospheric Research (NCAR) sponsored by the National Science Foundation (NSF)) for assisting this study.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


## References

Andela, N., Morton, D. C., Giglio, L., Chen, Y., van der Werf, G. R., Kasibhatla, P. S., DeFries, R. S., Collatz, G. J., Hantson, S., Kloster, S., Bachelet, D., Forrest, M., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Yue, C., and Randerson, J. T.: A human-driven decline in global burned area, Science, 356, 1356–1362, https://doi.org/10.1126/science.aal4108, 2017.

Björnsson, H. and Venegas, S. A.: A manual for EOF and SVD analyses of climate data, 1997.

Bryant, H. and Stevenson, D.: The influence of hydrogen on carbon monoxide in the troposphere, Weather, 79, 190–195, https://doi.org/10.1002/wea.4567, 2024.

Buchholz, R. R., Deeter, M. N., Worden, H. M., Gille, J., Edwards, D. P., Hannigan, J. W., Jones, N. B., Paton-Walsh, C., Griffith, D. W. T., Smale, D., Robinson, J., Strong, K., Conway, S., Sussmann, R., Hase, F., Blumenstock, T., Mahieu, E., and Langerock, B.: Validation of MOPITT carbon monoxide using ground-based Fourier transform infrared spectrometer data from NDACC, Atmospheric Measurement Techniques, 10, 1927–1956, https://doi.org/10.5194/amt-10-1927-2017, publisher: Copernicus GmbH, 2017.

Buchholz, R. R., Hammerling, D., Worden, H. M., Deeter, M. N., Emmons, L. K., Edwards, D. P., and Monks, S. A.: Links Between Carbon Monoxide and Climate Indices for the Southern Hemisphere and Tropical Fire Regions, Journal of Geophysical Research: Atmospheres, 123, 9786–9800, https://doi.org/10.1029/2018JD028438, 2018.

Buchholz, R. R., Worden, H. M., Park, M., Francis, G., Deeter, M. N., Edwards, D. P., Emmons, L. K., Gaubert, B., Gille, J., Martínez-Alonso, S., Tang, W., Kumar, R., Drummond, J. R., Clerbaux, C., George, M., Coheur, P.-F., Hurtmans, D., Bowman, K. W., Luo, M., Payne, V. H., Worden, J. R., Chin, M., Levy, R. C., Warner, J., Wei, Z., and Kulawik, S. S.: Air pollution trends measured from Terra: CO and AOD over industrial, fire-prone, and background regions, Remote Sensing of Environment, 256, 112 275, https://doi.org/10.1016/j.rse.2020.112275, 2021.

Buell, C. E.: The Number of Significant Proper Functions of Two-Dimensional Fields, Journal of Applied Meteorology and Climatology, 17, 717–722, https://doi.org/10.1175/1520-0450(1978)017<0717:TNOSPF>2.0.CO;2, 1978.

Chen, S., Xu, L., Zhang, Y., Chen, B., Wang, X., Zhang, X., Zheng, M., Chen, J., Wang, W., Sun, Y., Fu, P., Wang, Z., and Li, W.: Direct observations of organic aerosols in common wintertime hazes in North China: insights into direct emissions from Chinese residential stoves, Atmospheric Chemistry and Physics, 17, 1259–1270, https://doi.org/10.5194/acp-17-1259-2017, 2017.

Cheng, M., Zhi, G., Tang, W., Liu, S., Dang, H., Guo, Z., Du, J., Du, X., Zhang, W., Zhang, Y., and Meng, F.: Air pollutant emission from the underestimated households' coal consumption source in China, Science of The Total Environment, 580, 641–650, https://doi.org/https://doi.org/10.1016/j.scitotenv.2016.12.143, 2017.

Cicerone, R.: The Changing Atmosphere cds. FS Rowland and ISA Isaksen, pp. 49-61 John Wiley & Sons Ltd. S. Bernhard, Dahlem Konferenzen, 1988, The Changing Atmosphere, 2, 49, 1988.

Conte, L., Szopa, S., Séférian, R., and Bopp, L.: The oceanic cycle of carbon monoxide and its emissions to the atmosphere, Biogeosciences, 16, 881–902, https://doi.org/10.5194/bg-16-881-2019, 2019.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martínez-Alonso, S., Worden, H. M., Ziskin, D., and Andreae, M. O.: Radiance-based retrieval bias mitigation for the MOPITT instrument: the version 8 product, Atmospheric Measurement Techniques, 12, 4561–4580, https://doi.org/10.5194/amt-12-4561-2019, publisher: Copernicus GmbH, 2019.

Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J.: A review of 9-year performance and operation of the MOPITT instrument, Advances in Space Research, 45, 760–774, https://doi.org/10.1016/j.asr.2009.11.019, 2010.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Edwards, D. P., Emmons, L. K., Hauglustaine, D. A., Chu, D. A., Gille, J. C., Kaufman, Y. J., Pétron, G., Yurganov, L. N., Giglio, L., Deeter, M. N., Yudin, V., Ziskin, D. C., Warner, J., Lamarque, J.-F., Francis, G. L., Ho, S. P., Mao, D., Chen, J., Grechko, E. I., and Drummond, J. R.: Observations of carbon monoxide and aerosols from the Terra satellite: Northern Hemisphere variability, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2004JD004727, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2004JD004727, 2004.

Edwards, D. P., Emmons, L. K., Gille, J. C., Chu, A., Attié, J.-L., Giglio, L., Wood, S. W., Haywood, J., Deeter, M. N., Massie, S. T., Ziskin, D. C., and Drummond, J. R.: Satellite-observed pollution from Southern Hemisphere biomass burning, 111, https://doi.org/10.1029/2005JD006655, 2006.

Espinosa, F., Bartolomé, A. B., Hernández, P. V., and Rodriguez-Sanchez, M. C.: Contribution of Singular Spectral Analysis to Forecasting and Anomalies Detection of Indoors Air Quality, Sensors, 22, 3054, https://doi.org/10.3390/s22083054, 2022.

Feng, S., Jiang, F., Wu, Z., Wang, H., Ju, W., and Wang, H.: CO Emissions Inferred From Surface CO Observations Over China in December 2013 and 2017, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 808, https://doi.org/10.1029/2019JD031808, 2020.

Fiore, A. M., Naik, V., Spracklen, D. V., Steiner, A., Unger, N., Prather, M., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsøren, S., Eyring, V., Folberth, G. A., Ginoux, P., Horowitz, L. W., Josse, B., Lamarque, J.-F., MacKenzie, I. A., Nagashima, T., O'Connor, F. M., Righi, M., Rumbold, S. T., Shindell, D. T., Skeie, R. B., Sudo, K., Szopa, S., Takemura, T., and Zeng, G.: Global air quality and climate, Chemical Society Reviews, 41, 6663–6683, https://doi.org/10.1039/C2CS35095E, publisher: The Royal Society of Chemistry, 2012.

Gaubert, B., Arellano Jr., A. F., Barré, J., Worden, H. M., Emmons, L. K., Tilmes, S., Buchholz, R. R., Vitt, F., Raeder, K., Collins, N., Anderson, J. L., Wiedinmyer, C., Martinez Alonso, S., Edwards, D. P., Andreae, M. O., Hannigan, J. W., Petri, C., Strong, K., and Jones, N.: Toward a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition, Journal of Geophysical Research: Atmospheres, 121, 7310–7343, https://doi.org/https://doi.org/10.1002/2016JD024863, 2016.

Gaubert, B., Worden, H. M., Arellano, A. F. J., Emmons, L. K., Tilmes, S., Barré, J., Martinez Alonso, S., Vitt, F., Anderson, J. L., Alkemade, F., Houweling, S., and Edwards, D. P.: Chemical Feedback From Decreasing Carbon Monoxide Emissions, Geophysical Research Letters, 44, 9985–9995, 2017.

Gaubert, B., Emmons, L. K., Raeder, K., Tilmes, S., Miyazaki, K., Arellano Jr., A. F., Elguindi, N., Granier, C., Tang, W., Barré, J., Worden, H. M., Buchholz, R. R., Edwards, D. P., Franke, P., Anderson, J. L., Saunois, M., Schroeder, J., Woo, J.-H., Simpson, I. J., Blake, D. R., Meinardi, S., Wennberg, P. O., Crounse, J., Teng, A., Kim, M., Dickerson, R. R., He, H., Ren, X., Pusede, S. E., and Diskin, G. S.: Correcting model biases of CO in East Asia: impact on oxidant distributions during KORUS-AQ, Atmospheric Chemistry and Physics, 20, 14 617–14 647, https://doi.org/10.5194/acp-20-14617-2020, 2020.

George, M., Clerbaux, C., Bouarar, I., Coheur, P.-F., Deeter, M. N., Edwards, D. P., Francis, G., Gille, J. C., Hadji-Lazaro, J., Hurtmans, D., Inness, A., Mao, D., and Worden, H. M.: An examination of the long-term CO records from MOPITT and IASI: comparison of retrieval methodology, Atmospheric Measurement Techniques, 8, 4313–4328, https://doi.org/10.5194/amt-8-4313-2015, 2015.

Golyandina, N. and Zhigljavsky, A.: Singular Spectrum Analysis for Time Series, SpringerBriefs in Statistics, https://doi.org/10.1007/978-3-642-34913-3_2, 2013.

Gruszczynska, M., Rosat, S., Klos, A., Gruszczynski, M., and Bogusz, J.: Multichannel Singular Spectrum Analysis in the Estimates of Common Environmental Effects Affecting GPS Observations, in: Geodynamics and Earth Tides Observations from Global to Micro

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Scale, edited by Braitenberg, C., Rossi, G., and Geodynamics and Earth Tides Editor group, pp. 211–228, https://doi.org/10.1007/978-3-319-96277-1_17, 2019.

Hannachi, A.: Patterns Identification and Data Mining in Weather and Climate, Springer Atmospheric Sciences, Springer International Publishing, https://doi.org/10.1007/978-3-030-67073-3, 2021.

Hannachi, A., Jolliffe, I. T., and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, International Journal of Climatology, 27, 1119–1152, https://doi.org/10.1002/joc.1499, 2007.

Hartmann, D.: ATM S 552 notes, https://atmos.uw.edu/~dennis/552_Notes_ftp.html, 2016.

Hassani, H.: Singular Spectrum Analysis: Methodology and Comparison, Journal of Data Science, 5, 239–257, https://doi.org/10.6339/JDS.2007.05(2).396, 2007.

Holloway, T., Levy II, H., and Kasibhatla, P.: Global distribution of carbon monoxide, Journal of Geophysical Research: Atmospheres, 105, 12 123–12 147, https://doi.org/10.1029/1999JD901173, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1999JD901173, 2000.

Horel, J. D.: A Rotated Principal Component Analysis of the Interannual Variability of the Northern Hemisphere 500 mb Height Field, Monthly Weather Review, 109, 2080–2092, https://doi.org/10.1175/1520-0493(1981)109<2080:ARPCAO>2.0.CO;2, place: Boston MA, USA Publisher: American Meteorological Society, 1981.

Kaiser, H.: The Varimax Criterion For Analytic Rotation in Factor Analysis, Psychometrika, 23, 187–200, 1958.

Kong, L., Tang, X., Zhu, J., Wang, Z., Fu, J. S., Wang, X., Itahashi, S., Yamaji, K., Nagashima, T., Lee, H.-J., Kim, C.-H., Lin, C.-Y., Chen, L., Zhang, M., Tao, Z., Li, J., Kajino, M., Liao, H., Wang, Z., Sudo, K., Wang, Y., Pan, Y., Tang, G., Li, M., Wu, Q., Ge, B., and Carmichael, G. R.: Evaluation and uncertainty investigation of the $NO_2$, CO and $NH_3$ modeling over China under the framework of MICS-Asia III, Atmospheric Chemistry and Physics, 20, 181–202, https://doi.org/10.5194/acp-20-181-2020, 2020.

Lehr, C. and Hohenbrink, T. L.: Technical Note: An illustrative introduction to the domain dependence of spatial Principal Component patterns, Hydrology and Earth System Sciences Discussions, pp. 1–32, https://doi.org/10.5194/hess-2024-172, 2024.

Li, J., Carlson, B. E., and Lacis, A. A.: A study on the temporal and spatial variability of absorbing aerosols using Total Ozone Mapping Spectrometer and Ozone Monitoring Instrument Aerosol Index data, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011278, 2009.

Li, J., Carlson, B. E., and Lacis, A. A.: Application of spectral analysis techniques in the intercomparison of aerosol data: 1. An EOF approach to analyze the spatial-temporal variability of aerosol optical depth using multiple remote sensing data sets, Journal of Geophysical Research: Atmospheres, 118, 8640–8648, https://doi.org/10.1002/jgrd.50686, 2013.

Li, M., Zhang, Q., Kurokawa, J.-I., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmospheric Chemistry and Physics, 17, 935–963, https://doi.org/10.5194/acp-17-935-2017, 2017.

Lin, C., Cohen, J. B., Wang, S., and Lan, R.: Application of a combined standard deviation and mean based approach to MOPITT CO column data, and resulting improved representation of biomass burning and urban air pollution sources, Remote Sensing of Environment, 241, 111 720, https://doi.org/10.1016/j.rse.2020.111720, 2020a.

Lin, C., Cohen, J. B., Wang, S., Lan, R., and Deng, W.: A new perspective on the spatial, temporal, and vertical distribution of biomass burning: quantifying a significant increase in CO emissions, Environmental Research Letters, 15, 104 091, https://doi.org/10.1088/1748-9326/abaa7a, 2020b.

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Macias, D., Stips, A., and Garcia-Gorriz, E.: Application of the Singular Spectrum Analysis Technique to Study the Recent Hiatus on the Global Surface Temperature Record, PLOS ONE, 9, 1–7, https://doi.org/10.1371/journal.pone.0107222, 2014.

Martínez-Alonso, S., Deeter, M., Worden, H., Borsdorff, T., Aben, I., Commane, R., Daube, B., Francis, G., George, M., Landgraf, J., et al.: 1.5 years of TROPOMI CO measurements: comparisons to MOPITT and ATom, Atmospheric Measurement Techniques, 13, 4841–4864, 2020.

Monahan, A. H., Fyfe, J. C., Ambaum, M. H. P., Stephenson, D. B., and North, G. R.: Empirical Orthogonal Functions: The Medium is the Message, Journal of Climate, 22, 6501 – 6514, https://doi.org/10.1175/2009JCLI3062.1, 2009.

Mottungan, K., Roychoudhury, C., Brocchi, V., Gaubert, B., Tang, W., Mirrezaei, M. A., McKinnon, J., Guo, Y., Griffith, D. W. T., Feist, D. G., Morino, I., Sha, M. K., Dubey, M. K., De Mazière, M., Deutscher, N. M., Wennberg, P. O., Sussmann, R., Kivi, R., Goo, T.-Y., Velazco, V. A., Wang, W., and Arellano Jr., A. F.: Local and regional enhancements of $CH_4$, CO, and $CO_2$ inferred from TCCON column measurements, Atmospheric Measurement Techniques, 17, 5861–5885, https://doi.org/10.5194/amt-17-5861-2024, 2024.

Murayama, S., Taguchi, S., and Higuchi, K.: Interannual Variation in the Atmospheric CO<sub>2</sub> Growth Rate: Role of Atmospheric Transport in the Northern Hemisphere, Journal of Geophysical Research Atmospheres, https://doi.org/10.1029/2003jd003729, 2004.

Myhre, G. and Shindell, D.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2013.

Naik, V., Voulgarakis, A., Fiore, A. M., Horowitz, L. W., Lamarque, J.-F., Lin, M., Prather, M. J., Young, P. J., Bergmann, D., Cameron-Smith, P. J., Cionni, I., Collins, W. J., Dalsøren, S. B., Doherty, R., Eyring, V., Faluvegi, G., Folberth, G. A., Josse, B., Lee, Y. H., MacKenzie, I. A., Nagashima, T., van Noije, T. P. C., Plummer, D. A., Righi, M., Rumbold, S. T., Skeie, R., Shindell, D. T., Stevenson, D. S., Strode, S., Sudo, K., Szopa, S., and Zeng, G.: Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmospheric Chemistry and Physics, 13, 5277–5298, https://doi.org/10.5194/acp-13-5277-2013, 2013.

Nguyen, N. H., Turner, A. J., Yin, Y., Prather, M. J., and Frankenberg, C.: Effects of Chemical Feedbacks on Decadal Methane Emissions Estimates, Geophysical Research Letters, 47, e2019GL085 706, https://doi.org/10.1029/2019GL085706, 2020.

North, G. R., Bell, T. L., Cahalan, R. F., and Moeng, F. J.: Sampling Errors in the Estimation of Empirical Orthogonal Functions, Monthly Weather Review, 110, 699–706, https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2, 1982.

Prather, M. J. and Holmes, C. D.: Overexplaining or underexplaining methane's role in climate change, Proceedings of the National Academy of Sciences, 114, 5324–5326, https://doi.org/10.1073/pnas.1704884114, 2017.

Raman, A. and Arellano, A. F. J.: Spatial and Temporal Variations in Characteristic Ratios of Elemental Carbon to Carbon Monoxide and Nitrogen Oxides across the United States, Environmental Science & Technology, 51, 6829–6838, https://doi.org/10.1021/acs.est.7b00161, 2017.

Richman, M. B.: Rotation of principal components, Journal of Climatology, 6, 293–335, https://doi.org/10.1002/joc.3370060305, 1986.

Saunois, M., Jackson, R. B., Bousquet, P., Poulter, B., and Canadell, J. G.: The growing role of methane in anthropogenic climate change, Environmental Research Letters, 11, 120 207, https://doi.org/10.1088/1748-9326/11/12/120207, 2016.

Shindell, D. T., Faluvegi, G., Stevenson, D. S., Krol, M. C., Emmons, L. K., Lamarque, J.-F., Pétron, G., Dentener, F. J., Ellingsen, K., Schultz, M. G., Wild, O., Amann, M., Atherton, C. S., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Lawrence, M. G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G.,

https://doi.org/10.5194/egusphere-2024-3440
Preprint. Discussion started: 6 January 2025


Savage, N. H., Strahan, S. E., Sudo, K., Szopa, S., Unger, N., van Noije, T. P. C., and Zeng, G.: Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/https://doi.org/10.1029/2006JD007100, 2006.

Silva, S. J. and Arellano, A. F.: Characterizing Regional-Scale Combustion Using Satellite Retrievals of CO, NO2 and CO2, Remote Sensing, 9, 744, https://doi.org/10.3390/rs9070744, 2017.

Silva, S. J., Arellano, A. F., and Worden, H. M.: Toward anthropogenic combustion emission constraints from space-based analysis of urban CO2/CO sensitivity, Geophysical Research Letters, 40, https://doi.org/10.1002/grl.50954, 2013.

Simmons, A. J., Wallace, J. M., and Branstator, G. W.: Barotropic Wave Propagation and Instability, and Atmospheric Teleconnection Patterns, Journal of Atmospheric Sciences, 40, 1363 – 1392, https://doi.org/10.1175/1520-0469(1983)040<1363:BWPAIA>2.0.CO;2, 1983.

Tang, W. and Arellano Jr., A. F.: Investigating dominant characteristics of fires across the Amazon during 2005–2014 through satellite data synthesis of combustion signatures, Journal of Geophysical Research: Atmospheres, 122, 1224–1245, https://doi.org/10.1002/2016JD025216, 2017.

Tang, W., Arellano, A. F., Gaubert, B., Miyazaki, K., and Worden, H. M.: Satellite data reveal a common combustion emission pathway for major cities in China, Atmospheric Chemistry and Physics, 19, 4269–4288, https://doi.org/10.5194/acp-19-4269-2019, 2019.

Toumazou, V. and Cretaux, J.-F.: Using a Lanczos Eigensolver in the Computation of Empirical Orthogonal Functions, Monthly Weather Review, 129, 1243–1250, https://doi.org/10.1175/1520-0493(2001)129<1243:UALEIT>2.0.CO;2, 2001.

Turner, A. J., Frankenberg, C., and Kort, E. A.: Interpreting contemporary trends in atmospheric methane, Proceedings of the National Academy of Sciences, 116, 2805–2813, https://doi.org/10.1073/pnas.1814297116, 2019.

von Storch ad F. Zwiers, H.: Statistical Analysis in Climate Research, Cambridge University Press, virtual edn., 2003.

Wallace, J. and Gutzler, D.: Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter, Monthly Weather Review, 109, 784–812, 1981.

Wiedinmyer, C., Kimura, Y., McDonald-Buller, E. C., Emmons, L. K., Buchholz, R. R., Tang, W., Seto, K., Joseph, M. B., Barsanti, K. C., Carlton, A. G., and Yokelson, R.: The Fire Inventory from NCAR version 2.5: an updated global fire emissions model for climate and chemistry applications, Geoscientific Model Development, 16, 3873–3891, https://doi.org/10.5194/gmd-16-3873-2023, 2023.

Wilks, D. S.: Statistical Method in the Atmospheric Sciences, Elsevier Inc., third edn., 2011.

Yin, Z., Cao, B., and Wang, H.: Dominant patterns of summer ozone pollution in eastern China and associated atmospheric circulations, Atmospheric Chemistry and Physics, 19, 13 933–13 943, https://doi.org/10.5194/acp-19-13933-2019, 2019.

Young, P. J.: Pre-industrial to end 21st century projections of tropospheric ozone from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP, Atmospheric Chemistry and Physics, pp. 2063–2090, 2013.

Zheng, B., Chevallier, F., Yin, Y., Ciais, P., Fortems-Cheiney, A., Deeter, M. N., Parker, R. J., Wang, Y., Worden, H. M., and Zhao, Y.: Global atmospheric carbon monoxide budget 2000–2017 inferred from multi-species atmospheric inversions, Earth System Science Data, 11, 1411–1436, https://doi.org/10.5194/essd-11-1411-2019, 2019.

Zhi, G., Zhang, Y., Sun, J., Cheng, M., Dang, H., Liu, S., Yang, J., Zhang, Y., Xue, Z., Li, S., and Meng, F.: Village energy survey reveals missing rural raw coal in northern China: Significance in science and policy, Environmental Pollution, 223, 705–712, https://doi.org/10.1016/j.envpol.2017.02.009, 2017.