Science of the Total Environment xxx (xxxx) xxx

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Understanding temporal and spatial changes of O_3 or NO_2 concentrations combining multivariate data analysis methods and air quality transport models

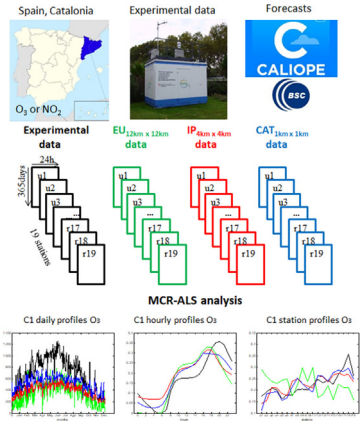
Stefan Platikanov [a], Marta Terrado [b], María Teresa Pay [c], Albert Soret [b], Romà Tauler [a,*]

[a] Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona, 18-26, 08034 Barcelona, Spain
[b] Earth Sciences Department, Barcelona Supercomputing Center (BSC), Jordi Girona, 31, 08034 Barcelona, Spain
[c] Department of Genetics, Microbiology and Statistics, University of Barcelona, Faculty of Biology, Diagonal, 643, 08028 Barcelona, Spain

HIGHLIGHTS

- CALIOPE O_3 and NO_2 concentration forecasts at different spatial resolutions in Catalonia (Spain).
- Experimental and CALIOPE O_3 and NO_2 concentration forecasts are analyzed by the MCR-ALS method
- The accuracy of CALIOPE forecasts depends on the spatial resolution, time and contaminant.
- MCR–ALS describes the most meaningful contributions to NO_2 and O_3 concentration changes.
- Sunlight, season, traffic and local station environment rule different patterns.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 July 2021
Received in revised form 7 October 2021
Accepted 7 October 2021
Available online xxxx

Editor: Philip K. Hopke

Keywords:
Air quality modelling
Chemometrics
MCR-ALS
Ozone
Nitrogen dioxide
CALIOPE system

ABSTRACT

The application of the multivariate curve resolution method to the analysis of temporal and spatial data variability of hourly measured O_3 and NO_2 concentrations at nineteen air quality monitoring stations across Catalonia, Spain, during 2015 is shown. Data analyzed included ground-based experimental measurements and predicted concentrations by the CALIOPE air quality modelling system at three horizontal resolutions (Europe at 12×12 km^2, Iberian Peninsula at 4×4 km^2 and Catalonia at 1×1 km^2). Results obtained in the analysis of these different data sets allowed a better understanding of O_3 and NO_2 concentration changes as a sum of a small number of different contributions related to daily sunlight radiation, seasonal dynamics, traffic emission patterns, and local station environments (urban, suburban and rural). The evaluation of O_3 and NO_2 concentrations predicted by the CALIOPE system revealed some differences among data sets at different spatial resolutions. NO_2 predictions, showed in general a better performance than O_3 predictions for the three model resolutions, specially at urban stations. Our results confirmed that the application of the trilinearity constraint during the multivariate curve resolution factor analysis decomposition of the analyzed data sets is a useful tool to facilitate the understanding of the resolved variability sources.

* Corresponding author.
E-mail address: Roma.Tauler@idaea.csic.es (R. Tauler).

https://doi.org/10.1016/j.scitotenv.2021.150923

1. Introduction

Air pollution is a serious threat to both human and environmental health, being currently one of the most pressing challenges for cities in Europe and around the globe (EEA, 2019; WHO, 2016). In 2016, 91% of the world population was living in places where the World Health Organization (WHO) air quality guideline levels were not met. Having information about air pollution contributions and dynamics is essential to design adequate policies to improve air quality and reduce the negative impacts of pollution on health (European Environmental Agency (EEA), 2011; Nieuwenhuijsen, 2018).

Monitoring and modelling of air pollutants are essential tasks to evaluate the impact from the continuous increase of human activities on the environment and public health. Therefore, current research efforts are mainly focused on the improvement of monitoring and modelling systems, and on accurately forecasting the behavior of hazardous air pollutants in order to understand their origin, transport, geographical distribution and time evolution (Schaap et al., 2015; Pay et al., 2019; Massagué et al., 2019). Ozone (O_3) and nitrogen dioxide (NO_2) are among the major air pollutants directly associated with negative effects on the human health (Bell et al., 2004; WHO, 2013; Vicedo-Cabrera et al., 2020).

Roughly 90% of O_3 is distributed in the stratosphere and around 10% in the troposphere. The origins of O_3 in the troposphere include its migration from the stratosphere and its formation from various photochemical reactions among carbon monoxide (CO), peroxy radicals (generated by the photochemical oxidation of volatile organic compounds, VOCs) and nitrogen oxides (NO_x) (Crutzen, 1974). High temperatures, high solar radiation and low precipitation also favor positively the formation of tropospheric O_3 (Otero et al., 2016). Photochemical and oxidation reactions take place specially when the sunlight interacts with NO_x and VOCs, mostly derived from anthropogenic emissions, thus converting the industrial and traffic emissions in additional contributions of O_3 formation. In these urban and industrialized areas, the concentrations of O_3 sharply peak at sunny middays, because of the rapid interconversion of O_3 and NO_x. Conversely, when the impact of sunlight decreases in the evenings, destruction of O_3 occurs because of its reaction with NO from different activities, such as the evening rush hour car traffic. The circulation of air masses also controls the short and long-range transport of O_3, affecting its lifetime in the atmosphere and its levels in rural and remote areas (Monks et al., 2015; Pay et al., 2019). For example, the transport of precursors emitted in urban and industrialized areas has been shown to cause O_3 production downwind (Holloway et al., 2003; Querol et al., 2017).

NO_2 is also considered to be an important trace gas in the atmosphere. Apart from its participation in the photocatalytic formation of O_3 in the troposphere, NO_2 can also produce secondary aerosols and acid rain (Chan et al., 2015). Generally, NO_2 can have both natural and anthropogenic origins, and most of the anthropogenic contributions are related to fossil fuel combustion, biomass burning, and various types of industrial emissions (Tack et al., 2015; Gratsea et al., 2016; Chan and Chan, 2017). Across industrial and urban areas, the air pollution is largely due to emissions of pollutants from motor vehicles, industrial plants and generation of energy (i.e. power plants) (Khan et al., 2018). Sudden increases in pollutants' concentrations are correlated with the intensive morning and evening car traffic.

The European Directive on Ambient Air Quality and Clean Air (European Commission, 2008) stands for air quality monitoring and modelling as useful tools to understand the dynamics of air pollutants, analyze and forecast air quality, develop plans to reduce emissions and alert the population when health-related episodes occur (EEA, 2011). In line with these objectives, the CALIOPE air quality forecasting system (Baldasano et al., 2008a, 2008b; http://www.bsc.es/caliope/es, last access: 15 March 2021) has been developed to forecast the air quality in Spain. CALIOPE, which runs on high-performance computing platforms, uses multi-model inputs about meteorology, atmospheric composition and emissions, to predict the air quality at three different spatial resolutions: 12×12 km^2 for the European domain (EU12); 4×4 km^2, for the Spanish Iberian Peninsula domain (IP4); and 1×1 km^2 for the Catalonian domain (CAT1); in all cases with 1 h time resolution for the next 48 h.

Air quality modelling requires the use of physicochemical parametric models strongly dependent on accurate prior knowledge about the atmospheric processes that involve air pollutants' transport dispersion and reaction. In the CALIOPE system, all possible local contributions of pollution for a given territorial domain, together with the possible influence from more distant geographical zones - transboundary transport models – should be explicitly considered and, in general, need long computation time, especially when they are run at high spatial resolutions.

Various studies have been conducted with the objective to compare the pollutants' concentrations predicted by CALIOPE with the real experimental concentrations measured by the ground-based monitoring stations (Pay et al., 2010, 2012; Baldasano et al., 2011; Aguilera et al., 2013). Satisfactory results have been observed in reproducing the temporal and spatial distribution of various pollutants, including O_3 and NO_2, at different resolutions (Pay et al., 2014).

In this work, a multivariate data analysis approach based on chemometrics soft modelling methods is used to evaluate the CALIOPE forecasts of O_3 and NO_2 concentrations. Unlike the more traditional parametric approaches, based on complex physicochemical and meteorological forecasting models, the proposed chemometrics approach offers the possibility to directly investigate the hidden information present in the experimental data without the assumption of a physical model. This is achieved by the application of a multilinearity constraint during the factor analysis decomposition of the experimental data arranged in multiway/multimode data structures (Norris et al., 2014; Belis et al., 2019; Malik and Tauler, 2013; Alier et al., 2009; Dadashi et al., 2020).

In this work, we apply a chemometric analysis to both the experimental O_3 and NO_2 concentrations and to the concentrations predicted by the CALIOPE system at three different spatial resolutions (EU12, IP4 and CAT1) during the year 2015 at 19 air quality monitoring stations across Catalonia. The comparison of the results of the analysis of experimental and CALIOPE predicted data allowed us to identify the applications for which the CALIOPE system is already fit-for-use and the aspects that can be improved for a better performance. This study upgrades previous investigations for Europe (Schaap et al., 2015) and for the Iberian Peninsula (Pay et al., 2014), as it includes a full year simulated data at higher spatial resolution (1×1 km^2) for the whole territory of Catalonia.

2. Datasets

2.1. Experimental dataset

Hourly O_3 and NO_2 concentration data were obtained from the Catalan air quality monitoring network (named Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica, XVPCA), which provides a relatively dense geographical coverage of the Catalan territory. The network contributes to the European Environment Information and Observation Network (EIONET; https://www.eionet.europa.eu/, last access: April 2019). For 2015, O_3 and NO_2 measurements were obtained from 19 monitoring stations across Catalonia with a temporal coverage higher than 85% on an hourly basis (Fig. 1). As defined by the European Environment Agency (EU, 2011), depending on the distribution/density of buildings, the area surrounding the station is classified as urban (6 stations, u1-u6 in Fig. 1), suburban (5 stations, s7-s11) and rural (8 stations, r12- r19), from coastal to mountain sites. All measurements were recorded in UTC time.

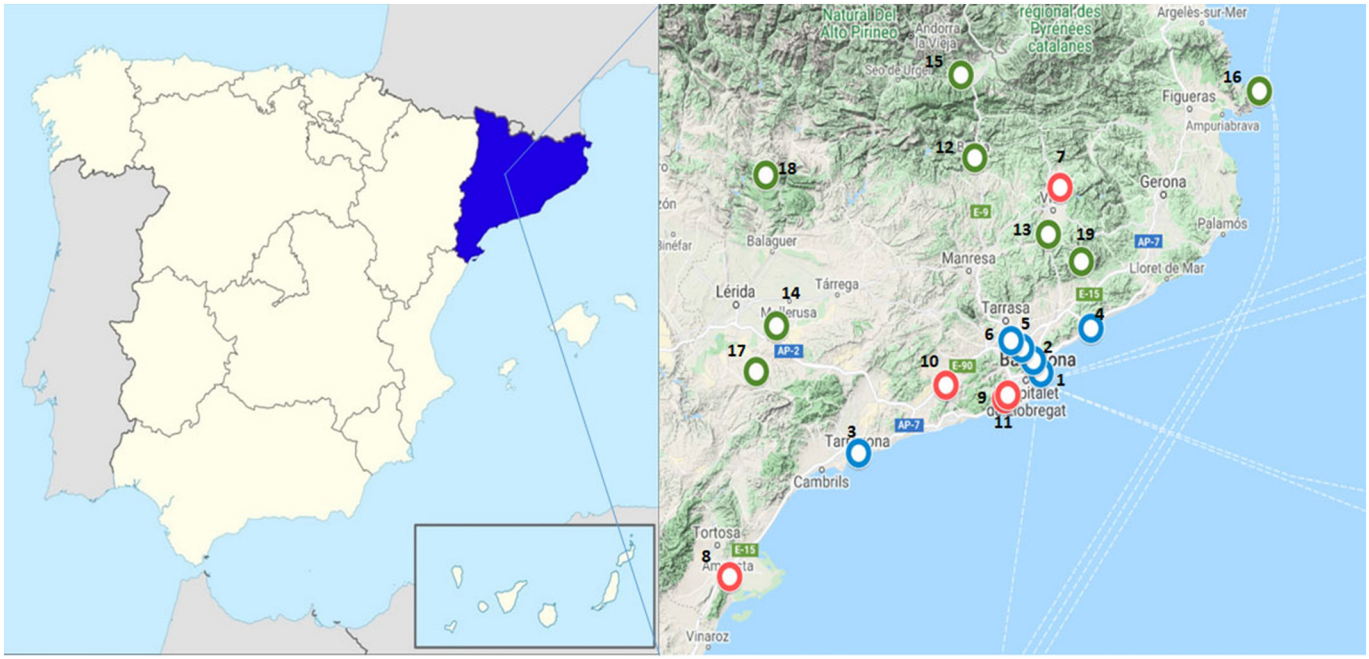


Fig. 1. Location of the EIONET O_3 and NO_2 monitoring stations in Catalonia, Spain. Different icons represent the three environment defined by the EEA. Urban stations: u1 - Cuitadella district (Barcelona), u2 - Vall d'Hebron district (Barcelona), u3 - Tarragona, u4 - Mataró, u5 - Sant Cugat, u6 - Rubí. Suburban stations: s7 - Manlleu, s8 - Amposta, s 9 - Gavà, s10 - Vilafranca de Penedès, s11 - Viladecans. Rural stations: r12 - Berga, r13 - Tona, r14 - Juneda, r15 - Bellver, r16 - Cap de Creus, r17 - Els Torms, r18 - Montsec, r19 - Montseny.

2.2. CALIOPE predicted datasets

The CALIOPE air quality forecasting system (http://www.bsc.es/caliope/es, last access: 15 March 2021) was used to predict the NO_2 and O_3 concentrations over Catalonia during year 2015 using three spatial grid resolutions. The system, described in previous works (Baldasano et al., 2008a, 2011; Pay et al., 2010), is based in numerical modelling including a full and precise description of the processes involved for gas-phase and particulate matter modelling, both important factors to obtain accurate results of air pollutant concentrations in a complex region as Spain. CALIOPE is composed of the HERMES v2.0 emission model (Guevara et al., 2013), the WRF-ARW v3.6 meteorological model (Skamarock and Klemp, 2008), the CMAQ v5.0.2 chemical transport model (Byun and Schere, 2006) and the BSC-DREAM8bv2 mineral dust model (Basart et al., 2012).

CALIOPE first runs over Europe at a 12×12 km^2 horizontal resolution (EU12 domain), then over the Iberian Peninsula at a 4×4 km^2 resolution (IP4 domain) and finally over Catalonia at a 1×1 km^2 resolution (CAT1 domain). The WRF-ARW system provides the meteorological fields required by the CMAQ chemical transport model. In the present work, this system has been configured with 38 sigma layers up to 50 hPa to resolve the troposphere-stratosphere exchanges properly. The planetary boundary layer (PBL) is characterized by approximately 11 layers, and the bottom layer's depth is 39 m. The EU12 domain uses meteorological initial and boundary conditions from the final analyses provided by the National Centers of Environmental Prediction (FNL/NCEP) at a $0.5° \times 0.5°$ resolution. The first 12 h of each meteorological run are treated as cold start, and the next 23 h are provided to the chemical transport model. The CMAQ chemical transport model combines current knowledge in atmospheric science and air quality modelling with multiprocessor computing techniques in an open-source framework to deliver concentration estimates of the main air pollutants. CMAQ v5.0.2 uses the CB05 chemical mechanism (Yarwood et al., 2005), the AERO5 for aerosol modelling, and the in-line photolysis calculation. CALIOPE considers desert dust contribution by means of the BSC-DREAM8bv2, which runs offline at a $0.5° \times 0.5°$ resolution covering Europe, North Africa and the Middle East.

The emissions model HERMESv2.0 uses information and the state-of-the-art methodologies for the estimation of emissions (Baldasano et al., 2008b; Guevara et al., 2013). For the non-Spanish countries included the IP4 domain, HERMESv2.0 processes the original annual European Monitoring and Evaluation Programme (EMEP) gridded emissions (50×50 km^2), performing a Selected Nomenclature for Air Pollution (SNAP) sector-dependent spatial and temporal (1 h) disaggregation as well as a speciation treatment (top-down approach). For the IP4 and CAT1 domains, HERMESv2.0 uses a combination of local information and state-of-the-art methodologies (mostly bottom-up approaches) to estimate anthropogenic and biogenic emissions at a high spatial (1×1 km^2) and temporal (1 h) resolution over the whole Spanish territory, which for the case of the IP4 are then aggregated to a 4×4 km^2 resolution (Guevara et al., 2013). Hourly simulations of NO_2 and O_3 concentrations for year 2015 covering the territory of Catalonia from the EU12, IP4 and CAT1 domains were used for this study. For each domain, the grid cells coinciding with the location of the monitoring stations were used for comparison of simulated results with observations.

2.3. Data arrangement

During 2015, each monitoring station ($k = 1,..,19$) produced two long data vectors, ($\mathbf{d_k}$), one for O_3 and another for NO_2, with 24-h recordings for the 365 days (with dimensions 1×8760). These row vectors for every station were concatenated one on the top of the other giving the data matrix $\mathbf{D^*}$ with size: 19 stations $\times$ 8760 recordings, as shown in Fig. 2a.

In a second type of data arrangement (Fig. 2b), the O_3 or NO_2 data vectors ($\mathbf{d_k}$) for every data station ($k = 1,.0.19$) were first individually arranged in a data matrix $\mathbf{D_k}$ with 365 rows (days) and 24 columns (hours). In the simultaneous analysis of the data from the 19 stations, these individual matrices $\mathbf{D_k}$, were vertically concatenated in a column-wise augmented data matrix, $\mathbf{D_{aug}}$ of size 6935 rows (365 days $\times$ 19 stations) and 24 columns (hours). These data arrangements were applied to the experimental and to the CALIOPE predicted data at the three spatial resolutions (EU12, IP4 and CAT1). Less than 5% of the O_3 and NO_2 measured concentrations were missing in the experimental

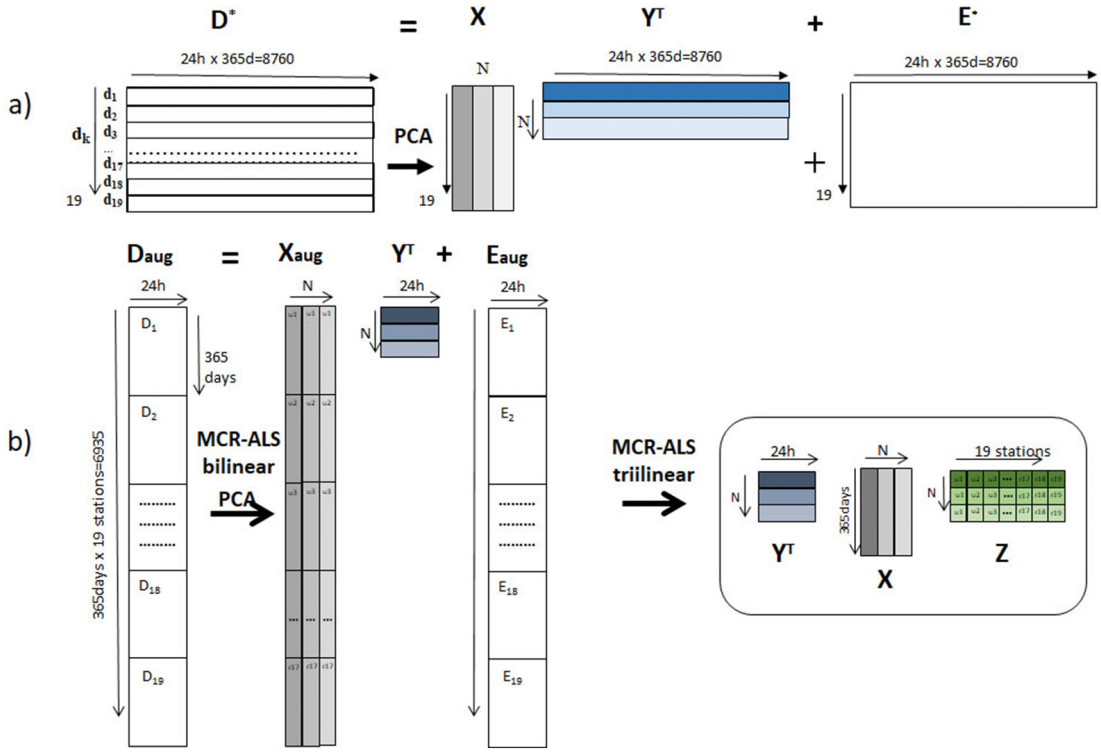


Fig. 2. Data arrangement and application of (a) PCA and (b) MCR-ALS bilinear and trilinear. Hourly O_3 and NO_2 concentrations experimentally observed during one year were arranged in long row vectors, $\mathbf{d_k}$ for the different monitoring stations (k = 1,... 19), in the wide data matrix $\mathbf{D^*}$ with 19 rows and 8760 columns, (as shown in Fig. 2a), in multiple individual station data matrices $\mathbf{D_k}$ with 365 rows and 24 columns and as column-wise augmented data matrix $\mathbf{D_{aug}}$ with 365 × 19 = 8760 rows and 24 columns, as shown in Fig. 2b). PCA, MCR bilinear and MCR-ALS trilinear decompositions are also shown graphically (see Sections 3.1 and 3.2 and Eqs. (1)–(4)).

data. For the CALIOPE data sets, missing values were below 1%. In this work, missing values in the individual data matrices $\mathbf{D_k}$ were replaced by the mean (average) in the same column of the data matrix. Since this number was relatively small, the results will not be affected significantly by the imputation method and more rigorous techniques for data missing imputation (Stanimirova, 2013) were not required. Additionally, the Whittaker smoother (Eilers, 2003) was used for offset correction in O_3 and NO_2 data matrices.

3. Methods

3.1. PCA and bilinear MCR-ALS

Different chemometric methods have been proposed in the literature for the analysis of environmental data. Principal Component Analysis (PCA) (Jolliffe, 2002) is the method most widely used to analyze the variance sources and patterns of variation of environmental datasets. Moreover, the US Environmental Protection Agency (EPA) recommends the application of Positive Matrix Factorization (PMF) for factor analysis in air quality studies (Hopke, 2008; Norris et al., 2014). Similarly to PMF, Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) is a frequently used method in spectrochemical mixture data analysis, which can also be easily extended to the analysis of environmental source apportionment data sets (Alier and Tauler, 2013; Alier et al., 2011). MCR-ALS is a flexible soft-modelling factor analysis tool which allows for the application of natural constraints (see below) and can be easily adapted to the analysis of complex multiway (multimode) data structures, such as three-way and four-way environmental data sets using trilinear and quadrilinear models (Tauler et al., 1998; Smilde et al., 2004; Malik and Tauler, 2013; Marín-García and Tauler, 2020). In this work, PCA and MCR-ALS methods are used to reveal and compare the information present in the experimental and CALIOPE predicted O_3 and NO_2 data sets.

Both, PCA and MCR-ALS methods, are based on a bilinear model that performs the factor decomposition of a two-way data set (i.e. a data table or a data matrix). Eq. (1) summarizes this bilinear model in its element-wise form, while Eq. (2) presents the same model in a matrix linear algebra format:

$$d_{ij} = \sum_{n=1}^{N} x_{i,n} y_{j,n} + e_{i,j} \quad i = 1,..I\ (\text{rows}),\ j = 1, \ldots J(\text{columns}) \tag{1}$$

$$D = X\,Y^T + E \tag{2}$$

In Eq. (1), the individual data values, $\mathbf{d}_{i,j}$ elements (in this case the O_3 or NO_2 concentration values) are decomposed as the sum of a number of components (contributions), $n = 1,..N$, each one of them defined by the product of two factors, $\mathbf{x}_{i,n}$ (scores) and $\mathbf{y}_{j,n}$ (loadings). In addition, the term $\mathbf{e}_{i,j}$ is the residual part of $\mathbf{d}_{i,j}$, which cannot be explained by these N components and accounts for the experimental noise and uncertainties. In Eq. (2), the data matrix, **D**, of dimensions IxJ is decomposed into the scores factor matrix **X** (IxN) and the loadings factor matrix, $\mathbf{Y^T}$ (NxJ). The number of components, N, is selected to explain as much as possible the data variance, while the unexplained small contributions of data variance and experimental noise are in **E**.

In the case of PCA, the components are obtained under the constraints of orthogonality and maximum variance, and they are ranked in descending order of explained variance (see Jolliffe, 2002 for more details). Thus, a reduced set of N (N ≪ J or I) principal components (PCs) are selected, which contain most of the relevant variance (not noise) in the original data matrix. The constraints used in PCA provide mathematical unique solutions, but they imply negative values in the resolved component profiles, which for example in the case of O_3 and NO_2 concentrations have not a direct physical meaning.

Multivariate Curve Resolution (Tauler, 1995) performs a similar bilinear model decomposition as shown in Eq. (2) but in this case different constraints are applied to provide more interpretable and meaningful factor profiles. The bilinear model factor decomposition is performed using an alternating least squares (ALS) optimization under constraints fulfilled naturally by the real data, for example, non-negativity constraints (Bro and de Jong, 1997; de Juan and Tauler, 2003). In the MCR-ALS method, an initial selection of the number of components is required. This can be gathered from PCA or singular value decomposition (Golub and van Loan, 1989) of the analyzed data set, by inspection of the sizes of the singular (or eigen) values of the analyzed data matrix. This selection can also be made by taking into account the amount of explained variance by the successive components, thus avoiding components associated with noise, usually with the smaller singular (eigen) values. Models with different numbers of components can be tested and a final decision is taken considering the data fit and the shapes of the resolved profiles. The ALS algorithm also needs initial estimates of either $\mathbf{X}$ or $\mathbf{Y^T}$ factor matrices. These initial estimates can be obtained from the 'more 'dissimilar' rows or columns of the original data matrix (Windig and Guilment, 1991). Eq. (2) for $\mathbf{D}$ is solved iteratively using an ALS optimization, which updates the solutions (vector profiles in $\mathbf{X}$ and $\mathbf{Y^T}$ matrices) until they fit the data optimally and fulfill the proposed constraints. The profiles in $\mathbf{X}$ and $\mathbf{Y^T}$ matrices are constrained to be non-negative and they are directly interpretable in physical terms. The bilinear data modelling used in Eqs. (1) and (2) can be applied to the experimental and CALIOPE predicted data sets explained in the previous data section (data matrices $\mathbf{D^*}$ and $\mathbf{D_{aug}}$) and in Fig. 2.

3.2. Trilinear MCR-ALS

The bilinear factor decomposition model given for a two-way data matrix can be extended to a three-way dataset, $\mathbf{D}$, using a trilinear model. This model can be expressed individually for every data value as given by Eq. (3).

$$d_{ijk} = \sum_{n=1}^{N} x_{i,n} y_{j,n} z_{k,n} + e_{i,j,k} \tag{3}$$

where $\mathbf{d}_{i,j,k}$ are the individual data values (concentrations of O_3 or NO_2) in the three experimental data modes: the day of the year $i = 1, \ldots,365$, the hour of the day $j = 1, \ldots,24$, and the station $k = 1,..,19$, (urban, u, suburban, s, and rural, r). Data are modeled as the sum of a number of components (contributions), $n = 1,..N$, defined by the product of three factors: $\mathbf{x}_{i,n}$, $\mathbf{y}_{j,n}$, and $\mathbf{z}_{k,n,}$ These factors are related with the three data modes respectively. $\mathbf{e}_{i,j,k}$ is the part of $\mathbf{d}_{ijk}$ not explained by the contribution of these N components. This trilinear model can be written in a matrix form using the decomposition of every individual $\mathbf{D_k}$ data slice (every individual matrix $\mathbf{D_k}$) in Eq. (4).

$$\mathbf{D_k} = \mathbf{X Z_k Y^T} + \mathbf{E_k} \quad \text{for } k = 1, \ldots 19 \tag{4}$$

Under the trilinear model, all individual data matrices, $\mathbf{D_k}(I,J)$ are simultaneously decomposed with the same number of components N and the same daily, $\mathbf{X}$ (I,N) and hourly $\mathbf{Y^T}$ (N,J) profiles. Thus, they differ only in a diagonal matrix $\mathbf{Z_k}(N,N)$ different for every one of the $k = 1,\ldots$ 19 stations (station profiles), which gives the relative amounts of the N components in every data matrix (station), $\mathbf{D_k}$. These N diagonal elements of the $\mathbf{Z_k}$ can also be grouped in the third factor matrix $\mathbf{Z}$ (K,N). Therefore, the proposed trilinear model takes advantage of the natural structure of the analyzed data sets, especially in relation to their different temporal modes (i.e. hourly, daily, seasonal) and to the different type of monitoring stations analyzed simultaneously.

Fig. 2b shows the practical implementation of the trilinear model in the MCR-ALS analysis of a three-way data set. Nineteen individual data sets (one per station), were arranged in the column-wise augmented data matrix $\mathbf{D_{aug}}$ of dimensions 365 days × 19 stations = 6935 rows and 24 hourly measures in columns. The application of the trilinearity constraint implies that the augmented profiles of every component n, $\mathbf{x^n_{aug}}$, of dimensions 6935 × 1, is first refolded (step 1) in a data matrix $\mathbf{X^n_{aug}}$ of dimensions 365 days × 19 stations. Then this matrix is decomposed by SVD considering only the first component into the product of two vector profiles, $\mathbf{x^n}$ (365 × 1) describing the 365 daily profiles, and $\mathbf{z^n}$ (19 × 1) describing differences among the different stations. The application of this trilinearity constraint implies that for every component, the daily changes are described by the same single $\mathbf{x^n}$ vector profile which changes station by station by the corresponding scalar value in $\mathbf{z^n}$. See previous works for a more detailed description of the algorithm used for the practical implementation of the trilinearity constraint in MCR-ALS (Tauler et al., 1998; de Juan and Tauler, 2001). MCR-ALS allows for the implementation of the trilinearity (or multililinearity) constraint separately for every component and it also allows for the relaxation of the synchronization of the component profiles, permitting therefore building mixed bilinear-trilinear models. It is worth also to mention that the application of the trilinearity constraint in MCR-ALS converges fast. MCR-ALS with the trilinearity constraint gives similar results to PARAFAC as we have shown in previous works (Tauler, 1995; Tauler et al., 1998; De Juan and Tauler, 2001), and solves the rotational ambiguity problem associated to MCR-ALS bilinear.

The evaluation of the MCR-ALS fitting results is performed calculating the explained data variances (R^2) using Eq. (5),

$$R^2 = 100 \times \left(1 - \frac{\sum_{k=1}^{19}\sum_{i=1}^{365}\sum_{j=1}^{24}\left(d_{ijk} - \hat{d}_{ijk}\right)^2}{\sum_{k}^{19}\sum_{i=1}^{365}\sum_{j=1}^{24} d_{ijk}^2}\right) \tag{5}$$

where $\mathbf{d}_{ijk}$ are the experimental or CALIOPE predicted O_3 (or NO_2) concentrations, and $\hat{d}_{ijk}$ are the corresponding calculated values by MCR-ALS using either the bilinear (Eqs. (1)–(2)) or the trilinear model (Eqs. (3)–(4)).

3.3. Accuracy of CALIOPE predicted data

The accuracy of the CALIOPE predicted data at three spatial resolutions can be assessed with the calculation of the Root Mean Square Error (RMSE) between the experimental and the CALIOPE model predicted data, averaging them according to Eqs. (6) and (7), respectively.

$$RMSE_{daily} = \sqrt{\frac{\sum_{k=1}^{19}\sum_{j=1}^{24}\left(d_{ijk} - \hat{d}_{ijk}\right)^2}{19x24}} \tag{6}$$

$$RMSE_{hourly} = \sqrt{\frac{\sum_{k=1}^{19}\sum_{i=1}^{365}\left(d_{ijk} - \hat{d}_{ijk}\right)^2}{19 \times 365}} \tag{7}$$

where d_{ijk} are the experimental or CALIOPE predicted O_3 or NO_2 concentrations at day $i = 1, \ldots,365$, hour $j = 1, \ldots,24$ and station $k = 1, \ldots,19$ and $\hat{d}_{ijk}$ are the corresponding experimental concentrations. $RMSE_{daily}$ and $RMSE_{hourly}$ profiles (Eqs. (6) and (7)) are the representative daily and hourly mean profiles averaged for all stations.

To further compare and summarize the prediction accuracy, additional mean scalar values were calculated for all the data values averaging all data values per day, hour and station.

$$RMSE_{avg} = \sqrt{\frac{\sum_{k=1}^{19}\sum_{i=1}^{365}\sum_{j=1}^{24}\left(d_{ijk} - \hat{d}_{ijk}\right)^2}{19 \times 365 \times 24}} \tag{8}$$

This gives a single scalar value $RMSE_{avg}$ for all stations or they can grouped also according to the urban, suburban and rural types for

comparison purposes (see experimental data set Section 2.1, Fig. 1 and Table 2 of results).

3.4. Chemometrics software

All data were transferred to MATLAB computer and visualization environment (The Mathworks, Natick, USA) for their analysis. PCA was performed using PLS 8.5 Toolbox software (Eigenvector Research Ltd., Manson, WA, USA) and MCR-ALS analysis of the different data sets was performed using the MCR toolbox (Jaumot et al., 2015; http://www.mcrals.info/).

4. Results and discussion

In the Supplementary Material section, a brief description of the experimental O_3 and NO_2 concentration data is given. Supplementary Fig. S1 shows the average profiles of O_3 and NO_2 experimental concentrations measured over 365 days (a and b), over the 24 h of the day (c and d), and for the different stations (e and f). Values from urban, suburban and rural stations have been marked in different colors. Supplementary Fig. S2 shows the average concentration profiles of O_3 (a) and NO_2 (b) experimental (black) and predicted by CALIOPE (EU12 green, IP4 red and CAT1 blue) over the 24 h.

4.1. PCA and MCR-ALS results

PCA results of the mean-centered O_3 and NO_2 column-wise augmented experimental data matrices (see Fig. 2a) of the 24 h × 365 days concentration values at the 19 stations in 2015, $\mathbf{D^*_{O3}}$ and $\mathbf{D^*_{NO2}}$, are given in Fig. 3. PC1 and PC2 of the O_3 matrix explain 52.2% and 9.9% of the total data variance, respectively. In the PC1vsPC2 scores plot (Fig. 3a), the 19 stations can be well distinguished in three groups according to the station type. The urban (blue triangles) and rural (green diamonds) type of stations are significantly separated one from another, while the suburban stations (red squares) occupy intermediate positions between them. This separation agrees well with the stations classification given by the local environmental agencies. In the loadings plot (Fig. 3c), the different symbols and colors refer to the different seasons. Loadings on PC1 are mostly positive in all seasons with a homogeneous distribution. In the case of NO_2 (Fig. 3b and d), PC1 and PC2 explain 67.8% and 5.7% of the total data variance. The 19 stations are again distinguished in three different groups for NO_2 concentrations (Fig. 3b). The dominating group in this plot corresponds to urban stations (blue triangles), showing a wider dispersion, while suburban and rural stations are distributed around the axis origin. No clear seasonal pattern can be observed in the loadings plot (see Fig. 3d). In summary, PCA results in Fig. 3 show some differences in the trends of O_3 and NO_2 concentrations according to the type of station whereas these differences are smaller among the different seasons. In a similar way, PCA was also applied to the three CALIOPE predicted data sets at different resolutions, $\mathbf{D^*_{EU12,O3}}$ and $\mathbf{D^*_{EU12,NO2}}$; $\mathbf{D^*_{IP4,O3}}$ and $\mathbf{D^*_{IP4,NO2}}$ and $\mathbf{D^*_{CAT1,O3}}$ and $\mathbf{D^*_{CAT1,NO2}}$. The results of PCA analysis of CALIOPE predicted data are given in more detail in the Supplementary material.

Table 1 summarizes the explained data variances obtained in the analysis of experimental and CALIOPE predicted data by MCR-ALS

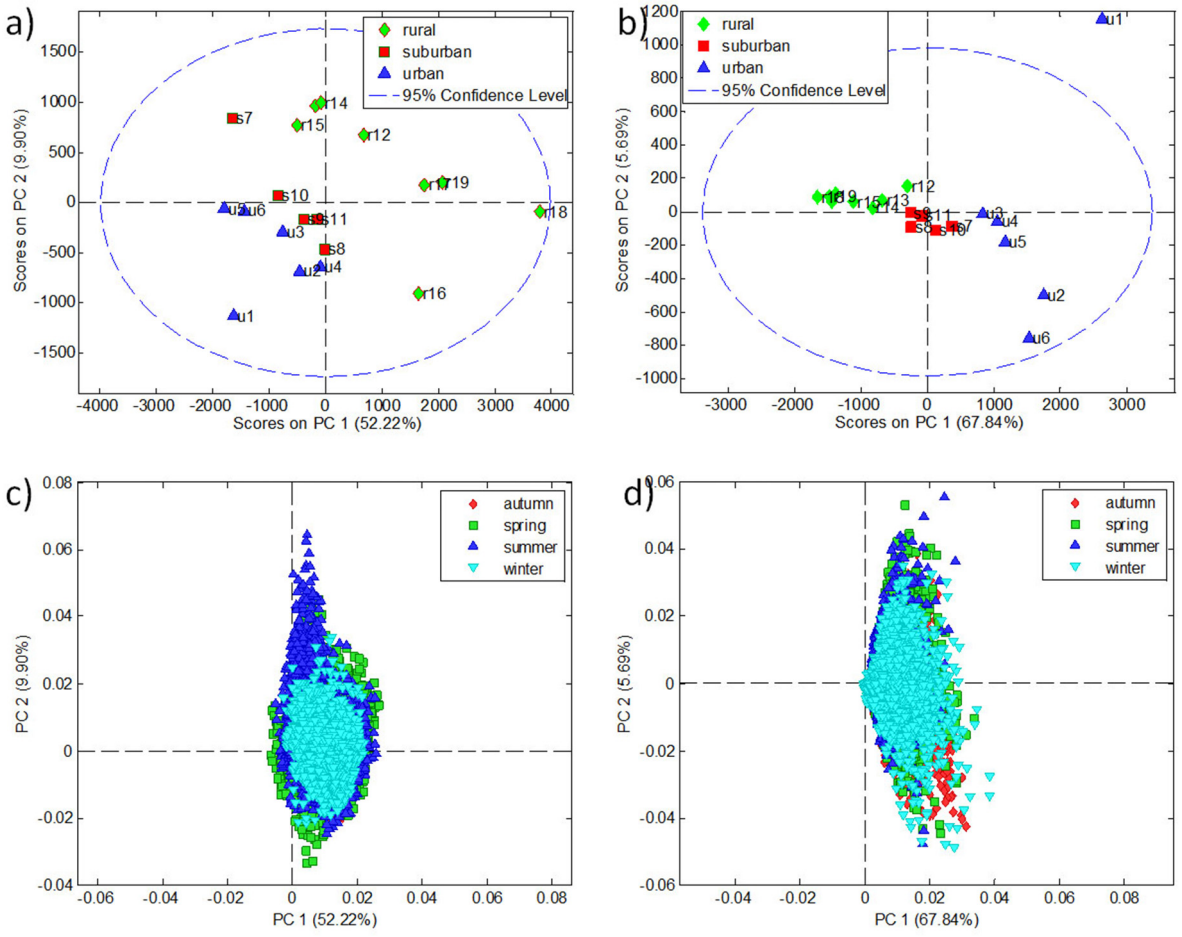


Fig. 3. PCA results of the mean-centered O_3 and NO_2 2015 experimental data matrix $\mathbf{D^*}$. Upper panels show PC1 vs PC2 scores for the analysis of O_3 (a) and NO_2 (b) colored by station type: urban, suburban and rural. Lower panels show PC1 vs PC2 loadings for the analysis of O_3 (c) and NO_2 (d), colored by season: autumn, winter, spring, summer.

Table 1
Summary of the PCA and MCR-ALS (bilinear and trilinear) explained variances (in %) for the different data sets (experimental, EU12, IP4 and CAT1), using three components in all cases.

Data	MCR-ALS Bilinear	MCR-ALS Trilinear
O_3 Experimental	Total[a] 98.4	Total[a] 93.9
	C1 49.6	C1 59.6
	C2 44.8	C2 42.1
	C3 43.3	C3 33.6
	Sum[b] 137.7	Sum[b] 135.3
O_3 EU12 model	Total[a] 99.5	Total[a] 98.5
	C1 61.3	C1 30.2
	C2 48.9	C2 90.7
	C3 36.9	C3 8.7
	Sum[b] 147.1	Sum[b] 129.6
O_3 IP4 model	Total[a] 99.3	Total[a] 97.7
	C1 56.1	C1 63.1
	C2 52.6	C2 56.1
	C3 36.5	C3 26.2
	Sum[b] 145.2	Sum[b] 145.4
O_3 CAT1 model	Total[a] 99.2	Total[a] 97.3
	C1 53.3	C1 68.6
	C2 54.3	C2 47.3
	C3 34.4	C3 14.3
	Sum[b] 142.0	Sum[b] 130.2
NO_2 Experimental	Total[a] 93.6	Total[a] 85.7
	C1 43.8	C1 41.6
	C2 44.9	C2 38.9
	C3 37.1	C3 36.9
	Sum[b] 125.8	Sum[b] 117.4
NO_2 EU12 model	Total[a] 96.8	Total[a] 92.8
	C1 37.3	C1 34.8
	C2 46.3	C2 42.3
	C3 41.9	C3 44.1
	Sum[b] 125.5	Sum[b] 121.2
NO_2 IP4 model	Total[a] 95.4	Total[a] 89.7
	C1 50.3	C1 41.6
	C2 34.0	C2 32.5
	C3 39.4	C3 45.8
	Sum[b] 123.7	Sum[b] 119.9
NO_2 CAT1 model	Total[a] 94.4	Total[a] 86.7
	C1 47.3	C1 40.6
	C2 36.0	C2 34.1
	C3 37.5	C3 37.4
	Sum[b] 120.8	Sum[b] 112.1

[a] Total states for the percentage of explained variance by MCR-ALS bilinear and trilinear models with all components.
[b] Sum states for the sum of the percentages of explained variances by the individual MCR-ALS components C1, C2 and C3. This Sum can be above 100% due to overlapping of information among the different MCR-ALS components.

bilinear and trilinear (i.e. with trilinearity constraint) using three components. In the MCR-ALS results, all three components explained relatively similar amounts of data variance for O_3 and NO_2. With the same number of components (three), the explained variances by the bilinear (98.4%, 93.6%) and trilinear (93.9%, 85.7%) MCR-ALS models for O_3 and NO_2 were rather similar. This suggests the convenience of applying the trilinearity constraint for this type of data, which has also the advantage of providing a more simplified interpretation of the results. MCR-ALS models for all O_3 datasets showed higher amounts of overlapping variance than the models for NO_2 data sets, suggesting a more complex behavior in the case of O_3 concentrations. Overlapping information for MCR-ALS is reduced with the application of the trilinearity constraint.

The results of MCR-ALS analysis of the experimental data and of the data predicted by the EU12, IP4 and CAT1 CALIOPE models are given in Fig. 4 (O_3) and Fig. 5 (NO_2). According to the MCR-ALS first component (C1), O_3 daily profiles (Fig. 4a) have a similar shape in all cases, although for the experimental data analyzed by MCR-ALS (in black in Fig. 4a), it has a higher intensity in the warmer seasons. C1 O_3 hourly profiles (Fig. 4b) are also similar in shape but a time shift of two hours is observed for the experimental data profile (peak maximum around 20:00 h, in the evening) compared with the homologous profiles for the CALIOPE predicted data (peak maximum around 18:00 h). C1 O3 station profiles for IP4 and CAT1 (Fig. 4c) resemble very much the profile obtained in the analysis of the experimental data, showing that this C1 contribution is more important for rural stations. However, the C1 O_3 EU12 profile in Fig. 4c has a rather constant contribution for all the monitoring stations, without differentiating rural from the other types of stations. This indicates that the EU12 data had not enough spatial resolution to capture the local changes among different monitoring stations.

C2 O_3 daily profiles in Fig. 4d have similar shapes for all data sets but with rather different intensities. In this case, the EU12 O_3 daily profile has the highest intensity, while the IP4 and CAT1 profiles have lower intensities than the experimental profile. C2 O_3 EU12 hourly profile (Fig. 4e) is different to the homologous experimental, IP4 and CAT1 profiles. While hourly profiles for IP4 and CAT1 have peak maxima at 02:00 h and 04:00 h respectively, the EU12 hourly profile keeps a flat shape. Again, in this case, a 2 h' time shift is observed for the experimental profile compared to IP4 and CAT1 profiles. C2 O_3 station profiles in Fig. 4f show no significant differences among the four different data sets. In all cases, the C2 O_3 station profile has higher values for the rural stations, with a more pronounced O_3 night peak.

Finally, C3 O_3 daily profile (Fig. 4g) has a more pronounced peak shape in the experimental profile than in the CALIOPE profiles. C3 O_3 hourly profiles in Fig. 4h show a pronounced peak maximum around noon, shifted again by 2 h in the case of the experimental data. This O_3 profile clearly shows the ozone formation due to the increasing solar radiation at noon and afternoon time. C3 O_3 station profiles in Fig. 4i show similar trends for the different monitoring stations, decreasing from urban to rural stations.

Fig. 5 shows the comparison of the daily, hourly and station MCR-ALS profiles obtained in the investigation of NO_2 data sets (experimental, EU12, IP4 and CAT1). C1 NO_2 daily profiles in Fig. 5a show some differences in their shape throughout the year. Whereas in the case of the experimental data (black line), the profile has a pronounced U-shape with their maxima in autumn and winter (increase of NO_2), the C1 NO_2 annual daily profiles for the CALIOPE predicted data are almost flat lines. The C1 hourly (Fig. 5b) and station (Fig. 5c) profiles are very similar in shape and intensity in all cases. In this case, no pronounced time shift was observed and the evening peak maximum of the C1 NO_2 hourly profile coincides around 21:00 h in all cases. The C1 NO_2 station profiles are also coincident, describing similar urban and suburban background NO_2 pollution.

C2 NO_2 daily profiles (Fig. 5d) are also similar for all the investigated data sets. Only the C2 NO_2 daily profile obtained for the experimental data presents slightly more pronounced changes, mainly in spring. Hourly and station C2 NO_2 profiles (Fig. 5e and f) are practically identical in all cases, suggesting a great similarity in the results obtained in the MCR-ALS analysis of the experimental and CALIOPE predicted data.

C3 NO_2 daily profiles in Fig. 5g are also coincident in all cases. As for C1 and C2 NO_2 daily profiles, a similar seasonal trend with a U shape was observed, with their peak minima in the summer and their peak maxima in autumn and winter. A small difference can be observed for the C3 NO_2 hourly CALIOPE profiles (Fig. 5h) compared to the experimental data. A double humped plateau with two peaks and a valley at 14:00 h is more pronounced for the profile obtained from experimental data, and it is not so well defined for the profiles obtained from the three CALIOPE models. In contrast, the latter have a more pronounced second peak maximum at 17:00 h, which is higher than the first peak at 09:00 h.

A more detailed comparison of the hourly profiles for O_3 and NO_2 detected in the MCR-ALS analysis of the experimental data, shows that

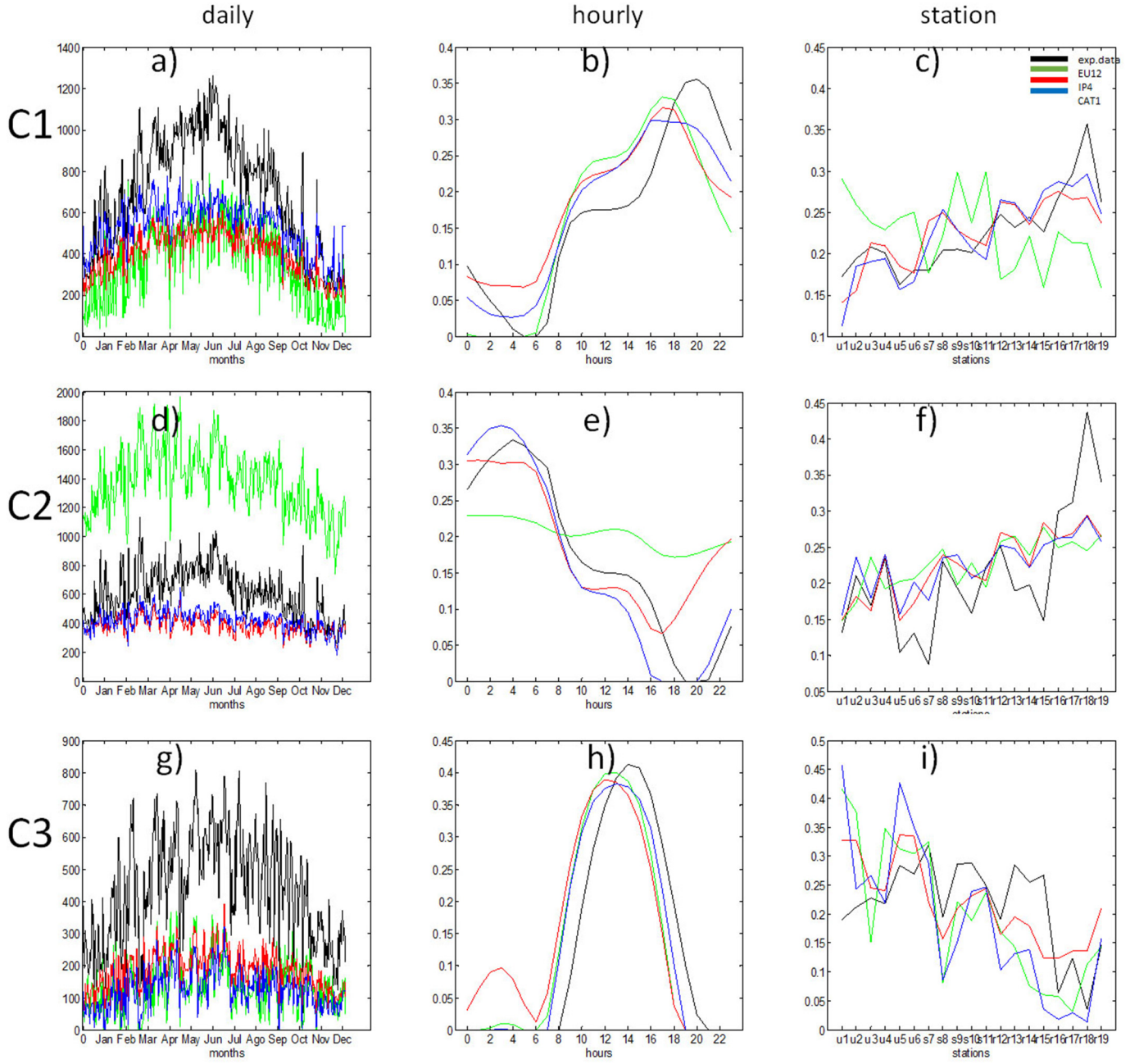


Fig. 4. MCR-ALS results obtained in the analysis of O_3 experimental data (black) and of the three CALIOPE EU12 (green), IP4 (red) and CAT1 (blue) predicted data. C1, C2 and C3 daily (a, d and g), hourly (b, e and h) and station (c, f, i) profiles of the three MCR-ALS components. MCR-ALS explained variances are given in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the first maximum of NO_2 profile at 08:00 h (see C2 Fig. 5e) comes before the C3 O_3 hourly profile starts to increase in the morning (see Fig. 4h). This confirms that the sunlight was triggering the reaction between NO (from morning traffic) and O_3 to form NO_2, thus resulting in O_3 concentration depletion (Alier et al., 2011). After reaching its maximum at the end of the rush traffic hours at 8:00 AM, with the accumulation of solar irradiance hours, the photochemical catalytic reactions turn again in favor of O_3 formation, which reached its daily maximum between 14:00 and 16:00 h. This phenomenon can be observed in detail in Fig. 6a where the maximum of the O_3 hourly profile at 14:00 h coincides with the bottom of the valley in the humpback NO_2 profile. Fig. 6b and c provide additional information about the behavior of the O_3-NO_2 interconnected patterns, e.g. the maximum of O_3 concentrations at 14:00 h occurred for all days, including working days and weekends (Fig. 6b), whereas the NO_2 humpback profile was more pronounced in working days, in agreement with the increased traffic and industrial activities (Fig. 6c). It is worth mentioning here that the simultaneous MCR-ALS analysis of both parameters, O_3 and NO_2, have been also attempted. However, the results did not show any other interaction between them than the one already explained above for the independent analysis of both.

4.2. Comparison of prediction errors using the different CALIOPE model resolutions

Comparison of the experimental data with the predicted data by CALIOPE at different model resolutions is performed using the $RMSE_{daily}$

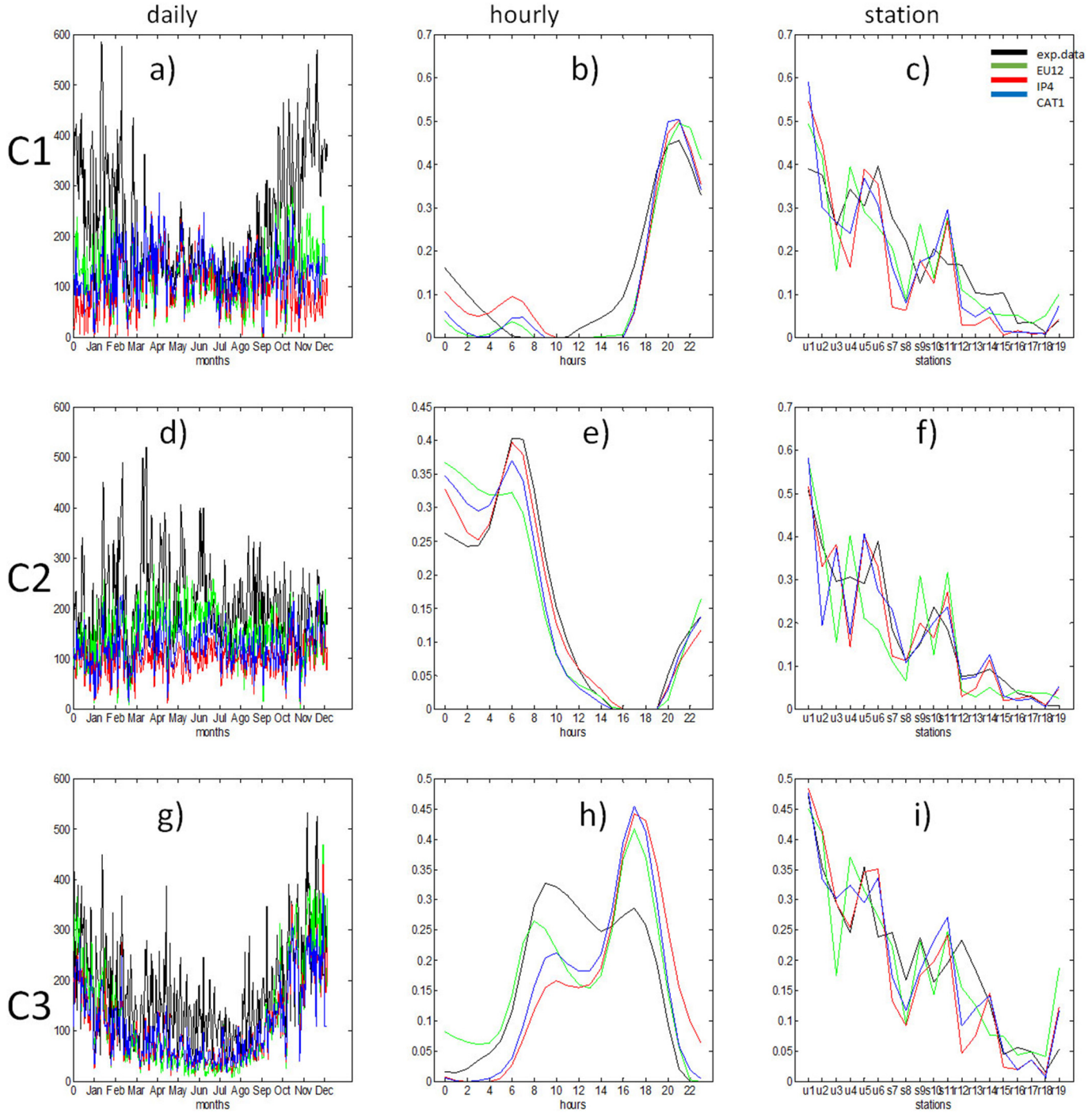


Fig. 5. MCR-ALS results obtained in the analysis of NO_2 experimental data (black) and of the three CALIOPE EU12 (green), IP4 (ref) and CAT1 (blue) predicted data. C1, C2 and C3 daily (a, d and g) hourly (b, e and h) and station (c, f, i) profiles of the three MCR-ALS components. MCR-ALS explained variances are given in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and $RMSE_{hourly}$ profiles calculated by Eqs. (6) and (7). Fig. 7a shows the $RMSE_{daily}$ O_3 profiles obtained for the CALIOPE predicted data. Results present some bias, with curved RMSE profiles in the case of CAT1 and IP4 models. The EU12 predictions gave a $RMSE_{daily}$ profile without a significant seasonal dependence. In summer, when the highest O_3 concentrations are reported, the CAT1 and IP4 prediction errors also increased. In winter, the CAT1 and IP4 models yield lower $RMSE_{daily}$ values than the EU12 model, meaning both models are working better in the lower range of O_3 concentrations. This can be related to the different spatial resolution adjustments used in the CALIOPE system. Fig. 7b shows $RMSE_{daily}$ NO_2 profiles obtained for the CALIOPE predicted data. Slight increases of the $RMSE_{daily}$ rofile can be observed for the winter season, coinciding with the higher NO_2 concentrations reported during this season. In this case, the discrepancies among the CALIOPE predicted data and the experimental data were lower and more similar. Fig. 7c shows the $RMSE_{hourly}$ (Eq. (7)) O_3 profiles obtained for the CALIOPE predicted data. Larger $RMSE_{hourly}$ peak shaped O_3 profiles were obtained for the EU12, IP4 and CAT1 data. The $RMSE_{hourly}$ O_3 profile for the EU12 data had a completely different shape to those obtained for the IP4 and CAT1 data. This suggests

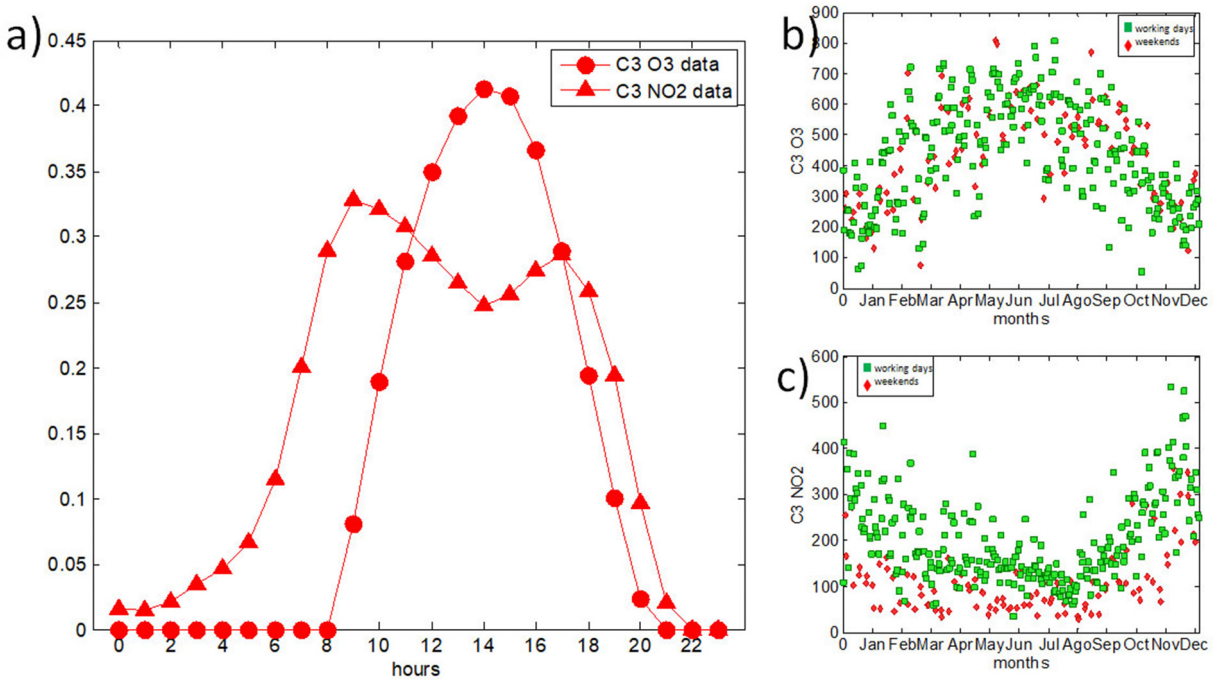


Fig. 6. Comparison of MCR-ALS third component (C3) O_3 and NO_2 hourly and daily profiles. (a) C3 hourly O_3 (red circles) and NO_2 (red triangles) profiles; (b) C3 daily O_3 profiles for weekends (red diamonds) and working days (green squares) and(c) C3 daily NO_2 profiles for weekends (red diamonds) and working days (green squares). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

different modelling abilities regarding the photochemical processes that occur at noon and in the afternoon, when the solar radiation tends to be higher.

Table 2 summarizes the prediction accuracies of the CALIOPE system at different model resolutions The $RMSE_{avg}$ value was calculated (Eq. (8)) between the modeled and experimental data, averaging RMSE values per hour, day and station for O_3 and NO_2. In average, for the three resolutions, CALIOPE predicted NO_2 concentrations better ($RMSE_{avg}$ values between 16 and 17) than O_3 concentrations ($RMSE_{avg}$ values between 33 and 34). EU12 slightly over performed the IP4 and CAT1 predictions for O_3 and NO_2, but only by just 1-1.5 units. $RMSE_{avg}$ values for CALIOPE predictions were also calculated for the different types of stations separately, as shown in Table 2. For O_3 in the urban areas, $RMSE_{avg}$ values were similar in the range 29-33. IP4 and CAT1 predicted slightly better than EU12 with 2-3 RMSE units lower. In the case of NO_2 in urban stations, $RMSE_{avg}$ values were higher than those calculated for all types of stations together. For the suburban areas, $RMSE_{avg}$ values were similar to those obtained for all stations, both for O_3 and NO_2. However, NO_2 suburban $RMSE_{avg}$ values were lower than those in urban areas (urban $RMSE_{avg}$) . For rural areas, O_3 $RMSE_{avg}$ EU12 values were lower than those of IP4 and CAT1, whereas NO2 rural $RMSE_{avg}$ for the three CALIOPE predicted data were similar.

5. Conclusions

The application of MCR-ALS to the analysis of hourly measured O_3 and NO_2 concentrations at nineteen air quality monitoring stations across Catalonia, Spain, during 2015 allowed the resolution of three major variability contributions of these two pollutants, described by their hourly, daily, and spatial profiles, which can be correlated with the major physicochemical and pollution patterns acting over the investigated region.

The results obtained by MCR-ALS analysis of the measured experimental data were compared with those obtained by the MCR-ALS analysis of the CALIOPE predicted data. In particular, a time lag of 2 h in the O_3 hourly resolved profiles was observed when compared to the analysis of experimental data recorded in UTC time. This delay could be attributed to a combination of uncertainties in the modelling of the planetary boundary layer in WRF and to the temporal emission profiles for medium-cities, which were adjusted by population density. In general, MCR-ALS hourly, daily and station profiles obtained in the analysis of concentrations predicted by CALIOPE IP4 and CAT1 models with higher spatial resolution were more similar to those obtained in the MCR-ALS analysis of experimental concentrations, than those obtained in the analysis of EU12 data. These similarities between IP4 and CAT1 predictions can be attributed to the fact that they are both based on the same emission estimation approach (mainly bottom-up), meanwhile EU12 predictions are based on spatial and temporal disaggregation of the EMEP gridded emissions. However, both datasets, IP4 and CAT1, still appeared to underestimate the O_3 daily profiles, compared to the ones obtained from the MCR-ALS analysis experimental data. The EU12 model predictions differed in the stations O_3 profile of the first component and in the hour profile of the second component, with the midnight maximum around 02:00 h not well described. Prediction accuracy of the three CALIOPE models was affected by the daytime, the season and the pollutant. Globally, the accuracy of NO_2 predictions was better than the accuracy of O_3 predictions in the three model resolutions, with low differences in their hourly and seasonal predictions. However, the station local environment influenced more the accuracy of NO_2 CALIOPE predictions than the O_3 CALIOPE ones, with larger errors in urban and suburban areas than in rural areas.

Future work could envisage the possibility of extending the same type of analysis to larger geographical regions in Spain and Europe and for longer time periods of several years. Moreover, it will be also interesting to compare the results of this study with those obtained using other modelling approaches, also including the chemical transportation and transformation of the investigated pollutants. Finally, another possible application of the proposed combined MCR-ALS analysis of experimental and CALIOPE predicted data could be its use as a real time monitoring warning tool of pollution episodes, when CALIOPE predicted

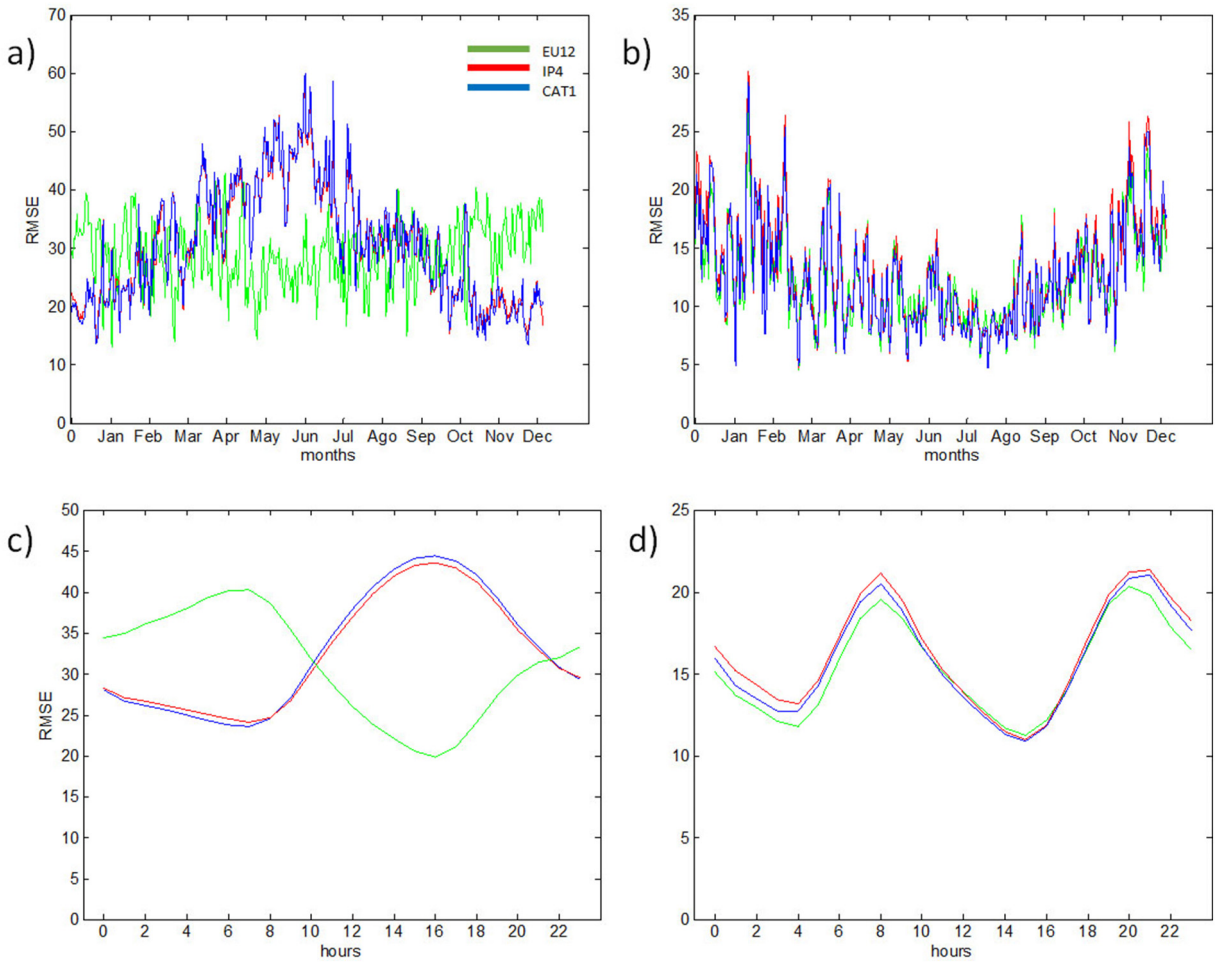


Fig. 7. O_3 and NO_2 $RMSE_{daily}$ (a and b) and (c and d) $RMSE_{hourly}$ profiles from EU12 (green), IP4 (red) and CAT1 (blue) CALIOPE predicted data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

concentrations differ significantly from experimental ones and urgent actions are needed to limit emission sources and to avoid critical episodes.

Table 2
O_3 and NO_2 $RMSE_{avg}$ values as $\mu g/m^3$ for MCR-ALS and CALIOPE predicted data for all stations and separately for urban suburban and rural stations (see experimental Section 2.1, Fig. 1 and Eq. (8)).

	O_3	NO_2
$RMSE_{avg}$		
EU12	31.8	15.6
IP4	33.2	16.6
CAT1	33.5	16.1
Urban $RMSE_{avg}$		
EU12	32.5	22.5
IP4	29.1	23.5
CAT1	29.9	23.6
Suburban $RMSE_{avg}$		
EU12	33.5	14.9
IP4	32.0	15.9
CAT1	32.4	14.5
Rural $RMSE_{avg}$		
EU12	30.0	8.0
IP4	36.6	8.9
CAT1	36.6	8.2

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper with title, Understanding temporal and spatial changes of O_3 and NO_2 concentrations combining multivariate data analysis methods and air quality transport models, coauthored by Stefan Platikanov, Marta Terrado, María Teresa Pay, Albert Soret and Romà Tauler.

Acknowledgements

SP and RT would like to acknowledge the Spanish Ministerio de Ciencia e Innovación for the project PID2019-105732GB-C21. SP and RT work at IDAEA-CSIC which is a Center of Excellence Severo Ochoa (Spanish Ministry of Science and Innovation, Project CEX2018-000794-S). MTP would like to acknowledge the Spanish Ministry of Economy and Competitiveness and FEDER funds under the PAISA (CGL2016-75725-R) project.

CRediT authorship contribution statement

Stefan Platikanov: preparation, data curation, writing - original draft, software, visualization. Marta Terrado: preparation, data curation, validation, reviewing and editing. María Teresa Pay: validation, reviewing and editing. Albert Soret: supervision, conceptualization, reviewing.

Romà Tauler: supervision, conceptualization, methodology, writing - reviewing and editing.

Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.150923.

References

Aguilera, I., Basagaña, X., Pay, M.T., Agis, D., Bouso, L., Foraster, M., Rivera, M., Baldasano, J.M., Künzli, N., 2013. Evaluation of the CALIOPE air quality forecasting system for epidemiological research: the example of NO2 in the province of Girona (Spain). Atmos. Environ. 72, 134–141.

Alier, M., Tauler, R., 2013. Multivariate curve resolution of incomplete data multisets. Chemom. Intell. Lab. Syst. 127, 17–28.

Alier, M., Felipe-Sotelo, M., Hernàndez, I., Tauler, R., 2009. Variation patterns of nitric oxide in Catalonia during the period from 2001 to 2006 using multivariate data analysis methods. Anal. Chim. Acta 642, 77–88.

Alier, M., Felipe-Sotelo, M., Hernàndez, I., Tauler, R., 2011. Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O3 pollution in Barcelona. Anal. Bioanal. Chem. 399, 2015–2029.

Baldasano, J.M., Jiménez-Guerrero, P., Jorba, O., Pérez, C., López, E., Güereca, P., Martín, F., Vivanco, M.G., Palomino, I., Querol, X., Pandolfi, M., Sanz, M.J., Diéguez, J.J., 2008a. CALIOPE: an operational air quality forecasting system for the Iberian Peninsula, Balearic Islands and Canary Islands – first annual evaluation and ongoing developments. Adv. Sci. Res. 2, 89–98.

Baldasano, J.M., Güereca, L.P., López, E., Gassó, S., Jimenez-Guerrero, P., 2008b. Development of a high-resolution (1 km x 1 km, 1 h) emission model for Spain: the high-elective resolution modelling emission system (HERMES). Atmos. Environ. 42, 7215–7233.

Baldasano, J.M., Pay, M.T., Jorba, O., Gassó, S., Jiménez-Guerrero, P., 2011. An annual assessment of air quality with the CALIOPE modeling system over Spain. Sci. Total Environ. 409, 2163–2178.

Basart, S., Pérez, C., Nickovic, S., Cuevas, E., Baldasano, J.M., 2012. Development and evaluation of the BSC-DREAM8b dust regional model over Northern Africa, the Mediterranean and the Middle East. Tellus Ser. B 64, 1–12.

Belis, C.A., Favez, O., Mircea, M., Diapouli, E., Manousakas, M.-I., Vratolis, S., Gilardoni, S., Paglione, M., Decesari, S., Mocnik, G., Mooibroek, D., Salvador, P., Takahama, S., Vecchi, R., Paatero, P., 2019. European Guide on Air Pollution Source Apportionment With Receptor Models-Revised Version 2019, EUR 29816EN. Publications Office of the European Union, Luxembourg 978-92-76-09001-4 https://doi.org/10.2760/439106 JRC117306.

Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M., Dominici, F., 2004. Ozone and short-term mortality in 95 US urban communities, 1987 -2000. J. Am. Med. Assoc. 292 (19), 2372–2378.

Bro, R., de Jong, S., 1997. A fast non-negativity-constrained least squares algorithm. J. Chemom. 11, 393–401.

Byun, D., Schere, K., 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. Appl. Mech. Rev. 59, 51–77.

Chan, K., Chan, K., 2017. Aerosol optical depths and their contribution sources in Taiwan. Atmos. Environ. 148, 364–375.

Chan, K., Hartl, A., Lam, Y., Xie, P., Liu, W., Cheung, H., Lampel, J., Pohler, D., Li, A., Xu, J., Zhou, H., Ning, Z., Wenig, M., 2015. Observations of tropospheric NO2 using ground based MAX-DOAS and OMI measurements during the Shanghai World Expo 2010. Atmos. Environ. 119, 45–58.

Crutzen, P.J., 1974. Photochemical reactions initiated by and influencing ozone in the troposphere. Tellua 26, 47–67.

Dadashi, M., Pages, D., Hernandez, I., Tauler, R., 2020. Chemometrics modelling of temporal changes of ozone half hourly concentrations in different monitoring station. Chemom. Intell. Lab. Syst. 201, 104015. https://doi.org/10.1016/j.chemolab.2020.104015.

De Juan, A., Tauler, R., 2001. 2001. Comparison of three-way resolution methods for non-trilinear chemical data sets. J. Chemometr. 15 (10), 749–771.

EEA, 2011. The Application of Models Under the European Union's Air Quality Directive: A Technical Reference Guide, EEA Technical Report 10/2011. ISSN Technical report series 1725-2237Publication Office of the European Union, Luxembourg 978-92-9213-223-1 https://doi.org/10.2800/80600 76 pp.

EEA, 2019. Air Quality in Europe – 2019 Report. European Environment Agency, Publications Office of the European Union, EEA Report No 10/2019, Luxembourg.

Eilers, P., 2003. A perfect smoother. Anal. Chem. 75, 3631–3636.

EU, 2011. Commission Implementing Decision No 2011/850/EU of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality (OJ L 335, 17.12.2011, p. 86-106). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32011D0850. (Accessed 19 March 2020).

European Commission, 2008. Directive 2008/50/EC of the European Parliament andof the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe. Technical Report 2008/50/EC, L152. Off. J. Eur. Comm. 51, 1–44. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32008L0050. (Accessed 15 October 2021).

European Environmental Agency (EEA), 2011. Air Quality in Europe -2011 Report. EEA, Copenhagen.

Golub, G., van Loan, C., 1989. Matrix Computations. Second ed. John Hopkins University Press, Baltimore, USA.

Gratsea, M., Vrekoussis, M., Richter, A., Wittrock, F., Schönhardt, A., Burrows, J., Kazadzis, S., Mihalopoulos, N., Gerasopoulos, E., 2016. Slant column MAX-DOAS measurements of nitrogen dioxide, fromaldehyde, glyoxal and oxygen dimer in the urban environment of Athens. Atmos. Environ. 135, 118–131.

Guevara, M., Martínez, F., Arévalo, G., Gassó, S., Baldasano, J.M., 2013. An improved system for modelling Spanish emissions: HERMESv2.0. Atmos. Environ. 81, 209–221.

Holloway, T., Fiore, A., Galanter Hastings, M., 2003. Intercontinental transport of air pollution: will emerging science lead to a new hemispheric treaty? Environ. Sci. Technol. 37, 4535–4542.

Hopke, P.K., 2008. The use of source apportionment for air quality management and health assessments. J. Toxic. Environ. Health A 71, 555–563.

Jaumot, J., de Juan, A., Tauler, R., 2015. G.U.I. MCR-ALS 2.0: new features and applications. Chemom. Intell. Lab. Syst. 140, 1–12.

Jolliffe, I.T., 2002. Principal Component Analysis. 2nd ed. Springer Verlag, Berlin, Germany.

de Juan, A., Tauler, R., 2003. Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution. Anal. Chim. Acta 500, 195–210.

Khan, J., Ketzel, M., Kakosimos, K., Sorensen, M., Jensen, S., 2018. Road traffic air and noise pollution exposure assessment – a review of tolos and techniques. Sci. Total Environ. 634, 661–676.

Malik, R., Tauler, R., 2013. Extension and application of multivariate curve resolution-alternating least squares to four-way quadrilinear data-obtained in the investigation of pollution patterns on Yamuna River, India—a case study. Anal. Chim. Acta 794, 20–28.

Marín-García, M., Tauler, R., 2020. Chemometrics characterization of the Llobregat river dissolved organic matter. Chemom. Intell. Lab. Syst. 201, 104018.

Massagué, J., Carnerero, C., Escudero, M., Baldasano, J.M., Alastuey, A., Querol, X., 2019. 2005–2017 ozone trends and potential benefits of local measures as deduced from air quality measurements in the north of the Barcelona metropolitan area. Atmos. Chem. Phys. 19 (11), 7445–7465.

Monks, P., Archibald, A., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K., Mills, G., et al., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. 15(15), 8889–8973.

Nieuwenhuijsen, M.J., 2018. Influence of urban and transport planning and the city environment on cardiovascular disease. Nat. Rev. Cardiol. 15, 432–438. https://doi.org/10.1038/s41569-018-0003-2.

Norris, G., Duvall, R., Brown, S., Bai, S., 2014. EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. Prepared for the U.S. Environmental Protection Agency Office of Research and Development, Washington, DC, USA.

Otero, N., Sillmann, J., Schnell, J.L., Rust, H., Butler, T., 2016. Synoptic and meteorological drivers of extreme ozoneconcentrations over Europe. Environ. Res. Lett. 11, 024005.

Pay, M.T., Piot, M., Jorba, O., Basart, S., Gassó, S., Jiménez-Guerrero, P., Gonçalves, M., Dabdub, D., Baldasano, J.M., 2010. A full year evaluation of the CALIOPE-EU air quality system in Europe for 2004: a model study. Atmos. Environ. 44, 3322–3342.

Pay, M.T., Jiménez-Guerrero, P., Jorba, O., Basart, S., Pandolfi, M., Querol, X., Baldasano, J.M., 2012. Spatio-temporal variability of levels and speciation of particulate matter across Spain in the CALIOPE modeling system. Atmos. Environ. 46, 376–396. https://doi.org/10.1016/j.atmosenv.2011.09.049.

Pay, M.T., Martínez, F., Guevara, M., Baldasano, J., 2014. Air quality forecast on a kilometer-scale grid over complex Spanish terrains. Geosci. Model Dev. 7, 1979–1999.

Pay, M.T., Gangoiti, G., Guevara, M., Napelenok, S., Querol, X., Jorba, O., Pérez García-Pando, C., 2019. Ozone source apportionment during peak summer events over southwestern Europe. Atmos. Chem. Phys. 19, 5467–5494.

Querol, X., Gangoiti, G., Mantilla, E., Alastuey, A., Minguillón, M.C., Amato, F., Reche, C., Viana, M., Moreno, T., Karanasiou, A., Rivas, I., Pérez, N., Ripoll, A., Brines, M., Ealo, M., Pandolfi, M., Lee, H.-K., Eun, H.-R., Park, Y.-H., Escudero, M., Beddows, D., Harrison, R.M., Bertrand, A., Marchand, N., Lyasota, A., Codina, B., Olid, M., Udina, M., Jiménez-Esteve, B., Soler, M.R., Alonso, L., Millán, M., Ahn, K.-H., 2017. Phenomenology of high-ozone episodes in NE Spain. Atmos. Chem. Phys. 17, 2817–2838.

Schaap, M., et al., 2015. Performance of european chemistry transport models as function of horizontal resolution. Atmos. Environ. 112, 90–105.

Skamarock, W., Klemp, J., 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. J. Comput. Phys. 227, 3465–3485.

Smilde, A.K., Bro, R., Geladi, P., 2004. Multi-way Analysis With Applications in the Chemical Sciences. J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ.

Stanimirova, I., 2013. Practical approaches to principal component analysis for simultaneously dealing with missing and censored elements in chemical data. Anal. Chim. Acta 796, 27–37.

Tack, F., Hendrick, F., Goutail, F., Fayt, C., Merlaud, A., Pinardi, G., Hermans, C., Pommereau, J., Roozendael, M., 2015. Tropospheric nitrogen dioxide column retrieval from ground-based zenith-sky DOAS observation. Atmos. Meas. Tech. 8, 2417–2435.

Tauler, R., 1995. Multivariate curve resolution applied to second order data. Chemom. Intell. Lab. Syst. 30, 133–146.

Tauler, R., Marqués, I., Casassas, E., 1998. Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid–base titration of salicylic acid at three excitation wavelengths. J. Chemom. 12, 55–75.

Vicedo-Cabrera, A.M., Sera, F., Liu, C., Armstrong, B., Milojevic, A., Guo, Y., Gasparrini, A., 2020. Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries. BMJ 368.

WHO, 2013. Review of evidence on health aspects of air pollution – REVIHAAP project: technical report. 302 pp., available at:WHO Regional Office for Europe, Copenhagen. http://www.euro.who.int/__data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1.

WHO, 2016. Ambient air pollution: a global assessment of exposure and burden of disease. available atWHO Library Cataloging-in-Publication Data, World Health Organization. https://www.who.int/phe/publications/air-pollution-global-assessment/en/.

Windig, W., Guilment, J., 1991. Interactive self-modeling mixture analysis. Anal. Chem. 63, 1425–1432.

Yarwood, G., Roa, S., Yocke, M., Whitten, G., 2005. Updates to the Carbon Bond Chemical Mechanism: CB05, Final Report to the USEPA, RT-0400675.