https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Long-term evaluation of surface air pollution in CAMSRA and MERRA-2 global reanalyses over Europe (2003-2020)

Aleks Lacima[1], Hervé Petetin[1], Albert Soret[1], Dene Bowdalo[1], Oriol Jorba[1], Zhaoyue Chen[2,3], Raúl Méndez Turrubiates[2], Hicham Achebak[2], Joan Ballester[2], and Carlos Pérez García-Pando[1,4]

[1]Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
[2]ISGlobal, Barcelona, Spain
[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[4]ICREA, Catalan Institution for Research and Advanced Studies, Barcelona, Spain

Correspondence: Aleksander Lacima (aleksander.lacima@bsc.es)

Abstract. Over the last century, our societies have experienced a sharp increase in urban population and fossil-fueled transportation, turning air pollution into one of the most critical issues of our time. It is therefore fundamental to accurately characterize the spatiotemporal variability of surface air pollution, in order to understand its effects upon human health and the environment, knowledge that can then be used to design effective pollution reduction policies. Global atmospheric composition reanalyses offer great capabilities towards this characterization through assimilation of satellite measurements. However, they do not integrate surface measurements and thus remain affected by significant biases at ground-level. In this study, we thoroughly evaluate two global atmospheric composition reanalyses, CAMSRA and MERRA-2, between 2003 and 2020, against independent surface measurements of O_3, NO_2, CO, SO_2, PM_{10} and $PM_{2.5}$ over the European continent. Overall, both reanalyses present significant and persistent biases for almost all examined pollutants. CAMSRA clearly outperforms MERRA-2 in capturing the spatiotemporal variability of O_3, CO, PM_{10} and $PM_{2.5}$ surface concentrations. Despite its higher spatial resolution and focus on aerosol representation, MERRA-2 only performs better than CAMSRA for SO_2. CAMSRA also outperforms MERRA-2 in capturing the annual trends found in all pollutants. Both reanalyses show a better performance in summer (JJA), in terms of biases and errors, than in winter (DJF), when pollutant concentrations peak, with the exception of O_3. Higher correlations are not necessarily found in JJA, particularly for reactive gases, which show greater correlation values in autumn (SON) and winter. Compared to MERRA-2, CAMSRA assimilates a wider range of satellite products which, while enhancing the performance of the reanalysis in the troposphere (as shown by other studies), has a limited impact on the surface. The biases found in both reanalyses are likely explained by a combination of factors, including errors in emission inventories and/or sinks, a lack of surface data assimilation and their relatively coarse resolution. Our results highlight the current limitations of reanalyses to represent surface pollution, which limits their applicability for health and environmental impact studies. When applied to reanalysis data, bias-correction methodologies based on surface observations should help constraining the spatiotemporal variability of surface pollution and its associated impacts.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

1 Introduction

In the last two decades, reanalyses have become a very powerful tool in modern Earth sciences as they combine both model- and observation-based information to provide physically consistent data of land, ocean and atmospheric variables with continuous spatial and temporal coverage. In the field of atmospheric composition (hereafter, AC), different reanalysis products are available at global scale, including the Copernicus Atmospheric Monitoring Service reanalysis (CAMSRA; Inness et al. (2019)), produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), and the Modern-Era Retrospective Analysis for Research and Applications v2 (MERRA-2; Gelaro et al. (2017), Randles et al. (2017), Buchard et al. (2017a)), produced by National Aeronautics and Space Administration (NASA)'s Global Modelling and Assimilation Office (GMAO). Both products assimilate a variety of space-based remote sensing observations (mostly total and tropospheric columns) obtained from a growing fleet of satellites measuring reactive gases such as ozone (O_3), nitrogen dioxide (NO_2) or carbon monoxide (CO), as well as aerosol optical depth (AOD). Such an extensive data assimilation of satellite observations is crucial for reducing the biases related to erroneous emission forcings and/or overly coarse representations of the physical and chemical processes that occur in the atmosphere. Data assimilation helps to better constrain the spatiotemporal variability and long-term trends of the most important chemical compounds, providing a physically consistent view of the Earth's atmospheric composition.

Considering the strong interest of atmospheric composition reanalyses for a variety of applications (e.g. climatological studies, initial and/or boundary conditions for regional-scale modeling systems, air pollution impact assessment, health studies), it is crucial to characterize the strengths and limitations of these global products, in particular at the surface, as no in situ chemical observations are assimilated. The most recent studies evaluating the CAMSRA and/or MERRA-2 reanalysis at ground-level are indicated in Table 1, highlighting the limited effort that has been made so far to evaluate and inter-compare these reanalysis products against in situ surface measurements.

The main findings of this more recent literature are briefly outlined here. Ryu and Min (2021) found significant and persistent biases in all the pollutants examined over South Korea, with CAMSRA outperforming MERRA-2 in all cases except for SO_2. By performing a global evaluation, Wagner et al. (2021) showed that CAMSRA provides an overall accurate representation of reactive gases over time, and highlighted the key role played by satellite data assimilation and emissions in improving AC reanalysis products. Both these two previous studies analyze a wide range of aerosols and reactive gases and cover the most extensive period possible at the time, 2003-2018, which is limited by the start of CAMSRA in 2003. Ma et al. (2021) found persistent biases in PM_{10} concentration over mainland China in MERRA-2 for the periods 2011-2013 and 2016-2017, with better performance during summer. Their results also found that adding nitrate compounds to MERRA-2 significantly improved the reanalysis performance. Navinya et al. (2020) found a systematic underestimation of $PM_{2.5}$ concentration in MERRA-2 over India for the period 2015-2018.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Table 1. Review of recent studies evaluating the CAMSRA and/or MERRA-2 reanalysis at the surface using in situ observations.

Author	Region	Period	Reanalysis	Pollutants
Ryu and Min (2021)	South Korea	2003-2018	CAMSRA; MERRA-2; TCR-2	CO, NO_2, SO_2, O_3, PM_{10}
Wagner et al. (2021)	Global	2003-2018	CAMSRA	NO_2, O_3, CO, HCHO
Ma et al. (2021)	China	2011-2013; 2016-2017	MERRA-2	PM_{10}
Navinya et al. (2020)	India	2015-2018	MERRA-2	$PM_{2.5}$
Provençal et al. (2017a)	Europe	2003-2014	MERRA-1	$PM_{2.5}$, PM_{10}
Provençal et al. (2017b)	Israel; Taiwan	2002-2015	MERRA-1	$PM_{2.5}$
Buchard et al. (2016)	USA	2003-2012	MERRA-1	$PM_{2.5}$

Our study evaluates CAMSRA and MERRA-2 against independent surface in situ measurements over the period 2003-2020, focusing on the European continent, a region still poorly covered by past evaluation studies (Table 1). It considers all major pollutants with recognized harmful effects on human health and sufficient observational data available at the surface, namely O_3, NO_2, CO, SO_2, PM_{10} and $PM_{2.5}$. The motivation behind this study arose in the context of the ERC project EARLY-ADAPT (https://early-adapt.eu/), in the frame of which a pioneer health dataset is currently being collected over Europe to investigate the time-varying health effects of climate and air pollution, and thus shed light into the early adaptation response to climate change in the field of human health. This impact will be quantified by fitting epidemiological models on historical local health, climate and air pollution data, which thus requires a long-term (multi-decadal) air quality database of the most harmful pollutants, at daily-scale and over the entire European domain. Despite their relatively coarse spatial resolution, which is the counterpart to a sufficiently long-term coverage, global-scale atmospheric composition reanalyses provide highly valuable information, though remain subject to biases and errors both in terms of spatial, seasonal and intra-annual variability, but also regarding long-term trends. It is worth mentioning here that the CAMS regional reanalysis (Marécal et al. (2015)), focused on Europe, assimilates surface in situ observations and provides air pollution fields at a finer spatial resolution than CAMSRA, but only over a limited period of time (2014-2018), for which reason we focus here on the global reanalysis.

In Sect. 2, we introduce the data (Subsect. 2.1) and provide details on the different methods employed for their analysis (Subsect. 2.2). Results are presented and discussed in Sect. 3, and summarized in Sect. 4.

2 Data & Methodology

In this section we briefly describe our observational and reanalysis datasets, while providing details on the different statistical methods employed for their analysis. Throughout this work, square brackets, [], are used to indicate concentration of a chemical compound (e.g. $[O_3]$ = O_3 concentration) measured in *parts per billion* (ppbv) for reactive gases and in $\mu g\,m^{-3}$ for aerosols.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions
Open Access
EGU

2.1 Data

Our model data comes from two global atmospheric composition reanalyses, CAMSRA and MERRA-2, whose main characteristics are summarized in Table 2. The reanalyses are evaluated against surface in situ measurements obtained from two European Environment Agency (EEA) databases, AIRBASE, for the period 2003-2012, and AQ_eReporting, for the period 2012-2020.

2.1.1 CAMSRA

Produced by ECMWF, the CAMS global atmospheric composition reanalysis consists of 3-dimensional time-consistent AC fields that include chemical species, aerosols and greenhouse gases (GHGs), and currently covers a temporal period extending from 2003 to mid-2021. The reanalysis starts in 2003, when space-based observational measurements, retrieved from a myriad of instruments on-board Envisat, Terra, Aura, MetOp and POES satellites, became available. The latest CAMSRA version was produced in cycle 42R1 of ECMWF's Integrated Forecast System (IFS) using 4DVar data assimilation of satellite measurements, including O_3, NO_2, CO and AOD. This IFS cycle includes the modified Carbon Bond 2005 Chemical Mechanism (CB05), which serves as the tropospheric chemistry scheme of the reanalysis (Flemming et al., 2015). Anthropogenic emissions come from the MACCity inventory data (Granier et al., 2011) for the period 2003-2010, and from 2010 onwards they are derived according to the 8.5 representative concentration pathway (RCP). Biomass burning emissions are obtained from the Global Fire Assimilation System (GFAS) v1.2 (Kaiser et al., 2012), whereas monthly mean biogenic VOC emissions are computed with the Model of Emissions of Gases and Aerosols from Nature (MEGAN) using MERRA-2 reanalysed meteorology (Sindelarova et al., 2014). Meteorological observations and fields are taken from ERA5 (Hersbach et al., 2020).

CAMSRA has a horizontal resolution of approximately 80 km (similar to a regular 0.75º x 0.75º latitude/longitude grid), with data being available both as spectral coefficients (T255) or on a reduced Gaussian grid (N128). Its vertical resolution consists of 60 hybrid sigma/pressure model levels, with the top level located at 0.1 hPa. CAMSRA products are available at a temporal resolution of 3 h, including 3-hourly analysis fields and 3-hourly forecast fields. The bias present in the different AC datasets employed to build CAMSRA is corrected through a variational bias correction scheme (Dee and Uppala, 2008). For a more thorough and detailed description of CAMSRA we direct the reader to Inness et al. (2019) and Wagner et al. (2021).

In CAMSRA, both PM_{10} and $PM_{2.5}$ are directly available and do not require to be reconstructed from its separate chemical compounds, which include black carbon (BC), organic carbon (OC), organic matter (OM), sulphates (SO_4), sea salt and dust. Notably, aerosol nitrates are, at this time, not included in the reanalysis, which can lead to significant underestimations in regions where nitrates represent an important part of total aerosol concentration. Secondary organic aerosols (SOA) of anthropogenic origin are parametrized according to Spracklen et al. (2011).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions
Open Access
EGU

Table 2. Summary of reanalysis products.

Reanalysis	CAMSRA	MERRA-2
Available pollutants	O_3, NO_2, CO, SO_2, PM_{10}, $PM_{2.5}$	O_3, CO, SO_2, PM_{10}, $PM_{2.5}$
Coverage period	2003–present	1980–present
Spatial resolution	~80 km (roughly 0.75º x 0.75º)	0.5° x 0.625°
Assimilation system	IFS Cycle 42r1 4D-Var	3D-Var Gridpoint Statistical Interpolation (GSI)
Meteorology	ERA5 (Hersbach et al., 2020)	GEOS-5 (Rienecker et al. (2008), Molod et al. (2012))
Chemistry	IFS(CB05) (Flemming et al., 2015)	GOCART (Chin et al. (2002), Colarco et al. (2010))
Anthropogenic emissions	MACCity (Granier et al., 2011)	AeroCom Phase II (HCA0 v1; Diehl et al. (2012)), EDGARv4.2 (https://edgar.jrc.ec.europa.eu/)
Biomass burning emissions	GFASv1.2 (Kaiser et al., 2012)	RETROv2 (Duncan et al., 2003), GFEDv3.1 (Randerson et al., 2006), QFED 2.4-r6 (Darmenov and da Silva, 2013)
Biogenic emissions	MEGAN (Sindelarova et al., 2014)	NVOC (Guenther et al., 1995)
Volcanic emissions	—	AeroCom Phase II (HCA0 v2; Diehl et al. (2012))
Assimilated O_3 products	SCIAMACHY, MIPAS, MLS OMI, GOME-2, SBUV/2	MLS, OMI, SBUV, SBUV/2
Assimilated NO_2 products	SCIAMACHY, OMI, GOME-2	—
Assimilated CO products	MOPITT	—
Assimilated SO_2 products	—	—
Assimilated aerosol products	AATSR, MODIS	AVHRR, AERONET, MISR, MODIS

2.1.2 MERRA-2

Developed by NASA's GMAO, the MERRA-2 atmospheric composition reanalysis is based on the Goddard Earth Observing System v5 (GEOS-5) atmospheric model. It is important to note at this stage that, in contrast with CAMSRA, which aims to simulate all major chemical compounds present in the atmosphere, the MERRA-2 reanalysis, despite being the first AC reanalysis that couples chemistry to global atmospheric circulation, focuses mainly on aerosols. Therefore, aside from meteorological data, only AOD observations and O_3 columns are assimilated in MERRA-2, based on both measurements from Terra, Aura, MetOp and POES satellites, and - unlike in CAMSRA - surface-based observations from the Aerosol Robotic Network (AERONET). Anthropogenic sulfate, black carbon (BC) and primary organic matter (POM) emissions are obtained from AEROsol COMparisons between Observations and Models (AeroCom) Phase II (HCA0 v1; Diehl et al. (2012)). Anthropogenic SO_2 emissions are taken from the Emissions Database for Global Atmospheric Research (EDGAR) v4.2, developed by the European Commission (https://edgar.jrc.ec.europa.eu/), whereas volcanic SO_2 is retrieved from AeroCom Phase II (HCA0 v2; Diehl et al. (2012)). CO is simulated by the GEOS-5 modeling system. Sea salt and dust emissions, both composed of five non-interacting size bins, are wind-driven. Aerosol chemistry is reproduced with a version of the Goddard Chemistry Aerosol

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

Radiation and Transport (GOCART; Chin et al. (2002), Colarco et al. (2010)) model, which simulates the processes, interactions, sources and sinks of the different chemical compounds included in MERRA-2, with the exception of O_3 and CO.

MERRA-2 currently covers a temporal period extending from 1980 to mid-2021. The reanalysis was produced using 3DVar data assimilation of AOD and several other meteorological fields. MERRA-2 uses cubed-sphere horizontal discretization, which serves to mitigate grid spacing singularities that appear in regular Gaussian grids, at an approximate resolution of 0.5° x 0.625° (~50 km), and has 72 hybrid-eta model levels from the surface to the top at 0.01 hPa. MERRA-2 includes 1-hourly and 3-hourly analysis fields for its aerosol diagnostics and meteorological data. For a more thorough and detailed description of MERRA-2 we direct the reader to Gelaro et al. (2017) and Randles et al. (2017).

Designed primarily for research focused on aerosols, the MERRA-2 reanalysis dataset also provides data of the most important trace gases, including O_3, CO and SO_2 (with only NO_2 being unavailable). In MERRA-2, both PM_{10} and $PM_{2.5}$ need to be reconstructed from the available aerosol chemical compounds, which include organic carbon (OC), black carbon (BC), dust (DS), sea salt (SS) and sulphates (SO_4). In this study, the PM_{10} and $PM_{2.5}$ concentrations are computed as follows:

$$[\mathrm{PM_{10}}] = 1.375 \times [\mathrm{SO_4}] + 1.8 \times [\mathrm{OC}] + [\mathrm{BC}] + [\mathrm{DS}] + [\mathrm{SS}] \quad (1a)$$

$$[\mathrm{PM_{2.5}}] = 1.375 \times [\mathrm{SO_4}] + 1.8 \times [\mathrm{OC}] + [\mathrm{BC}] + [\mathrm{DS_{2.5}}] + [\mathrm{SS_{2.5}}] \quad (1b)$$

The 1.375 factor applied to $[SO_4]$ is used here to convert sulfate into ammonium sulfate (assuming full neutralization). The 1.8 factor applied to [OC] accounts for other organic compounds found in organic matter (OM). In recent literature, Eq. 1a and 1b are the most frequently used to reconstruct the PM fields. Equation 1a is used by Provençal et al. (2017b) and also in Ma et al. (2021), though with an additional term to account for aerosol nitrates in the latter. Equation 1b is used by Provençal et al. (2017a, b) and in Ryu and Min (2021), where it is also employed to reconstruct $[PM_{10}]$ by multiplying it with a measurement-based $[PM_{10}]/[PM_{2.5}]$ ratio of 1.75 (computed over the period 2003-2018). Note also that there are large uncertainties in the [OM]/[OC] ratio as it varies in time and space, and other studies have chosen a different value (e.g. 1.4 in Buchard et al. (2016) and Buchard et al. (2017b)) for this factor. Notably, nitrates are currently not available in MERRA-2, even though they can make up a considerable portion of total [PM] Aldabe et al. (2011). To overcome this limitation, some authors such as Ma et al. (2021) have introduced an additional term partly based on observations.

In our study aerosol nitrates are not included in the PM_{10} and $PM_{2.5}$ concentration fields, neither in MERRA-2 nor in CAMS-RA. The potential underestimation due to the absence of nitrates is at least partially compensated by the fact that both reanalyses assimilate total AOD observations, which corrects all PM chemical compounds proportionally and thus minimizes the biases due to the absence of aerosol nitrates.

2.1.3 Air quality observations and GHOST

The EEA observations are accessed from the Globally Harmonised Observational Surface Treatment (GHOST) initiative, a BSC in-house project dedicated to the harmonisation of global air pollution surface observations and its metadata, with the purpose of facilitating a greater quality of observational/model comparison in the atmospheric chemistry community. Besides

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions
Open Access
EGU

the chemical concentration data originally available in the EEA databases, GHOST provides an extended set of metadata, including a variety of quality-assurance (QA) flags, that is used here to eliminate doubtful, non-physical or other faulty data (see Appendix B for a detailed description of the QA filters applied here). To ensure a good temporal representativeness, only daily averages based on at least 18 hourly values (75% threshold) are retained in our study. Given the relatively coarse spatial resolution of both reanalyses, only background stations, of greater spatial representativeness, are considered in the evaluation. This includes stations classified as rural, regional, remote, urban or suburban, thus discarding traffic and industrial point source stations. The latter are generally located in areas with restricted/limited air flow and close to emission sources, thus being heavily affected by day-to-day variability in pollution levels.

2.2 Methodology

Our domain of study extends from 25°W to 45°E in longitude, and from 27°N to 72°N in latitude, thus covering all continental Europe as well as the Canary Islands, Iceland, Western/European Russia, North Africa and the westernmost regions of the Middle East and the Caucasus. For convenience, both CAMSRA and MERRA-2 are regridded over this domain on a common regular longitude-latitude grid at a resolution of 0.2º x 0.2º (roughly 20 km) through bilinear interpolation. The (pointwise) observations are also gridded to this same resolution by averaging (at daily-scale) all the stations available within a given grid cell. Compared to a pointwise-to-gridded comparison, this is expected to partly overcome the issues of spatial representativeness and spatial heterogeneity, although we acknowledge here that more sophisticated methods such as those proposed by Souri et al. (2022) (which employ geostatistical approaches by making use of semivariograms and kriging) might be worth implementing in the future. Table 3 and Fig. 1 provide some information on the observations available over our European domain during 2003-2020, in terms of both pointwise and gridded observations (the total number of observations is typically reduced by a 2-3 factor after the gridding operation). Unfortunately, in situ observations from GHOST are not available for several countries falling within the domain considered in this study, located in Northern Africa (e.g. Morocco, Algeria, Tunis, Lybia, Egypt), Eastern Europe (e.g. Russia, Belarus, Ukraine) and the Middle East (e.g. Israel, Lebanon, Jordan, Syria), thus somewhat limiting the scope of the evaluation, particularly in terms of spatial variability and pollution hot-spots.

Table 3. Number of EEA background stations, number of gridded stations and percentage of domain covered by gridded cells over the period 2003-2020.

Pollutant	EEA stations	Gridded stations	Covered area (%)
O_3	4325	1922	2.44
NO_2	4672	1991	2.53
CO	1074	636	0.81
SO_2	3154	1526	1.94
PM_{10}	5507	2029	2.58
$PM_{2.5}$	2520	1179	1.50

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

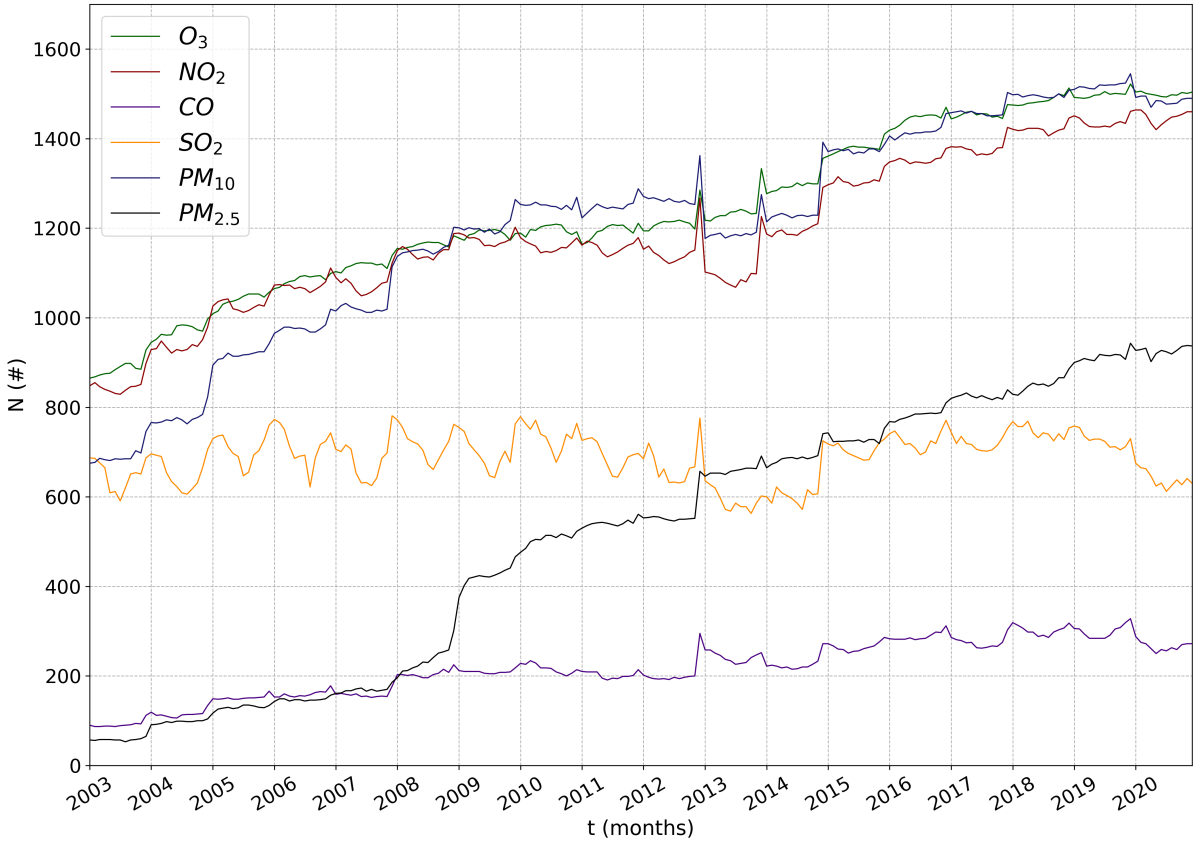

Figure 1. Monthly number of gridded cells with available observational data for O_3, NO_2, CO, SO_2, PM_{10} and $PM_{2.5}$ over the period 2003-2020.

The evaluation is performed on a set of metrics including the (normalized) Mean Bias ((n)MB), the (normalized) Root Mean Square Error ((n)RMSE) and the Pearson Correlation Coefficient (PCC), defined as follows:

$$\mathrm{MB} = \frac{1}{N}\sum_{i=1}^{N}(m_i - o_i) \tag{2a}$$

$$\mathrm{nMB} = \frac{\mathrm{MB}}{\overline{o}} \times 100\% \tag{2b}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(m_i - o_i)^2}{N}} \tag{2c}$$

$$\mathrm{nRMSE} = \frac{\mathrm{RMSE}}{\overline{o}} \times 100\% \tag{2d}$$

$$\mathrm{PCC} = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(m_i - \overline{m})(o_i - \overline{o})}{\sigma_m \sigma_o} \tag{2e}$$

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

Where m_i and o_i are the predicted and observed concentrations, $\overline{m}$ and $\overline{o}$ their means, σ_m and σ_o their standard deviations, and N is the number of points employed to compute the statistics (i.e. number of daily values across all stations). The overlines in Eq. 2a–2e indicate a time-averaged variable.

Annual trends, based on monthly averages over the entire domain (considering only cells and days with available observations to allow for fair comparisons) and reported in Sect. 3, have been computed using Seasonal Theil-Sen estimators, which account for seasonal variability. Statistical significance has been analyzed through Correlated Seasonal Mann-Kendall trend tests, considering both seasonality and autocorrelation. For more detailed information on how the annual trends are computed we refer the reader to Appendix A.

3 Results & Discussion

The evaluation results, alongside its analysis and discussion, are presented in this section. Overall statistics obtained over the European continent during 2003-2020 are provided in Table 4. Annual trends are reported in Table 5 for the different pollutants.

Different aspects of the evaluation results are provided for each pollutant in Fig. 2-7, including (1) monthly time series of concentrations and evaluation statistics, (2) bar plots of country-scale statistics, and (3) maps of mean concentrations (and differences between both reanalyses) over the domain. Each point in the monthly time series corresponds to the median of the monthly-mean values across all individual cells with available observations over the domain. In order to highlight potential spatial differences in pollution patterns across the European continent, country-scale statistics computed over the entire time period and country area are provided for 37 European countries which either are part of, or report data to the EEA, namely Albania (AL), Austria (AT), Bosnia and Herzegovina (BA), Belgium (BE), Bulgaria (BG), Switzerland (CH), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Spain (ES), Finland (FI), France (FR), Hungary (HR), Ireland (IE), Iceland (IS), Italy (IT), Lithuania (LT), Luxembourg (LU), Latvia (LV), Montenegro (ME), North Macedonia (MK), Malta (MT), Netherlands (NL), Norway (NO), Poland (PL), Romania (RO), Serbia (RS), Sweden (SE), Slovenia (SI), Slovakia (SK), Turkey (TR) and the United Kingdom (UK). Additional results are provided in the Appendix A, including seasonal-scale statistics (Tables A1-A6), mean monthly profiles (Fig. A1) and overall statistics for all EEA member countries (Fig. A2-A4).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Table 4. Overall statistics obtained over the period 2003-2020 across Europe, for CAMSRA (subscript C) and MERRA-2 (subscript M). Statistics are shown both on a daily scale (d; over all cells and days in the period 2003-2020) and on a monthly scale (m; weight-averaged by N over all median monthly values). OBS and MOD stand for observational and model concentration, respectively. Reactive gases concentrations are expressed in ppbv, aerosol concentrations in $\mu g\,m^{-3}$ and normalized statistics in %.

Scale	Pollutant	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	O_3	27.4	27.3	41.6	-0.5	51.5	34.9	64.1	0.66	0.53	$7.94\ 10^6$
	NO_2	9.5	6.7	—	-30.1	—	70.7	—	0.52	—	$7.37\ 10^6$
	CO	334.1	190.9	125.6	-42.9	-62.4	91.2	105.4	0.38	0.19	$1.28\ 10^6$
	SO_2	2.8	2.1	2.3	-24.0	-17.1	249.6	253.6	0.19	0.09	$3.05\ 10^6$
	PM_{10}	24.7	21.2	24.5	-14.3	-1.0	85.5	115.6	0.37	0.19	$7.48\ 10^6$
	$PM_{2.5}$	14.9	13.7	11.0	-8.0	-25.6	88.1	96.7	0.41	0.25	$3.05\ 10^6$
m	O_3	26.4	26.6	41.6	3.0	66.2	30.5	72.8	0.59	0.23	216
	NO_2	8.9	6.6	—	-26.4	—	54.1	—	0.53	—	216
	CO	289.3	188.3	121.5	-32.5	-54.1	48.8	64.1	0.53	0.40	216
	SO_2	1.7	1.4	2.1	-13.7	29.3	67.6	85.8	0.28	0.27	216
	PM_{10}	21.2	20.5	21.3	-2.3	4.7	53.4	73.0	0.51	0.25	216
	$PM_{2.5}$	13.1	13.1	10.7	4.2	-13.8	59.2	58.1	0.53	0.45	216

Table 5. Annual trends (Seasonal Theil-Sen estimators, b) over the period 2003-2020 across Europe, for observations (subscript O), CAMSRA (subscript C) and MERRA-2 (subscript M), together with corresponding 99 % confidence intervals (ϵ_-, ϵ_+). Statistically significant annual trends are highlighted in bold. Trends and uncertainty ranges are expressed in $ppbv\ y^{-1}$ and $\mu g\,m^{-3}\ y^{-1}$ for reactive gases and aerosols, respectively. Relative trends (normalized by the mean concentration over 2003-2020) are also indicated in parenthesis.

Pollutant	b_O	ϵ_{O-}	ϵ_{O+}	b_C	ϵ_{C-}	ϵ_{C+}	b_M	ϵ_{M-}	ϵ_{M+}
O_3	+0.08 (+0.31 %/yr)	-0.20	+0.28	**+0.24** (+0.9 %/yr)	+0.02	+0.46	-0.06 (-0.14 %/yr)	-0.22	+0.11
NO_2	**-0.21** (-2.4 %/yr)	-0.31	-0.15	**-0.17** (-2.6 %/yr)	-0.23	-0.12	—	—	—
CO	**-5.32** (-1.8 %/yr)	-7.99	-2.58	**-4.27** (-2.2 %/yr)	-6.10	-3.14	**-0.70** (-0.57 %/yr)	-0.99	-0.29
SO_2	**-0.036** (-2.2 %/yr)	-0.045	-0.028	**-0.073** (-5.0 %/yr)	-0.076	-0.066	**-0.032** (-1.5 %/yr)	-0.047	-0.017
PM_{10}	**-0.39** (-1.8 %/yr)	-0.52	-0.27	**-0.68** (-3.2 %/yr)	-0.82	-0.59	-0.05 (-0.22 %/yr)	-0.20	+0.04
$PM_{2.5}$	**-0.18** (-1.4 %/yr)	-0.27	-0.09	**-0.44** (-3.0 %/yr)	-0.55	-0.39	-0.03 (-0.27 %/yr)	-0.10	+0.02

3.1 Ozone (O_3)

Over the entire period of study, CAMSRA reproduces the observed O_3 concentration reasonably well, with virtually no bias, and reasonable error and correlation (35 % and 0.66, respectively). In comparison, MERRA-2 systematically overestimates $[O_3]$ (+52 %) and shows a larger error and lower correlation (64 % and 0.53, respectively). While CAMSRA reproduces the seasonality of O_3, MERRA-2 substantially underestimates the seasonal amplitude (around 15 ppbv, against more than 20 ppbv

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

in CAMSRA and observations).

Throughout all seasons, median monthly-scale nMB in CAMSRA remains in the range $\pm$ 15 %, with slight negative and positive biases occurring typically in winter and autumn, respectively. Regarding the other metrics, median monthly nRMSE in CAMSRA reaches its worst values in winter (36 %), when the PCC is conversely the best (0.75), whereas an opposite behaviour with low nRMSE and poor PCC can be observed in summer (25 % and 0.43, respectively). A strong seasonal variability is also found in MERRA-2 statistics, although limited to nMB and nRMSE, which are worst in SON (+94 %) and DJF (102 %), respectively. While the reasonable PCC obtained over the entire dataset (0.53) is likely driven by a good ability of MERRA-2 to capture the O_3 seasonality, the much lower monthly PCC values (oscillating around 0.25) suggest that MERRA-2 represents the intra-monthly variability of daily O_3 concentrations very poorly over a large part of the domain. Nonetheless, MERRA-2 is able to reproduce the drop in O_3 concentration typically seen in European observations during JJA, as shown in Fig. A1. In contrast, CAMSRA presents an O_3 concentration peak in summer, a common behaviour of global models due to an inadequate representation of ozone's dry deposition mechanisms, which tend to increase O_3 biases during JJA, as suggested in Valmartin et al. (2014). Over 2003-2020, no statistically significant annual trend (estimated as a Seasonal Theil-Sen slope) of mean [O_3] is observed over Europe, neither in MERRA-2 nor in the observations. However, a significant though low positive increase of +0.24 $\mathrm{ppbv\ y^{-1}}$ is found in CAMSRA (Table 5).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

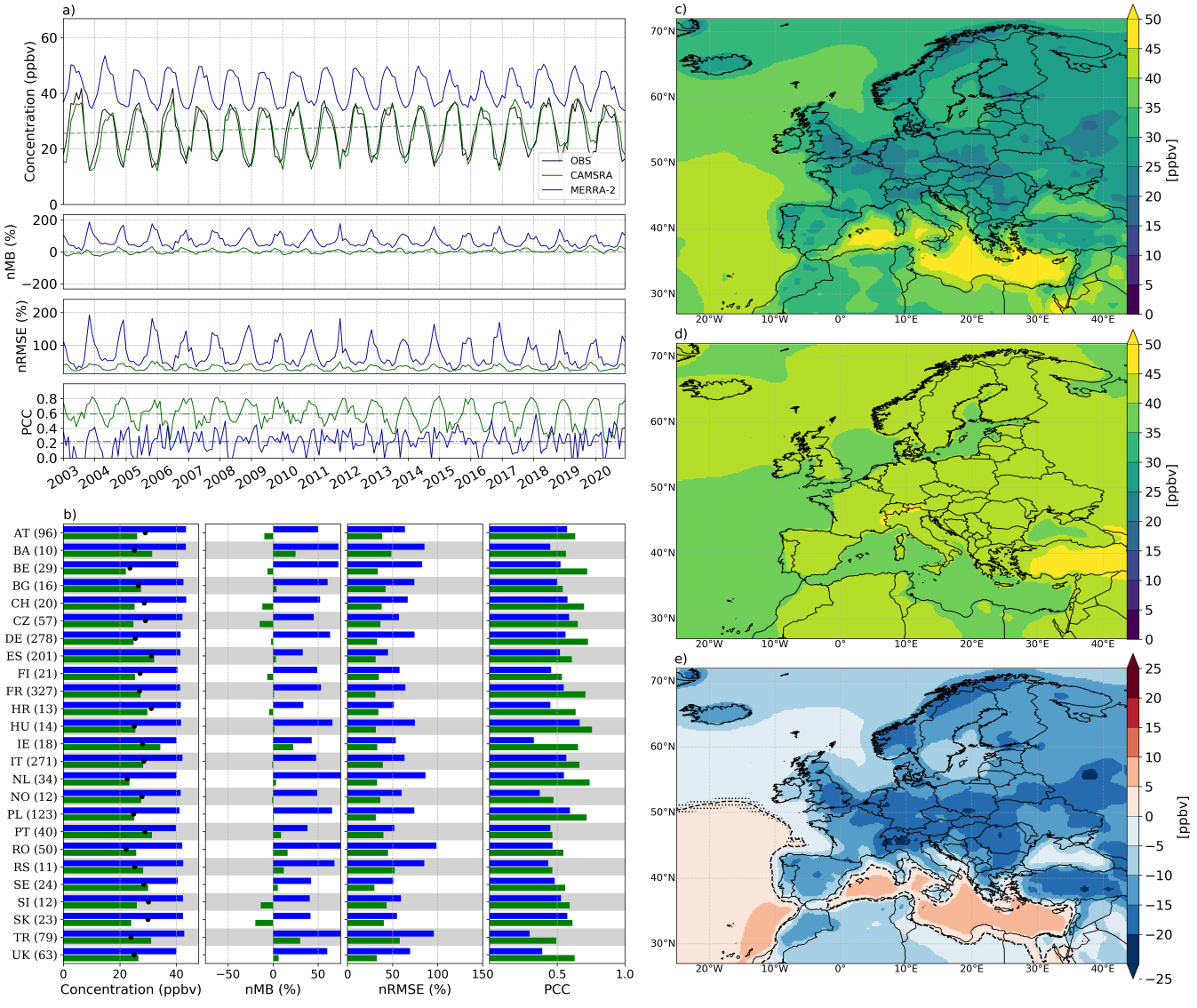

Figure 2. Evaluation of O_3 over Europe depicting: a) Monthly time series of [O_3], nMB, nRMSE and PCC over the period 2003-2020; b) Spatially-averaged [O_3], nMB, nRMSE and PCC for countries with at least 5 cells with observations; c) Mean [O_3] climatology in CAMSRA; d) Mean [O_3] climatology in MERRA-2; e) Differences in Mean [O_3] climatology between CAMSRA and MERRA-2. Black, green and blue colors in a) and b) indicate observations, CAMSRA and MERRA-2, respectively. Numbers between parentheses in b) indicate the cells with available observations. Statistically significant trends, at a 99 % confidence level, are displayed in a). Dotted areas in e) indicate where the differences are not statistically significant at a 99 % confidence level, whereas the black dashed contour stands for a zero difference in concentration between reanalyses.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

The country-level evaluation highlights how CAMSRA outperforms MERRA-2 in every single country across the European continent for every computed statistic, with the greatest differences appearing in Romania (RO) and the Netherlands (NL), and the smallest ones in Greece (EL; only 7 cells with observations are available). In CAMSRA the nMB oscillates roughly between $\pm$ 10 %, with several countries showing virtually no bias, while MERRA-2 displays values in the range +30–70 %. As for the nRMSE, in CAMSRA it remains constrained between 30 and 50 % for all evaluated countries, whereas in MERRA-2 it generally surpasses 50 %, even reaching a value close to a 100 % for Romania (RO). In most countries the PCC does not differ considerably between reanalyses, remaining in the range 0.4–0.7 and slightly higher values for CAMSRA. Ireland (IE) is the only notable exception, as the PCC in CAMSRA is greater by a factor of 2. Compared to the correlations obtained for individual countries, the overall correlation at European scale in CAMSRA is substantially higher (0.7), driven by a good ability to capture the spatial variability of O_3 concentration from one country to another, and in contrast with MERRA-2, as shown in Fig. 2b.

Despite its greater original resolution, MERRA-2 fails to capture the spatial variability of the [O_3] field, with highly homogeneous concentration values over land, ranging from 40 to 45 ppbv (Fig. 2d), likely a result of the lack of accurate ozone sources in the parameterized chemistry and limited sensitivity of OMI measurements to lower tropospheric ozone (note that neither MLS nor OMI provide ozone profile information in the troposphere). A wider range of assimilated products, as seen in Table 2, likely accounts for CAMSRA's better overall performance and greater spatial variability. Nevertheless, we expect the MERRA-2 ozone profile product to be useful for scientific studies that focus on the upper troposphere and the stratosphere, given the high correlations found by Bosilovich et al. (2015) against independent ozonesonde data at these altitudes.

Inness et al. (2019) evaluated surface O_3 against the World Meteorological Office (WMO)'s Global Atmosphere Watch (GAW) background stations, and noticed slightly higher negative biases in winter (with modified nMB down to -40%), though based on a different and smaller set of stations (45 GAW stations, against 4325 EEA background stations gridded into 1922 cells here). Over 2003-2018, Wagner et al. (2021) evaluated CAMSRA surface O_3 concentrations against European Monitoring and Evaluation Programme (EMEP) observations, and found biases within -30 % in winter (driven by underestimated O_3 mostly at higher latitudes) and +30 % in autumn (at all latitudes), in good agreement with our results over the European continent. Although satellite O_3 measurements are extensively assimilated in CAMSRA (11 space-based O_3 products included), Wagner et al. (2021) already demonstrated their minor impact on surface O_3. This may be at least partly due to the relatively low sensitivity of space-borne instruments to lowermost tropospheric O_3 (e.g.Cuesta et al. (2013)). All in all, likely due to a more detailed representation of the tropospheric chemistry, CAMSRA clearly outperforms MERRA-2 in simulating surface O_3 concentrations.

3.2 Nitrogen dioxide (NO_2)

CAMSRA systematically underestimates the concentration of NO_2 (Fig. 3a), with an overall bias of -30 % (Table 4). The seasonal variability is relatively well captured, though with the monthly-scale bias increasing from -23 % in summer to -34 % in winter (Table A2). Such negative biases are expected given the short chemical lifetime of NO_2 and the coarse spatial resolution of CAMSRA. Note that Ryu and Min (2021) also found a large underestimation of NO_2 in winter over South Korea

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

(around -10 ppbv, against -2 ppbv in summer). Additionally, Inness et al. (2019) evaluated NO_2 against surface data from 4 GAW stations in Europe over 2003-2016, finding very limited biases with a larger underestimation of NO_2 concentration during winter. Our results show good agreement with the wintertime peak underestimation, though biases are significantly larger (evaluated here against 4672 EEA background stations, gridded into 1991 cells). Wagner et al. (2021) evaluated NO_2 tropospheric columns against retrievals from the SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCHIAMANY) and the Global Ozone Monitoring Experiment-2 (GOME-2), finding an underestimation of European NO_2 hotspots which peaks in winter, a behaviour similar to the one observed here for surface concentration. CAMSRA also shows a relatively large overall nRMSE of 71 %, reduced to around 53 % at seasonal scale, without strong differences from one season to another. The PCC ranges between 0.52 in winter and 0.38 in summer, though a slight correlation improval is found at monthly-scale, with PCC values in the range 0.47–0.60 (Fig. 3a). In terms of long-term trends, the significant decrease of NO_2 concentration observed over 2003-2020 (-0.21 ppbv y^{-1}) is slightly underestimated by the reanalysis (-0.17 ppbv y^{-1}, i.e. differing by a 1.2 factor). In relative terms, these decreasing concentration trends found for NO_2 in the observations and CAMSRA (-2.4 and -2.6 % y^{-1}, respectively) are close to the -2.0 % y^{-1} NOx emission trend reported by the EEA over the period 1990-2019 in its emission inventory report (European Environment Agency, 2021). Although the observed NO_2 concentrations shown on Fig. 3a) decreased in 2020 due to the COVID-19 pandemic (Bauwens et al. (2020), Vîrghileanu et al. (2020)), the observations accessed through GHOST do not reflect this sharp decrease, as background stations used in the present study include not only urban but also rural stations. The change in CAMSRA also appears less pronounced, potentially due to the coarse resolution of the reanalysis, but most likely due to CAMSRA following the RCP8.5 for emissions after 2010 Granier et al. (2011).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

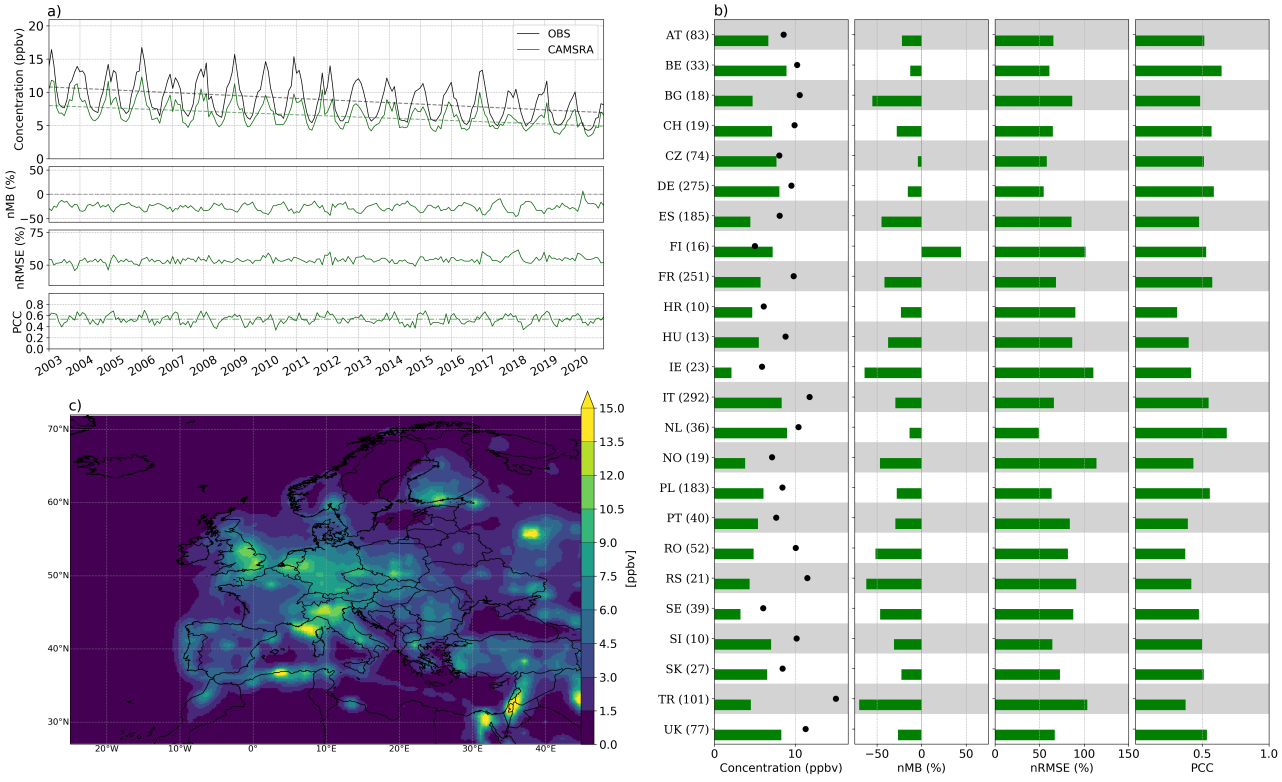

Figure 3. Evaluation of NO_2 over Europe depicting: a) Monthly time series of [NO_2], nMB, nRMSE and PCC over the period 2003-2020; b) Spatially-averaged [NO_2], nMB, nRMSE and PCC for countries with at least 5 cells with observations; c) Mean [NO_2] climatology in CAMSRA. Black and green colors in a) and b) indicate observations and CAMSRA, respectively. Numbers between parentheses in b) indicate the cells with available observations. Statistically significant trends, at a 99 % confidence level, are displayed in a).

At country-level (considering only countries with more than 5 cells containing observations), most nMBs fall roughly between -10 and -60 %, with the notable exception of Finland (FI), where a strong overestimation (+40 %) is found. The nRMSE ranges from 50 to 110 %, depending on the country considered. The PCC remains generally around 0.5, though countries with fewer measuring stations available tend to present lower PCC values (Fig. 3b).

The spatial variability of the [NO_2] field across the European continent is consistent with the location of dense urban areas (e.g. Paris, Moscow, Barcelona, Oslo, Algiers), highly industrialized regions (e.g. Po River basin, Rhine-Rühr valley, Silesia) and busy shipping lanes (e.g. Mediterranean, English Channel, Portuguese coastline). In sparsely populated areas, low industrialized regions and the open seas [NO_2] levels remain below 3 or even 1.5 ppbv (Fig. 3c).

3.3 Carbon monoxide (CO)

As shown in Fig. 4a, both CAMSRA and MERRA-2 systematically underestimate the concentration of CO (overall nMB of -43 and -62 %, respectively), with amplified negative biases during DJF. The overall nRMSE is high in both reanalyses (91 and

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions
Open Access
EGU

105 %, respectively), with again a lower performance in winter. Wagner et al. (2021) evaluated CO in Europe against data from GAW stations over the period 2003-2018, reporting a persistent underestimation (modified nMB ranging from -10 to -20 %) of surface CO. Ryu and Min (2021) also found similar results in South Korea, with a systematic underestimation that increases significantly in winter. In contrast, Inness et al. (2019) reported an overall overestimation of around 10 ppbv for the period 2003-2017, which is likely due to the different set of stations taken into account (15 GAW stations, most of them regional and several of them located at high altitude).

At monthly-scale, the median CO concentration, nMB and nRMSE in CAMSRA partially capture the seasonality, showing a better performance in spring (MAM; -24 % and 43 %, respectively) and a strong wintertime deterioration (-42 % and 56 %, respectively). As seen for O_3, the PCC follows the opposite behaviour, with better performance in DJF (0.59) and a late springtime deterioration (0.47). In contrast to CAMSRA, MERRA-2 is unable to reproduce the seasonal variability of surface CO concentration, despite the nMB and nRMSE displaying significant variability throughout the different seasons. Moreover, the strong statistically significant decrease in CO observed across Europe over 2003-2020 (-5.32 $\mathrm{ppbv\ y^{-1}}$) is moderately underestimated in CAMSRA (-4.27 $\mathrm{ppbv\ y^{-1}}$), although less dramatically than in MERRA-2, where CO remains roughly constant over all the period study, displaying a small negative trend (-0.70 $\mathrm{ppbv\ y^{-1}}$). Similarly, Ryu and Min (2021) report a severe underestimation along with an absence of variability in surface CO over the period 2003-2020 in MERRA-2. In European Environment Agency (2021) the EEA reports a CO emission trend of -2.3 % $\mathrm{y^{-1}}$ over 1990-2019, relatively close to the concentration trends found in CAMSRA, -2.2 % $\mathrm{y^{-1}}$, and the observations, -1.8 % $\mathrm{y^{-1}}$. In 2020, MERRA-2 shows a very large increase of CO concentration across most of Europe, in contrast to both CAMSRA and the observations. The overall PCC in MERRA-2 is poor (0.19), although better PCC values (~0.40) are found at monthly-scale. The overall moderate PCC of CAMSRA (0.38) is only comparable to MERRA-2 during summertime, being also slightly better when computed on a monthly basis (0.53).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

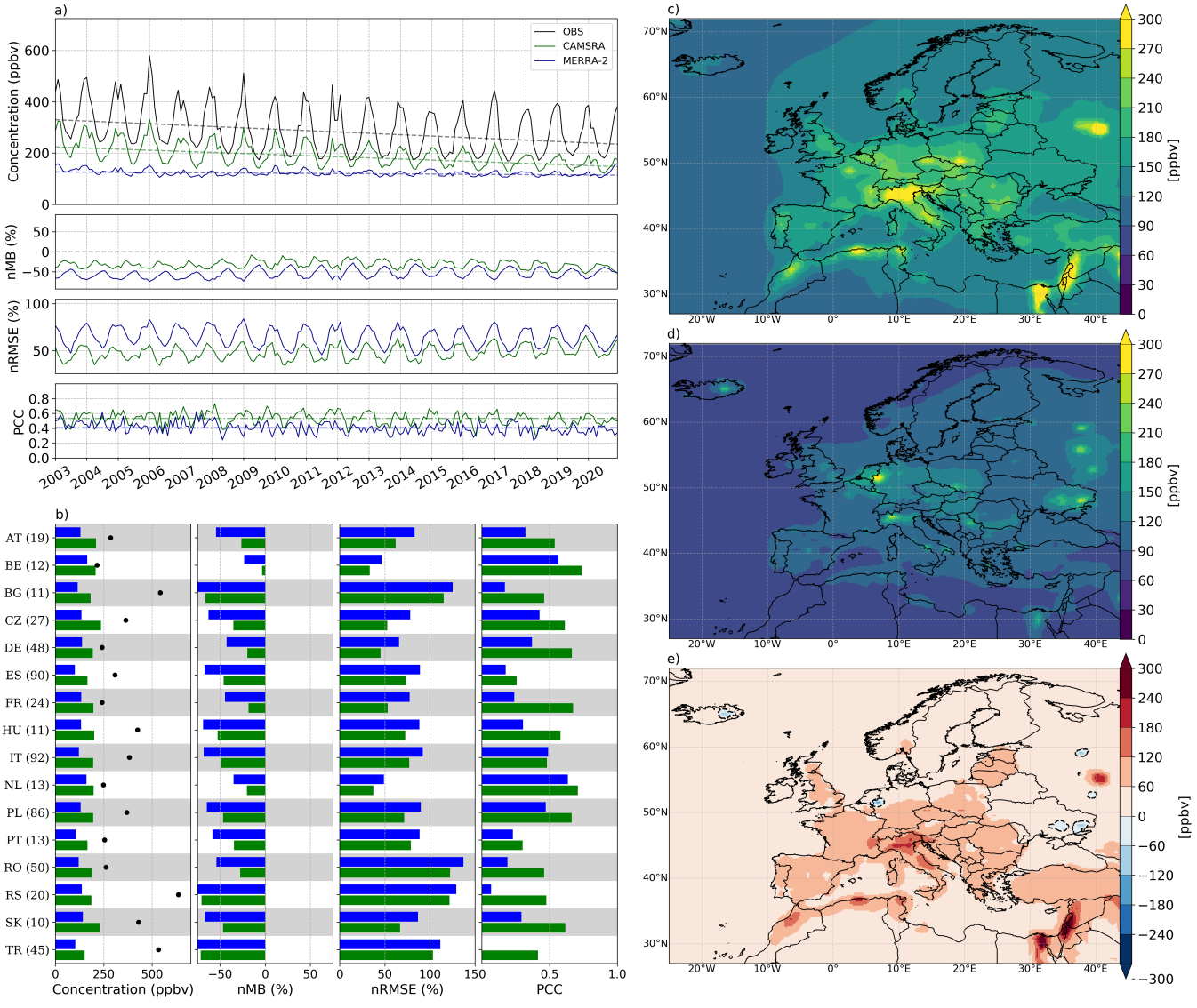

Figure 4. Similar to Fig. 2 but for CO.

This CO underestimation typically spreads over all the European continent, with strong differences across countries. As CO is not assimilated in MERRA-2, but simulated by the GEOS-5 modeling system, it is likely this underestimation originates due to the model considering too low CO emissions and/or excessively large CO sinks. In both reanalyses, the best scores in terms of bias, PCC and nRMSE are found in Belgium, and to a lesser extent in the Netherlands. Conversely, far poorer results are obtained in Bulgaria, Serbia and Romania. Although different, the nMB and nRMSE in both reanalyses typically show comparable variations from one country to another. Both CAMSRA and MERRA-2 show CO hot-spots over large

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

315 urban areas and/or highly industrialised regions (e.g. Moscow, Po River basin). However, compared to CAMSRA, MERRA-2 highlights some additional hot-spots, for instance on the Vatnajökull ice cap, located in Iceland, a region well known for its sub-glacial volcanoes (e.g. Grímsvötn) which experience frequent degassing. Another significant hot-spot is found in the Donets Basin (eastern Ukraine), an important coal-mining region. Two other CO hot-spots can be seen south and north of Moscow, corresponding to the cities of Voronezh and Yaroslavl, respectively, but it is unlikely that CO levels comparable to those of 320 Moscow are found in these intermediate sized cities (Fig. 4c,d).

The reanalyses also differ in the locations where CO concentration is higher across Europe (Po River basin in CAMSRA; Rhine-Rühr valley in MERRA-2). CAMSRA highlights the highest CO concentrations in Europe in the Po River basin and displays moderate concentration values in the Rhine-Rühr area, which suggests a longer CO lifetime in the former given that European Environment Agency (2021) reports the highest CO emissions, over all the period 1990-2019, in Germany. Therefore, 325 in sharp contrast with CAMSRA, MERRA-2 obviously fails to capture the chemistry processes of surface CO, with a likely underestimation of emission sources and/or too large CO sinks, thus being unable to reproduce the spatiotemporal variability of surface CO observed over Europe.

3.4 Sulphur dioxide (SO_2)

When computed over the entire dataset (Table 4), the statistics of CAMSRA and MERRA-2 show extremely poor nRMSE and 330 PCC (around 250 % and 0.09–0.19, respectively), and moderate though comparable negative bias (around -20 %). Surprisingly, on a monthly basis (Fig. 5a), the median nMB at monthly-scale in MERRA-2 appears mostly positive (+29 %) and increasing along time, which appears to be due to the fact that these median values hide a very strong underestimation of SO_2 in some parts of Europe (e.g. Turkey).

The median monthly-scale nMB in CAMSRA tends to improve between late spring and early autumn, even reaching values 335 close to 0 %, though it remains negative throughout the rest of the year, dropping to -31 % in winter. Ryu and Min (2021), though finding a large SO_2 concentration overestimation over South Korea, as opposed to the underestimation shown here for Europe, found a similar nMB seasonality, with nMB improving (~+2 ppbv) and worsening (~+6 ppbv) in warm and cold months, respectively. In MERRA-2 the median nMB oscillates roughly around +30 % (with a $\pm$ 15 % range), though it suffers a large increase (with significant intra-annual variability) from 2013 onwards due to an important decrease in observed 340 SO_2 concentration. A similar increase is also observed for the nRMSE. The monthly-scale nRMSE and PCC remain roughly constant (when averaged across all months) throughout all seasons, both in CAMSRA (around 68 % and 0.27, respectively) and in MERRA-2 (around 85 % and 0.27, respectively), though the latter displays much stronger seasonal variability. Note also the large difference between the monthly-scale nRMSE (68–85 %) and the overall nRMSE (around 250 %). The statistically significant negative trend found in the observations (-0.036 ppbv y^{-1}) is largely overestimated by CAMSRA (-0.073 ppbv 345 y^{-1}), and well reproduced by MERRA-2 (-0.032 ppbv y^{-1}). In European Environment Agency (2021) the EEA reports a SO_2 emission trend of -3.2 % y^{-1} over 1990-2019, falling between the concentration trend found in CAMSRA, -5.0 % y^{-1}, and the one found in the observations, -2.2 % y^{-1}, and MERRA-2, -1.5 % y^{-1}.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

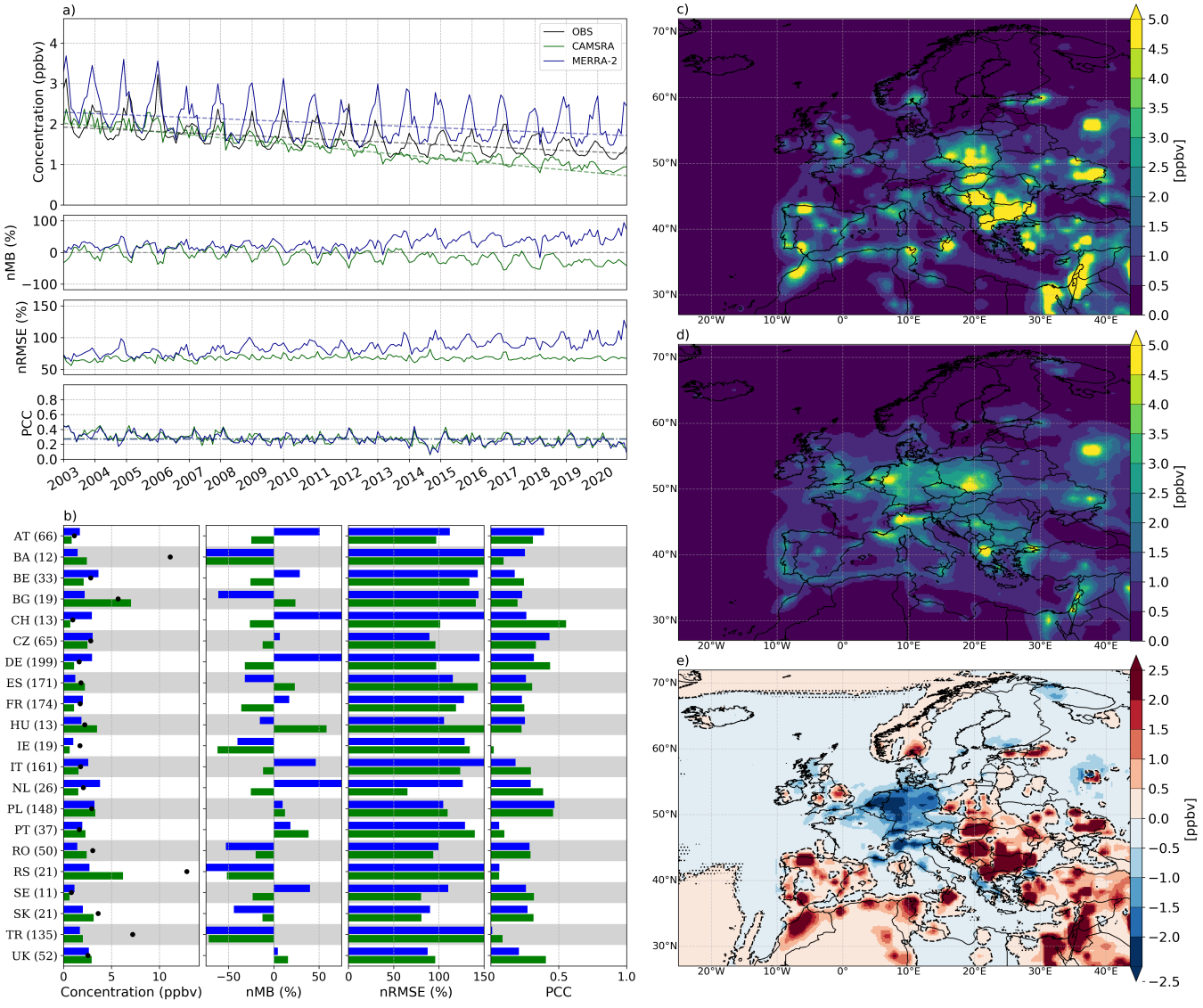

Figure 5. Similar to Fig. 2 but for SO_2.

The country-level evaluation for SO_2 shows very heterogeneous results across countries, differing substantially from the observed behaviour in previously examined reactive gases. The nMB presents a wide range of variation, with certain countries showing virtually no bias (e.g. Poland, Czechia, United Kingdom) and others presenting biases well over ± 50 % (e.g. Turkey, Serbia, Bosnia, Ireland). Both the nRMSE and PCC display a poor performance, ranging roughly within 75–150 % and 0.10–0.50, respectively (Fig. 5b).

In both reanalyses, the heterogeneous distribution of [SO_2] is consistent with the location of highly industrialized areas (e.g.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

Po River basin, Rhine-Rühr valley) and coal-mining regions (e.g. Silesia, Donets Basin, Balkans). To a minor extent, there are also significant SO_2 concentrations in dense urban areas and along shipping lanes. Surprisingly, no SO_2 hot-spot is detected over the Icelandic Vatnajökull ice cap in MERRA-2, which contrasts with the fact that SO_2 emissions represent a large fraction of volcanic gases. The reanalyses show sharp differences in the regions where highest concentrations of SO_2 are present, with CAMSRA favouring coal-mining regions and dense urban areas, and MERRA-2 showing a more balanced distribution between them (Fig 5c,d,e). Overall, MERRA-2 clearly outperforms CAMSRA in Europe over the period 2003-2020, though a significant worsening of nMB and nRMSE is observed from 2013 onwards. Anthropogenic SO_2 emissions in MERRA-2 are obtained from AeroCom Phase II (Diehl et al. (2012)) and EDGAR v4.2 (European Commission, 2011 [https://edgar.jrc.ec.europa.eu/]) inventories, with emissions being persisted in the model according to the ending year of each inventory (Randles et al., 2017). Thus, the observed deterioration after 2013 likely arises due to an emission overestimation which propagates throughout the time period where no updated SO_2 emissions are available.

3.5 Coarse particulate matter (PM_{10})

Overall, CAMSRA and MERRA-2 reanalyses represent moderately well surface PM_{10} concentrations over Europe (Table 4), with a reasonable negative nMB (-14 %) for CAMSRA and virtually no bias for MERRA-2 (-1 %), but poor nRMSE (85 and 115 %, respectively) and PCC (0.37 and 0.19, respectively).

At monthly-scale, the median nMB in CAMSRA oscillates roughly around 0 %, with positive bias during spring (+13 %) and a strong underestimation in DJF (-20 %), while the nRMSE and PCC show a strong and complex intra-annual variability without a clear seasonal pattern (remaining in the range 73–96 % and 0.30–0.44, respectively). In comparison, nRMSE and PCC in MERRA-2 follow a clear seasonal behaviour, with strongly deteriorated results during winter (87 % and 0.04, respectively) but better summertime performance (59 % and 0.40, respectively). Surprisingly, the median nMB in MERRA-2 also peaks in JJA (+20 %), with virtually no bias in SON, and a wintertime low (-18 %). Ryu and Min (2021) found a slightly positive PM_{10} bias for CAMSRA in South Korea over 2003-2018, while for MERRA-2 their findings suggest a clear underestimation that worsens significantly in winter, the latter being in good agreement with our results over Europe. The statistically significant negative trend present in the observations (-0.39 $\mu g\,m^{-3}\,y^{-1}$) is strongly overestimated by CAMSRA (-0.68 $\mu g\,m^{-3}\,y^{-1}$) and severely underestimated by MERRA-2 (-0.05 $\mu g\,m^{-3}\,y^{-1}$), being the latter not statistically significant (at a 99 % confidence level). In European Environment Agency (2021) the EEA reports a PM_{10} emission trend of -1.7 % y^{-1} over 2000-2019, far from the concentration trend of CAMSRA, -3.2 % y^{-1}, and close to the one found in the observations, -1.8 % y^{-1}.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

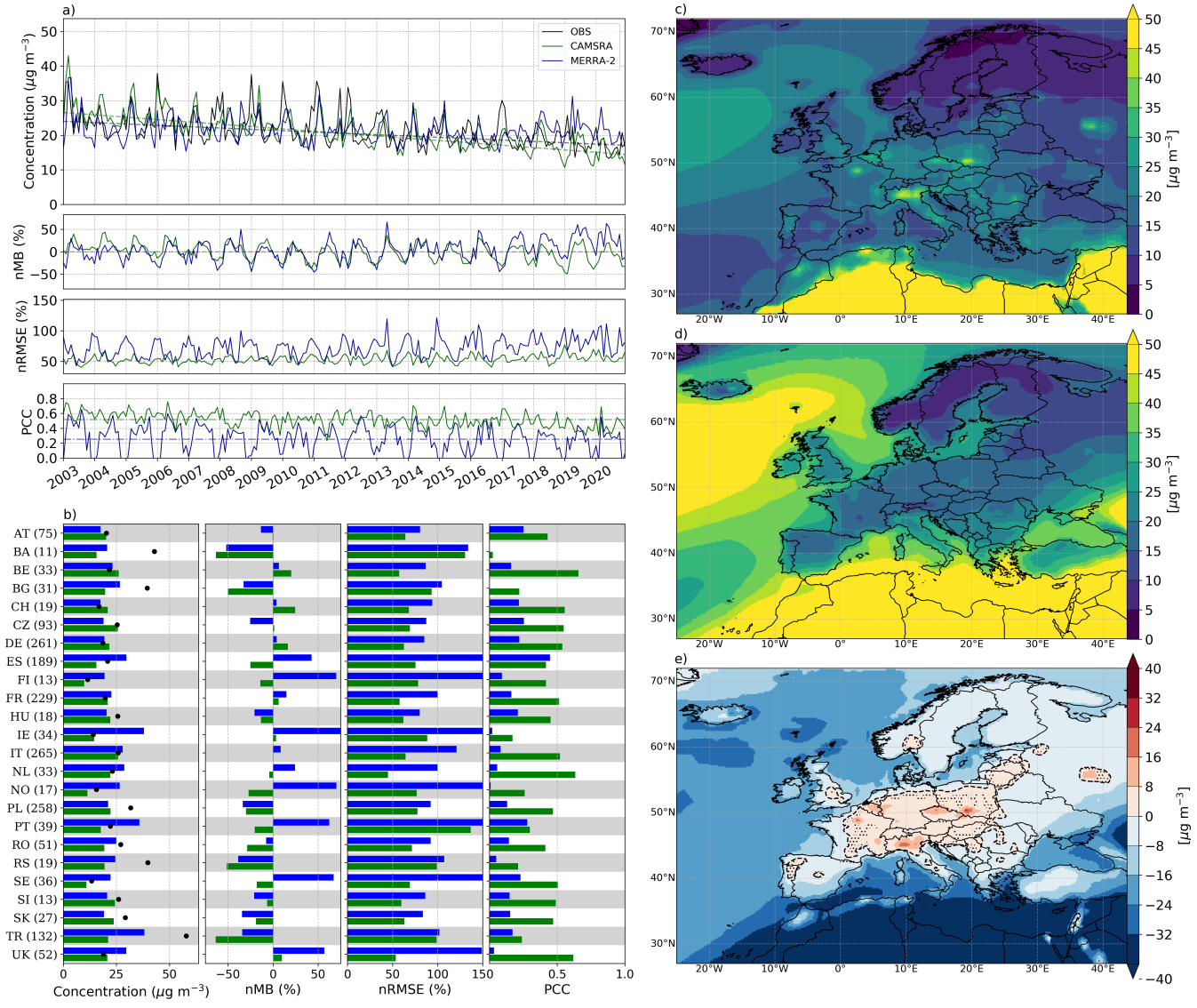

Figure 6. Similar to Fig. 2 but for PM_{10}.

At country-level, CAMSRA tends to outperform MERRA-2 in most countries, with lower nRMSE (50–100 % and 75–150 %, respectively) and higher PCC values (0.3–0.6 against 0.1–0.4, respectively). The nMB presents a wide range of variation in both reanalyses, with certain countries showing virtually no bias (e.g. Austria, Italy, Germany, France) and other countries presenting biases well over ± 50 %. Though MERRA-2 presents lower nMB values than CAMSRA in several countries (e.g. Romania, Turkey, Germany, Belgium, Bulgaria, Bosnia), both the nRMSE and PCC point towards a greater performance by CAMSRA in all cases (Fig. 6b).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

Again, despite its finer resolution, MERRA-2 displays a more homogeneous concentration over land in which the multiple PM_{10} hotspots found in CAMSRA - in industrialized regions (e.g. Po River basin, Silesia) and in certain urban areas (e.g. Paris, Moscow, Madrid) - are missing. In addition, it also shows much higher PM_{10} concentrations over the open seas and Northern Africa, where sea salt and dust sources are predominant. It thus seems that Eq. 1a severely overestimates the surface concentrations of sea salt and dust, as shown in Fig. 6d), with MERRA-2 displaying differences of more than a 100 $\mu g\,m^{-3}$, particularly over desert areas. Overall, CAMSRA unambiguously outperforms MERRA-2 in capturing the spatiotemporal variability of PM_{10} surface concentrations over Europe.

3.6 Fine particulate matter ($PM_{2.5}$)

CAMSRA reproduces moderately well surface $PM_{2.5}$ concentrations over Europe (Table 4), with a low negative nMB (-8 %) but poor nRMSE (88 %) and moderate PCC (0.41), while MERRA-2 presents overall worst statistics, with poor nMB (-26 %), nRMSE (97 %) and PCC (0.25).

The median monthly-scale nMB in CAMSRA remains in the range $\pm$ 30 %, with the greater over- and underestimations occurring in summer (+28 %) and winter (-26 %), respectively. In contrast, MERRA-2 shows limited biases ($\pm$ 10 %) during spring and summer, and a strong wintertime deterioration (-40 %). Interestingly, the MERRA-2 and CAMSRA nMB time series', while initially displaying an absolute difference of ~50 %, converge from 2017 onwards. Similarly to the behaviour observed for PM_{10}, the median nRMSE and PCC in CAMSRA show a strong intra-annual variability without a clear seasonal pattern (remaining in the range 54–65 % and 0.50–0.55, respectively). As for MERRA-2, both the nRMSE and the PCC present significant seasonal variability, with better performance in summer (44 % and 0.55, respectively) and a heavy wintertime deterioration (76 % and 0.32, respectively). Similar results are reported by Provençal et al. (2017a) when evaluating MERRA-1 over Europe, with an overall limited negative bias and a deterioration in winter. Note also that Navinya et al. (2020) evaluated $PM_{2.5}$ in MERRA-2 against 20 background stations in India, finding a moderate negative nMB (-34 % ; -27 $\mu g\,m^{-3}$) and a larger wintertime underestimation, in agreement with our results over Europe. As for $PM_{2.5}$, the statistically significant negative trend present in the observations (-0.18 $\mu g\,m^{-3}\,y^{-1}$) is heavily overestimated by CAMSRA (-0.44 $\mu g\,m^{-3}\,y^{-1}$), and completely missed by MERRA-2. In European Environment Agency (2021) the EEA reports a $PM_{2.5}$ emission trend of -1.9 % y^{-1} over 2000-2019 which, while not strictly comparable to a concentration trend as previously mentioned, falls between the trend found in CAMSRA, -3.0 % y^{-1}, and the one found in the observations, -1.4 % y^{-1}.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

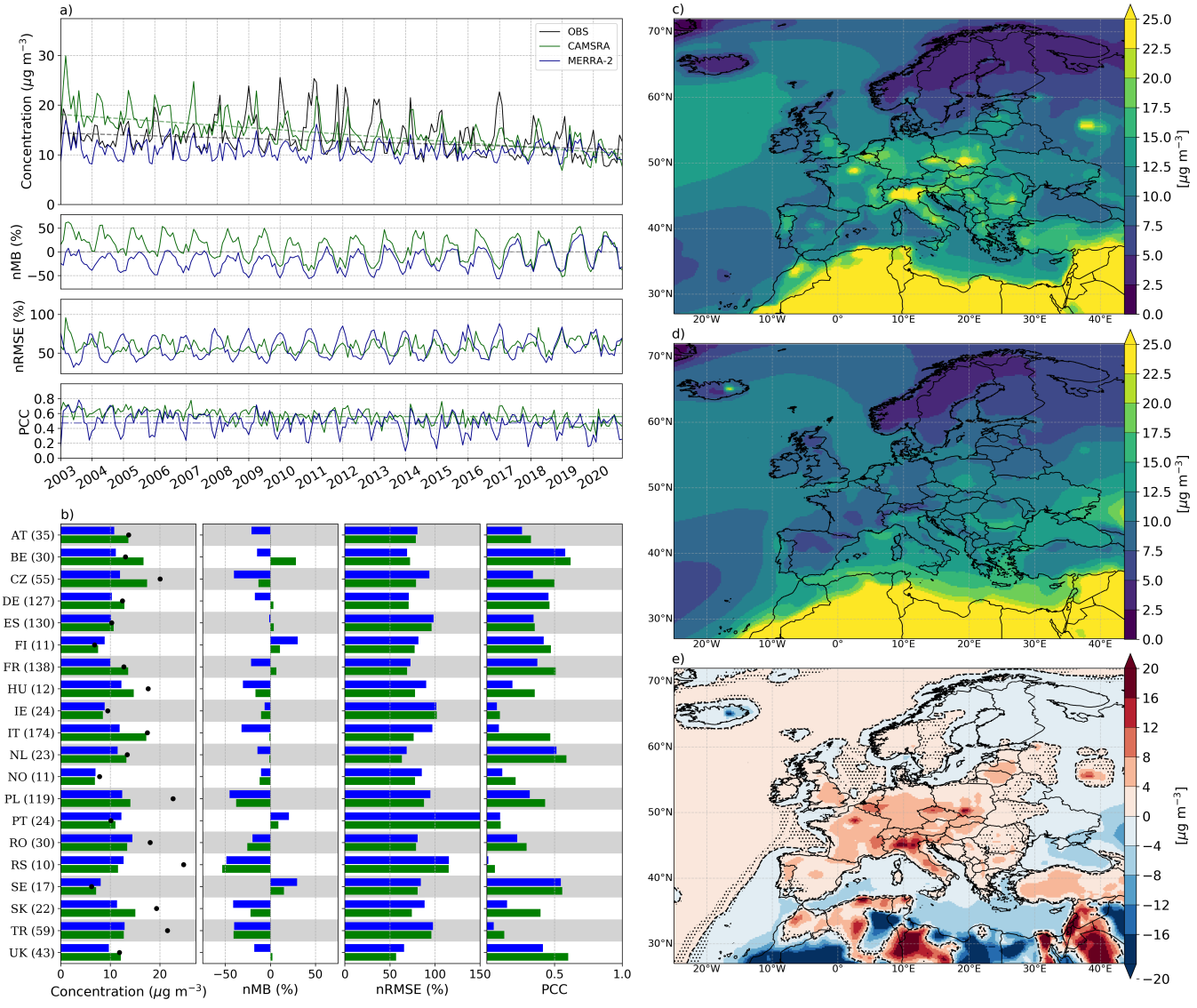

Figure 7. Similar to Fig. 2 but for $PM_{2.5}$.

At country-level (Fig. 7b), the differences in $PM_{2.5}$ between CAMSRA and MERRA-2 are less pronounced than for PM_{10}, especially for the PCC (with most values in the range 0.1–0.5), and to a lesser extent for the nRMSE (with most values in the range 70–100 %). The nMB presents a similar behaviour to the one observed for PM_{10}, with certain countries showing virtually no bias (e.g. Spain, Ireland, Netherlands) and other countries presenting large negative biases (e.g. Turkey, Serbia, Poland).

The spatial variability of $PM_{2.5}$ concentration remains close to the one obtained for PM_{10} in all regions and in both reanalyses,

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

except over the open seas, where MERRA-2 no longer shows exceedingly large sea salt levels (which thus prevail mostly in the coarse mode). The surface pollution hot-spots present in Fig. 7 are essentially the same ones that appear in Fig. 6, though a notable exception is observed in MERRA-2 over Iceland. A large $PM_{2.5}$ concentration peak, also visible for PM_{10}, can be spotted in Iceland's time series during 2010, surpassing 100 $\mu g\,m^{-3}$, likely due to the Eyjafjallajökull volcanic eruption, which emitted very large amounts of volcanic ash (Thorsteinsson et al., 2012).

4 Summary and conclusions

In this work we have performed a long-term (2003-2020) multi-pollutant evaluation of CAMSRA and MERRA-2 global atmospheric composition reanalyses against in situ surface measurements over the European continent. In contrast to past evaluation studies, we have included a more extended set of surface stations (21252 EEA background stations between all pollutants, gridded into 9283 cells), quality-assured using GHOST metadata and gridded in order to limit, to some extent, representativeness issues.

As a summary, CAMSRA unambiguously outperforms MERRA-2 in representing surface pollutant concentrations across Europe. Differences are particularly clear for O_3 and CO, but also persist for PM_{10} and $PM_{2.5}$. CAMSRA clearly achieves the best results for O_3, while statistics for the other pollutants show more mixed results: reasonable error and correlation for NO_2 but strong underestimation, low biases but relatively poor errors and correlations on SO_2, PM_{10} and $PM_{2.5}$. SO_2 is the only pollutant for which MERRA-2 clearly outperforms CAMSRA. With MERRA-2 being designed mainly for research on aerosols, the reanalysis indeed provides statistics on PM_{10} and $PM_{2.5}$ in line with CAMSRA, but the latter still gives slightly better results over Europe, especially for $PM_{2.5}$, with overall lower biases and a better characterization of its spatial variability.
Compared to CAMSRA, MERRA-2 benefits from a slightly finer spatial resolution, but assimilates a much less diversified set of satellite products. However, recent evaluations of CAMSRA have noticed that this assimilation only partially improves the representation of pollutant concentrations at the surface, despite a clear improvement being found in the entire troposphere. Although at least partly due to the still coarse spatial resolution of CAMSRA, a large if not dominant part of the model-versus-observation differences found here at the surface are likely explained by errors on emissions and/or sinks. Therefore these global reanalysis datasets need to be carefully bias-corrected with surface observations in order to be used in long-term air pollution and impact studies.

The surface pollution evaluation carried out in this work can serve as a milestone for future air quality and other pollution-related studies. In that regard, further advancements in the field could be focused in developing new statistical approaches to merge surface observations with reanalysis data. As global atmospheric composition reanalyses do not assimilate data at the surface, ground level measurements can be employed, through different statistical methods, to bias correct and improve raw model output statistics, thus leading to more robust reanalysis products. This improved characterization of the spatiotemporal variability of surface air pollution would open the door to improved health impact and air quality assessments, while also helping design and implement more effective air pollution reduction policies.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022


Geoscientific Model Development Discussions

Eventually, if reanalyses are to be used in long-term health impact studies, consistent statistical approaches to combine observational data with reanalysis data need to be developed.

Data availability. The observational data, obtained from EEA AIRBASE and AQ e-Reporting air quality datasets, and reanalysis data, obtained from CAMSRA and MERRA-2, used in this study are publicly available.

Appendix A: Trends

Given our monthly time series does not contain tied or missing values, the Seasonal Mann-Kendall statistic, S', and its variance, $Var[S']$, can be obtained as follows:

$$\mathrm{S'} = \sum_{g=1}^{m} S_g = \sum_{g=1}^{m} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_{jg} - x_{ig}) \tag{A1a}$$

$$\mathrm{Var[S']} = \sum_{g=1}^{m} \sigma_g^2 + \sum_{g,h} \sigma_{gh} = \frac{1}{18}[n(n-1)(2n+5)] + \frac{1}{3}[K_{gh} + 4\sum_{j=1}^{n} R_{jg}R_{jh} - n(n+1)^2] \tag{A1b}$$

$$\mathrm{K_{gh}} = \sum_{i=1}^{n-1} \sum_{j=1}^{n} sgn[(x_{jg} - x_{ig})(x_{jh} - x_{ih})] \tag{A1c}$$

Where n and m are the number of years and seasons (i.e. here monthly values), respectively, S_g is the Mann-Kendall statistic for each g_{th} season, R_g and R_h are Spearman's correlation coefficients for seasons g and h, respectively, and $sgn(x)$ is the sign function. Seasonal Theil-Sen slopes (i.e. annual trends) are then derived from S' (Hussain and Mahmud (2019); Hipel and McLeod (1994); Hirsch and Slack (1984)). The confidence intervals, derived from $Var[S']$, are computed accounting for seasonality but not for autocorrelation, mainly due to the detection of a potential bug in the function *correlated_multivariate_test* from the *Python* library *pyMannKendall* (Hussain and Mahmud, 2019), which at the date of this work's submission remained unresolved.

Appendix B: QA flags

Using the metadata available in GHOST, a quality assurance screening is applied by removing all air quality observations associated with a set of flags detailed in Table B1. In addition, we detected a few very low CO concentrations in specific regions during specific time periods, which we suspect originate from errors of units when the Member State reported its observations to the EEA. Therefore, as a precautionary measure, all CO hourly observations below 1 ppbv were discarded in this study.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Table B1. Description of the GHOST quality-assurance flags used on the EEA air quality observational dataset.

Flag	Description
0	Measurement is missing (i.e. NaN).
1	Value is infinite – occurs when data values are outside of the range that *float32* data type can handle (-3.4E+38 to +3.4E+38).
2	Measurement is negative in absolute terms.
3	Measurement is equal to zero.
6	Measurements are associated with data quality flags given by the data provider which have been decreed by the GHOST project architects as being associated with substantial uncertainty/bias.
8	After screening by key QA flags, no valid data remains to average in the temporal window.
10	The measurement methodology used has not yet been mapped to standardised dictionaries of measurement methodologies.
18	The specific name of the measurement method is unknown.
20	The primary sampling is not appropriate to prepare the specific parameter for subsequent measurement.
21	The sample preparation is not appropriate to prepare the specific parameter for subsequent measurement.
22	The measurement methodology used is not known to be able to measure the specific parameter.
23	The specific measurement methodology has been decreed not to conform to QA standards as the method is not sufficiently proven/ subject to substantial biases/uncertainty.
72	Measurement is below or equal to the preferential lower limit of detection.
75	Measurement is above or equal to the preferential upper limit of detection.
82	The preferential resolution for the measurement is coarser than a set limit (variable by measured parameter).
83	The resolution of the measurement is analysed month by month. If the minimum difference between observations is coarser than a set limit (variable by measured parameter), measurements are flagged.
90	Check for persistently recurring values. Check is done by using a moving window of 9 measurements. If 5/6 (i.e. 83.33%) of values in the window are the same then the entire window is flagged.
91	Check for persistently recurring values. Check is done by using a moving window of 12 measurements. If 9/12 (i.e. 75%) of values in the window are the same, then the entire window is flagged.
92	Check for persistently recurring values. Check is done by using a moving window of 24 measurements. If 16/24 (i.e. 66.66%) of values in the window are the same, then the entire window is flagged.
110	The measured value is below or greater than scientifically feasible lower/upper limits (variable by parameter).
111	The median of the measurements in a month is greater than a scientifically feasible limit (variable by parameter).
112	Data has been reported to be an outlier through data flags by the network data reporters (and not manually checked and verified as valid).
113	Data has been found and decreed manually to be an outlier.
131	2 out of 3 months' distributions are classed as Zone 6 or higher, suggesting there are potentially systematic reasons for the inconsistent distributions across the 3 months.
132	4 out of 6 months' distributions are classed as Zone 6 or higher, suggesting there are potentially systematic reasons for the inconsistent distributions across the 6 months.
133	8 out of 12 months' distributions are classed as Zone 6 or higher, suggesting there are potentially systematic reasons for the inconsistent distributions across the 12 months.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

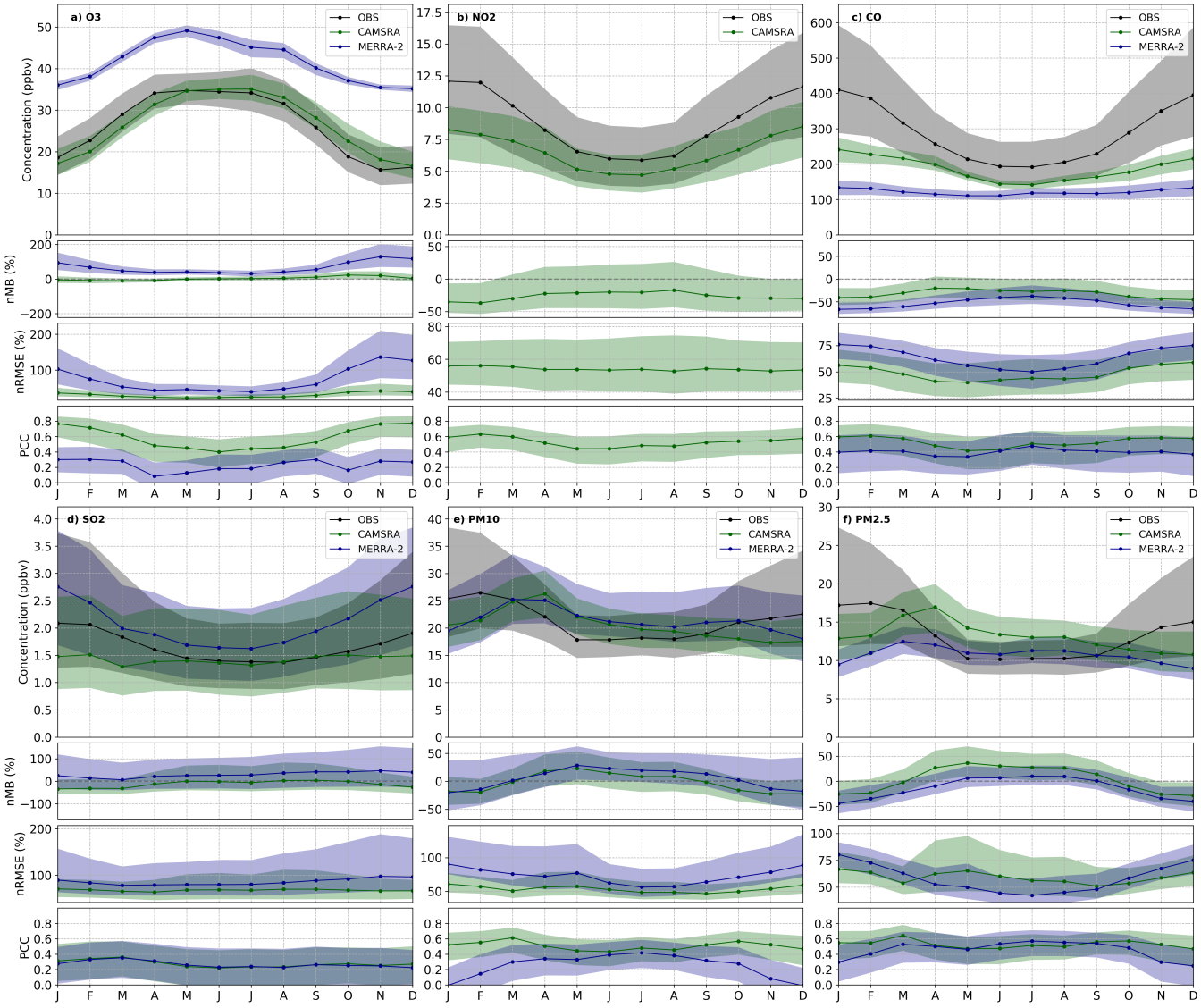

Figure A1. Seasonal variation of [O_3], [NO_2], [CO], [SO_2], [PM_{10}] and [$PM_{2.5}$] over the period 2003-2020 across Europe. For each pollutant the panels show, from top to bottom, concentration, nMB, nRMSE and PCC. The black, green and blue lines represent observations, CAMSRA and MERRA-2, respectively. Shaded contours indicate the 25th (bottom) and 75th (top) percentiles.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

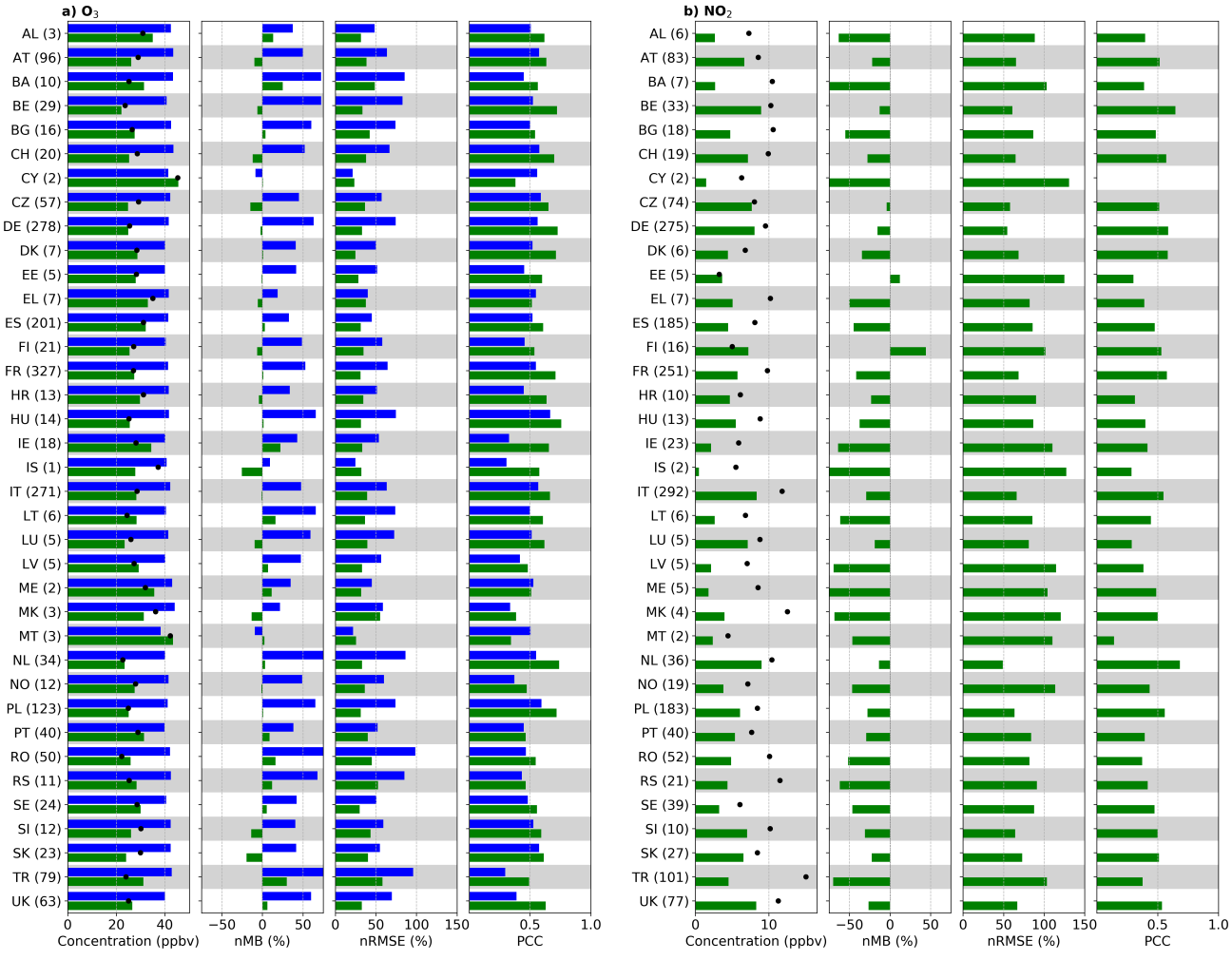

Figure A2. a) Spatially-averaged [O_3], nMB, nRMSE and PCC for all countries reporting data to the EEA; b) Same as a) but for NO_2. Black, green and blue colors indicate observations, CAMSRA and MERRA-2, respectively. Numbers between parentheses indicate the cells with available observations.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

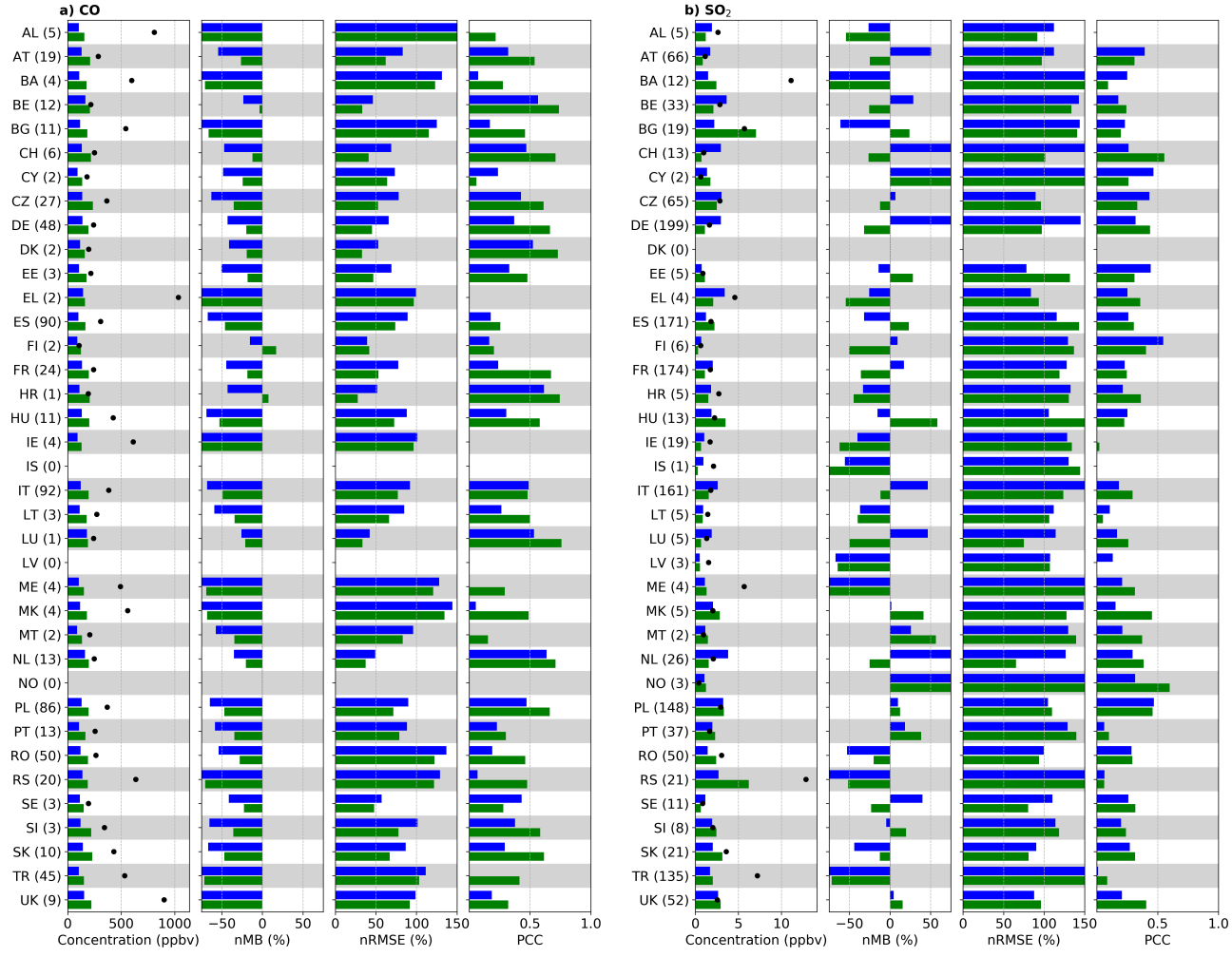

Figure A3. Same as Fig. A2 but for CO, a), and SO_2, b).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

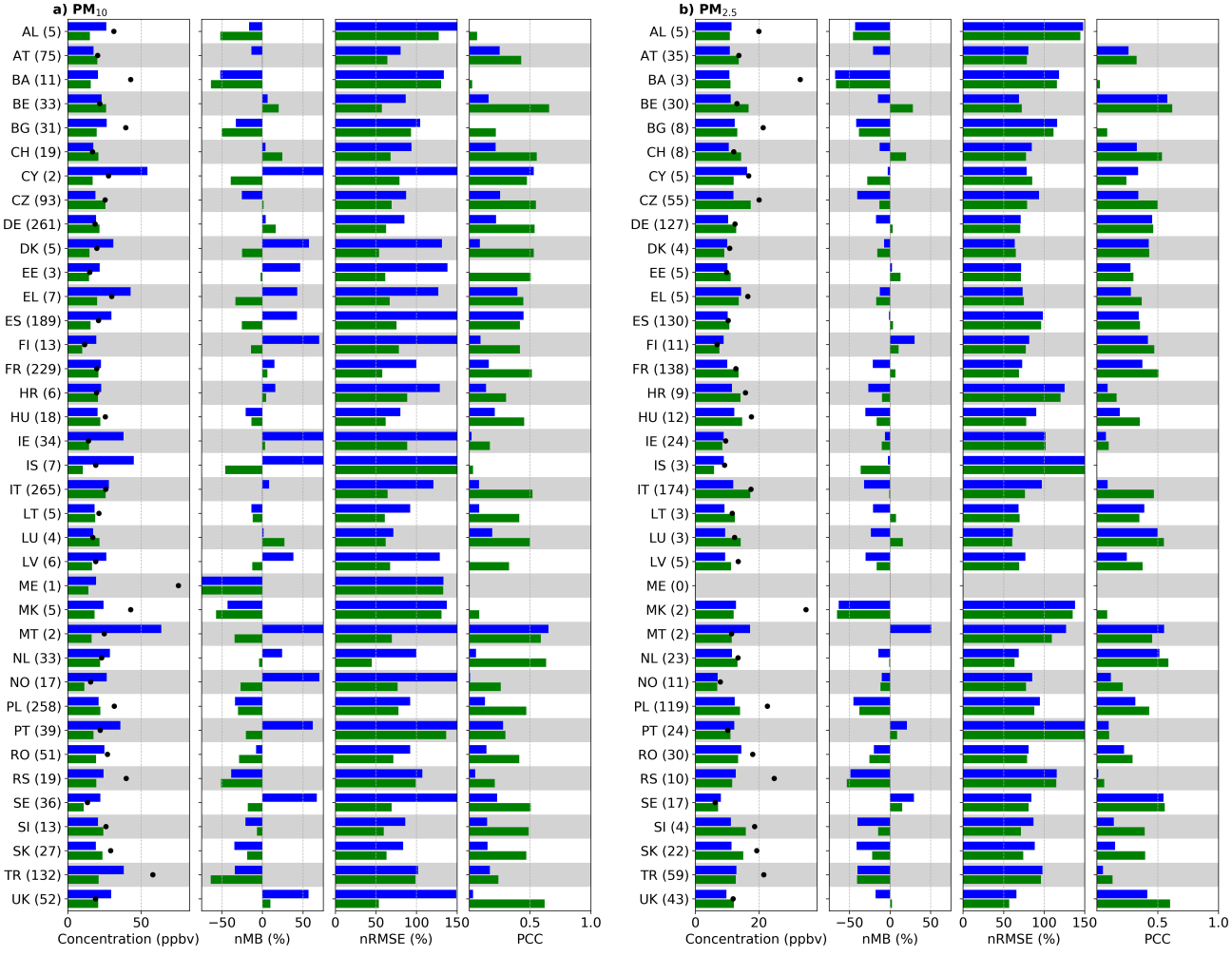

Figure A4. Same as Fig. A2 but for PM_{10}, a), and $PM_{2.5}$, b).

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Table A1. O_3 seasonal statistics over the period 2003-2020 across Europe, for CAMSRA (subscript C) and MERRA-2 (subscript M). Statistics are shown both on a daily scale (d; over all cells and days in the period 2003-2020) and on a monthly scale (m; weight-averaged over all median monthly values). Reactive gases concentrations are expressed in ppbv and normalized statistics in %.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	33.3	31.0	46.4	-7.1	39.3	30.0	49.5	0.45	0.23	2.01 10^6
	JJA	34.3	35.0	45.5	2.0	32.7	29.8	45.0	0.43	0.26	2.03 10^6
	SON	21.6	24.2	37.7	12.2	74.6	41.8	87.9	0.63	0.33	1.98 10^6
	DJF	20.1	18.6	36.4	-7.6	80.7	44.4	95.9	0.61	0.23	1.93 10^6
m	MAM	32.6	30.7	46.5	-6.2	42.3	24.7	47.7	0.52	0.16	54
	JJA	33.4	34.4	45.7	2.9	36.6	24.6	43.4	0.43	0.21	54
	SON	20.1	23.0	37.6	18.1	93.8	36.8	99.9	0.66	0.25	54
	DJF	19.2	17.9	36.4	-3.1	93.3	36.2	101.6	0.75	0.29	54

Table A2. Same as Table A1 but for NO_2.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	8.9	6.5	—	-27.3	—	69.8	—	0.48	—	1.85 10^6
	JJA	6.7	5.2	—	-22.7	—	71.0	—	0.38	—	1.83 10^6
	SON	9.8	6.8	—	-30.5	—	68.0	—	0.47	—	1.85 10^6
	DJF	12.6	8.1	—	-35.6	—	68.8	—	0.52	—	1.84 10^6
m	MAM	8.3	6.3	—	-24.5	—	54.3	—	0.52	—	54
	JJA	6.0	4.9	—	-19.2	—	53.3	—	0.47	—	54
	SON	9.3	6.8	—	-27.8	—	53.5	—	0.54	—	54
	DJF	11.9	8.2	—	-33.9	—	55.2	—	0.60	—	54

Table A3. Same as Table A1 but for CO.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	296.1	197.7	120.4	-33.2	-59.3	72.1	89.1	0.32	0.18	0.32 10^6
	JJA	225.9	147.9	118.4	-34.6	-47.6	80.1	86.2	0.15	0.11	0.31 10^6
	SON	334.0	181.8	126.0	-45.6	-62.3	89.0	101.2	0.34	0.16	0.32 10^6
	DJF	470.3	232.8	136.9	-50.5	-70.9	94.7	109.5	0.32	0.14	0.33 10^6
m	MAM	262.5	193.8	115.6	-23.9	-53.1	42.9	62.0	0.49	0.36	54
	JJA	197.0	146.5	115.5	-25.9	-40.1	43.1	51.7	0.47	0.44	54
	SON	289.3	180.2	121.3	-36.9	-55.7	51.8	66.1	0.56	0.40	54
	DJF	397.0	228.2	132.5	-42.0	-65.8	56.3	75.2	0.59	0.39	54

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

Table A4. Same as Table A1 but for SO_2.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	2.5	2.0	2.0	-22.1	-20.5	228.9	231.3	0.18	0.09	$0.78\ 10^6$
	JJA	1.9	2.0	1.9	6.7	-0.7	216.0	213.3	0.19	0.08	$0.69\ 10^6$
	SON	2.5	2.3	2.4	-7.1	-2.0	246.9	250.8	0.19	0.07	$0.74\ 10^6$
	DJF	4.0	2.2	2.8	-46.0	-29.5	245.0	250.4	0.22	0.06	$0.83\ 10^6$
m	MAM	1.6	1.4	1.9	-14.8	18.0	65.7	79.0	0.30	0.31	54
	JJA	1.4	1.4	1.7	-1.5	30.4	68.6	81.2	0.23	0.23	54
	SON	1.6	1.5	2.2	-3.5	43.6	67.9	92.6	0.26	0.25	54
	DJF	2.0	1.5	2.7	-30.7	26.5	68.5	89.7	0.31	0.28	54

Table A5. Same as Table A1 but for PM_{10}. Aerosol concentrations are expressed in $\mu g\,m^{-3}$.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	24.2	24.9	27.0	2.8	11.5	73.5	110.1	0.44	0.32	$1.89\ 10^6$
	JJA	20.0	20.0	23.9	-0.1	19.2	73.9	87.1	0.30	0.41	$1.88\ 10^6$
	SON	24.3	18.8	23.5	-22.5	-3.5	83.3	107.9	0.37	0.19	$1.87\ 10^6$
	DJF	30.5	21.0	23.6	-31.0	-22.5	95.7	130.4	0.40	0.08	$1.84\ 10^6$
m	MAM	21.7	24.4	24.2	13.3	14.8	54.9	75.1	0.52	0.32	54
	JJA	18.0	19.9	20.7	10.6	20.3	49.8	58.5	0.46	0.40	54
	SON	20.6	18.0	20.7	-13.4	0.8	49.9	71.2	0.54	0.23	54
	DJF	24.8	19.8	19.8	-20.0	-17.5	59.0	87.2	0.51	0.04	54

Table A6. Same as Table A5 but for $PM_{2.5}$.

Scale	Season	OBS	MOD_C	MOD_M	nMB_C	nMB_M	$nRMSE_C$	$nRMSE_M$	PCC_C	PCC_M	N
d	MAM	14.3	16.0	12.1	11.8	-15.8	80.6	83.0	0.46	0.37	$0.77\ 10^6$
	JJA	10.6	13.3	11.3	24.7	6.5	78.8	66.5	0.39	0.39	$0.77\ 10^6$
	SON	14.2	12.1	10.6	-15.1	-25.7	80.6	86.6	0.42	0.31	$0.77\ 10^6$
	DJF	20.3	13.3	10.2	-34.7	-49.9	92.5	108.7	0.48	0.21	$0.75\ 10^6$
m	MAM	13.3	15.7	11.8	20.4	-8.6	60.5	55.1	0.54	0.50	54
	JJA	10.2	13.1	11.1	28.4	8.9	57.3	44.0	0.50	0.55	54
	SON	12.4	11.5	10.2	-7.0	-16.6	54.4	57.9	0.55	0.44	54
	DJF	16.5	12.3	9.8	-25.6	-39.6	64.7	76.0	0.52	0.32	54

Author contributions. AL carried out the analysis. AL and HP contributed to the conception and design of the study. DB was responsible for the acquisition and preprocessing of the air quality data through the GHOST project. AL, HP, ZC, RFMT, HA, CPGP, OJ, AS and JB contributed to the interpretation of results. AL and HP were responsible for writing the article, with a review from CPGP and AS.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Geoscientific Model Development Discussions

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

Acknowledgements. This research has received funding from the European Research Council (ERC), in the frame of the EARLY-ADAPT project (https://early-adapt.eu/), under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 865564), as well as the MITIGATE project (PID2020-116324RA-I00 / AEI / 10.13039/501100011033) from the Agencia Estatal de Investigacion (AEI). We also acknowledge support by the the AXA Research Fund and Red Temática ACTRIS España (CGL2017-90884-REDT), PRACE and RES for awarding us access to MareNostrum Supercomputer in the Barcelona Supercomputing Center, and H2020 ACTRIS IMP (#871115).

Appendix: References

Aldabe, J., Elustondo, D., Santamaría, C., Lasheras, E., Pandolfi, M., Alastuey, A., Querol, X., and Santamaría, J. M.: Chemical characterisation and source apportionment of PM2.5 and PM10 at rural, urban and traffic sites in Navarra (North of Spain), Atmospheric Research, 102, 191–205, https://doi.org/10.1016/j.atmosres.2011.07.003, 2011.

Bauwens, M., Compernolle, S., Stavrakou, T., Müller, J. F., van Gent, J., Eskes, H., Levelt, P. F., van der A, R., Veefkind, J. P., Vlietinck, J., Yu, H., and Zehner, C.: Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using TROPOMI and OMI Observations, Geophysical Research Letters, 47, 1–9, https://doi.org/10.1029/2020GL087978, 2020.

Bosilovich, M., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D., and Suarez, M.: MERRA-2 : Initial Evaluation of the Climate, NASA Technical Report Series on Global Modeling and Data Assimilation, 43, 139, 2015.

Buchard, V., da Silva, A. M., Randles, C. A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., and Winker, D.: Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States, Atmospheric Environment, 125, 100–111, https://doi.org/10.1016/j.atmosenv.2015.11.004, 2016.

Buchard, V., Randles, C. A., da Silva, A. M., Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., and Yu, H.: The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies, Journal of Climate, 30, 6851–6872, https://doi.org/10.1175/JCLI-D-16-0613.1, 2017a.

Buchard, V., Randles, C. A., da Silva, A. M., Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., and Yu, H.: The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies, Journal of Climate, 30, 6851–6872, https://doi.org/10.1175/JCLI-D-16-0613.1, 2017b.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, Journal of the Atmospheric Sciences, 59, 461–483, https://doi.org/10.1175/1520-0469(2002)059<0461:taotft>2.0.co;2, 2002.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Colarco, P., Da Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, Journal of Geophysical Research Atmospheres, 115, https://doi.org/10.1029/2009JD012820, 2010.

Cuesta, J., Eremenko, M., Liu, X., Dufour, G., Cai, Z., Höpfner, M., Von Clarmann, T., Sellitto, P., Foret, G., Gaubert, B., Beekmann, M., Orphal, J., Chance, K., Spurr, R., and Flaud, J. M.: Satellite observation of lowermost tropospheric ozone by multispectral synergism of IASI thermal infrared and GOME-2 ultraviolet measurements over Europe, Atmospheric Chemistry and Physics, 13, 9675–9693, https://doi.org/10.5194/acp-13-9675-2013, 2013.

Darmenov, A. and da Silva, A.: The Quick Fire Emissions Dataset (QFED) - Documentation of versions 2.1, 2.2 and 2.4., NASA Technical Report Series on Global Modeling and Data Assimilation, 32, 2013.

Dee, D. and Uppala, S.: Variational bias correction in ERA-Interim, p. 26, http://rda.ucar.edu/datasets/ds627.1/docs/IFS_documentation/ifs.2b.variational_bias_correction.pdf, 2008.

Diehl, T., Heil, A., Chin, M., Pan, X., Streets, D., Schultz, M., and Kinne, S.: Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and SO_2 from 1980 to 2010 for hindcast model experiments, Atmospheric Chemistry and Physics Discussions, 12, 24 895–24 954, https://doi.org/10.5194/acpd-12-24895-2012, 2012.

Duncan, B. N., Martin, R. V., Staudt, A. C., Yevich, R., and Logan, J. A.: Interannual and seasonal variability of biomass burning emissions constrained by satellite observations, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002jd002378, 2003.

European Environment Agency: European Union emission inventory report 1990-2019, EEA Report No 05/2021, Tech. Rep. 6, https://www.eea.europa.eu/publications/european-union-emission-inventory-report-1, 2021.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A. M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V. H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the integrated forecasting system of ECMWF, Geoscientific Model Development, 8, 975–1003, https://doi.org/10.5194/gmd-8-975-2015, 2015.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., van der Gon, H. D., Frost, G. J., Heil, A., Kaiser, J. W., Kinne, S., Klimont, Z., Kloster, S., Lamarque, J. F., Liousse, C., Masui, T., Meleux, F., Mieville, A., Ohara, T., Raut, J. C., Riahi, K., Schultz, M. G., Smith, S. J., Thompson, A., van Aardenne, J., van der Werf, G. R., and van Vuuren, D. P.: Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period, Climatic Change 2011 109:1, 109, 163–190, https://doi.org/10.1007/S10584-011-0154-1, 2011.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Guenther, A., Nicholas, C., Fall, R., Klinger, L., Mckay, W. A., and Scholes, B.: A global model of natural volatile organic compound emissions s Raja the balance Triangle changes in the atmospheric accumulation rates of greenhouse Triangle Several inventories of natural and Exposure Assessment global scales have been two classes Fores, J. Geophys. Res., 100, 8873–8892, 1995.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hipel, K. W. and McLeod, A. I.: Time series modelling of water resources and environmental systems, https://doi.org/10.1016/0022-1694(95)90010-1, 1994.

Hirsch, R. M. and Slack, J. R.: A Nonparametric Trend Test for Seasonal Data With Serial Dependenc, Water Resources Research, 20, 727–732, 1984.

Hussain, M. and Mahmud, I.: pyMannKendall: a python package for non parametric Mann Kendall family of trend tests., Journal of Open Source Software, 4, 1556, https://doi.org/10.21105/joss.01556, 2019.

Inness, A., Ades, M., Agustí-Panareda, A., Barr, J., Benedictow, A., Blechschmidt, A. M., Jose Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V. H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmospheric Chemistry and Physics, 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J. J., Razinger, M., Schultz, M. G., Suttie, M., and Van Der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, https://doi.org/10.5194/BG-9-527-2012, 2012.

Ma, X., Yan, P., Zhao, T., Jia, X., Jiao, J., Ma, Q., Wu, D., Shu, Z., Sun, X., and Habtemicheal, B. A.: Article evaluations of surface pm10 concentration and chemical compositions in merra-2 aerosol reanalysis over central and eastern china, Remote Sensing, 13, https://doi.org/10.3390/rs13071317, 2021.

Marécal, V., Peuch, V. H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Van Der Gon, H. A., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Van Velthoven, P., Van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: The MACC-II daily ensemble production, Geoscientific Model Development, 8, 2777–2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., Song, I.-S., and Eichmann, A.: NASA / TM – 2012-104606 / Vol 28 Technical Report Series on Global Modeling and Data Assimilation , Volume 28 The GEOS-5 Atmospheric General Circulation Model : Mean Climate and Development from MERRA to Fortuna April 2012, 2012.

Navinya, C. D., Vinoj, V., and Pandey, S. K.: Evaluation of pm2.5 surface concentrations simulated by nasa's merra version 2 aerosol reanalysis over india and its relation to the air quality index, Aerosol and Air Quality Research, 20, 1329–1339, https://doi.org/10.4209/aaqr.2019.12.0615, 2020.

Provençal, S., Buchard, V., da Silva, A. M., Leduc, R., and Barrette, N.: Evaluation of PM surface concentrations simulated by Version 1 of NASA's MERRA Aerosol Reanalysis over Europe, Atmospheric Pollution Research, 8, 374–382, https://doi.org/10.1016/j.apr.2016.10.009, 2017a.

Provençal, S., Buchard, V., da Silva, A. M., Leduc, R., Barrette, N., Elhacham, E., and Wang, S. H.: Evaluation of PM2.5 surface concentrations simulated by version 1 of NASA's MERRA aerosol reanalysis over Israel and Taiwan, Aerosol and Air Quality Research, 17, 253–261, https://doi.org/10.4209/aaqr.2016.04.0145, 2017b.

Randerson, J. T., Liu, H., Flanner, M. G., Chambers, S. D., Jin, Y., Hess, P. G., Pfister, G., Mack, M. C., Treseder, K. K., Welp, L. R., Chapin, F. S., Harden, J. W., Goulden, M. L., Lyons, E., Neff, J. C., Schuur, E. A., and Zender, C. S.: The impact of boreal forest fire on climate warming, Science, 314, 1130–1132, https://doi.org/10.1126/science.1132075, 2006.

Randles, C. A., da Silva, A. M., Buchard, V., Colarco, P. R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., and Flynn, C. J.: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation, Journal of Climate, 30, 6823–6850, https://doi.org/10.1175/JCLI-D-16-0609.1, 2017.

Rienecker, M., Suarez, M., Todling, R., Bacmeister, J., Takacs, L., Liu, H.-C., Gu, W., Sienkiewicz, M., Koster, R., Gelaro, R., and Nielsen, J.: The GEOS-5 Data Assimilation System-Documentation of Versions 5.0. 1, 5.1. 0, and 5.2. 0, NASA Technical Report, 27, 118 pp., 2008.

Ryu, Y. H. and Min, S. K.: Long-term evaluation of atmospheric composition reanalyses from CAMS, TCR-2, and MERRA-2 over South Korea: Insights into applications, implications, and limitations, Atmospheric Environment, 246, 118 062, https://doi.org/10.1016/j.atmosenv.2020.118062, 2021.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmospheric Chemistry and Physics, 14, 9317–9341, https://doi.org/10.5194/ACP-14-9317-2014, 2014.

Souri, A. H., Chance, K., Sun, K., Liu, X., and Johnson, M. S.: Dealing with spatial heterogeneity in pointwise-to-gridded-data comparisons, Atmospheric Measurement Techniques, 15, 41–59, https://doi.org/10.5194/amt-15-41-2022, 2022.

Spracklen, D. V., Jimenez, J. L., Carslaw, K. S., Worsnop, D. R., Evans, M. J., Mann, G. W., Zhang, Q., Canagaratna, M. R., Allan, J., Coe, H., McFiggans, G., Rap, A., and Forster, P.: Aerosol mass spectrometer constraint on the global secondary organic aerosol budget, Atmospheric Chemistry and Physics, 11, 12 109–12 136, https://doi.org/10.5194/acp-11-12109-2011, 2011.

https://doi.org/10.5194/gmd-2022-197
Preprint. Discussion started: 6 September 2022

Thorsteinsson, T., Jóhannsson, T., Stohl, A., and Kristiansen, N. I.: High levels of particulate matter in Iceland due to direct ash emissions by the Eyjafjallajkull eruption and resuspension of deposited ash, Journal of Geophysical Research: Solid Earth, 117, 1–9, https://doi.org/10.1029/2011JB008756, 2012.

Valmartin, M., Heald, C. L., and Arnold, S. R.: Coupling dry deposition to vegetation phenology in the Community Earth System Model: Implications for the simulation of surface O3, Geophysical Research Letters, 41, 2988–2996, https://doi.org/10.1002/2014GL059651, 2014.

Vîrghileanu, M., Săvulescu, I., Mihai, B. A., Nistor, C., and Dobre, R.: Nitrogen dioxide (No2) pollution monitoring with sentinel-5p satellite imagery over europe during the coronavirus pandemic outbreak, Remote Sensing, 12, 1–29, https://doi.org/10.3390/rs12213575, 2020.

Wagner, A., Bennouna, Y., Blechschmidt, A. M., Brasseur, G., Chabrillat, S., Christophe, Y., Errera, Q., Eskes, H., Flemming, J., Hansen, K. M., Inness, A., Kapsomenakis, J., Langerock, B., Richter, A., Sudarchikova, N., Thouret, V., and Zerefos, C.: Comprehensive evaluation of the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis against independent observations: Reactive gases, Elementa, 9, https://doi.org/10.1525/elementa.2020.00171, 2021.