CoCO2

Prototype system for a Copernicus CO_2 service

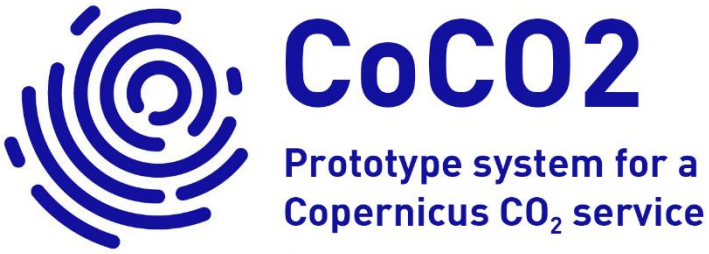
D2.7 PED uncertainty 2018 v2

Ingrid Super (TNO)

Margarita Choulga (ECMWF)

Tilman Hohenberger (TNO)

Co-ordinated by ECMWF



D2.7 PED uncertainty 2018 and uncertainties based on Monte Carlo simulation using the emission model from D2.5 v2

Dissemination Level:	Public
Author(s):	Ingrid Super (TNO), Margarita Choulga (ECMWF), Tilman Hohenberger (TNO)
Date:	27/07/2023
Version:	1.0
Contractual Delivery Date:	30/06/2023
Work Package/ Task:	WP2/ T2.5
Document Owner:	TNO
Contributors:	TNO, ECMWF
Status:	Issued

CoCO2: Prototype system for a Copernicus CO_2 service

Coordination and Support Action (CSA)
H2020-IBA-SPACE-CHE2-2019 Copernicus evolution – Research activities in support of a European operational monitoring support capacity for fossil CO2 emissions

Project Coordinator: Dr Richard Engelen (ECMWF)
Project Start Date: 01/01/2021
Project Duration: 36 months

Published by the CoCO2 Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX,
richard.engelen@ecmwf.int

The CoCO2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958927.

Table of Contents

Figures

Tables

1. Executive Summary and introduction

This deliverable report describes the development of the prior uncertainty dataset related to the European and global prior emission datasets (PED) for the year 2018. We provide an overview of the developed methodology to estimate prior uncertainties in the aggregated PED starting from a detailed set of uncertainties. This also includes the estimation of gridded uncertainties, uncertainties in temporal profiles and spatial and temporal error correlation lengths. For the global emissions only CO_2 is considered, whereas for the European emissions next to CO_2, CO and NO_x are also included. Since the data underlying the global and European datasets is different, a slightly different approach is used for the datasets, but where possible we kept our methods consistent.

This dataset is an update of CoCO2 D2.6 from June 2022. In addition to the previous dataset the uncertainties in temporal profiles and their correlation lengths are now added. The methodology to estimate these error statistics are described in this report and we show some results. Furthermore, we will shortly describe any alterations to what was described in D2.6, so that a clear documentation of the current dataset is provided. Finally, a separate note will be provided describing the uncertainties related to the emission model from D2.5.

1.1 Data access

The data are available through an FTP site (coco2@ftp.ecmwf.int), following the directory structure data-exchange/WP2/D2-7-prior_emission_uncertainties. Please contact the CoCO2-Coordinator to obtain the FTP password access.

1.2 Background

Prior emission data is important input data for the modelling efforts done in CoCO2. For inverse modelling, and also to understand discrepancies between models and observations, the uncertainties in the prior emissions also need to be quantified. The challenge is to get a consistent set of uncertainties for the emission datasets, making use of uncertainty estimates at a very detailed level and considering different error distributions and correlations.

First efforts have been made in the CHE project to quantify these uncertainties, which form the basis for the work in T2.5 (Choulga et al. (2021), Super et al. (2020)). Whereas in CHE work was done mostly on country-level uncertainties for CO_2, we want to extend this work by including other sources of uncertainties and co-emitted species. Main focus points are:

- Create a consistent set of country-level uncertainties, both globally and for Europe.
- Include CO_2 and co-emitted species CO and NO_x.
- Assess error correlations between species.
- Assess uncertainties in spatial/temporal proxies, including spatial/temporal correlation lengths.

1.3 Scope of this deliverable

1.3.1 Objectives of this deliverable

The aim is to provide a set of uncertainties for the PED products from WP2, including spatiotemporal uncertainties and error correlations. Other WPs can use these data in inverse modelling efforts and provide feedback on the current product. Although this is the final deliverable of CoCO2 WP2 task 2.5 (T2.5), work will continue in e.g. CORSO (https://corso-project.eu/), and feedback will be taken into account for future updates.

1.3.2 Work performed in this deliverable

There are two deliverables for T2.5 (D2.6 and D2.7), of which this report (and accompanying dataset and user documentation) is the final one. We made a list of priorities and started from there. This report describes the work that has been done, acknowledging potential challenges for future work.

1.3.3 Deviations and counter measures

Originally, the focus year (base year) in the proposal was 2016, but throughout the project this has been updated to 2018 as for 2018 much more satellite data are available, especially since TROPOMI was launched at the end of 2017. As a result, the WP2 PED and uncertainties are provided for 2018. The PED data for 2021 is based on extrapolation and therefore uncertainties are expected to be higher. Therefore, we stick to 2018, but we believe that uncertainties will not change a lot between different years.

This deliverable also requires a contribution from the developers of the FFDAS modelling system, which is part of T2.4. However, due to time limitations their contribution will follow in October 2023 as an additional note to this deliverable.

Originally, JRC had the lead for this deliverable. However, for practical reasons this was changed to TNO. The work has been done by ECMWF and TNO, as planned.

2 Temporal profiles

In this deliverable, we add uncertainties in temporal profiles, including temporal correlation lengths. This chapter describes input data, methodology and results.

2.1 Input data: CAMS-TEMPO

As a starting point for temporal profile uncertainties we use the CAMS-TEMPO v3.2 (Guevara et al., 2020a, 2020b, 2021) for the year 2018. The temporal profiles (as weight factors) are specified per sector, with the regional profiles following the GNFR (Gridded Nomenclature for Reporting) categorisation and the global profiles following the sectoral categorisation of the EDGAR emission inventory.

How the profiles are defined differs per sector. For some sectors fixed weekly and monthly profiles are provided, whereas for some sectors daily factors are provided for a whole year. Some sectors have country-specific profiles, some have pollutant-specific profiles, others have both.

For the global domain the following profiles are available:

- Energy (ENE): monthly and weekly profiles, pollutant- and country-specific
- Residential and commercial (RES): daily profiles, country-specific
- Road transport (TRO): monthly and weekly profiles, country-specific
- Industry (IND): monthly profiles, country-specific
- Agricultural waste burning (AWB): monthly profiles, country-specific
- Off-road transport (TNR): weekly profiles, pollutant-specific
- Aviation (AVI): monthly and weekly profiles, country-specific
- Agricultural livestock (AGL): monthly profiles (only for NOx), country-specific

For the European domain the following profiles are available:

- Public power (GNFR A): monthly and weekly profiles, pollutant- and country-specific
- Industry (GNFR B): monthly profiles, country-specific
- Other stationary combustion (GNFR C): daily profiles, pollutant- and country-specific
- Road transport (GNFR F): monthly and weekly profiles per sub-sector, pollutant- and country-specific
- Shipping (GNFR G): monthly profiles, pollutant- and country-specific
- Aviation (GNFR H): monthly and weekly profiles, country-specific
- Off-road transport (GNFR I): monthly and weekly profiles, pollutant-specific
- Agriculture (GNFR K (livestock)/L (other)): daily profiles (only for NOx), country-specific

Whereas the global CAMS-TEMPO contains specific profiles for CO_2, this is not the case for the European CAMS-TEMPO. Here, we only have CO_2-specific profiles for shipping. Therefore, some pre-processing is needed to get temporal profiles for CO_2. We make the following assumptions:

- For other stationary combustion and road transport (LPG exhaust) the profiles for CO and NO_x look similar and we take the average to represent CO_2
- For road transport (gasoline exhaust) the profile for CO is temperature-dependent and the profile for NO_x is not; hence we use the profile of NO_x to represent CO_2
- For road transport (diesel exhaust) the profile for NO_x is temperature-dependent and the profile for CO is not; hence we use the profile of CO to represent CO_2
- For public power an alternative dataset is used, which includes country- and pollutant-specific monthly and weekly profiles using as a basis the CoCO2 global point source

database. This dataset only contains profiles for CO_2 and NO_x, and for the CO profile we take the average of the CO_2 and NO_x profiles

For the sectors not listed here information is too scarce to build country- or pollutant specific profiles and in that case default profiles from TNO are used (Denier van der Gon et al., 2011). Also, in case only monthly profiles are given, the weekly disaggregation follows the default profiles.

2.2 Methodology

To calculate the uncertainties in the profiles we make the assumption that modellers often use one average profile for the whole domain, which was discussed with global modellers from ECMWF. Therefore, the spread in the profiles between countries can be a reasonable indication of the uncertainty in that average profile. Of course, there are more sources of uncertainty, for example in the proxy data to make the profiles and how representative these are for the temporal variations in the activity causing emissions. Moreover, to make weekly profiles all weeks in a year are averaged, which also contains a certain spread. Therefore, with this approach we are likely to underestimate the actual errors. In the framework of the CORSO project more work will be done to include other sources of error by the developers of the CAMS-TEMPO data, which have more detailed knowledge on the underlying data and assumptions.

2.2.1 Uncertainties in temporal profiles

2.2.1.1 Global approach

Firstly, the CAMS-TEMPO data is processed from monthly and weekly temporal profiles to daily profiles for CO_2 per sector (in total 7 sectors, as RES is already provided in daily format). Secondly, daily profiles per country are aggregated into one global profile. The number of countries with temporal profile information depends on the sector – in general it is 218, with some exceptions. RES profiles are available for 203 countries. TRN and AGL have temporal information for two additional countries (i.e. 220 in total), namely Guam (GUM) and Montserrat (MRS). To examine if the temporal profiles and their uncertainty ranges differ strongly for regions/continents, as well as to better compare results over Europe to see if minor differences in the methodologies applied (i.e. over the global and European data) lead to noticeable differences in the results – we have additionally grouped all countries into 7 main regions (see Table 1 for the full list of geographical regions included).

Table 1: List of all geographical regions included in this study and their grouping.

Continent (parts)	Name of countries/ parts (Alpha-3 code/ ISO 3166)
Africa (52)	Algeria (DZA), Angola (AGO), Botswana (BWA), Burkina Faso (BFA), Burundi (BDI), Cabo Verde (CPV), Cameroon (CMR), Central African Rep (CAF), Chad (TCD), Comoros (COM), Cote d'Ivoire (CIV), Djibouti (DJI), Egypt (EGY), Equatorial Guinea (GNQ), Eritrea (ERI), Ethiopia (ETH), Gabon (GAB), Ghana (GHA), Guinea (GIN), Guinea-Bissau (GNB), Kenya (KEN), Lesotho (LSO), Liberia (LBR), Libya (LBY), Madagascar (MDG), Malawi (MWI), Mali (MLI), Mauritania (MRT), Mauritius (MUS), Morocco (MAR), Mozambique (MOZ), Namibia (NAM), Rwanda (RWA), Sao Tome & Principe (STP), Senegal (SEN), Seychelles (SYC), Sierra Leone (SLE), Somalia (SOM), South Africa (ZAF), Sudan (SDN), Swaziland (SWZ), Tanzania (TZA), Togo (TGO), Tunisia (TUN), Uganda (UGA), Western Sahara (ESH, disputed), Zambia (ZMB), Zimbabwe (ZWE), Benin (BEN), Niger (NER), Nigeria (NGA)

Continent (parts)	Name of countries/ parts (Alpha-3 code/ ISO 3166)
Asia (46)	Afghanistan (AFG), Armenia (ARM), Azerbaijan (AZE), Bhutan (BTN), Brunei (BRN), Cambodia (KHM), China (CHN), Georgia (GEO), Hong Kong (China) (HKG), Indonesia (IDN), Iran (IRN), Iraq (IRQ), Japan (JPN), Jordan (JOR), Kazakhstan (KAZ), Kyrgyzstan (KGZ), Laos (LAO), Lebanon (LBN), Macau (China) (MAC), Malaysia (MYS), Maldives (MDV), Mongolia (MNG), Nepal (NPL), Oman (OMN), Pakistan (PAK), Philippines (PHL), Saudi Arabia (SAU), Singapore (SGP), Sri Lanka (LKA), Syria (SYR), Taiwan (TWN), Tajikistan (TJK), Thailand (THA), Timor-Leste (TLS), Turkmenistan (TKM), Uzbekistan (UZB), Vietnam (VNM), Yemen (YEM), Bahrain (BHR), Kuwait (KWT), Qatar (QAT), United Arab Emirates (ARE), Israel (ISR), Bangladesh (BGD), Burma (MMR), India (IND)
Europe (44)	Albania (ALB), Andorra (AND), Austria (AUT), Belarus (BLR), Belgium (BEL), Bosnia & Herzegovina (BIH), Bulgaria (BGR), Croatia (HRV), Cyprus (CYP), Czechia (CZE), Denmark (DNK), Estonia (EST), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Hungary (HUN), Iceland (ISL), Ireland (IRL), Italy (ITA), Latvia (LVA), Liechtenstein (LIE),Lithuania (LTU), Luxembourg (LUX), Macedonia (MKD), Malta (MLT), Moldova (MDA), Monaco (MCO), Netherlands (NLD), Norway (NOR), Poland (POL), Portugal (PRT), Romania (ROU), Russia (RUS), San Marino (SMR), Serbia (SRB), Slovakia (SVK), Slovenia (SVN), Spain (ESP), Sweden (SWE), Switzerland (CHE), United Kingdom (GBR), Ukraine (UKR), Turkey (TUR)
North America (24)	Antigua & Barbuda (ATG), Bahamas (BHS), Barbados (BRB), Belize (BLZ), Canada (CAN), Costa Rica (CRI), Cuba (CUB), Dominica (DMA), Dominican Republic (DOM), El Salvador (SLV), Grenada (GRD), Guatemala (GTM), Haiti (HTI), Honduras (HND), Jamaica (JAM), Mexico (MEX), Netherlands [Caribbean] (ANT), Nicaragua (NIC), Panama (PAN), St Kitts & Nevis (KNA), St Lucia (LCA), St Vincent & the Grenadines (VCT), Trinidad & Tobago (TTO), United States (USA)
North America extra (14)	Anguilla (UK) (AIA), Aruba (Netherlands) (ABW), Bermuda (UK) (BMU), Br Virgin Is (UK) (VGB), Br Virgin Islands (UK) (VGB), Cayman Is (UK) (CYM), Greenland (Denmark) (GRL), Guadeloupe (France) (GLP), Martinique (France) (MTQ), Montserrat (UK) (MSR), Puerto Rico (US) (PRI), St Pierre & Miquelon (France) (SPM), Turks & Caicos Is (UK) (TCA), US Virgin Is (US) (VIR)
Oceania (13)	Australia (AUS), Fiji (FJI), Kiribati (KIR), Marshall Is (MHL), Nauru (NRU), New Zealand (NZL), Palau (PLW), Papua New Guinea (PNG), Samoa (WSM), Solomon Is (SLB), Tonga (TON), Tuvalu (TUV), Vanuatu (VUT)
South America (12)	Argentina (ARG), Bolivia (BOL), Brazil (BRA), Chile (CHL), Colombia (COL), Ecuador (ECU), Guyana (GUY), Paraguay (PRY), Peru (PER), Suriname (SUR), Uruguay (URY), Venezuela (VEN)

2.2.1.2 European approach

Similar to the global approach, for Europe the CAMS-TEMPO data are also first processed to make daily profiles for all sectors, such that the format is the same for all sectors. For the road transport sector we make a weighted average of the 3 sub-sector profiles by fuel type based on the emission shares in each sub-sector. We do this per pollutant and per country. For the uncertainty calculation, we only include the 42 countries for which we also have a spatial

uncertainty estimate. Note that the temporal profiles for shipping are assigned to sea regions and not to countries.

2.2.2 Temporal correlation lengths

To calculate the temporal correlation lengths for each sector for the European domain we use two different methods. One follows the same methodology as used to calculate the spatial correlation lengths, namely fitting a semi-variogram to the median profiles. The other uses the autocorrelation function (Schepanski et al., 2015). We used different maximum length values. Also globally, we investigated the temporal correlation length by computing the Pearson's correlation *r* between the original and lagged temporal profiles (lag of N days, N between 0 and 180 days).

2.3 Results

2.3.1 Uncertainties in temporal profiles

For each sector, the daily temporal profiles were examined to determine the shape of their distributions for the globe and the different regions. We calculated mean, standard deviation and 23 different percentile values (i.e. 0.0, 2.5, 5.0, 10.0, 15.0, …, 85.0, 90.0, 95.0, 97.5, 100.0) of (i) the actual distribution, (ii) a constructed triangular distribution (computed 2.5, 50.0, 97.5 percentiles of the real distribution were used as triangular lower limit, mode, and upper limit respectively), and (iii) a constructed normal distribution (based on computed mean and standard deviation of the real distribution). Next, we calculate Pearson's correlation *r* (and its statistical significance *p* value) between the real and constructed distribution percentiles. Finally, we computed the number of days per sector when the correlation was at least 0.8 (and *p* not greater 0.05), which is assumed to be a good threshold to accept or reject the assumed distribution. Figure 1 shows the share of days per sector and per region when daily temporal profiles based on all countries have a normal (a) or triangular (b) distribution shape. In general, across all sectors and regions a triangular distribution has a higher correlation with the real data than a normal distribution. The AWB and AVI sectors' temporal profile distributions are not well represented with either normal or triangular distributions, but emissions from these sectors are very modest compared to other sectors. For Europe, all sectors follow a triangular distribution almost every day of the year. Based on these results it was decided to assume a triangular distribution for the temporal profiles, and to provide to the user the 2.5, 50.0, 97.5 percentiles of the real data to construct the triangular distribution for use in ensemble simulations.

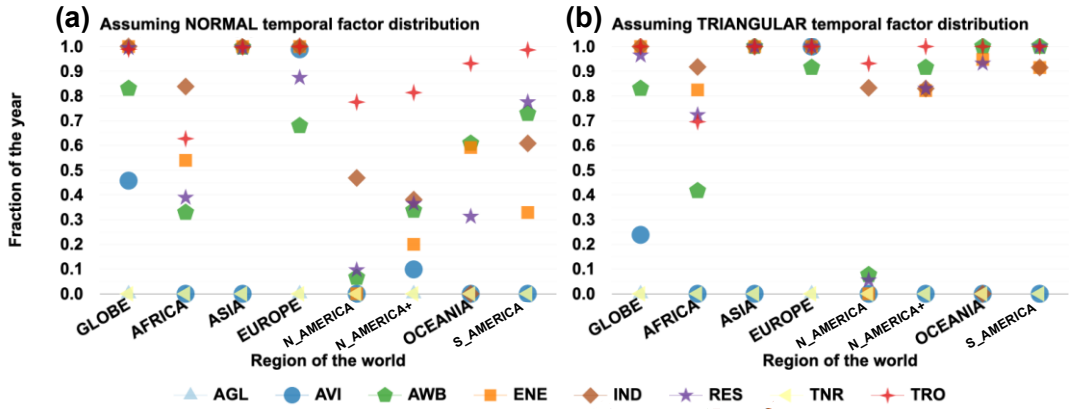


Figure 1: Fractions of the year, for eight sectors and seven regions, when daily temporal factor distributions can be assumed normal (a) or triangular (b).

The TNR sector has the same weekly profile for all countries due to a lack of data. Similarly, the AGL sector has the same monthly profile for all countries. Consequently, we are unable to calculate their distributions and for these two sectors the 2.5 and 97.5 percentile values are specified as 0.2 and 1.8 of the median. For all sectors, it should be also noted that for daily

temporal factor distributions based on all countries, “average” actually is the median (i.e. 50.0 percentile value).

Figure 2 shows the global average real and reconstructed (based on triangular distribution) temporal profiles for eight sectors. When real daily distribution of temporal factors is close to symmetric triangular distribution, median values coincide. The range of the reconstructed distributions are always narrower compared to the real data due to the triangular distribution assumptions.

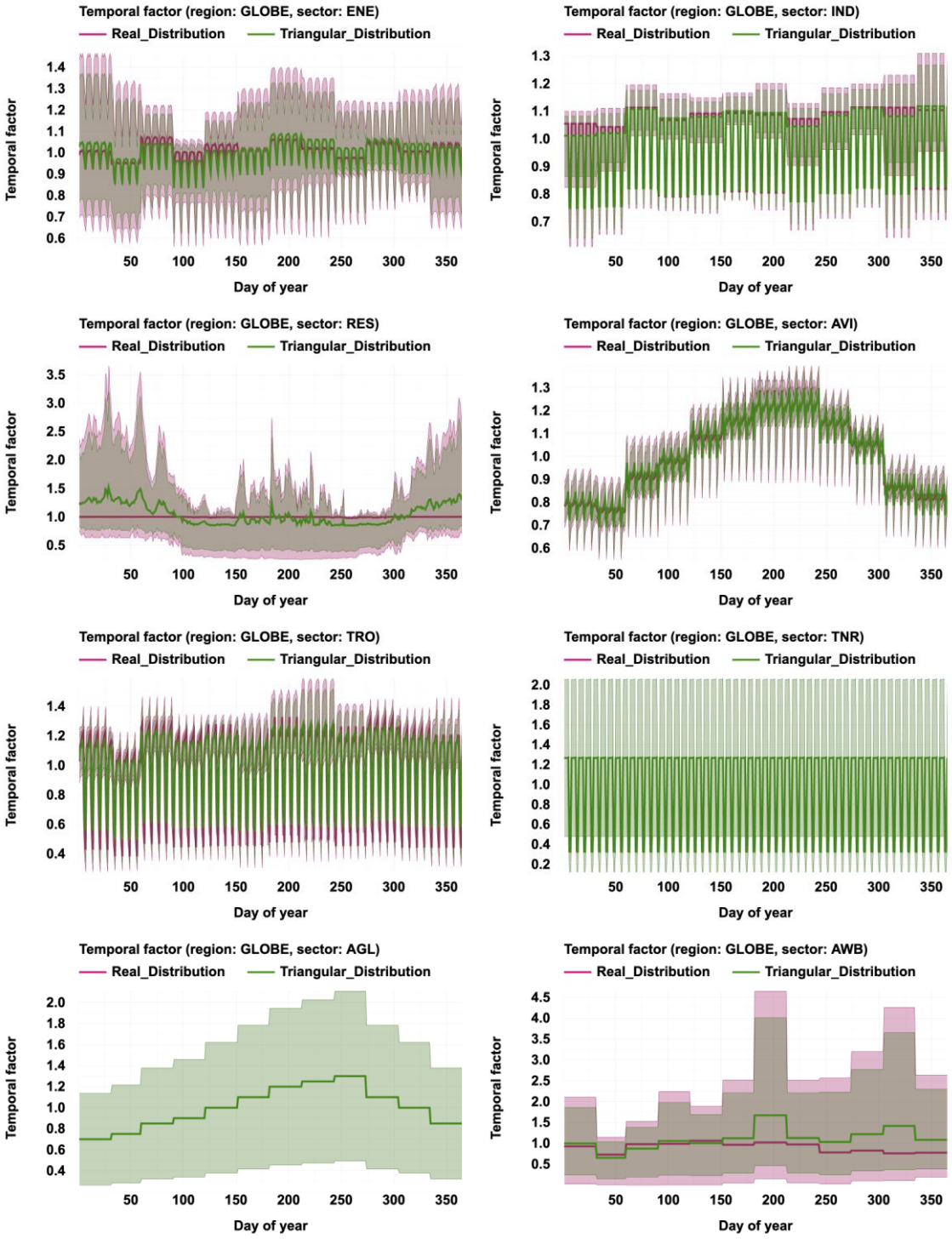


Figure 2: Global average real (pink) and reconstructed (green; assumed triangular distribution) temporal profiles for eight sectors. Lines show median of the distribution, shading shows range (2.5 percentile to 97.5 percentile).

The globally aggregated temporal profile distribution for RES is practically constant and equals ~1 due to caveats of the underlying data (Figure 3). First of all, 65 out of 203 countries have a constant yearly profile. Secondly, for 186 days the median of all profiles equals 1. Even though 179 days of the year have more than half of the countries with daily factors less than 1, their median's minimum value is 0.979 – which is still close to 1. Therefore, we investigated if simple separation over continents would bring some more information. Based on Figure 3 we conclude that only Asia and Europe are well represented as individual regions for RES emissions. The median values of their temporal profile distributions have some seasonal cycle as expected for a sector where emissions are largely driven by the weather which has a yearly cycle (winter vs. summer). In contrast, the yearly profile for North America is aggregated over 24 countries, of which one has no data and 13 have flat profiles. Therefore, the median of all 365 days of the year equals 1. It is evident that for more informative analysis more country specific yearly profiles are needed. With gridded country-specific data analysis can be more detailed, e.g. based on global maps of the Köppen-Geiger climate classification presented in Beck at al. (2018). Similar analyses were also made for other sectors, but the differences between regions are most clearly illustrated by the RES sector.

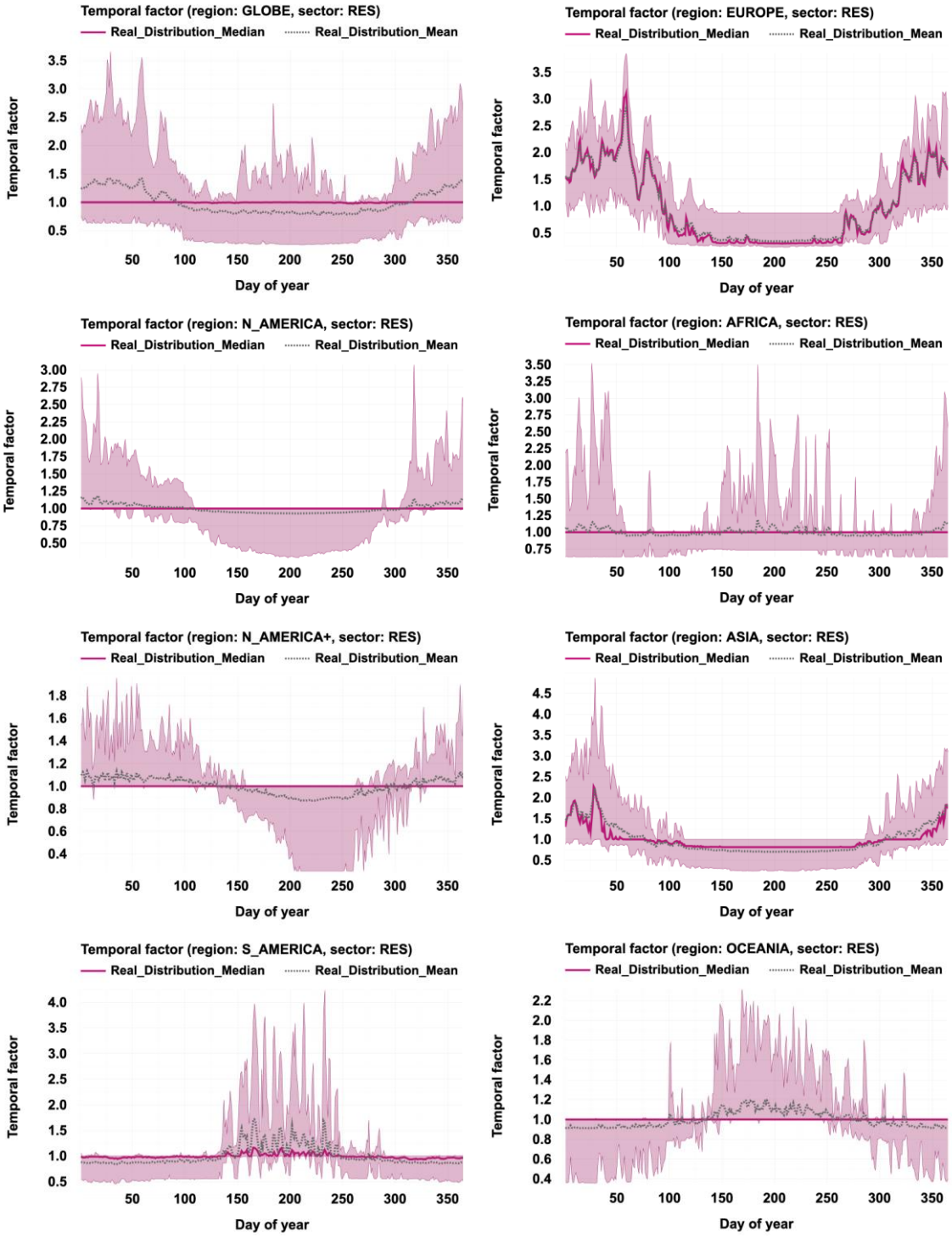


Figure 3: Real aggregated temporal profile for RES over different regions. Solid line shows the median of the distribution, the dashed line the mean, and shading shows range (2.5 percentile to 97.5 percentile).

Figure 4 shows the median temporal profile per GNFR sector for the European CAMS-TEMPO data, including the 95% confidence interval. The amount of variability differs per country and also the time scales over which large variations occur is different. For example, the temporal variations in the road transport sector are dominated by the weekly cycle, with especially much lower activities during weekends. The seasonal variations are small. In contrast, the aviation sector shows a strong seasonal cycle, with higher activities during the summer holidays. The largest variations are visible for the industry, other stationary combustion and agriculture.

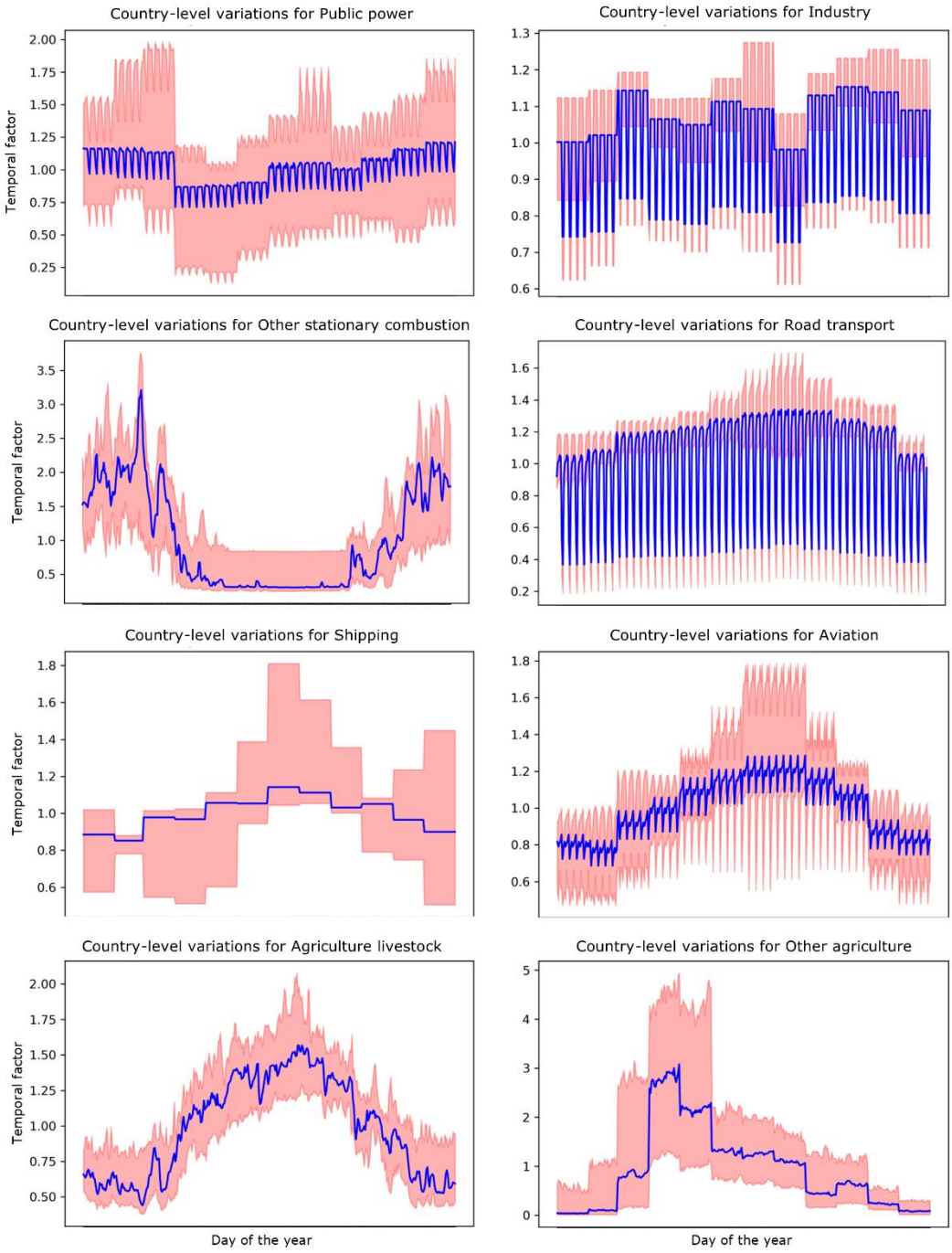


Figure 4: Country-averaged temporal profiles for the year 2018 per GNFR sector. Pink areas shows 95% confidence interval, blue line is the median profile.

We can clearly see that often the spread in the country-specific time profiles is not Gaussian. When we look at the distribution for specific times and sectors there is no distinct distribution visible. Mostly there is a peak around the median value and then it drops towards the limits with some ups and downs. Therefore, we decided to calculate a daily 95% confidence interval and describe the lower and upper limits of these intervals in the final dataset, including the median profile. We suggest to use these as the limits of a triangular distribution with the median as expected value, in line with the global approach. Of course, it is also possible to simply use the country-specific profiles directly as an ensemble.

For the off-road sector we have no country-specific profiles. Also for sectors for which default profiles are used we cannot estimate an uncertainty with this approach. Therefore, we make the lower and upper limits equal to a 80% deviation from the median, similar to the global approach.

For the industry sector the variations between countries is small and therefore the uncertainty is on average only 10% (difference between lower/upper limit and median profile). The largest uncertainties are found for other stationary combustion and other agriculture, which both have a strongly skewed distribution. For other stationary combustion the lower limit varies between 10-74 % from the median and the upper limit lies 13-254 % from the median. This skewness is most pronounced during summer. For other agriculture the lower limit varies between 34-87 % from the median and the upper limit lies 35-1740 % from the median. Here, the skewness is most pronounced during the first months of the year due to differences in the start of the growing season between countries, which is the only factor determining the shape of the temporal profile (i.e., the remainder of the year emissions are very small). This range seems unrealistically large, but reflects the variations in the current data. The average deviations from the median for the other sectors are about 45 % for public power, 20 % for road transport, 25 % for shipping, 22 % for aviation and 25 % for agriculture livestock.

2.3.2 Temporal correlation lengths

Table 2 shows the estimated temporal correlation lengths for CO_2 (only for agriculture results are based on NO_x profiles to make sure the results are connected to combustion engines) for the European CAMS-TEMPO data. Generally, the results are very similar for CO and NO_x, except for off-road transport. However, we find strange artefacts due to the nature of the temporal profiles, which mostly follow a fixed weekly pattern. This means that each Monday is strongly correlated to any other Monday, whereas in reality this correlation does not exist throughout the year. Therefore, it is worthwhile to invest more in better quantifying the temporal patterns with actual data, such that we have daily weight factors and avoid strange autocorrelation patterns. Nevertheless, we discuss some initial results here, as they give an indication of relevant time scales. The results are not yet included in the dataset due to their ambiguity.

The correlation lengths for industry, solvents and road transport are relatively consistent across the different methods and maximum length scale settings. The correlation lengths suggest that the dominant time scale for these sectors is around a week, which is in line with the temporal profiles shown in Figure 4. For sectors with a dominant seasonal cycle, i.e. fugitives, shipping, aviation and agriculture, the semi-variogram has no convergence and gives the maximum length scale as the correlation length. The autocorrelation method gives an estimate of close to 50 days for the maximum length of 50 days, but somewhere around 90 days for the maximum length of 180 days. This suggests that correlations occur within seasons, but not between seasons.

The public power sector shows significant variations between months, but also within a week. Therefore, the estimated correlation length scale is shorter than for sectors with a dominating seasonal cycle. However, the estimates are quite different depending on the methodology and maximum length scale. This probably has to do with multiple modalities in the data. Generally, the activity of power plants shows large deviations over time, which are related to many factors (energy demand, availability of renewable energy, fuel prices, etc.).

Table 2: Calculated temporal correlation lengths (in days) per sector for CO_2 (or for NO_x for Agriculture other).

Sector	Max. length 50 days		Max. length 180 days	
	Auto-correlation	Semi-variogram	Auto-correlation	Semi-variogram
Public power	33.3	49.6	60.1	128.3
Industry	7.3	4.1	9.3	11.6
Other stationary combustion	42.9	50	87.8	180
Fugitives	42.4	50	84.6	180
Solvents	3.6	4	4.6	9
Road transport	6	4.1	9.8	10.5
Shipping	45.2	50	96.2	180
Aviation	40.4	50	85.7	180
Off-road transport	40.6	50	81.6	180
Agriculture other	45.3	50	95.7	180

Finally, other stationary combustion is a complex case. Like the public power sector, it has multiple dominant length scales. There is a strong seasonal cycle, but there are also variations on shorter time scales related to mesoscale weather changes. This is clearly visible in the semi-variograms in Figure 5. In the left plot, the maximum length scale is set to 180 days and then there is no convergence. However, we do see a sort of plateau around 20 days. If we take that as the maximum length scale, we do get convergence and the estimated length scale is 19 days. The autocorrelation function gives a similar estimate of 18.7 days. Moreover, during summer the emission profile is more or less flat, whereas during the winter we see strong short-term variations. When we only select winter months (December – February) we find a correlation length scales of 9 days with a cut-off at 20 days.

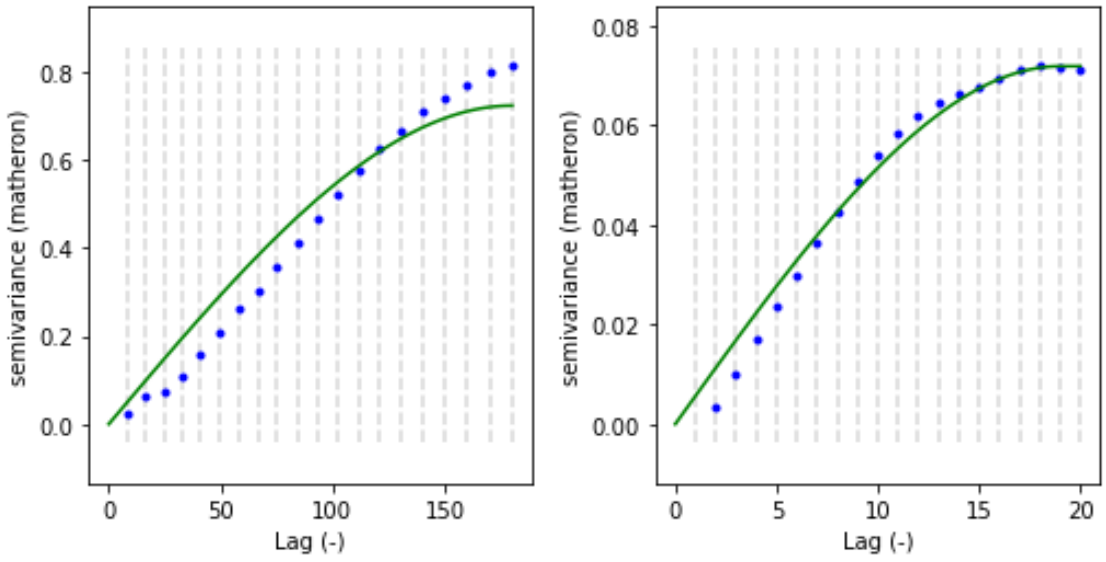


Figure 5: Semi-variograms for the other stationary combustion sector with maximum length scales of 180 days (left) and 20 days (right).

Also in the global CAMS-TEMPO data we notice that sectors could be divided into two groups – (i) where correlation values follow a weekly cycle, meaning weekly fluctuations are much stronger than yearly fluctuations, and (ii) where correlation values strongly follow yearly cycle with small weekly deviations (see Figure 6 for examples). Strong domination of the weekly cycle is observed in ENE, IND, TRO, and TRN sectors; strong domination of the yearly cycle is seen in AVI, AWB, AGL, and RES sectors. The peak in the ENE sector correlation around 60 days also reflects patterns that are visible in the temporal profile (Figure 2). Yet more investigation should be done in the future using actual daily sector temporal profiles.

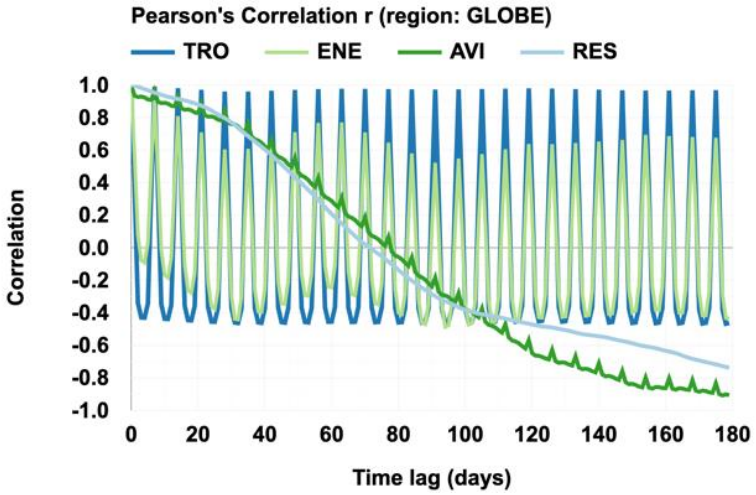


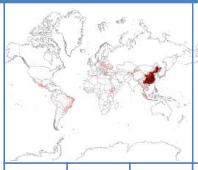
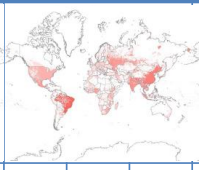
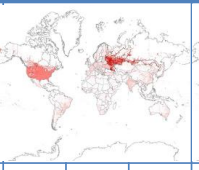
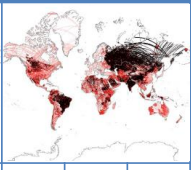
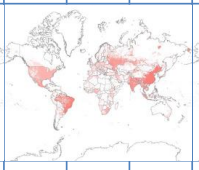
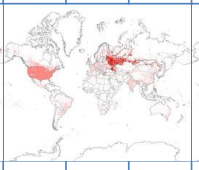
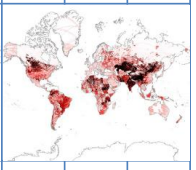
Figure 6: Pearson's Correlation coefficient depending on time lag from 0 to 180 days for ENE, TRO, AVI, and RES sectors.

3 Alterations to D2.6

3.1 Global emissions

Some possible improvement discussed in the previous global gridded uncertainty section of D2.6 deliverable report have been implemented, and revised global gridded yearly uncertainty maps per six emission groups were generated.

Main issue with the previously applied methodology was in cases where a substantial amount of grid-cells in a country had quite small emission values (e.g. emissions are distributed according to population density, rather than allocated to a certain grid cell). In such cases the country yearly uncertainty gridding approach was producing unrealistically high values for normalized standard deviations per grid-cell (main problematic countries were China, Russia, USA, Canada). The updated methodology filters grid-cells with less than 1 kt CO_2 emission from calculations, which is negligible for the overall country budget (applied only over countries where this is true), but prevents from overestimation of yearly gridded CO_2 uncertainty per emission group. Figure 7 summarises results geographically (maps) and statistically (min, mean, max values over the globe) for both previous and updated methodologies.

	Energy			Manufacturing industries			Residential/ commercial combustion			Road transport			Shipping			Other (agriculture, aviation, gas/oil)		
No emission filtering																		
	min	mean	max	min	mean	max	min	mean	max	min	mean	max	min	mean	max	min	mean	max
	0.05	6.61	17.12	0.02	0.85	7.56	0.10	3.15	6.62	0.04	3.33	35.48	0.03	2.13	9.35	0.06	20.13	89.13
	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$
	0.09	4.31	17.12	0.20	0.80	2.39	0.64	3.16	6.62	0.37	1.67	14.20	0.10	0.95	7.21	0.73	9.24	89.13
Emissions < 1 kt filtered																		
	min	mean	max	min	mean	max	min	mean	max	min	mean	max	min	mean	max	min	mean	max
	0.05	2.16	6.73	0.02	0.78	7.33	0.10	2.80	5.38	0.03	2.92	26.90	0.03	1.71	6.91	0.05	8.67	67.60
	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$	$P_{2.5}$	$P_{50.0}$	$P_{97.5}$
	0.09	2.06	4.36	0.17	0.69	2.27	0.28	3.01	5.38	0.23	1.56	11.99	0.05	0.74	6.91	0.36	3.74	41.59

Normalised standard deviation of anthropogenic CO_2 yearly emission

Figure 7: Summary of the methodology change impact on resulting yearly gridded CO_2 uncertainty per emission group.

3.2 European emissions

Some improvements to the previous error definition of D2.6 have been implemented, mainly meant to ease the use of the error data by modellers. Most of these improvements have been suggested by the modellers themselves and provided useful feedback to improve the product. All changes are shortly summarized here.

3.2.1 Spatial errors and additional sectors

One major change is that we improved the error definition in the proxy maps, based on literature, and we also included the representativeness error. This results in much larger gridded uncertainties than used previously (see Table 3).

In the D2.6 product spatial information was only included for two sectors: road transport and other stationary combustion. For the other sectors, the following information is now also included:

- Weighted proxy maps
- Gridded uncertainties (standard deviation), based on expert judgement
    - Public power and Industry: 100% only for non-point sources
    - Shipping: 30% for inland shipping
    - Other sectors: 100%
- Spatial error correlation length, based on expert judgement
    - Public power and Industry: no spatial correlation length, because this sector is dominated by point sources
    - Shipping: 100 km
    - Other sector: no spatial correlation length, as there are many sub-activities that are uncorrelated

In the previous product the ‘other sector’ was an aggregate sector of all GNFR sectors not listed separately. In the new product all GNFR sectors are presented as separate sectors, but

the GNFR sectors that were previously aggregated do get the same error estimates. Nevertheless, having separate sectors allows us to include sector-specific temporal profiles.

Moreover, in the previous product spatial errors were based on CO_2 only. Now, we included pollutant-specific spatial errors.

Table 3: Overview of proxy maps used for downscaling other stationary combustion and road transport, their 95% CI and correlation length.

Proxy map	Uncertainty (95% CI)	Correlation length (km)
RoadTransport_Urban_PC	0.6	15
RoadTransport_Urban_Mopeds	0.6	15
RoadTransport_Urban_Motorcycles	0.6	15
RoadTransport_Highway_HDV	0.6	28
RoadTransport_Highway_LDV	0.6	28
RoadTransport_Highway_Buses	0.6	28
RoadTransport_Highway_PC	0.6	28
RoadTransport_Highway_Motorcycles	0.6	28
RoadTransport_Highway_Mopeds	0.6	28
RoadTransport_Rural_Buses	0.6	21
RoadTransport_Rural_LDV	0.6	21
RoadTransport_Rural_HDV	0.6	21
RoadTransport_Rural_Motorcycles	0.6	21
RoadTransport_Rural_Mopeds	0.6	21
RoadTransport_Rural_PC	0.6	21
RoadTransport_Urban_HDV	0.6	15
RoadTransport_Urban_LDV	0.6	15
RoadTransport_Urban_Buses	0.6	15
Population_total_2015	0.32	23
Population_rural_2015	0.32	23
Population_urban_2015	0.32	23
Wood_use_2014	1.0	26

3.2.2 Spatial correlation length

The spatial error correlation length for road transport and other stationary combustion were also updated. In the previous dataset these correlation lengths were calculated per country and the median for all countries was used to represent the whole domain. Now we analysed the whole domain at once. For this, we used the actual data used to build the proxy maps instead of the proxy maps themselves, as fractions may show large gradients where small and large countries border on each other.

For population density, the Landscan dataset (Bright et al., 2016) of the year 2020 was used. This dataset was available in a resolution of 0.1x0.05 decimal degrees, and gives the number of people residing in a given grid cell. Only population data inside our study area was used for further calculations. Land use data comes from the Corine dataset for the year 2018, with an original resolution of 100x100 metres. The dataset was downscaled to a 5x5km resolution using mode resampling to match the resolution to the gridded emission database. Only the "Industrial or commercial units" and "Non-irrigated arable land" land use classes were used. Road intensity data was based on the OpenTransportMap, a dataset compiled from a combination of OpenStreetMap data, population data and transport modelling (Jedlička et al., 2016). The resolution of the dataset is 0.1x0.05 decimal degrees. We base our findings on the total vehicle kilometres for all road classes.

Based on the outlined spatial data, variograms were calculated using the gstat package (version 2.0.8) in the R programming language. For continuous spatial data (population density and road intensity), variograms were directly calculated from the aforementioned datasets. For the categorical land use data, we derived indicator variograms for the selected layers of the Corine land use dataset. Here, each of the layers of interest (L) was first transformed into an own binary variable (indicator I) as given in Equation 1 (Maleki et al., 2017):

$$I_k(x) = \begin{cases} 1, if\ x\ belongs\ to\ L_x \\ 0, otherwise \end{cases} \quad (1)$$

Model fitting was done with a spherical model and by employing the standard initial parameters for range, nugget and partial sill values as outlined in the gstat package manual (Pebesma, 2004).

The results show much larger correlation lengths than previously calculated (Figure 8). For population data, the spatial autocorrelation range of 496 km was calculated, which is thought to capture the larger population-density patterns within Europe. Using total vehicle kilometres for all vehicles and all road types, we found a range of 2019 km. When considering the total vehicle kilometres of passenger cars on all road types alone, the range parameter does not change considerably (2058 km).

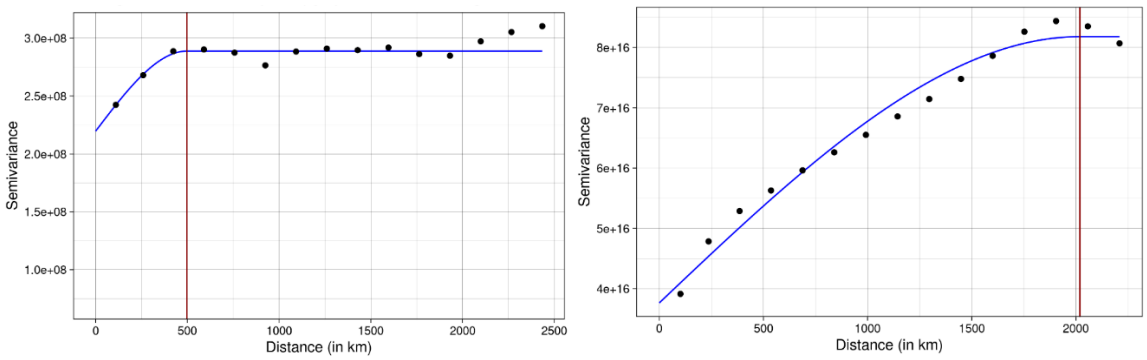


Figure 8: Variogram for Landscan (2020) population data, with a range of 496 km (left) and for total vehicle kilometres for all vehicle classes and all road types based on Open Transport Map (OTM) data, with a range of 2019 km (right).

Spatial autocorrelation ranges for land use classes were dominated by large-scale variations. We calculated ranges between 1900 km-2100 km for the analysed land use classes. This is generally in line with previous findings of geospatial patterns with the same dataset (Kallimanis & Koutsias, 2013). The *'Non-irrigated arable land'* layer, which also includes arable land with non-permanent irrigation, is the biggest land use class within the study area, with 26.37 %. The correlation length for this layer was calculated at 1971 km (Figure 9, left). In contrast, the *'Industrial or commercial units'* layer made up 0.17 % of the study area, and resulted in a correlation range of 2136 km (Figure 9, right). The large correlation length of these land use types can be understood in the context of the continuous patterns for many land use indicators

within Europe. As can be seen from Figure 10 for the example of *'non-irrigated arable land'*, an interconnected surface of arable land exists throughout most of the study domain.

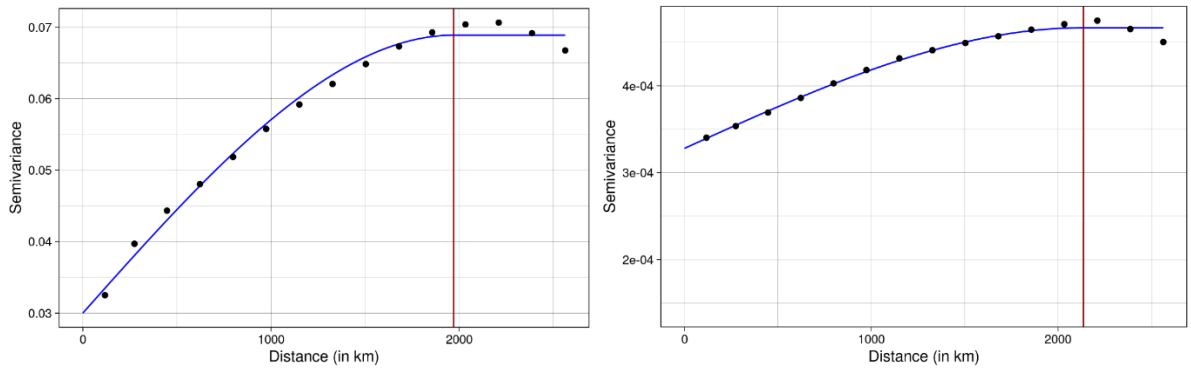


Figure 9: Indicator variogram for the *Non-irrigated arable land* layer of the Corine dataset (2018) dataset, with a range of 1971 km (left) and for the *Industrial or commercial units* layer of the Corine dataset (2018) dataset, with a range of 2136 km (right).

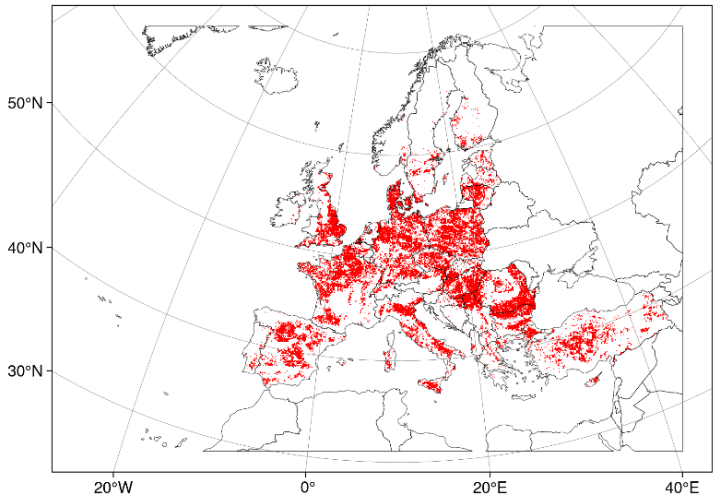


Figure 10: Areas dominated by non-irrigated arable land use (red) in Europe.

These results show that the methodological choices have a considerable impact on the calculated spatial correlation length. We believe that the optimal choice is one that takes into account the underlying spatial data, but also the set-up of the inversion framework as to make optimal use of the available observations. Hence, we now leave the correlation lengths as in the previous dataset and will continue optimizing it in the future.

3.2.3 CO:CO$_2$ error correlations

Finally, we started with estimating gridded CO:CO$_2$ error correlations for some sectors. For each pollutant the same proxy map is used for spatial downscaling of an NFR (Nomenclature for Reporting) sector. Hence, spatial errors are correlated at this level, but the correlation decreases when aggregating multiple NFR sectors due to different mixes of NFR contributions per pollutant. We tried to define a predictor that estimates the CO:CO$_2$ error corelation per grid cell for road transport and other stationary combustion. We plot the predictor against actual correlation coefficients based on a Monte-Carlo simulation for 7 countries and estimate the relationship between the predictor and the correlation strength. Combined results are shown in Figure 11. Although these figures illustrate that we may find a useful predictor for some sectors, some more work is needed. Therefore, these results are not yet included in the dataset. We plan to continue this work in the CORSO project.

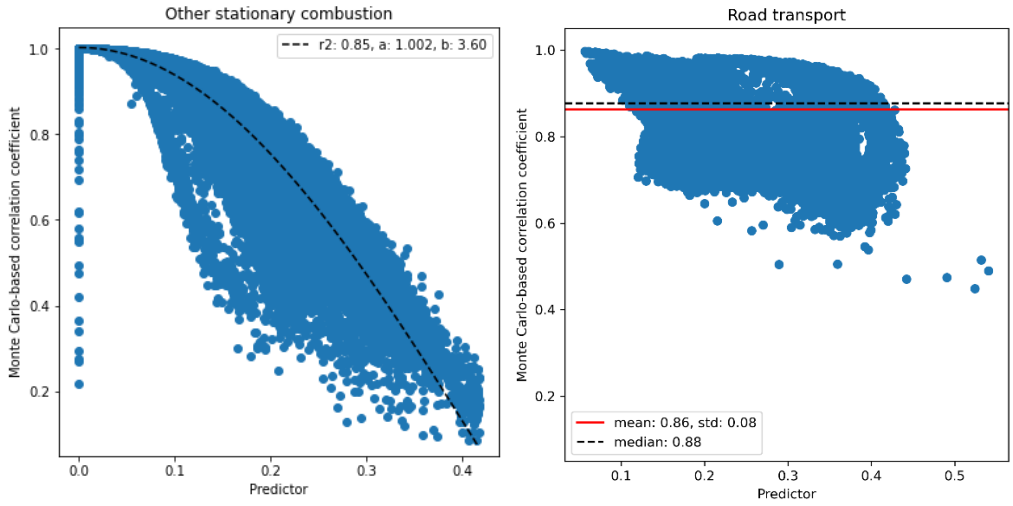


Figure 11: Scatter plots of Monte-Carlo (N=500) based correlation coefficient (r) per grid cell against the predictor. In the left panel the fit (R^2) and cosine function parameters are shown. In the right panel the mean, median and standard deviation (std) of the correlation coefficients are shown.

4 Conclusion and discussion

In this report we describe additional efforts and updates to D2.6 to quantify the different aspects of prior emission uncertainties. Here, we mainly focused on temporal errors, but Section 2 illustrates that also our previous work is still being updated and improved as new insights arise. This will be an ongoing effort for the next years, but we aim to provide the modellers with useful intermediate products. This will help us to understand what the modellers need by receiving their feedback.

A more extensive analysis of the time profile distributions will be made in the framework of the CORSO project, but for the temporal correlation lengths different methods will also need to be compared. Potentially, the dominant time scales will depend on the questions asked by the modellers. For example, in an inversion framework focussing on monthly fluxes the correlations between months might be more relevant than the weekly correlations, given the constraint provided by the scarce observations. Hence, we will continue our dialogue with the users on this topic. The same is valid for the choice of spatial correlation lengths.

Finally, in October an additional note to this report will follow, describing a first effort to propagate global uncertainties with the FFDAS model developed under T2.4.

References

Beck, H., Zimmermann, N., McVicar, T. et al. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data 5, 180214 (2018). https://doi.org/10.1038/sdata.2018.214

Bright, E., Rose, A., and Urban, M.: LandScan Global 2015, https://doi.org/https://doi.org/10.48690/1524210, 2016.

Choulga, M., Janssens-Maenhout, G., Super, I., Solazzo, E., Agusti-Panareda, A., Balsamo, G., Bousserez, N., Crippa, M., Denier van der Gon, H., Engelen, R., Guizzardi, D., Kuenen, J., McNorton, J., Oreggioni, G., and Visschedijk, A.: Global anthropogenic CO_2 emissions and uncertainties as a prior for Earth system modelling and data assimilation, Earth Syst. Sci. Data, 13, 5311–5335, https://doi.org/10.5194/essd-13-5311-2021, 2021.

Denier van der Gon, H. A. C., Hendriks, C., Kuenen, J., Segers, A., and Visschedijk, A. J. H.: Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns, EU FP7 MACC deliverable report D_D-EMIS_1.3, available at: https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC_TNO_del_1_3_v2.pdf, 2011.

Guevara, M., Jorba, O., Tena, C., Denier van der Gon, H., Kuenen, J., Elguindi, N., Darras, S., Granier, C., and Pérez García-Pando, C.: Copernicus Atmosphere Monitoring Service TEMPOral profiles for the Regional domain version 2.1 (CAMS-REG-TEMPOv2.1) Copernicus Atmosphere Monitoring Service [publisher] ECCAD [distributor], https://doi.org/10.24380/1cx4-zy68, 2020a.

Guevara, M., Jorba, O., Tena, C., Denier van der Gon, H., Kuenen, J., Elguindi, N., Darras, S., Granier, C., and Pérez García-Pando, C.: Copernicus Atmosphere Monitoring Service TEMPOral profiles for the Global domain version 2.1 (CAMS-GLOB-TEMPOv2.1) Copernicus Atmosphere Monitoring Service [publisher] ECCAD [distributor], https://doi.org/10.24380/ks45-9147, 2020b.

Guevara, M., Jorba, O., Tena, C., Denier van der Gon, H., Kuenen, J., Elguindi, N., Darras, S., Granier, C., and Pérez García-Pando, C.: Copernicus Atmosphere Monitoring Service TEMPOral profiles (CAMS-TEMPO): global and European emission temporal profile maps for atmospheric chemistry modelling, Earth Syst. Sci. Data, 13, 367–404, https://doi.org/10.5194/essd-13-367-2021, 2021.

Jedlička, K., Hájek, P., Ježek, J., Kolovský, F., Mildorf, T., Charvát, K., Kozhukh, D., Martolos, J., Šťastný, J. and Beran, D. (2016) Open Transport Map: open, harmonized dataset or road network, International Symposium on Computer-Assisted Cartography 2016.

Kallimanis, A. & Koutsias, N. (2013), Geographical patterns of Corine land cover diversity across Europe: The effect of grain size and thematic resolution, Progress in Physical Geography 37(2)

Maleki, M., Emergy, X. & Mery, N. (2017), Indicator variograms as an aid for geological interpretation and modeling of ore deposits, Minerals 7.12

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30: 683-691

Schepanski, K., Klüser, L., Heinold, B., and Tegen, I.: Spatial and temporal correlation length as a measure for the stationarity of atmospheric dust aerosol distribution, Atmospheric Environment, 122, 10-21, https://doi.org/10.1016/j.atmosenv.2015.09.034, 2015.

Super, I., Dellaert, S. N. C., Visschedijk, A. J. H., and Denier van der Gon, H. A. C.: Uncertainty analysis of a European high-resolution emission inventory of CO_2 and CO to support inverse modelling and network design, Atmos. Chem. Phys., 20, 1795–1816, https://doi.org/10.5194/acp-20-1795-2020, 2020.

Document History

Version	Author(s)	Date	Changes
	Name (Organisation)	dd/mm/yyyy	
V0.1	Ingrid Super (TNO)	9/5/2023	First version containing input TNO
V0.2	Ingrid Super (TNO), Margarita Choulga (ECMWF), Tilman Hohenberger (TNO)	14/6/2023	Second version including input ECMWF
V0.3	Hugo Denier van der Gon (TNO)	29/06/2023	TNO internal review
V0.4	Roxane Petrescu (VU)	05/07/2023	VU review
	Konstantinos Politakos (FORTH)	25/07/2023	FORTH review
	Rhona Phipps	25/07/2023	
V1.0	Ingrid Super (TNO)	26/07/2023	Review comments processed

Internal Review History

Internal Reviewers	Date	Comments
Ingrid Super (TNO), Margarita Choulga (ECMWF), Tilman Hohenberger (TNO), Hugo Denier van der Gon (TNO), Roxane Petrescu (VU), Konstantinos Politakos (FORTH)	May-July 2023	Input and updates to create v1.0