Emission ensemble approach to improve the development of multi-scale emission inventories

Philippe Thunis[1], Jeroen Kuenen[2], Enrico Pisoni[1], Bertrand Bessagnet[1], Manjola Banja[1], Lech Gawuc[3], Karol Szymankiewicz[3], Diego Guizardi[1], Monica Crippa[1,4], Susana Lopez-Aparicio[5], Marc Guevara[6], Alexander De Meij[7], Sabine Schindlbacher[8], Alain Clappier[9]

[1] European Commission, Joint Research Centre, Ispra, Italy
[2] TNO, Department of Air, Climate and Sustainability, Utrecht, The Netherlands
[3] Institute of Environmental Protection – National Research Institute (IEP-NRI), Słowicza 32, 02-170 Warsaw, Poland
[4] Unisystem S.A., Milan, Italy
[5] NILU – Norwegian Institute for Air Research, 2027 Kjeller, Norway
[6] Barcelona Supercomputing Center, Barcelona, Spain
[7] MetClim, Varese, 21025, Italy
[8] Environnent Agency Austria, Spittelauer Lände 5, 1090 Vienna, Austria
[9] Université de Strasbourg, Laboratoire Image Ville Environnement, Strasbourg, France

Correspondence to: Philippe Thunis (Philippe.THUNIS@ec.europa.eu)

Abstract

Many studies have shown that emission inventories are one of the input with the most critical influences on the results of air quality modeling. Comparing emission inventories among themselves is therefore essential to build confidence in emission estimates. In this work we extend the approach of Thunis et al. (2022) to compare emission inventories by building a benchmark that serves as reference for comparisons. This benchmark is an ensemble that is based on three state-of-the-art EU-wide inventories: CAMS-REG, EMEP and EDGAR. The ensemble-based methodology screens differences between inventories and the ensemble. It excludes differences that are not relevant and identifies among the remaining ones, those that need special attention. We applied the ensemble-based screening to both a EU-wide and a local (Poland) inventory.

The EU-wide analysis highlighted a large number of inconsistencies. While the origin of some differences between EDGAR and the ensemble can be identified, their magnitude remains to be explained. These differences mostly occur for SO_2, PM and NMVOC, for the industrial and residential sectors, and reach a factor 10 in some instances. Spatial inconsistencies mostly occur for the industry and other sectors.

At the local scale, inconsistencies relate mostly to differences in country sectorial shares that result from different sectors/activities being accounted for in the two types of inventories. This is explained by the fact that some emission sources are omitted in the local inventory due to lack of appropriate geographically allocated activity data. We identified sectors and pollutants for which discussion between local and EU-wide emission compilers would be needed in order to reduce the magnitude of the observed differences (e.g. in the residential and industrial sectors).

The ensemble-based screening proved to be a useful approach to spot inconsistencies by reducing the number of necessary inventory comparisons. With the progressive resolution of inconsistencies and associated inventory improvements, the ensemble will improve. In this sense, we see the ensemble as a useful tool to motivate the community around a single common benchmark and monitor progress towards the improvement of regional and locally developed emission inventories.

Keywords: emission inventories, quality assurance, quality control, screening, urban emissions, ensemble

1. Introduction

Many studies have shown that emission inventories are one of the inputs with the most critical influences on the results of air quality modeling (Kryza et al., 2015, Zhang et al., 2015). Even more concerning, certain studies have shown that important uncertainties affect emission inventories, which may impeach conclusions based on air quality model results (Trombetti et al., 2018, Markakis et al., 2015). These uncertainties result from the need to compile a wide variety of information to develop an emission inventory. For the many pollutants and activity sectors to cover, the spatial and temporal distribution of emissions is typically based on proxies that can be estimated through different methods.

In Thunis et al. (2022), we showed that comparing emission inventories is an effective way to detect inconsistencies when differences are very large. A methodology was designed to compare two emission inventories, one against the other. This methodology identifies disparities between the two inventories by assessing country totals, their sectorial share and the proportion of the country emissions attributed to the urban areas. In this work, we adhere to the same principle of analyzing differences while introducing a novel ensemble concept to facilitate the simultaneous comparison of a larger number of inventories.

Ensemble of models have widely been used in climate (Kotlarski et al., 2014) and air quality modelling fields throughout the world (Stevenson et al., 2006; Vautard et al, 2009; Marecal et al. 2015; Brasseur et al., 2019) as they generally provide better and more robust results. While in some instances, reference values (e.g., measurements) exist against which models can be compared, this is unfortunately not the case for emissions, and hence the emission ensemble is not necessarily better than any of its members. The emission ensemble is therefore not a more accurate inventory. This is, however, not an issue as the ensemble is used here as a common benchmark for comparison. Moreover, our focus is on differences between emission estimates rather than on their absolute values, for which accuracy and robustness is of secondary importance. The underlying concept is that above a certain threshold, differences are so large that one or both inventories can be considered wrong. The choice of this vocabulary, i.e. wrong is intentional and is meant here to foster the process of reviewing the data when differences exceed a given threshold. In other words, a factor 100 difference between inventories for a given sector/pollutant most likely reveals one or more significant errors (or inconsistencies) which are relatively straightforward to identify and must be addressed in either one or both inventories. The methodology screens differences between inventories, excludes differences that are not relevant (i.e., large differences on low emission values are disregarded) and identifies among the remaining ones, those that need special attention.

In addition to this key advantage, several other objectives are pursued by introducing the ensemble for EU wide emission inventories, namely (1) to create a unique common benchmark to monitor and quantify the current level of agreement among the ensemble members; (2) to identify and characterize the largest mismatches in terms of pollutant, sector among them; (3) to foster interactions between EU wide emission inventory developers around identified inconsistencies and (4) to allow for comparing additional inventories (e.g. bottom-up ones) with the ensemble. A comparison of the ensemble with local (intended here as national or sub-national) inventories can be indeed helpful, as they are independent estimates, which methods are based on local knowledge and understanding of the activities and processes that result on emissions.

The work is structured as follows. In Section 2, we review the screening methodology proposed in Thunis et al. (2022) and discuss the construction of the ensemble in the frame of this screening approach. In Section 3, we apply the ensemble-based screening approach to one European-wide inventory whereas in Section 4 we illustrate how this ensemble can then be compared to local inventories in a bilateral manner. For the latter, a local inventory developed for Poland is used. In Section 5, we discuss the main findings from both type of comparisons and conclude in Section 6.

2. Description of the methodology

2.1 Overview of the screening methodology

In this section, we provide a brief summary of the screening method detailed in Thunis et al. (2022). The approach aims at comparing two emission inventories over a series of urban areas over which the consistency is assessed for all sectors and pollutants. Based on gridded annual emissions detailed in terms of pollutants (“p”) and sectors of activity (“s”), the data required for each pollutant and sector ([p,s] couple) are twofold and consist of (1) emissions aggregated over specific urban areas (lowercase notation $e_{p,s}$) and country scale emissions (uppercase notation $E_{p,s}$).

The consistency between emissions in both inventories is assessed around three aspects: (1) the total pollutant emissions assigned at country level; (2) the way these country emissions are distributed across sector and 3) the way country emissions are distributed spatially, and therefore, allocated to main urban areas. To address these three aspects, we decompose the ratio of the known pollutant-sector emissions for each city as follows:

$$\frac{e_{p,s}^1}{e_{p,s}^2} = \frac{\frac{e_{p,s}^1}{E_{p,s}^1}}{\frac{e_{p,s}^2}{E_{p,s}^2}} * \frac{\frac{E_{p,s}^1}{\bar{E}_p^1}}{\frac{E_{p,s}^2}{\bar{E}_p^2}} * \frac{\bar{E}_p^1}{\bar{E}_p^2} \tag{1}$$

where $\bar{E}_p$ represents the country scale emissions summed over all sector for a given pollutant. Superscripts refer to the two inventories used for the screening. Equation (1) is an identity where all terms are known from input quantities, i.e. the city and country scale emissions detailed in terms of pollutants and sectors. The three terms on the right-hand side of the identity provide information on spatial distribution (*FAS*, Focus Area Share*)*, on the country sectorial share (*LSS*, Large Scale Sectorial share*)* and on the country pollutant totals (*LPT*, Large scale Pollutant Total*)*.

For convenience, we rewrite equation (1) in logarithm form as:

$$log\left(\frac{e_{p,s}^1}{e_{p,s}^2}\right) = log\left(\frac{\frac{e_{p,s}^1}{E_{p,s}^1}}{\frac{e_{p,s}^2}{E_{p,s}^2}}\right) + log\left(\frac{\frac{E_{p,s}^1}{\bar{E}_p^1}}{\frac{E_{p,s}^2}{\bar{E}_p^2}}\right) + log\left(\frac{\bar{E}_p^1}{\bar{E}_p^2}\right) \tag{2}$$

Which can be rewritten as equation (3) with simplified notations:

$$\hat{e} = \widehat{FAS} + \widehat{LSS} + \widehat{LPT} \tag{3}$$

where the hat symbol (^) indicates that quantities are expressed as logarithmic ratios. These three quantities form the basis of the screening methodology and serve as input information for a graphical representation that facilitates the interpretation of the results.

As the number of [*p*,*s*] points under screening, equivalent to the product of the number of pollutants and sectors further multiplied by the number of urban areas (i.e. $N \times N_p \times N_s$), may become overwhelming, we adopt a series of steps to concentrate the screening on priority aspects. First, we restrict the screening to emissions that are relevant, i.e. large enough. As shown in Thunis et al. (2022), this exclusion step leads to eliminating a large fraction of the [*p*,*s*] couples from the screening process (between 80 and 90%). Second, we flag, among the remaining emissions, only those for which inventory emission ratios are larger than a given threshold (β_t).

When differences are small, it is not possible to tell whether they originate from methodological choices or from errors. We refer to these small differences as “uncertainty”. Although very large differences may result from methodological choices as well (e.g., inclusion or not of particulate matter condensable emissions for the residential sector), they are more likely to be associated to errors. Given the magnitude of the differences, it will in most cases be possible to identify one best value out of the two inventory estimates, even though the true emissions are unknown. These large differences are named “inconsistencies”. In the proposed screening methodology, a β_t threshold of 2 (free parameter) is introduced to distinguish inconsistencies from uncertainties.

As a follow-up step, all [p,s] couples that remain after the relevance test and inconsistency detection steps ($\beta_{p,s} > \beta_t$), are used to calculate an “Emission Consistency Indicator (ECI)” as follows:

$$ECI = \max_{\{relevant\ emissions\}} \frac{\log(\beta_{p,s})}{\log(\beta_t)} \quad (4)$$

The ECI quantifies the maximum difference among all relevant [p,s], normalized by the inconsistency level (β_t). It therefore quantifies the ratio between the maximum inconsistency and the assumed level of uncertainty. A value of ECI less than one means that all differences are considered as uncertainty (in other words none of the inventory can be identified as best performing). Together with the ECI, which quantifies this maximum difference, we associate the percentage of inconsistent [p,s] with respect to the total number of relevant data, to provide information on the number of detected inconsistencies.

Finally, we prioritise inconsistencies following the LPT – LSS – FAS hierarchy. In other words, if large scale inconsistencies are spotted for LPT, they are flagged as the priority, regardless of the magnitude of inconsistencies calculated for LSS and/or FAS. If no inconsistency is flagged for LPT, the same holds for LSS regardless of the level of inconsistency calculated for FAS. Consequently, the inconsistency flagged as priority might not be the largest inconsistency. This hierarchy is motivated by the fact that addressing large scale inconsistencies will lead to potentially resolving several issues at once (e.g. all urban areas within a given country). Inconsistencies are counted when the individual terms in equation (3) are larger than the threshold β_t but also when the indicators sums (i.e., $\widehat{FAS} + \widehat{LSS} + \widehat{LPT}, \widehat{LSS} + \widehat{LPT}$) exceed this threshold.

It is important to note that the method follows a bottom-up approach, i.e., we assess the three types of inconsistencies for each city, pollutant and sector. This means that the same LPT inconsistency is counted for all cities within a given country or for all sectors for a given pollutant. Similarly, a LSS inconsistency is counted for each city belonging to the same country. While this might be seen as double counting of some inconsistencies, the approach allows comparing local vs country scale indicators.

2.2 Construction of an ensemble as reference

This work aims at applying a novel ensemble concept to extend the Thunis et al. (2022) methodology to several inventories. The ensemble is calculated from EU-wide inventories that have been developed and regularly updated over several years within the EU[1]. While either the mean or the median of these inventories could be used to calculate the ensemble, we choose to use the median as it has been shown to be a more robust indicator compared to the mean (Riccio et al. 2007). Indeed, if one of the inventories is a strong outlier (i.e., much larger or much smaller values), the mean would be strongly influenced by these extreme values and would differ from the values of most of the inventories. On the other hand, the median is not affected by extreme values and therefore takes a value closer to the values taken by most of the inventories. It therefore remains further away from outliers, which become easier to identify.

[1] Note that EDGAR is designed as a global inventory but we consider here its European coverage only in this analysis and refer to it as a European wide inventory

In this work, the ensemble is created from three state-of-the-art Europe wide inventories: CAMS-REG (Copernicus Atmospheric Monitoring Service), EMEP and EDGAR.

EDGAR is a comprehensive global emission inventory providing country and sector specific greenhouse gas and air pollutant emissions from 1970 up to date. EDGAR is becoming a global reference for anthropogenic emissions, in particular contributing to the IPCC AR6 (Sixth Assessment Report) and to the annual UNEP emissions gap reports (UNEP2023) tackling global climate change issues. In the context of air pollution, EDGAR is also widely used by air quality modellers, playing an important role as gap-filling inventory in the Hemispheric Transport of Air Pollution mosaic compilation. Emissions are computed using a consistent methodology for all world countries, following the IPCC Guidelines (IPCC 2006, 2019) and EMEP/EEA Guidebook (EMEP/EEA, 2016, 2019) for greenhouse gases (GHGs) and air pollutants, respectively. Emissions are calculated for all anthropogenic sectors outlined by the IPCC excluding Land Use, Land Use Change and Forestry. This computation utilizes international statistics and default emission factors complemented with state-of-the-art information. Subsequently, annual emissions specific to each sector and country are downscaled globally at 0.1x0.1 degree employing a multitude of spatial proxies. Comprehensive insights into the EDGAR methodology and the underlying assumptions regarding the spatial data used for downscaling national emissions are available in several scientific publications (Janssens-Maenhout et al. 2015, 2019; Crippa et al. 2018, 2021; Crippa et al. 2020; Oreggioni et al. 2022). Additionally, the yearly emission data are further disaggregated into monthly emissions to further support atmospheric modellers in capturing the seasonality of anthropogenic emissions (Crippa et al. 2020).

CAMS-REG version 5.1 is an emission inventory developed as part of CAMS to support European scale air quality modelling (Kuenen et al. 2022). The inventory builds on the officially reported emission data to EMEP in the year 2020, which are complemented by other sources where reported data are not available or deemed of insufficient quality. The data are spatially distributed consistently across the entire domain at a resolution of 0.05x0.1 degrees (latitude-longitude). The spatial distribution takes into account specific point source emissions as reported in the European Pollutant Release and Transfer Register (EPTR2022) to correctly represent point source emissions to the extent possible. The emissions are provided in GNFR (Gridded Nomenclature For Reporting) format. The emission dataset is used in support of the CAMS regional modelling activities, but is also publicly available to support air quality assessment at European level. CAMS-REG-v5.1 is an update of version 4.2 that includes official national emission submissions for the year 2020.

The EMEP-GNFR emissions (Mareckova et al., 2017), based on 2017 reporting, are compiled within the “UNECE co-operative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe”, or also known as EMEP. EMEP is a scientifically based and policy driven programme under the Convention on Long-range Transboundary Air Pollution (CLRTAP) for international co-operation, that has the final aim of solving transboundary air pollution problems. Emissions are built from officially reported data provided to CEIP (Centre of Emission Inventory and Projection by the Member States in Europe) and follow the EMEP/EEA guidebook guidelines (EMEP/EEA 2019) to define the annual totals. The emissions are gap-filled with gridded TNO data from CAMS and EDGAR. The dataset consists of gridded emissions for SO_x, NO_x, NMVOC, NH_3, CO, $PM_{2.5}$, PM_{10} and PMcoarse at 0.1° x

0.1° resolution. More information on the emissions and where to download can be found in the User Guide (https://emep-ctm.readthedocs.io/en/latest/) and in Mareckova et al., (2017). The EMEP domain covers the geographic area between 30°N-82°N latitude and 30°W-90°E longitude.

Based on these three inventories, the ensemble is defined on a yearly basis (here 2018). Urban ($e_{p,s}$) and country emissions ($E_{p,s}$) for the selected year are required as input. Independent ensemble values for *E* and *e* are defined for each pollutant-sector couple [p,s] as the median of the three inventory values. For a given area, the urban and country scale emission ensembles for a given year read as:

$$\begin{aligned} e_{p,s}^{ens} &= median\left\{e_{p,s}^{CAMS}, e_{p,s}^{EMEP}, e_{p,s}^{EDGAR}\right\} \\ E_{p,s}^{ens} &= median\left\{E_{p,s}^{CAMS}, E_{p,s}^{EMEP}, E_{p,s}^{EDGAR}\right\} \end{aligned} \tag{5}$$

Note that this calculation implies that $e_{p,s}^{ens}$ and $E_{p,s}^{ens}$ might not belong to the same inventory for a given area and pollutant-sector couple [p,s]. It is also worth mentioning that should one inventory pollutant-sector value behave as an outlier; its value will not be selected in the ensemble.

As the three emission inventories are characterised by different grid resolutions and sector aggregations, harmonisation is required to construct the ensemble. This is done in 2 steps:

- by grouping the initial emission categories into common categories based on the GNFR classification (NFR-I, 2023 and Table 1 in supplementary material). The original GNFR sectors have been aggregated in 5 categories: road transport (F), residential (C), power plants (A), industry (B) and others. The latter category includes fugitive emissions (D), solvents (E), shipping (G), aviation (H), off-road transport (I), waste (J) and agriculture (K-L).
- by aggregating gridded emissions on common polygons that delineate the area covered by an urban area or by a country. Urban area emissions ($e_{p,s}$) are calculated over functional urban areas (FUA, OECD 2012), composed of a core city plus its wider commuting zone, consisting of the surrounding travel-to-work areas. About 150 FUAs across Europe are selected for this screening. Details on these urban areas are provided in Thunis et al. (2018). The larger scale emissions ($E_{p,s}$) are defined at country level, level at which emissions are initially reported for these emission inventories.

In terms of pollutants, we consider NOx, NMVOC, PM2.5, PMco (coarse PM, calculated as the difference between PM10 and PM2.5 emissions), SO2 and NH3.

The approach then consists in comparing a given inventory with the ensemble to identify inconsistencies. It is important to note that while the approach likely highlight errors in the inventory under screening, it is however not possible to exclude that the inconsistency originates from the ensemble (i.e., be present in all other inventories). Despite this inconveniency, the method remains an efficient way to identify, among the large amount of data from several inventories, those that are most likely to be problematic and therefore need to be verified in priority.

3. Application to EU-wide inventories

The first objective of the ensemble-based screening is to systematically monitor and quantify existing uncertainties and inconsistencies within EU-wide inventories. It aims to identify the sources of discrepancies in terms of pollutant, sector and location. To perform this task, we compare bilaterally each of the three inventories to the ensemble and present the findings in Figure 1 (left). This figure provides for all ensemble members an overview of existing inconsistencies, i.e. for emissions that are relevant (i.e., large enough values) and that differ from the ensemble by more than a factor 2 ($\beta_t = 2$). Each inconsistent emission [p, s] is represented by a point that has larger-scale emissions ($\widehat{LSS} + \widehat{LPT}$) as abscissa and spatial distribution of emissions ($\widehat{FAS}$) as ordinate. The sum of these two terms is equal for points that lie on "−1" slope diagonals. The diamond shape (in the middle of the diagram) delineates the inconsistency limits. Therefore, each [p, s] point lying outside this shape is an inconsistency. In this diamond diagram, shapes are used to differentiate activity sectors, while colors indicate pollutants. The size of the symbol is proportional to the relevance of the emission contribution. Finally, we use symbol filling to distinguish the type of inconsistencies (i.e., LPT, LSS, and FAS). We refer to Thunis et al. (2021) for details.

The summary report (bottom part of Figure 1) provides overview information about inconsistencies. More than 21% (number within brackets beside the ECI indicator) of the relevant emission ratios show inconsistencies. The ECI indicator is equal to 132, meaning that the largest inconsistency is more than two orders of magnitude larger than the level associated to uncertainties. The EDGAR inventory is flagged for two thirds of them (the total number of inconsistencies, denoted as NI is 227 out of 357), with the largest part of them associated to industry for SO_2 and PM_{co} (see numbers within brackets besides the sectors/pollutants in the bottom legend: Figure 1). Most of the inconsistencies are obtained within the allocation of emissions at urban scale (218), although an important number of them also occur at country scale (LSS+LPT=80+59). The diagram also shows that EDGAR reports larger residential and industrial emissions at country level (yellow squares on the right of the X-axis). It is important to remember that flagging one particular inventory does not necessarily indicates that this inventory is the problematic one. But this flagging means that this inventory and/or the others show an important inconsistency for that city, pollutant and sector which requires further checking.

In addition to providing a useful summary that details the current state of variability, the diagram can also serve as basis to monitor progress, through the ECI indicator and associated percentage.

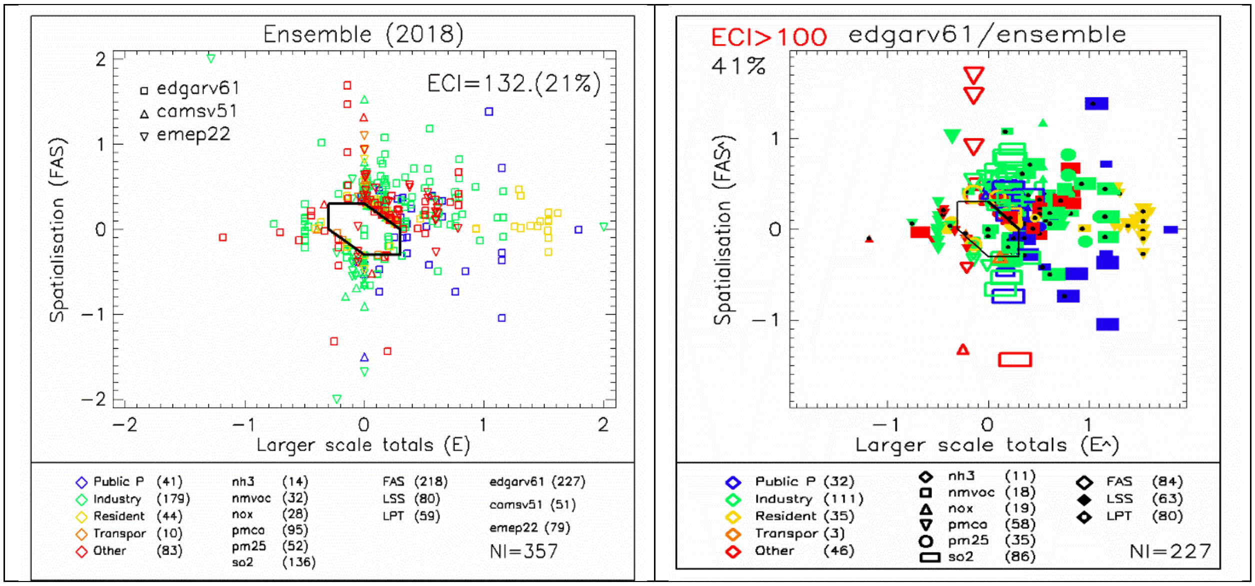

Figure 1: Overview diamonds. The left diagram shows the comparison of the three ensemble members (CAMS-REG, EDGAR, EMEP) with the ensemble for 2018. The right picture isolate the bilateral comparison between EDGAR and the ensemble. Symbols and colours are as specified in the legend. Please note that symbols/colors differ between the right and left figures. In both diagrams, only inconsistencies are displayed. For visualization purposes, we limit the axis to a factor 2 in terms of magnitude (from -2 to 2) and bound the ECI to 100 (e.g. values of ECI larger than 100 are plotted with a value of 2). Numbers within bracket in the bottom legend are the total number of inconsistencies for a given pollutant, sector or type.

The ensemble-based screening methodology also serves as a benchmark to compare individual inventories. It is applied here (Figure 1 - right) to one of the three state of the art inventories used to build the ensemble, EDGAR v.6.1 (Crippa et al. 2022). Results for the two other ensemble members: CAMS-REG v5.1 and EMEP (2022 gridding) are discussed in the supplementary material (Section 1).
The ECI (>100) indicates that the maximum inconsistency is at least a factor 100 larger than the estimated level of uncertainty. Moreover, about 41% of the relevant emission points show an inconsistency. As indicated in the overview table, these 41% amount to 227 inconsistencies (NI) which are shared into about 35% within the spatial distribution of emissions (FAS=84) and 65% at country scale (LPT+LSS=83+80). Most of the inconsistencies are identified, as for SO_2, PM_{co} and $PM_{2.5}$ from the industry sector, in line with the findings of De Meij et al. (2024). There are also an important number of inconsistencies related to the other (46), residential (35) and public power sectors (32). In general, for all inconsistencies, EDGAR estimates are larger than those represented by the ensemble (all points on the right and/or top of the diagram).

To prioritize the inconsistency analysis, Figure 2 (right side) shows the largest differences for LPT (country pollutant total), LSS (country sectorial share) and FAS (spatial distribution), which are also identified on the map (left on Figure 2).

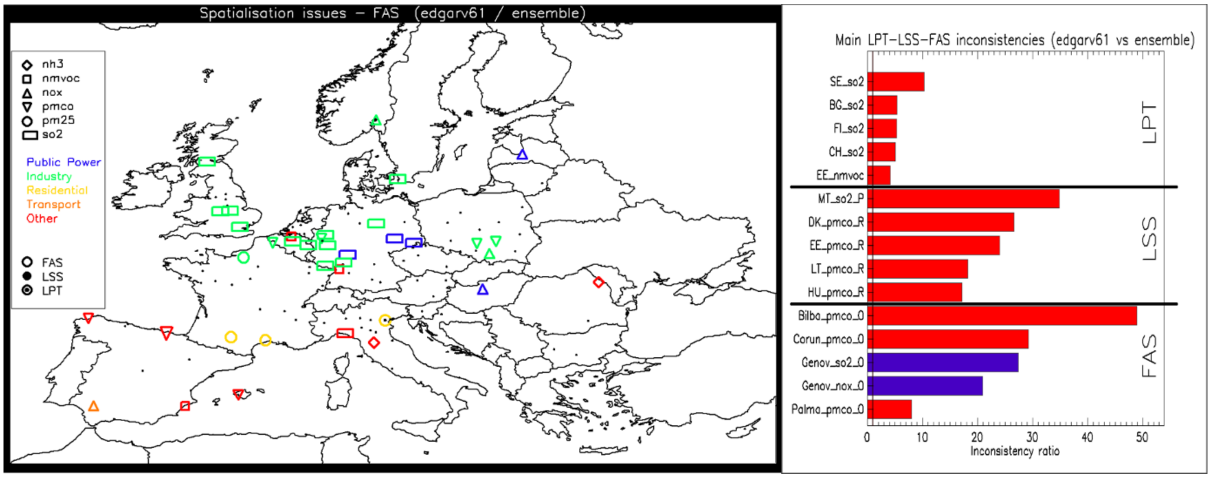

Figure 2: Left: Main inconsistencies spotted at urban scale for EDGAR when compared to the ensemble. Only the main spatial inconsistency (FAS) for each city is plotted. See explanation of symbols on the top left of the figure. Right: Major LPT (top 5), LSS (middle 5) and FAS (lower 5) inconsistencies. The two first letters indicate the country code for LSS and LPT whereas the 4 first city letters are given for FAS. Red shading indicates an overestimation and blue shading an underestimation for the EDGAR inventory

The following main issues can be extracted from Figure 2 for EDGAR:

- Inconsistencies in SO_2 country totals (LPT) are notably observed in Sweden (factor 10), Bulgaria, Finland and Switzerland (factor 5). In the case of Sweden and Finland, we could identify that the main difference comes from the industry sector, particularly the pulp, paper and print sub-sector, for which the inclusion of black liquor use for energy purposes in EDGAR need to be revised. For Bulgaria, the SO_2 total is dominated by the public power sector for which the activity data, sourced from IEA energy balances, subject to regular updates, influence the magnitude of the differences. According to the Bulgarian Informative Inventory Report (IIR) of emissions in 2022, SO2 emissions are regularly updated with measurements, which is not the case for the EDGAR emissions estimates, explaining part of the differences. Work is in progress to update SO_2 abatement measures in EDGAR. Another issue that can explain these inconsistencies relates to the different emission factors applied for SO_2 that are based on the sulphur content of fuels, usually not reported regularly by countries, values which are integral to CAMS-REG and EMEP[2]. As a follow-up of this analysis, the SO_2 emission factors for the power sector in EDGAR have been revised taking into account the limits established by the implementation of the large Combustion Directive (Directive 2001/80/EC).

- A larger sectorial share (LSS) at the country level for SO_2 in Malta for Public Power (factor 30), for residential PMco emissions in Denmark, Estonia (above a factor 20) and Lithuania and Hungary (about a factor 10) is found. The large differences in the residential sector is related to biomass burning emissions, both in terms of technology allocation and emission

[2] The default EMEP/EEA Guidebook 2019 emission factor for SO_2 are w/o abatements and only for 1% mass sulphur content for coal and oil and 0.01 g/m3 for gas (EMEP/EEA guidebook 2019).

factors applied. Given the large differences with the ensemble, the review of the EDGAR methodology led to the indication that EDGAR estimates needed to be updated, especially in terms of technology allocation. This adjustment is important to accurately reflect the current technological structure within that sector. Although the filter on low emission values (relevance test) is applied, it is not effective in the case of Malta because it is a small country where national totals are composed of few power plants only. The large LSS ratios obtained there are not significant as the values estimated for the power plant sector appear to be very small.

- A few large inconsistencies also appear at the local scale (FAS) due to the use of different proxies to spatially distribute emissions. The largest inconsistencies occur for the other sector (likely originating from the waste treatment installations). This can probably be explained by the approach followed in EDGAR for the waste sector for which all emissions are distributed over a few locations only, using E-PRTR locations for landfilling and incineration and population in case of missing information. This results in large differences with other inventories due to the proportion of the emissions being placed within the city area (see Figure 7 and following in supplementary material, section 3). A similar issue appears in many north west European cities for SO_2 for public power (green rectangles in the left Figure). Work is in progress to update the spatial allocation of the public power and waste sectors emissions (personal communication M. Crippa 2023).

The ensemble-based comparison highlights an important number of inconsistencies at country level. It is important to note that two ensemble members (EMEP and CAMS-REG) use officially reported emissions and therefore rely on similar total emissions per country. On the other hand, EDGAR estimates emissions in an independent bottom-up approach, starting from activity levels and emissions factors from international agencies and bodies (Crippa et al., 2018, Oreggioni et al. 2022). This difference in approach can explain a large number of inconsistencies identified for EDGAR but some of them are very large, especially for SO_2 and PM in the industrial sector. For this particular sector, estimates mostly come from the LPS and E-PRTR databases in EMEP/CAMS-REG, with emissions being mostly based on measurements or facility-level estimates. Such information is not used in EDGAR, where estimates are based on fuel consumption and emission factors that are very general and not plant specific.

4. Application to local inventories: a case-study over Poland

4.1 The high resolution Poland emission inventory

The ensemble-based screening methodology also serves as a benchmark to compare local inventories. In this section, it is applied to the inventory for Poland.

The Central Emission Database (CED) is a local emission inventory designed for Polish national air quality modelling. The CED is based on source location and provides accurate resolution-free data, which can be gridded depending on the requested target resolution for different computational grid configurations over Poland (typically 2.5 km over the entire country and 0.5

km for agglomeration zones). The majority of data is processed with respect to its exact geographical location. Priority is given to the most critical sectors, like residential combustion (described in detail in Gawuc et al., 2021) and road transport. The road transport data presented in this paper (relative to 2019) was based on a traffic model for the major roads in the country. Emissions on minor roads were distributed using the residue values taken from subtracting emission on major roads from the national totals. The current methodology is based on smartphone car navigation app which provides GPS data on road traffic and annual average car speed.

One of the essential components of CED is the "National database on greenhouse gases and other substances emission" (so-called national database – NB). NB consists of information on installations and sources' location responsible for emission into the atmosphere. NB has similarities to E-PRTR, but unlike it, it covers all emission sources regardless of type, power or production level. Registered NB users provide information on emission volumes resulting directly from the exploitation of their installations, as well as ancillary processes, which may cause fugitive emissions. To be applied for CED and air quality modelling, the reported data is categorized into SNAP (Selected Nomenclature for Air Pollution) and converted to GNFR if needed (Table 1, supplementary material).
NB is a basis for GNRF A (public power), B (industry), D (fugitive), E (solvents), and J (waste) emission estimations contributing to CED. Two approaches are applied to evaluating CED data. Firstly, as part of each modelling stream (i.e., operational air quality forecast, annual air quality assessment, station representativeness analysis), a comprehensive evaluation is undertaken (station-by-station time series for over 100 monitoring sites for each pollutant). Moreover, spatial patterns of the increments calculated in the assimilation procedure led to identify and improve the assumptions behind CED. The database is updated every year and there is a continuous attempt to improve emission estimates both – for total load and spatial distribution of sources. Modelling results helped to identify missing sources (e.g. resuspension, underestimated agriculture sector, domestic water heating). All sectors in CED are constantly improved using the best available activity data.

Note that the CED reference year (2019) differs from the ensemble one (2018). Inconsistencies are however generally large enough to justify explanations other than those originating from the difference in terms of reference year.

4.2 Comparison of the CED inventory to the ensemble

The ensemble-based screening applied to Poland is performed for 14 cities (see city locations in Figure 5), 5 sectors and 6 pollutants, leading to 420 emission ratios being tested.

Before proceeding with the screening of the local data, we first analyse the level of consistency among EU-wide inventory over Poland (Figure 3 is a zoom of Figure 1 over Poland). Among the 420 available data, 84 remain after the relevance test ($\gamma_t > 0.5$). These 84 [p,s] points serve as basis to identify inconsistencies ($\beta_t > 2$). Inconsistencies occur for about 13% of the relevant [p,s] points, with a maximum inconsistency (ECI) 2.5 times larger than the assumed level of uncertainty. As seen from the overview table, most of the issues are related to the EDGAR (20) and EMEP (6) inventories, in particular to the residential sector for EDGAR, to the industry sector for CAMS-REG and to the other sector for EMEP. Additional details are provided in the supplementary material (Section 2).

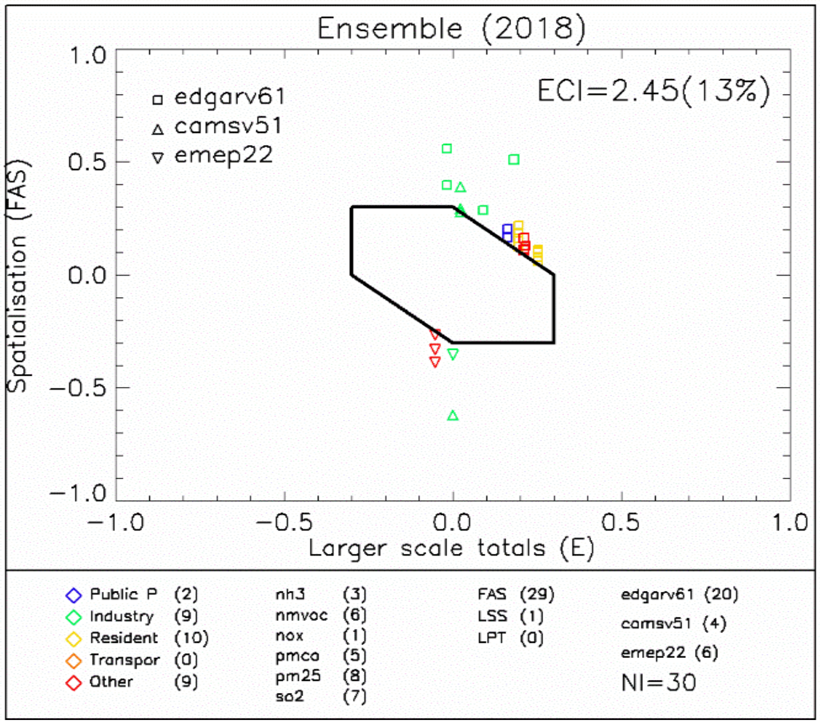

Figure 3: Overview diamonds. The diagram shows the comparison of the three ensemble members (CAMS-REG, EDGAR, EMEP) with the ensemble inventory over Poland. Symbols and colours are as specified in the legend. In all diagrams, only inconsistencies are displayed.

The overview diamond diagram (Figure 4 - left) shows the comparison of the CED local inventory with the ensemble. It indicates that out of the 420 emission ratios being tested, only 73 are associated to relevant emissions among which 49 (i.e. 67%) are identified as inconsistencies. The consistency indicator (ECI) is around 14, indicating that the maximum inconsistency is larger than the assumed level of uncertainty by a factor 14. The summary table (at bottom of the diamond, Figure 4) points to the residential and other sectors as the main issues with NMVOC and $PM_{2.5}$ in terms of pollutants. Most inconsistencies originate at country level, and mostly related to the country sectorial share.

PM residential emissions are systematically larger in CED than in the ensemble for $PM_{2.5}$, whereas smaller for PM_{co}. This can be partially explained by the inclusion of condensable emissions in CED (not included in EU-wide ensemble). Note that including or not condensable emissions lead to doubling the total PM2.5 emissions over Poland due to the importance of residential wood combustion. Note that in this case, the CED inventory likely performs better than the ensemble, highlighting the fact that ensemble estimates are not necessarily more accurate. Despite this, inconsistencies are flagged and paths for improvements are identified.

Relatively less important but yet about a factor between 2 and 5, low values occur for SO_2 emissions from power-generation sector (blue rectangles, Figure 4). As none of the three Europe-wide inventory shows an inconsistency for this sector/pollutant, this indicates a general issue between local and EU-wide inventories. This might be explained by the fact that CED is solely based on NB, supplied directly with users' data, while Europe wide inventories (EMEP) likely include additional emissions as they are based on overall fuel sales. In addition, point source emissions from E-PRTR may be different from point source emissions used in national inventories.

The transport and industry sectors show the lowest number of inconsistencies, which is observed by few points related to those sectors in the diagram (Figure 4 left). While this is expected for transport which is a diffuse source, this is surprising for the industry as this sector was the main source of inconsistencies at Europe wide level (see Figure 3).

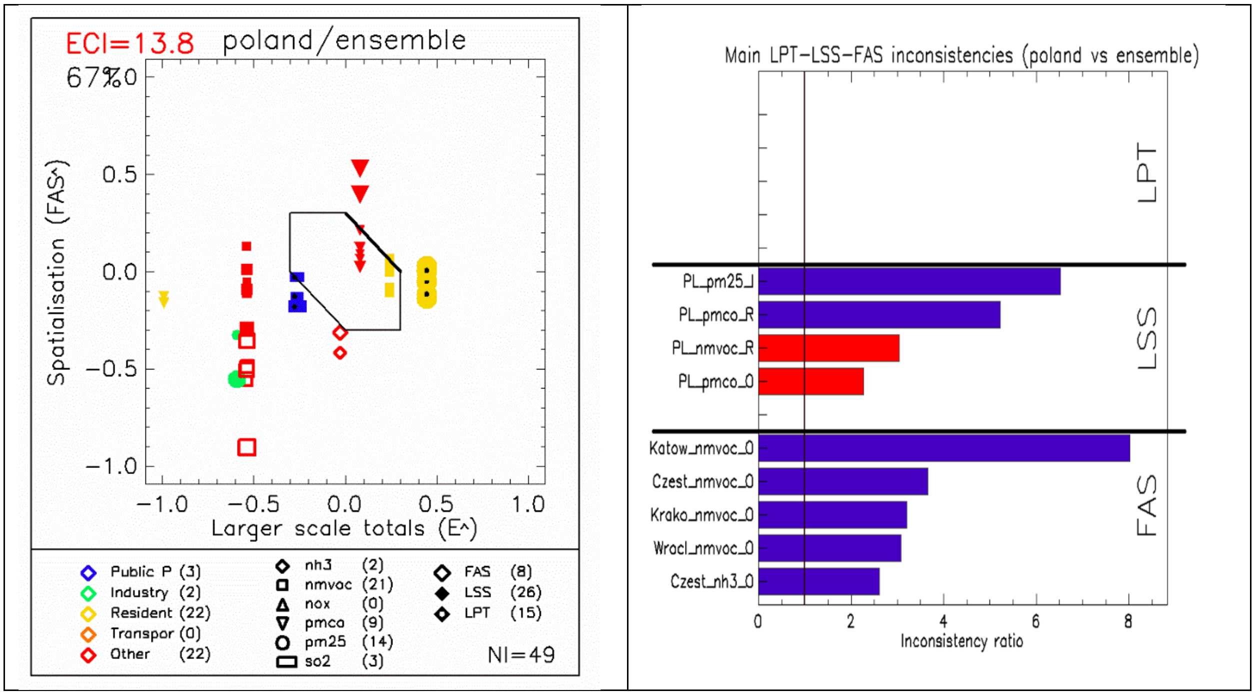

Figure 4: Diamond comparison of the local Polish vs ensemble inventory (left) and comparison of the ensemble top-down members vs the ensemble restricted to the Polish territory.

Figure 4 (right) highlights the priorities for the analysis. At country scale, the largest inconsistency occurs for the industrial share of PM2.5 (factor 6 larger in the Polish inventory, LSS, Figure 4), for PMco and NMVOC from the residential sector by a factor 5 lower and 3 larger in the Polish inventory, respectively, as well as for PMco from the other sector (factor 3 lower in the Polish inventory). In the case of PM2.5, the difference can be explained by the fact that the reports provided to NB are based on user-specific permits which specify the list of pollutants to be reported whereas in EU wide inventories, emissions are generally calculated using official EMEP/EEA emission factors. A comparison of EMEP and CED country totals per pollutant and GNRF sector is available in Table 2 of the supplementary material.

At the local scale (Figure 5), the spatial allocation of NMVOC emissions for the other sector leads to important differences in cities like Katowice (factor 8, Figure 4 – right), Czestochowa and Krakow. A similar situation is found for PM in Kielce. We see from Figure 4 that this issue occurs for many cities in the southern part of Poland. The large differences spotted in some cities (e.g. Kielce) are likely caused by emissions from heaps and excavations. While in CED, emissions from these sources are accounted for, only emissions from brown coal excavations (part of NFR 1B1a) are included in the EMEP inventory. Hence, including all heap and excavations emissions in EU-wide inventories would be advisable.

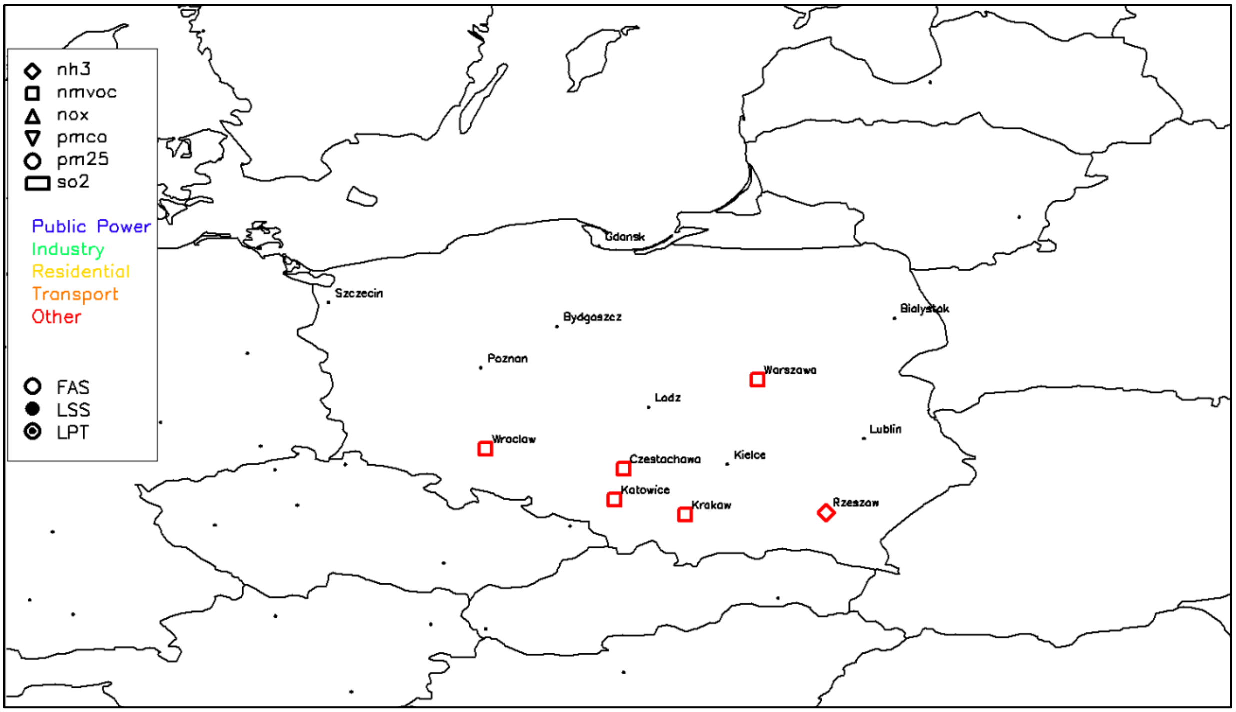

Figure 5: overview of inconsistencies for the comparison between local emission inventory in Poland and the Europe wide emission inventory ensemble

In conclusion, the comparison of the Polish inventory with the ensemble mostly spots issues that are related to a difference in terms of sectorial share at country level, explained by the accounting of different sources in the two types of inventories. A similar argumentation can explain part of the large discrepancies observed in some cities. Most of the issues occur for the residential and other sectors and mostly for PM and NMVOC. Although the number of inconsistencies may seem large, many of these are similar for all cities.
Inconsistencies in the spatial distribution of the emissions are relatively minor. This is due to the fact that EMEP reports for Poland, used in two out of three EU-wide inventories in the ensemble, are gridded by Polish experts, utilizing spatial proxies based on CED activity data for several sectors like stationary combustion, road transport and livestock (last updated in 2021, Bebkiewicz et al. 2022).

5. Added value and limitations of the ensemble approach

European wide inventories are not totally independent of each other. Interlinkages between the CAMS-REG, EDGAR and EMEP inventories exist. For example, the link between EMEP and CAMS-REG is that (1) both inventories rely on country reported data and may use the same spatial proxies when the country does not report. EMEP is also linked to EDGAR as it uses in some cases EDGAR distribution as a proxy for gridding in case a country is not reporting

(CEIP2022). Consequently, these interlinkages hide some of the inconsistencies, when all inventories behave similarly. It is however expected that repeated screenings lead to improvements and to a progressive convergence among inventories, hence reducing the number of flagged inconsistencies.

In our work, the number of members of the ensemble is limited to three. This would be an issue if the goal were to obtain more accurate and robust results with the ensemble. In such a case, the more members, the more robust the results of the ensemble. Our goal is however different and consists in creating a benchmark for comparison. Rather than looking at absolute values, we assess differences (between an inventory and the ensemble), for which the accuracy and robustness of the absolute values is of secondary importance.

As emission inventories are characterized by different grid resolution and sector aggregations, harmonization is required prior to the screening process for a meaningful comparison. Conversion to a common grid resolution might result in point sources shifted by one grid cell and be in the urban area in one inventory and not in another, although having the same geographical coordinates in both inventories. However, city specific diamond diagrams can be used to check if this issue occurs.

While it is more effective for inventory teams to meet and compare approaches in detail to understand and correct differences between inventories, this can be challenging at times, especially in the absence of a specific project to support the work. It must however be noted, that in many instances the reporting of an inconsistency, especially when it is very large, leads to a generally straightforward identification of the underlying cause without requiring too detailed information regarding the inventories.

The settings used in this work, e.g. the choice of 150 urban areas or the way sectors are aggregated are arbitrarily fixed. The method allows for flexible choices and could be applied to other areas than urban (e.g. complex industrial areas or intensive agriculture land) to assess the consistency with respect to other types of emissions. In terms of sectors, a further disaggregation of the other sector will be performed in future to better understand where inconsistencies originate from.

6. Conclusions

The approach presented in this work supports the screening and flagging of inconsistencies among inventories, through the construction of an ensemble benchmark. This ensemble is created to monitor the status and progress made with the development of Europe-wide inventories, but also to facilitate the comparison among inventories in a relatively simple manner.

The analysis of the EU-wide ensemble and the comparison with its individual members highlighted a large number of inconsistencies. While two out of the three inventories constituting the ensemble behave more closely to each other (CAMS-REG and EMEP), they yet show inconsistencies in terms of the spatial distribution of emissions. The origin of some differences between these inventories and EDGAR can be identified but their magnitude remains to be explained. These differences mostly occur for SO_2, PM and NMVOC, for the industrial and residential sectors, and reach a factor of 10 in some instances. The results of the screening provided useful information that allowed identifying necessary improvements on the estimation of air pollutants emissions, in particular for EDGAR, with the PM emissions from the small-scale combustion sector and SO_2 from the industry and power plant sectors. Spatial inconsistencies mostly occur for the industry and other sectors. The fact that the largest inconsistencies are found for sectors where point sources play a major role was expected. Indeed, while a diffuse sector like transport may be distributed quite differently, outliers would not appear as strongly as for point sources.

The application of the ensemble-screening approach to the local inventory for Poland leads to identifying another type of inconsistencies. While we would intuitively expect differences between local and European-wide inventories to be driven mainly by the spatial distribution of the emissions, this is not always the case in our analysis. Inconsistencies indeed relate mostly to differences in country sectorial shares that result from different sectors/activities being accounted for in the two types of inventories. This can be explained by the fact that some emission sources are omitted in the local inventory due to lack of appropriate geographically allocated activity data. We identified sectors and pollutants for which discussion between local and EU-wide emission compilers would be needed in order to reduce the magnitude of the observed differences (e.g. in the residential and industrial sectors mostly for NMVOC, PM2.5 and PM10).

It is also interesting to note that the comparison at local and European-wide scale lead to different types of inconsistencies. While the comparison to one local inventory is presented in this work as an example, these comparisons can be systematized to improve the quality of the ensemble.

The ensemble is not meant to be a static entity. It will evolve as inconsistencies are progressively discussed and solved and emission inventories get improved. The ensemble is therefore associated with reference inventory versions as well as with a reference year. In this sense the ensemble represents a useful tool to motivate the community around a single common benchmark and monitor progress towards the improvement of regional and locally developed emission inventories. It also ensures that improvements become permanent, as forgotten improvements would indeed be flagged again by the system.

Table of abbreviations

CAMS-REG	Copernicus Atmospheric Monitoring Services - Regional
CED	Central Emission Database
CEIP	Centre of Emission Inventory and Projection
CLRTAP	Convention on Long-range Transboundary Air Pollution
CO	Carbon Oxides
ECI	Emission Consistency Indicator
EEA	European Environment Agency
E-PTR	European Pollutant Release and Transfer Register
EU	European Union
FAS	Focus Area Share
FUA	Functional Urban Area
GHG	GreenHouse Gases
GNFR	Gridded Nomenclature For Reporting
GPS	Global Positioning System
IIR	Informative Inventory Report
IPCC – AR6	Intergovernmental Panel on Climate Change - Sixth Assessment Report
LPT	Large-scale Pollutant totals
LSS	Large-scale Sectorial Share
NMVOC	Non-Methane Volatile Organic Carbons
NFR	Nomenclature For Reporting
NH3	Ammonia
NOX	Nitrogen Oxides
OECD	Organisation for Economic Co-operation and Development
NB	National dataBase
PM	Particulate matter
PM2.5	Particulate matter with diameter less than 2.5 μm
PM10	Particulate matter with diameter less than 10 μm
SNAP	Selected Nomenclature for Air Pollution
SO2	Sulfur Oxides
UNECE	United Nations Economic Commission for Europe
UNEP	United Nations Environment Program

Code and data availability.
Supporting data and source code are available at: “Philippe Thunis. (2023). Supporting data for the publication "Emission ensemble approach to improve the development of multi-scale emission inventories" [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7940402”
.

Author contributions.
PT and AC contributed to the study conception and design. Material preparation, data collection and analysis were performed by PT, EP, ADM, JK, MB, LG, KS, and AC. All authors reviewed the manuscript. All authors read and approved the final manuscript.

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

ReferencesBebkiewicz, K., Boryń, E., Chłopek, Z., Chojacka, K., Kanafa, M., Kargulewicz, I., Rutkowski, J., Zasina, D., Zimakowska-Laskowska, M., Żaczek, M., Waśniewska, S.: Poland's Informative Inventory Report 2022. Institute of Environmental Protection – National Research Institute. KOBiZE. https://cdr.eionet.europa.eu/pl/un/clrtap/iir/envyi8lmq/IIR_2022_Poland.pdf (accessed on 09/12/2022).

Brasseur, G. P., Xie, Y., Petersen, A. K., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., Timmermans, R., van der A, R., Walters, S., Xu, J., and Zhou, G.: Ensemble forecasts of air quality in eastern China – Part 1: Model description and implementation of the MarcoPolo–Panda prediction system, version 1, Geosci. Model Dev., 12, 33–67, https://doi.org/10.5194/gmd-12-33-2019, 2019.

ChIEP 2021: Air Quality Assessment for the Year 2020: Model Evaluation. Chief Inspectorate for Environmental Protection. https://powietrze.gios.gov.pl/pjp/publications/card/34101 (accessed on 8/12/2022).

CEIP2022: Methodologies applied to the CEIP GNFR gap-filling 2022, Part I: Main Pollutants (NOx, NMVOCs, SOx, NH3, CO), Particulate Matter (PM2.5, PM10, PMcoarse) and Black Carbon (BC) for the years 1990 to 2020. Technical report CEIP 01/2022. https://www.ceip.at/ceip-reports (accessed on 05/05/2023).

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., van Aardenne, J. A., Monni, S., Doering, U., Olivier, J. G. J., Pagliari, V., & Janssens-Maenhout, G.: Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth System Science Data*, *10*(4), 1987–2013. https://doi.org/10.5194/essd-10-1987-2018, 2018.

Crippa, M., Guizzardi, D., Pisoni, E., Solazzo, E., Guion, A., Muntean, M., Florczyk, A., Schiavina, M., Melchiorri, M. and Fuentes Hutfilte, A.: Global anthropogenic emissions in urban areas: patterns, trends, and challenges. Environmental Research Letters, 16(7), 074033, http://dx.doi.org/10.1088/1748-9326/ac00e2, 2021.

Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., Schieberle, C., Friedrich, R. and Janssens-Maenhout, G. : High resolution temporal profiles in the Emissions Database for Global Atmospheric Research. 7(1), 1-17, 2020.

Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Monforti-Ferrario, F., Banja, M., Pagani, F., Solazzo and E.: EDGAR v6.1 Global Air Pollutant Emissions. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/df521e05-6a3b-461c-965a-b703fb62313e, 2022.

de Meij, A., Cuvelier, C., Thunis, P., Pisoni, E., and Bessagnet, B.: Sensitivity of air quality model responses to emission changes: comparison of results based on four EU inventories through FAIRMODE benchmarking methodology, Geosci. Model Dev., 17, 587–606, https://doi.org/10.5194/gmd-17-587-2024, 2024..

EMEP/EEA 2019. EMEP/EEA air pollutant emission inventory guidebook 2019. EEA Report No 13/201. https://www.eea.europa.eu/publications/emep-eea-guidebook-2019 (accessed on 24/05/2023)

EPTR2022, Industrial Reporting under the Industrial Emissions Directive 2010/75/EU and European Pollutant Release and Transfer Register Regulation (EC) No 166/2006. https://www.eea.europa.eu/data-and-maps/data/industrial-reporting-under-the-industrial-6. (Accessed on 5 January 2023)

Gawuc, L., Szymankiewicz, K., Kawicka, D., Mielczarek, E., Marek, K., Soliwoda, M., Maciejewska, J.:. Bottom–Up Inventory of Residential Combustion Emissions in Poland for National Air Quality Modelling: Current Status and Perspectives. Atmosphere 12, 1460. https://doi.org/10.3390/atmos12111460, 2021.

IIR 2022, Swedish Environmental Protection Agency Report 2022, Informative Inventory Report Sweden 2022. https://www.naturvardsverket.se/490927/contentassets/650c7f0c1e3446369baf84934c59873c/informative-inventory-report-sweden-2022.pdf (accessed on 26/04/2023)

Kryza, M., Józwicka, M., Dore, A. J., and Werner, M.: The uncertainty in modelled air concentrations of NOx due to choice of emission inventory, Int. J. Environ. Pollut, 57, 3–4, 2015.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K. and Wulfmeyer, V.: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, *7*(4), 1297–1333. https://doi.org/10.5194/GMD-7-1297-2014, 2014.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmüller, R., Denier van der Gon, H. Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B. and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution. Atmos. Chem. Phys., 15(19), 11411-11432, https://acp.copernicus.org/articles/15/11411/2015/, 2015.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., et al. (2019). EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012. Earth Syst. Sci. Data, 11(3), 959-1002. https://essd.copernicus.org/articles/11/959/2019/

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I. and Denier van der Gon, H.: CAMS-REG-v4: a state-of-the-art high resolution European emission inventory for air quality modelling. *Earth System Science Data*, 14, 491-515, https://doi.org/10.5194/essd-14-491-2022, 2022

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., Denier van der Gon, H.A.C., 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. Mareckova K., Pinterits, M., Ullrich, B., Wankmueller, R., Mandl, N.: Review of emission data reported under the LRTAP Convention and NEC Directive Centre Emission Inventories Project., 2, p. 52, 10.1029/2009JD011823, 2017.

Markakis, K., Valari, M., Perrussel, O., Sanchez, O. and Honore, C.: Climate-forced air-quality modeling at the urban scale: sensitivity to model resolution, emissions and meteorology, Atmos. Chem. Phys., 15, 7703–7723, https://doi.org/10.5194/acp15-7703-2015, 2015.

Thomas M., Timmermans, R., Valdebenito, A., van Velthoven, P., van Versendaal, R., Vira, J., Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev. (GMD), 8 (9) , pp. 2777-2813, 2015.

NFR-I, 2023. Annex I NFR reporting template, https://www.ceip.at/reporting-instructions/annexes-to-the-2023-reporting-guidelines (last accessed on 24/04/2023)

Oreggioni, G. D., Mahiques, O., Monforti-Ferrario, F., Schaaf, E., Muntean, M., Guizzardi, D., Vignati, E. and Crippa, M. : The impacts of technological changes and regulatory frameworks on global air pollutant emissions from the energy industry and road transport. Energy Policy, 168, 113021, https://www.sciencedirect.com/science/article/pii/S0301421522002464, 2022.

Riccio, A., Giunta, G. and Galmarini, S.: Seeking for the rational basis of the Median Model: The optimal combination of multi-model ensemble results, Atmos. Chem. Phys., 7, 6085–6098, 2007.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., Noije, T. P. C. van, Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J.-F., Lawrence, M. G., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K. and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone. *Journal of Geophysical Research: Atmospheres*, *111*(D8), 8301, https://doi.org/10.1029/2005JD006338, 2006.

Thunis, P., Clappier, A., Pisoni, E., Bessagnet, B., Kuenen, J., Guevara, M., and Lopez-Aparicio, S.: A multi-pollutant and multi-sectorial approach to screening the consistency of emission inventories, Geosci. Model Dev., 15, 5271–5286, https://doi.org/10.5194/gmd-15-5271-2022, 2022

Thunis, P., Degraeuwe, B., Pisoni, E., Trombetti, M., Peduzzi, E., Belis, C. A., Wilson, J., Clappier, A. and Vignati, E. (2018). PM2.5 source allocation in European cities: A SHERPA modelling study. *Atmospheric Environment*, *187*, 93–106. https://doi.org/10.1016/J.ATMOSENV.2018.05.062

Trombetti, M., Thunis, P., Bessagnet, B., Clappier, A., Couvidat, F., Guevara, M., Kuenen, J., and López-Aparicio, S.: Spatial intercomparison of Top-down emission inventories in European urban areas, Atmos. Environ., 173, 142–156, 2018.

UNEP2023, United Nations Environment Programme. Emissions Gap Report 2023: Broken record – Temperature hit new highs, yet world fails to cut emissions (again). Nairobi. https://doi.org/10.59117/20.500.11822/43922, 2023.

Vautard, R., Schaap, M., Bergström, R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J.H., Cuvelier, C., Foltescu, V., Graff, A., Kerschbaumer, A., Krol, M., Roberts, P., Rouïl, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E. and Wind, P.: Skill and uncertainty of a regional air quality model ensemble, Atmospheric Environment, Volume 43, Issue 31, 4822-4832, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2008.09.083, 2009.

Zhang, W., Trail, M. A., Hu, Y., Nenes, A., and Russell, A. G.: Use of high-order sensitivity analysis and reduced-form modeling to quantify uncertainty in particulate matter simulations in the presence of uncertain emissions rates: A case study in Houston, Atmos. Environ., 122, 103–113, 2015