https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

Open Access Earth System Science Data Discussions

Global Covenant of Mayors, a dataset of GHG emissions for 6,200 cities in Europe and the Southern Mediterranean

Albana Kona[1], Paolo Bertoldi[1], Fabio Monforti-Ferrario[1], Marta Giulia Baldi[1], Eleonora Lo Vullo[1], Georgia Kakoulaki[1], Nadja Vetters[1], Christian Thiel[1], Giulia Melica[1], Alessandra Sgobbi[2], Christofer Ahlgren[2], Brieuc Posnic[3]

[1]European Commission – Joint Research Centre - Ispra
[2]European Commission – Directorate General for Climate Action
[3]European Commission – Directorate General for Energy

Correspondence to: Fabio.MONFORTI-FERRARIO@ec.europa.eu

Abstract. The Paris Agreement has underlined the role of cities in combating climate change. The Global Covenant of Mayors for Climate and Energy (GCoM) is the largest dedicated international initiative to promote climate action at city level, covering globally over 10,000 cities and almost half the population of the European Union (EU) by end of March 2020. The latest Intergovernmental Panel on Climate Change (IPCC) report denotes that there is a lack of comprehensive, consistent datasets of cities' Greenhouse Gas (GHG) emissions inventories. To address part of this gap, we present here a harmonised, complete and verified dataset of GHG inventories for 6,200 cities in European and Southern Mediterranean countries, signatories of the GCoM initiative. To complement the emission data reported, a set of ancillary data that have a direct or indirect potential impact on cities' climate action plans were collected from other databases, supporting further research on local climate action and monitoring the EU's progress on Sustainable Development Goal 13 on Climate Action. The dataset is archived and publicly available with the DOI number https://doi.org/10.2905/57A615EB-CFBC-435A-A8C5-553BD40F76C9

1. Background & Summary

Cities consume over two-thirds of the world's energy and generate about 70% of global GHG emissions (IPCC, 2014), and are at the same time particularly vulnerable to the impacts of climate change (Reckien et al., 2018). An increasing number of cities have voluntarily adhered to transnational networks active in climate action (Busch et al., 2018; Heidrich et al., 2016; van der Ven et al., 2017). As these networks and initiatives have evolved, cities' ambition and climate targets have increased to match or even go beyond the ambition of countries (Bertoldi et al., 2018d)

However, the scientific community denotes the current lack of systemic knowledge of cities' quantified contribution to combating climate change (Acuto et al., 2018; IPCC, 2015). This knowledge gap originates from many issues, including dissimilarities in the methodologies used for developing local emission accountings and reference scenarios and for setting ambition targets, as well as the absence of a global, open and harmonised dataset of cities' emissions inventories (Kona et al., 2018). Only as

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

recently as in 2019, the first datasets were published in academic literature, aiming to fill regional gaps (Adami et al., 2020; Kilkis, 2019; Palermo et al., 2020).

The dataset presented in this paper aims to fill these gaps in Europe and Southern Mediterranean countries. It consists of a harmonised, comprehensive and verified dataset of GHG emissions based on data produced by 6,200 cities in the EU, European Free Trade Association (EFTA) countries and UK, Western Balkans, Eastern and Southern EU neighbourhoods. The dataset is extremely valuable for communities engaged in climate action, including policy makers at all levels pursuing informed decisions. Along with the dataset, we provide the method of producing the data, their metadata, as well as the technical validation performed.

The European Commission launched the Covenant of Mayors (CoM) in 2008 to endorse and support the effort of EU local authorities in mitigating climate change. In 2015, the Covenant expanded to also include climate adaptation. In 2011, the initiative was launched in the EU's Eastern and Southern neighbourhood , and in 2016 the initiative became global, through the launch of the Global Covenant of Mayors for Climate and Energy (GCoM). The initiative registered a very rapid growth from 241 signatories in 2008 in the EU to more than 10,000 covering more than 869 million inhabitants worldwide as of March 2020.

The Joint Research Centre (JRC), the science and knowledge service of the European Commission, provides scientific and technical support to GCoM cities in the development and implementation of their climate action plans. The scientific support is given through guidance on methodologies for emission accounting and climate adaptation, as well as through the development of urban policy tools for climate action (Bertoldi et al., 2018a, 2018b; Kovac et al., 2020; Monforti-Ferrario et al., 2018; Peduzzi et al., 2020), While the technical support consists of checking and validating the data reported by cities in the MyCovenant platform (www.covenantofmayors.eu).

The published dataset (Kona et al., 2020) contains verified reported GHG emissions for 6,200 European and South Mediterranean cities for a set of reference years. Given the voluntary nature of the CoM and the difficulty of local authorities to report using a harmonised framework, a statistical method for checking the reliability, cleaning and validating the reported data was developed and applied. The method allows building a coherent dataset and consists of four steps:

- Data reporting principles, extraction and clustering: accounting principles of GHG reporting framework, data extraction and clustering of signatories into two groups (large /small areas) based on population size and degree of urbanization (threshold 50,000 inhabitants);
- Data coherence and completeness: data reported in the platform are checked in terms of coherence and completeness with the official Sustainable Energy and Climate Action Plan (SECAP) document in large urban areas;
- Detection of outliers: statistical method applied for the identification and correction of outliers in small urban areas in small medium towns;
- Matching emission data with ancillary data: signatories from the EU are matched with their respective administrative units in the EU official statistics for cities. Harmonized statistical information on signatories allows building a referenced structure for collecting, processing, storing, analysing and aggregating data to support the monitoring of the EU progresses on the Sustainable Development Goal 13 on Climate Action (Eurostat, 2020).

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

The dataset (Figure 1) thus contains adjusted self-reported data from cities (i.e. CoM dataset 2019: Emission Inventories) coupled with ancillary data (CoM dataset 2019: Ancillary data) related to geographic attributes (area, latitude and longitude, local administrative codes, heating degree-days), socio-economic aspects (GDP per capita) and demographic characteristics at city level (degree of urbanisation, population time series). A detailed technical evaluation at the city level was also performed against the independent estimates provided by the Emissions Database for Global Atmospheric Research (EDGAR: provides time series of global anthropogenic emissions of greenhouse gases and air pollutants by country on a spatial grid (Crippa et al., 2020)).

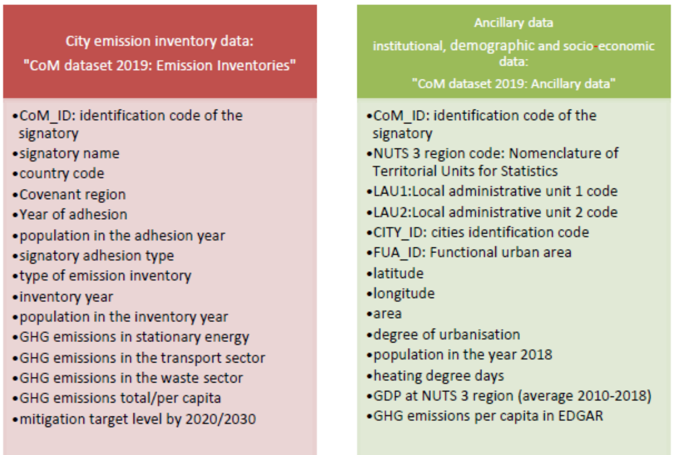

Figure 1. Datasets and their attributes. Source of the "CoM dataset 2019: Emission Inventories" MyCovenant of Mayors platform, source of "CoM dataset 2019: Ancillary data" Eurostat and JRC

In compliance with the EU data policy, we are now in a position to share with the community a ten years' dataset complete, cleaned, validated and harmonised with the EU statistical database on local authorities. Given the overall good quality of the dataset, some limitations and uncertainties remain and are described in the "Limitation and future work" section.

The resulting dataset is of great value and interest and responds to the clear needs expressed by the scientific and academic community and governmental institutions. Last but not least, the peculiarity of the CoM initiative is the participation of small towns interested and engaged in climate action, often absent from other initiatives. This dataset therefore offers cities of all sizes a mean for comparative analysis of the magnitude, efficiency and intensity of energy use and GHG emissions, which can serve as a catalyst for learning and collaboration among cities.

2. Methods

Hereafter we describe the methods used to produce and consolidate the final dataset. Due to local authorities' difficulties in harnessing and reporting data within a harmonised framework, which may

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

differ from the national emission reporting, not all the self-reported data could be considered reliable. Therefore, a method was developed to construct a robust dataset of emission inventories, organised into four steps:

- Step 1: Data reporting principles, extraction and clustering: accounting principles of GHG reporting framework, data extraction and clustering of signatories into two groups (large/small areas) based on degree of urbanization and/or population size (threshold 50,000 inhabitants);
- Step 2: Data coherence and completeness: Digital curation of data reported in the platform were performed in terms of completeness and coherence with the official climate action plan document, the so-called SECAP in large urban areas;
- Step 3: Detection of outliers: statistical method for the identification and detection of outliers in the GHG emission dataset in small medium towns;
- Step 4: Matching emission data with ancillary data: signatories from the EU are matched with their respective local administrative units of the Geographic Information System of the European Commission.

2.1 Data principles, extraction and clustering

The GCoM has two officially recognised platforms for data reporting, the CDP-ICLEI Unified Reporting System and the My Covenant platform. To streamline measurement and reporting procedures, a Common Reporting Framework (CRF) was developed during 2018 in consultation with partners and signatories of the GCoM. The dataset provided in the current study is based on the information reported by signatories through the MyCovenant platform, following the JRC methodology (Bertoldi et al., 2018a). Hereafter we report a brief description of the data collected on MyCovenant in alignment with the CRF.

The protocols for accounting the cities' emissions differ mainly in the principles and minimum reporting requirements on sources, the type of gases and the boundary of the inventory to be reported. The protocol for accounting the emissions used by Covenant signatories is closely aligned with the IPCC guidelines regarding the source category of the in-boundary emissions. It includes "sources" and "activities" rather than the scope framework used in other city protocols. Nevertheless, the emission inventory is not meant to be an exhaustive inventory of all emission sources in the territory. It focuses mainly on GHG emissions related to sectors (stationary energy, transport and waste/wastewater) upon which the local authority could intervene through sectoral measures and urban policies. They also can report GHG emissions from Industrial Processes and Product Use (IPPU) and Agriculture, Forestry and Other Land Use (AFOLU) sectors where these are significant (Table 1).

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

Open Access Earth System Science Data Discussions

Table 1. Mapping of emission source categories in GCoM reporting framework with IPCC guidance

Sectors and subsectors in GCoM reporting framework		IPCC (ref no.)	Description
Stationary energy	Residential buildings	1A4b; 1A1	All activities and related GHG emissions (direct emission from fuel combustion and indirect emission due to consumption of grid-supplied energy) occurring in stationary sources within the local authority boundary are reported. GHG emissions from sources covered by a regional or national emissions trading scheme (ETS), or similar (i.e. industries with thermal energy in input below or equal to 20 MW) when ETS does not exists, are not accounted in the inventory. In addition, "energy generation" industries/facilities are not reported under this sector to avoid double counting with indirect emissions.
	Commercial building and facilities	1A4a; 1A1	
	Institutional buildings and facilities	1A4a; 1A1	
	Manufacturing, construction industries	1A1, 1A2; 1A1	
	Agriculture	1A4c; 1A1	
	Fugitive emissions	1B1, 1B2	
Transportation	On-road	1A3b; 1A1	All activities and related GHG emissions (direct emission from fuel combustion and indirect emission due to consumption of grid-supplied energy) occurring for transportation purposes within the local authority boundary will be reported. NB. The CoM dataset reported in this dataset does not reflect only the on road fraction of the transport sector, but all the emission in this sector.
	Rail	1A3c; 1A1	
	Waterborne navigation	1A3d.; 1A1	
	Aviation	1A3a;	
	Off-road	1A3e; 1A1	
Waste	Solid waste disposal	4A	Sources related to disposal and treatment of waste and wastewater generating emissions within the city boundary are reported under Waste sector. Where waste/wastewater is used for energy generation, emissions are not reported under this sector to avoid double counting of indirect emission.
	Biological treatment	4B	
	Incineration and open burning	4C	
	Wastewater	4D	

The geographical boundaries of the "local territory" are the administrative boundaries of the entity (municipality, region) governed by the local authority which is a signatory to the GCoM. Regarding the type of gases, GCoM signatories shall report emissions of carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) converted into CO_2-equivalents ($CO_{2\text{-eq.}}$), according to their global warming potential. The three main GHG emission categories included in the inventories are:

- Direct emissions due to final energy consumption, excluding those from industrial plants involved in Emissions Trading Systems.
- Indirect emissions related to grid supplied energy (electricity, heat, or cold) consumed in the local territory (Kona et al., 2019).

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

Open Access Earth System Science Data Discussions

- Non-energy related direct emissions (such as from waste, wastewater) that occur in the local territory, if the climate action plan contains measures to reduce such GHG emissions.

The GHG emissions are automatically derived in the platform as the product of activity data (detailing the energy consumption/waste per carrier/type) and emission factors, as reported by the signatories (Table 1). The emission factors are coefficients, which quantify the emissions per unit of activity, and one out of three approaches can be used:

- IPCC – emission factors for fuel combustion – default values mostly based on the carbon content of each fuel;
- National or subnational emission factors for fuel combustion when these are different from the IPCC's.
- Life Cycle Assessment (LCA) – emission factors for the overall life cycle of each energy carrier, i.e. including not only the GHG emissions due to fuel combustion but also emissions of the entire energy supply chain – exploitation, transport and processing.

The procedure to verify and improve the coherence of the dataset starts with the extraction of complete emission inventories stored in a PostgreSQL database. At the closing date of this study, (September 2019) 6,239 climate action plans with complete inventories had been submitted by cities in the EU27, EFTA countries and UK, Western Balkans, Eastern and Southern EU neighbourhoods.

Inventories and other data are self-reported to the online platform and must accurately reflect the content of the official climate action plan (SECAP) document. The SECAP document is a separated file, usually in PDF format and publicly available that represents the official action plan endorsed and signed by the local council. As a first step to address the quality of the data reported, yearly GHG emission per capita are plotted for each signatory.
The first step to understand the degree to which this is true and the quality of the data reported, yearly GHG emission per capita are plotted for each signatory. Figure 2 shows the frequency distribution of GHG emissions per capita in emission inventories dataset in 2019, with observations that range from 0 to 80 tCO_2-eq/cap, with a mean of 5.18 tCO_2-eq/cap. In the vertical axis the density values are reported, i.e. the share of signatories with the same range of GHG emission per capita (i.e. emission from all the CoM sectors, excluding manufacturing and construction industries), by the width of the class (0.01 in this case).

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

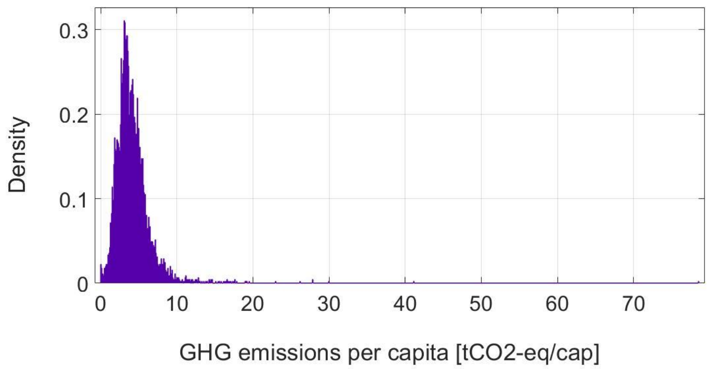

Figure 2 GHG emissions per capita of signatories in the "CoM dataset 2019". (Emission from all the CoM sectors, excluding manufacturing and construction industries)

The occurrence of outliers is a clear indication of errors in the data, likely due to inputting error, therefore not all the data collected in the platform are consistent with the SECAP document. As the calculations of performance indicators for the dataset, such as the mean and standard, can be distorted by a single grossly inaccurate data point, checking and treating outliers is a routine part of data analysis.

Due to the high volume of information, it is not feasible to check individually the consistency of all the data objects with the SECAP document. The collection of the attributes (i.e. the variables: 15 energy carriers and 16 subsectors) describes the data objects (it is also known as record, point, case, sample, entity, or instance), which visually corresponds to the rows in the excel files.

The original dataset comprises 6,239 signatories with a baseline inventory, out of which 1,845 with an additional monitoring inventory. In each inventory the cities report at maximum data for 15 energy carriers grouped into 16 subsectors, resulting therefore into 1.94 million data objects. The 16 subsectors have been grouped into 6 sectors (i.e. municipal, residential, tertiary, manufacturing and construction industries, transportation and waste sector) and null objects were deleted, leading in total to 61,207 data objects.

We therefore adopted a rule to treat the outliers, based on the benefits expected when scrutinizing the dataset for the overall assessment of the initiative. The overall assessment consists of producing performance indicators on the impact and the contribution of climate actions planned and implemented by CoM signatories. In this context, it is evident that, the bigger a city is, the more impact any errors will have on the overall dataset. In order to have an accurate representation, it is then of utmost importance that large cities have highly accurate data.

Hence, we decided to adopt a customized method to treat outliers based on the signatories' degree of urbanization and population size (source https://ec.europa.eu/eurostat/web/nuts/local-administrative-units). The 6,239 signatories and their data were clustered into two groups:

- Large urban areas (densely populated area with a population density of at least 1,500 inhabitants per km^2 and a minimum population of 50,000): for this group manual curation of imputed errors in inventories was implemented, which significantly increased the performance indicators of the database by increasing their robustness (described in step 2).

- Small towns and rural areas (intermediate and thinly populated areas): for this group an automatic routine to identify and remove the outliers is applied. The rules governing the automatic detection and treatments of the outliers are detailed in Step 3.

2.2 Data coherence and completeness

In this section, we describe the steps followed to detect and treat the outliers in inventories from large urban areas (i.e. cities and greater cities, with a population density of at least 1,500 inhabitants per km^2 and a minimum population of 50,000) along with correctness and completeness checks in the overall dataset. The identification and treatment of outliers in this group of cities has been performed qualitatively.

Because of the harmonization process of CoM administrative data and local administrative units Eurostat database, 430 signatories covering 116.2 million inhabitants, classifies as cities and greater cities. In addition, in the other regions out of the EU and EFTA regions (i.e. Eastern Europe; Western Balkans and Southern Mediterranean) where the classification was not available, we adopted as criteria only the population size as threshold (i.e. a minimum population of 50,000). Hence, within the CoM 2019 dataset there are 701 baseline inventories presented by large urban areas, covering a total population of 165.26 million inhabitants.

As part of the evaluation process carried out by JRC on individual SECAPs, activity data were compared against the national/ EU averages (available at national/EU statistical systems). In case of reported data that ranged out of one or more units higher than the average of the sectors national average, we double checked the accuracy of the data reported on the platform with the SECAP document. As a result of the digital curation of outliers, identified through the comparison of self – reported data in the MyCovenant platform against the same data declared in the SECAP, twenty inventories (i.e., about 3%) have been corrected manually. The SECAP document represents the official action plan endorsed and signed by the local council; therefore here the assumption is that the data reported within the SECAP are the valid ones. The errors were often due to the misinterpretation of the unit measure to be reported in the online template (e.g. kWh/year instead of MWh/year, etc.).

At this point in the procedure, with the help of the statistical routine and the digital curation, we have consolidated the dataset related to activity data. The next step consists of comparing the emission factors used in CoM inventories against the reference values from IPCC (4th Assessment Report) and the JRC databases(Lo Vullo et al., 2020) and their completeness (i.e. missing data on emissions were derived from reported activity data and vice versa). In case of reported emission factors that ranged out of ± 50% of the reference value, we corrected them with the corresponding reference value. As a result of this procedure, there were 153 inventories from large urban areas, where 9.7% (i.e. 526 out of 5433 objects) of the data objects were corrected.

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

2.3 Detection of outliers

In this section we describe the automatic routine implemented to detect and treat the outliers in inventories from small medium towns (number of inventories = 5,538 covering a total population of 46.78 million inhabitants).

Urban GHG emissions per capita may deviate significantly from national averages, due to the tendency of emissions to concentrate around human activities. Therefore, setting exclusion ranges of outliers in the per capita GHG emissions based on the national averages may lead to the exclusion of a high number of valid emission inventories from the CoM dataset. To avoid this bias, we apply a statistical method based on intrinsic properties of the distribution of the emissions in the CoM database. This allows identifying more accurately potentially unreliable emission inventories and the outliers likely to be the results of incorrect data entry.

The procedure starts with dividing the data into two groups based on the normalization process: the activity data in the residential/municipal/institutional/tertiary buildings and transport sector were normalised with the population size, whereas the activity data in manufacturing and construction industries were normalised with the GDP values. The majority of these industries are already governed by the cap and trade system (EU-Emission Trading scheme), therefore they are not recommended to be reported in the CoM platform, although exceptions exist. In addition, signatories that report manufacturing emissions are generally large urban areas (80% of the activity data within this sector is reported by cities and greater cities), which we have been already examined individually to check for outlying data.

The outliers identification method is based on a generalised ESD (extreme studentized deviate) procedure for the detection of abnormal energy consumptions. The ESD is commonly used in literature (Cerquitelli et al., 2019; Gant., 2013; Rosner, 1983; Seem., 2007), because of its excellent performance under a variety of conditions to detect one or more outliers in a dataset that follows an approximately normal distribution. The per capita activity data in the residential/municipal/institutional/tertiary buildings and transport sector follows approximately a normal distribution.

The procedure iteratively identifies the extreme values in the dataset and then selects to remove those observations which are higher than the extreme values with a confidence level of 95%. A detailed description of the routine is available at Supplementary File 1 and Supplementary File 2.

Applying this approach, 39 inventories were removed from the initial dataset (i.e. from initial 5,538 inventories). These signatories have been approached through the feedback report to check and correct the data in the online platform. The clean and robust dataset thus contains 5,499 inventories. As a result, the original inventory containing 6,239 entries was reduced to a clean dataset of 6,200 signatories (i.e. 99% of the original data), referred to hereafter as the “CoM dataset 2019: Emission Inventories”.

To conclude, also a non-parametric statistical procedure, i.e. the Median Absolute Deviation (MAD), has been applied to identify outliers in dataset that are non normal distributed. This method is more robust than the ESD, but less efficient, and its validity increases as data approach normal distribution. Similarly to the ESD, the choice of the critical value is motivated by the reasoning that if the observations other than outliers have an approximately normal distribution, it picks up as an outlier any observations more than about three standard deviations from the means. The results of the MAD procedure produce the same

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

Open Access Earth System Science Data Discussions

outliers as the ESD procedure; therefore, we argue that the assumption on the quasi normal distribution is correct.

The next step consists of verifying the emission factors used in the inventories, against the reference values from IPCC and the JRC databases (Lo Vullo et al., 2020), and their completeness (i.e. missing data on emission were derived from reported activity data and vice versa). In case of reported emission factors that ranged out of ± 50% of the reference value, we corrected them with the corresponding reference value. Because of this procedure, there were 3019 inventories from small towns, where 15 % (i.e. 8,008/52,496 records) of the data records were corrected. Table 2 compares the main descriptive parameters of the two datasets. The main difference can be noted in the skewness parameter. Both frequency distributions have a positive skewness, meaning that the right tail is longer and the mass of the distribution is concentrated on the left of the figure.

Table 2. Statistical parameters of the GHG emissions in the CoM datasets 2019

Parameters	All CoM dataset 2019	Clean CoM dataset 2019
Number of signatories with complete GHG inventories in the Baseline year	6,239	6,200
Total population in the baseline year [Million inhabitants]	216.61	216.25
Mean [tCO_2-eq/cap]	5.18	4.69
Median [tCO_2-eq/cap]	4.78	4.75
Standard deviation [tCO_2-eq/cap]	82.35	3.68
Skewness [tCO_2-eq/cap]	76.39	2.37

2.4 Matching emission data with ancillary data

GCoM signatories, when submitting their data in the MyCovenant platform, report the local authority name, the country and their centroids' coordinates. Through these three attributes, we have been able to digitally match the signatories with their corresponding local administrative units in the Geographic Information System of the European Commission (GISCO). Harmonized statistical information on signatories allows building a referenced structure for collecting, processing, storing, analyzing and aggregating data.

In this way, we can derive all ancillary data related to institutional, demographic and socio-economic dimensions:

- Institutional dimension: the CoM signatories are associated with their correspondent NUTS codes; the Local administrative units' codes; their Functional urban area and cities codes; the geographical coordinates; the area and shape files of their local administrative units;

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

- Demographic dimension: the CoM signatories are associated with the population data in 2018; the degree of urbanisation;
- Socio-economic and climate dimension: the CoM signatories are associated with the GDP at NUTS 3 level; and heating degree-days at NUTS 3 level.

The aim of the ancillary data is also to support the monitoring of the SDG 13 on climate action in an EU context, which focuses on climate mitigation, climate impacts and on initiatives that provide support to climate action, as the Global Covenant of Mayors. More broadly, the ancillary data could support further research on investigating drivers of climate action at city level and the development of urban policy design. In addition, we extracted the national values of GHG emissions per capita from EDGAR for the corresponding CoM activity sectors (Table 3).

Table 3. Mapping of emission source categories with IPCC categories

GCoM sector	**Sector code**	**Sector name**	**Inventory years**	**Country ISO code**
Stationary energy/indirect emissions	1.A.1.a	1.A.1.a - Public Electricity and Heat Production	1990-2018	52 ISO codes
Stationary energy/residential	1.A.4.b	1.A.4.b - Residential	1990-2018	52 ISO codes
Stationary energy/ Commercial/Institutional	1.A.4.a	1.A.4.a - Commercial/Institutional	1990-2018	52 ISO codes
Transportation	1.A.3.b	1.A.3.b - Road Transportation	1990-2018	52 ISO codes

3. Data availability

The dataset is archived and publicly available with the DOI number https://doi.org/10.2905/57A615EB-CFBC-435A-A8C5-553BD40F76C9 (Kona et al., 2020).

4. Comparison with other emission inventories

Matching reported emissions with cities' ancillary data enables an additional robustness check of the accuracy of inventories. The uncertainty of reported emissions is particularly difficult to estimate since non-formal uncertainty analysis is applied by cities on the activity data and the emission factors. Hence, given this limitation, we argue that the best practical way to assess the uncertainty is a detailed comparison against international dataset such as EDGAR. A similar approach has been applied to validate cities emission data in United States (Nangini et al., 2019).

EDGAR provides past and present global anthropogenic emissions of greenhouse gases and air pollutants by country on a spatial grid (Crippa et al., 2020). The methods used in EDGAR downscale the emissions from a national or subnational scale to finer scales using spatial proxies and present results in gridded maps. EDGAR combines several proxies ranging from population density to specific point source location maps for estimating emissions of different economic sectors. The potential use of EDGAR gridded data for the examination of emission in large sample of cities worldwide has been already noticed in literature (Marcotullio et al., 2014).

Using ArcGIS, we overlaid the LAU urban spatial boundaries onto the EDGAR emissions grids. We then used the built-in Spatial Zonal Statistics tool to estimate total emissions for each urban area for two source categories: energy in buildings (RCO) and road transportation. EDGAR includes emissions from a variety of sources at the aggregate level of at least 0.1° spatial resolution (representing about 10 x 10 km2 at the equator). Here we use the EDGAR global grids of estimated emissions in metric tons for the year 2005 for the most prevalent GHGs: carbon dioxide excluding short cycle organic carbon (i.e. CO2_excl_short-cycle_org_C). Emissions of CO2_excl_short-cycle_org_C include all fossil CO2 (such as fossil fuel combustion) and exclude all sources and sinks from land-use, land-use change and forestry (LULUCF) (Crippa et al., 2020). Overall, we compared data from 1945 signatories from EU27 + UK countries with EDGAR corresponding data on direct emissions in Energy in Buildings sector and road transportation (Online-only Table 3).

The correlation coefficient (R2) between the CoM direct emission in the residential/municipal/institutional building sector and their corresponding data in EDGAR was 0.92, whereas for road transportation data was 0.66. We also calculated the root-mean-square error (RSME) between the CoM dataset and EDGAR direct emissions in the two sectors: in the residential/municipal/institutional building sector RSMEBuildings = 0.08; and in the road transport sector RSMETransp = 0.11.

Regarding the direct emissions in the “Energy in building” sector, the low value of the uncertainty (8%) shows a good agreement between the two datasets. This might be attributed to the fact that, in this sector EDGAR assumes that each resident is responsible for the same amount of emissions. Whereas CoM collects direct observations (data collected from utilities mainly), that are good proxies of national of energy usage pattern and fuels deployed.

About the direct emissions in the transport sector, the slightly higher level of uncertainty (11%) might be attributed to the fact that emission in the Covenant framework covers only the urban fraction of the sector (i.e. in the EU urban mobility accounts for 40% of all CO_2 emissions of road transport). Moreover, the CoM dataset reported here does not reflect only the on road fraction of the transport sector, but all the emission in this sector. This is due to the old version of the reporting platform that collected data without distinguishing the modal share. Further on limitation is discussed in the next section. Overall, considering the completely different origin of EDGAR and CoM primary data the agreement has to be considered fully satisfactory

5. Conclusions and future work

Despite the data mining and verification process, few limitations and uncertainties remain in relation to the data quality. To start with, it is important to highlight the fact that the overall quality of the data reported in the platform depends mostly on the city’s capacity to gather and report into the harmonized framework of CoM. The JRC does not correct or adjust the reported data itself, but is the responsibility

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

of the signatory, on receipt of the feedback analysis from experts, to check and possible revise its data according to the climate action plan. Indeed, we have noted an increasing quality of data reported from cities since 2010, mainly thanks to this feedback-rechecking system.

Therefore, the aim of the approach adopted here is not to validate the data as such (they are collected and reported by the signatories), but to guarantee as far as possible the internal consistency and completeness of the data reported in the online platform with the climate action plan documents (i.e. the SECAP).

Second, there is a limited knowledge on the methods used by cities in determining the emissions especially within the transport sector. The aim of this technical validation is to compare the CoM dataset against international datasets such as EDGAR, being well aware of the fact that CoM reports direct observations, whereas EDGAR represents modeled values. In small medium urban areas we suppose that local authorities use the territorial approach based on the activity data collected. Indeed, there is a good match with EDGAR data, whereas in large urban areas, we note a significant deviation from EDGAR proxies. We argue that most probably this is due to the fact that CoM reports real data supplied from the transportation department, which are not caught in the modelling exercise of EDGAR. More precisely, EDGAR uses the average national fleet that could be quite different from a local one. Moreover, due to the uncertainty on the methodological differences for accounting the emissions, embedded in the nature of the sector, the emissions in this sector can differ widely between cities with similar patterns or sizes.

About waste, the mapping of emissions in this sector has only been added in the last revision of the reporting framework, therefore we expect more data in this sector to become available as cities integrate it in their inventories.

To conclude, a major source of uncertainty originates from the use of emission factors developed for the national or sometimes even international scale, especially for the electricity and waste sectors. On the contrary, deploying city level emission factors for electricity supplied through the grid, taking also into account local renewable energy production, would greatly increase the accuracy of the data.

Future work envisage the possibility of undertaking a comparable analysis also of the data reported by GCoM signatories through the CDP-ICLEI Unified Reporting System with a view to expanding the coverage of a harmonised, complete and verified dataset of GHG inventories at city level.

Code Availability: Most data handling in the methods and technical validation was done in MATLAB (available at Supplementary File 1) and Microsoft Excel (Supplementary File 2). Data and codes also available in GitHub: https://github.com/PattoScripts/ComScripts

Acknowledgements: We are grateful to local authorities who make public their engagement in climate action planning, through their participation in the Global Covenant of Mayors initiative. The authors would like to thank European Commission Directorate General for Energy, the CoM Office and JRC colleagues working in the CoM initiative for their support in giving visibility and effectiveness to the

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

effort of cities and local governments in the climate change action. A special acknowledgment to our colleagues Monica Lanzoni for her deep reviewing of the statistical analyses.

Author contributions: Kona A., Bertoldi P. and Ferrario-Monforti F. designed the research. Kona A. prepared the manuscript. Kona A., Baldi M.G., Kakoulaki G. and Ahlgren C. assembled and prepared the dataset. All the authors revised the methodological steps and the manuscript.

Competing interests: The authors declare no competing interests.

Disclaimer: The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

References

Acuto, M., Parnell, S. and Seto, K. C.: Building a global urban science, Nat. Sustain., 1(January), 4–6, doi:10.1038/s41893-017-0013-9, 2018.

Adami, L., Tubino, M., Ragazzi, M., Conti, F. and Rada, E. C.: Local actions for reducing global greenhouse gas footprint: 10 years of covenant of mayors initiative, Int. J. Sustain. Dev. Plan., 15(2), 247–252, doi:10.18280/ijsdp.150216, 2020.

Bertoldi, P., Iancu, A., Kona, A., Suvi, M., Muntean, M., Lah, O. and Rivas, S.: Guidebook "How to develop a Sustainable Energy and Climate Action Plan (SECAP)": PART 2 – Baseline Emission Inventory (BEI) and Risk and Vulnerability Assessment (RVA), EUR 29412., Publications Office of the European Union, Luxembourg. [online] Available from: https://doi.org/10.2760/118857, 2018a.

Bertoldi, P., Kona, A., Palermo, V., Zangheri, P., Serrenho, T., Rivas, S., Labanca, N., Kilkis, S., Lah, O., Glancy, R., Follador, M., Barbosa, P. and Andreanidou, K.: Guidebook "How to develop a Sustainable Energy and Climate Action Plan (SECAP)" PART 3-Policies, key actions, good practices for mitigation and adaptation to climate change and Financing SECAP(s), Publication Office of the European Union, Luxembourg. [online] Available from: https://doi.org/10.2760/58898, 2018b.

Bertoldi, P., Kona, A., Rivas, S. and Dallemand, J. F.: Towards a global comprehensive and transparent framework for cities and local governments enabling an effective contribution to the Paris climate agreement, Curr. Opin. Environ. Sustain., 30, 67–74, doi:https://doi.org/10.1016/j.cosust.2018.03.009, 2018c.

Bertoldi, P., Kona, A., Rivas, S. and Dallemand, J. F.: Towards a Global Comprehensive and Transparent Framework for Cities and Local Governments enabling an Effective Contribution to the Paris Climate Agreement, Curr. Opin. Environ. Sustain., 2018d.

Busch, H., Bendlin, L. and Fenton, P.: Urban Climate Shaping local response – The influence of transnational municipal climate networks on urban climate governance, Urban Clim., 24(December 2017), 221–230, doi:10.1016/j.uclim.2018.03.004, 2018.

Cerquitelli, T., Corso, E. Di, Proto, S., Capozzoli, A., Bellotti, F., Cassese, M. G., Baralis, E., Mellia, M., Casagrande, S. and Tamburini, M.: Exploring energy performance certificates through visualization, in Proceedings of the Workshops of the EDBT/ICDT 2019 Joint Conference,EDBT/ICDT 2019, Lisbon, Portugal, March 26, 2019., edited by CEUR-WS.org, p. 7. [online] Available from: http://ceur-ws.org/Vol-2322/BigVis_2.pdf, 2019.

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, Sci. Data, 1–17 [online] Available from: http://dx.doi.org/10.1038/s41597-020-0462-2, 2020.

Eurostat: Sustainable development in the European Union, Luxembourg. [online] Available from: https://doi.org/10.2785/555257, 2020.

Gant., F.: Understanding Statistics: Basic Theory and Practice. Basic Theory and Practise., 2013.

Heidrich, O., Reckien, D., Olazabal, M., Foley, A., Salvia, M., Hurtado, S. D. G., Orru, H., Flacke, J., Geneletti, D., Pietrapertosa, F., Hamann, J. J. and Tiwary, A.: National climate policies across Europe and their impacts on cities strategies, J. Environ. Manage., 168, 36–45, doi:10.1016/j.jenvman.2015.11.043, 2016.

IPCC: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel, and J. C. Minx, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. [online] Available from: http://www.ipcc.ch/pdf/assessment-report/ar5/wg3/ipcc_wg3_ar5_full.pdf, 2014.

IPCC, 2015: Human Settlements, Infrastructure, and Spatial Planning. In Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report, pp. 923–1000, Cambridge University Press, Cambridge. [online] Available from: https://doi.org/10.1017/CBO9781107415416.018, 2015.

Kilkis, S.: Data on cities that are benchmarked with the sustainable development of energy , water and environment systems index and related cross- sectoral scenario, Data Br., 24, doi:10.1016/j.dib.2019.103856, 2019.

Kona, A., Bertoldi, P., Monforti-Ferrario, F., Rivas, S. and Dallemand, J. F.: Covenant of mayors signatories leading the way towards 1.5 degree global warming pathway, Sustain. Cities Soc., 41, 568–575, doi:10.1016/J.SCS.2018.05.017, 2018.

Kona, A., Bertoldi, P. and Kılkış, Ş.: Covenant of Mayors: Local Energy Generation, Methodology, Policies and Good Practice Examples, Energies, 12(985), doi:10.3390/en12060985, 2019.

Kona, A., Bertoldi, P., Baldi, M., Monforti-Ferrario, F., Palermo, V., Lo Vullo, E., Rivas-Calvete, S.,

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

Kakoulaki, G., Vetters, N., Thiel, C., Sgobbi, A., Ahlgren, C. and Posnic, B.: A dataset of GHG emissions for 6,200 cities in Europe and the Southern Mediterranean countries, Publications Office European Union, Luxembourg. [online] Available from: https://doi.org/10.2905/57A615EB-CFBC-435A-A8C5-553BD40F76C9, 2020.

Kovac, A., Mcdaniel, S., Kona, A., Bertoldi, P. and Chavara, C.: Aggregating Cities' GHG Mitigation Targets with Modeled Emissions Scenarios, Washington, DC. [online] Available from: https://files.wri.org/s3fs-public/aggregating-cities-ghg-mitigation-targets.pdf, 2020.

Marcotullio, P. J., Sarzynski, A., Albrecht, J. and Schulz, N.: A Top-Down Regional Assessment of Urban Greenhouse Gas Emissions in Europe, Ambio, 43, 957–968, doi:10.1007/s13280-013-0467-6, 2014.

Monforti-Ferrario, F., Kona, A., Peduzzi, E., Pernigotti, D. and Pisoni, E.: The impact on air quality of energy saving measures in the major cities signatories of the Covenant of Mayors initiative, Environ. Int., 118(February 2018), 222–234, doi:10.1016/j.envint.2018.06.001, 2018.

Palermo, V., Bertoldi, P., Apostolou, M., Kona, A. and Rivas, S.: Data on mitigation policies at local level within the Covenant of Mayors ' monitoring emission inventories, Data Br., 32, 106217, doi:10.1016/j.dib.2020.106217, 2020.

Peduzzi, E., Baldi, M. G., Pisoni, E., Kona, A., Bertoldi, P. and Monforti-ferrario, F.: Impacts of a climate change initiative on air pollutant emissions : Insights from the Covenant of Mayors, Environ. Int., 145(April), 106029, doi:10.1016/j.envint.2020.106029, 2020.

Reckien, D., Salvia, M., Heidrich, O., Church, J. M., Pietrapertosa, F., De Gregorio-Hurtado, S., D'Alonzo, V., Foley, A., Simoes, S. G., Lorencová, E. K., Orru, H., Orru, K., Wejs, A., Flacke, J., Olazabal, M., Geneletti, D., Feliu, E., Vasilie, S., Nador, C., Krook-Riekkola, A., Matosović, M., Fokaides, P. A., Ioannou, B. I., Flamos, A., Spyridaki, N., Balzan, M. V., Fülöp, O., Paspaldzhiev, I., Grafakos, S. and Dawson, R.: How are cities planning to respond to climate change? Assessment of local climate plans from 885 cities in the EU-28, J. Clean. Prod., 191, 207–219, doi:https://doi.org/10.1016/j.jclepro.2018.03.220, 2018.

Rosner, B.: Percentage Points for a Generalized ESD Many-Outlier Procedure, Technometrics, 25(2), 165–172, 1983.

Seem.: Using intelligent data analysis to detect abnormal energy consumption in buildings., Energy Build., 39, 52–58, 2007.

van der Ven, H., Bernstein, S. and Hoffmann, M.: Valuing the Contributions of Nonstate and Subnational Actors to Climate Governance, Glob. Environ. Polit., 17(1), 1–20, 2017.

Lo Vullo, E., Muntean, M., Duerr, M., Kona, A. and Bertoldi, P.: National and European Emission Factors for Electricity - Covenant of Mayors, Luxembourg. [online] Available from:

https://doi.org/10.5194/essd-2021-67
Preprint. Discussion started: 15 March 2021

https://data.jrc.ec.europa.eu/collection/id-00172, 2020.