Project no. 606953

MarcoPolo

Monitoring and Assessment of Regional air quality in China using space Observations.

Project Of Long-term sino-european co-Operation

Type of funding scheme: Collaborative Project - Small or medium-scale focused research project
Work programme topics addressed: SPA.2013.3.2-01: Cooperation with third countries

Deliverable D 4.2 (PU)

Marco Polo Emission Inventory for East-China: Basic Description

Version: 1.0

Due date of final deliverable: project month 30 (June 2016)
Actual submission date: project month 30 (June 2016)

Organisation name of lead contractor for this deliverable: VITO

Document Status sheet	
Lead Author	Hans Hooyberghs (VITO)
Contributing Authors	Nele Veldeman (VITO), Bino Maiheu (VITO)
Reviewed by	
Distribution list	Project partners and EC

Document Change Record			
Date	***Issue***	***Pages affected***	***Description***
June 15th	1.0	All	Initial draft
June 27th	1.1	All	Added high resolution inventory

Table of Contents

1. Introduction

In Task 4.2 of the MarcoPolo project, a new low-resolution emission inventory for the East of China has been composed, as well as three high-resolution inventories for three zoom-in regions (Beijing-Tianjin region, Yangtze Delta and Pearl Delta).

The new low-resolution inventory contains monthly anthropogenic and biogenic emissions at a 0.25x0.25 degree resolution for 2014, for a domain containing most major Chinese cities, the Korean peninsula and parts of Mongolia. This inventory is mainly based on the emission estimates obtained using satellite data in work package (WP) 3 The source sector information is based on existing emission inventories available for China (the 2010 MIX- and the 2012 MEIC-inventory) and bottom-up local data. The high resolution inventories each provide anthropogenic emissions at a 0.01x0.01 degree resolution, an appropriate resolution for the modellers involved in Work Package 5. To downscale to the higher resolution, bottom-up and top-down proxy data have been used. A basic schematic overview of the methodology is provided in Figure 1.

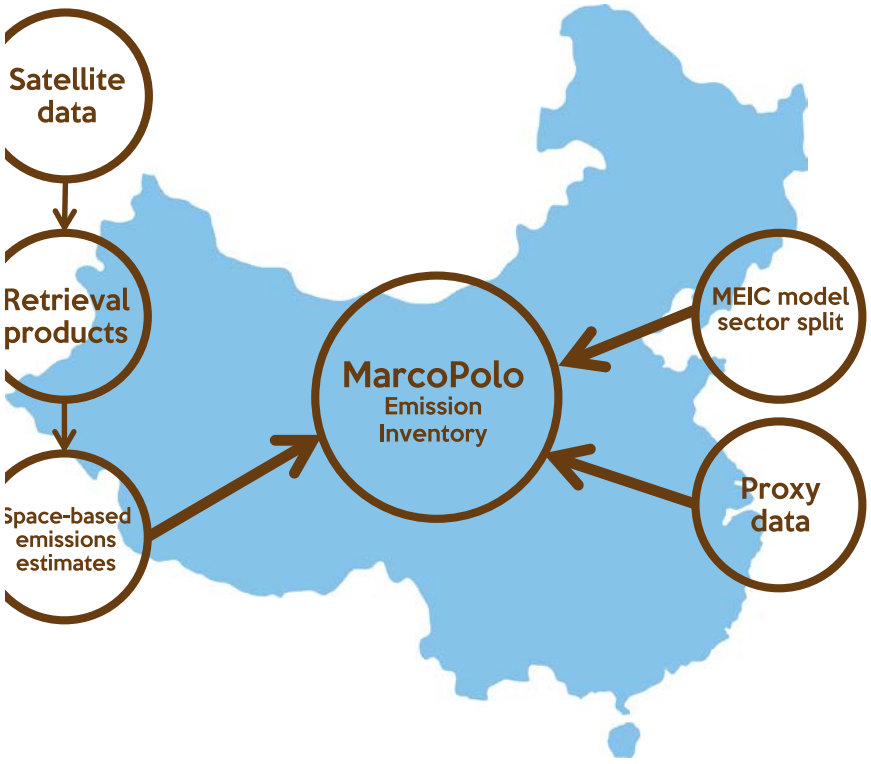

Figure 1: Schematic overview of the methodology.

This short description describes both the low and the high resolution inventory. The input and methodology are briefly discussed, and the new inventory is compared with the results of the MIX-and MEIC-inventory.

2. Low resolution inventory

2.1. General description

The new low-resolution inventory contains anthropogenic and biogenic emissions at a 0.25 degree resolution for 2014, for a domain containing most major Chinese cities, the Korean peninsula and a part of Mongolia. Figure 2 visualizes the entire domain, which lies between 102 and 132 eastern longitude, and between 18 and 50 northern latitude. The coordinates refer to the centre of the grid cells (the centre of the top-left grid cell is thus located at 102E 50N). There are in total 121 by 129 grid cells in the domain. In this way, the grid mimics the grid properties of the NO_x-emissions of KNMI, and the SO_2-emissions of AUTH.

The inventory contains emission estimates for nitrogen oxides (NO_x), sulphur dioxide (SO_2), particulate matter (both $PM_{2.5}$ and PM_{10}), ammonia (NH_3), carbon monoxide (CO), non-methane volatile organic compound (VOC) and black carbon. The anthropogenic fraction provides emissions for six sectors: industry, power, residential, transportation, shipping and agriculture (only NH_3), while the biogenic emissions focus on wildfires.

Figure 2: Extent of the domains of the low resolution (red) and high resolution inventories (green) composed in Task 4.2. The high resolution inventories contain, from top to bottom, the Beijing area, the Yangtze Delta and the Pearl Delta. Background image: Open Street Maps.

The new inventory is mainly based on the emission estimates obtained using satellite data. In WP3, estimates for four pollutants have been composed: NO_x, SO_2, VOC and $PM_{2.5}$. The source sector information for these pollutants is based on existing emission inventories available for China (the 2010 MIX- and the 2012 MEIC-inventory). Complementary bottom-up local data has been used to fill the gaps, providing, amongst other, shipping emissions in ports. For pollutants without satellite retrievals, we either combine the sector split data of the existing inventories with satellite estimates for related pollutants (PM_{10} and BC), or we simply use the emissions contained in the existing inventories (NH_3 and CO). The biogenic emissions are based on the satellite estimates for $PM_{2.5}$ wildfire emissions and emission factors per dry matter (DM) burned. In the remainder of the paragraph, the methodology is described more in detail. We also compare the new emission inventory with the existing MIX- and MEIX-inventories.

2.2. Anthropogenic emissions for pollutants with satellite estimates

2.2.1. Satellite emission estimates

In WP3 of the MarcoPolo project, anthropogenic emission estimates for various pollutants have been composed based on inverse modelling and satellite retrievals. For some pollutants, multiple emission estimates have been obtained, often using different techniques. In consultation with all the partners involved in the MarcoPolo project, it has been decided to use the NO_x-emissions of KNMI, the SO_2-emissions of AUTH, the VOC-emissions of BIRA and the $PM_{2.5}$-emissions of FMI. All these estimates focus on the year 2014 and are considered to be the best available products. A detailed description of these datasets is provided in the reports of WP3.

All emission estimates are provided on a 0.25x0.25 degree resolution, but the grids and spatial extents differ. We limit all the emission to the spatial extent outlined in Figure 2, and resample using bilinear resampling to the grid of the KNMI NO_x-emissions.

At the moment, the satellite based estimates of AUTH only provide emissions for mainland China, as their inversion routine did not yield trustworthy results above sea and in the countries bordering China. Therefore, we use the emissions contained in the MIX-MEIC 2012 inventory (described in section 2.2.2) for the countries bordering China. These emissions are by far dominated by the emissions in mainland China, thus the impact on the total emissions in the domain is negligible. Since the MIX-MEIC 2012 inventory does not contain any shipping emissions, there are currently no SO_2 emissions for oversea grid cells. We hope to change this in a future version of the inventory.

2.2.2. Anthropogenic sector split

MIX-MEIC 2012 inventory

To determine the sector split of the low resolution MarcoPolo emission inventory, we make use of the MIX-inventory (Li et al. 2015) that has been composed in the third phase of the Model Inter-Comparison Study for Asia (MICS-Asia). This inventory contains monthly 0.25x0.25 degree gridded emissions for overland regions in Asia for 2008 and 2010 for five sectors (residential, power, industry, transport and agriculture). Since there are no shipping emissions, the inventory only contains overland emissions. The inventory is composed using a mosaicking procedure with various local emission inventories and the REAS v2.0 inventory as input. The MIX-inventory has been incorporated in the HTAP v2 gridded emission inventory. Gridded emissions are available on the website of the MEIC-inventory, http://www.meicmodel.org/dataset-mix.html.

For China, the MIX-inventory contains the 2010 0.25x0.25 degree gridded MEIC-emissions (Multi-resolution Emission Inventory for China), constructed by THU in the first task of WP4. Within the same task, THU has also compiled an updated version of the MEIC-inventory using 2012 data. Since the ultimate goal of WP4 is providing a new, updated emission inventory, we try to use as recent data as possible. We thus replace the 2010 MEIC-emissions with the 2012 MEIC-emissions in the MIX-inventory, thereby obtaining an "updated" version of the MIX-inventory. We apply this procedure for all pollutants except NH_3. For ammonium, the MIX-inventory contains the process-based (and thus more correct) inventory of PKU, another partner involved in WP 4 (Huang et al. 2012), and we therefore keep the original 2010 MIX-emissions for this pollutant.

In the remainder of the text, we refer to this modified MIX-inventory as the *MIX-MEIC 2012 inventory*.

Source sector contribution

An important aspect of an emission inventory deals with the sector contributions. The satellite based estimates lack such information; to remedy this shortcoming we make use of the MIX-MEIC inventory. For each grid cell, the sector split contained in the MIX-MEIC 2012 inventory has been used to calculate the local sector split in the new MarcoPolo inventory. Since there is a shift between the grid of the MIX-MEIC 2012 inventory and the grid of the MarcoPolo inventory (the grid centers of both inventories are shifted by half a grid cell with respect to each other in both east-west and north-south direction), we first regrid the sector splits of the MIX-MEIC 2012 inventory using a simple bilinear resampling procedure[1].

In this way, the relative importance of the power generating, residential, industrial, agriculture and transportation sector are derived. Another major anthropogenic emission source deals with shipping emissions. As both the MIX- and the MEIC-inventory lack these, we have to rely on

[1] There are however two small caveats in this procedure:

- Since the MIX-MEIC 2012 inventory contains only overland emissions, the bilinear resampling causes some small issues close to the coastline (as it would retract the coastline dealing with the NaN-values above sea). To remedy, for all affected grid cells, nearest-neighbor interpolation (using the nearest overland grid cell) is preferred above bilinear interpolation.
- For a small region within the domain (in the Gobi desert) the total emissions are zero for some pollutants. For these locations, we use the sector split of the nearest grid cell that contains non-zero total emissions.

other source. Of course, almost all - if not all - anthropogenic emissions at sea are related to shipping. All emissions in grid cells that lie completely above sea, are thus classified as shipping emissions. Moreover, as detailed bottom-up data on inland shipping emissions is missing and other existing inventories for China contain only very limited shipping emissions for overland grid cells, we assume that these inland shipping emission are dominated by the other sectors. Hence, inland shipping will be neglected.

In major seaports, shipping emissions are of course non-negligible. We first tried to estimate these using other emission inventories. However, a comparison with bottom-up data indicates that most of the inventories are underestimating the importance of the shipping emissions. According to bottom-up data (Fung et al. 2014), approximately one third of the particulate matter and NO_x-emissions in the HongKong port region is related to shipping emissions. Somewhat lower, but still significant values are observed in Shenzhen and Shanghai. For SO_2, shipping is the source of half of the emissions in HongKong, and even two thirds of the emissions in Shenzhen. Again, lower values are observed in Shanghai, where the industrial SO_2-emissions are much higher. A complete overview can be found in Table 1. Shipping emissions in the main seaports are hence certainly significant, and we therefore compose them using a bottom-up estimate, as explained in detail in the next paragraph.

Table 1: Relative importance of shipping emissions in three major ports in China (source: (Fung et al. 2014)).

Port	NOx (%)	PM (%)	SO2 (%)	VOC (%)	CO (%)
HongKong	32	34	50	11	17
Shenzhen	14	6	66		
Shanghai	12	6	12		

Shipping emissions in ports

As shipping causes a significant fraction of the emissions in ports, we construct a bottom-up inventory for the major ports in China. For simplicity and computational complexity, we only retain ports that are part of the hundred largest container traffic ports in the world, based on the 2014 list of the American Association of Port Authorities[2] (AAPA). For the Chinese ports, this list is based on data of the Chinese Waterborne Transport Institute for ports in China. In total, we consider 14 ports, listed in Table 2. Their locations are shown in Figure 3.

Table 2: List of the ports for which bottom-up emissions have been constructed. The first column provides the rank of the port on the World Ports Index for 2014, the last column provides the container capacity of the port (in 1000 Twenty Foot Equivalent Units (TEUs).

Rank	Port	Country	Capacity (1000 TEU)

[2] Other datasets exists as well, as for instance the Lloyd's list with the largest 100 container ports for 2014. However, there is a great resemblance between both lists, and the actual TEUs for Chinese ports approximately match each other.

1	Shanghai	China	35,286
3	Shenzhen	China	23,798
4	Hong Kong	China	22,374
5	Ningbo	China	19,450
7	Qingdao	China	16,624
8	Guangzhou	China	16,160
10	Tianjin	China	14,050
14	Dalian	China	10,128
17	Xiamen	China	8,572
25	Yingkou	China	5,768
27	Lianyungang	China	5,005
65	Yantai	China	2,361
66	Rizhao	China	2,349
70	Fuzhou	China	2,237

Figure 3: Location of the fourteen ports that are considered in this study. The actual location has been based on open source information and satellite imagery.

Bottom-up emission information is only available for a small number of ports in China (Fung et al. 2014). Detailed inventories are available for HongKong (C. C. Chang 2014), and Shanghai (Fu et al. 2012), and there is also some data for the port of Shenzhen (Yang et al. 2015) and for the entire Guangdong region (which comprises, amongst other, the ports of Shenzhen and Guangzhou) (Siqi et al. 2014). For the former three ports, we make use of the actual emissions contained in these local, bottom-up inventories. The details can be found in Table 3. For all other ports, we estimate the emissions based on the container capacity of the port, i.e. we linearly rescale the emissions of the port of Shanghai to a port of a similar size. Although this method provides only rough estimates for the shipping emissions of the main ports, it is a more accurate reflection of the reality than the one contained in existing inventories.

Table 3: Shipping emissions in three major ports in China. Source: HongKong (C. C. Chang 2014), Shanghai (Fu et al. 2012) and Shenzhen (Yang et al. 2015).

Port	NOx ton/year	PM2.5 ton/year	SO2 ton/year	VOC ton/year	CO ton/year
HongKong	36500	2250	16500	3480	11800
Shenzhen	23300	1700	13600	1000	2200
Shanghai	57300	3700	35400	5000	4900

2.3. Anthropogenic emissions for other pollutants

Only satellite estimates for NO_x, $PM_{2.5}$, SO_2 and VOC have been composed in WP3. For the other pollutants in the inventory, we have to rely on other techniques. For PM_{10} and BC, we exploit the link between these pollutants and $PM_{2.5}$. For ammonium, we simply use the inventory of Peking University contained in the MIX-inventory. Finally, also the CO-emissions are taken directly from the MIX-MEIC 2012 inventory, as other options failed to compose a reliable dataset.

2.3.1. PM_{10} and BC

For PM_{10} and BC, we exploit the close link between these pollutants and $PM_{2.5}$. For each grid cell, the local emissions for PM_{10} and BC are based on the local $PM_{2.5}$-emissions contained in the satellite emission estimates, and the local relation between the BC- / PM_{10}-emissions and the $PM_{2.5}$-emissions contained in the MIX-MEIC 2012 inventory. This procedure is applied sector by sector and month by month, resulting in a best estimate for the PM_{10}- and BC-emissions.

Since the MIX-MEIC 2012 inventory does not contain any shipping emissions, the method outlined in the previous paragraph is unable to estimate those. Shipping emissions are therefore estimated based on the ratio of the total monthly BC- / PM_{10}- and $PM_{2.5}$-emissions contained in the MIX-MEIC 2012 inventory. The same ratio is used for the oversea emissions and the emissions in the ports. As a consequence, the ratio between the BC / PM_{10}- and the PM_{25} shipping emissions is constant over the domain.

2.3.2. NH_3 and CO

For ammonium, the inventory of PKU is a detailed, process based inventory (Huang et al. 2012; Li et al. 2015). Because of this, and the lack of satellite emission estimates, we use the PKU NH_3-emission estimates in the MarcoPolo inventory. Note that these emissions stem from 2006, but since the agricultural sector has developed much slower than other sectors, the emissions in 2014 will not differ a lot from these in 2006 (Li et al. 2015). Moreover, we assume that there are no shipping NH_3 emissions.

For CO, we have tried a procedure similar to the one for PM_{10} and BC, but it was impossible to generate an accurate inventory in this way, as the relationship between CO and other pollutants is too delicate. Hence, we have copied the MIX-MEIC 2012 emissions to the MarcoPolo emission inventory. Shipping emissions in ports for for CO are based on bottom-up estimates, following

the procedure outlined in section 2.2.2. Oversea shipping emissions for CO are currently not considered.

2.4. Biogenic emissions

Emissions from natural sources have many different origins. In the MarcoPolo inventory, we focus on emissions caused by wildfires, as we presume that they will be the dominating factor for the pollutants under consideration and for the domain under consideration. Locally, other biogenic emissions may be important as well during some episodes (f.i. direct emissions from vegetation, volcanic emissions, agricultural waste burnings[3]…), but it's hard to estimate these.

2.4.1. Particulate matter

Within WP3, wildfire $PM_{2.5}$-emission estimates have been composed by FMI. Mostly, these wildfire emissions are rather small. Only during specific months the emissions become important, at some specific locations in the domain.

The majority of the biogenic emissions in the MarcoPolo domain are emitted in March and April. These emissions occur mostly outside China, in the south-east of the domain (mainly in the north-east of Laos and the north of Vietnam), with a smaller contribution in the Yunnan region in China. In April, there is also a substantial contribution in North-Korea. Minor peaks occur in January, June and October. In January, the peak is related to higher emissions in the south of China, mostly in the provinces Guangdong, Hunan and Jiangxi. In June, the largest share of the emissions occurs at rural locations in the Eastern plains, with large peaks in the Shangdong, Jiansu and (the north of the) Anhui provinces. In October, the fires mostly take place in the North-East of China, in the Heilongjiang and the Jilin provinces. Hence, focussing on the Chinese part of the domain, the peaks in June, January and September are the most important. Given their locations, these biogenic emissions will also have the largest impact on the air quality.

These periods and regions with substantial wildfire emissions correspond with those in literature. We have compared the satellite based emissions with emission estimates contained in the GFEDv4 (Global Fire Emissions Database) inventory, and with MODIS-based fire datasets. The former provides global estimates of monthly burned area, monthly carbon emissions and fractional contributions of different fire types at a 0.25x0.25 degree spatial resolution. For the MODIS-based products, we used the burned area monthly product with a 500m resolution. Although there are some differences on the local scale between these two datasets and the satellite based emissions of FMI, the general locations of fires in both datasets agree qualitatively, especially when focussing on the peaks in China that occur in January, June and September.

2.4.2. Other pollutants

[3] It could be discussed whether these emissions caused by agricultural waste burning are "biogenic" or "anthropogenic" emissions, but it's sure that they will be very important in some areas in China. Especially in the north-eastern planes, they'll have some significant impact. In a future version of the inventory, we hope to include these emissions as well.

There are no satellite emission estimates for the other pollutants in the inventory. Hence, we rely anew upon the GFEDv4 database, which also contains emission factors relating the emissions of many pollutants to dry matter emissions. The database contains factors for different type of fires (savanna, grassland, and shrubland fires, boreal forest, temperate forest fires, tropical deforestation & degradation, peat fires and agricultural waste burning). Here, we make a combination of these factors in order to represent the diverse types of burning that occur in the domain. Based on this new factor, we deduce the factors relating the $PM_{2.5}$ emissions to the emissions of the other pollutants in the MarcoPolo inventory (see Table 4 for an overview).

Pollutant	**CO**	**NO_x**	**PC**	**BC**	**SO_2**	**VOC**	**NH_3**	**$PM_{2.5}$**	**OC**	**PM_{10}**
Emission factor	11.42	0.223	0.582	0.045	0.07	0.165	0.15	1	0.58	1

Table 4: Emission factors for wildfires that have been used in this study. These factors are based on the factors of the GFEDv4 database.

2.5. Results

2.5.1. MarcoPolo low resolution emission inventory

Figure 4 shows the total anthropogenic and biogenic emissions in 2014 for all pollutants in the MarcoPolo domain, and Figure 5 shows the sector split based on the yearly total emissions in 2014. Figure 6 shows the monthly profiles for the anthropogenic emissions for the four main pollutants (NO_x, SO_2, VOC and $PM_{2.5}$), Figure 7 shows the profile for the biogenic emissions (which is identical for all the pollutants, because of the methodology outlined in the section 2.4). These profiles are again based on total yearly emissions.

For all pollutants, the anthropogenic emissions are clearly larger than the biogenic emissions. The largest biogenic fraction is observed for CO, for which 17% of the emissions are related to wildfires. Also for particulate matter (12% for $PM_{2.5}$ and 9% for PM_{10}) there are substantial wildfire emissions. For the other pollutants, the biogenic emission considered in this inventory correspond to at most 5% of the total emissions. Figure 7 shows that most of the biogenic emissions are related to wildfires in the March and April, with minor peaks in January, June and October. This pattern matches the one described in section 2.4.

The main source of the anthropogenic emissions differs between the pollutants. For all pollutants except NO_x and NH_3, the industrial and residential sector are the main contributors. The share of industrial emissions ranges from 30 % (for CO and BC) up to 60% (for SO_2), while the share of the residential sector varies between 20% (for SO_2) and 40% (for BC). The remaining emissions (at most 30%) are mostly attributed to power generation and transportation. The pattern is completely different for NO_x and NH_3. Traffic emissions and the industrial sector dominate the nitrogen dioxide emissions. Finally, ammonium emissions are dominated by the agricultural sector.

Also the monthly profiles differ substantially amongst the pollutants. Whereas SO_2- and VOC-emissions attain their highest values in the winter months, the profiles for particulate matter and NO_x are much more chaotic, with large differences between adjacent months. These profiles of course mimic the behaviour of the underlying satellite based emissions.

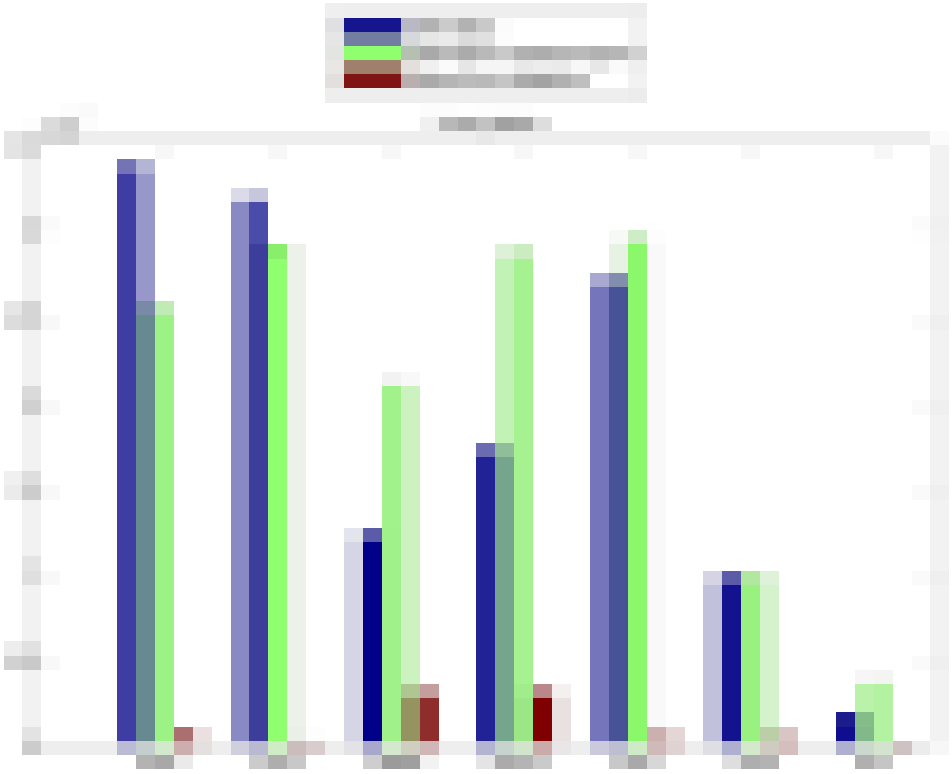

Figure 4: Comparison of the yearly total anthropogenic emissions contained in the MIX-MEIC 2012 inventory (blue) and the 2014 MarcoPolo low resolution inventory (green). Also the biogenic emissions in the MarcoPolo inventory are shown (dark red). For lay-out reasons, the results for CO are not shown. All emissions are provided in ton/year.

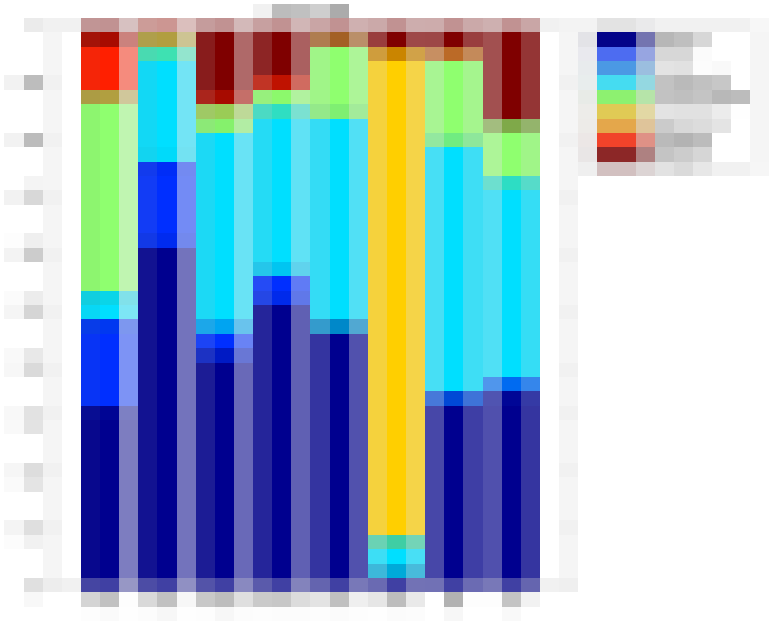

Figure 5: Sector split of the yearly total emissions contained in the low resolution MarcoPolo emission inventory. The figure shows, for all pollutants, the contribution of the six anthropogenic sectors (bottom to top: industry, power, residential, transportation, agriculture and shipping) and the biogenic fraction (top: brown).

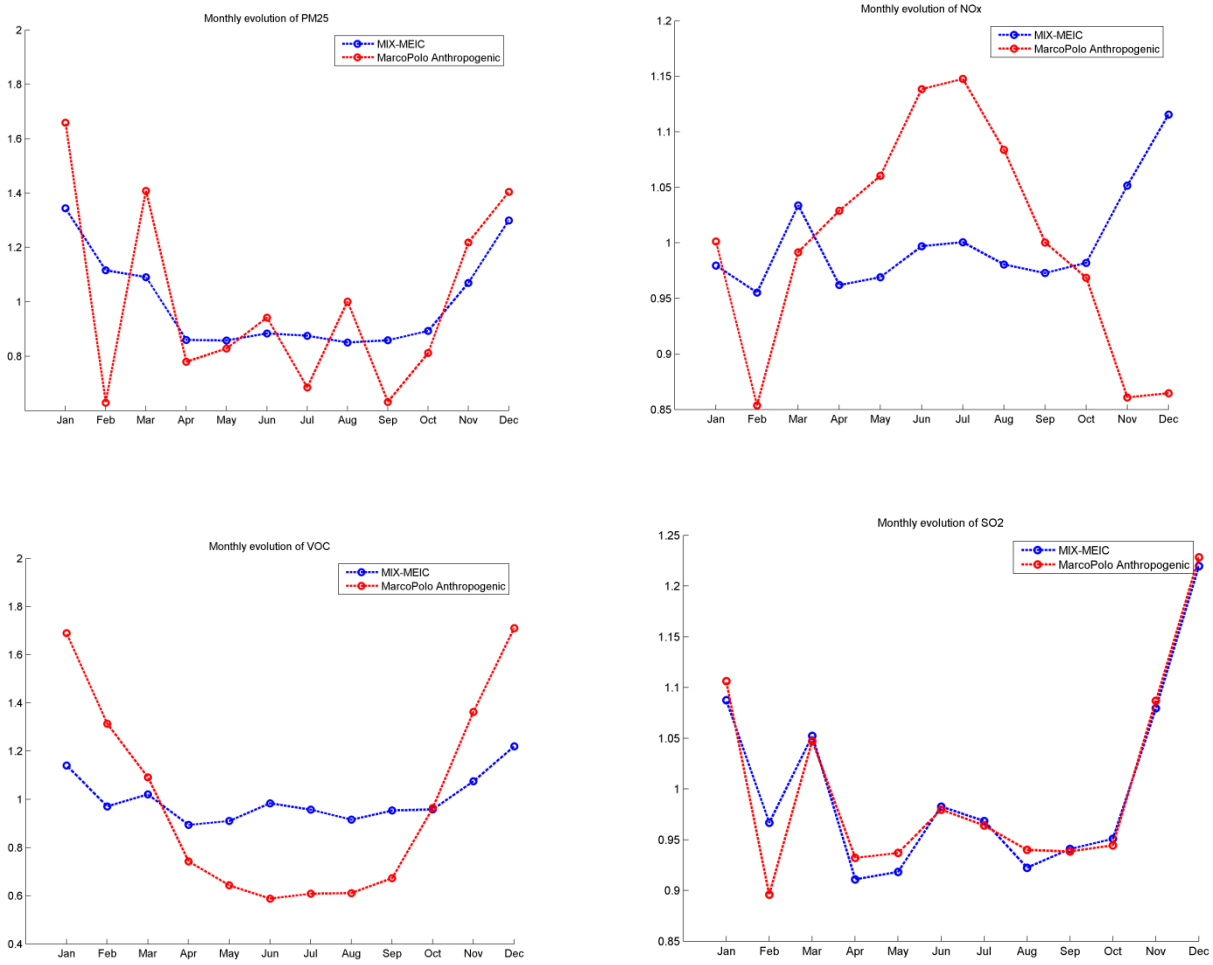

Figure 6: Monthly profiles of the four pollutants for which emissions are directly based on satellite estimates (top left: $PM_{2.5}$, top right: NO_x, bottom left: VOC, bottom right: SO_2). For each pollutant, the figure shows the profile for the MIX-MEIC 2012 inventory (blue) and the anthropogenic emissions contained in the MarcoPolo 2014 low resolution inventory (red).

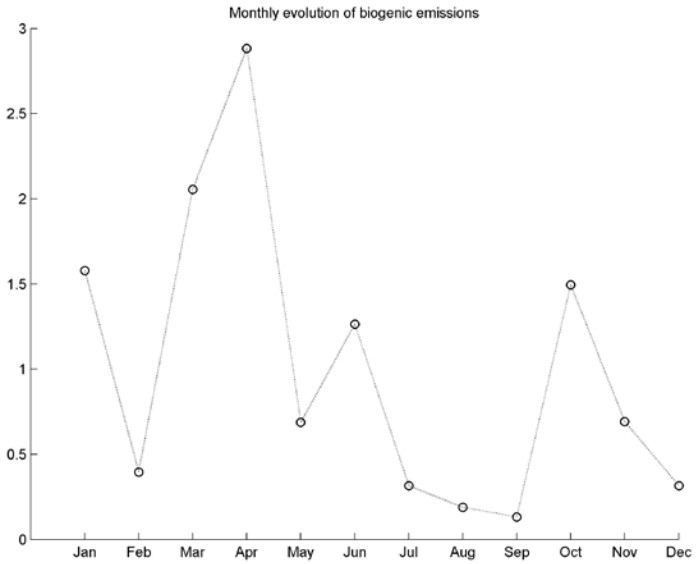

Figure 7: Monthly profiles of the total biogenic emissions contained in the MarcoPolo 2014 low resolution inventory.

2.5.2. Comparison with MIX-MEIC 2012 inventory

Focussing on the yearly total emissions in the domain, the largest deviations are observed for particulate. The $PM_{2.5}$ and PM_{10} emissions differ by approximately 40%, with the larger value observed in the satellite based emissions. For nitrogen dioxide, the difference is approximately 25%, with lower totals for the satellite based emissions. For SO_2 and VOC, both emission datasets differ at most 15%.

There are also major discrepancies between the monthly profiles of the MIX-MEIC 2012 inventory, and those of the new MarcoPolo inventory, especially for NO_x and $PM_{2.5.}$ For the former pollutant, the seasonality in the MarcoPolo inventory is much larger than the one in the MIX-MEIC 2012, for which there is not much difference between the different months. For nitrogen dioxide, the monthly profile is completely different, as the maximal and minimal emissions occur in different months in both inventories. Note however that, for both inventories, emissions differ at most 20% between the months. The general profiles for particulate matter match each other more closely, but the profile of the MarcoPolo inventory is much more chaotic (with much larger differences between months). For SO_2, both profiles are almost identical.

Figure 8 provides a comparison between the sector split of the total anthropogenic annual emissions contained in the MarcoPolo and the MIX-MEIC 2012 inventory. Although the local, monthly sector split in the former inventory is reminiscent of the one in the latter, there are substantial differences in the source attribution of the total annual emissions. For particulate matter, the differences are rather small, but for the other pollutants there are major distinctions. In comparison with the MIX-MEIC 2012 inventory, the MarcoPolo inventory contains in general more residential emissions and less industrial emissions. The reason hereof is the different spatial distribution of the emissions, as the satellite based emissions are much more concentrated on the

large cities. This trade-off between the industrial and residential emissions is the largest for VOC, but it is also clearly existing for SO_2 and particulate matter, and, to a lesser extent, for nitrogen dioxide. Finally, the relative importance of transportation is much larger in the MarcoPolo inventory, in comparison with the MIX-MEIC 2012 inventory.

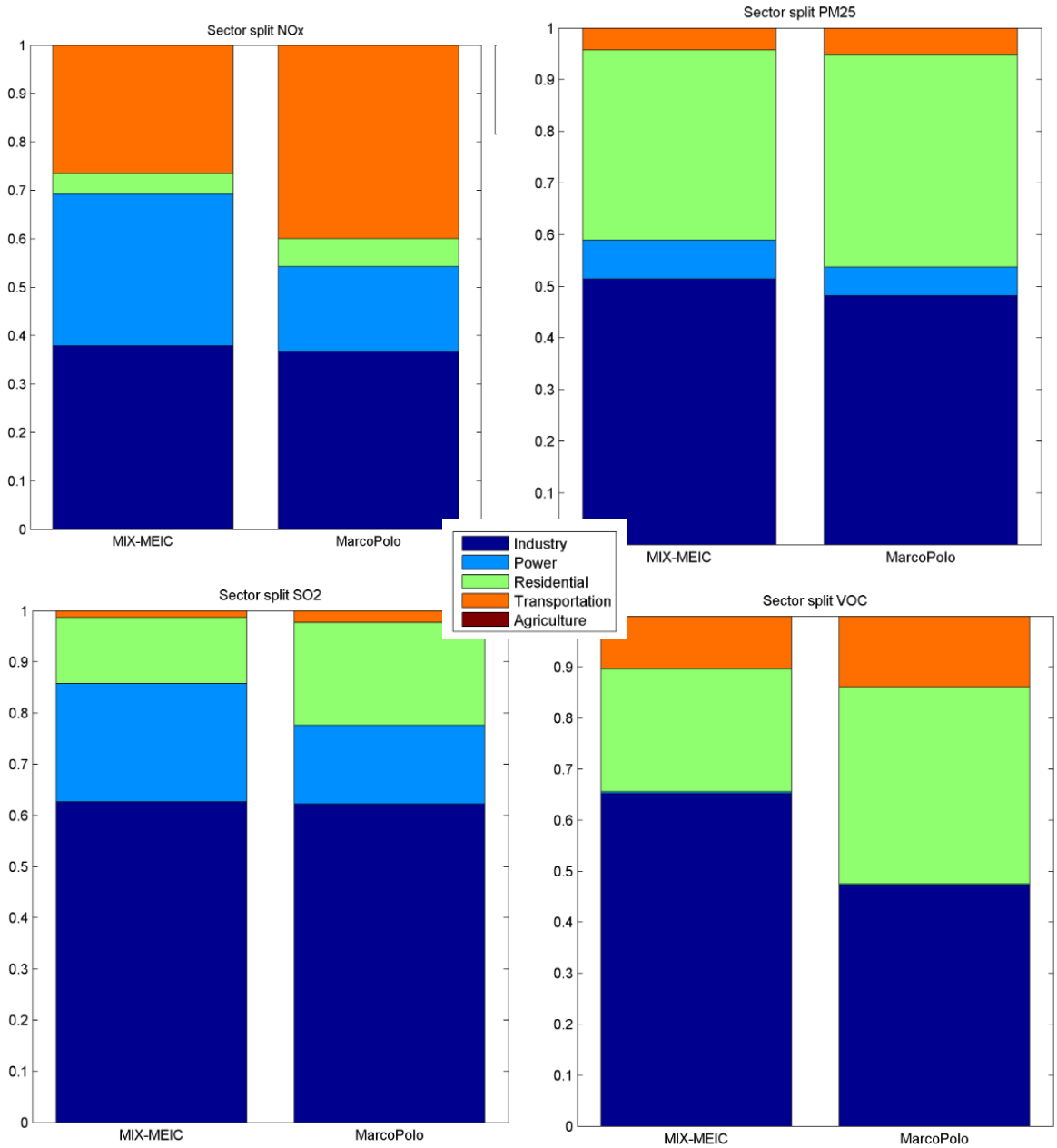

Figure 8: Comparison of the sector split of the yearly total emissions contained in the low resolution MarcoPolo emission inventory (right) and the MIX-MEIC 2012 inventory (left), for the five sectors contained in both inventories (from bottom to top: industry, power, residential, transportation and agriculture). Results are shown for NO_x (top left), $PM_{2.5}$ (top right), SO_2 (bottom left) and VOC (bottom right).

3. High resolution inventory

3.1. General description

For three selected regions, a more detailed inventory with a 0.01 degree resolution has been composed. These high resolution inventories only contain anthropogenic emissions, because the high resolution domains focus on urban areas (where anthropogenic emissions dominate), and because it is challenging to compose biogenic emissions with a resolution of 1 kilometre.

The location of the high-resolution regions, the Beijing-Tianjin area, the Yangtze River Delta and the Pearl River Delta, is shown in Figure 2. Figure 9 shows a zoom on the three regions. These domains are chosen in consultation with the partners involved in work package five, in which the emission inventory will be used as input for concentration modelling using CTMs and dispersion modelling. The Beijing-Tianjin area lies in the north-east of China and comprises the capital Beijing and the city of Tianjin (including its port located in the Binhai New Area). The high resolution grid spans a domain of 2 x 2 degrees, thus containing 200 x 200 grid cells. It lies between 115.88 and 117.87 eastern longitude and between 38.38 and 40.37 northern latitude (the coordinates anew refer to the centre of the grid cells). The exact extent is chosen such that its boundary coincides with the boundary between two low-resolution grid cells. The Yangtze River Delta area in the east of China comprises the major cities in the vicinity of the Yangtze river: Shanghai, Ningbo, Hangzhou, Suzhou, Wuxi, Changzhou and Nanjing. The high resolution grid spans a domain of 4 x 3 degrees, thus containing 400 x 300 grid cells. It lies between 118.13 and 122.12 eastern longitude and between 29.63 and 32.62 northern latitude (the coordinates anew refer to the centre of the grid cells). The Pearl River Delta area in the south of China comprises the major cities in the Guangdong – Hong Kong region, including Guangzhou, Fosham, Dongguan, Shenzhen and Hong Kong. The high resolution grid spans a domain of 2 x 2 degrees, thus containing 200 x 200 grid cells. It lies between 112.63 and 114.62 eastern longitude and between 21.88 and 23.87 northern latitude (the coordinates anew refer to the centre of the grid cells).

The high resolution emission inventory is compiled by combining the low resolution Marco Polo inventory with top-down and bottom-up proxy data. In the next sections we consecutively describe the downscaling methodology and the proxy data.

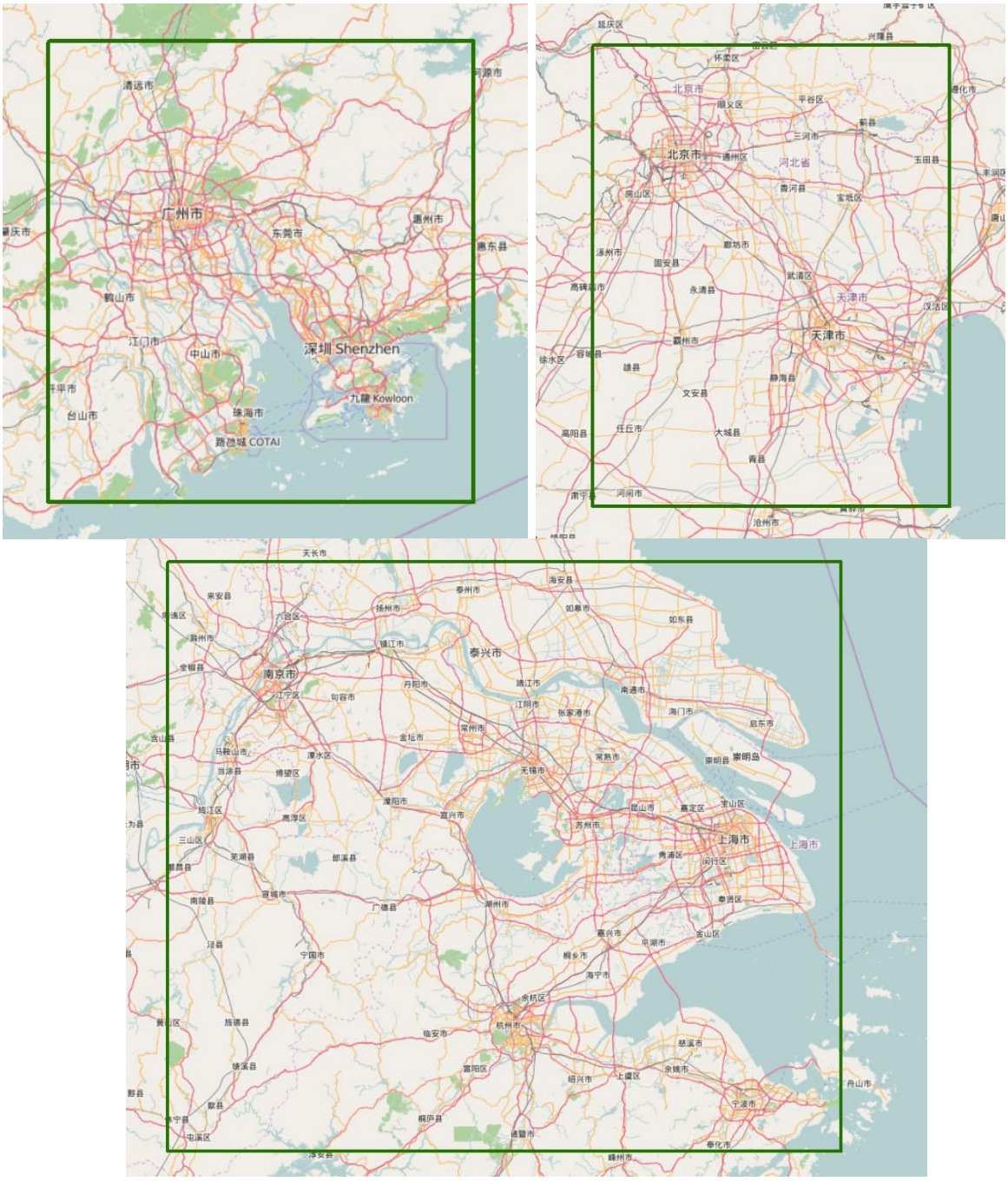

Figure 9: Domain of the three high resolution domains, focussing on the Pearl Delta (top left, containing Guangzhou, Shenzhen, Macau and Hong Kong), Beijing (top right, containing both Beijing and Tianjin), and the Yangtze Delta (bottom, containing Shanghai and Nanjing). Background image: Open Street Maps.

3.2. Downscaling algorithm

The general methodology consists of three major steps[4]. Initially, we start from the Marco Polo low resolution inventory, which provides emissions on a 0.25 x 0.25 degree grid. In a first step, the emissions are resampled to the high resolution 0.01 x 0.01 degree resolution grids using bilinear resampling. This step hinders the formation of large discontinuities at the borders of the low resolution grid cells.

In a second step, the resampled emissions are combined with weight factors based on proxy data. We rely on one top-down proxy per anthropogenic sector, and, in downscaling the power plant emissions, we have also used some bottom-up data to improve the top-down dataset manually. A basic overview of the proxies is provided in Table 5. In the next paragraph, we describe the different datasets in detail. For each high resolution grid cell, we calculate the fraction of the total proxy value in the high resolution domain contained in the grid cell, and use this value as a weight factor. As an example, to downscale the residential emissions, we calculate the fraction of the inhabitants in the high resolution domain that are living in each high resolution grid cell, and use these values as weight factors. The emissions in each high resolution grid cell are subsequently obtained by multiplying the resampled emissions (composed in step 1), with the weight factors.

Steps one and two change both the total emissions in the entire high resolution domain and the emissions in each low resolution grid cell. It is however impossible to keep both the "total" and the "local" emissions unchanged without introducing large discontinuities at the borders of the large resolution grid cells. Hence, we opt to only preserve the total emissions in the high resolution domains, and, as a consequence, the local emissions in each low resolution grid cell are not preserved in the corresponding high resolution grid cells. In sum, in the third step, we rescale all the emissions uniformly in order to match the *total* emissions in the high resolution domain with the *total* emissions in the corresponding cells of the low resolution domain.

Table 5: Overview of the top-down proxy data sets used.

Sector	Proxy
Agriculture	GlobCover land use (agriculture classes)
Industrial	GlobCover land use (built-up classes)
Powerplants	Enipedia (with manual editing)
Residential	WorldPop
Road Traffic	Open Street Maps
Shipping	Land-sea mask

3.3. Proxy data

3.3.1. Road traffic: OpenStreetMaps

[4] Note that the methodology is similar to the one used in the GlobEmission project, where high resolution emissions for Qatar have been composed.

For the transport sector, we assume that all emissions can be spread using a road traffic proxy. Hence, we assume that other transport emissions are negligible or at least spread similarly across the high resolution cells as the emissions of road transport. It is hard to find correct numbers on the ratio of the road and railway emissions, but according to the calculations based on the Statistical Communiqué on Road and Waterway Transport Industries issued by China's Ministry of Transport in 2012, the consumption of standard fuels by rail transport was only 7-9% of that by road transport (Hong et al. 2013). Hence we assume that, for the purpose of the downscaling, the rail emissions can be neglected.

To downscale the road emissions, we use the road map for China contained in the OpenStreetMap database, which was obtained on January 15th, 2016. For the high resolution region, a visual comparison with Google Maps has been performed for selected areas. In all cases, the OSM data matched the actual situation very closely. According to quality checks by our local partner, the quality of the OSM database for Beijing is very good, often outperforming Google Maps with respect to newly built roads.

For each road, the database contains a specific road type. We only consider the largest roads, those of the classes "motorway", "trunk", "primary" and "secondary". All smaller roads (for instance of type "local") are neglected, as we assume that the emissions on these roads are dominated by the emissions on the larger roads. For the downscaling, the simplest proxy would be the total road length in each grid cell. However, as higher emissions are expected on the larger roads, we apply a weight factor and calculate a weighted length in all grid cells. The weight factors are defined in Table 6. The weighted road length is then used in the methodology outlined in section 3.2.

Road Type	Value
Motorway	10
Trunk	5
Primary	2
Secondary	1

Table 6: Weight factors for the OpenStreetMaps roads.

3.3.2. Agriculture: GlobCover

As the low resolution inventory only contains NH_3-emissions for the agriculture sector, the proxy related to agriculture preferably indicates where these ammonium emissions are to be found. As very detailed local land use data for agricultural regions is missing, we make use of a top-down dataset based on satellite retrievals, GlobCover. The GlobCover dataset provides worldwide land cover data at 300 m resolution, its latest revision occurred in 2009. In practice, the resolution is somewhat lower in China, as a visual inspection of the data suggests. GlobCover makes use of 23 classes, of which 21 natural and one related to artificial structures (see Table 7).

Class	Label
11	Post-flooding or irrigated croplands (or aquatic)
14	Rainfed croplands
20	Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)
30	Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)
40	Closed to open (>15%) broadleaved evergreen or semi-deciduous forest (>5m)
50	Closed (>40%) broadleaved deciduous forest (>5m)
60	Open (15-40%) broadleaved deciduous forest/woodland (>5m)
70	Closed (>40%) needleleaved evergreen forest (>5m)
90	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)
110	Mosaic forest or shrubland (50-70%) / grassland (20-50%)
120	Mosaic grassland (50-70%) / forest or shrubland (20-50%)
130	Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)
140	Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)
150	Sparse (<15%) vegetation
160	Closed to open (>15%) broadleaved forest regularly flooded (semi-permanently or temporarily) - Fresh or brackish water
170	Closed (>40%) broadleaved forest or shrubland permanently flooded - Saline or brackish water
180	Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil - Fresh, brackish or saline water
190	Artificial surfaces and associated areas (Urban areas >50%)
200	Bare areas
210	Water bodies
220	Permanent snow and ice
230	No data (burnt areas, clouds,…)

Table 7: List of GlobCover classes.

To downscale the local agricultural emissions, we only make use of the classes related to agricultural land use. More in detail, we consider the classes 11, 14 and 20 respectively corresponding to “Post-flooding or irrigated croplands (or aquatic)”, “Rainfed croplands” and “Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)”. The downscaling routine resamples the input data to a 0.01 degree resolution, and uses this proxy in the methodology outlined in section 3.2.

3.3.3. Industrial: GlobCover

Also for the industrial proxy, we rely on the GlobCover land use data. This database has only one type of artificial structures, a feature which it unfortunately shares with other land use data sets for China. As an alternative proxy, we have considered the MODIS land use data (which also only contains one type of artificial structures). There are some local differences between both

datasets, but in general, there is a good correspondence between both datasets. The largest differences are observed in the Pearl River Delta, where the city of Guangzhou is much larger in the MODIS dataset, while the reverse happens for Hong Kong and Shenzhen. A comparison with satellite images indicates that both datasets still underestimate the actual size of the (current) built-up area. Because both datasets have similar drawbacks (although in different regions), we opt for the GlobCover dataset since it provides internal consistency with the proxy for the agricultural emissions.

To downscale the industrial emissions, we make use of the class 190 "Artificial surfaces and associated areas (Urban areas >50%)" which contains both the location of industrial sites and residential built-up areas. The downscaling routine resamples the input data to a 0.01 degree resolution, and calculates which fraction of the built-up areas in the total domain is located in each grid cell.

There are two drawbacks associated with the use of the GlobCover data as an industrial proxy. Firstly, the industrial emissions are probably spread over a slightly too small area, since the built-up area in GlobCover underestimates the actual extent of the Chinese cities. Secondly, the proxy does not distinct between residential and industrial built-up areas, and industrial emissions are thus spread over entire cities (instead of only over industrial areas). However, these issues also occur in other land use datasets which cover the entire domain at a suitable resolution. In the absence of a more detailed dataset, we thus rely on the GlobCover dataset.

3.3.4. Residential: WorldPop database

To downscale the residential emissions, we make use of a population density dataset. We use the results of the WorldPop project (http://www.worldpop.org.uk/). The project provides various datasets with different resolutions, the most detailed being 100m x100m gridded data (although the actual resolution is slightly lower). We start from a dataset for Asia at a resolution of 0.0083 degrees. We resample the input data to a 0.01 degree resolution, and calculate which fraction of the population in the total high resolution domain is living in each grid cell.

3.3.5. Powerplants: Enipedia

To infer the locations of the most important power plants, we make use of the Enipedia project of TU Delft (http://enipedia.tudelft.nl/wiki/Main_Page) . This database combines many open-source datasets, among which the well-known Carma (Carbon Monitoring for Action) dataset (http://carma.org/). Some manual processing of the raw on-line data has been performed. At first, we have removed all power stations without coordinates or with highly unrealistic (and probably faulty) emission data. Moreover, we neglect all "green" power plants (for instance wind and water energy) and all plants with zero CO_2-exhaust (we assume that they are green). Finally, we are mainly interested in the large plants, therefore we have removed all the power plants with an exhaust smaller than 1% of the maximal exhaust, as they will contribute a negligible fraction to the total emissions. In total, we retain 818 power plants in China.

There are still some issues with this processed dataset. THU did a comparison between the Carma database (which is contained in the Enipedia dataset) and their own high resolution dataset, which

is based upon the China coal-fired Power plant Emissions Database (CPED) and validated using visual inspection. They found that the total emissions of both datasets more or less match each other, but that many small power plants are missing in the Carma database. However, as we are mainly interested in the large power plants and even discard the small ones, we do not consider this as a major issue. Secondly, the locations of many power plants in Carma are incorrect (Li et al., 2015), which is much more of a problem when using the coordinates for downscaling the large-scale emissions.

To analyse the importance of this issue. we have compared the total yearly emissions (for 2012) contained in the MEIC-inventory cell by cell with the power output contained in Enipedia. The results are shown in , for NO_x and $PM_{2.5}$. There is some correlation in the plots (R^2 = 0.42 for NO_x), but also a lot of scatter. The scatter could have many underlying reasons, but probably it is mostly related to the uncertainties in the coordinates contained in both datasets, and in the (bad) correlation between power output and NO_x-emissions. The analysis has also been performed for SO_2, but this leads to even lower correlations (R^2 = 0.2).

We have repeated this analysis using only the power plants located in the regions for the high resolution emission inventory (as we will only use the proxy data for these regions). This does improve the correlation, but not by very much, as is shown in Figure 11. A worrying aspect are the large number of 0.25x0.25 degree grid cells in which there are no powerplants according to Enipedia, but in which there are power plants according to the MEIC inventory and vice versa. A graphical analysis mainly indicates problems for the Pearl River delta and (in lesser extent) for the Yangtze Delta. For the Beijing area, the comparison is much better.

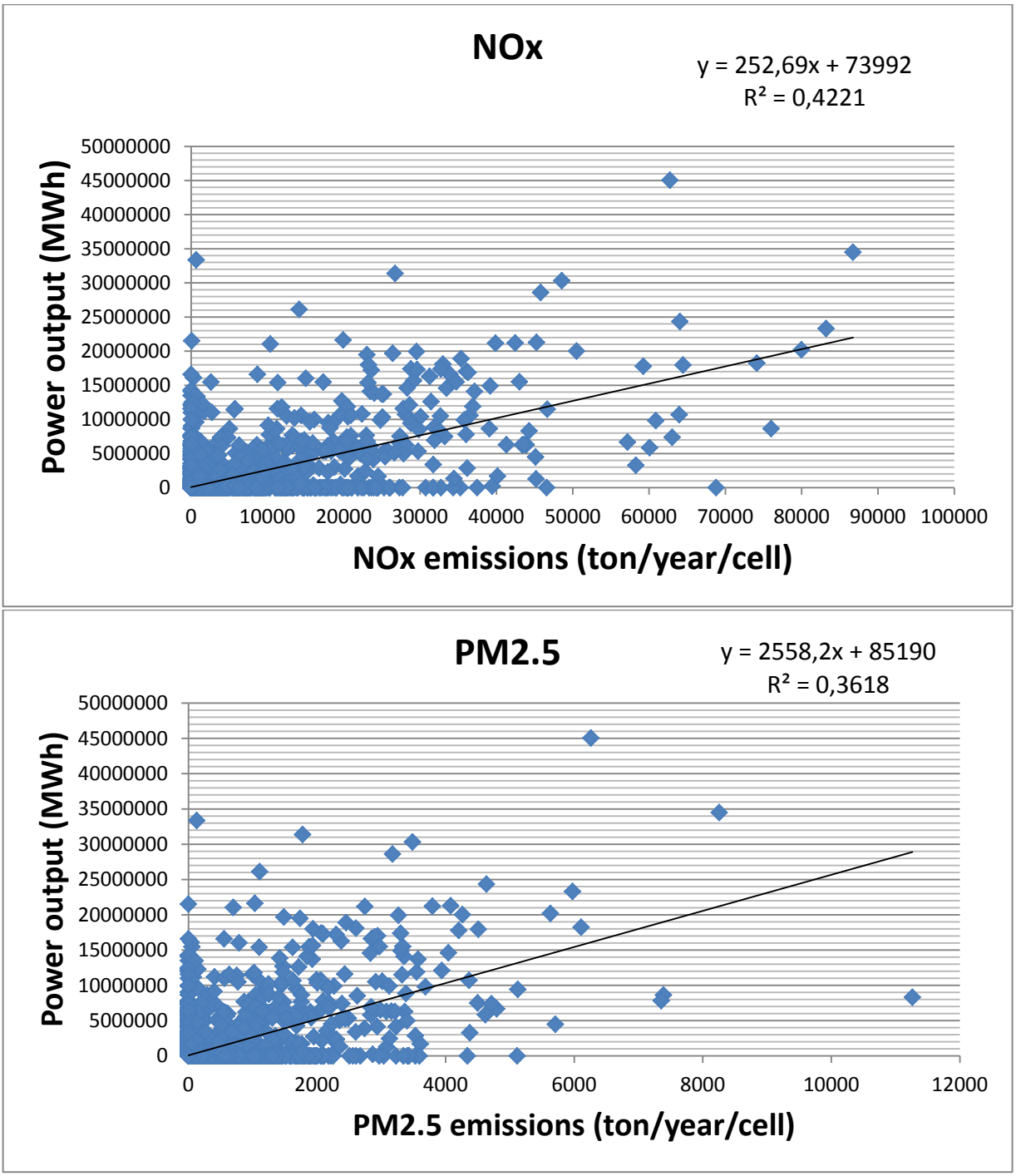

Figure 10: Power output per MEIC-grid cell in function of the MEIC power emissions for the same grid cell. The above figure presents the data for NO_x, below are the data for $PM_{2.5}$. In this figure, all the grid cells of the MEIC-inventory that lay in the MarcoPolo domain are considered.

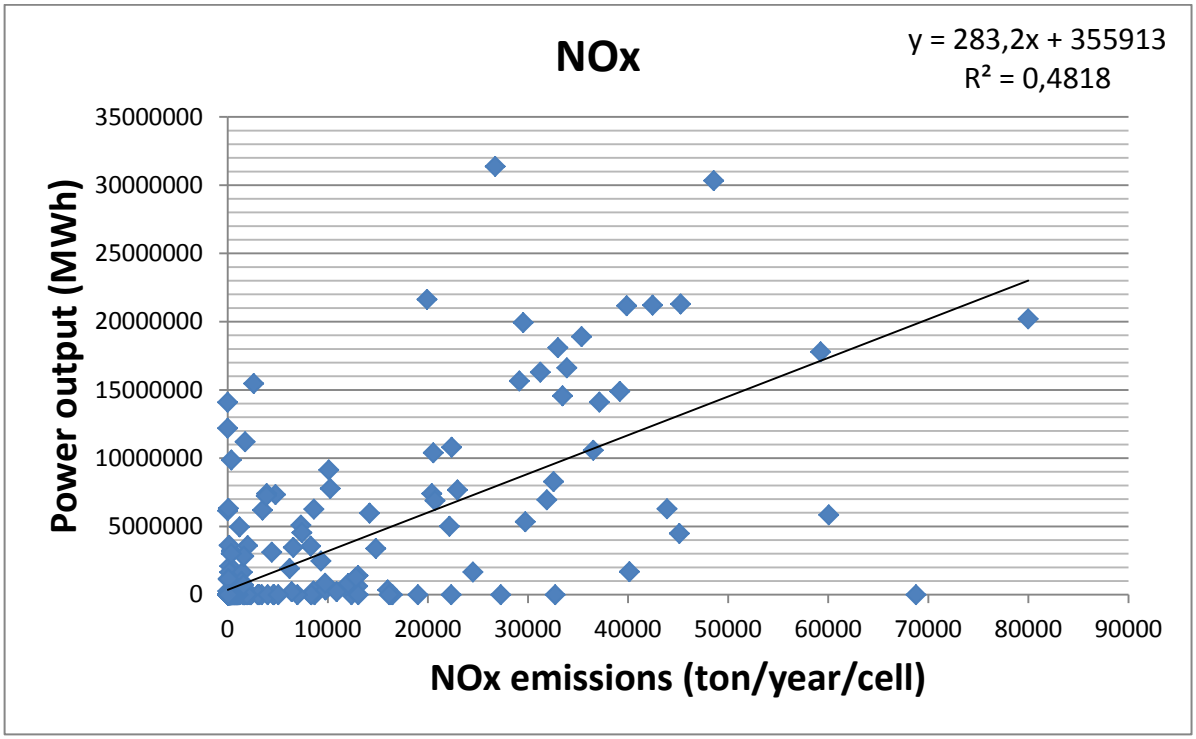

Figure 11: Power output per MEIC-grid cell in function of the NOx-MEIC power emissions for that grid cell, for the original Enipedia dataset. In this figure, only the grid cells of the MEIC-inventory that lay in one of the domains for the downscaling are considered.

The analysis shows that there are major issues with the Enipedia database (and probably also with the CPED data set), and that especially the coordinates for the power plants are sometimes inaccurate. Hence, we have processed the database manually for the three high resolution domains. Coordinates have been examined using visual inspection based on satellite images. Often, the incorrect coordinates were replaced with correct coordinates contained in other databases, as for instance the SourceWatch database,[5] which provides accurate locations for most power stations in the high resolution domain. Small powerstations for which no accurate location could be inferred, have been discarded. Duplicates have also been removed. In total, we have retained 74 power stations (44 in the Yangtze River Delta, 17 in the Pearl River Delta, and 13 in the Beijing-Tianjin region). The final dataset has anew be compared with the data contained in the MEIC-inventory. There are stille quite some incompabilities between both datasets (f.i. powerplants that are only present in one of the datasets), but the general accuracy has increased. Results are shown in Figure 12.

[5] http://www.sourcewatch.org/index.php/SourceWatch

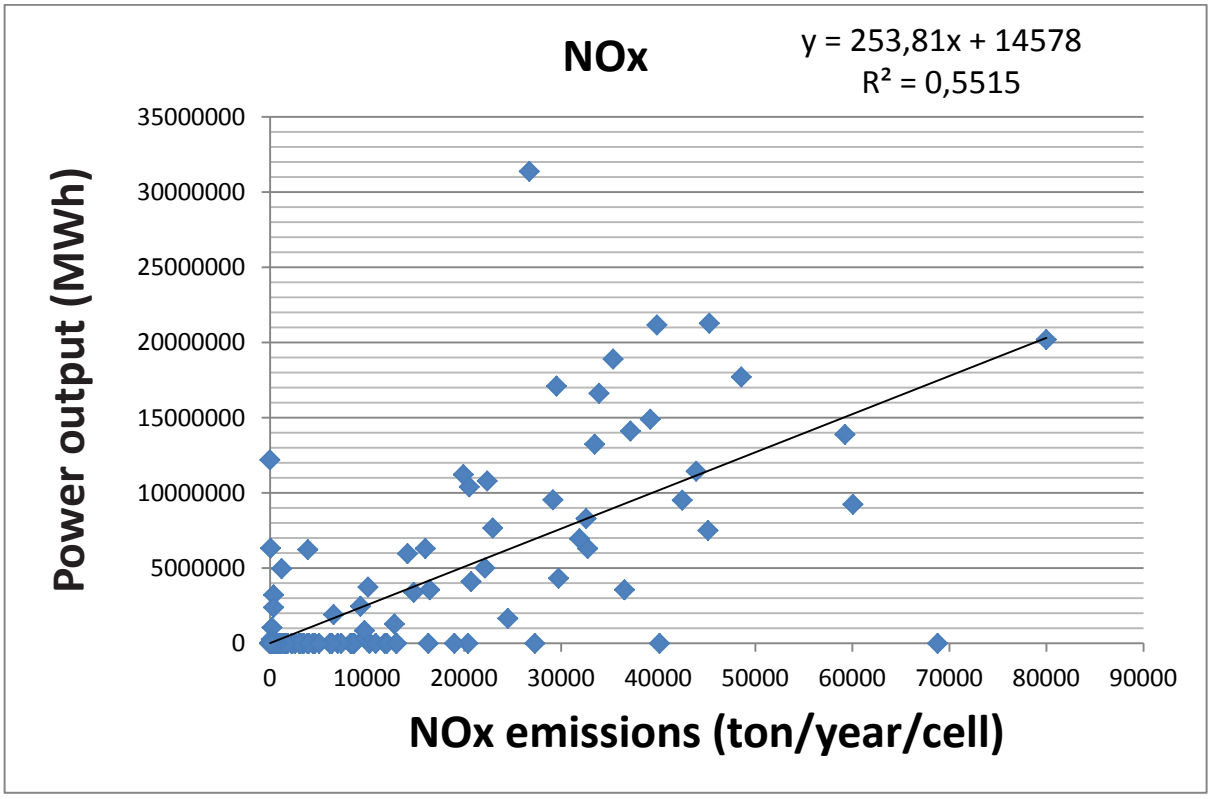

Figure 12: Power output per MEIC-grid cell in function of the NOx-MEIC power emissions for that grid cell, for the manually improved dataset. In this figure, only the grid cells of the MEIC-inventory that lay in one of the domains for the downscaling are considered.

For most of the 818 power plants in China, the Enipedia database provides the CO_2-exhaust and the power output, both of which could be used as a proxy in the downscaling. Since there is a very good correlation between the exhaust and the power output, as visualised Figure 13, the choice between both is not crucial. Hence, we choose the CO_2-emissions as a proxy. The emissions for the low resolution grid cells are thus spread over the high resolution grid cells according to the procedure outlined in section 3.2. Since there are (for the three domains in total) 74 power plants, the emissions of the power sector are spread over 74 cells. This procedure causes high concentrations in the grid cells in which power stations are located. Given the size of the power plants at stake, these values are presumably realistic.

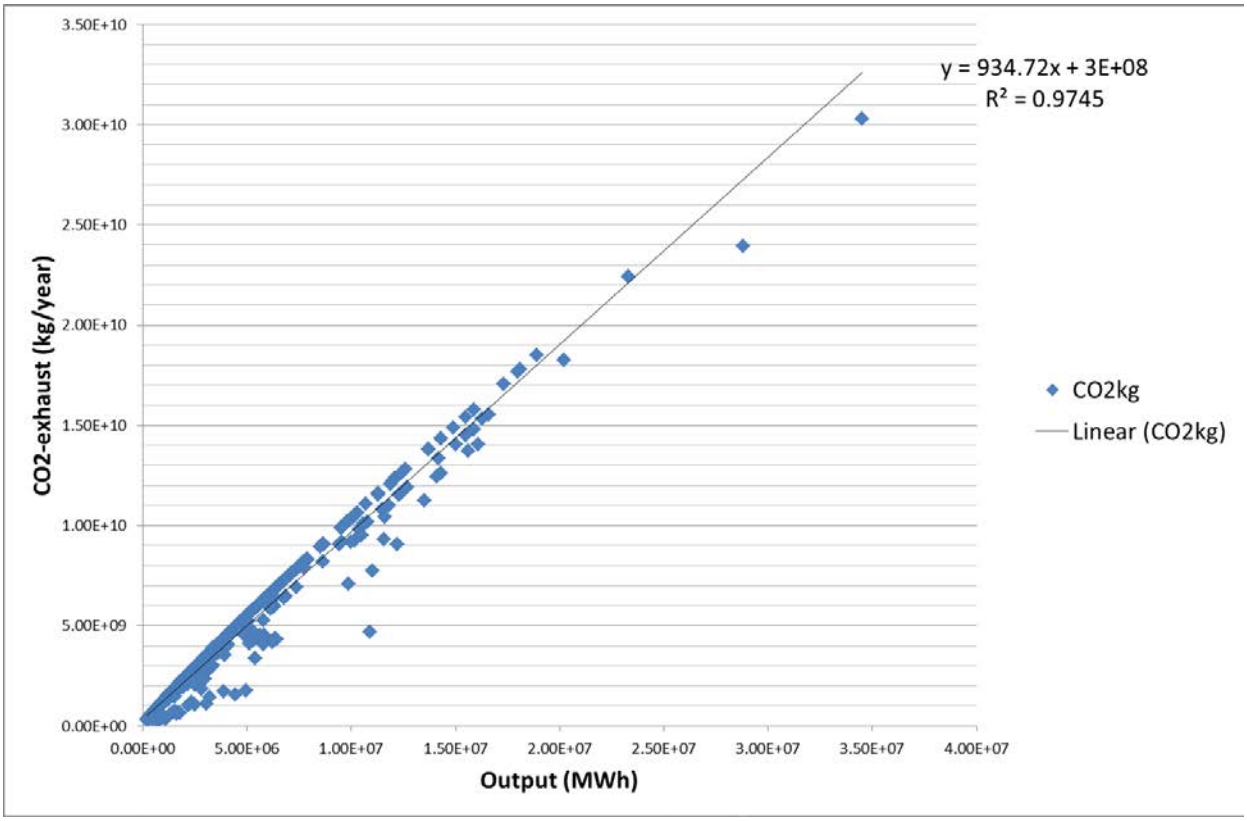

Figure 13:CO_2-exhaust (in kg/year) of the 818 power plants considered in this study as a function of the power output (in MWh) of these power plants.

3.3.6. Shipping

Since it is hard to locate the exact locations of the berths and shipping lanes, we simply make use of the land-sea mask to spread the shipping emissions over the high resolution grid cells according to the fraction of water in the grid cell. Since only a small portion of the high resolution domains covers sea grid cells, we assume that this is a reasonable assumption. Note that for general high resolution domains, serious deviations could occur, especially around shipping lanes in non-coastal waters.

4. Data format description

The low resolution inventory and the three high resolution inventories are provided in separate NetCDF-file. The file size is approximately 80 MB for the low resolution inventory, 210 MB for the high resolution inventory for Beijing and the Pearl River Delta and 630 MB for the inventory for the Yangtze River Delta.

The file contains three dimensions: longitude (in degrees east), latitude (in degrees north), and time (in days since January 1st, 2014). The emissions are stored in many variables (in Mg/cell/month), with each variable containing the emissions for one pollutant and one sector. The exact definition and formats should be self-explaining using the attributes contained in the file.

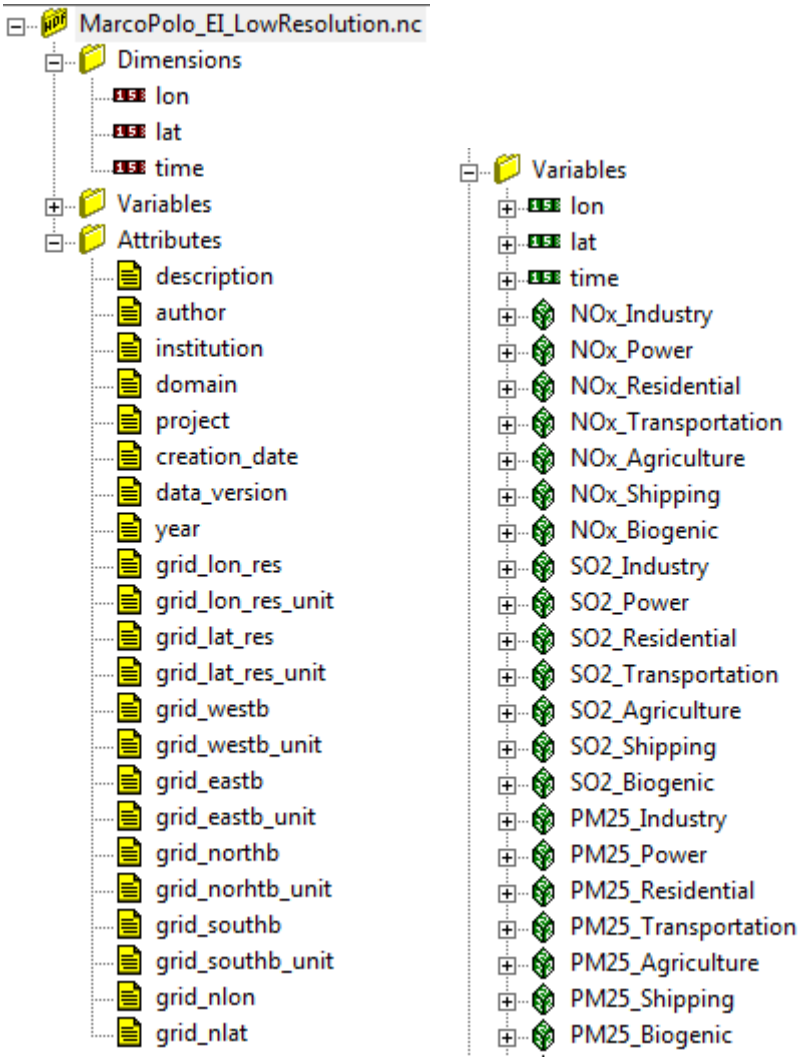

Figure 14: Lay-out of the NetCDF-file containing the MarcoPolo inventory. Only a subset of the variables is shown.

References

C. C. Chang, 2014. *2012 Hong Kong Emission Inventory Report*,

Fu, Q.-Y., Yin, S. & Zhang, J., 2012. On the ship pollutant emission inventory in Shanghai port. *Journal of Safety and Environment*, 05.

Fung, F. et al., 2014. *Prevention and control of shipping and Port Air emissions in china*, Available at: www.suerossi.com [Accessed June 14, 2016].

Hong, Z. et al., 2013. *The Competitiveness of Global Port-Cities: The Case of Shanghai, China*, OECD Publishing.

Huang, X. et al., 2012. A high-resolution ammonia emission inventory in China. *Global Biogeochemical Cycles*, 26(1), p.n/a–n/a. Available at:

http://doi.wiley.com/10.1029/2011GB004161 [Accessed June 14, 2016].

Li, M. et al., 2015. MIX: a mosaic Asian anthropogenic emission inventory for the MICS-Asia and the HTAP projects. *Atmospheric Chemistry and Physics Discussions*, 15, pp.34813–34869. Available at: http://www.atmos-chem-phys-discuss.net/15/34813/2015/.

Siqi, Y. et al., 2014. Marine emission inventory and its temporal and spatial characteristics in Guangdong Province. *Acta Scientiae Circumstantiae*, (3), pp.537–547.

Yang, J. et al., 2015. [Marine Emission Inventory and Its Temporal and Spatial Characteristics in the City of Shenzhen]. *Huan jing ke xue*, 36(4), pp.1217–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26164893 [Accessed June 14, 2016].