

## Comparison and assessment of two emission inventories for the Madrid region

M. Vedrenne, R. Borge, D. De la Paz,  
J. Lumbreras & M. E. Rodríguez  
*Department of Chemical and Environmental Engineering,  
Technical University of Madrid (UPM), Spain*

### Abstract

Emission inventories are databases that aim to describe the polluting activities that occur across a certain geographic domain. According to the spatial scale, the availability of information will vary as well as the applied assumptions, which will strongly influence its quality, accuracy and representativeness. This study compared and contrasted two emission inventories describing the Greater Madrid Region (GMR) under an air quality simulation approach. The chosen inventories were the National Emissions Inventory (NEI) and the Regional Emissions Inventory of the Greater Madrid Region (REI). Both of them were used to feed air quality simulations with the CMAQ modelling system, and the results were compared with observations from the air quality monitoring network in the modelled domain. Through the application of statistical tools, the analysis of emissions at domain and cell level, it was observed that the National Inventory showed better results for describing on-road traffic activities and agriculture, SNAP07 and SNAP10. The accurate description of activities, the good characterization of the vehicle fleet and the correct use of traffic emission factors were the main causes of such a good correlation. On the other hand, the Regional Inventory showed better descriptions for non-industrial combustion (SNAP02) and industrial activities (SNAP03). It incorporated realistic emission factors, a reasonable fuel mix and it drew upon local information sources to describe these activities, while NEI relied on surrogation and national datasets which led to a poorer representation. Off-road transportation (SNAP08) was similarly described by both inventories, while the rest of the SNAP activities showed a marginal contribution to the overall emissions.

*Keywords: air quality modelling, emission inventory, scale interaction.*



## 1 Introduction

Emission inventories are geographically specific databases. Therefore, the scale upon which they are compiled is an important issue to study. A given region might be described by several inventories differently, depending on the scale upon which each one has been built [1].

Emission inventories may be used for different objectives, namely policy purposes or scientific activity. To this respect, policy makers might be interested in the fact that a given inventory is recognised as valid by the institutions involved in the policy making process. When inventories are analysed under a scientific scope, the evaluation of inventory data will depend on their accuracy describing the emissions and their suitability to support scientific studies such as air quality modelling [2]. Under both perspectives, the scale factor is a very challenging issue. The difference in the development of scaled inventories lies in the level of detail of the input data, considered sources, hypotheses, and analysed parameters. National inventories require a broader approach for emissions estimation, as they encompass sources related to larger geographic areas, including air and maritime transport as well as the national energy grid. An emission inventory at national level also depends on many factors such as vehicle technology, socio-economic characteristics, transport policies, etc. [3].

In order to evaluate the scale issue, the use of an air quality simulation approach seems reasonable since air quality models are a valuable tool that links emission inventories and measurements at air quality monitoring stations [4]. The robustness of the inventory is strictly related with the consistency, within an acceptable uncertainty margin, between the air quality model predictions and real observations [5]. If this does not occur, it is necessary to identify the causes of the discrepancies between data pairs. This evaluation process may include the analysis of emission compilation methods (bottom-up, top-down) as well as the original sources of the emission factors used to construct the inventory through testing the sample set for appropriateness, reproducibility, statistical variance, etc. The assessment of the robustness of the survey and compilation methodologies might give strong hints about systematic differences [6].

## 2 Materials and methods

This study relies on the baseline hypothesis that the accuracy of an emission estimate is related to the degree of correspondence between ambient observations and the results of an air quality model (AQM) fed with a particular emission dataset. The analysis of the differential response of the AQM at representative points (monitoring stations) along with the analysis of the difference on alternative emission estimates is used to find out which of the underlying methods and information sets used in any of the inventories corresponds better with reality.



## 2.1 Modelled domain

The Greater Madrid Region (GMR) was modelled as a grid of 2304 cells (48×48) of a 4 km size each. This grid was centred approximately between the 5.0° W–2.5° W longitudes and the 39.5° N–41.5° N latitudes. The studied region included the entire Autonomous Community of Madrid (CAM) as well as some portions of the provinces of Avila, Segovia, Valladolid, Guadalajara, Cuenca and Toledo, (fig. 1). The before mentioned geographic domain was studied from January 1<sup>st</sup> to December 31<sup>st</sup>, 2007 with an hourly resolution for the 365 days of the year.

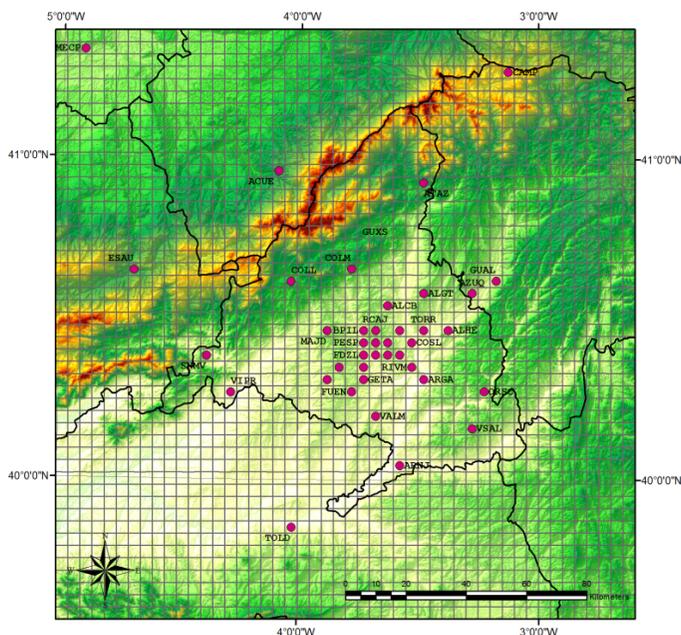


Figure 1: Modelled domain and monitoring network over the GMR.

## 2.2 Emission inventories

This study compared two official emission inventories available for the GMR. The National Emission Inventory is compiled by the Spanish Ministry for the Environment, and will be referred with the acronym NEI (National Emission Inventory) throughout this work. The Regional Emission Inventory has been produced by the Environment, Housing and Territory Council of the Autonomous Community of Madrid. References to this work will be made through the acronym REI (Regional Emission Inventory).

### 2.3 Air quality modelling

Two annual runs were carried out over the GMR domain to find out any differences in using alternative inventories when estimating air quality. The implemented AQM system is the Community Multiscale Air Quality Model (CMAQ) [7]. Emissions were processed by the Sparse Matrix Operator Kernel Emissions (SMOKE) modelling system as described in Borge *et al.* [8]. The meteorological fields needed to simulate the air pollution processes have been generated through the Weather Research and Forecasting (WRF) modelling system. Both emission datasets and meteorology corresponded to the year 2007. The applied initial and boundary conditions were the same for both runs and came from a 4-domain nested simulation as described in Borge *et al.* [9].

### 2.4 Pollutants

Four pollutants were followed throughout the domain: nitrogen oxides (NO<sub>x</sub>), sulphur dioxide (SO<sub>2</sub>), and two fractions of particulate matter, PM<sub>10</sub> and PM<sub>2.5</sub>. Although the reported pollutant at the emission inventories are total nitrogen oxides (NO<sub>x</sub>), the assumption that mostly the entire NO has been transformed into NO<sub>2</sub> will be considered as valid in order to make comparisons easier. Such an assumption seems reasonable since at urban entourage NO rapidly oxidises to NO<sub>2</sub>.

### 2.5 Statistical analysis

Using statistic indicators is useful to analyse the general performance of the inventories due to the great number of data being processed. The selected statistic indicators for this work aimed to characterise the performance of the modelling system under different perspectives (table 1). First, the ability to reproduce the temporal and geographical evolution of the prediction was evaluated by the Pearson's correlation coefficient (*r*). Second, the characterisation of the average behaviour through the mean bias (MB). Finally, the diagnosis of the general tendency for over or underprediction was examined from the mean fractional bias (MFB) values.

Table 1: Statistics used for model evaluation and experiment comparison.

Statistic	Units	Definition
Mean Bias (MB)	(ppm/μg·m <sup>-3</sup> )	$MB = \frac{1}{N} \cdot \sum_{i=1}^N (M_i - O_i)$
Mean Fractional Bias (MFB)	%	$MFB = \frac{2}{N} \cdot \sum_{i=1}^N \frac{M_i - O_i}{M_i + O_i}$
Pearson's correlation coefficient ( <i>r</i> )	-	$r = \frac{\sum_{i=1}^N (M_i - \bar{M})}{\left[ \sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \right] \cdot \left[ \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \right]}$

Additionally, Taylor diagrams were drawn as to have a better overview of the general behaviour of the simulation looking to find a compromise between the complexity of the evaluation and the need of having simple and straightforward diagnosis tools. These diagrams were drawn according to Thunis *et al.* [10].

## 2.6 Air quality monitoring stations

Comparisons were carried out with observations coming from a 55-monitoring-station network conveniently placed over the GMR (Figure 1) and measuring the before mentioned pollutants on an hourly resolution. Four types of stations were evaluated: *traffic*, *urban background*, *industrial* and *rural*. The type of station would condition the representativeness and validity of the conclusions being drawn.

## 2.7 Emission analysis at domain and cell-level

This analysis was conducted in *two* main aspects that should not be studied alone, namely (i) absolute emissions ( $E_{SNAPi}$ ) for each pollutant and SNAP group, in yearly emitted metric tons (ton/year) and (ii) the relative emissions of every SNAP group ( $P_{SNAPi}$ ).

## 3 Results

In general, it was shown that there are significant differences between both inventories describing the very same geographic domain. For the general purposes of this work, the term *well correlated station* means a location at which

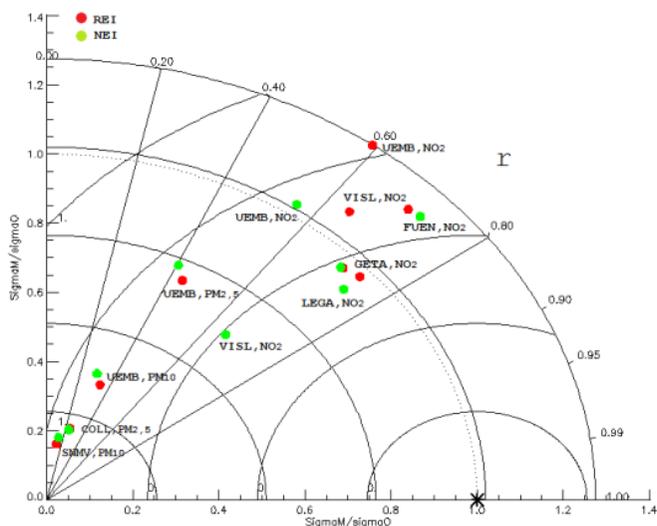


Figure 2: Taylor diagram for the analysed stations.

both emission inventories perform very similarly in addition to a fair statistical correlation degree. A *badly correlated station* is an emplacement at which the two emission inventories behave differently, regardless of its statistical correlation degree.

Figure 2 shows the well correlated stations chosen for the description of the GMR, where it is also evident that the points for both inventories, NEI and REI lie close to each other. For nitrogen dioxide, most of these well-correlated stations showed high correlation coefficients ( $r$ ). As for particulate matter ( $PM_{10}$ ,  $PM_{2.5}$ ), stations classified as well-correlated according to the before mentioned criterion, exhibited poor correlation coefficients in every case below  $r = 0.4$ . Furthermore, these stations presented systematic errors ( $\sigma_M < \sigma_0$ ) and negative mean bias values (MB), as Figure 2 suggests. However, they were selected for presenting the best available performance for particles.

Badly correlated stations in most cases exhibited both, high departures between points and low correlation coefficients. In the case of particles, the presence of highly negative mean bias values (MB) was an additional feature that was regarded to classify stations as such.

The chosen stations for sulphur dioxide have not been presented in Figure 2 because of their high mean bias values, which would affect the scale of the diagrams. For example, Arganda del Rey (ARGA) shows a mean bias value of 13.43 ppb for the REI, with a very low correlation coefficient ( $r = 0.02$ ). On the contrary, Leganés (LEGA) presents a high correlation coefficient for both inventories ( $r = 0.60$ ).

## 4 Discussion

### 4.1 Non-industrial combustion plants (SNAP02)

This group includes combustion devices with low thermal capacities namely those used for heat generation at residential and service buildings. Through inspection, it seems that both, the use of boilers at commercial and institutional sectors (020103) as well as in residences (020302) are the cause of the observed distortions. The analyses for this activity group have been carried out over the cell where the Urbanización Embajada (UEMB) station (traffic type) is located. Nitrogen dioxide emissions produced by the problematic activities have been differently accounted by both inventories. While the REI computed a total emission of 4.1 ton/year, the NEI quantified a threefold emission, namely 12.4 ton/year.

The REI apportioned 5% of such emissions to the commercial and institutional sectors while the rest, a 95% was assigned to residences. On the other hand, the NEI assigned a 15% to the first activity and an 85% to the latter. A similar behaviour was observed with  $PM_{10}$  and  $PM_{2.5}$  emissions at this precise location. The NEI estimated  $PM_{10}$  emissions to be 13.2 ton/year and the REI a total of 4.1 ton/year, while for  $PM_{2.5}$  the emissions were 12.1 ton/year and 6.2 ton/year respectively. The share of each respective sector is exactly the same as the one for  $NO_2$ .

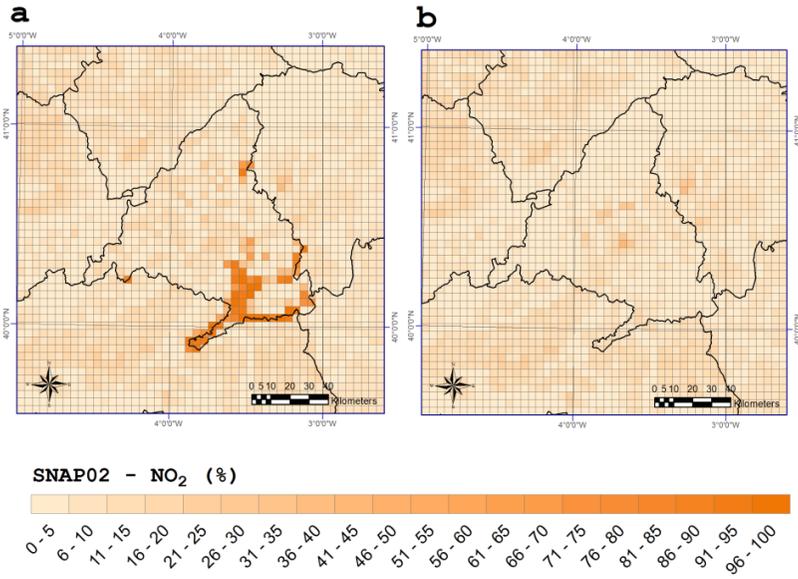


Figure 3: Relative SNAP02 emissions ( $P_{SNAP02}$ ) for NO<sub>2</sub> over the GMR according to a) REI and b) NEI.

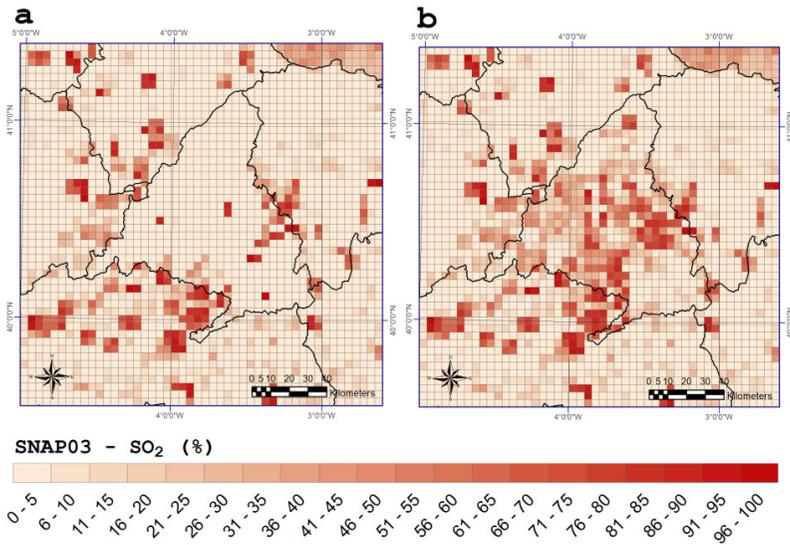


Figure 4: Relative SNAP03 emissions ( $P_{SNAP03}$ ) for SO<sub>2</sub> over the GMR according to a) REI and b) NEI.

Slightly differences were observed for SO<sub>2</sub>, for which the NEI distributed 52% of the emissions for the commercial and institutional sectors and 48% to residential heating for a total emission rate of 56.9 ton/year. Conversely, the REI maintained the same proportions between sectors even for this pollutant, of which 61.3 ton/year are produced.

Figure 3 shows the spatial distribution of  $P_{SNAP02}$  for NO<sub>2</sub>, for the entire GMR. As it is evident, the REI is able to differentiate between cells that are highly influenced by SNAP02, basically suburban and rural locations where road-traffic tends to contribute less; the NEI on the other hand, showed a uniform contribution degree of SNAP02 throughout the studied domain. The fact that this inventory is not able to differentiate between zones is an indicator of its poor resolution degree for the chosen scale. Yet the absolute emissions ( $E_{SNAPi}$ ) for each cell have been considered and evaluated to complement the analysis.

In general, the differences observed in this category are a consequence of the fact that the NEI considers consumptions of a representative Spanish fuel mix composed of coal, natural gas, LPG, among others. On its behalf, the REI has made a series of assumptions which distributed certain fuel types exclusively to certain sectors; in this case for the residential and commercial sectors LPG and natural gas were assigned as the only fuels being consumed.

#### 4.2 Industrial combustion plants (SNAP03)

The analysis for this SNAP group was best described by the behaviour of SO<sub>2</sub> emissions. There are indeed differences between inventories as depicted by the spatial distribution of  $E_{SNAP03}$  at cell level in Figure 4. While the REI inventory identifies SNAP03 contribution at much localised points, the NEI inventory considers a heavy contribution of SNAP03 on SO<sub>2</sub> emissions over a wide area of the domain. This very first assumption might not be completely right, since the Madrid region is not reputed for hosting extensive industrial zones. The analysis was carried out over Arganda del Rey (ARGA), which is of industrial type. The NEI accounted for this location a total of 94.9 ton/year, of which combustion plants (030103) had a 41%, gas turbines (030104) an 11%, stationary engines (030105) and plaster furnaces (030204) a 6% each and lime processes (030312) a 34%. On the contrary, the REI computed a total emission of 897.9 ton/year for this location which is 9 times higher than the NEI value and is caused in a 99% by lime processes.

It seems that the NEI has the provincial level as the maximum data resolution. To this respect, emission processing has been carried out through a spatial allocation procedure based on surrogate data, as described in Borge *et al.* [8] applied over industrial and commercial land uses specified by the CORINE land cover database. The REI relied on a direct on-site activity variable collecting campaign, accomplished through the use of questionnaires and surveys such as PRTR and the request of private information.

This spatial allocation process is deemed to be the cause of the large discrepancies observed between inventories, experienced by the high overestimations suggested by the MB values for the NEI against the REI. Hence, the use of surrogate disaggregation is not recommendable for finer resolutions,

encouraging the use of local or regional emission inventories instead, when available. Typically,  $\text{SO}_2$  is not a good pollutant to be surrogated since it is known that few large point sources dominate the spatial emission pattern for such pollutant.

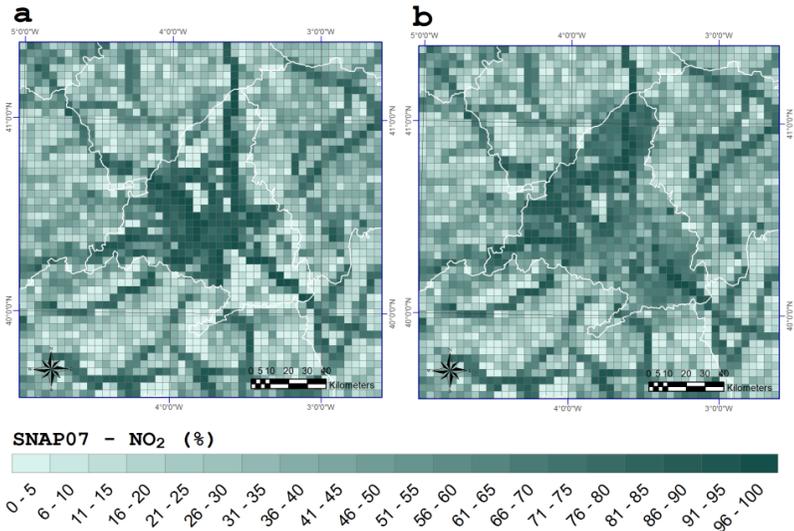


Figure 5: Relative SNAP07 emissions ( $P_{SNAP07}$ ) for  $\text{NO}_2$  over the GMR according to a) REI and b) NEI.

### 4.3 Road traffic (SNAP07)

Emissions coming from road traffic are the most relevant throughout the entire GMR. Figure 5 shows the spatial distribution of the SNAP07 contribution ( $P_{SNAP07}$ ) at cell level for  $\text{NO}_2$ , for the studied domain. The NEI shows lower SNAP07 percentages for the city centre and cells where highways are located show moderate values; the REI on the contrary shows heavy SNAP07 contributions along highways and the city centre, while no continuity with the neighbouring provinces is evident.

The Getafe (GETA) station is clearly classified as a well-correlated location. It is also a traffic-type station and is located at a zone with intense vehicle circulation. Passenger cars under highway driving (070101) have a 12% share at the NEI and 13% at the REI, accounting both a 120.4 ton/year emission. Passenger cars under urban driving (070103) are heavily considered by the NEI (328.5 ton/year–33%) while the REI made an analogous consideration with heavy duty vehicles under highway driving (357.7 ton/year–38%). Although these categories were accounted differently, the total emissions were computed similarly: the NEI reporting a 974.5 ton/year rate and the REI 923.5 ton/year.

On the contrary, for the rural station of Villarejo de Salvanes (VISL), the source of differences is basically the overwhelming consideration of HDV under highway driving by the REI. For this category, this inventory quantified

386.9 ton/year (69%) against 54.7 ton/year (20%) reported by the NEI, computing a total 554.8 ton/year for the first inventory and 270.1 ton/year for the second.

The main source of discrepancies between both inventories for this activity seems to be an excessive contribution of heavy duty vehicles in highway driving situation. Activity ratios (traffic intensity) should be more accurately depicted by the REI, since it is based on link-specific traffic intensities, yet it appears as if COPERT has not been run at road-level, being secondary emission factors used instead. This fact points out the need of a purely bottom-up computation approach, whenever detailed traffic data are available.

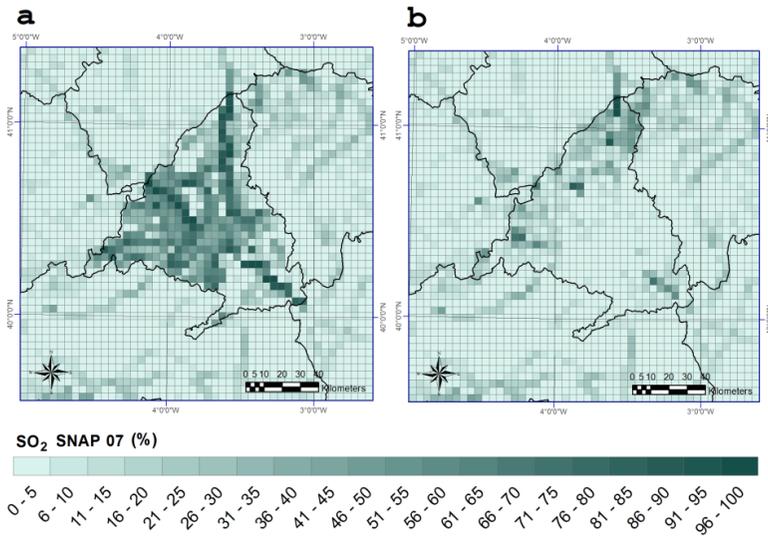


Figure 6: Relative SNAP07 emissions ( $P_{SNAP07}$ ) for SO<sub>2</sub> over the GMR according to a) REI and b) NEI.

Finally for SO<sub>2</sub>, an analysis carried out over the Fuenlabrada (FUEN) station (industrial type) exhibited sensitive differences between inventories for SNAP07. The total emissions reported by the REI are 20.4 ton/year while the NEI presented a 3.28 ton/year value, approximately sixfold. The NEI yielded a 271 ton emission for 2007, while the REI gave a 2.876 ton emission for the entire domain. The cause of discrepancies is the fact that the REI incorporated emission factors from the NEI for 2004 whose sulphur content in fuels was higher than for 2007. This is such a situation in which the predicted concentrations are directly proportional to the used emission factors, which led to high on-road SO<sub>2</sub> emissions, as it can be seen in Figure 6. This fact stresses out the need of keeping emission factors up to date with any modifications.

#### 4.4 Agriculture (SNAP10)

Although agricultural activities exhibited low absolute emissions, the considered activities showed sound differences among inventories. Analysis of PM<sub>10</sub> at the

rural station of San Martín de Valdeiglesias (SNMV) revealed that the NEI clearly considers more categories than the REI version; manure management (100500) is the main PM<sub>10</sub> source with a total contribution of 0.06 ton/year. If comparing the gross totals for both inventories, the emissions compiled by the NEI inventory are 60 times higher than the emissions considered by the REI, namely 0.001 ton/year. A tendency to underestimation was shown by the very low MFB at this location, -139.9 % for the REI and -135.8 % for the NEI. For the whole domain, the causes for differences within this category are basically related with the considered activity variables and emission factors.

## 5 Conclusions

According to the general issues discussed above, the estimates produced by the REI are better for almost every SNAP category except for road traffic (SNAP07), which is by far the most important emission category, and agriculture (SNAP10). At these weak points, the NEI provides a more accurate description. This study has shown that the sources of disagreement between the predicted concentrations obtained from the emission inventories and the actual observations are due to a series of methodological issues. Moreover, it was useful to gain some understanding on the differences and causes of discrepancies between NEI and REI. The data and information scale upon which each of the inventories has been based their assumptions is different; the NEI incorporated data from national and supra-national entities (national ministries, international agencies, etc.), the REI used information coming from national and regional sources.

When describing a given domain, available information for the scale of this domain should be chosen over information available for larger or smaller scales. Although spatial allocation is a common procedure to transfer these large-scale datasets to smaller scales, a limit scale exists under which the resulting information is no longer adequate.

Given the fact that emission inventory compilations need to be fed with a huge amount of data, the quality of this information is vital to guarantee their reliability. It is therefore essential to assure that the information is updated, valid and representative, so that in general emission inventories may be improved.

## References

- [1] Winiwarter, W., Kuhlbusch, T.A.J., Viana, M. & Hitzenberger, R. Quality considerations of European PM emission inventories. *Atmospheric Environment*, **43**, pp. 3819–3828, 2010.
- [2] Pulles, T. Can and should European inventories be improved? *Proceedings of EUROTRAC Symposium '99*, eds. P.M. Borrell & P. Borrell, Transactions on Ecology and the Environment, **28**, WIT Press: Southampton, pp. 9–15, 1999.



- [3] Alonso, M.F., Longo, K.M., Freitas, S.R., Mello da Fonseca, R., Marécal, V., Pirre, M. & Gallardo Klenner, L. An urban emissions inventory for South America and its application in numerical modelling of atmospheric chemical composition at local and regional scales. *Atmospheric Environment*, **44**, pp. 5072–5083, 2010.
- [4] Bellasio, R., Bianconi, R., Corda, G. & Cucca, P. Emission inventory for the road transport sector in Sardinia (Italy). *Atmospheric Environment*, **41**, pp. 677–691, 2007.
- [5] Norm–Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance. ASTM D6589, 2010.
- [6] Lim, B., Boileau, P., Bonduki, Y., van Amstel, A.R., Janssen, L.H.J.M., Olivier, J.G.J. & Kroeze, C. Improving the quality of national greenhouse gas inventories. *Environmental Science & Policy*, **2**, pp. 335–346, 1999.
- [7] Byun, D.W. & Schere, K.L. Review of the governing equations, computational algorithms, and other components of the Models–3 community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, **59**, pp. 51–77, 2006.
- [8] Borge, R., Lumbreras, J. & Rodríguez, M.E. Development of a high–resolution emission inventory for Spain using the SMOKE modelling system: a case study for the years 2000 and 2010. *Environmental Modelling and Software*, **23**, 1026–1044, 2008.
- [9] Borge, R., López, J., Lumbreras, J., Narros, A. & Rodríguez, M.E. Influence of boundary conditions on CMAQ simulations over the Iberian Peninsula. *Atmospheric Environment*, **44**, pp. 2681–2695. 2010.
- [10] Thunis, P., Georgieva, E. & Galmarini, S. A procedure for air quality models benchmarking. Joint Research Centre, Ispra, Italy, pp. 20–25, 2011.