Prioritising the sources of pollution in European cities: do air quality modelling applications provide consistent responses?

Bart Degraeuwe, Enrico Pisoni, Philippe Thunis

European Commission, Joint Research Centre (JRC), Ispra, Italy

Correspondence to: E. Pisoni (enrico.pisoni@ec.europa.eu)

Anonymous Referee #1

Received and published: 24 June 2020

The paper address relevant scientific modelling questions regarding the application and feasibility of air quality models to support air quality plans, which is quite new and advanced research. The methodology is appropriate and well-described. Nevertheless, there are some major (and minor) critical points that should be addressed before publication. They are listed below.

Major changes Line 70-77: please comment about the different baseline year for the emission and meteorology data and its implication on the results. Authors could highlight here (what is said at the end of the paper) that these differences in input data are interesting to the analysis of results since this is the usual way to define air quality plans, etc:…Nevertheless, these differences should be analysed in detail in order to understand their role in the SHERPA results. At least, authors should give some information regarding the validation (with observational data) of the 2 different models application (EMEP and CHIMERE). It is different if we are talking about 2 models with good performance or 2 models with completely different skills

We thank the reviewer for the comments. As suggested, we highlighted (at the beginning of Section 2) the importance of accounting for the variability of the input data (in terms of emissions, meteorology, etc…) when assessing potential impacts, something that is not performed in practice for air quality plans that are often based on a unique set of input. Also, as suggested by the reviewer, we added graphs in the Supplementary Material with model base-case validations (against observations) for the CHIMERE and EMEP configurations, that show similar skills. We however highlight here the fact that similar behaviour on base case concentrations do not imply similar source contributions (see Supplementary Material).

In section 2, we propose to add this text:

“Validation results for the two model configurations are presented in the Supplementary Material, showing similar performance (for CHIMERE and EMEP) in terms of comparison against observations. For CHIMERE the relation between predictions and observations at background stations is best characterised by a line through the origin with slope of 1.05, indicating a slight under-prediction. The standard error is 5.7 µg/m3 and uniform over the range of concentrations. The R2 is 0.9. Concentrations at traffic and industrial stations are underestimated by roughly 10%. For EMEP the relation between predictions and observations is best characterised by a power low with exponent 0.66. The data show a relative standard error is constant over the range of concentrations and equal to 30%. Traffic stations are under-predicted by 9% and industrial stations over-predicted by 7%.

It is important to note that the use of different input and model set-up (as listed before) represents the usual practice when air quality models are used, at the local scale, to assess the impact of air quality plans. This is why it is important (in this manuscript) to analyse how this choice influences the results and the subsequent design of an air quality plan; an issue that is often not tackled in *the scientific literature. Some differences in results might be due to trends in emissions and concentrations between 2010 and 2014. During this period, concentrations at Airbase stations decreased 2.2% per year on average (σ = 2.7%/year). Starting from these configurations, two set of SRRs have been built to model yearly average PM2.5 concentrations, based respectively on CHIMERE and EMEP data. The focus of this study is on PM2.5 yearly averages, as this is the pollutant with the highest impact on human health, and a key focus of policy makers in Europe. Before looking at the source allocation results, in the next section a brief description of the SHERPA methodology is proposed."*

Line 249: Again, regarding the sentence "Probably for these areas the differences in terms of meteorology, emissions, and their impact on concentrations through the air quality models, is higher than in other areas." This should be explored and analysed to better support the interpretations and conclusions and shouldn't be only a hypothesis to mention.
We elaborated a bit more in this section, linking also to the validation of the base case for 2 models setup, now presented in the Supplementary Material.
We propose to modify the paper as follows:
"Probably for these areas the differences in terms of meteorology, emissions, and their impact on concentrations through the air quality models, is higher than in other areas (in the Supplementary Material we show i.e. how the validation results, for the base case, are quite different for Spain in the 2 model implementation, and this could also have an impact on the correlation results shown in Figure)."

Minor changes Abstract: please add more details in the last sentence ("But there are also cases where results are contradictory". it is not mention which was the pollutant studied: PM2.5
We clarify now, in the abstract, that the paper is on PM2.5 yearly averages.
The paper has been modified as follows:
"But there are also cases where results (in terms of source allocation for PM2.5 yearly averages) are contradictory."

Line 18: all instead of al
We fixed the typo.

Line 19: Please review the sentence: "FAIRMODE (the Forum for air quality modelling in Europe) i.e. provides tools to assess the:"
The sentence has been reviewed, and modified as follows:
"For example, FAIRMODE (the Forum for air quality modelling in Europe) provides tools to assess the quality of the models, as the Model Quality Indicator and Model Quality Objective (Pernigotti el al., 2013b; Viaene et al., 2016)."

Line 82: please explain why PM2.5 is the focus, and why only this one
We specified that the focus is PM2.5, because we want to concentrate on the pollutants with the highest burden on human health. We also stress the fact that because a large number of sources contribute to PM2.5 concentrations, this is the most challenging pollutant to manage in air quality plans. It is therefore important to assess the different model contributions for that pollutant in particular.
This is how we propose to modify the text:
"The focus of this study is on PM2.5 yearly averages, as this is the pollutant with the highest impact on human health, and a key focus of policy makers in Europe. We also stress the fact that because a large number of sources contribute to PM2.5 concentrations, this is the most challenging pollutant to manage in air quality plans."

Line 145: please use subscript on the compound's chemical formulas
This issue has been fixed.

Line 182: "chimere.rank"
This issue has been fixed.

Line 184: what do the author mean with "for the different types of considered aggregations (area, sector, area-sector, ...)"? It is not obvious
This has been now better explained in the text. Text has been modified as follows:
"In addition to this, Figures 1 to 4 show the 'relative potentials' for the 2 models (S-CHIMERE and S-EMEP), for the different types of scenarios (considering emission reductions for the selected geographical area, for the chosen sector, or for combinations of geographical areas - sectors, …) and their corresponding correlations, for the same cities."

Line 185: Before starting to analyse the results for specific cities, the authors should identify and present which were the 4 cities selected (and their different behaviours associated)
Text has been rephrased to reflect the reviewer's comment. This has been explained in the text:
"We present results for 4 cities (Liege, Genova, Turin and Madrid) that are representative of the different behaviours found in our results."

Line 194/209/: Tables caption should be on the top of the table
This issue has been fixed.

Anonymous Referee #2

Received and published: 30 June 2020

Degraeuwe et al. describe the application of the SHERPA technique for determining Source/Receptor Relationships (SRRs) to the assessment of mitigation options for annual average PM2.5 concentrations in 150 European cities. SRRs are calculated from the output of two Chemical Transport Models (CTMs), CHIMERE and EMEP, which are commonly used in Europe for air quality simulation. The benefit of using pre-calculated SRRs instead of directly using the CTMs themselves is that the SRRs effectively emulate the relationship between emissions in each CTM grid cell and concentrations in other grid cells without having to simulate the full set of physical and chemical processes involved. SHERPA in particular provides an efficient way of calculating cell-to-cell SRRs without having to run a large number of training simulations, by making some assumptions about the degree to which grid cells can influence each other based on their separation. The authors use the two different sets of SRRs to determine the most effective options for mitigation of annual average PM2.5 in the 150 selected cities. They find that despite the use of different CTMs, emission inventories, and base meteorological years, the mitigation options identified for each of the cities are generally very similar. A few cases are however identified where the use of the different SRRs produces contradictory recommendations. While the topic is certainly within the scope of GMD, and the results as presented should be of interest to the community, it seems to me that the authors have gone to an extremely minimal amount of effort with this manuscript. The quality of the manuscript in its present form is not high enough to meet the standards that this reviewer would expect from GMD. Major revisions are required before the manuscript can be published.

Firstly, the authors appear to cite mostly their own work, or the work of their colleagues. This approach may be acceptable for an internal technical report, but in the peer-reviewed literature, authors must place their work in the broader context of the work that has come earlier, and clearly explain its novelty. The use of SRRs in air quality assessment has been prevalent for a long time, and SHERPA is not the only way that exists to calculate SRRs. It is not the job of this reviewer to perform the literature survey that the authors of this manuscript have neglected, so I will not suggest any specific references. But more context is certainly needed, and not only in the introduction; while the results are new and interesting, this is no excuse for not discussing them with appropriate reference to the existing literature.
As suggested by the reviewer, we extended the literature review. Note however that although many SRRs have been developed for air quality, we are not aware of a methodology that is flexible and fast enough to assess so many sensitivities at the urban scale. We stressed these points in the Introduction by adding the text below (for the discussion of the results and technical aspect, see our 'reply' to your next comment):
"The most precise way to use an AQM to produce source-receptor relationships for the model domain would be with an independent grid cell-to-grid cell approach. While this approach would allow a high level of flexibility in defining the zones over which emissions are spatially reduced, it involves simulating independently the effect of emissions changes in each single grid cell that has pollutant emissions in the model domain. It would require changing precursor emissions in individual grid cells one at a time and looking at the resulting change in concentrations in each receptor cell. While theoretically very simple, the resulting number of unknown parameters describing the transfers between source and receptor cells that need to be identified is very large. For example, for a domain with Ngrid = 50 × 50 and Nprec = 5, the identification of a maximum of 12,500 parameters would be required (if emissions occur in, and concentration changes need to be calculated for, all grid cells in the domain) to calculate the change of concentration at a given receptor cell. Therefore 12,500 equations, each connecting concentration changes and emission changes are necessary to identify these 12,500 unknown parameters. Because each of these

equations requires an independent AQM run, this independent grid cell-to-grid cell option is very costly, and simplifying assumptions that reduce the number of AQM runs are required (Clappier et al., 2015).

In GAINS ("Greenhouse gas - Air pollution Interactions and Synergies", Amann et al., 2011) the grid-cell to grid-cell relation is simplified by aggregating source cells into countries. The number of unknown parameters that need to be identified for one receptor cell equals the number of countries multiplied by the number of precursors. This system can only be solved if at least "N_prec x N_country" equations are available, requiring a similar number of independent AQM scenarios. In GAINS, about 50 countries and 5 precursors lead to the need of 250 independent AQM scenarios to identify 250 unknowns. Because they are derived from emission reductions at country level, these SRRs are not applicable at the urban scale.

In the RIAT + tool ("Regional Integrated Assessment Tool", Carnevale et al., 2014). Emissions are aggregated into 'quadrants' that are defined relatively to each grid cell within the domain. The 'quadrant' emissions and their related grid cell concentrations are then used to feed a neural network that delivers the SRR (Carnevale et al., 2009). Although the approach requires a limited number of full CTM simulations (around 20), the set-up of the SRR remains complex due to the need of implementing sophisticated neural networks.

In SHERPA (Thunis et al., 2016; Pisoni et al., 2017), a different approach is taken, that reproduces the grid cell-to-grid cell approach but does not require anywhere near as many AQM model runs. SHERPA assumes that the unknown parameters vary on a cell-by-cell basis but are no longer independent of each other. Instead these coefficients are assumed to be related through a bell shape function. With the SHERPA approach, the number of unknown parameters is then equal to 2 for each precursor and receptor cell. Consequently, for the five precursors of $PM_{2.5}$ (VOC, SO_2, NO_x, PPM and NH_3), ten independent AQM simulations are needed for a given receptor cell. Provided that they deliver independent information, the same AQM scenarios can be used to identify both parameters for all cells within the domain (see details in Pisoni et al. 2017). Based on these 10 CTM simulations the SHERPA approach allows to quickly assess the impact of emission reductions for many combinations of sectors, geographical areas and precursors. Because it is currently the only one existing to perform a systematic analysis in about 150 EU cities in terms of sectors and precursors, we use the SHERPA approach in this work to approximate two CTMs: CHIMERE and EMEP and compare their responses."

Secondly, for a technical journal such as GMD, the paper is extremely short on technical detail. In Section 3, the reader is referred to Pisoni et al. (2019) for all but a few of the relevant details. Of course the reference is appropriate in this section, but the paper should also contain enough detail to stand on its own. The authors need to summarise the key points from this earlier work. For example, readers need to know how the SHERPA technique differs from other approaches to calculating SRRs, and how well it has been shown to work. Have mitigation options identified with SHERPA been compared with actual CTM simulations of the same mitigation options? What are the strengths and weaknesses of the approach as identified by earlier work, and what are their implications for the present manuscript?

As suggested by the reviewer, we now provide more details on the methodology and validation results of the SRR. In particular, in the Supplementary Material, we added information on the base case validation for the 2 model set-up (validation against observations), and also on how the SRRs behave in comparison to CTM simulations (validation against CTM results). Please find attached to this reply, the Supplementary Material, with the aforementioned contents.

Furthermore, in the manuscript we propose to add this text (in Section 2), to better detail the technical capabilities of the SRR, and validation results:

"More details on the model simulations and settings can be found in Clappier et al., 2015 and Pisoni et al., 2019. Validation results for the two model configurations are presented in the Supplementary Material, showing similar performances (for CHIMERE and EMEP) in terms of

comparison against observations. For CHIMERE the relation between predictions and observations at background stations is best characterised by a line through the origin with slope of 1.05, indicating a slight under-prediction. The standard error is 5.7 µg/m3 and uniform over the range of concentrations. The R2 is 0.9. Concentrations at traffic and industrial stations are underestimated by roughly 10%. For EMEP the relation between predictions and observations is best characterised by a power low with exponent 0.66. The data show a relative standard error constant over the range of concentrations and equal to 30%. Concentrations at traffic stations are under-predicted by 9% and over-predicted at industrial stations by 7%. It is important to note that the use of different input and model set-up (as listed before) represents the usual practice when air quality models are used, at the local scale, to assess the impact of air quality plans. This is why it is important (in this manuscript) to analyse how this choice influences the results and the subsequent design of an air quality plan; an issue that is often not tackled in the scientific literature. Some differences in results might be due to trends in emissions and concentrations between 2010 and 2014. During this period, concentrations in Airbase stations decrease yearly by 2.2% on average (σ = 2.7%/year). Finally, differences can arise from the SRR approximation, even if (as shown in the Supplementary Material) validation against CTM simulations show similar results for the 2 considered model set-up. Starting from these configurations, two set of SRRs have been built to model yearly average PM2.5 concentrations, based respectively on CHIMERE and EMEP data."

I also have a couple of minor comments. It would be nice to see a short explanation of how the four cities shown in detail were chosen. It's good to see an example of a situation in which the approach works well, and a situation in which it doesn't (Liege and Madrid). But what about the other two cities (Genova and Torino)? Were these chosen to highlight specific points? Or for some other reason?

In section 5, we propose to add this text:

"Figures 1 to 4 show the 'relative potentials' for the 2 models (S-CHIMERE and S-EMEP), for the different types of performed scenarios (considering emission reductions for the selected geographical area, for the chosen sector, or for combinations of geographical areas - sectors, …) and their corresponding correlations, for the same cities. We present results for 4 cities (Liege, Genova, Turin and Madrid) that are representative of the different behaviours found in our results."
"

For the cases when the use of the two sets of SRRs from different CTMs yields different mitigation options, the authors take the position that their method is simply unable to explain the differences. I find this somewhat lazy. Actually the disagreement could point the way to targeted CTM simulations (or other analysis) designed to specifically understand the relevant processes. It would add a lot to the paper to see some more discussion of this.

We now better explain the possible reasons for disagreement, referring to the Supplementary Material. Even if further investigation would be required to understand precisely why these differences occur.

We propose to add this text, at the end of Section 5:

"The overall correlation map of Europe (Figure 6) shows that cities with the highest variability are mostly located in Spain, Northern Italy as well as the Baltic countries. For these areas, meteorological factors, emissions, and/or the impact of these input on concentrations in the air quality models is higher than in other areas. In the Supplementary Material we show i.e. how the validation results, for the base case, are quite different for Spain in the 2 model implementation, and this could also have an impact on the correlation results shown in the Figure."

Referee #3

Received and published: 7 July 2020

This paper presents a comparison between the results obtained with two different setup of the SHERPA Source Receptor Relationship (SRR): S-CHIMERE and S-EMEP. Each of these two SHERPA configurations is used to compute the impact of different emission reductions (per activity sectors, per areas and per precursors) for 150 cities in Europe. The authors compare all the impacts provided by the two SHERPA configurations to evaluate the variability resulting from the use of two model systems (CHIMERE and EMEP). This work is without any doubts very interesting because it provides information about the robustness of model results which could be directly used by decision makers to design abatement strategies. The authors take advantage of the capacity of SHERPA to simulate a very large number of scenarios concerning spatial as well as sectorial emission reductions. 150 cities have been considered and 100 scenarios have been computed for each of these cities. As far as I know, SHERPA is the only tool able of such performances and it is the first time that so many cities and scenarios have been tested. This is why I think that the most interesting results of this article concerns the analysis of all cities and all scenarios (graphic of figure 5 and map of figure 6). The graphic of figure 5 and the map of figure 6 shows that a large part of the impacts computed by the two SHERPA configurations are closed to each other. 67% of the 150 cities are evaluated as Fair, Good or Very Good (Pearson coefficients above 0.85 in figure 5). Moreover, these cities are located in the largest part of Europe (all Europe except the Iberian Peninsula, southern Italy, extreme North Europe and some points like Milan or Lyon). It indicates that the results are robust, which may reassure decision-makers. Unfortunately, even if two models give similar results, they can both be wrong. For this reason, a diagnosis of good robustness remains difficult to exploit. On the contrary, large differences between the results of two models shows that, at least, one of the models is wrong. In such case, the information provided by the comparison may worry decision-makers but become very valuable for model developers and data providers. Observing the map of figure 6 shows clearly that the Iberian Peninsula and the southern Italy are not well simulated by at least one of the SHERPA configurations. This should encourage the developers of CHIMERE and EMEP to control their models and their data in these regions. I advise the authors to insist on this point which seems to me one of the major contributions of their work.

Although precise suggestions directly linked to the exact causes of differences between S-EMEP and S-CHIMERE (emissions, meteorology, CTM, SHERPA approximation…) are not possible with the current methodology, we agree that locations where models diverge can be used to trigger further discussion by the model developers. This is indeed one of the contributions of this work and we better stressed this point in the revised version of the paper (in the Conclusions part).

But the evaluation of the difference between two CTM like EMEP and CHIMERE required some wariness. Indeed, SHERPA does not reproduce exactly the results of a CTM generating some errors which will be probably different for EMEP and CHIMERE. The differences which appear between EMEP and CHIMERE will be amplified or damped by SHERPA. So that, high differences between the two SHERPA configurations could hide low differences between EMEP and CHIMERE and vice et versa. This problem has not been commented and is even not mentioned in this article. I advise the authors to address this point. I suppose they can easily refer to the SHERPA accuracy that have been estimated in their previous publications.

In the revised Supplementary Material, we now included more discussion about the errors attached to the SHERPA approximation. In particular, Figures 4 and 5 show the percentage bias errors for different validation scenarios, for the S-CHIMERE and S-EMEP SRR. However, it is not possible to extrapolate these average 'percentage bias errors' into specific city errors because these depend on the sector considered, on the area over which emission reductions are applied, etc…

The authors use the Pearson correlation to evaluate the differences between the two SHERPA configurations, which is perhaps not the best statistical indicator. The Pearson coefficient does not spot situations where the results of one of the models are proportional to the other. Let suppose, for example, that the results of one of the models is constantly twice the results of the other model. The Pearson coefficient will then be equal to 1. I advise the author to use another indicator, like the RMSE, it will probably not change their conclusions but should avoid the problem just mentioned.

The main aim of this work is to assess the policy implications of using a model rather than another. This is why we focus on the ranking of the contributions rather than on their absolute values. The ranking is indeed the information that is used to start designing an air quality plans. The Pearson coefficient is a good indicator for this purpose whereas the RMSE might give misleading information (the example given by the Reviewer would lead to different information while the decision would remain unchanged). We now stressed this point in the revised document, at line 246.

Then, it could be interesting to evaluate (even roughly) a threshold above which the differences observed between the two SHERPA configurations reflect significant differences between the two systems of models EMEP and CHIMERE. This would help locate the areas where the differences between EMEP and CHIMERE are proven with near certainty.

We agree with the reviewer. However, it is not possible to evaluate this threshold at this stage. For doing this, we would need an estimate of the SHERPA uncertainty for each city, sector and precursor, something we only have for some validation simulations.

Prioritising the sources of pollution in European cities: do air quality modelling applications provide consistent responses?

Bart Degraeuwe, Enrico Pisoni, Philippe Thunis

European Commission, Joint Research Centre (JRC), Ispra, Italy

Correspondence to: E. Pisoni (enrico.pisoni@ec.europa.eu)

Abstract. To take decisions on how to improve air quality, it is useful to perform a source allocation study that identifies the main sources of pollution for the area of interest. Often source allocation is ~~implemented~~ performed with a Chemical Transport Model (CTM) but unfortunately, even if accurate, this technique is time consuming and complex. Comparing the results of different CTMs to assess the uncertainty on ~~the~~ source allocation results is even more difficult. In this work, we compare the source allocation (for PM2.5 yearly averages) ~~on~~ in 150 major cities in Europe, based on the results of two CTMs (CHIMERE and EMEP), approximated ~~through~~ with the SHERPA (Screening for High Emission Reduction Potential on Air) approach. Although contradictory results occur in some ~~instances~~cities, the source allocation results obtained with the two SHERPA simplified models lead to similar results in most cases, e~~E~~ven though the two CTMs use different input data and configurations~~, in most cases the source allocations with the SHERPA simplified models give similar results. But there are also cases where results (in terms of source allocation for PM2.5 yearly averages) are contradictory~~.

1. Introduction

Air quality models are useful tools to perform a variety of tasks like assessment (simulating ~~the~~ concentrations fields at a given moment), forecasting (~~reproducing~~ predicting future concentrations) and source allocation/planning (evaluating priorities of interventions, and the impact of potential emission reduction policies on concentrations). For assessment (Alvaro Gomez-Losada et al., 2018) and forecasting (Corani et al., 2016), it is possible to compare the model results with observations. For example, FAIRMODE[1] (the Forum for air quality modelling in Europe) ~~i.e. provides~~ proposes methods~~tools~~ as the Model Quality Indicator and Model Quality Objective (Pernigotti el al., 2013b; Viaene et al., 2016) to assess the quality of the model results for a given application.~~, as like the Model Quality Indicator and Model Quality Objective (Pernigotti el al., 2013b; Viaene et al., 2016).~~ However, ~~for source allocation and planning,~~ there is no benchmark against which to compare ~~the~~ model results for source allocation and planning, as~~. In this context air quality models are simulating~~ no measurements are available to test the impact of theoretical emission reduction scenarios on concentrations~~., for which no measurements are available. These scenarios are usually implemented considering alternative policy options that might never become real.~~ So, even if ~~they are~~ very useful to

[1] The Forum for Air quality Modeling (FAIRMODE) was launched in 2007 as a joint response initiative of the European Environment Agency (EEA) and the European Commission Joint Research Centre (JRC). The forum is currently chaired by the Joint Research Centre. Its aim is to bring together air quality modelers and users in order to promote and support the harmonized use of models by EU Member States, with emphasis on model application under the European Air Quality Directives. For more details, see https://fairmode.jrc.ec.europa.eu/.

evaluate ex-ante the impact of possible policy options, it is hard to judge the ~~uncertainty~~ quality ~~associated to~~of these results. ~~So~~On the other hand,~~,~~ the uncertainty ~~on~~ associated to~~the~~ source allocation results ~~s given by an air quality model~~ can be ~~evaluated~~ assessed by comparing ~~it~~ them with ~~the~~ results ~~of~~ from other ~~-~~air quality models (Thunis et al., 2007; Cuvelier et al., 2010; Pernigotti et al., 2013). Both the absolute and relative impacts of emission reductions can then be compared. ~~Even if models disagree about the absolute concentration reductions, they might still identify the same sources as main contributors to the air pollution in the area of interest. If model results are consistent one can assume that policies based on these results will be effective.~~

As an initial phase to design an air quality plan, one ~~can be~~is interested in ~~checking~~ identifying the main sources over a given domain that are responsible for the ~~of~~ pollution at a given location ~~for a given domain~~ (Isakov et al., 2017). This step is defined in literature as source allocation~~. By 'source allocation'~~ (Thunis et al., 2019), i.e. ~~we mean the~~a techniques~~s~~ applied to understand the key contributors to air pollution at a given location. ~~This s~~Source allocation then serves as the corner stone to choose the target sector or geographical area ~~on which to focus~~ when designing measures for an air quality plan.~~ Following this initial phase, a model can then be run in 'planning mode', to evaluate the impact of specific emission reduction scenarios on air quality.~~

~~The problem to use a CTM for source allocation is the long computation time. Hence, the number of sources that can be analysed, both in terms of locations, sectors and precursors is limited.~~ The ~~most precise way~~ ideal to perform source allocation would be to use directly a~~n~~ Chemical Transport Model (~~AQM~~CTM) ~~to~~ but this technique is unfortunately too time consuming to differentiate the impacts of many sources at the same time for various cities in Europe. An alternative is to simplify the CTM with a so-called ~~produce~~ source-receptor relationships (SRR) approach, that mimics the CTM relationships between emission and concentration changes. The most precise SRR would consist ~~for the model domain would be with an~~in an independent grid cell-to-grid cell approach. While this approach would allow a high level of flexibility in defining the zones over which emissions are spatially reduced, it involves simulating independently the effect of emissions changes in each single grid cell that has pollutant emissions in the model domain. It would require changing precursor emissions in individual grid cells one at a time and looking at the resulting change in concentrations in each receptor cell. While theoretically very simple, the resulting number of unknown parameters describing the transfers between source and receptor cells that need to be identified is very large. For example, for a domain with ~~Ngrid =~~ 50 × 50 grid cells (Ngrid=2500) and 5 precursors (Nprec = 5), the identification of a maximum of 12,500 parameters would be required (if emissions occur in, and concentration changes need to be calculated for, all grid cells in the domain) to calculate the change of concentration at a given receptor cell. Therefore 12,500 equations, each connecting concentration changes and emission changes are necessary to identify these 12,500 unknown parameters. Because each of these equations requires an independent ~~AQM~~CTM run, this independent grid cell-to-grid cell option is very costly, and simplifying assumptions that reduce the number of ~~AQM~~CTM runs are required (Clappier et al., 2015).

In GAINS ("Greenhouse gas - Air pollution Interactions and Synergies", Amann et al., 2011) the grid-cell to grid-cell relation is simplified by aggregating source cells into countries. The number of unknown parameters that need to be identified for one receptor cell equals the number of countries (Ncountry) multiplied by the number of precursors.

This system can only be solved if at least “Nprec x Ncountry” equations are available, requiring a similar number of independent ~~AQM~~CTM scenarios. In GAINS, about 50 countries and 5 precursors lead to the need of 250 independent ~~AQM~~CTM scenarios to identify 250 unknowns. However, b~~B~~ecause they are derived from emission reductions at country level, these SRRs are not applicable at the urban scale.

In the RIAT + tool (“Regional Integrated Assessment Tool”, Carnevale et al., 2014). Emissions are aggregated into ‘quadrants’ that are defined relatively to each grid cell within the domain. The ‘quadrant’ emissions and their related grid cell concentrations are then used to feed a neural network that delivers the SRR (Carnevale et al., 2009). Although the approach requires a limited number of full CTM simulations (around 20), the set-up of the SRR remains complex due to the need of implementing sophisticated neural networks.

In SHERPA (Thunis et al., 2016; Pisoni et al., 2017), a different approach is taken that reproduces the grid cell-to-grid cell approach but does not require anywhere near as many ~~AQM~~CTM ~~model~~ runs. SHERPA assumes that the unknown parameters vary on a cell-by-cell basis but are no longer independent of each other. ~~Instead~~Instead, these coefficients are assumed to be related through a bell shape function. With the SHERPA approach, the number of unknown parameters is then equal to 2 for each precursor and receptor cell. Consequently, for the five precursors of PM2.5 (VOC, SO_2, NO_x, PPM and NH_3), ten independent ~~AQM~~CTM simulations are needed for a given receptor cell. Provided that they deliver independent information, the same ~~AQM~~CTM scenarios can be used to identify both parameters for all cells within the domain (see details in Pisoni et al. 2017). Based on these 10 CTM simulations the SHERPA approach allows to quickly assess the impact of emission reductions for many combinations of sectors, geographical areas and precursors. ~~Because i~~It is currently the only ~~one~~approach ~~existing to~~that allows performing a systematic analysis ~~in~~for about 150 EU cities in terms of sectors and precursors.~~, we use the SHERPA approach in this work to approximate two CTMs: CHIMERE and EMEP and compare their responses.~~ ~~The SHERPA (Screening for High Emission Reduction Potential on Air) approach (Thunis et al., 2016; Pisoni et al., 2017) has been developed with the aim of providing information on source allocation. SHERPA implements a source-receptor relationship approach, to mimic the behaviour of a full Chemical Transport Model. Its main advantage is the important reduction of the computational time required to perform one simulation, in comparison to a CTM. With this approach the impact of emission reductions for many different combinations of sectors, geographical areas and precursors can be determined quickly. This would be impossible with a full Chemical Transport Model due to time constraints.~~

~~In this work, we used the SHERPA approach to produce a source allocation for 150 cities in Europe.~~

~~So, in this manuscript, A~~aFirst, the SHERPA SRR approximation of the two CTMs, CHIMERE and EMEP, ~~was~~ is built~~d~~. With these two SRR models the contribution of 100 sector-area-precursor combinations ~~on~~ to the concentration in the city centre ~~was~~ is determined and w~~. W~~e assess~~ed~~ the similarities and differences between these two set of results. Obviously some of the differences are caused by the fact that the two CTM models rely on different formulations and parametrisations but also ~~on~~ by the fact that they ~~are~~ use different input data (emissions, meteorology…). The objective of this work is ~~therefore not~~ to assess the overall uncertainty (or better, variability) attached to source allocation rather than to assess the sensitivity of the results to a given parameter (e.g. emissions) ~~but rather to assess the overall uncertainty (or better, variability) attached to source allocation~~.

The focus of this study is on PM2.5 yearly averages, ~~as~~because this is the pollutant with the highest impact on human health, and is therefore a key focus ~~of~~for policy makers in Europe. ~~We also stress the fact that b~~Because a large number of sources contribute to PM2.5 concentrations at one location, this is also the most challenging pollutant to manage in air quality plans.

The paper is structured as follows. We briefly present the two Chemical Transport Model and their set-up in Section 2. We then describe the SHERPA methodology and its assumptions in Section 3. Section 4 details the methodology followed for the source allocation, while the inter-comparison of the results is presented in Section 5. Conclusions are proposed in Section 6.

2. CHIMERE and EMEP Chemical Transport Models: set-up and simulations

In this work, we use~~d~~ two set of model simulations, performed with two of the leading ~~air quality models~~chemical transport models in Europe: CHIMERE and EMEP. More details on the models can be found in Mailler et al., 2017 and Couvida~~e~~t et al., 2018 (for CHIMERE) and Simpson et al., 2012 (for EMEP). Because a~~A~~ brute force source allocation for 150 cities with these models would be too time consuming, ~~; instead here~~ we use two sets of SHERPA Source Receptor Relationships (SRR), each based on a training set of about 20 CHIMERE and EMEP CTM simulations ~~to develop a set of SHERPA Source Receptor Relationships (SRR)~~. These~~is~~ SRR ~~set is~~are then used to perform ~~directly~~ the source allocation. Details on the SHERPA training ~~and validation~~ for CHIMERE can be found in Clappier et al., 2015, and for EMEP in Pisoni et al., 2019.

The CHIMERE and EMEP modelling set-up ~~are different~~differ in the following aspects:~~. The key differences between the two modelling configurations are detailed below:~~

- Grid setting: CHIMERE uses a grid of 0.125 degrees longitude by 0.0625 degrees latitude, corresponding to rectangular cells of more or less 9 by 7 km (in the centre of the domain) whereas EMEP uses a regular grid of 0.1 by 0.1 degrees, corresponding to rectangular cells of more or less 7 by 11 km.
- Emissions: The CHIMERE emission reference year is 2010 with a gridding based on the EC4MACS project proxies (Terrenoire et al., 2015) while EMEP uses a JRC set of emissions (Trombetti et al., 2017) based on 2014 as reference year.
- Boundary conditions: ~~The size of the modelling domains differs.~~ The CHIMERE domain extends from 10.5° East to 37.5° West and between 34° and 62° North while the EMEP domain extends from 30° East to 90° West and between 30° and 82° North.
- Meteorology: The two models use a different reference meteorological year; 2009 for CHIMERE and 2014 for EMEP; both meteorological fields are modelled through the Integrated Forecasting System (IFS) of ECMWF.
- Model Parameterization: Apart from the vertical and/or horizontal resolutions, transport, deposition, chemical processes ~~might be~~are reproduced with different levels of complexity in the two models.

More details on the model simulations and settings can be found in Clappier et al., 2015 and Pisoni et al., 2019. Some of the v~~V~~alidation results for the two model configurations (CHIMERE and EMEP) are briefly presented in the Supplementary Material, showing similar performances ~~(for CHIMERE and EMEP)~~ in terms of comparison against observations. For CHIMERE the relation between predictions and observations at background stations is best characterised by a line through the origin with slope of 1.05, indicating a slight under-prediction. The standard error is 5.7 μg/m^3 and uniform over the range of concentrations. The R2 is 0.9. Concentrations at traffic and industrial stations are underestimated by roughly 10%. For EMEP the relation between predictions and observations is best characterised by a power low with exponent 0.66. The data show a relative standard error constant over the range of concentrations and equal to 30%. Concentrations at traffic stations are under~~-~~predicted by 9% and over-predicted at industrial sites by 7%. It is important to note that the use of different input and model set-up (as listed before) represents the usual practice when air quality models are used, at the local scale, to assess the impact of air quality plans. This is why it is important ~~(in this manuscript)~~here to analyse how this choice influences the results and the subsequent design of an air quality plan; an issue that is often not tackled in the ~~scientific~~ literature. ~~Some differences in results might be due to trends in emissions and concentrations between 2010 and 2014. During this period, concentrations in Airbase stations decrease yearly by 2.2% on average (σ = 2.7%/year). Hence, only differences larger than about 10% in source apportionment should be considered as significant.~~ Finally, differences can arise from the SRR approximations themselves, even if ~~(as shown in the Supplementary Material)~~ validation against CTM simulations show similar results for the 2 considered model set-up (see Supplementary Material).

Starting from these ~~results~~configurations, two set of SRRs ~~have been~~are built ~~to model~~for yearly average PM2.5 concentrations, based respectively on CHIMERE and EMEP data.

~~The focus of this study is on PM2.5 yearly averages, as this is the pollutant with the highest impact on human health, and a key focus of policy makers in Europe. We also stress the fact that because a large number of sources contribute to PM2.5 concentrations, this is the most challenging pollutant to manage in air quality plans.~~ Before looking at the source allocation results, in the next section a brief description of the SHERPA methodology is proposed.

3. SHERPA methodology

Starting from the simulations performed with CHIMERE and EMEP, two sets of SHERPA ~~source-receptor relationships~~SRR are built.

Here we briefly summarise how the SHERPA methodology works; ~~please~~ we refer to Pisoni et al., 2019 for more details.

In the SHERPA approach, the PM concentration change in receptor cell “j” is computed as follows:

$$\Delta PM_j = \sum_{p}^{N_{prec}} \sum_{i}^{N_{grid}} a_{ij}^{p} \, \Delta E_i^{p} \tag{1}$$

where N_{grid} is the number of grid cells within the domain, N_{prec} is the number of precursors, ΔE_i^p are the emission changes, and a_{ij}^p are the unknown parameters to be identified, representing the transfer coefficients between each source cell i and receptor cell j. In SHERPA the a_{ij}^p coefficients are cell-dependent, and assume a ‘bell shape function’. This bell shape function accounts for variation in terms of distance but is directionally isotropic, and can be defined as follows:

Formatted: Not Superscript/ Subscript

Formatted: Not Highlight

$$a_{ij}^{p} = \alpha_j^{p}\left(1 + d_{ij}\right)^{-\omega_j^{p}} \qquad (2)$$

where d_{ij} is the distance between a receptor cell "j" and a source cell "i". Thus, in SHERPA the matrix of transfer coefficients is known when the two parameters α and ω are identified for a given receptor cell j and a given precursor p (see Equation 2). The final formulation implemented in SHERPA is:

$$\Delta PM_j = \sum_{p}^{N_{prec}} \sum_{i}^{N_{grid}} \alpha_j^{p}\left(1 + d_{ij}\right)^{-\omega_j^{p}} \Delta E_i^{p} \qquad (3)$$

With the SHERPA approach, the key step is so to find the optimal α, ω coefficients. As the number of unknown parameters is equal to 2 (α, ω) for each precursor and receptor cell "j", for the five precursors of PM2.5 (VOC – volatile organic compounds, SO_2 – sulphur dioxide, NO_x – nitrogen oxides, PPM – primary particulate matter and NH_3 – ammonia), ten independent CTM simulations are needed for a given receptor cell. We refer to ~~(~~Pisoni et al. (~~,~~ 2018) and ~~;~~ Thunis et al.~~,~~ (2016) for additional details about the SHERPA formulation and evaluation process.

Given its cell-to-cell characteristics (Equation 3), the SHERPA formulation can be used to assess the impact of emission reductions over any given set of grid cells. Different geographical entities can therefore be freely defined in terms of boundaries~~, and simulated through SHERPA~~.

As ~~previously said, in this work~~mentioned earlier, the SHERPA approach is used in this work to analyse the differences in source allocation results between two ~~air quality modelling setting~~CTM:~~,~~ ~~based on~~ CHIMERE and EMEP, referred to in this paper as S-CHIMERE and S-EMEP, respectively. The "S-" first letter in these acronyms reminds that we compare the EMEP and CHIMERE SRR rather than the models themselves.

4. Source allocation methodology

~~Starting from the S-CHIMERE and S-EMEP SRRs, the~~ The aim of this work is to ~~analyse~~ compare the main contributors to urban pollution in terms of sectors, geographical areas and precursors, obtained with S-CHIMERE and S-EMEP~~, as modelled by the 2 modelling configurations~~. We focus on the PM2.5 yearly average concentrations as target indicator, because PM2.5 is responsible for most of the health related burden in the EU urban areas (EEA 2019). The approach is applied to ~~the~~ 150 European cities, those analysed in the 'PM2.5 Urban Atlas' (Thunis et al., 2018). As mentioned above, the cell-to-cell characteristics of the SHERPA approach allows assessing the impact of emission reductions over any given set of grid cells ~~(to be assessed. C~~cities , regions or countries can ~~therefore~~ be freely defined in terms of boundaries) and~~. E~~ emission reductions can ~~also~~ be freely defined in terms of precursors or sectors. The following single (or combination of) sectors, source areas and precursors are considered as sources.

In terms of sectors, ~~emissions~~ the source categories follow the CORINAIR SNAP nomenclature for emissions:

- Combustion in energy and transformation industries (SNAP 1),
- Non-industrial combustion plants (SNAP 2),
- Combustion in manufacturing industry (SNAP 3),
- Production processes (SNAP 4),
- Extraction and distribution of fossil fuels and geothermal energy (SNAP 5),

- Solvent use and other product use (SNAP 6),
- Road transport (SNAP 7),
- Other mobile sources and machinery (SNAP 8),
- Waste treatment and disposal (SNAP 9) and
- Agriculture (SNAP 10).

which have been aggregated in this work into five sectors:

- industry (SNAP 1, 3 and 4),
- residential (SNAP 2),
- traffic (SNAP 7),
- agriculture (SNAP 10), and
- others (SNAP 5, 6, 8 and 9).

In terms of geographical sources, four areas are considered for the analysis:

- the core city,
- the commuting zone,
- the rest of the country and
- international (what is outside the considered country).

The commuting zone is defined as th~~eat~~ area surrounding the city where at least 15% of the population commutes daily to the core city. The combination of the core city and the commuting zone is referred to as the functional urban area, or FUA[2].

Finally, the precursors considered are NO_X, VOC, NH_3, PPM and SO_2.

This leads to 100 (4 areas x 5 precursors x 5 sectors) runs for each ~~model~~ SRR and city. For small cities (66 out of 150) the core city covers too few grid cells which would lead to discretization errors. In such ~~case~~case, the analysis is restricted to the FUA. For these cities, 75 runs (3 areas x 5 precursors x 5 sectors) per city and model were therefore performed. With 150 analysed cities for two CTM models, ~~it is interesting~~we ~~to~~ note that the SHERPA approach allows for a comparison that would have implied 26700 ((66x75 + 84x100) x 2 models) independent air quality simulations with a full ~~Chemical Transport Model~~CTM. ~~Note that t~~The same amount of runs ~~has been done~~ with the SHERPA simplified model~~, but with~~ only ~~requested a~~takes few ~~minutes~~ seconds ~~required to perform one~~per scenario. The results for S-CHIMERE were published in the 'Urban PM2.5 Atlas' (Pisoni et al., 2018). ~~For~~ In this paper, the same runs are done with S-EMEP, and a comparison between the 2 is provided.

Each run performed with the SHERPA SRRs provides a concentration change (ΔC) that results from an emission reduction (ΔE) with an intensity α ~~imposed~~ applied ~~on~~ to a given precursor, for a given sector and within a given area. ~~While~~ The 'relative potential' of a given precursor-sector-area ~~combination~~source is expressed as $\Delta C/\alpha C$, (Thunis and Clappier, 2014). This indicator represents the share of a particular emission source to the concentration. From a policy point of view, high 'relative potential' sources are the ones to be addressed ~~at~~ first to achieve the largest improvements. In this work, the SRRs ΔC ~~from SRRs~~ are ~~representative~~ obtained for emission reductions of α=50%~~, results are then~~

[2]See https://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm for details.

~~scaled to 100% to obtain the total impact of a given source ($\Delta C/\alpha$)~~, a level that ~~. The 50%~~ represents a threshold below which the quasi-linearity of the model responses is preserved~~, at least when considering yearly average concentrations of PM2.5~~ (Thunis et al., 2015). In other words, with this approach the model response in terms of concentration change ~~is~~ remains proportional to the emission change ~~of a given source~~. It is important to stress that this threshold is only valid for PM2.5 and for yearly average~~s~~ concentrations, as considered here. Because of this 50% threshold, it is also worthwhile to note that the source allocation results discussed here provide information on the impact of potential emission reductions up to that level, ~~of 50% (~~ not beyond~~)~~.

~~The 'relative potential' of a given precursor-sector-area combination is expressed as $\Delta C/\alpha C$, (Thunis and Clappier, 2014). This indicator represents the share of a particular emission source to the concentration. From a policy point of view, high 'relative potential' sources are the ones to be addressed at first to achieve the largest improvements.~~ To compare the 'relative potentials' from S-CHIMERE and S-EMEP ~~from S-CHIMERE and S-EMEP~~, we calculate the correlation ~~between the relative potentials~~. A high correlation means that both models agree well on the emission sources (sectoral and/or geographic) that contribute most to the concentration for a given city. The main advantage of a correlation indicator is that it ignores systematic differences. In other words, ~~the fact that~~if one model systematically ~~might~~ predicts ~~systematically~~ higher concentration changes for all sources than the other, this is ~~will~~ not ~~be~~ detected by the correlation metric. This is a desirable characteristic because from a policy perspective ~~from the policy perspective~~, it is the 'relative ranking' among the sources contributions that counts rather than their absolute values.

5. Comparison of the results

In this study, we compare the ~~contributions~~relative potentials for 150 cities, based on the two SHERPA implementations, S-CHIMERE and S-EMEP. ~~The s~~Source allocation is ~~provided~~ calculated ~~for~~ at the city location characterised by the worst target indicator value, i~~ of its target indicator (i~~.e. the most polluted cell in the considered city~~)~~. We first discuss the results for a few cities, before moving to an EU wide perspective.

Tables 1 to 4 show, for each emission area, sector and precursor, the 'relative potential' expressed in percentage of the total concentration for the 2 models (~~in % of the total concentration,~~ 'chimere_rp' and 'emep_rp') and the resulting ranking in terms of importance ('emep.rank' and 'chimere.rank')~~,~~ for 4 cities: Liege, Genova, Turin and Madrid. These cities~~,~~ are selected as representative samples to illustrate the ~~to represent different~~ characteristic behaviours obtained in ~~in terms of SRRs~~our comparison. In addition to this, Figures 1 to 4 show the S-CHIMERE/S-EMEP correlations ~~show the 'relative potentials'~~obtained ~~for the 2 models (S-CHIMERE and S-EMEP),~~ for ~~the different~~ various types ~~of~~relative potentials defined in terms of ~~considered aggregations~~~~performed scenarios (considering emission reductions for the selected~~ geographical area, ~~for the chosen~~ sector, or ~~for~~ their combinations ~~of geographical areas - sectors, …) and their corresponding correlations, for the same cities~~.

~~As said, we present results for 4 cities (Liege, Genova, Turin and Madrid) selected as representative of the different behaviours identified in our analysis.~~

For Liege (Belgium), the overall (all individual sectors, precursors and areas included, i.e. about 15000 relative potentials) Pearson correlation[3] between the relative potentials of both ~~models~~ SRR is the highest among the 150 cities (r=0.99, see Figure 1). Both models identify ammonia emissions from agriculture, outside Belgium, as the main contributor to local PM2.5 concentrations. Primary PM from local industry comes second and NO_x from international traffic third. Although the lower ranked combinations are not identical, they are quite similar. From a policy perspective, the fact that both ~~modelling applications~~SRR provide similar information is a sign of robustness. It increases our confidence in the priority of interventions (which sectors-areas to act at first to achieve the maximum air quality improvement)~~ proposed by each model.~~. The values ~~of~~ for the ~~the different~~ main sector-precursor-areas ~~contributions (expressed as~~ relative potentials~~)~~ are reported in Table 1.

Table 1: Top 10 area-sector-precursor ~~combinations contributi~~relative potentials~~ng~~ to ~~the~~ PM2.5 concentrations in Liege (B).

Formatted: Font color: Auto

area	sector	precursor	emep_rp	emep.rank	chimere_rp	chimere.rank
International	Agriculture	NH3	22.9	1	20.6	1
FUA	Industry	PPM	12.6	2	12.4	2
International	Road Transport	NOx	7.5	3	6.9	3
International	Industry	NOx	4.9	5	5.2	4
National	Agriculture	NH3	4.2	6	4.6	5
International	Industry	SOx	5.1	4	2.3	10
International	Residential	PPM	2.2	7	2.5	8
FUA	Road Transport	PPM	2.1	10	2.9	6
International	Industry	PPM	2.2	8	2.4	9
FUA	Industry	SOx	1.9	15	2.7	7
International	Other	NOx	2.2	9	1.9	13

A breakdown analysis for Liege is proposed in Figure 1 where correlations are ~~expressed~~ calculated for ~~different data~~ relative potentials that are aggregated in terms of sectors (~~aggregations. In addition to the overall correlation (75000 values), values are also proposed for data grouped by sectors (150 cities x~~ 5 relative potentials~~sectors~~), ~~by~~ area (~~150 cities x~~ 4 ~~areas~~relative potentials) or ~~by~~ area/sectors (~~150 cities x~~ 5 ~~pr3ecursors~~ x 5 ~~pollutants~~relative potentials). In the case of Liege, all correlations are ~~consistently~~ very good.

[3] The main aim of this work is to assess the policy implications (i.e. which source to tackle first) of using a model rather than another. This is why we focus on the ranking of the contributions (Pearson correlation) rather than on their absolute values~~; that means, this is way we use in this context the Pearson correlation~~.

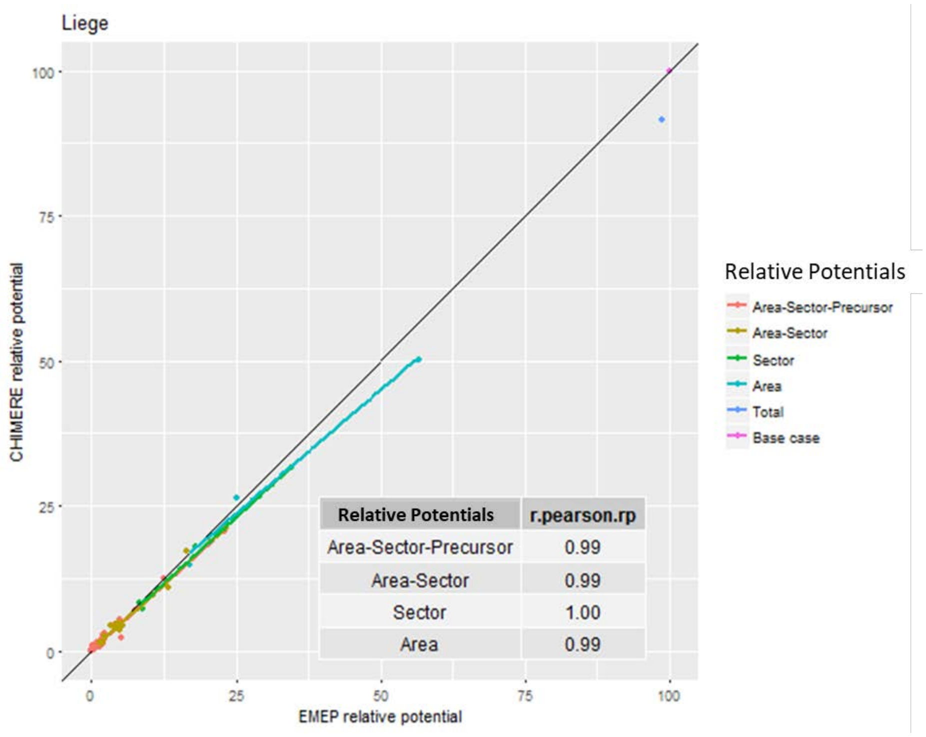

Figure 1: Correlation between S-EMEP and S-CHIMERE relative potentials ~~relative potentials of S-EMEP and S-CHIMERE~~ for different sector-area-precursor source aggregations in Liege (B).

Unfortunately, the agreement is not always ~~as~~ so good. For the city of Genova (Table 2 and Figure 2), both models agree that national/international ammonia emissions from agriculture ~~areas~~ are the largest contributor to local PM2.5 (see Table 2). But the third position in the priority ranking is occupied by NO_x from national traffic for S-EMEP while it is PPM from the national residential sector for S-CHIMERE. However, the overall correlation still reaches 89% and the ~~absolute values of the third ranked sectors are quite close~~two main sources are similar. The agreement between the two models is therefore still satisfactory. It is interesting to note that for ~~relative potentials~~ area-aggregated relative potentials~~aggregated per area~~, the correlation drops to 42%, ~~pointing~~ highlighting possible ~~to~~ differences in the way emission inventories are spatially distributed~~ion~~ ~~of~~in the two models ~~the two emission inventories~~.

Table 2: Top 10 area-sector-precursor ~~combinations contributi~~relative potentials~~ng~~ to ~~the~~ PM2.5 concentrations in Genova (IT).

Relative Potentials

area	sector	precursor	emep_rp	emep.rank	chimere_rp	chimere.rank
National	Agriculture	NH3	14.5	1	11.3	1
International	Agriculture	NH3	6.8	2	10.1	2
National	Residential	PPM	4.3	4	4.7	3
FUA	Residential	PPM	3.2	5	3.5	4
National	Road Transport	NOx	4.9	3	2.6	8
FUA	Road Transport	NOx	3.2	6	2.8	7
International	Industry	SOx	2.2	10	3.4	5
National	Industry	SOx	1.7	15	2.5	9
International	Residential	PPM	1.4	18	2.8	6
FUA	Road Transport	PPM	1.4	17	2.1	10
FUA	Other	NOx	2.5	8	0.7	21
FUA	Industry	NOx	2.4	9	0.0	59
FUA	Industry	SOx	3.1	7	0.0	62

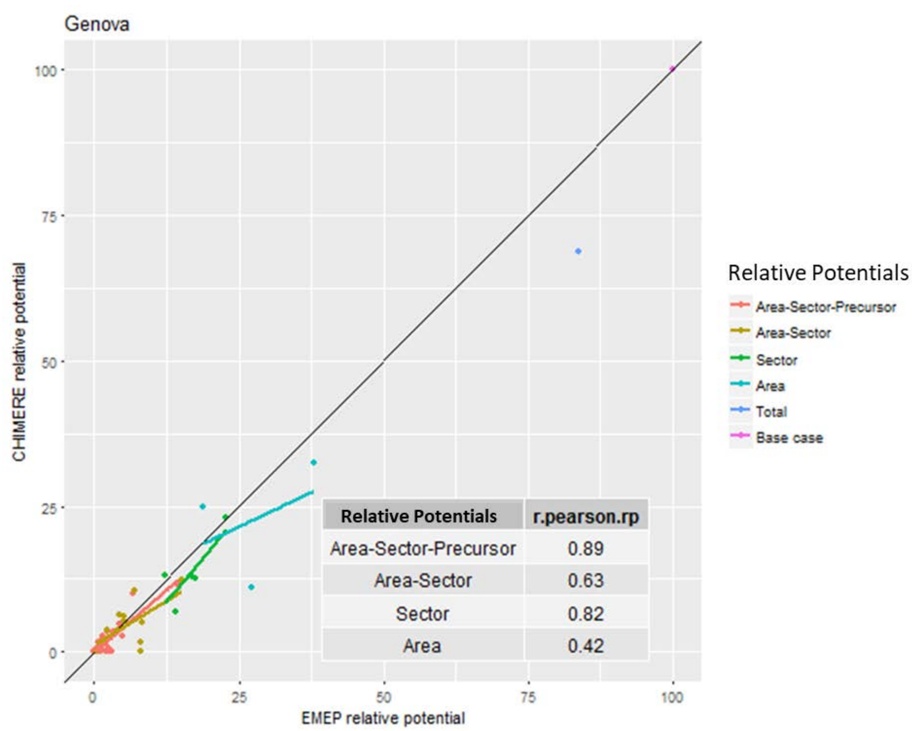

Figure 2: Correlation between ~~relative potentials~~ S-EMEP and S-CHIMERE relative potentials ~~of S-EMEP and S-CHIMERE~~ for different sector-area-precursor source aggregations in Genova (I).

In the case of Torino (Table 3 and Figure 3), the two models give contradicting recommendations. While S-CHIMERE points to city residential heating as main contributor to PM2.5, S-EMEP points to national agriculture ammonia emissions. The model disagreement extends to the top 5 relative potentials~~top 5 ranking~~. As indicated, the problem is probably related to the sectoral (R2=0.78) rather than to the geographical dimension (R2=0.97). Nevertheless, the overall correlation (0.81) is not too bad, and can be explained by the fact that the ~~relative potential~~contribution values are not too different from each other (although the ranking is quite different).

Formatted: Font: 9 pt, Font color: Auto, English (United Kingdom)

Formatted: Font: (Default) Times New Roman, Font color: Auto

Formatted: Font: (Default) Times New Roman, Font color: Auto

Formatted: Font: (Default) Times New Roman, Font color: Auto

Table 3: Top 10 area-sector-precursor ~~combinations contributi~~relative potentials~~ng~~ to ~~the~~ PM2.5 concentrations in Torino (I).

Relative Potentials						
area	**sector**	**precursor**	**emep_rp**	**emep.rank**	**chimere_rp**	**chimere.rank**
FUA	Residential	PPM	8.6	2	13.3	1
National	Agriculture	NH3	10.6	1	5.9	4
FUA	Industry	PPM	6.4	3	13.3	2
FUA	Road Transport	NOx	6.2	4	4.8	6
National	Residential	PPM	4.9	7	5.4	5
International	Agriculture	NH3	6.1	5	4.2	8
FUA	Industry	NOx	5.2	6	4.7	7
FUA	Road Transport	PPM	2.6	13	8.4	3
FUA	Other	PPM	2.9	12	3.5	10
International	Residential	PPM	2.0	16	4.0	9
National	Road Transport	NOx	4.3	8	1.3	18
FUA	Residential	NOx	3.8	9	1.0	23
International	Road Transport	NOx	3.1	10	0.8	25

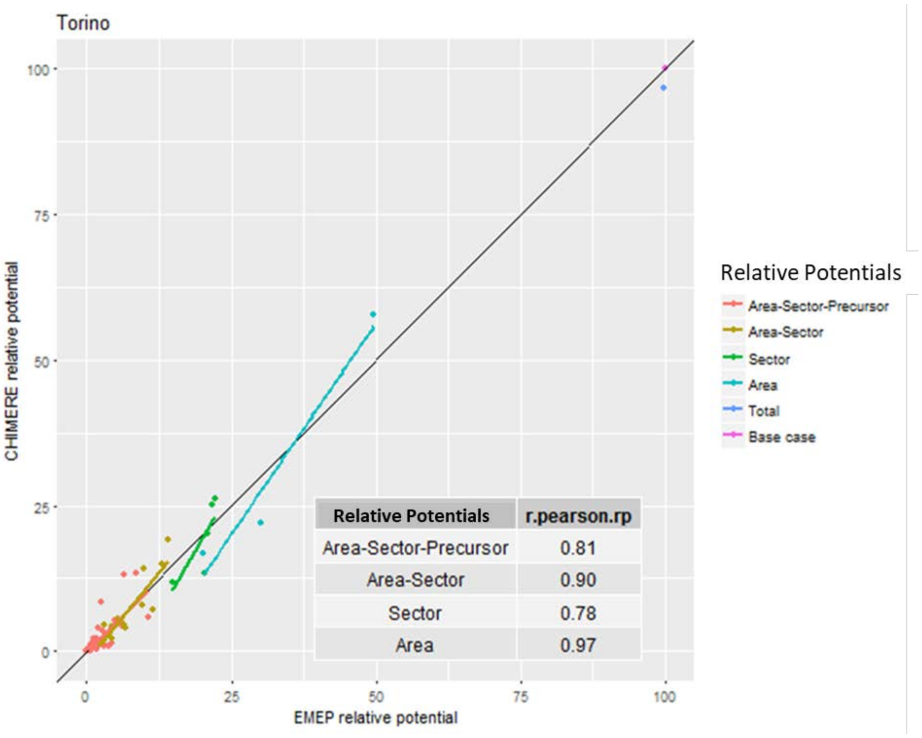

Figure 3: Correlation between S-EMEP and S-CHIMERE ~~relative potentials~~relative potentials ~~of S-EMEP and S-CHIMERE~~ for different sector-area-precursor source aggregations in Torino (I).

In our last example (Madrid - Table 4 and Figure 4), differences are extremely important in terms of ~~relative potentials~~relative potentials and ranking, leading to an overall correlation of 41%. All other correlations, with the exception of the spatial ones are extremely poor. Uncertainties for this city are important, and the choice among policy options ~~shows important variability~~is not robust.

Table 4: Top 10 area-sector-precursor relative potentials to ~~the~~ PM2.5 concentrations in Madrid (E).

Relative Potentials						
area	**sector**	**precursor**	**emep_rp**	**emep.rank**	**chimere_rp**	**chimere.rank**
City	Road Transport	PPM	9.9	2	24.6	1
City	Residential	PPM	6.2	3	8.9	2
City	Other	PPM	2.0	9	5.0	4
National	Agriculture	NH3	2.5	6	2.4	8
Comm	Road Transport	PPM	1.7	11	5.3	3
National	Agriculture	PPM	0.9	13	4.3	5
City	Industry	PPM	2.4	7	1.4	12
City	Other	NH3	2.3	8	1.8	11
Comm	Residential	PPM	1.0	12	2.3	9
City	Industry	SOx	25.4	1	0.8	21
City	Road Transport	NOx	0.8	16	2.7	6
City	Residential	SOx	4.7	4	0.9	20
National	Residential	PPM	0.7	18	2.4	7
National	Road Transport	PPM	0.8	15	2.2	10
National	Industry	SOx	1.8	10	0.8	22
Comm	Industry	SOx	2.8	5	0.4	28

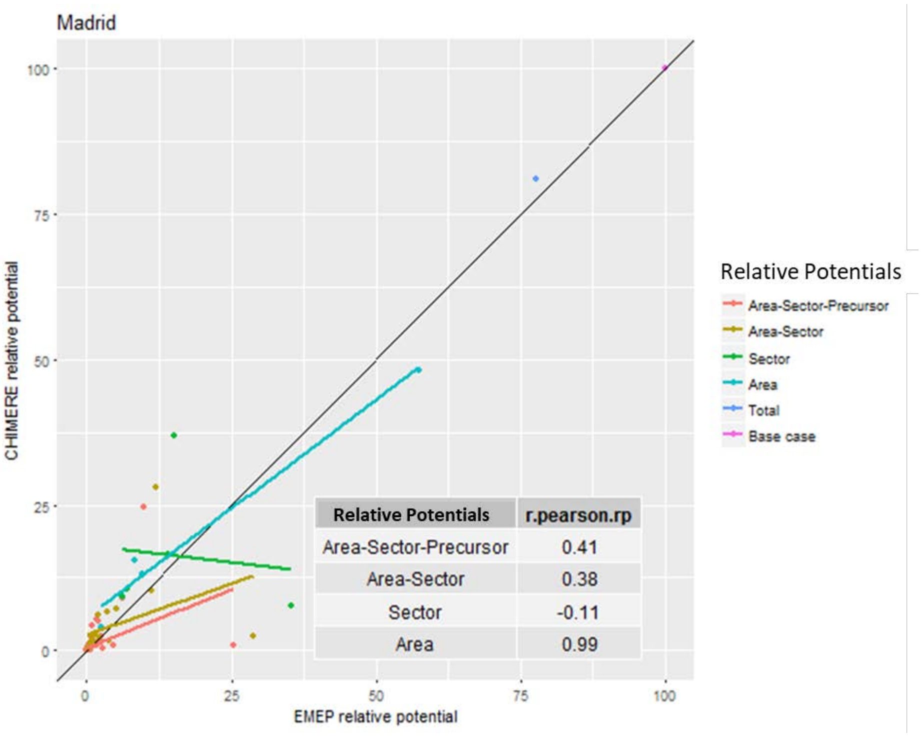

Figure 4: Correlation between S-EMEP and S-CHIMERE relative potentials ~~of S-EMEP and S-CHIMERE~~ for different sector-area-precursor source aggregations for Madrid (E).

As seen from the city examples presented above, we can have both strong (Liege) and weak (Madrid) agreement between the two modelling set-up.

The analysis presented above was done for all 150 cities, and we can here ~~the results are presented in an aggregated~~ present the results in an aggregated way. We will consider here that an overall correlation is very good above 95%~~ as very good~~, good between 90 and 95%~~ as good~~, fair between 85 and 90%~~ as fair~~, poor between 70% and 85% ~~bad~~ and very poor below 70%~~ very bad~~. This is an arbitrary choice, but it is~~can be~~ useful to start grouping and classifying the results. The histogram of the overall correlations for all 150 cities (Figure 5:) shows that the model agreement is good or very good for about half of the cities, satisfactory for another 21%, leaving 32% of doubtful/problematic cities.

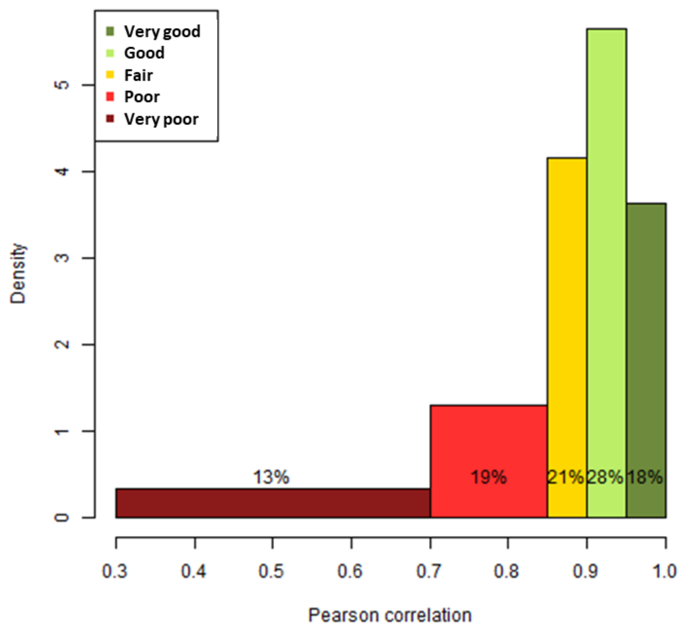

Figure 5: Distribution of the ~~overall~~ Pearson correlation coefficients between relative potentials, for 150 cities.

The mapping of the overall correlations ~~map of Europe~~ (Figure 6) shows that cities with the highest variability are mostly located in Spain, Northern Italy ~~as well as~~and in the Baltic countries. ~~Probably for~~ For these areas, ~~the differences in terms of meteorology~~meteorological factors, emissions, and/or thei~~r~~ impact of these input on concentrations ~~through~~ in the air quality models, is ~~higher~~ larger than in other areas. I~~.~~n the Supplementary Material we show that even for the base case, results~~i.e. how the validation results, for the basecase,~~ are quite different for countries like Spain. ~~in the 2 model implementation, and t~~This ~~could~~might also have an impact on the correlation results shown in this Figure.

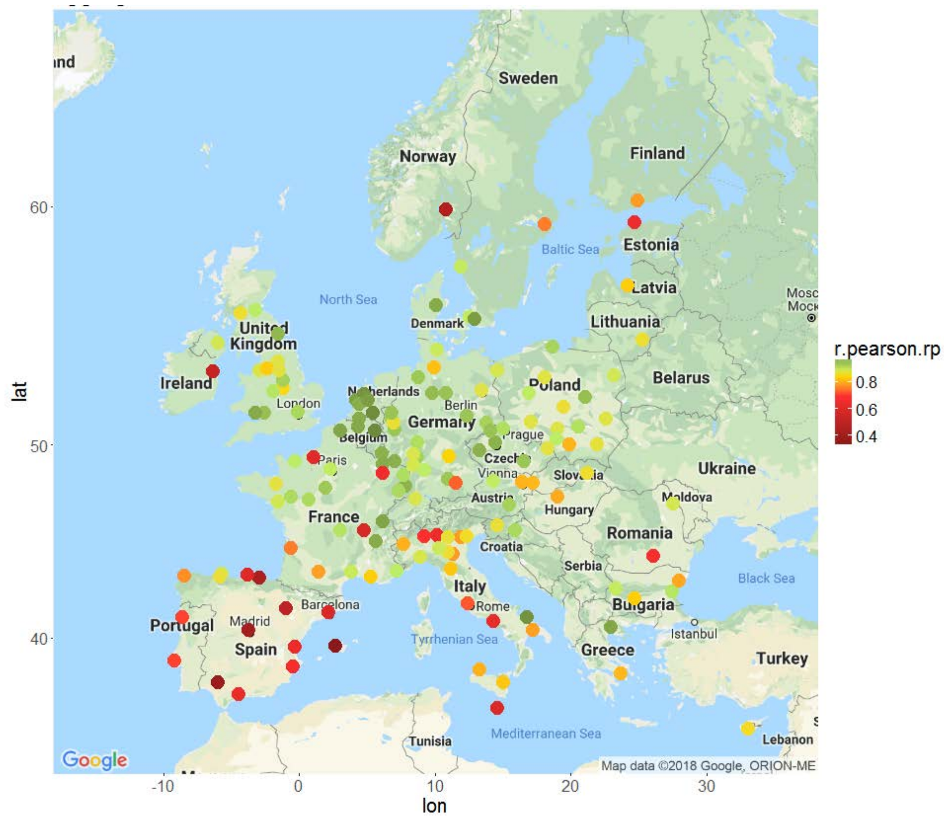

Figure 6: Pearson overall correlation between EMEP and CHIMERE relative potentials.

To the knowledge of the authors, this is one of the first attempts to systematically compare the sources and causes of pollution in European cities, using a harmonized approach. The reasons for ~~these~~ the differences between cities highlighted above are however not easy to identify. This is because the SRRs used in this study are based on different meteorological years (2009 vs 2014), emissions (2010 vs 2014) and air quality models (CHIMERE vs EMEP). ~~So,~~

~~even~~ Although ~~if~~ this analysis provides an overall estimate of the variability ~~of~~ between policy responses and~~, it~~ does not allow ~~us to~~ identifying the~~a~~ specific cause for the observed differences, ~~.~~it indicates where modelling improvements need to be made. Modelling inconsistencies are indeed categorised in terms of geographical area, sectors and precursors, a useful information to trigger discussion among modelling groups and direct the investigations towards the most problematic issues.

It is also worth reminding that ~~However, this situation (that is to say, the~~ using~~e~~ ~~of~~ different input and model set-up~~)~~ represents the usual practice whenever air quality models are used at the local scale to assess the impact of air quality plans. Indeed, each local/regional authority~~ies~~ generally uses ~~only one given~~its own set of data and ~~,~~ applies~~ying~~ ~~a~~ its ~~particular~~ own model.~~, due to a lack of resources and information.~~ Therefore, only a ~~given~~ single meteorology, a ~~given~~ single emission inventory for a ~~given~~ single reference year and a specific model are used to identify the sources of pollution to target. The impact of these choices ~~How this choice influences~~on source allocation ~~the results~~ and on the subsequent design of an air quality plan is an issue that is ~~often~~ not often tackled.

It is probably unreasonable to think that a local authority can evaluate in a comprehensive way the variability of a particular modelling pathway (too ~~much~~ demanding in terms of sensitivity analysis). We however believe that this work can be used to develop further guidance ~~should be provided~~ to select the proper modelling set-up (choice of meteorological year, emission, model to use) to reduce the uncertainty attached to the results and increase their robustness.

The ~~final~~ ultimate goal of this work would be to help decision makers to properly define key sources, so that only 'no-regret' policies are selected. As mentioned above, the present ~~work~~ approach flags out potential issues and a possible lack of robustness (by ~~aims to~~ quantifying the overall~~is~~ variability) but it cannot provide explanations for the observed differences. The only process to identify the causes of differences, is to perform regular inter-comparison exercises where the responses of models to emission changes are systematically tested via sensitivity analysis. While exercises of this type occurred in the past years (Colette et al., 2017, Cuvelier et al., 2007, Pernigotti et al., 2013), it is crucial that these are performed on a regular basis as models and input data continuously evolve.

6. Conclusions

Before applying emission reduction measures to improve air quality, it is important to evaluate the importance of the key sources contributing to pollution in a given area. The main methodology to perform this task is referred to as 'source allocation'.

Source allocation can be implemented in various ways. In this paper we use the SHERPA model, a source-receptor relationship mimicking the behaviour of a fully-fledged CTM. With SHERPA one can perform hundreds of simulations in few minutes to test the impact of various geographical, sectoral or precursor-based emission sources, on the concentration at a ~~point~~ location of interest. The result is a complete source-allocation study for a given domain explaining the key sources of pollution ~~for a~~at a given ~~area~~location.

In this work, we developed two SHERPA versions, based on two modelling set-up using different meteorological reference year, emission inventories and air quality models. Even if these setting are quite different and difficult to compare, they represent what happens in the real-world when designing air quality plans. Indeed~~fact~~, ~~different~~ local authorities in Europe are free to use different reference meteorological years, emissions and models. The comparison of these results therefore provide an estimate of the variability attached to source allocation results for a given area. ~~As this is the current practice in air quality modelling for planning in Europe (in fact one can freely choose meteorological reference years, emissions, models, when building a plan) we conclude that~~ The results can also be used to provide further guidance ~~is needed to understand how~~ to ~~properly~~ define the~~is~~ modelling set-up and~~; and to~~ understand how this choice ~~could~~ impact the selection of priorities ~~for intervention and the variability of the results~~when designing air quality plans.

The two SHERPA SRRs versions (based on CHIMERE and EMEP) have then been used to perform source allocation on 150 main cities in Europe, and results have been presented in terms of priorities of interventions (i.e.: which are the sector/geographical areas/pollutants that are more relevant for air quality in a given city?).

The results are consistent for some cities, i.e. ~~consistent (changing~~ the modelling set-up ~~we get~~produces the same ranking in terms of ~~priorities~~contributions,~~),~~ ~~while~~ whereas for other cities (~~a minority~~about 30%) the two SRRs deliver different results. Even if it is not possible in this work to identify the causes for these differences as ~~(as the two modelling set-ups are too different~~additional sensitivity simulations would be needed for this, this work indicates where modelling improvements need to be made. Modelling inconsistencies are indeed categorised in terms of geographical area, sectors and precursors, a useful information to trigger discussion among modelling groups and direct the investigations towards the most problematic issues.~~)~~ ~~the paper shed light on the fact that one can get quite different ranking of sectors-areas depending on the modelling set-up considered.~~ Although differences in terms of results were expected (different assumptions deliver different results), it is comforting to see that similar policy decisions would be taken in about 75% of cities considered in this study. ~~This is quite logical (different assumptions will deliver different results) but at the same time it is an important issue to be underlined. As this is the current practice in air quality modelling for planning in Europe (in fact one can freely choose meteorological reference years, emissions, models, when building a plan) we conclude that further guidance is needed to understand how to properly define this modelling set-up; and to understand how this choice could impact the selection of priorities for intervention and the variability of the results. Furthermore, locations where models diverge could be used to trigger further discussion by the model developers and users.~~

Thanks to the limited number of required simulations to build SHERPA, future work could envisage the implementation of 'constrained setting' to build SRR (i.e. keeping the same air quality model but changing emissions, or keeping the same emissions but changing the model) to be able to discriminate ~~on~~ the ~~relative contributions~~ role of these ~~different~~ factors ~~involved~~. Also, further model inter-comparison works should be fostered.

Code and data availability

The code and data used to perform the analysis presented in this paper is available at https://github.com/enricopisoni/SRR_comparison (Last access: 7th of April 2020). The SHERPA model, providing the source-receptor relationships applied in this paper, is available at https://aqm.jrc.ec.europa.eu/sherpa.aspx (Last access: 7th of April 2020).

Authors contribution

BD developed the methodology, performed ~~all~~ the analysis and drafted a first version of the paper. PT conceived the initial development of SHERPA, and contributed to the structuring and revision of the paper. EP developed the SHERPA tool, contributed to the interpretation of the results and to the preparation of the final version of the paper.

Acknowledgements

We acknowledge A. Colette (INERIS), H. Fagerli and S. Tsyro (The Norwegian Meteorological Institute) for their work in performing CTM simulations, and for exchange of views on the content of this paper.

References

Amann M. et al.: Cost-effective control of air quality and greenhouse gases in Europe: modelling and policy applications, Environ. Model. Softw, 26, 1489-1501, 2011.

Carnevale C., Finzi G., Pisoni E., Volta M.: Neuro-fuzzy and neural network systems for air quality control, Atmos. Environ., 43, 4811-4821, 2009.

Carnevale C., Finzi G., Pederzoli A., Turrini E., Volta M., Guariso G., Gianfreda R., Maffeis G., Pisoni E., Thunis P., Markl-Hummel L., Blond N., Clappier A., Dujardin V., Weber C., Perron G.: Exploring trade-offs between air pollutants through an integrated assessment model, Sci. Total Environ., 481, 7-16, 2014.

Clappier, A., Pisoni, E., Thunis, P.: A new approach to design source-receptor relationships for air quality modelling, Environ. Model. Softw., 74, pp. 66-74, 2015.

Colette, A., Andersson, C., Manders, A., Mar, K., Mircea, M., Pay, M.-T., Raffort, V., Tsyro, S., Cuvelier, C., Adani, M., Bessagnet, B., Bergström, R., Briganti, G., Butler, T., Cappelletti, A., Couvidat, F., D'Isidoro, M., Doumbia, T., Fagerli, H., Granier, C., Heyes, C., Klimont, Z., Ojha, N., Otero, N., Schaap, M., Sindelarova, K., Stegehuis, A. I., Roustan, Y., Vautard, R., van Meijgaard, E., Vivanco, M. G., and Wind, P: EURODELTA-Trends, a multi-model experiment of air quality hindcast in Europe over 1990–2010, Geosci. Model Dev., 10, 3255–3276, https://doi.org/10.5194/gmd-10-3255-2017, 2017.

Corani G., Scanagatta M.: 2016. Air pollution prediction via multi-label classification, Environ. Model. Softw., 80, 259-264, 2016.

Couvidat, F., Bessagnet, B., Garcia-Vivanco, M., Real, E., Menut, L., and Colette, A.: Development of an inorganic and organic aerosol model (CHIMERE 2017β v1.0): seasonal and spatial evaluation over Europe, Geosci. Model Dev., 11, 165–194, https://doi.org/10.5194/gmd-11-165-2018, 2018.

Cuvelier, C., Thunis, P., Vautard, R., Amann, M., Bessagnet, B., Bedogni, M., Berkowicz, R., Brandt, J., Brocheton, F., Builtjes, P., Carnavale, C., Coppalle, A., Denby, B., Douros, J., Graf, A., Hellmuth, O., Hodzic, A., Honoré, C., Jonson, J., Kerschbaumer, A., de Leeuw, F., Minguzzi, E., Moussiopoulos, N., Pertot, C., Peuch, V.H., Pirovano, G., Rouil, L., Sauter, F., Schaap, M., Stern, R., Tarrason, L., Vignati, E., Volta, M., White, L., Wind, P., Zuber, A.,: CityDelta: A model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmos. Environ. 41, 189–207. doi:10.1016/j.atmosenv.2006.07.036, 2007.

Gómez-Losada A., José Carlos M. Pires, Rafael Pino-Mejías: Modelling background air pollution exposure in urban environments: Implications for epidemiological research, Environ. Model. Softw., 106, 13-21, 2018.

Isakov V., Barzyk T.M., Smith E.R., Arunachalam S., Naess B., Venkatram A.: A web-based screening tool for near-port air quality assessments, Environ. Model. Softw., 98, 21-34, 2017.

Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., Meleux, F.: CHIMERE-2017: From urban to hemispheric chemistry-transport modeling. Geosci. Model Dev. 10, 2397–2423. doi:10.5194/gmd-10-2397-2017, 2017.

Pernigotti, D., Thunis, P., Cuvelier, C. et al.: POMI: a model inter-comparison exercise over the Po Valley, Air Qual. Atmos. & Health, 6 (4), 701-715, 2013.

Pernigotti, D., Gerboles, M., Belis, C.A., Thunis, P.: Model quality objectives based on measurement uncertainty. PartII: NO2 and PM10, Atmos. Environ., 79, pp. 869-878, 2013b.

Pisoni, E., Clappier, A., Degraeuwe, B., Thunis, P.: Adding spatial flexibility to source-receptor relationships for air quality modeling, Environ. Model. Softw., 90, 68-77, 2017.

Pisoni, E., Thunis, P., Clappier, A.: Application of the SHERPA source-receptor relationships, based on the EMEP MSC-W model, for the assessment of air quality policy scenarios, Atmos. Environ: X, 4, art. no. 100047, 2109.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L.D., Fagerli, H., Flechard, C.R., Hayman, G.D., Gauss, M., Jonson, J.E., Jenkin, M.E., Nyúri, A., Richter, C., Semeena, V.S., Tsyro, S., Tuovinen, J.P., Valdebenito, A., Wind, P:. The EMEP MSC-W chemical transport model – Technical description. Atmos. Chem. Phys. 12, 7825–7865. doi:10.5194/acp-12-7825-2012, 2012.

Sorte et al., Assessment of source contribution to air quality in an urban area close to a harbor: case-study in Porto, Portugal, Sci. Total Environ., 662, 347-360, 2019.

Terrenoire, et al.: High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE, Geosci. Model. Dev., 8, 21-42, 2015.

Thunis, P., Rouil, L., Cuvelier, C., Stern, R., Kerschbaumer, A., Bessagnet, B., Schaap, M., Builtjes, P., Tarrason, L., Douros, J., Moussiopoulos, N., Pirovano, G., Bedogni, M.: Analysis of model responses to emission-reduction scenarios within the CityDelta project. Atmos. Environ. 41, 208–220. doi:10.1016/j.atmosenv.2006.09.001, 2007.

Thunis, P., Clappier, A.: Indicators to support the dynamic evaluation of air quality models. Atmos. Environ. 98, 402–409. doi:10.1016/j.atmosenv.2014.09.016, 2014.

Thunis, P., Pisoni, E., Degraeuwe, B., Kranenburg, R., Schaap, M., Clappier, A.: Dynamic evaluation of air quality models over European regions. Atmos. Environ. 111. doi:10.1016/j.atmosenv.2015.04.016, 2015.

Thunis, P., Degraeuwe, B., Pisoni, E., Ferrari, F., Clappier, A.: On the design and assessment of regional air quality plans: The SHERPA approach. J. Environ. Manage. 183, 952–958. doi:10.1016/j.jenvman.2016.09.049, 2016.

Thunis, P., Degraeuwe, B., Pisoni, E., Trombetti, M., Peduzzi, E., Belis, C.A., Wilson, J., Clappier, A., Vignati, E.: PM 2.5 source allocation in European cities: A SHERPA modelling study, Atmos. Environ, 187, pp. 93-106, 2018.

Thunis, P., Clappier, A., Tarrason, L., Cuvelier, C., Monteiro, A., Pisoni, E., Wesseling, J., Belis, C.A., Pirovano, G., Janssen, S., Guerreiro, C., Peduzzi, E.: Source apportionment to support air quality planning: Strengths and weaknesses of existing approaches, Environ. Int., 130, art. no. 104825, 2019.

Trombetti et al.: Downscaling methodology to produce a high resolution gridded emission inventory to support local/city level air quality policies, JRC Technical Report, 10.2760/51058, 2017.

Viaene, P. , C.A. Belis, N. Blond, C. Bouland, K. Juda-Rezler, N. Karvosenoja, A. Martilli, A. Miranda, E. Pisoni, M. Volta: Air quality integrated assessment modelling in the context of EU policy: A way forward, Environ. Sci. Policy, 65, 22-28, 2016.

Prioritising the sources of pollution in European cities: do air quality modelling applications provide consistent responses?

Bart Degraeuwe, Enrico Pisoni, Philippe Thunis

European Commission, Joint Research Centre (JRC), Ispra, Italy

Correspondence to: E. Pisoni (enrico.pisoni@ec.europa.eu)

Supplementary material

In this section we show the results of the basecase validation for the 2 model configurations considered in this paper. At first, comparing measured PM2.5 yearly averages (from the AirBase European Environmental Agency, EEA, database) VS the model results, with CHIMERE referring to the 2009 year and EMEP to 2014. Then, we also compare the results of the Source Receptor Relationships (SRR) against the model results, for a selected number of emission reduction scenarios.

In both Figures 1 and 2 (related respectively to CHIMERE and EMEP configurations) the scatter plot show the measurements (x-axis) VS the modelled results (y-axis), with PM2.5 in μg/m3. Each point represents the PM2.5 yearly average for a station, and each scatter represents a specific country. Results are quite comparable, with Poland and Austria being slightly better for CHIMERE, and Spain and Italy being slightly better for EMEP.

Figure 3 shows modelled versus measured PM2.5 concentrations at background stations for CHIMERE in 2009 (left) end EMEP 2014 (right).

Figures 4 and 5 show the validation of the SRR when used for simulating scenarios. For these two Figures, the validation is not done against observations (we do not have observations for scenarios) but against the results of the CHIMERE and EMEP runs. Both validations include simulations with reductions of various precursors and sectors, performed over different spatial geographical entities:

- at EU scale;
- at national scale (France and Poland);
- at regional scale (Katowice, Milan, London, Barcelona, Athens and Stockholm, with domains of roughly 100 km^2 around the cities);
- at local scale (Katowice, Milan, London, Barcelona, Athens, Stockholm, Antwerp, Porto, Paris, Berlin, Clermont-Ferrand, Copenhagen and Sofia, with domains of few tens of km^2 around the cities).

More details on the validation strategy can be found in Pisoni et al., 2017 and Pisoni et al., 2019.

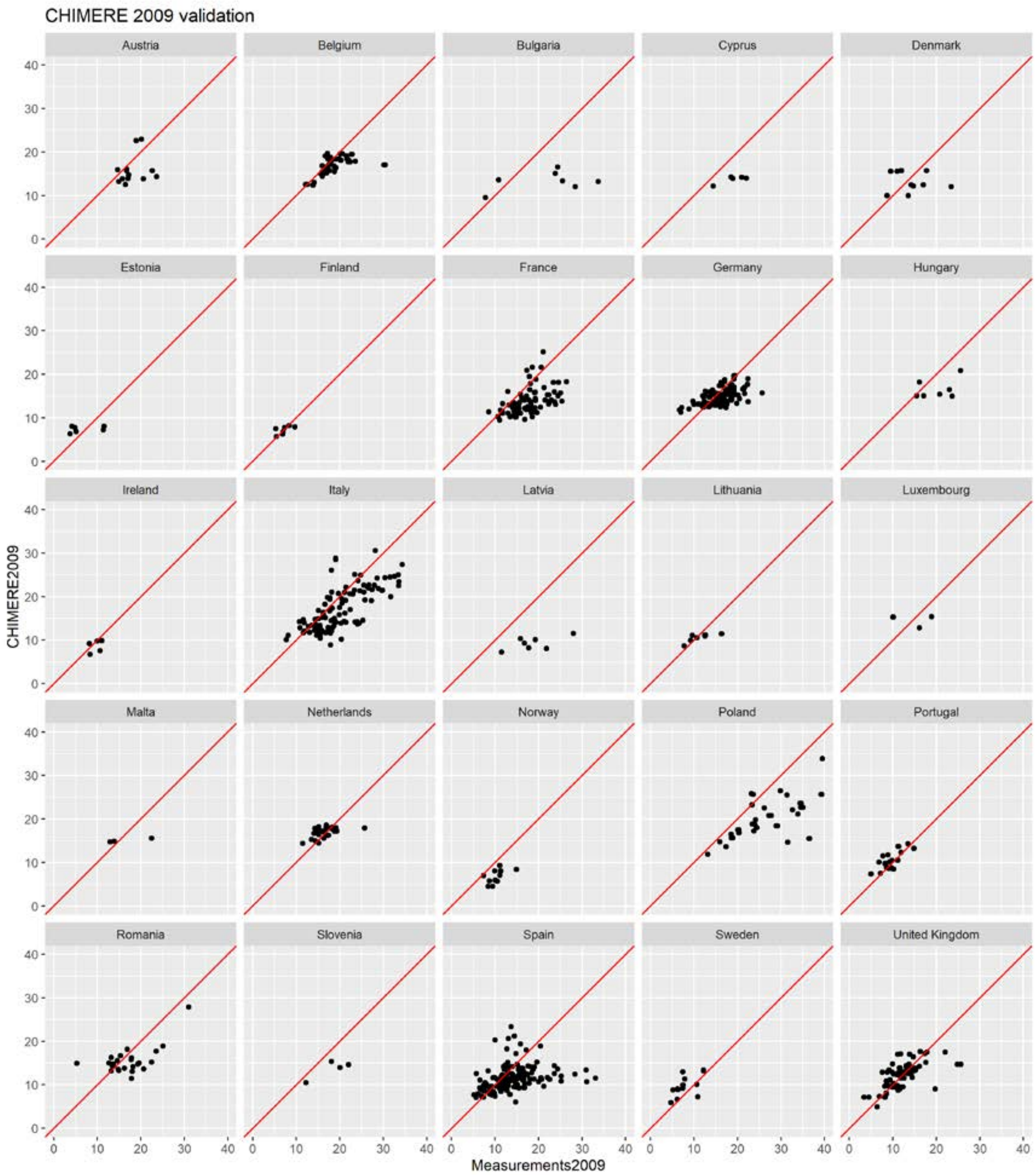

Figure 1: CHIMERE 2009 validation results, comparing observations VS model results.

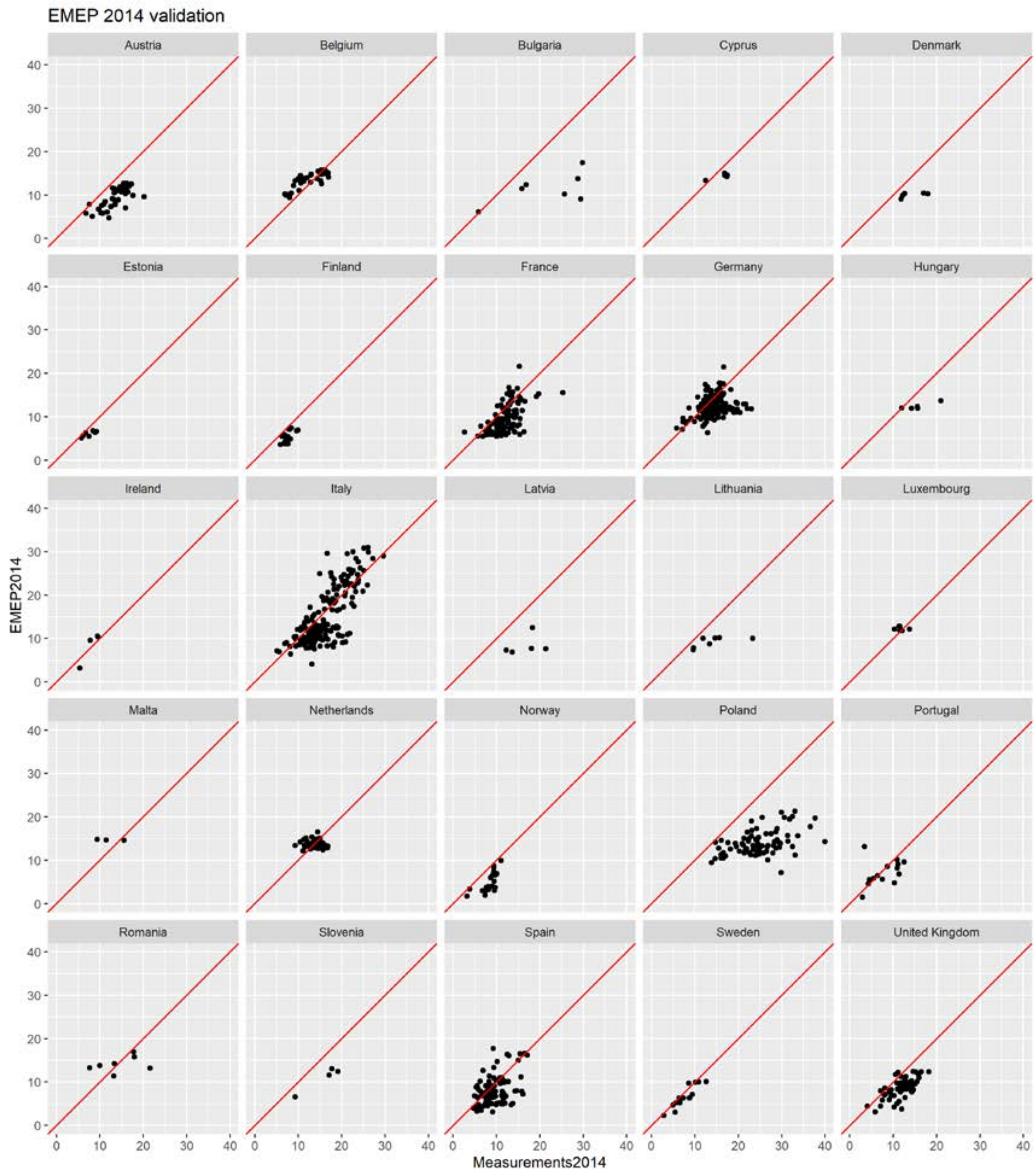

Figure 2: EMEP 2014 validation results, comparing observations VS model results.

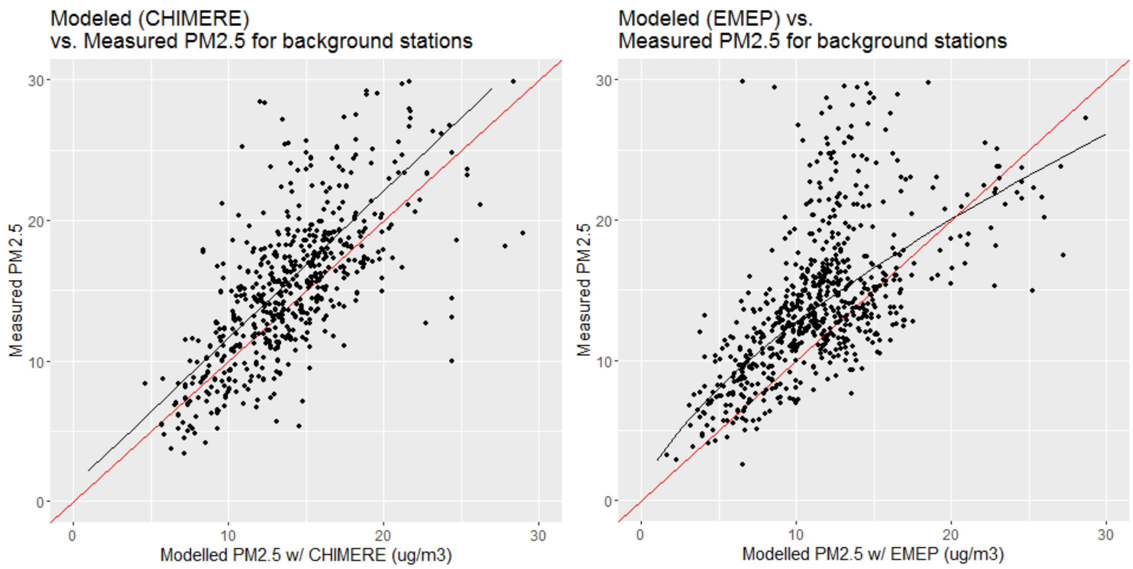

Figure 3: Modelled versus measured PM2.5 concentration at background stations for CHIMERE in 2009 (left) end EMEP 2014 (right).

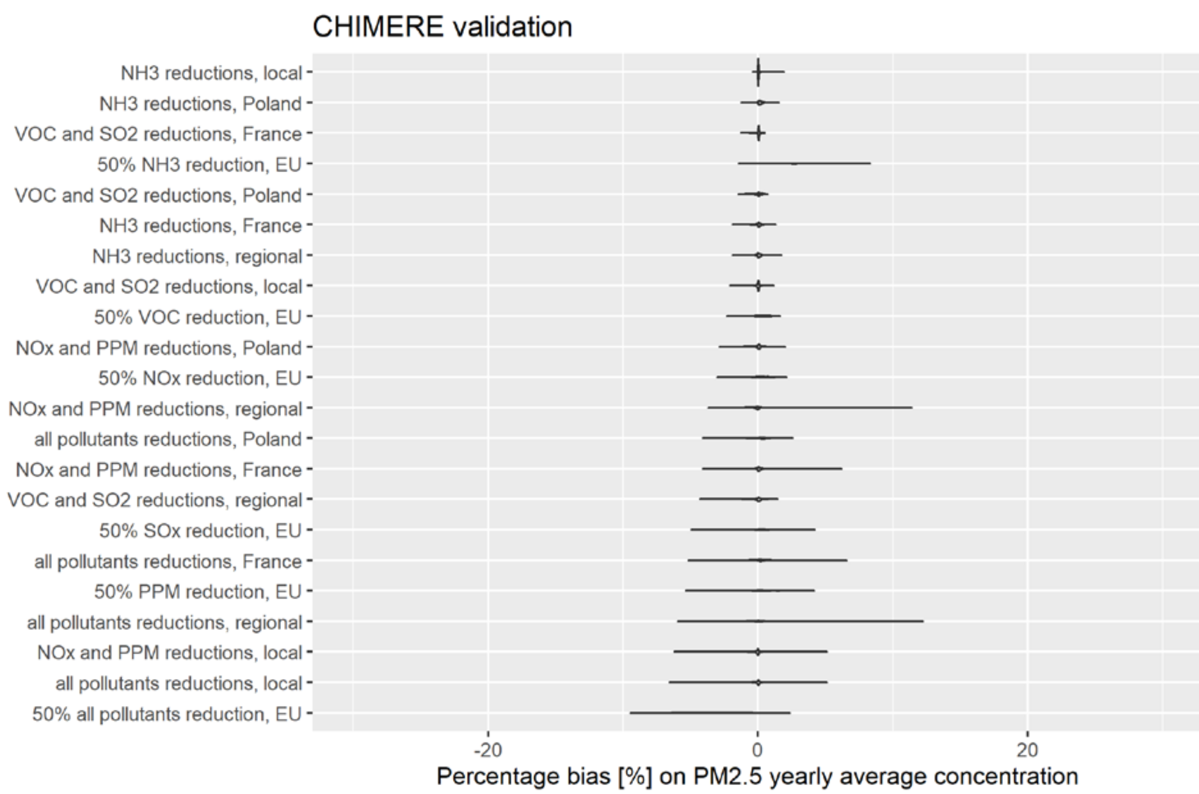

Figure 4: validation results for the SRR (in terms of percentage bias) on a number of CHIMERE scenarios. The percentage bias is computed by comparing the SRR and CTM results, for each point of the domain. On the y-axis all the considered scenarios are listed specifying, for each scenario, the reduced pollutants (NOx, PPM, …) and the domain over which reductions are applied (EU, France, …). Note that for the regional and local scenarios, all domains are included in the same visualisation.

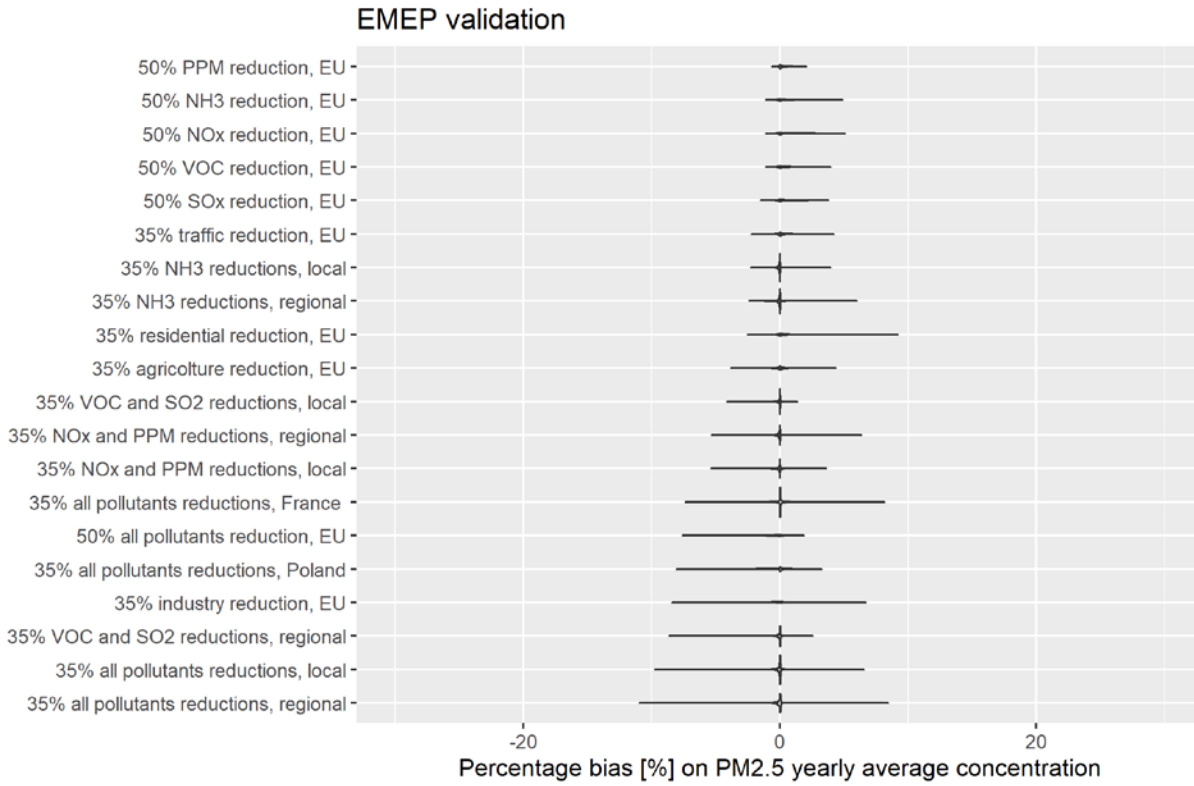

Figure 5: validation results for the SRR (in terms of percentage bias) on a number of EMEP scenarios. The percentage bias is computed by comparing the SRR and CTM results, for each point of the domain. On the y-axis all the considered scenarios are listed specifying, for each scenario, the reduced pollutants (NOx, PPM, …) and the domain over which reductions are applied (EU, France, …). Note that for the regional and local scenarios, all domains are included in the same visualisation.