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)

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 main contributions of this work and we will better stress this point in the revised version of the paper.

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 implementations. 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… We will refer to this point in the revised document.

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.

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.