The study by Pommier et al., demonstrates the ability of two modelling setups to identify source contributions of particulate matter from different countries to multiple capital cities in Europe during a pollution episode. Overall, the paper does indeed demonstrate this and after some major revisions it should be suitable for publication in GMD. The main concerns I have with the manuscript is its lack of clarity in places. Firstly, the description of the source-receptor calculations needs to be more clearly discussed as it is not easy to follow to non-experts of this methodology. Secondly, the manuscript is compiled of lots of short sentences which lead to a stop-start flow with makes the manuscript more difficult to read. Thirdly, the comparisons between the models and the observed PM concentrations are satisfactory at best. For instance, many of the correlation statistics between the models and observations are below 0.5 and the mean biases (magnitude) are sizable. Therefore, I feel the authors really need to stress that the model comparisons are "satisfactory" and clearly state whether the metrics presented (e.g. P11) show the models are doing well or badly when compared to the observations.

The authors would like to thank the reviewer 1 for his comments which help to improve our study. We have tried to clarify the points raised by the reviewer and to answer all remarks. Our responses are written in blue in this document.

Finally, some of the figures are too busy and need to be made clearer. For instance, Figure 6 is overly complicated and takes a long time to fully digest.
The Fig 6 and the text have been changed (see your last point).

Also, the "agreement in the dominant contributor" in figure 8 between the models is not clear. How is this agreement determined? What statistical metrics are used? If this is already stated, then please make it clearer!
We agreed it was a missing information. It has been added in the text.
"This rate corresponds to the number of occurrences in the dominant contributor calculated for each hourly concentration in the 4-day forecast over each city. So, a number as 100% over a city shows that both models predict the same dominant country contributor during a 4-day forecast."
And (in bold):
The mean agreement increases up to 75% for determination in the top 5 of the main country contributors to $PM_{10}$ (Fig 11). **In that case, the rate is calculated for the five main country contributors. A score of 100% means both models predict the same five main country contributors for each hourly concentration, but not necessarily in the same order.**"

Minor Comments: P3 L71-73: Provide reference for the WHO health metric stated.
The reference has been added.

P3 L84: Space between "VOCs). The".
Done

P4 L99-101: Please explain in detail how "source" and "receptor" are related in this work to make it clear for readers not familiar with this method.
Additional information has been added (in bold):
"With a such simulation comparison, the simulation with reduced emissions over a source region **(e.g. a country)** allows to highlight the impact of this source on the concentrations over a receptor, **hereafter a city**".

P5 L158: What is the new land-cover dataset used?
The land-cover dataset merges information from GLC-2000 data-set (http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php) and CLM database (http://www.cgd.ucar.edu/models/clm/).

The GLC2000 and CLM data-sets through the following procedure:
1. GLC2000 is used to define water, ice, urban and bare surfaces, and then 'high' and 'low' vegetation (HV, LV).
2. Where high vegetation is labelled as sparse, we allocate 50% as HV, 50% as LV.
3. Where low vegetation is labelled as sparse, we allocate 50% as LV, 50% as bare.
4. For each grid square we then allocate the HV and LV vegetation according to CLM categories.

This information is provided in Simpson et al. (2017) – see pages from 116 to 118.

P6 L168-169: Make it clear whether or not other BVOC emissions are used in the model other than isoprene and monoterpenes.
We have added these following sentences:
"The soil-NO emissions of seminatural ecosystems are specified as a function of the N-deposition and temperature (Simpson et al., 2012). The biogenic DMS emissions are calculated dynamically during the model calculation and vary with the meteorological conditions (Simpson et al., 2016)."

P6 L170: The definition of the "remainder" is unclear.
It has been changed. Now it reads:
"… the rest of primary PM defined as the remainder".

P7 L212-213: What do the authors mean by "we have harmonized the used of different parameters"? Do you mean that the setup and input/outputs of the model are been made as consistent as possible?
The following information has been added (in bold):
"To perform properly the analysis between both models, we have harmonized the use of different parameters **such as the horizontal resolution, the anthropogenic emissions used, the definition of the city area and meteorological data used** (Tab. 1)."

P8 L 222: Worth saying that the ECMWF operational system does not archive 3D precipitation fields when this is first discussed on P6.
The information has been added (in bold):
"An estimation of this 3D precipitation can be calculated by EMEP if this parameter is missing in the meteorological fields **as in the data used in this work (see Section 2.4)**."

P8 L 247: Can the authors please elaborate on what they mean by "medium intensity".
The information "no more than three consecutive days beyond the WHO $PM_{10}$ threshold" has been added to the sentence.

P9 L256: What quantifies as "large concentrations"?
It has been added:
"Large concentrations **(>60 µg/m$^3$)** were also predicted…"

P10 L277: I suggest the authors change the word "enormous" to something more scientific.

It has been changed (in bold):
"...to calculate **very large** concentrations **(e.g. hourly concentration higher than 200 µg/m$^3$)**…"

P11: L307-312: I suggest the authors re-write this paragraph as it is unclear and difficult to follow.
It has been re-written. The main corrections are highlighted in bold:
"LOTOS-EUROS **is less correlated** with the concentrations measured by the rural stations than EMEP (Fig. 4). However, **as EMEP,** LOTOS-EUROS also presents **a lower bias with these rural stations in comparison with the urban stations. This** is predictable since with such resolution, the model calculates mainly the urban background concentrations. By comparing the 5 cities having urban and rural stations, as done with EMEP, only the bias and the FGE between the predictions and the **urban** measurements are improved (Fig S4). It is also worth noting that the concentrations predicted by LOTOS-EUROS over these 5 cities are lower than **the ones calculated by** the EMEP model **(in Fig. S3).**"

P11: The discussion of the different metrics is a bit over-kill here. If all this discussion is to be kept in the manuscript, can the authors at least specify what the numbers mean in terms of model performance (e.g. R=0.72 is reasonable and R=0.25 is poor).
This information has been added in Section "3.2.1 Methodology".
"By knowing this point, we have stated that a comparison with the observations presenting for example a correlation coefficient equal to 0.5 or NMB lower than 15% are reasonable results ($r \geq 0.7$ and NMB $\leq 10\%$ are good results)."

P13 L365-371: As mentioned above, I think the authors need to discuss in more detail the source-receptor methodology to make it clear to non-experts of this approach.
It has been rewritten. The changes are highlighted in bold:
"There are in total 31 runs **for each date with reduced anthropogenic emissions. Each run** corresponds to the perturbations for **one of the** 28 countries **related to the 28 EU capitals**, plus Iceland, Norway and Switzerland, giving the contribution for each country.
To calculate the concentration of the pollutant integrated over the studied area, i.e. a selected city, coming from a source, we follow the equation (5):

$$C_{source} = \frac{C_{reference} - C_{pertubation}}{x} \quad (5)$$

With x the reduction in % (i.e. 0.15), $C_{reference}$ is the concentration of the pollutant integrated over the studied area from the reference run and $C_{pertubation}$ is the concentration of the pollutant integrated over the studied area from the perturbation run. **Thus, by differentiating over the studied area, the concentration from the perturbated run with the concentration provided by the reference run, we have an estimation of the influence of the source (i.e. country). By scaling with the reduction used (parameter x), it gives the estimated concentration related to the source**."

P14 L415: "For the positive correlation, a clear feature appears" is an example of these short sentences which break the flow of the text.
We have supposed that the reviewer wanted to quote "For the positive concentrations, a clear feature appears".
It has been deleted and replaced by the part in bold:

"The main contributors to the "Domestic" $PM_{10}$ are POM (~20%) and rest PPM (~30%) (which corresponds to the remainder of coarse and fine PPM), **as noticed for the positive concentrations** (Fig. 6a)."

P15 Section 5.2: If the LOTUS model is using a different approach to that of EMEP, how are the emissions perturbed? This is not overly clear from the text as it stands.
As explained in the introduction, LOTOS-EUROS does not use an emission perturbation scenario but a labelling technique. Thus, the model traces the pollutants through conserved atoms (C, N, S) related to emission sources.
This technique is described in Section 4.2.

P16 L458: What do the authors mean by "each at the end of the EMEP model"?
We have supposed the reviewer wanted to quote "at the opposite of the EMEP model".
We agreed that this sentence was confusing. It has been changed as below:
"As reminder, the EMEP model predicted a slightly larger influence from the "30 European countries" (35%) than from "Others" (25%)."

P16 L459-462: The term "Rest" appears to represent the difference between the total PM and the sum of all its components. Is this the metric used to explain the "non-linearity in the chemistry? If so or if not, I think this sentence need to be rewritten to clear emphasis the definition of "Rest".
The sentence has been changed as:
"In the list of LOTOS-EUROS $PM_{10}$ components there is one named "Rest". "Rest" corresponds to the difference between the total $PM_{10}$ and the sum of all the components, and Fig. 8 shows that it is also a large component of this "Domestic" $PM_{10}$".

P16 L467-8: Is this true? In section 3 I got the impression there was substantial disagreement between the models.
That is certain that both models underestimate the larger peaks observed over the cities. However, both models agree between their predictions.
The reader must remind that the predictions from both models are representative for a large area and will obviously underestimate the concentrations and the contributions for the larger peaks measured by a specific station.
Thus, we have added this sentence in Section 6:
"It has also been shown in Section 3 that both models are representative for a large area and the predictions can underestimate the concentrations and the contributions for the larger concentrations measured by a specific station."
And in the conclusion:
"It may suggest that the both models, which calculate the country contributions over the cities, defined by a large area, may underestimate the contribution measured by a specific station for the higher concentrations."

Figure 2: Could the country outlines be more clearly plotted.
It has been changed.

Figure 6: There is a lot of stuff is this plot, so could be good to make it simpler or bigger at least so easier to see everything. The calculation of non-linearity need to be explained more clearly in the manuscript.
The figure has been split into two parts. Moreover, an explanation in the text has been added:

"This non-linearity has been calculated for each hourly concentration as the standard deviation of the hourly contribution (which can be positive or negative) obtained by the three reduced emissions scenarios and weighted by the hourly total concentration by following the equation (6):

$$NONLIN_{Contrib} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Ccontrib_i - \overline{Ccontrib})^2}{n}}}{Ctot} \times 100\% \quad (6)$$

n corresponds to the number of perturbations used (n=3), Ccontrib is the hourly $PM_{10}$ concentration for a specific contribution ("Domestic" or "30 European countries" or "Others") and Ctot is the hourly $PM_{10}$ concentration."