Dear Editor,

We have answered all the points given by both reviewers and the comments provided by A. Clappier, P. Thunis and the executive editor D. Ham.
All our answers are written in blue.

Best regards,
The authors.

Reviewer 1

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):

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

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."

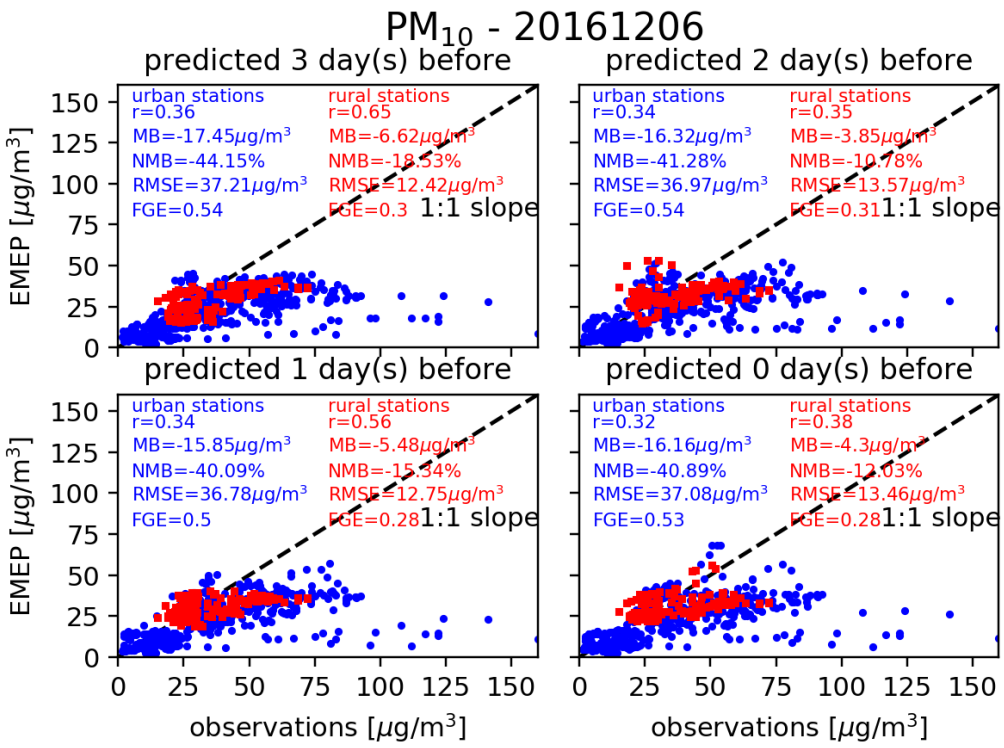
Reviewer 2

We would like to thank the reviewer 2 for the helpful comments and corrections. We have answered the different points by highlighting our comments in blue.

This paper compares two source apportionment methods. The methods are not clearly explained. Some clarifications are needed and there are some methodological flaws. Also the English used in this paper needs to be revised.
Some parts of the introduction have been rewritten.

Specific comments: 1. About the comparison between measured and modeled concentration. I understand that the author wants the compare the average concentration over an urban area. From model results it is easy to obtain this, averaging concentrations over some grid cells. Unfortunately you cannot obtain a comparable number from measurements. The stations are not equally distributed over the area of interest and the number of urban, rural and traffic stations might be different. Comparing an average of the stations with an average of the grid cells will introduce an additional uncertainty. Why not interpolate the model results at the station locations and compare with the measurements. A separate comparison for different station types should be made. I think the analysis for one cell, 9 cells and the GADM. I would restrict the analysis to stations inside the GADM.
The comparison by interpolating the model grid cells to the stations present similar results as shown in the paper. The following examples present such comparison and they are comparable to Figs 3 & 4 in the ACPD manuscript.

EMEP



LOTOS-EUROS

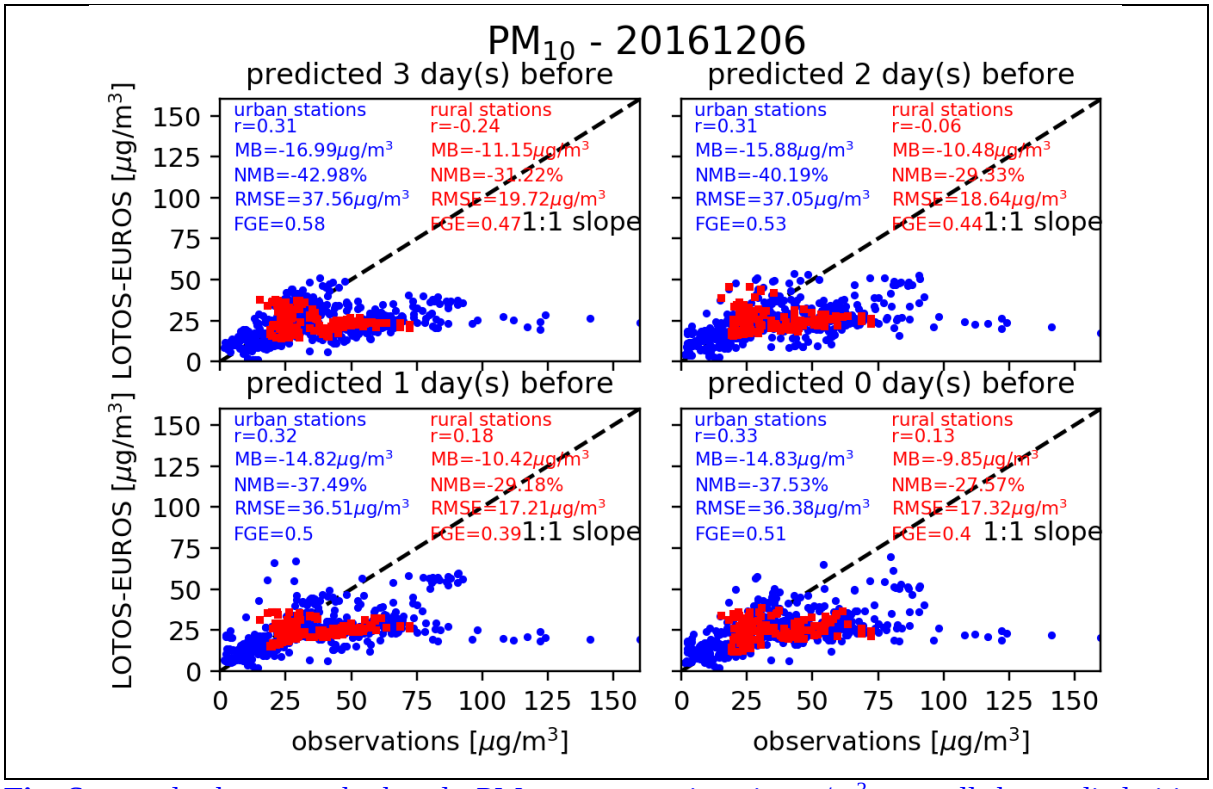


Fig. Scatterplot between the hourly PM_{10} concentrations in µg/m^3 over all the studied cities using the 9 grid cells definition, predicted by the EMEP model (top), LOTOS-EUROS (bottom) on December 06th 2016 and the observations of the urban sites (blue dot) and rural sites (red square). The EMEP predictions are interpolated to the observations. The four panels correspond to the different predictions from 3 days before the December 06th to the actual day, i.e. December 06th. The correlation coefficient (r), the mean bias (MB), the normalized mean bias (NMB), the root-mean-square error (RMSE) and the fractional gross error (FGE) are provided on each panel.

However, such comparison does not answer the question about the reliability of our predictions over the cities since the objective was to compare the average concentration and thus the average contribution over each city domain.
We also agreed with a limited number of stations per city, the comparison remains difficult and a regional model will never predict the same concentrations than these sparse measurements.

We have also decided to present the evaluation of the prediction over the cities for the 3 definitions, since we aimed to test different definitions in our products. We also wanted to highlight the importance of the definition of the city boundaries to determine the country contribution.

2. The non-linearity discussed Line 374 and following.
The contributions of individual countries don't have to theoretically sum up to the contribution of all countries reduced together. Even for small reductions there is some non-linearity. But the non-linearity is small for small reductions. The difference between the sum of individual contributions and the joint contribution can be positive or negative. I would not speak about

negative concentrations. You scale up to 100% but in fact you do a source apportionment of the top 15% of the PM10 column. That's perfect and useful for policy. Achieving small emission reductions is already hard enough.
It is an interesting comment from the reviewer. We have however preferred to keep the negative concentrations, since these concentrations highlight the compounds involved in the non-linearity (NO3, H2O and NH4).

3. Validation versus measurements.
The validation shows quite big differences between model and measurements. What is the impact of this error on the source apportionment? To which extent can it be trusted? In regard of this error, which differences between the two methodologies are significant? How certain is it that the biggest contribution is really the biggest?
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."

4. Figure 6 Maybe it is more useful to present the analysis for some selected cities (and the others in Annex) than for all cities together. The behavior can be quite different across Europe. If non-linearity is small plots for one reduction percentage are sufficient. It is not clear to me which runs were done to obtain these plots.
We have decided to keep the overall description in Fig. 6.
However, the part describing the non-linearity (black horizontal bars) has been shown in another figure (now Fig. 7).

By providing a figure as Fig. 6 for each city will add complexity.
It is right that the impact of non-linearity is not similar for each city. Thus, we have decided to add an additional figure in the supplement with the following text, showing this non-linearity over each city.

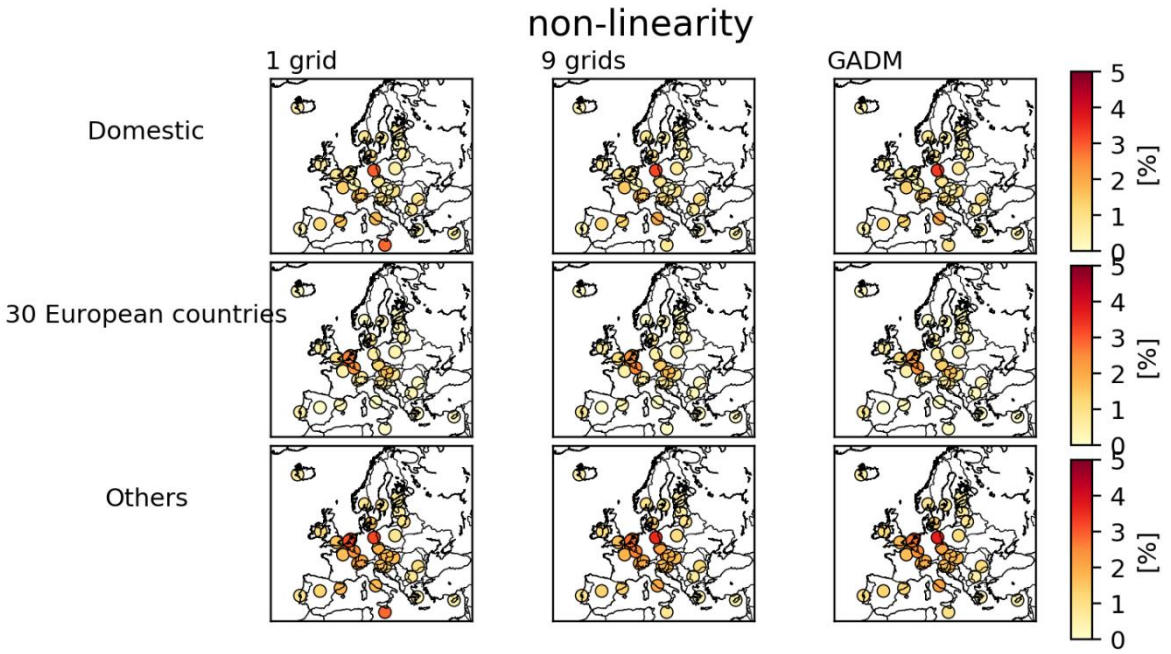


Figure. S10 Mean hourly non-linearity in percent calculated for the "Domestic", "30 European countries" and "Others" contributions, over the 34 European cities and for all 4-day forecasts (i.e. from 01-04 Dec to 09-12 Dec 2016). The non-linearity is presented for the cities defined by 1 grid cell (left row), 9 grid cells (middle row) and by the GADM (right row).

In Section 5.1:
"The mean non-linearity is not homogenously distributed over all cities as shown in Figure S10 and may vary from date to date (not shown). It has remained limited even if some hourly contributions show higher non-linearity. In maximum, 3% of the calculated hourly contributions for all 4-day forecasts over the selected cities have a non-linearity higher than 5% (not shown)."

Indeed, even if some hourly non-linearities may present larger values, the amount of these large non-linearities is limited as shows with this distribution for the 9 grid cells definition:

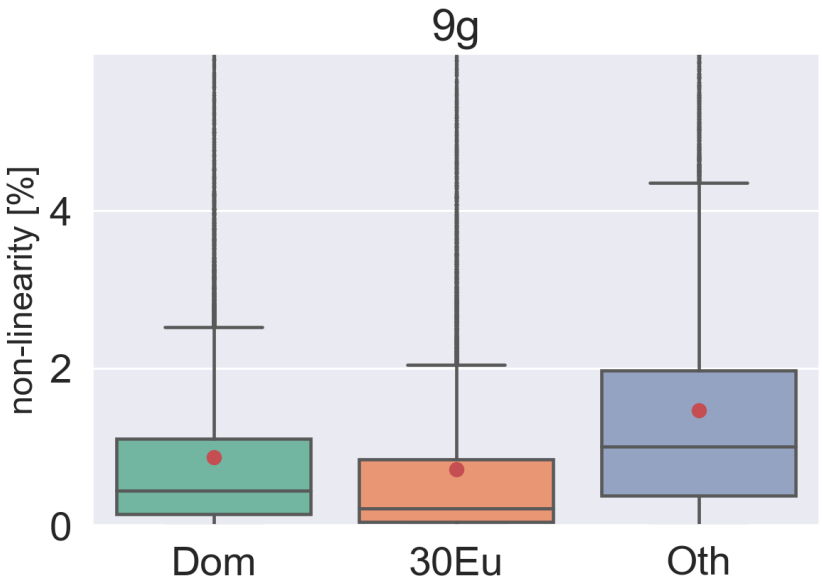


Figure: Distribution of the non-linearity for the "Domestic", "30 European countries" and "Others" contribution. The line that divides the boxes into two parts represents the median of the data. The end of the boxes shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers which are represented by grey diamonds (almost seen as a line). The red dots correspond to the mean of each data set.

We have also added these sentences in the conclusion:
"Even if this non-linearity is not identical for all cities and for the different dates, the larger non-linearities (>5%) impact only 3% of all the calculated hourly contributions. However, the non-linearity related to the reduction of each emission precursor has not been calculated in the study for computational reason."

Concerning the runs used for this figure, we have added an additional information in the text:
"This figure is a result of the perturbation runs by separating the positive and the negative concentrations obtained in the calculations. The concentrations have also been gathered by their calculated origin".

On line 358 is mentioned that emission per country where reduced with 15%. Are precursors reduced one by one or all together?
All the anthropogenic emissions are reduced simultaneously. It was explained in the following sentences (lines 361-363). However, we have added the word "anthropogenic" since it was missing.
"The perturbation runs are done for **anthropogenic** emissions of CO, SO_x, NO_x, NH_3, NMVOC and PPM (primary particulate matter). For computational efficiency, in the perturbation calculations, all anthropogenic emissions in the perturbation runs have been reduced here simultaneously."

As also mentioned previously, there is now this sentence in the conclusion:
"However, the non-linearity related to the reduction of each emission precursor has not been calculated in the study for computational reason."

How is the non-linearity calculated? Is it calculated as a share of the total concentration (Line 506). In my opinion it is more correct to use the concentration change as reference?
In the section 5.1, we have added an explanation about the calculation in the non-linearity:

"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):

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

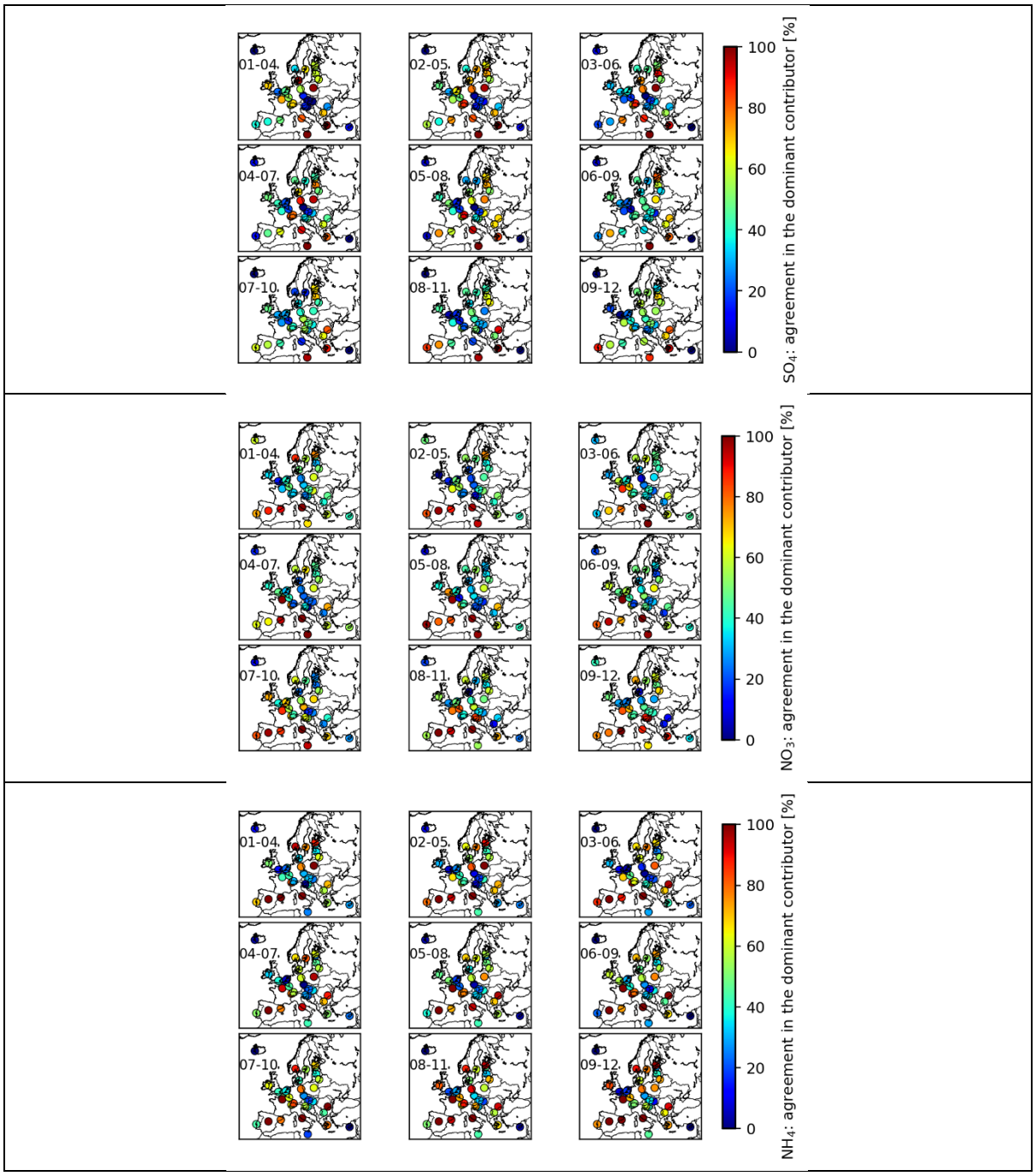
n corresponds to the number of perturbations used (n=3), Ccontrib is the hourly PM_{10} concentration for a specific contribution ("Domestic" or "30 European countries" or "Others") and Ctot is the hourly PM_{10} concentration."

It is important to remind, that our calculated contribution (Eq. 5), corresponds to the change in concentrations related to the change in emissions.

5. Comparability of the two methodologies. For primary pollutants both source apportionment methodologies are comparable. Differences are due to differences in the models (transport, deposition,..). But for secondary PM the methods don't necessarily give the same result. E.g. an amount of NOx emitted somewhere can result in a certain ammonium nitrate concentration in the receptor. If NOx is emitted in excess (ammonia limited regime) an emission reduction will have little effect at the receptor point. On the other hand, in the NOx limited regime the same NOx reduction will have a big impact. The labeling method will give the same result in both cases while the 'perturbations' method will give different results. Hence, comparing contributions calculated by the two models is not very useful. The statement on line 513 is not complete: differences are not only due to differences in aerosol chemistry between the models.
The reviewer raises an important point.
The following sentences have been added in Section 6:
"It is also related to the differences in both methodologies (e.g. Clappier et al, 2017b). Indeed, an emission reduction and a labelling technique will not necessarily provide the same results for the secondary PM. An emission reduction depends on the atmospheric composition already present. For example, an amount of NO_x emitted over a source can result in a certain NH_4NO_3 concentration in the receptor. If this NO_x is emitted in excess (NH_3 limited regime), a NO_x emission reduction will have a small effect at the receptor point. On the other hand, in the NO_x limited regime, the same NO_x reduction will have a large impact. The labelling method will give the same result in both cases while the scenario approach will give different results."

6. Figure 8 How is the percentage of agreement defined? I think it's more useful to present this for individual cities.
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.**"

In Fig. 8 we have decided to show the mean agreement since to present 34 figures will be unreadable. However, the agreement was calculated for each other compound, as shown below:



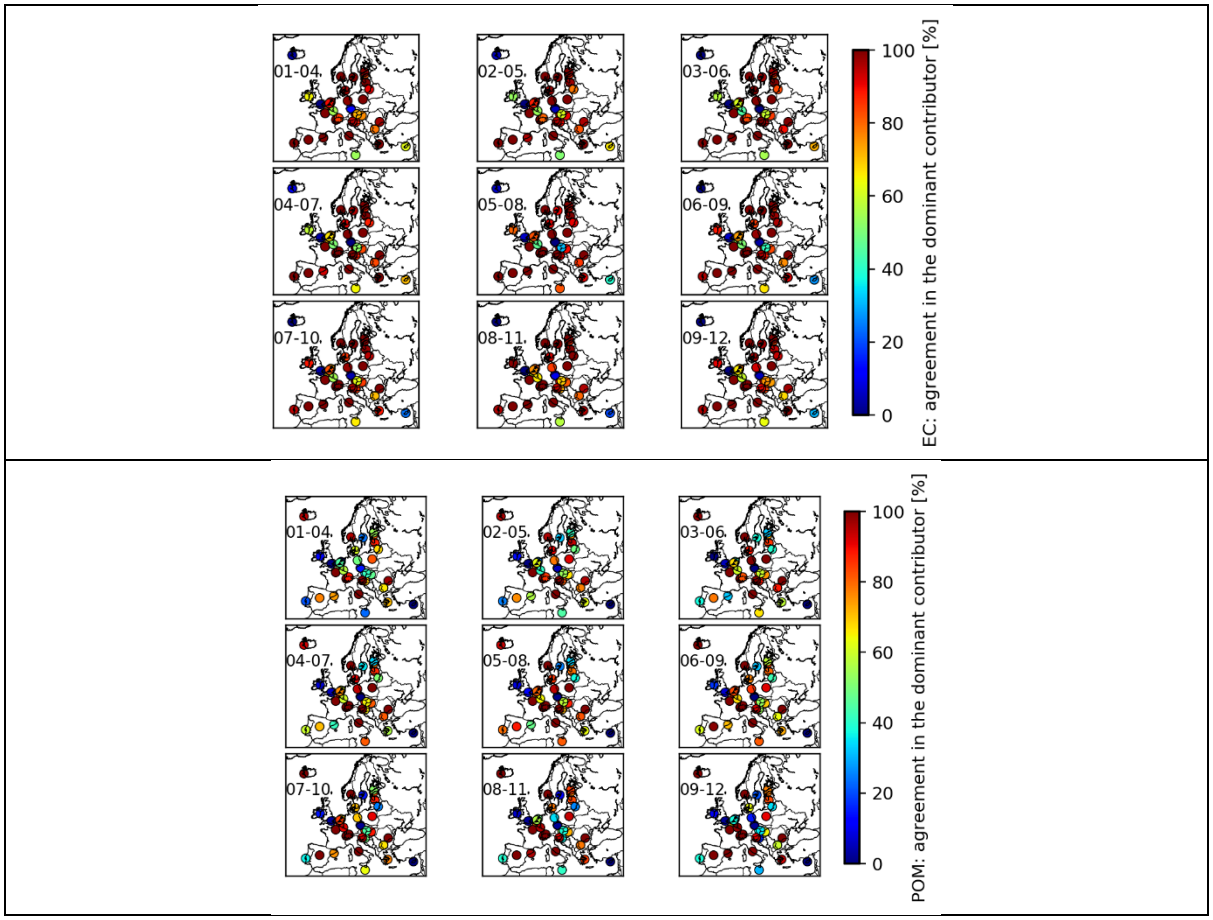


Fig. Agreement in the determination of the dominant country contributor for SO_4, NO_3, NH_4, EC and POM in percent, and for each 4-day forecast (01-04 Dec 2016, 02-05 Dec 2016, 03-06 Dec 2016, 04-07 Dec 2016, 05-08 Dec 2016, 06-09 Dec 2016, 07-10 Dec 2016, 08-11 Dec 2016, 09-12 Dec 2016) over all the cities using the 9 grid cells definition.

Technical comments:
Line 30: change '15% factor' > '15% emission reduction'. This 15% is not a factor. I think it's confusing. Change this in the whole paper.
It has been changed.

Line 3033: revise grammar, sentence too long. We found that the combination of a 15% reduction and a larger domain help to reduce...
It has been corrected.

Line 36: split sentence
Done.

Line 68: crops yields > crop yields
It has been corrected.

Line 71: states > better established/proposed a PM10 limit value

Changed.

Line 77-79: very unclear contradictory sentence. If a pollutant has a short life time it's impact is close to the source and long-range transport doesn't matter much. Is PM10 really so short lived compared to other pollutants (like NO2). The concentration of PM10 is rather uniform compared to the latter.
We have added the following information (in bold):
"Due to the **relative** short atmospheric life time **(from some hours to days),** the variability is impacted by local sources, meteorological conditions affecting dispersion and long-range transport as well as chemical regimes controlling the efficiency of secondary formation."

Line 81: atmospheric processing? > formed by chemical reactions in the atmosphere.
It has been changed as requested.

Line 85: traffic and transport, all traffic is transport
The word "traffic" has been deleted.

Line 86: biomass burning refers to burning wood for heating. It is an anthropogenic source. You mean wild fires?
It has been replaced by "forest fires".

Line 95: Revise grammar and content. With a country source calculation...????
It has been changed.
"A country source calculation allows to tackle the emissions from the countries responsible for the air pollution episode."

Line 98: revise. Something like: The EMEP calculations use reductions of anthropogenic emissions...
Now it reads:
"The EMEP calculations use reduced anthropogenic emission scenario and compare to a reference run where no changes are applied."

Line 115: Both models are part of...
It has been changed.

Line 122: Use a consistent terminology. You say 'SR system' and the next paragraph is called 'SA system'. A source receptor (SR) model is not a source apportionment (SA) system. Check this through the whole paper. Line 123 and 126: SR? I think it you mean source apportionment. Line 130: be consistent. SR product (should be source apportionment product)... real-time source allocation (=? source apportionment)
Thank you to notice this error.
To be consistent, we have used the term of "source contribution" as presented on the website.
https://policy.atmosphere.copernicus.eu/SourceContribution.php

Line 132: for the 28 EU capitals, plus Bern, Oslo and R.
Changed.

Line 145: too long, split up
Done.

Line 153: ...but the model has also been used...
It has been changed.

Line 161: sigma coordinates? There are... > grammatically incorrect sentence
It has been replaced by:
"The PBL is located within approximately the 10 lowest model levels…"

Line 212: word order! ...cover a slightly different domain...
Corrected.

Line 119: ...by the IFS
"the" has been added.
Since it was the first time that IFS was defined in this paper, we kept the definition and now it reads: "by the Integrated Forecasting System (IFS) of ECMWF".

Line 236: 1 (grid) cell ... 9 (grid) cell... There is only one grid.
It has been corrected.

Line 237: The latter...
It has been changed.

Line 238: ... living area... better 'urban area' or 'build up area'
It has been changed as "build-up area" as requested.

Line 244: BCs ? boundary conditions?
Yes, it corresponds to "boundary conditions". We forgot to define this abbreviation. It has been added when we explained the BCs used, i.e. at the former line 225.

Line 250: repetition
It has been corrected.

Line 255: until fronts moved in
It has been corrected.

Line 258: metrics
Corrected.

Line 259: To properly estimate
Corrected

Line 267: N is the number of the reference dataset? The number of what? Hours? Days?
We have added the following information (in bold):
"number of the reference data set **(e.g. number of observations)**."

Line 283: grid cell
Changed.

Line 285: city edge > city boundary
Changed.

Line 298: ...smoothed over a large domain... Do you mean smoothed over a grid cell? Mis-interpretation > underestimation. So, the correlation is similar for urban and rural stations but urban stations have a bigger bias. That's because peaks are smoothed out over the full cell.
It has been changed. Now it reads:
"In Figure 3, it is also clear that the EMEP model has difficulties to reproduce the highest concentrations measured by the urban stations which are probably smoothed by the model over **the large grid cells as the ones** defining the cities. **The underestimation in** the largest urban concentrations is highlighted by the comparison with the rural stations."

Line 303: By comparing only the 5... remove the comma
Done.

Line 309: grammatically incorrect
It has been corrected.

Line 312: ...than the ones from the EMEP model
Now it reads:
"…than the ones calculated by the EMEP model".

Line 315: globally > In general
It has been changed.

Line 318: ...at the urban...
It has been changed.

Line 332: negative correlation coefficients? Can you explain this better?
It is tricky to interpret these negative coefficients with such limited number of stations. We have added a sentence:
"The correlation coefficient with the rural stations remains difficult to interpret related to the limited number of stations available."

Line 393: only ... as well as... confusing formulation
"as well as" has been replaced by "and".

Line 401: ...one source area.
Corrected.

Line 445: Averaging out over more cells reduces non-linearity. I would not use the term 'negative concentrations'
It has been changed. Now it reads:
"Averaging out over the larger grids reduces globally the non-linearity."

Line 454: confirms the global feature > shows the same trend (?)
It has been replaced by "general trend".

Line 475: reformulate
Done. Now it reads:
"showing that the mean value in the agreement for both compounds **is reduced** by a few low values"

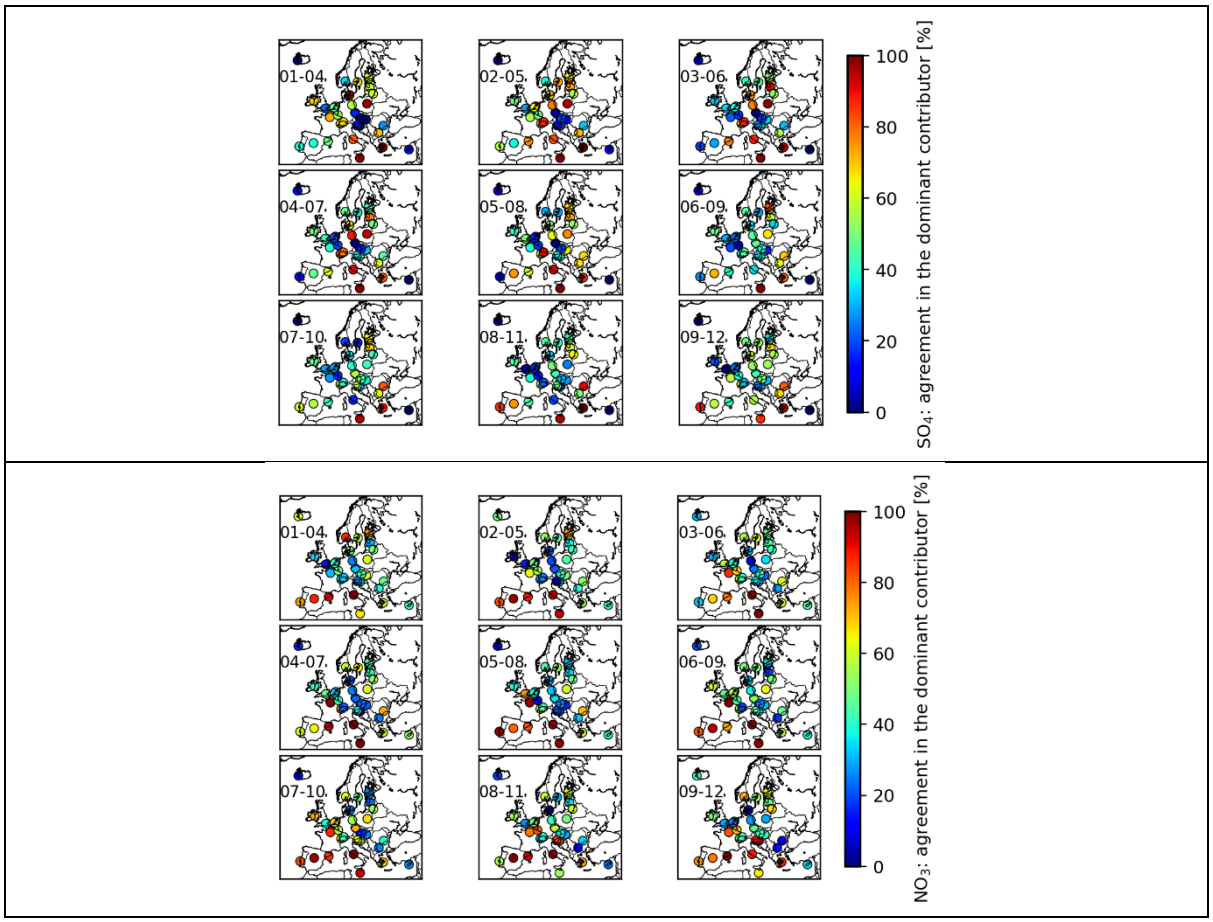
Line 522: ...probably foresees an underestimation... unclear formulation.
“foresees” has been replaced by “suggests”

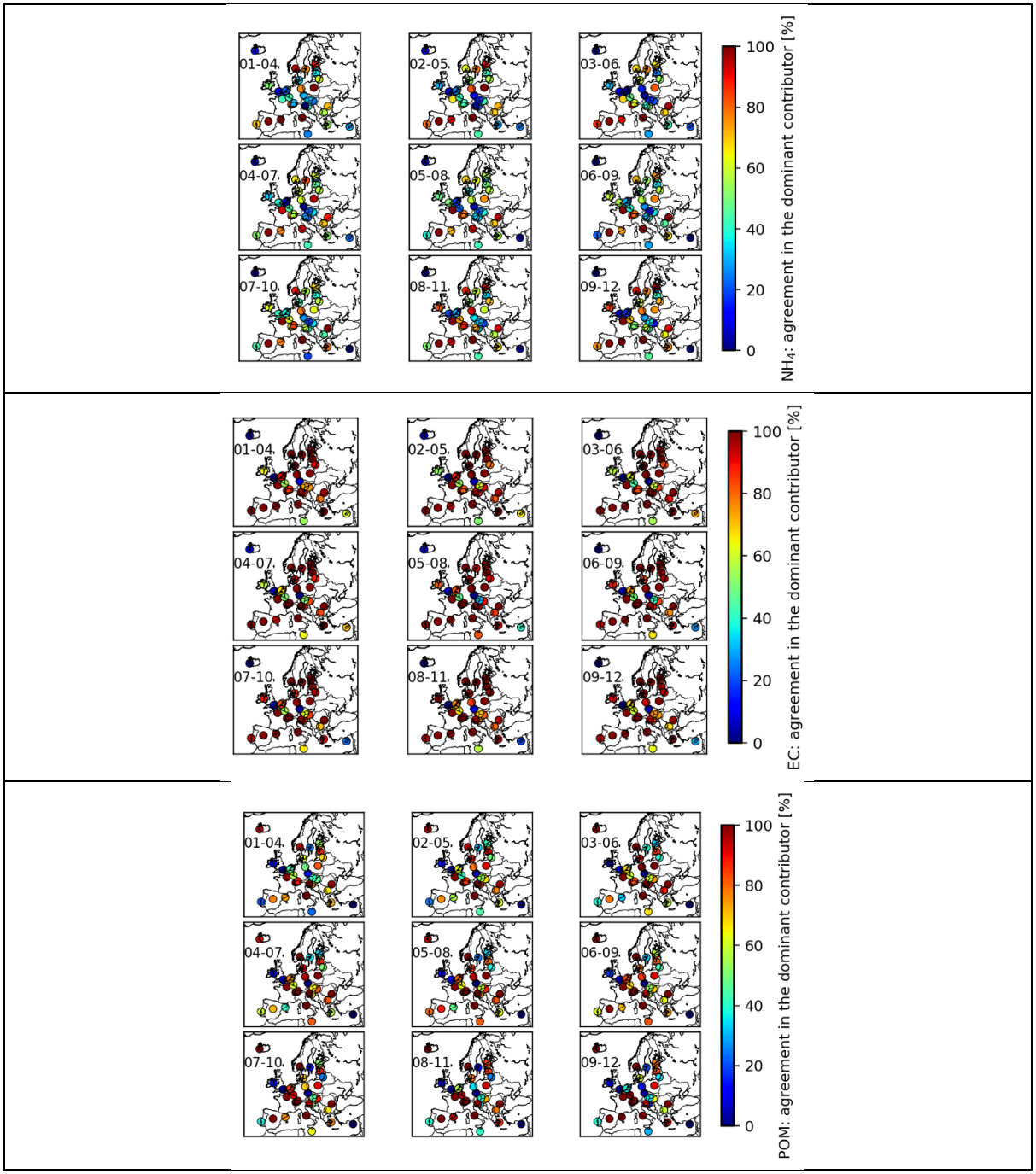
P. Thunis

We would like to thank Philippe Thunis for his helpful comments. We have tried to answer all his remarks below.
We also apologize for our missing points in our analysis and we are very grateful for all the details provided by Philippe Thunis.
We also would like to remind that we aimed to provide in this study, the origin of the PM under current conditions and thus to provide the information which sectors/regions to target. Effectivity of measures as such is not aimed for here.

In this publication, the Authors compare two different source apportionment approaches (referred as "scenario" and "labeling") and conclude that they reach a satisfactory agreement (68% for PM10, 50% for SIA) between the two methods for a few-days episode analyzed in more than 30 cities. This claim of a satisfactory agreement is based on numbers that represent average results across cities and forecast days. While this average agreement/non-agreement probably represents a necessary first step, it is not sufficient in my view. Given the capacity of air quality models to deliver highly resolved data in terms of space and time, a user will mostly be interested in the results for a specific city and specific day. When looking at results detailed in terms of city and forecast day (Figure 9 and box-plots 8 and 10), the agreement is quite low for some cities/days.

In Fig. 8 we have decided to show the mean agreement since to present 34 figures (one for each city) will be unreadable. However, in the preparation of our study, the agreement was also calculated for each component, as shown below:





Agreement in the determination of the dominant country contributor for SO_4, NO_3, NH_4, EC and POM in percent, and for each 4-day forecast (01-04 Dec 2016, 02-05 Dec 2016, 03-06 Dec 2016, 04-07 Dec 2016, 05-08 Dec 2016, 06-09 Dec 2016, 07-10 Dec 2016, 08-11 Dec 2016, 09-12 Dec 2016) over all the cities using the 9 grids definition.

Moreover, a low agreement between the two methods for some cities/days is not surprising. Indeed, conceptual differences do exist between the scenario and tagging/labeling approaches that have been shown to generate important differences in terms of results for non-linear compounds (see for example Burr and Zang 2011, Grewe et al. 2010, Thunis et al. 2019). This point was also made by Kranenburg et al. (2013) themselves. In their work, Clappier et al. (2017) and Grewe et al. (2010) explained conceptually why these two approaches do not lead to comparable results for non-linear compounds and concluded that these two methods should serve different purposes. Given the above points, I find the Author's conclusions surprising and also misleading in terms of their implications on air quality management practices as they

suggest that both methods are equally suited for calculating source contributions (see e.g. lines 113-114) when this is known not to be the case.
We have changed the sentence in the abstract:
"Better results are found in the determination the dominant country contributor for the primary component (70% for POM and 80% for EC) than for the secondary inorganic aerosols (50%) **which is predictable due the conceptual differences in the source attribution used by both models."**

We also added this sentence in the introduction:
"Thus, the scenario approach is more appropriate in the calculation of the source contribution for the primary PM components than for non-linear species such as the secondary components (e.g. Burr and Zhang, 2011, Thunis et al., 2019)."
and the following information (in bold):
"Even if both methodologies mainly aim to answer two different questions, i.e. the emission control scenarios with the scenario approach and the attribution of concentrations from a source by the labelling technique, it is still useful to estimate the reliability of both methodologies in the estimation of the source contribution to PM_{10} concentrations. For example, it is important to ensure that the non-linearity, related to the perturbation used in the scenario approach, has a limited impact on the calculated contributions and to show that both methodologies **may** present similar results **in the country source attribution**."

A few other points are raised below.
- As shown by Clappier et al. (2017) or Kranenburg et al. (2013), the results of the scenario and labeling techniques would lead to identical results for the linear fraction of PM (primary), if obtained with the same underlying air quality model. The level of agreement obtained for primary compounds like EC therefore provides quantitative information on the difference caused by the underlying model (LOTOS vs. EMEP). On the other hand the difference in agreement between primary and secondary (NO3, SO4 or NH4 Figures 8 and 10) is a direct consequence of the apportionment of the secondary fraction which conceptually differs in the two methods. The lower agreement for SIA than for primary is not only due to differences between EMEP and LOTOS as suggested at lines 477-478, but also, according to me, because of the conceptual differences between the labeling and scenario approaches.
We agreed. The following sentences have been added in Section 6:
"It is also related to the differences in both methodologies (e.g. Clappier et al, 2017b). Indeed, an emission reduction and a labelling technique will not necessarily provide the same results for the secondary PM. An emission reduction depends on the atmospheric composition already present. For example, an amount of NO_x emitted over a source can result in a certain NH_4NO_3 concentration in the receptor. If this NO_x is emitted in excess (NH_3 limited regime), a NO_x emission reduction will have a small effect at the receptor point. On the other hand, in the NO_x limited regime, the same NO_x reduction will have a large impact. The labelling method will give the same result in both cases while the scenario approach will give different results."

We have also added this sentence in the conclusion:
"The differences seen are mainly related to the SIA and is a direct consequence of the difference between both methodologies used."

-The impact of the reduction percentage used in the scenarios is shown as an average over cities and forecast days (Figure 6). It is unclear how the average indicator has been obtained (have negative and positive differences been summed-up in absolute term?) but even if in absolute

terms, the average process does not show the real level of non-linearity obtained for specific cities and days. Thunis et al. (2016) have shown, based on LOTOS_EUROS simulations, that non-linearities could reach more that 5 to 10% on daily values and that the "interaction non-linearity (ignored in the current work) was the dominating factor (up to 20% in some cities). The level of non-linearity obtained here (around 1%) is very low and therefore surprising. It would interesting to see detailed values of this non-linear indicator for each city/day.

In the section 5.1, we have added an explanation about the calculation in the non-linearity:
"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):

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

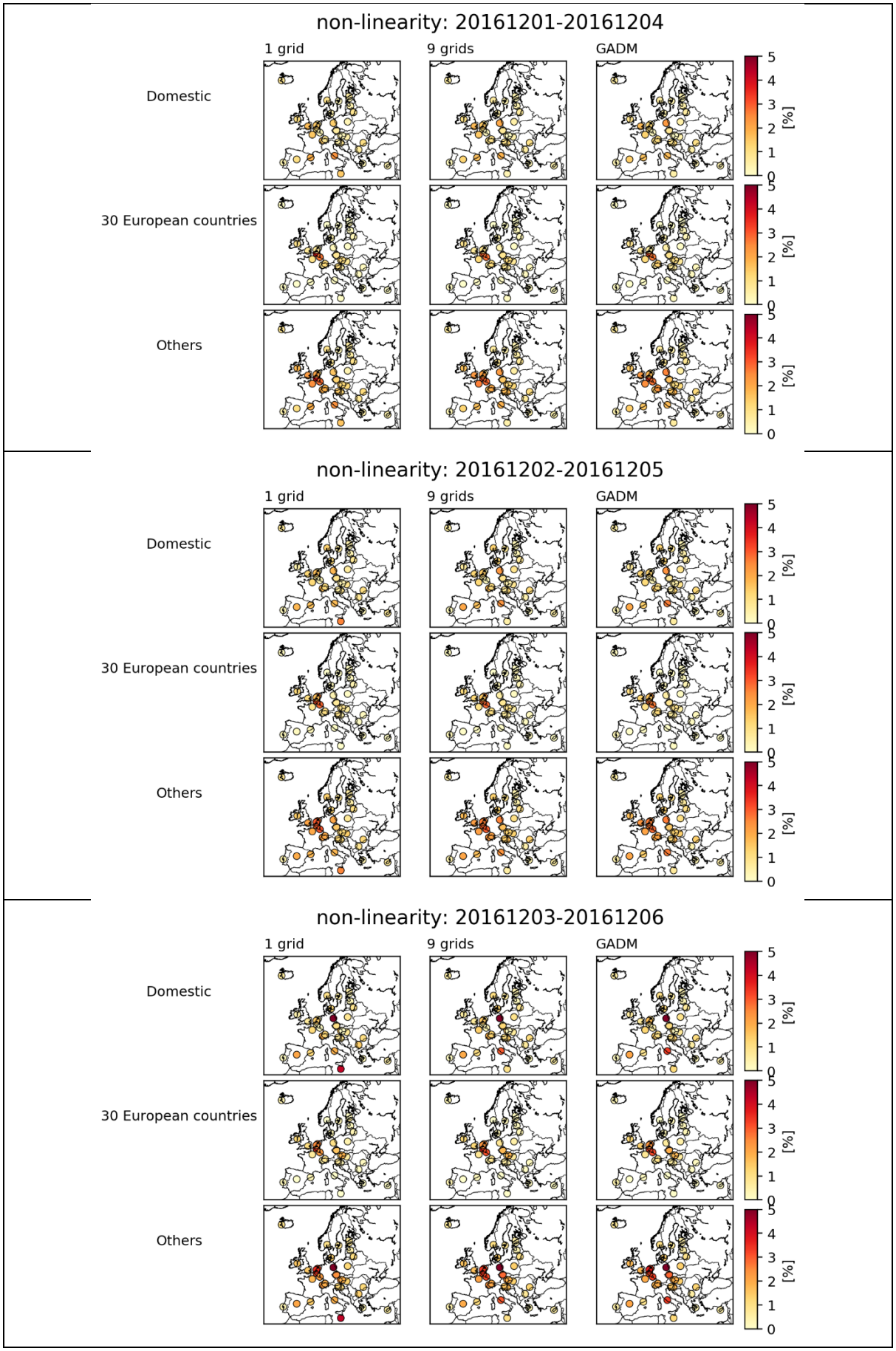
n corresponds to the number of perturbations used (n=3), Ccontrib is the hourly PM_{10} concentration for a specific contribution ("Domestic" or "30 European countries" or "Others") and Ctot is the hourly PM_{10} concentration."

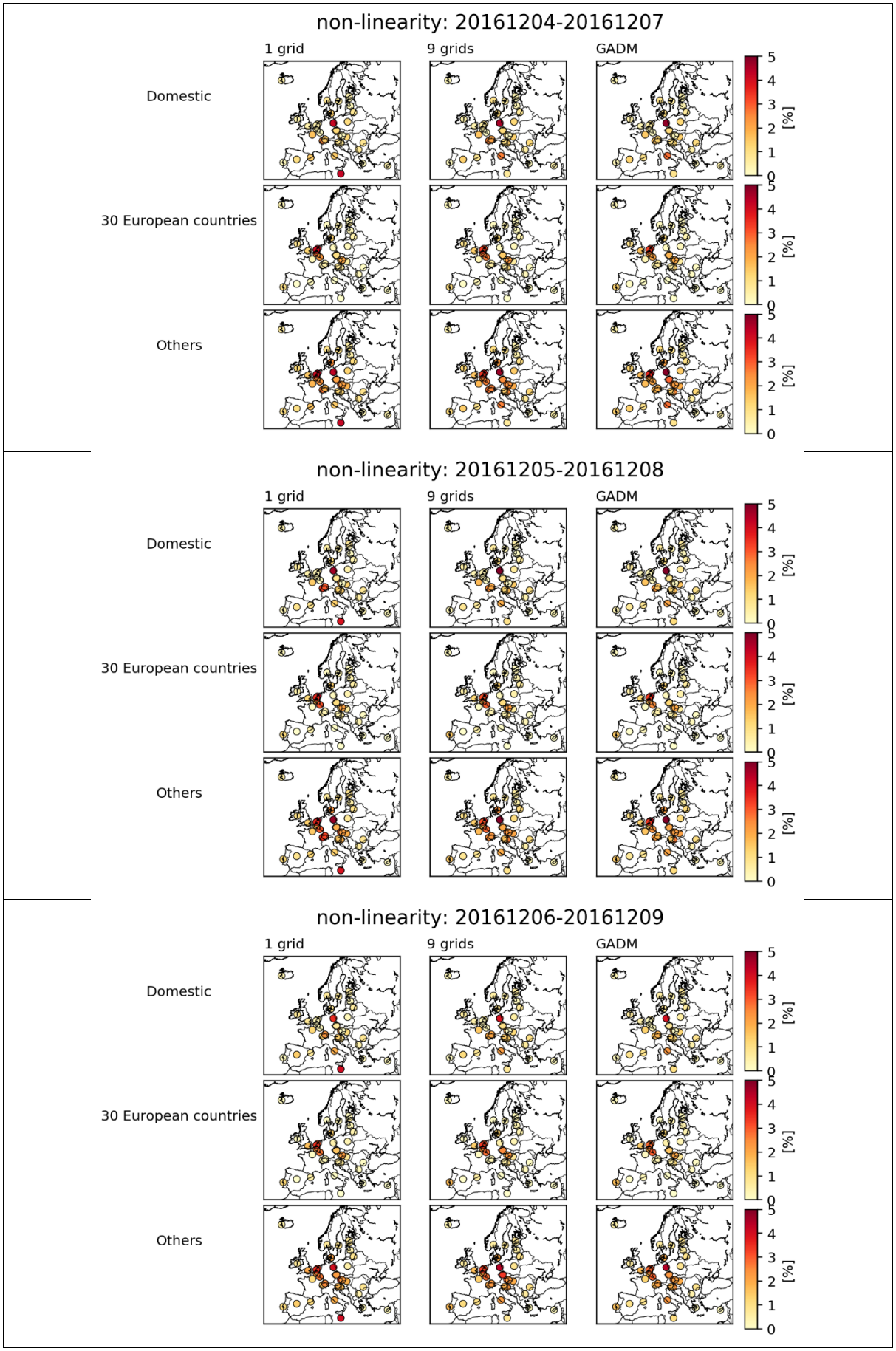
It is important to remind, that our calculated contribution (Eq. 5), corresponds to the change in concentrations related to the change in emissions.

It is also right that we did not study in this work the non-linearity between the different PM_{10} precursors. As explained in the paper, to perform a test related to the non-linearity in the reduction of each individual precursor will be too time consuming.
Without to count the 9 reference runs corresponding to each date; it will represent in total 1395 runs: 9 dates * 31 countries (sources) * 5 anthropogenic emissions (CO, SOx, NOx, NH3, NMVOC and PPM). To perform a complete analysis, it should also be done for the three perturbations, namely 5, 15 and 50%, and it will result to 4185 runs in total.
To perform our study, we already did 837 4-day runs: 9 dates * 31 countries * 3 perturbations + 9 reference runs.
These numbers of runs do not take into account the postprocessing of the simulations over the 34 studied cities.

We have added this sentence in section 4.1.2:
"In total, 847 4-day runs have been performed in this work (9 reference runs, and 9 dates × 31 countries × 3 perturbations runs)."
And the following information (in bold) in Section 5.1:
"This also shows that the responses to perturbation runs are robust, **even if only the non-linearity in the chemistry related the perturbation used, and not the one related to the reduction of each emission precursor, has been estimated in this study as mentioned in Section 4.1."**

The non-linearity presented below, shows the variability of this non-linearity for each date over the different cities. The non-linearity remains low.


non-linearity: 20161201-20161204
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161202-20161205
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161203-20161206
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]


non-linearity: 20161204-20161207
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161205-20161208
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161206-20161209
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]

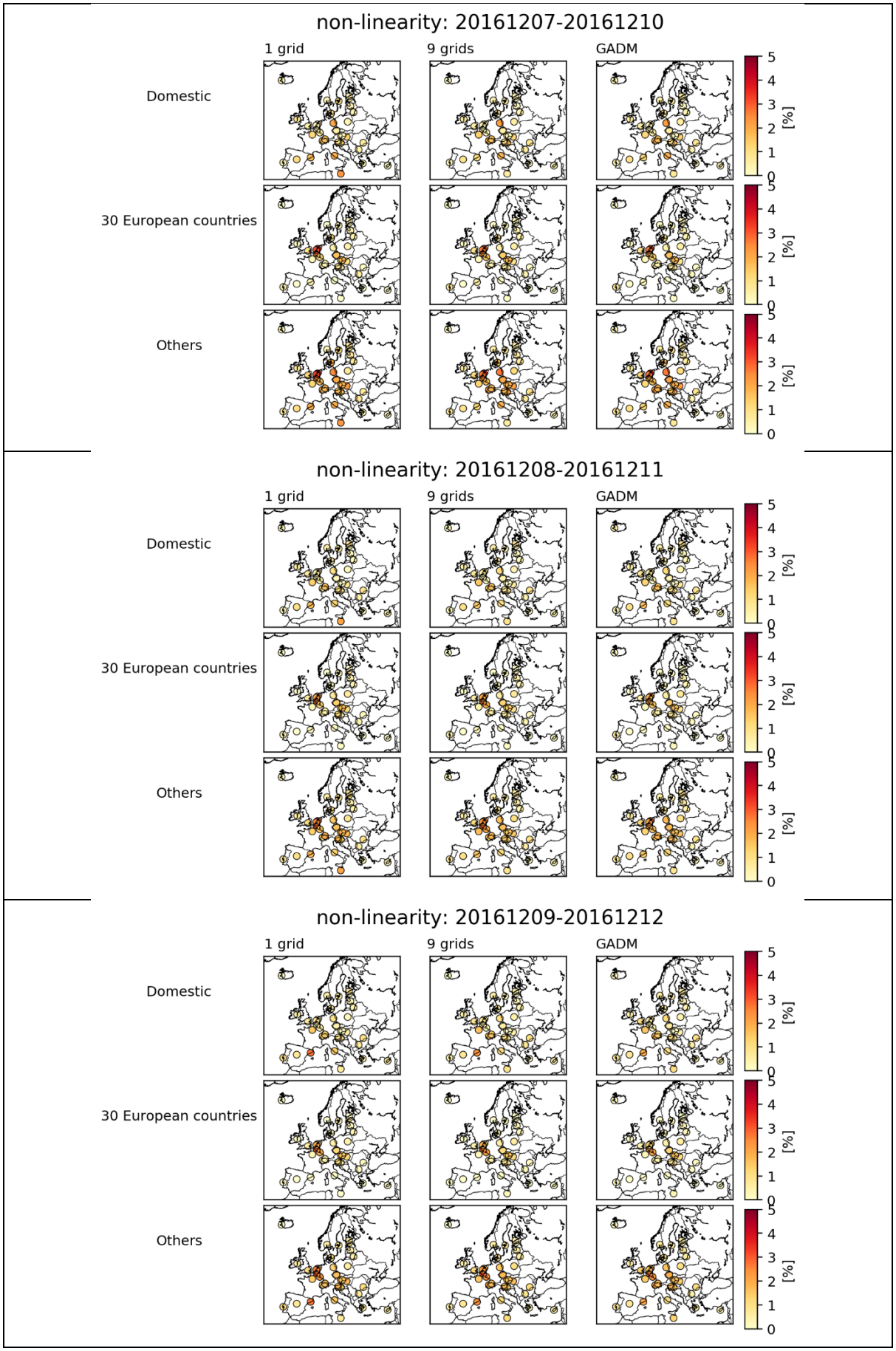

non-linearity: 20161207-20161210
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161208-20161211
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
non-linearity: 20161209-20161212
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]

Figure: Mean hourly non-linearity in percent calculated for the "Domestic", "30 European countries" and "Others" contributions, over the 34 European cities and for each 4-day forecast individually (from 01-04 Dec to 09-12 Dec 2016). The non-linearity is presented for the cities defined by 1 grid cell (left row), 9 grid cells (middle row) and by the GADM (right row).

Even some hourly non-linearities may present larger values, the amount of these large non-linearities is limited as shows with this distribution for the 9 grid cells definition:

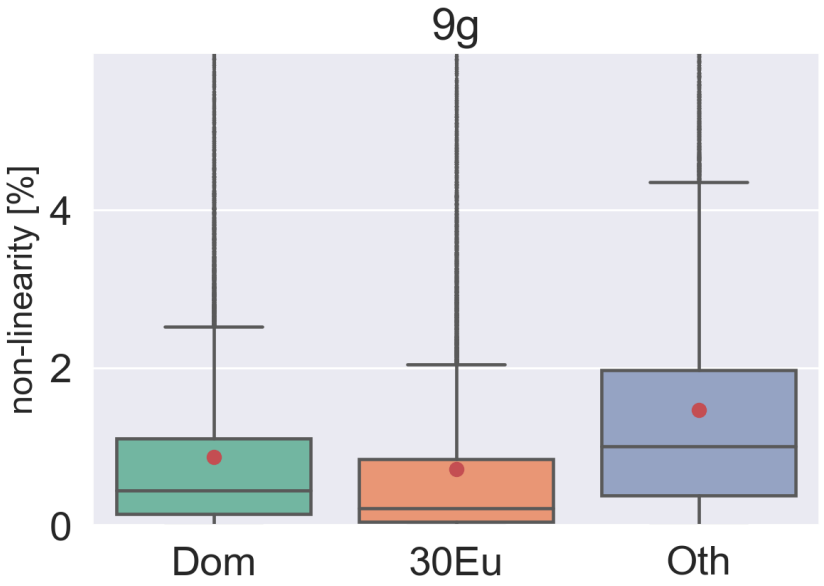


Figure: Distribution of the non-linearity for the "Domestic", "30 European countries" and "Others" contribution. The line that divides the boxes into two parts represents the median of the data. The end of the boxes shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers which are represented by grey diamonds (almost seen as a line). The red dots correspond to the mean of each data set.

We have also added this sentence in the manuscript:
"The mean non-linearity is not homogenously distributed over all cities as shown in Figure S10 and may vary from date to date (not shown). It has remained limited even if some hourly contributions show higher non-linearity. In maximum, 3% of the calculated hourly contributions for all 4-day forecasts over the selected cities have a non-linearity higher than 5% (not shown)."

With this following figure:

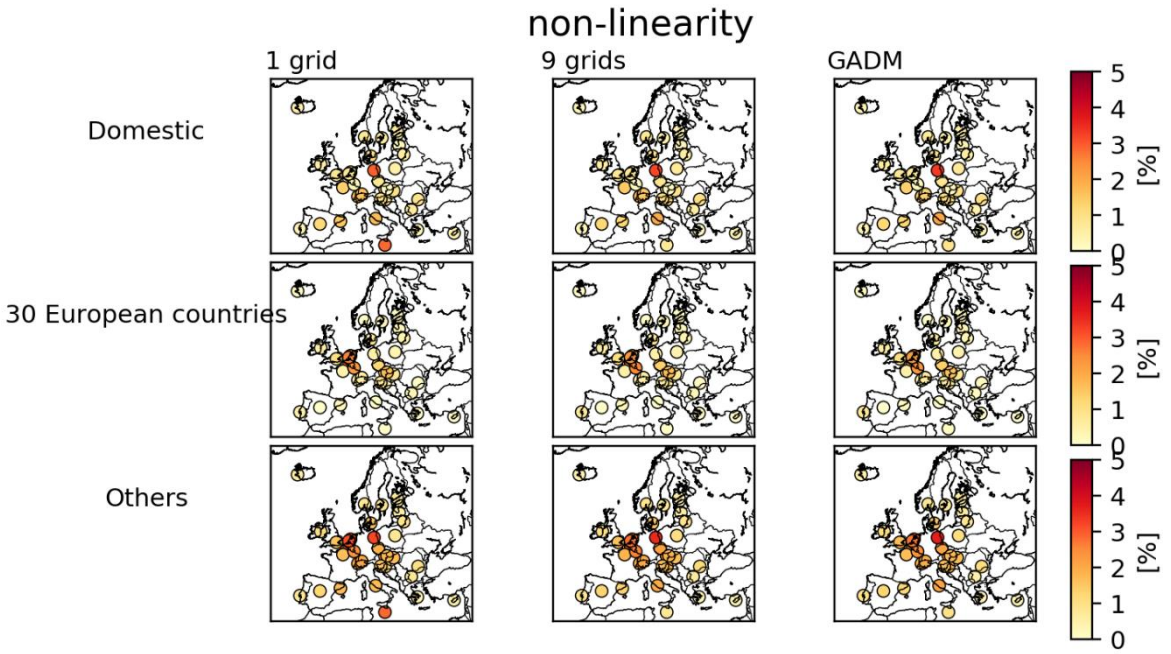


Figure. S10 Mean hourly non-linearity in percent calculated for the "Domestic", "30 European countries" and "Others" contributions, over the 34 European cities and for all 4-day forecasts (i.e. from 01-04 Dec to 09-12 Dec 2016). The non-linearity is presented for the cities defined by 1 grid cell (left row), 9 grid cells (middle row) and by the GADM (right row).

We have also added these sentences in the conclusion:
"Even if this non-linearity is not identical for all cities and for the different dates, the larger non-linearities (>5%) impact only 3% of all the calculated hourly contributions. However, the non-linearity related to the reduction of each emission precursor has not been calculated in the study for computational reason."

- When noting that the results of scenario and labeling differed for non-linear species, Kranenburg et al. (2013) compared a 5% scenario reduction with a simulation where only 5% of the emissions where labeled. Could the Authors explain why they did not label only 15% of the emissions in this comparison?

The answer here is simply that labelling 5% would give the same signal as 15%, only a factor three different in absolute terms. We chose the 5% scenario simulations to keep to the same pollution regime and illustrate its behaviour. With that belongs the 5 % labelling contribution, which would be the same as labelling 15% and divide by three. In this comparison, we multiplied the EMEP results to go from 15 to 100%, we could also divide our labelling results by 100/15 to get the same comparison. Hence, it would not change our results. And by going to 100% you could check the non-linearity issue.

A. Clappier

We would like to thank Alain Clappier for his comments, and we apologize that our analyse could mislead him to unclear conclusions.
We have answered the remarks below, highlighted in blue.

In this article, the authors compare two different source apportionment methods, both able to evaluate how different emission sources contribute to the formation of PM concentrations. The first method is a scenario approach method. It is implemented using the EMEP model to calculate the impact of the reduction of each individual source. The second method is a labelling approach. It is implemented using the LOTOS-EUROS model to calculate the contribution of different sources tracing the mass of the emitted pollutants throughout the different processes computed by the model.
The authors explain that the two methods are comparable only if the concentrations changes related to the scenario approach are not impacted by the non-linearity: Lines 111 to 114 "This highlights the importance to estimate the reliability of both methodologies in the attribution of sources to PM10 concentrations, e.g. to ensure that the concentrations changes related to the scenario approach are not impacted by the non-linearity and to show that both methodologies present similar results."

I have a first serious concern with the way the authors are testing the linearity using the scenario approach: In their article Thunis P. and Clappier A. (2014) show that the non-linearity between emissions and concentrations can affect the impact of the reduction of each individual emission precursors (the concentration reduction is not proportional to the emission reduction) as well as the impact of the reduction of all the emission precursors (the concentration reduction resulting from the reduction of all the precursors simultaneously is not equal to the sum of the concentration reductions resulting from each individual precursor emission).
To test the linearity the authors performed different simulations with EMEP reducing of 5, 15 and 50% all the precursors simultaneously. They claim that reducing the emissions simultaneously or separately may lead to a slight different results.
Lines 383 to 384: "Furthermore, by reducing the emissions simultaneously or separately may lead to a slight different result in the concentrations, but as mentioned previously, this effect is not addressed in this work for computational reason."
How can they claim that the difference between simultaneous reductions and individual reductions is slight. They did not show any results of such test which quantify this difference. Thunis P. et al (2015) show that the non-linearity resulting from the interactions between the different emission precursors is higher than the non-linearity resulting from different reduction percentages. The test performed by the authors can evaluate only a part of the non-linearity which is most likely not the most important part. This test is clearly not sufficient to evaluate the degree of non-linearity. If I refer to what the authors claim lines 111 to 114, they are unable to ensure that the scenario approach and the labelling approaches will give similar results.

Alain Clappier has highlighted an important point.

We agreed that the word "slight" was inappropriate, especially without to show a comparison. It has been deleted.

As also explained in the paper, to perform a test related to the non-linearity in the reduction of each individual precursor will be too time consuming.

Without to count the 9 reference runs corresponding to each date; it will represent in total 1395 runs: 9 dates * 31 countries (sources) * 5 anthropogenic emissions (CO, SOx, NOx, NH3, NMVOC and PPM). To perform a complete analysis, it should also be done for the three perturbations, namely 5, 15 and 50%, and it will result to 4185 runs in total.
To perform our study, we already performed 837 4-day runs: 9 dates * 31 countries * 3 perturbations + 9 reference runs.
These numbers of runs do not consider the postprocessing of the simulations over the 34 studied cities.

We have added this sentence in section 4.1.2:
"In total, 847 4-day runs have been performed in this work (9 reference runs, and 9 dates × 31 countries × 3 perturbations runs)."
And the following information (in bold) in Section 5.1:
"This also shows that the responses to perturbation runs are robust, **even if only the non-linearity in the chemistry related the perturbation used, and not the one related to the reduction of each emission precursor, has been estimated in this study as mentioned in Section 4.1**."

The sentence from line 111 to 114 which was:
"This highlights the importance to estimate the reliability of both methodologies in the attribution of sources to PM_{10} concentrations, e.g. to ensure that the concentrations changes related to the scenario approach are not impacted by the non-linearity and to show that both methodologies present similar results."
aimed to pinpoint the importance to compare both methodologies in our study.
However, to clarify our point, we have modified it. Now it reads:
"Even if both methodologies mainly aim to answer two different questions, i.e. the emission control scenarios with the scenario approach and the attribution of concentrations from a source by the labelling technique, it is still useful to estimate the reliability of both methodologies in the estimation of the source contribution to PM_{10} concentrations. For example, it is important to ensure that the non-linearity, related to the perturbation used in the scenario approach, has a limited impact on the calculated contributions and to show that both methodologies **may** present similar results **in the country source attribution**."

We have also added these sentences in the conclusion:
"Even if this non-linearity is not identical for all cities and for the different dates, the larger non-linearities (>5%) impact only 3% of all the calculated hourly contributions. However, the non-linearity related to the reduction of each emission precursor has not been calculated in the study for computational reason."

I have a second serious concern with the way the authors have interpreted the conclusions of the article of Clappier et al. (2017):
In their article, Clappier et al. (2017) illustrate with simple examples that the scenario approach and the labelling approaches gives similar results only if the concentrations changes related to the scenario approach are not impacted by the non-linearity for any kind of percentage reductions from 0 to 100%. This happens for non-reactive species.
Clappier et al. (2017) illustrate also that, even if the scenario approach often shows linearity between emissions and concentrations for a limited reduction fraction (below 50% for example), the results provided by the scenario approach and the labelling approaches are

different. That means it is expected that the two methods tested in this article will give different results, even before to start complex simulations.
If I refer again to what the authors claim lines 111 to 114, they should not compare the results of the scenario approach and the labelling approaches because we know they are different. Moreover, comparing different methods using different models ensure with a great certainty that the results will be different. Then, how can we interpret the authors' conclusions? Lines 518 to 519 "It was shown that the results from both source apportionment methodologies agree in average by 68% in the determination the dominant country contributor to the hourly PM10 concentrations and 75% for the top 5 of these country contributors". Are the disagreements shown by the results due to the discrepancy between the methods or to the difference between the models?
As mentioned previously, the sentence in lines 111-114 has been changed. We have also added (in bold), the following information in the abstract:
"Better results are found in the determination the dominant country contributor for the primary component (70% for POM and 80% for EC) than for the secondary inorganic aerosols (50%) **which is predictable due the conceptual differences in the source attribution used by both models**."
And in the introduction:
"Thus, the scenario approach is more appropriate in the calculation of the source contribution for the primary PM components than for non-linear species such as the secondary components (e.g. Burr and Zhang, 2011, Thunis et al., 2019)."

Since the difference in the contributions is mainly seen in the SIA, it shows that it is a clear result related to the difference between both methodologies.
For this reason, we have added this sentence in the conclusion:
"The differences seen are mainly related to the SIA and is a direct consequence of the difference between both methodologies used."

It is also important to note that we have added this following information in the Section 6:
"It is also related to the differences in both methodologies (e.g. Clappier et al, 2017b). Indeed, an emission reduction and a labelling technique will not necessarily provide the same results for the secondary PM. An emission reduction depends on the atmospheric composition already present. For example, an amount of NO_x emitted over a source can result in a certain NH_4NO_3 concentration in the receptor. If this NO_x is emitted in excess (NH_3 limited regime), a NO_x emission reduction will have a small effect at the receptor point. On the other hand, in the NO_x limited regime, the same NO_x reduction will have a large impact. The labelling method will give the same result in both cases while the scenario approach will give different results."

I have a third serious concern with the way the authors interpret the capacity of the labelling and the scenario approaches to represent the reality:
Lines 386 to 388 the authors mention that: "In their study, Kranenburg et al. (2013) have shown that this technique [the labelling approach] provides more accurate information about the source contributions than using a brute force approach with scenario runs as the chemical regime remains unchanged."
The relation between emissions and concentrations is non linear is the real world as well as in the numerical models. If the results of the scenario approach are changing according to the percentage of reduction and/or the number of reduced emission sources, it is simply because

this method is able to reflect reality. Since the reality is non-linear, the scenario approach method behaves non-linearity. If it is used correctly the method can even quantify the degree of non-linearity.
The labelling approach gives always one unique result, regardless of the degree of non-linearity of the system under study. Because they are not impacted by the nonlinearity, the results are certainly much easier to show. But they give no information about non-linearity showing that the method does not reflect how the system change when the emission change. I fully agree with the authors when they write about the labelling approach: lines 108 to 109, "However, it is not designed to study the impact of emission abatement policies to pollutants concentrations...". It appears clearly that it is nonsense to claim that the labelling approach provides more accurate information about the source contributions than using a brute force approach with scenario runs as the chemical regime remains unchanged.
We have deleted this sentence in our manuscript, which was given in Kranenburg et al. (2013).

As mentioned in our answer to the previous comment, the difference between both methodologies, based on the non-linearity in the "reality" and in the models is explained in Section 6.

To conclude: This article shows significant gaps in the design of the different test as well as in the analysis of the results. I do not understand the usefulness to compare results if it is known in advance that they will be different and if it is know it will be not possible to find the origin of the differences.
Even their respective limitations, we still think that both approaches provide complementary information on source contributions and PM_{10} composition.
The use of both techniques is very useful for quality assurance purposes and our study demonstrates the ability of two modelling approaches to identify source contributions of particulate matter from different countries to several cities in Europe during a pollution episode. The results show a large degree of similarity which is a key result and should be appreciated as there is no way to arrive at the true source apportionment (see FAIRMODE documentation by Mircea et al, in prep)

Thunis, P. and A. Clappier, 2014. Indicators to support the dynamic evaluation of air quality models, Atmos. Environ., 98, 402-409 Thunis P., A. Clappier, E.Pisoni, B.Degraeuwe, 2015: Quantification of non-linearities as a function of time averaging in regional air quality modeling applications, Atmospheric Environment, 103, 263-275. Clappier A., C. Belis, D. Pernigotti and P. Thunis, 2017: Source apportionment and sensitivity analysis: two methodologies with two different purposes. Geosci. Model Dev., 10, 4245-4256.

D. Ham

We would like to thank the Editor David Ham to highlight these issues and we would like to apologize for our previous lack.
We have corrected the missing points. Please find our answers written in blue.

I am writing as executive editor to bring to your attention several ways in which this manuscript currently does not comply with GMD policy. These issues will need to be addressed before any revised manuscript could be accepted for publication.

Title
As per https://www.geoscientific-model-development.net/about/manuscript_types.html#item4, the names and version numbers of the models being evaluated need to be included in the title of the manuscript. Please change the title accordingly.
It has been corrected. The title is now:
"Prediction of source contributions to urban background PM_{10} concentrations in European cities: a case study for an episode in December 2016 by using EMEP/MSC-W rv4.15 and LOTOS-EUROS v2.0 - Part.1 The country contributions"

EMEP code on GitHub
GitHub is an excellent development platform, but it is not an archival location suitable for the code used in a paper. Indeed, even GitHub themselves tell you to use Zenodo (https://guides.github.com/activities/citable-code/). Please therefore archive (probably using Zenodo) the precise version of EMEP used in this manuscript, and cite it from the code and data availability section. The Zenodo archive will give you the entry to paste into BibTeX or another reference manager.
LOTOS-EUROS code archive
LOTUS-EUROS causes more issues because the code is not open source. It claims to be open source, however reading its licence indicates that this is not actually true because redistribution is prohibited (see https://opensource.org/osd item 1). This means that properly archiving the version of LOTUS-EUROS used will not be possible. Instead, GMD policy requires you to point out that LOTUS-EUROS is only available under a restricted licence. It remains critical that the manuscript identifies the exact version used in order to enable the results to be reproduced. The manuscript should also not make the incorrect claim that the software is open source. It is unfortunate that this claim is made on the website, however that is not a reason to reproduce the error in the manuscript.
Data availability is missing
This manuscript describes a model evaluation campaign. The models were driven using data, and evaluated using data. The code and data availability section needs to point the user at the persistent public archives for the precisely identified code that was used. For a model evaluation paper, this is likely to make this section quite expansive, in contrast with its current brief extent.
Configuration files, run scripts and evaluation scripts
Reproducibility also demands that the exact configuration files and scripts used to run the models are presented, along the scripts used to process and evaluate the model output. Please also archive and cite this data.
For a fuller description of GMD's code and data evaluation policy, please see: https://opensource.org/osd

It has been corrected. Now it reads:

“The EMEP model is an open source model available on https://doi.org/10.5281/zenodo.3355041. The base-code of LOTOS-EUROS is available under the license on https://lotos-euros.tno.nl/, but the code used for this study, including the source apportionment is only available in cooperation with TNO. The data processing and analysis scripts are available upon request.”

Revised manuscript:

Prediction of source contributions to urban background PM_{10} concentrations in European cities: a case study for an episode in December 2016 by using EMEP/MSC-W rv4.15 and LOTOS-EUROS v2.0 - Part.1 The country contributions

Matthieu Pommier[1], Hilde Fagerli[1], Michael Schulz[1], Alvaro Valdebenito[1], Richard Kranenburg[2], Martijn Schaap[2,3]

1 Norwegian Meteorological Institute, Oslo, Norway
2 TNO, PO Box 80015, 3508TA Utrecht, The Netherlands
3 FUB – Free University Berlin, Institut für Meteorologie, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany
Correspondence to: matthieu.pommier@met.no

Abstract.
A large fraction of the urban population in Europe is exposed to particulate matter levels above the WHO guideline. To make more effective mitigation strategies, it is important to understand the influence on particulate matter (PM) from pollutants emitted in different European nations. In this study, we evaluate a country source contribution forecasting system aimed to assess the domestic and transboundary contributions to PM in major European cities for an episode in December 2016. The system is composed of two models (EMEP/MSC-W rv4.15 and LOTOS-EUROS v2.0) which allows to consider differences in the source attribution.
We also compared the PM_{10} concentrations and both models present satisfactory agreement in the 4-day forecasts of the surface concentrations, since the hourly concentrations can be highly correlated with in-situ observations. The correlation coefficients reach values up to 0.58 for LOTOS-EUROS and 0.50 for EMEP for the urban stations; and 0.58 for LOTOS-EUROS and 0.72 for EMEP for the rural stations. However, the models under-predict the highest hourly concentrations measured by the urban stations (mean underestimation by 36%), predictable with the relatively coarse model resolution used (0.25° longitude × 0.125° latitude).
For the source attribution calculations, LOTOS-EUROS uses a labelling technique, while the EMEP/MSC-W model uses a scenario having reduced anthropogenic emissions and then it is compared to a reference run where no changes are applied. Different percentages (5%, 15% and 50%) in the reduced emissions for the EMEP/MSC-W model were used to test the robustness of the methodology. The impact of the different ways to define the urban area for the studied cities was also investigated (i.e. 1 model grid cell, 9 grid cells and the grid cells covering the definition given by the Global Administrative Area - GADM). We found that the combination of a 15% emission reduction and a larger domain (9 grid cells or GADM) help to reduce the impact of non-linearity on the chemistry. This non-linearity, related to the perturbation used, is seen in the mismatch between the total concentration and the sum of the concentrations from different calculated sources. Even limited, this non-linearity is observed in the NO_3^-, NH_4^+ and H_2O concentrations, which is related to gas-aerosol partitioning of the species. The use of a 15% emission reduction and of a larger city domain also gives a better agreement in the determination of the main country contributors between both country source calculations.

The studied episode over the 34 European cities investigated, occurring from December 01st to 09th 2016, was dominated by the domestic emissions. The two models agree 68% of the time (on hourly resolution) on the country, having been the dominant contributor to PM_{10} concentrations. 75% of the hourly predicted PM_{10} concentrations by both models, have the same top 5 main country contributors. Better results are found in the determination the dominant country contributor for the primary component (70% for POM and 80% for EC) than for the secondary inorganic aerosols (50%) which is predictable due the conceptual differences in the source attribution used by both models.

1. Introduction.

The adverse health impacts from air pollution and especially from particulate matter (PM) is a well-documented problem (e.g. Keuken et al., 2011; REVIHAAP, 2013; Mukherjee and Agrawal, 2017; Segersson et al., 2017). Furthermore, it affects crop yields (e.g. Crippa et al., 2016), visibility (e.g. Founda et al., 2016) and even the economy (e.g. Meyer and Pagel, 2017). The mass of particulate matter with an aerodynamic diameter lower than 10 μm (PM_{10}) is an air quality metric linked to premature mortality at high exposure (e.g. Dockery and Pope, 1994). The World Health Organization (WHO) has established a short-term exposure PM_{10} guideline value of 50 $\mu g/m^3$ daily mean that should not be exceeded in order to ensure healthy conditions (the long-term exposure guideline is 20 $\mu g/m^3$ for annual-mean PM_{10}) (WHO, 2005). Although policies have been proposed and implemented at the international (e.g. Amann et al., 2011) and national (e.g.D'Elia et al., 2009) levels, European cities still suffer from poor air quality (EEA report 2017), especially due to high PM_{10} concentrations. In short, to further decrease the adverse health impacts of PM in Europe its concentrations need to be reduced further.

PM_{10} concentrations in the atmosphere are highly variable in space and time. Due to the relative short atmospheric life time (from some hours to days), the variability is impacted by local sources, meteorological conditions affecting dispersion and long-range transport as well as chemical regimes controlling the efficiency of secondary formation. PM_{10} consists of both primary and secondary components. Primary PM_{10} components include organic matter (OM), elemental carbon (EC), dust, sea salt (SS) and other compounds. Secondary PM_{10} comprises compounds formed by chemical reactions in the atmosphere from gas-phase precursors. This includes various compounds as nitrate (NO_3^-) from nitrogen oxide (NO_x) emissions, ammonium (NH_4^+) from ammonia (NH_3) emissions, sulphate (SO_4^{2-}) from sulphur dioxide (SO_2) emissions, and a large range of secondary organic aerosol (SOA) compounds from both anthropogenic and biogenic volatile organic compounds (VOCs). The sources for PM and its precursors are multiple but the main anthropogenic sources are the transport, industries, energy production and agriculture. The main natural sources are composed of forest fires, mineral dust and sea salt. The main sink is the wet deposition. The dry deposition can also be important and depends on the type of land surface such as grass, tree leaves and others; and on meteorological conditions. With these components deriving from various sources, we understand the importance to reflect properly the source contributions in the modelling for policy support.

Many studies have already focused on source receptor relationships to calculate the transport of atmospheric pollutants, with country-to-country relationships (e.g. EMEP Status Report 1/2018) but also over cities (e.g. Thunis et al., 2016; 2018). However, these studies focus on annual means, whereas information is also required on exposure from episodes which cause short-term limit value exceedances throughout Europe. Source apportionment provides valuable information on the attribution of different sources to PM_{10} concentrations. A country source calculation allows to tackle the emissions from the countries responsible for the air pollution episode. Two distinct methodologies have been compared in this study. Indeed, the

country source contribution presented hereafter is performed by two regional models, the EMEP/MSC-W model (Simpson et al., 2012) and LOTOS-EUROS (Manders et al., 2017).
The EMEP calculations use reduced anthropogenic emission scenario and compare to a reference run where no changes are applied. It is also known as the scenario approach. 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. Hence, the scenario approach is useful for analyzing the concentration changes due to emission reductions. On the other hand, one simulation per source is needed to calculate the impact of each source, as done on annual means for each country in each EMEP report (e.g. EMEP Status Report 1/2018). The scenario approach may also lead to a non-linearity in the calculated concentrations, i.e. a slight difference between the concentrations over a receptor and the sum of the estimated concentrations from different sources over this same receptor, as shown by Clappier et al. (2017a). Thus, the scenario approach is more appropriate in the calculation of the source contribution for the primary PM components than for non-linear species such as the secondary components (e.g. Burr and Zhang, 2011, Thunis et al., 2019).
LOTOS-EUROS traces the origin of air pollutants throughout a simulation using a labelling approach. The advantage of the labelling technique is the reduction of the computational time, in comparison to the scenario approach. It also quantifies the contribution of an emission source to the concentration of one pollutant at one given location. However, it is not designed to study the impact of emission abatement policies to pollutants concentrations (Grewe et al., 2010; Clappier et al., 2017b) and only traceable atoms can be used in labelling approach, i.e. only conserved atoms (C, N, S), directly related to emission sources, in their different oxidation states. Thus, for example, the origin of ozone (O_3) cannot be studied, which can be done with the scenario approach. Even if both methodologies mainly aim to answer two different questions, i.e. the emission control scenarios with the scenario approach and the attribution of concentrations from a source by the labelling technique, it is still useful to estimate the reliability of both methodologies in the estimation of the source contribution to PM_{10} concentrations. For example, it is important to ensure that the non-linearity, related to the perturbation used in the scenario approach, has a limited impact on the calculated contributions and to show that both methodologies may present similar results in the country source attribution.
Both models are part of the operational country source contribution (SC) prediction system for the European cities within the Copernicus Atmosphere Monitoring Service (CAMS). This system aims at attributing country contribution to surface PM_{10} in European cities for 4-day forecasts. The objective of this study is to evaluate the robustness of a new system that provides forecasts of source region resolved PM for European cities. The evaluation of the system is focused on an event occurring between the December 01st and 09th 2016, which corresponds to the first event listed from the beginning of the development of our system. To do so, the predicted PM_{10} concentrations are compared with observations. The simulations from both models, for the concentrations and the SC calculations are also inter-compared.
Section 2 describes the country SC system composed by the two models and the experiment. Section 3 describes the studied episode and it presents the evaluation of both predictions in terms of PM_{10} concentrations. The methodology used for the SC calculations by both models is explained in Section 4. Then Section 5 gives an overview of the composition and the origin of PM_{10} over the cities predicted by both models, and the issue regarding the non-linearity in the chemistry related to the EMEP SC calculation. Section 6 is a comparison between the two country SC calculations. Finally, the conclusions are provided in Section 7.

2. Description of the country source apportionment system

2.1. Overview of the system

Within CAMS, a country SC product has been developed. This is a new forecasting and near-real time source allocation system for surface PM_{10} concentrations and its different components over all European capitals. The predictions are available online on https://policy.atmosphere.copernicus.eu/SourceContribution.php. The concentrations are calculated over the 28 EU capitals plus Bern, Oslo and Reykjavik. Forecasts for Barcelona, Rotterdam and Zurich are also provided. In addition to providing information about the air quality over the selected cities by focusing on PM_{10}, this product aims at quantifying the contributions of emissions from different countries in each city (Fig. 1).
The system is composed of predictions from two regional models (the EMEP/MSC-W model and LOTOS-EUROS), using two distinct source contribution calculation methodologies. The EMEP/MSC-W chemistry transport model (Simpson et al., 2012) has been used for decades to calculate source receptor relationships between European countries (including Russia) (e.g. EMEP Status Report 1/2018) and the LOTOS-EUROS chemistry transport model (Manders et al., 2017) has also been used in several source apportionment studies over Europe, especially for PM (Hendriks et. al., 2013; 2016; Schaap et al., 2013). Both models are involved in the operational air quality analysis and forecasting for Europe in the CAMS regional ensemble system (Marécal et al., 2015) and for China (Brasseur et al., 2019). For the simplicity of the reading, the EMEP/MSC-W model is hereafter referred to as EMEP model.
Both models are Eulerian models but there are differences between these two models such as the calculation of the planetary boundary layer (PBL) and of the advection, the vertical resolution. There are also differences in the presence of the secondary organic aerosol (included in the EMEP model and not in LOTOS-EUROS), PM_{10} diagnosing particle water explicitly in the EMEP model and not in LOTOS-EUROS, the calculation of the biogenic emissions, the description of the gas-phase chemistry and the treatment of dust (from agriculture and traffic are included in LOTOS-EUROS and not in the EMEP model).
The main details about the models and the experiment are provided in the Table 1 and a more complete description is provided in the following Sections.

2.2. Description of the EMEP model

The EMEP model is a 3-D Eulerian chemistry-transport model described in detail in Simpson et al. (2012). Initially, the model has been aimed at European simulations, but the model has also been used over other regions and at global scale for many years (e.g. Jonson et al., 2010). The EMEP model version rv4.15 has been used here in the forecast mode. The version rv4.15 has been described in Simpson et al. (2017) and references cited therein. The main updates since Simpson et al. (2012) and used in this work, concern a new calculation of aerosol surface area (now based upon the semi-empirical scheme of Gerber, 1985), revised parameterizations of N_2O_5 hydrolysis on aerosols, additional gas-aerosol loss processes for O_3, HNO_3 and HO_2, a new scheme for ship NO_x emissions, a new calculated natural marine emissions of dimethyl sulphid (DMS), and the use of a new land-cover (used to calculate biogenic VOC emissions and the dry deposition) (Simpson et al., 2017). This version is the official EMEP Open Source version that was released in September 2017 (Tab. 1)
Vertically, the model uses 20 levels defined as sigma coordinates (Simpson et al., 2012). The PBL is located within approximately the 10 lowest model levels (~5 levels below 500 m), and the top of the model domain is at 100 hPa. The PBL height is calculated, based on the turbulent diffusivity coefficient as described in the EMEP Status Report (2003). The numerical solution of the advection terms is based upon the scheme of Bott (1989).
The chemical scheme couples the sulphur and nitrogen chemistry to the photochemistry using about 140 reactions between 70 species (Andersson-Sköld and Simpson, 1999; Simpson et al.

2012). The chemical mechanism is based on the "EMEP scheme" described in Simpson et al. (2012) and references therein.
The biogenic emissions of isoprene and monoterpene are calculated in the model by emission factors as a function of temperature and solar radiation (Simpson et al., 2012). 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).
PM emissions are split into EC, OM (here assumed inert) and the rest of primary PM defined as the remainder, for both fine and coarse PM. The OM emissions are further divided into fossil-fuel and wood-burning compounds for each source sector. As in Bergström et al. (2012), the OM/OC ratios of emissions by mass are assumed to be 1.3 for fossil-fuel sources and 1.7 for wood-burning sources. The model also calculates windblown dust emissions from soil erosion. Secondary aerosol consists of inorganic sulphate, nitrate and ammonium, and SOA; the latter is generated from both anthropogenic and biogenic emissions, using the 'VBS' scheme detailed in Bergström et al (2012) and Simpson et al. (2012).
The main loss process for particles is wet-deposition, and the model calculates in-cloud and sub-cloud scavenging of gases and particles as detailed in Simpson et al. (2012). Wet scavenging is treated with simple scavenging ratios, taking into account in-cloud and sub-cloud processes.
In the EMEP model, the 3D precipitation is needed. 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). This estimate is derived from large scale precipitation and convective precipitation. The height of the precipitation is derived from the cloud water. Then, it is defined as the highest altitude above the lowest level, where the cloud water is larger than a threshold taken as 1.0×10^{-7} kg water per kg air. Precipitations are only defined in areas where surface precipitations occur. The intensity of the precipitation is assumed constant over all heights where they are non-zero
Gas and particle species are also removed from the atmosphere by dry deposition. This dry deposition parameterization follows standard resistance-formulations, accounting for diffusion, impaction, interception, and sedimentation.

2.3. Description of LOTOS-EUROS

The LOTOS-EUROS model is an off-line Eulerian chemistry-transport model which simulates air pollution concentrations in the lower troposphere solving the advection-diffusion equation on a regular latitude-longitude-grid with variable resolution over Europe (Manders et al., 2017) (Tab. 1).
The vertical grid is based on terrain following vertical coordinates and extends to 5 km above sea level. The model uses a dynamic mixing layer approach to determine the vertical structure, meaning that the vertical layers vary in space and time. The layer on top of a 25 m surface layer follows the mixing layer height, which is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological input data that is used to force the model. The horizontal advection of pollutants is calculated applying a monotonic advection scheme developed by Walcek and Aleksic (1998).
Gas-phase chemistry is simulated using the TNO CBM-IV scheme, which is a condensed version of the original scheme (Whitten et al, 1980). Hydrolysis of N_2O_5 is explicitly described following Schaap et al. (2004).
LOTOS-EUROS explicitly accounts for cloud chemistry computing sulphate formation as a function of cloud liquid water content and cloud droplet pH as described in Banzhaf et al.

(2012). For aerosol chemistry the thermodynamic equilibrium module ISORROPIA2 is used (Fountoukis and Nenes, 2007).
The biogenic emission routine is based on detailed information on tree species over Europe (Schaap et al., 2009). The emission algorithm is described in Schaap et al. (2009) and is very similar to the simultaneously developed routine by Steinbrecher et al. (2009). Dust emissions from soil erosion, agricultural activities and resuspension of particles from traffic are included following Schaap et al. (2009).
As in the EMEP model, the 3D precipitation is needed and cloud liquid water profiles are used to diagnose cloud base height and where below and incloud scavenging takes place. The wet deposition module accounts for droplet saturation following Banzhaf et al. (2012). Dry Deposition fluxes are calculated using the resistance approach as implemented in the DEPAC (DEPosition of Acidifying Compounds) module (van Zanten et al., 2011). Furthermore, a compensation point approach for NH_3 is included in the dry deposition module (Wichink Kruit et al., 2012).

2.4. Description of the experiment

The study focuses on the period from December 01st to 09th 2016. In our system, the forecasts provided by the EMEP model cover a slightly different regional domain than LOTOS-EUROS (Tab. 1). 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). This harmonization has been revealed important for such comparison and increases the consistency of the model results. The impact of such choices is illustrated for the city definitions, for which subjective choices can be made causing inconsistencies.
An initial spin-up of 10 days was conducted. Both models provide four-day air quality forecasts, and the simulations have been defined as “forecast-cycling experiments”, i.e. the predicted fields have been used to initialize successive four-day forecasts (e.g Morcrette et al., 2009). The pollution transport in both models is based on forecasted meteorological fields at 12 UTC from the previous day, with a 3-hour resolution, calculated by the Integrated Forecasting System (IFS) of ECMWF. These forecasted meteorological fields correspond to the fields which were used in the online SC production for these dates. The ECMWF operational system does not archive 3D precipitation forecasts, which is needed by the EMEP model and LOTOS-EUROS as mentioned in Sections 2.2 and 2.3. Therefore, a 3D precipitation estimate is derived from IFS surface variables (large scale and convective precipitations) in the EMEP model and the 3D field is based on the cloud liquid water profile in LOTOS-EUROS.
The boundary conditions (BCs) at 00UTC of the current day from the atmospheric Composition module (C-IFS) have been used. These BCs are specified for ozone (O_3), carbon monoxide (CO), nitrogen oxides (NO and NO_2), methane (CH_4), nitric acid (HNO_3), peroxy-acetyl nitrate (PAN), SO_2, ISOP, ethane (C_2H_6), some VOCs, sea salt, Saharan dust and SO_4. In LOTOS-EUROS, sea salt BCs have not been used as these are shown to be overestimated in comparison with the model. In the EMEP model, the sea salt parameter has been used. This may cause a difference between both models in the estimation of the contribution from sea salt especially for the coastal cities.
Both models use the TNO-MACC emission data set for 2011 on 0.25° × 0.125° (longitude-latitude) resolution (Kuenen et al., 2014, see http://drdsi.jrc.ec.europa.eu/dataset/tno-macc-iii-european-anthropogenic-emissions) and the forest fire emissions are from GFASv1.2 inventory (Kaiser et al., 2012).
Since the study aims to quantify the contributions of long-range transport in each city to the urban background PM_{10}, the effect of the choice of the receptor, i.e. the city domain, has been tested. The city receptor has been defined by three definitions: 1 grid cell (i.e. 0.25° lon ×

0.125° lat, corresponding to the emissions data set resolution), 9 grid cells and the all the grid cells covering the administrative area provided by the database of Global Administrative Areas (GADM, https://gadm.org/data.html). The latter is the more precise definition in terms of build-up area, however it may represent a large region for a definition of a city as shown in Fig. S1 (e.g. London, Nicosia, Riga, Sofia). It is important to explain that this study does not aim to quantify the contribution to PM_{10} at a street scale as done in Kiesewetter et al. (2015) but over the full area defining the cities. The relatively coarse definition of the cities is comparable to the definition used in previous studies as in Thunis et al. (2016) who used an area of 35 × 35 km^2 or in Skyllakou et al. (2014) who used a radius of 50 km from the city center.
For the contribution, we also have harmonized the definition of the natural contributions. The natural contributions are defined in this study as the sum of the contributions from sea salt, dust and forest fires, except for the BCs. In LOTOS-EUROS, the natural sources (e.g. dust) coming from the boundaries are classified as BCs and not natural.

3. Evaluation of the predicted surface concentrations during the episode

During December 2016, a PM episode with medium intensity (no more than three consecutive days beyond the WHO PM_{10} threshold) developed across North-Western Europe. As a consequence of a high pressure system over central Europe pollutants concentrations were built up over western Europe (see http://policy.atmosphere.copernicus.eu/reports/CAMSReportDec2016-episode.pdf).
From December 1st to 2nd, high concentrations were measured and predicted over Paris (Figures 1 & 2). In Figure 2, we can also see from December 3rd to December 8th, that levels of PM_{10} were elevated in Western Europe. Especially on December 6th and 7th, concentrations at some measurement stations in France, Belgium, the Netherlands, Germany and Poland, exceeded the daily limit value of 50 $\mu g/m^3$ (e.g Fig. S2 – see Section 3.2 for more details about the observations).
During the following days relatively stable conditions with slow southerly winds characterized the episode until fronts moved in western Europe at the 9th. Large concentrations (>60 $\mu g/m^3$) were also predicted between December 6th and 9th over the Po Valley and over UK on December 6th (Figs. 2 and S2).

3.1. Statistical metrics used

To properly estimate the quality of these forecasts, five statistical parameters have been used, such as the Pearson correlation (r), the Mean Bias (MB), the Normalized Mean Bias (NMB), the Root-Mean-Square Error (RMSE) and the Fractional Gross Error (FGE). The ideal score of these parameters is 0 except for the correlation which is 1.
The MB provides the information about the absolute bias of the model, with negative values indicating underestimation and positive values indicating overestimation by the model. The NMB represents the model bias relative to the reference. The RMSE considers error compensation due to opposite sign differences and encapsulates the average error produced by the model. The FGE is a measure of model error, ranging between 0 and 2 and behaves symmetrically with respect to under- and overestimation, without over emphasizing outliers.
We have used M and R as notation to refer, respectively, to model and the reference data (e.g. observations), and N is the number of the reference data set (e.g. number of observations).
Thus, MB is calculated by equation (1) and expressed in $\mu g/m^3$:

$$MB = \frac{\sum_{i=1}^{N}(M_i - R_i)}{N} \quad (1)$$

NMB is calculated by equation (2):

$$\text{NMB} = \frac{\sum_{i=1}^{N}(M_i - R_i)}{\sum_{i=1}^{N} R_i} \times 100\% \quad (2)$$

RMSE is calculated by equation (3) and expressed in µg/m^3:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(M_i - R_i)^2}{N}} \quad (3)$$

and FGE is calculated by equation (4) and dimensionless:

$$\text{FGE} = \frac{2}{N} \sum_{i=1}^{N} \frac{|M_i - R_i|}{|M_i + R_i|} \quad (4)$$

3.2. Comparison with observations

3.2.1. Methodology

In order to evaluate the reliability of the predictions over each city, the modelled hourly PM_{10} concentrations have been compared with the AirBase data (see https://acm.eionet.europa.eu/databases/airbase/). The traffic stations were not included in the comparison since a regional model with a somewhat coarse resolution will not be able to calculate very large concentrations (e.g. hourly concentration higher than 200 µg/m^3) which may be measured by these stations. Indeed, the concentrations calculated by a regional model over cities are mostly representative of the urban background. 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). The observations have also been categorized into two sets of data by differentiating the rural stations to the urban stations (as shown in Fig. S2). This follows the procedure done in the yearly evaluation of the EMEP model over Europe (e.g. EMEP Status Report 1/2018). Due to the relatively coarse definition of a city, it appears that stations classified as rural may be present in our city domain.

It was noticed that for the smaller definition of the city edges, i.e. 1 grid cell, there were no rural stations within the city domain. Obviously, by increasing the size of the city domain, to 9 grid cells or by using the GADM definition, the number of rural stations present within the city domain increases. Indeed, all the hourly measurements are averaged within the city boundary, by separating the urban and the rural stations. A comparison with these two types of stations can highlight a difference between the urban background and the urban concentrations. For such comparison, the model concentrations are also averaged over the city domain.

3.2.2. Results

Figures 3 and 4 show the comparison between the hourly averaged observations within the city edges defined by the 9 grid cells definition, and the predictions from EMEP and from LOTOS-EUROS respectively.

Figures 3 and 4 show that for the urban stations, the different predictions from a same model, for the same date, are consistent since the values for the statistical parameters are relatively constant. It is noticed; however, that the bias is slightly reduced when the starting date of the forecast is closer to the target date. The available observations and thus the stations may also differ from day to day (e.g. Fig. S2a). Figures 3 and 4 also show that despite many differences, the models have very similar performances in comparison with the urban stations.

In Figure 3, it is also clear that the EMEP model has difficulties to reproduce the highest concentrations measured by the urban stations which are probably smoothed by the model over the large grid cells as the ones defining the cities. The underestimation in the largest urban concentrations is highlighted by the comparison with the rural stations. This also shows that over the area defining the cities there is a large variability in the measured PM_{10} concentrations

and that few stations are not necessarily representative of the model grids. It also shows with such resolution; the model represents urban background concentrations.
Only 5 cities have measurements defined as rural stations by using the 9 grids definition (i.e. Amsterdam, Berlin, Luxembourg, Rotterdam and Vienna) while there are up to 19 cities with urban stations. By comparing only the 5 cities having urban and rural stations, the agreement between EMEP and the urban stations is largely improved as shown in Fig. S3. We can also notice that the difference in concentrations predicted by the EMEP model between both types of stations is also reduced. This shows that for these five cities, the predicted PM_{10} concentrations on December 6th are higher than over the other cities.
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).
By using the GADM definition, the number of cities having rural stations decreases to 2 while the number of cities with the urban stations remains identical.
In general, both models present similar performance with the observations especially for the NMB, RMSE and FGE as presented in Figures S5 and S6. These figures show an overview of the statistical parameters for all 4-d forecasts, i.e. the dates from December 01st to 12th 2016 with a starting date from December 01st to 09th, for all the cities defined by 9 grid cells, in comparison with the concentrations measured at the urban and the rural stations, respectively.
As already shown by Figs. 3 and 4, LOTOS-EUROS shows slightly better correlation coefficients with the urban stations than EMEP (Fig. S5, in average $R_{LOTOS\text{-}EUROS}$=0.31, R_{EMEP}=0.25; with a maximum of 0.58 for LOTOS-EUROS and 0.5 for EMEP) and EMEP presents better correlations with the few rural stations (Fig. S6, in average $R_{LOTOS\text{-}EUROS}$=0.23, R_{EMEP}=0.35; with a maximum of 0.58 for LOTOS-EUROS and 0.72 for EMEP). However, the limited number of cities having rural stations explain the larger variability in the correlations compared to the correlations found with the urban stations. Similar results are found by using the GADM definition (not shown) while by using only 1 grid to define the city edges, the correlation coefficients with the urban stations are larger (up to 0.8), with an increase in the bias and a decrease in the RMSE (Fig. S7).
In average, both models have a FGE equal to 0.5 over the cities defined by 9 grid cells with the urban stations and 0.4 with the rural stations. For the RMSE, it is 33 µg/m^3 with the urban stations and 11 µg/m^3 with the rural stations. While both models underestimate the PM_{10} concentrations by 36% in average by using the urban sites, EMEP overestimates by 6% with the rural stations and LOTOS-EUROS underestimates by 6%.
Performances of both models are improved with daily means, especially with better correlation coefficients (not shown). For example, with the cities defined by 9 grid cells, the correlation coefficients reach 0.8 with the urban stations for EMEP and LOTOS-EUROS and 0.98 with the rural stations for EMEP. However, a lot of negative correlation coefficients between LOTOS-EUROS and the rural stations are noticed. The correlation coefficient with the rural stations remains difficult to interpret related to the limited number of stations available. Thus, EMEP presents a mean correlation coefficient equal to 0.4 with the urban and rural stations, and LOTOS-EUROS has a mean correlation of 0.5 with the urban stations and only 0.06 with the rural stations. Better scores with the FGE and the RMSE are also noticed in comparison to the hourly evaluation (not shown). Both models present with these 9 grids definition a mean

FGE of 0.5 with the urban stations and 0.3 for the rural stations and a mean RMSE of 21 µg/m^3 with the urban stations and 10 µg/m^3 with the rural stations.

3.3. Inter-comparison in the concentrations predicted by both models

The second analysis has been focused on the agreement between both models. During the episode, all 4-d forecasts present a high correlation between the PM_{10} predicted by the EMEP model and LOTOS-EUROS as shown by Figure 5a. These correlations vary from day to day and city by city but remain large for the different simulated periods (median = 0.7).
There is no clear geographical pattern in terms of performance between the two models, even if the central European cities (e.g. Budapest, Vienna, Warsaw) presented the larger differences (Fig. 5b). These differences may be explained by slightly lower Secondary Inorganic Aerosols (SIA = NO_3^- + NH_4^+ +SO_4^{2-}) in LOTOS-EUROS for these cities but also by lack of water in LOTOS-EUROS (which is not diagnosed as mentioned in Sect. 2). Moreover, it confirms the larger PM_{10} concentrations predicted by EMEP than by LOTOS-EUROS for the five cities plotted in Figs. S3 and S4. It is also worth noting that LOTOS-EUROS predicts more sea salt and dust for almost all the cities during the studied period (Fig. S8) which is representative of the overall feature over the regional domain (not shown). Actually, it was noticed that for the predicted PM_{10} with the larger positive NMB (EMEP predicting larger PM_{10} concentrations), EMEP has more SIA and "other" than LOTOS-EUROS (Figure S9a), while the PM_{10} from LOTOS-EUROS is dominated by natural components when a larger negative NMB is predicted (Figure. S9b).

4. Methodology of the source contribution calculation

4.1 The EMEP model

4.1.1 Emission reductions

The SC calculation follows the methodology uses in each EMEP annual report to quantify the annual country-to-country source receptor relationships (e.g. EMEP Status Report 1/2018). The experiment is based on a reference run, where all the anthropogenic emissions are included. The other runs are the perturbation runs. These runs correspond to the simulations where the emissions from every considered country are reduced by 15%. As explained in Wind et al. (2004), a reduction by 15% is sufficient to give a clear signal in the pollution changes. It also gives a negligible effect from non-linearity in the chemistry even if in this work it has been estimated.
The perturbation runs are done for anthropogenic emissions of CO, SO_x, NO_x, NH_3, NMVOC and PPM (primary particulate matter). For computational efficiency, in the perturbation calculations, all anthropogenic emissions in the perturbation runs have been reduced here simultaneously. This simultaneous reduction differs from the methodology uses in each EMEP annual report where the emissions are reduced individually.
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.

4.1.2 Issue concerning the chemical non-linearity

The reason why emissions should not be perturbed by 100% in the model simulations is to stay within the linear regime of involved chemistry. Even limited, such methodology may still introduce a non-linearity in the chemistry. The total PM_{10} over the receptor should be identical theoretically to the sum of the PM_{10} originated from the different sources. This is not always the case and the difference between the total PM_{10} and the sum from the various sources may lead to negative or positive concentrations. This is a result of the perturbation used which is assumed to be linear to a 100% perturbation.
The 15% emission reduction has been used during many years for the annual country-to-country source receptor relationships calculations (e.g. EMEP Status Report 1/2018). Clappier et al. (2017a) have already shown the robustness of the methodology at the country scale on yearly averages and for the highest daily concentrations. However, this emission reduction was not used for smaller areas. Thus this 15% emission reduction for the study over a city and on hourly basis has been tested, in order to assess the robustness of the calculations. 5% and 50% were the other selected emission reductions. In total, 847 4-day runs have been performed in this work (9 reference runs, and 9 dates × 31 countries × 3 perturbations runs).
Furthermore, by reducing the emissions simultaneously or separately may lead to a different result in the concentrations, but as mentioned previously, this effect is not addressed in this work for computational reason.

4.2. LOTOS-EUROS

A labelling technique has been developed within each LOTOS-EUROS simulation (Kranenburg et al., 2013). An important advantage of the labelling technique is the reduction of computation costs and analysis work associated with the calculations. The source apportionment technique has been previously used to investigate the origin of PM (Hendriks et al., 2013; 2016), NO_2 (Schaap et al., 2013), and nitrogen deposition (Schaap et al., 2018).
Besides the concentrations of all species the contributions of a number of sources to all components are calculated. The labelling routine is only implemented for primary, inert aerosol tracers and chemically active tracers containing a C, N (reduced and oxidized) or S atom, as these are conserved and traceable. This technique is therefore not suitable to investigate the origin of e.g. O_3 and H_2O_2, as they do not contain a traceable atom. The source apportionment module for LOTOS-EUROS provides a source attribution valid for current atmospheric conditions as all chemical conversions occur under the same oxidant levels. For details and validation of this source apportionment module we refer to Kranenburg et al. (2013).
To avoid violating the memory size and avoid excessive computation times it was chosen to trace the EU-28 countries, supplemented by Norway and Switzerland. For convenience, a number of small countries was combined with a neighboring state. For example, Switzerland and Liechtenstein as well as Luxembourg and Belgium were combined. In addition, all sea areas were combined into one source area. To be mass consistent, all non-specified regions (denoted Rest), natural emissions and as well as the combined impact of initial conditions and boundary conditions were given labels as well.

5. Information provided by the Source Contribution calculations

5.1 In the EMEP calculations

As presented in Fig. 1, the country contributions to the predicted PM_{10} concentrations in the cities is provided in our products.
Figure 6 presents the mean composition for the "Domestic", "30 European countries" and "Others" PM_{10} contributions for all cities, for all 4-d predictions and split into negative and positive concentrations. This figure is a result of the perturbation runs by separating the positive and the negative concentrations obtained in the calculations. The concentrations have also been gathered by their calculated origin. The "Domestic" contribution corresponds to the contribution from the domestic country to the city (for example from France to Paris). The "30 European countries" corresponds to the other 30 European countries used in the study. "Others" gathers mainly natural sources, the other European countries included in the regional domain (and not included in our SC calculations, e.g. Turkey) and the boundary conditions. This figure gives a graphical illustration of the composition of the different contributions and presents the effect of the non-linearity. Indeed, the positive concentrations shows the overall composition for each contribution, while the chemical reason of the non-linearity is highlighted by the negative contribution to the predicted PM_{10} concentrations.
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). Actually, the variation in the mean concentrations is mainly influenced by the variation in these primary components. NO_3^- is also an important component of these "Domestic" PM_{10}. The value of the mean concentration depends on the city definition and so on the average of the concentrations over different size of city. The mean PM_{10} concentration over a smaller area is larger showing that with a smaller grid, the PM_{10} is less diffused over the integrated area. The "30 European countries" PM_{10} is mainly influenced by NO_3^- (by 38%) (Fig. 6b).
Overall, 45% of the contributions to the PM_{10} calculated over the selected cities for this episode are "Domestic" and essentially due to primary components. 35% are from the "30 European countries", essentially NO_3^- and 25% are from "Others" mainly composed by natural sources (representing 50% of "Others"). Obviously, this feature is an overview of all selected cities for all the studied dates and it can vary from city to city and from date to date.
By comparing the PM_{10} concentrations calculated over the same city edges but by using different percentages in the perturbation runs, we have calculated the impact of the non-linearity for each contribution and presented in Figure 7. 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):

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

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. This mean non-linearity due to the "Domestic" contribution represents in maximum 0.9% of the total PM_{10}. This non-linearity from the "30 European countries" contribution, counts for 0.7% of the total PM_{10} and 1.5% from "Others". Actually, the non-linearity from the "Others" depends on the non-linearity from the two other

contributions. The mean non-linearity is not homogenously distributed over all cities as shown in Figure S10 and may vary from date to date (not shown). It has remained limited even if some hourly contributions show higher non-linearity. In maximum, 3% of the calculated hourly contributions for all 4-day forecasts over the selected cities have a non-linearity higher than 5% (not shown). This shows that due to the methodology used in the EMEP model, based on a reduced emission scenario, the non-linearity in the chemistry has a limited impact on the SC calculation. This non-linearity is slightly reduced by using the larger domains to define the cities (e.g. 9 grids) (Fig. 7). This also shows that the responses to perturbation runs are robust, even if only the non-linearity in the chemistry related the perturbation used, and not the one related to the reduction of each emission precursor, has been estimated in this study as mentioned in Section 4.1.

Negligible negative contributions have been calculated for the "Domestic" and "30 European countries" contributions (Figs. 6a & b) and small negative contributions are predicted in "Others" (Fig. 6c). These negative PM_{10} are a result of negative values in NO_3^-, NH_4^+ and H_2O which is a consequence of gas-aerosol partitioning of the species. Indeed, NH_3 reacts with nitric acid (HNO_3) to form ammonium nitrate (NH_4NO_3). This is an equilibrium reaction, and thus the transition from solid to gaseous phase depend on relative humidity (e.g. Fagerli and Ass, 2008; Pakkanen, 1996). This shows that, for example, a reduction in NO_x over a country which impacts the selected city, does not necessarily only impact the NO_3^- over this city, but may also have an effect on NH_3 chemistry over a second region. This second region may also have itself an impact on the selected city. This combination of NO_x and NH_3 chemistry from different regions may lead at the end to these negative concentrations.

The impact of the percentage used in the perturbation runs and the size of the city edges have no significant impact in the amount of negative "Others" PM_{10} concentrations. The impact of both parameters is more visible on the "Domestic" and "Rest of Europe" concentrations but it remains very small.

Averaging out over the larger grids reduces globally the non-linearity. The 15% emission reduction also reduces the negative non-linearity in the "Domestic" concentrations (e.g. H_2O for the 9 grids and GADM runs).

5.2 In the LOTOS-EUROS calculations

As presented with the EMEP predictions, Figure 8 presents the mean composition for the "Domestic", "30 European countries" and "Others" PM_{10} contributions for all cities, for all 4-d predictions provided by LOTOS-EUROS. The definition of "Others" is slightly different than the EMEP one since e.g. the dust from agriculture and traffic is included (see Sect. 2). For an easier comparison, the result for the EMEP model using the 15% emission reduction has also been plotted with thinner charts, even if, as just mentioned, the definition of "Others" slightly differs between both models.

First of all, during the episode, LOTOS-EUROS confirms the general trend calculated by the EMEP model, i.e. the dominant contribution to the surface PM_{10} is "Domestic", ranging between 40% and 48% of the predicted PM_{10} over all selected cities and for all the studied dates. However, LOTOS-EUROS always presents more "Domestic" PM_{10} than the EMEP model. LOTOS-EUROS also predicted slightly more influence from "Others" than the "30 European countries" with a ratio close to 25-30% each. As reminder, the EMEP model predicted a slightly larger influence from the "30 European countries" (35%) than from "Others" (25%).

As with the EMEP model, the mean PM_{10} concentration over the smaller city definition is larger and the "Domestic" PM_{10} is largely driven by POM. 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}. POM and "Rest" represent each between 25% and 30% of these "Domestic" PM_{10}.
The large influence of NO_3^- (48%) in the "30 European countries" PM_{10} is also calculated by LOTOS-EUROS, as well as the large contribution of the natural components (60%) in "Others". It is noteworthy to see that, even small, the dust emitted by the road traffic and the agriculture is not negligible in these "Others" PM_{10} (~10%).

6. Comparison between both country source contribution calculations

Section 3 has highlighted the similar performance from both models in the prediction of the PM_{10} concentrations over the European cities with observations. 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. Section 5 has shown similar results in terms of composition of these PM_{10}. It is also noteworthy to see in Figure 9 that both SC calculations present a high rate of agreement over the selected period with the common simulated components and the PM_{10} calculated by both models. 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. In Fig. 9, both models show that, by using the 9 grid cells definition, in average 68% of the hourly predicted PM_{10} concentrations have the same dominant country contributor. In average, 50% of the secondary inorganic aerosols predicted by both models over all the cities and all 4-day forecasts have the same main contributor. This value goes up to 70% for POM and 80% for EC. For the two primary components (POM and EC) the median is larger, with a value of 77% and 93% respectively, showing that the mean value in the agreement for both compounds is reduced by a few low values (Fig. 9). On a daily basis, the mean agreement is slightly improved, e.g. 70% of agreement for the PM_{10} (Fig. S11). The main improvement is calculated for EC, with a median equal to 100% (Fig. S11).
The lower agreement for the SIA is predictable due to the various origins (chemistry and primary emissions) for these particulates and the different aerosols treatment (gas-aerosols partitioning) in both models. It is also related to the differences in both methodologies (e.g. Clappier et al, 2017b). Indeed, an emission reduction and a labelling technique will not necessarily provide the same results for the secondary PM. An emission reduction depends on the atmospheric composition already present. For example, an amount of NO_x emitted over a source can result in a certain NH_4NO_3 concentration in the receptor. If this NO_x is emitted in excess (NH_3 limited regime), a NO_x emission reduction will have a small effect at the receptor point. On the other hand, in the NO_x limited regime, the same NO_x reduction will have a large impact. The labelling method will give the same result in both cases while the scenario approach will give different results.
This agreement varies from city to city (Fig. 10) but it is shown, in addition to the example of PM_{10} (Fig. 5), that central European cities often present a limited agreement due to their central location and the influence of various countries. This limited agreement is also sometimes observable for the cities close to the edge of the regional domain (Fig. 10), which could be explained by the influence of the boundary conditions as the dust transported from other regions (e.g. Valetta influenced by dust from Sahara).
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. This rate is

around 70% for SO_4^{2-}, EC and POM and close to 60% for NO_3^- and equal to 65% for NH_4^+ (Fig. 11). As for the dominant country contributor, the agreement is slightly improved by using daily means, e.g. we found 76% of agreement with the PM_{10} (not shown).
It is also important to notice that these overall agreements are neither significantly influenced by the definition of the cities area nor on the perturbation percentage tested for the EMEP SC calculations (Fig. S12). The agreement is slightly better by using the smaller area (1 grid) in the determination of the dominant country contributor and slightly better by using a large domain (9 grids or GADM) in the determination of the 2 and 5 main contributors.
Overall, a perturbation run using a reduction of 15% and the use of a larger city area (e.g. GADM or 9 grids) allow a better determination in the country contributors, with a better agreement with LOTOS-EUROS and limit the impact of the non-linearity in the chemistry.

7. Conclusions

By focusing on a specific event, occurring from December 01st to 09th 2016 over Europe, this work is the first attempt to evaluate the source contribution calculations provided by two regional models (EMEP and LOTOS-EUROS) in a forecast mode. Together, the models compose the operational source contribution prediction system for the European cities within the Copernicus Atmosphere Monitoring Service (CAMS) and aim to estimate the impact of the long-range transport to urban PM_{10}. These models also use two distinct source apportionment methodologies, a labeling technique for LOTOS-EUROS and the use of perturbation runs for EMEP.
The methodology used for the EMEP model was tested by using three different percentages (5%, 15% and 50%) in the perturbation runs. The importance in the choice of the domain defining the edges of the studied cities was also investigated in terms of predicted concentrations and calculated contributors. It was concluded that the 15% emission reduction and the use of large city areas (9 grids or GADM) were the more efficient. It reduces the impact of non-linearity, which especially impacts the NO_3^-, NH_4^+ and H_2O concentrations, and it presents a better agreement in the determination of main country contributors. The mean non-linearity always represents less than 2% of the total modelled PM_{10} for each contribution calculated by the EMEP SC and is caused by the perturbation used which is assumed to be linear to a 100% perturbation. Even if this non-linearity is not identical for all cities and for the different dates, the larger non-linearities (>5%) impact only 3% of all the calculated hourly contributions. However, the non-linearity related to the reduction of each emission precursor has not been calculated in the study for computational reason.
The predicted PM_{10} concentrations were compared with AirBase observations showing fair agreement even if the models remain perfectible since they have difficulties to reproduce the highest hourly concentrations measured by the urban stations (mean underestimation by 36%). 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. It was also noticed the bias is slightly reduced when the forecast is closer to the studied date. An inter-comparison between both models was also performed showing satisfactory results with few discrepancies in the predictions of the PM_{10} concentrations, mainly explained by an underestimation in sea salt and dust by the EMEP model (compared to LOTOS-EUROS); and differences in SIA, caused by different chemical aerosols treatment in both models.
During the episode, both models have shown that 45% of the predicted PM_{10} over the selected cities were from “Domestic” sources and essentially composed of primary components. The rest of the contribution was roughly equitably split into an influence from the others 30

countries used in the regional domain, essentially composed of NO_3^- and from "Others" mainly composed of natural sources.
It was shown that the results from both source apportionment methodologies agree in average by 68% in the determination the dominant country contributor to the hourly PM_{10} concentrations and 75% for the top 5 of these country contributors. The daily country attribution also presents similar agreement. The differences seen are mainly related to the SIA and is a direct consequence of the difference between both methodologies used.
A full year of evaluation will be necessary to confirm our satisfactory results. Moreover, the bias of the predicted PM_{10} concentrations with the urban observations probably suggests an underestimation of the "Local" contribution (from the city) which is also predicted by the EMEP model. This is investigated in a companion paper (in preparation), also focusing on the same event.

Data availability

The EMEP model is an open source model available on https://doi.org/10.5281/zenodo.3355041. The base-code of LOTOS-EUROS is available under the license on https://lotos-euros.tno.nl/, but the code used for this study, including the source apportionment is only available in cooperation with TNO. The data processing and analysis scripts are available upon request.

Author contribution

MP, HF and M Schulz designed the research. MP performed the experiment. MP developed the analyzing codes and analyzed the data. AV developed the EMEP part of the forecasting system. RK and M Schaap performed and provided the LOTOS-EUROS results. MP wrote the paper with the inputs from all coauthors.

Acknowledgments

This work is partly funded by the EU Copernicus project CAMS 71 to provide policy support. This work has also received support from the Research Council of Norway (Programme for Supercomputing) through the EMEP project (NN2890K) for CPU and the Norstore project "European Monitoring and Evaluation Programme" (NS9005K) for storage. The EMEP project itself is supported by the Convention on the Long Range Transmission of Air Pollutants, under UN-ECE. The authors thank A. Mortier (Norwegian Meteorological Institute) for the development and the design of the website (https://policy.atmosphere.copernicus.eu/SourceContribution.php).

References

Amann, M., Bertok, I., Borken-Kleefeld, J., Cofala, J., Heyes, C., Höglund-Isaksson, L., Klimont, Z., Nguyen, B., Posch, M., Rafaj, P., Sandler, R., Schöpp, W., Wagner, F., Winiwarter, W.: Cost-effective Control of Air Quality and Greenhouse Gases in Europe: Modeling and Policy Applications, 26 ,1489-1501, 2011.

Andersson-Sköld, Y. and Simpson, D.: Comparison of the chemical schemes of the EMEP MSC-W and the IVL photochemical trajectory models, Atmos. Env., 33(7), 1111-1129, doi:10.1016/S1352-2310(98)00296-9, 1999.

Banzhaf, S., Schaap, M., Kerschbaumer, A., Reimer, E., Stern, R., van der Swaluw, E., Builtjes, P.: Implementation and evaluation of pH-dependent cloud chemistry and wet deposition in the chemical transport model REM-Calgrid. Atmos. Env., 49, 378–390, doi:10.1016/j.atmosenv.2011.10.069, 2012.

Bergström, R., Denier van der Gon, H. A. C., Prévôt, A. S. H., Yttri, K. E., and Simpson, D.: Modelling of organic aerosols over Europe (2002–2007) using a volatility basis set (VBS) framework: application of different assumptions regarding the formation of secondary organic aerosol, Atmos. Chem. Phys., 12, 8499-8527, https://doi.org/10.5194/acp-12-8499-2012, 2012.

Binkowski, F. S., and Shankar, U.: The Regional Particulate Matter Model 1. Model description and preliminary results, J. Geophys.Res., 100(D12), 26,191–26,209, doi:10.1029/95JD02093, 1995.

Bott, A.: A positive definite advection scheme obtained by nonlinear renormalization of the advective fluxes, Mon. Weather Rev.,117(5), 1006-1016, doi:10.1175/1520-0493(1989)117(1006:APDASO)2.0.CO;2, 1989.

Brasseur, G. P., Xie, Y., Petersen, A. K., Bouarar, I., Flemming, J., Gauss, M., Jiang, F., Kouznetsov, R., Kranenburg, R., Mijling, B., Peuch, V.-H., Pommier, M., Segers, A., Sofiev, M., Timmermans, R., van der A, R., Walters, S., Xu, J., and Zhou, G.: Ensemble forecasts of air quality in eastern China – Part 1: Model description and implementation of the MarcoPolo–Panda prediction system, version 1, Geosci. Model Dev., 12, 33-67, https://doi.org/10.5194/gmd-12-33-2019, 2019.

Burr, M.J. and Zhang, Y.: Source apportionment of fine particulate matter over the Eastern U.S. Part II: source sensitivity simulations using CAMX/PSAT and comparisons with CMAQ source sensitivity simulations, Atmospheric Pollution Research, 2, 318-336, 2011.

Callaghan, A., de Leeuw, G., Cohen, L., and O'Dowd, C. D.: Relationship of oceanic whitecap coverage to wind speed and wind history, Geophys. Res. Lett., 35, L23609, https://doi.org/10.1029/2008GL036165, 2008.

Clappier, A., Fagerli, H., Thunis, P.: Screening of the EMEP source receptor relationships: application to five European countries, Air Qual. Atmos. Health, 10, 4, 497-507, DOI: 10.1007/s11869-016-0443-y, 2017a.

Clappier, A., Belis, C. A., Pernigotti, D., and Thunis, P.: Source apportionment and sensitivity analysis: two methodologies with two different purposes, Geosci. Model Dev., 10, 4245-4256, https://doi.org/10.5194/gmd-10-4245-2017, 2017b.

Crippa, M., Janssens-Maenhout, G., Dentener, F., Guizzardi, D., Sindelarova, K., Muntean, M., Van Dingenen, R., and Granier, C.: Forty years of improvements in European air quality: regional policy-industry interactions with global impacts, Atmos. Chem. Phys., 16, 3825-3841, https://doi.org/10.5194/acp-16-3825-2016, 2016.

Dockery, D. W. and Pope III, C. A.: "Acute respiratory effects of particulate air pollution", Ann. Rev. Public Health, 15, 107-132, DOI: 10.1146/annurev.pu.15.050194.000543, 1994.

D'Elia, I., Bencardino, M., Ciancarella, L., Contaldi, M., Vialetto, G.:Technical and Non-Technical Measures for air pollution emission reduction: The integrated assessment of the regional Air Quality Management Plans through the Italian national model, Atmos. Env., 43, 6182-6189, https://doi.org/10.1016/j.atmosenv.2009.09.003, 2009.

EEA Report No 13/2017, Air quality in Europe 2017, https://www.eea.europa.eu/publications/air-quality-in-europe-2017

EMEP Status Report 1/2003: "Transboundary acidification and eutrophication and ground level ozone in Europe: Unified EMEP model description", The Norwegian Meteorological Institute, Oslo, Norway, ISSN 0806-4520, 2003.

EMEP Status Report 1/2018: "Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components", Joint MSC-W & CCC & CEIP Report, ISSN 1504-6109, 2018.

Fagerli, H. and Aas, W.: Trends of nitrogen in air and precipitation: Model results and observations at EMEP sites in Europe, 1980–2003, Env. Poll., 154, 3, 448-461, https://doi.org/10.1016/j.envpol.2008.01.024, 2008.

Founda, D., Kazadzis, S., Mihalopoulos, N., Gerasopoulos, E., Lianou, M., and Raptis, P. I.: Long-term visibility variation in Athens (1931–2013): a proxy for local and regional atmospheric aerosol loads, Atmos. Chem. Phys., 16, 11219-11236, https://doi.org/10.5194/acp-16-11219-2016, 2016.

Fountoukis, C. and Nenes, A.: ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K^+–Ca^{2+}–Mg^{2+}–NH_4^+–Na^+–SO_4^{2-}–NO_3^-–Cl^-–H_2O aerosols, Atmos. Chem. Phys., 7, 4639-4659, https://doi.org/10.5194/acp-7-4639-2007, 2007.

Gerber, H. E.: Relative-Humidity Parameterization of the Navy Aerosol Model (NAM), Naval Research Laboratory, NRL report 8956, 1985.

Grewe, V., Tsati, E., and Hoor, P.: On the attribution of contributions of atmospheric trace gases to emissions in atmospheric model applications, Geosci. Model Dev., 3, 487-499, https://doi.org/10.5194/gmd-3-487-2010, 2010.

Hendriks, C., Kranenburg, R., Kuenen, J., van Gijlswijk, R., Wichink Kruit, R., Segers, A., Denier van der Gon, H., Schaap, M.: The origin of ambient particulate matter concentrations in the Netherlands, Atmos. Env., 69, 289–303, doi:10.1016/j.atmosenv.2012.12.017, 2013.

Hendriks, C., Kranenburg, R., Kuenen, J.J.P., Van den Bril, B., Verguts, V., Schaap, M.: Ammonia emission time profiles based on manure transport data improve ammonia modelling across north western Europe. Atmos. Environ. 131, 83–96, doi:10.1016/j.atmosenv.2016.01.043, 2016.

Jonson, J. E., Stohl, A., Fiore, A. M., Hess, P., Szopa, S., Wild, O., Zeng, G., Dentener, F. J., Lupu, A., Schultz, M. G., Duncan, B. N., Sudo, K., Wind, P., Schulz, M., Marmer, E., Cuvelier, C., Keating, T., Zuber, A., Valdebenito, A., Dorokhov, V., De Backer, H., Davies, J., Chen, G. H., Johnson, B., Tarasick, D. W., Stübi, R., Newchurch, M. J., von der Gathen, P., Steinbrecht, W., and Claude, H.: A multi-model analysis of vertical ozone profiles, Atmos. Chem. Phys., 10, 5759-5783, doi:10.5194/acp-10-5759-2010, 2010.

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527-554, https://doi.org/10.5194/bg-9-527-2012, 2012.

Keuken, M, Zandveld, P., van den Elshout, S., Janssen, N. A. H., Hoek, G.: Air quality and health impact of PM_{10} and EC in the city of Rotterdam, the Netherlands in 1985–2008. Atmos Env., 45, 5294–5301, doi:10.1016/j.atmosenv.2011.06.058, 2011.

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., Terrenoire, E., Fagerli, H., Nyiri, A., and Amann, M.: Modelling street level PM_{10} concentrations across Europe: source apportionment and possible futures, Atmos. Chem. Phys., 15, 1539-1553, https://doi.org/10.5194/acp-15-1539-2015, 2015.

Kranenburg, R., Segers, A. J., Hendriks, C., and Schaap, M.: Source apportionment using LOTOS-EUROS: module description and evaluation, Geosci. Model Dev., 6, 721-733, https://doi.org/10.5194/gmd-6-721-2013, 2013.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10963-10976, https://doi.org/10.5194/acp-14-10963-2014, 2014.

Manders, A. M. M., Builtjes, P. J. H., Curier, L., Denier van der Gon, H. A. C., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J. J. P., Segers, A. J., Timmermans, R. M. A., Visschedijk, A. J. H., Wichink Kruit, R. J., van Pul, W. A. J., Sauter, F. J., van der Swaluw, E., Swart, D. P. J., Douros, J., Eskes, H., van Meijgaard, E., van Ulft, B., van Velthoven, P., Banzhaf, S., Mues, A. C., Stern, R., Fu, G., Lu, S., Heemink, A., van Velzen, N., and Schaap, M.: Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model, Geosci. Model Dev., 10, 4145-4173, https://doi.org/10.5194/gmd-10-4145-2017, 2017.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geosci. Model Dev., 8, 2777-2813, https://doi.org/10.5194/gmd-8-2777-2015, 2015.

Mårtensson, E.M., Nilsson, E.D., de Leeuw, G., Cohen, L.H. and Hansson, H.C.: Laboratory simulations and parameterization of the primary marine aerosol production. J. Geophys. Res. Atmos., 108., D9, 4297, doi:10.1029/2002JD002263,2003.

Meyer, S. and Pagel, M.: Fresh Air Eases Work - The Effect of Air Quality on Individual Investor Activity, NBER Working Paper No. 24048, DOI: 10.3386/w24048, 2017.

Monahan, E., Spiel, D., and Davidson, K.: A model of marine aerosol generation via white caps and wave disruption, in:
Oceanic whitecaps, edited by: Monahan, E. and MacNiochaill, G., 167–193, Dordrecht: Reidel, The Netherlands, 1986.

Morcrette, J.-J, Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M., Serrar, S., Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M., Untch, A.: Aerosol analysis and forecast in the ECMWF Integrated Forecast System: Forward modeling, J. Geophys. Res., 114, D06206, doi:10.1029/2008JD011235, 2009.

Mukherjee, A., and Agrawal, M.: World air particulate matter: sources, distribution and health effects, Environmental Chemistry Letters, 15, 2,283-309, DOI: 10.1007/s10311-017-0611-9, 2017.

Pakkanen, T. A.: Study of formation of coarse particle nitrate aerosol, Atmos. Env., 30, 2475-2482, doi: 1352-2310(95)00492-0, 1996

REVIHAAP, Review of Evidence on Health Aspects of Air Pollution - REVIHAAP Project Technical Report, World Health Organization (WHO) Regional Office for Europe, Bonn, 2013, http://www.euro.who.int/__data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf

Schaap, M., van Loon, M., ten Brink, H. M., Dentener, F. J., and Builtjes, P. J. H.: Secondary inorganic aerosol simulations for Europe with special attention to nitrate, Atmos. Chem. Phys., 4, 857-874, https://doi.org/10.5194/acp-4-857-2004, 2004.

Schaap, M., Manders, A. M. M., Hendriks, E. C. J., Cnossen, J. M., Segers, A. J. S., Denier van der Gon, H. A. C., Jozwicka, M., Sauter, F. J., Velders, G. J. M., Matthijsen J., and Builtjes, P.J.H. : Regional modelling of particulate matter for the Netherlands, PBL-rapport 500099008, Den Haag/Bilthoven: PBL, 2009.

Schaap, M., Kranenburg, R., Curier, L., Jozwicka, M., Dammers, E., and Timmermans, R.: Assessing the Sensitivity of the OMI-NO_2 Product to Emission Changes across Europe, Remote Sens. 5, 4187–4208, doi:10.3390/rs5094187, 2013.

Schaap, M., Hendriks, C., Kranenburg, R., Kuenen, J., Segers, A., Schlutow, A., Nagel, H.-D., Ritter, A., and Banzhaf, S.: PINETI-III: Modellierung und Kartierung atmosphäri-scher Stoffeinträge von 2000 bis 2015 zur Bewer-tung der ökosystem-spezifischen Gefährdung von Biodiversität in Deutschland. UBA-Texte. Retrieved from https://gis.uba.de/website/depo1/download/PINETI-3_Abschlussbericht_fuer_mapserver_vorl.pdf, 2018.

Segersson, D., Eneroth, K., Gidhagen, L., Johansson, C., Omstedt, G., Engström Nylén, A. and Forsberg, B.: Health Impact of PM_{10}, $PM_{2.5}$ and Black Carbon Exposure Due to Different Source Sectors in Stockholm, Gothenburg and Umea, Sweden, Int. J. Environ. Res. Public Health, 14(7), 742; doi:10.3390/ijerph14070742, 2017.

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., Nyíri, A., Richter, C., Semeena, V. S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., and Wind, P.: The EMEP MSC-W chemical transport model – technical description, Atmos. Chem. Phys., 12, 7825-7865, https://doi.org/10.5194/acp-12-7825-2012, 2012.

Simpson, D., Ágnes Nyíri, A., Tsyro, S., Valdebenito, Á, and Wind, P.: Updates to the EMEP/MSC-W model, 2015-2016 Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components. EMEP Status Report 1/2016, The Norwegian Meteorological Institute, Oslo, Norway, 15-36, ISSN 1504-6109, 2016.

Simpson, D., Bergström, R., Imhof, H., and Wind, P.: Updates to the EMEP/MSC-W model, 2016-2017 Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components. EMEP Status Report 1/2017, The Norwegian Meteorological Institute, Oslo, Norway, 15-36, ISSN 1504-6109, 2017.

Skyllakou, K., Murphy, B. N., Megaritis, A. G., Fountoukis, C., and Pandis, S. N.: Contributions of local and regional sources to fine PM in the megacity of Paris, Atmos. Chem. Phys., 14, 2343-2352, https://doi.org/10.5194/acp-14-2343-2014, 2014.

Steinbrecher, R., Smiatek, G., Köble, R., Seufert, G., Theloke, J., Hauff, K., Ciccioli, P., Vautard, R., Curci, G.: 2009. Intra- and inter-annual variability of VOC emissions from natural and semi-natural vegetation in Europe and neighbouring countries. Atmos. Env., 43, 1380–1391, doi:10.1016/j.atmosenv.2008.09.072, 2009.

Thunis, P., Degraeuwe, B., Pisoni, E., Ferrari, F., Clappier, A.:On the design and assessment of regional air quality plans: The SHERPA approach, J. Env. Management, 183, 952-958, http://dx.doi.org/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. Env., 187, 93-106, doi.org/10.1016/j.atmosenv.2018.05.062, 2018.

Thunis, P., Clappier, A., Tarrason, L., Cuvelier, C., Monteiro, A., Pisoni, E., Wesseling, J., Belis, C. A., Pirovano, G., Janssen, S., Guerreiro, C., and Peduzzi, E.: Source apportionment to support air quality planning: Strengths and weaknesses of existing approaches, Env. Int., 130, 104825, https://doi.org/10.1016/j.envint.2019.05.019, 2019.

Tuovinen, J.-P., Ashmore, M. R., Emberson, L. D., Simpson, D.: Testing and improving the EMEP ozone deposition module, Atmos. Env., 38, 2373-2385, https://doi.org/10.1016/j.atmosenv.2004.01.026, 2004.

Van Zanten, M. C., Sauter, F. J., Wichink Kruit, R. J., Van Jaarsveld, J. A., and Van Pul, W. A. J.: Description of the DEPAC module: Dry deposition modelling with DEPAC GCN2010, RIVM report 680180001/2010, Bilthoven, the Netherlands, 74., 2010.

Walcek, C. J. and Aleksic, N. M.: A simple but accurate mass conservative peak-preserving, mixing ratio bounded advection algorithm with fortran code, Atmos. Environ., 32, 3863–3880, https://doi.org/10.1016/S1352-2310(98)00099-5, 1998.

Wichink Kruit, R. J., Schaap, M., Sauter, F. J., van Zanten, M. C., and van Pul, W. A. J.: Modeling the distribution of ammonia across Europe including bi-directional surface–atmosphere exchange, Biogeosciences, 9, 5261-5277, https://doi.org/10.5194/bg-9-5261-2012, 2012.

Whitten, G., Hogo, H., Killus, J.: The carbon bond mechanism for photochemical smog, Environ. Sci. Technol., 14 (6), 690–700, DOI: 10.1021/es60166a008, 1980.

WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide - Global update 2005 - Summary of risk assessment (available on https://apps.who.int/iris/bitstream/handle/10665/69477/WHO_SDE_PHE_OEH_06.02_eng.pdf?sequence=1).

Wind, P., Simpson, D., Tarrasón, L.: Chap. 4 Source-receptor calculations, EMEP Status Report 1/2004: "Transboundary acidification, eutrophication and ground level ozone in Europe", Joint MSC-W & CCC & CIAM & ICP-M&M & CCE Report, ISSN 0806-4520, 2004.

Zhang, L., Gong, S., Padro, J., and Barrie, L.: A size-segregated particle dry deposition scheme for an atmospheric aerosol module, Atmos. Env., 35, 549–560, https://doi.org/10.1016/S1352-2310(00)00326-5, 2001.

Table. 1 Technical description of both models used in the SC calculation system.

Model	**EMEP/MSC-W**	**LOTOS-EUROS**
Model version	rv4.15 (open source version Sept 2017)	V2.0 (open source version 2016)
Horizontal resolution	0.25° × 0.125° lon-lat	0.25° × 0.125° lon-lat
Regional domain	30°N-76°N 30°W-45°E	31°N-68.875°N 24°W-43.75°E
PBL	Calculation based on turbulent	From ECMWF

	diffusion coefficients (Kz) (EMEP Status Report 1/2003)	
Vertical resolution	20 sigma layers up to 100 hPa, with about 10 in the Planetary Boundary Layer	Mixing layer approach with a 25m surface layer. Model top at 5 km.
Gas phase chemistry	Evolution of the "EMEP scheme" (Andersson-Sköld and Simpson, 1999; Simpson et al. 2012)	TNO-CBM-IV (Schaap et al., 2009)
Nitrate formation	Oxidation of NO_2 by O_3 on aerosols (night and winter) N_2O_5 hydrolysis on aerosol (Simpson et al., 2012)	N_2O_5 hydrolysis on aerosol (Schaap et al., 2004)
Sulphate production	SO_2 oxidation by O_3 and H_2O_2	SO_2 oxidation by O_3 and H_2O_2
Inorganic aerosols	MARS (Binkowski and Shankar, 1995)	ISORROPIA-II (Fountoukis and Nenes, 2007)
Secondary organic aerosols	EmChem09soa (Bergström et al, 2012)	Not included in this model version
Water	PM_{10} particle water at 50% relative humidity	Not diagnosed
Advection	Scheme of Bott (1989)	Monotonic advection scheme (Walceck and Aleksic, 1998)
Dry deposition/sedimentation	Resistance approach for gases and for aerosol, including non-stomatal deposition of NH_3 (EMEP Status Report 1/2003)	Resistance approach for gases and for aerosol, including compensation point for NH_3 (van Zanten et al., 2011; Wichnik Kuit et al., 2012; Zhang, 2001)
Wet deposition	wash out ratio's	pH dependent wash out ratio's accounting for saturation

Dust	Boundary conditions + windblown dust	Boundary conditions + Soil, traffic and agriculture (Schaap et al., 2009)
Sea Salt	Mårtensson et al. (2003), Monahan (1986) production accounting for whitecap area fractions (Callaghan et al., 2008)	Mårtensson et al. (2003), Monahan (1986)
Boundary values	global C-IFS 00UTC	global C-IFS 00UTC, except for sea salt
Initial values	24h forecast from the day before	24h forecast from the day before
Anthropogenic emissions	TNO-MACC-III for 2011	TNO-MACC-III for 2011
Fire emissions	CAMS product: GFAS	CAMS product: GFAS
Biogenic emissions	Emission factors as a function of temperature and solar radiation (Simpson et al., 2012)	Emission factors as a function of temperature and solar radiation (Schaap et al., 2009)
Meteorological driver	12:00 UTC operational IFS forecast (yesterday's)	12:00 UTC operational IFS forecast (yesterday's)

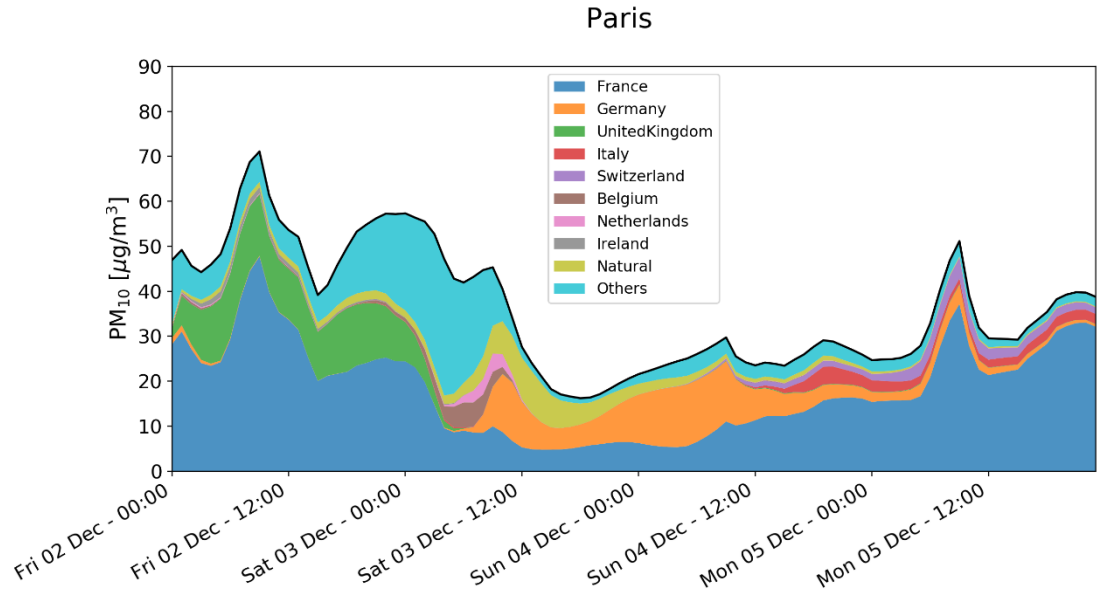


Figure 1: Hourly PM_{10} concentrations in µg/m^3 over Paris predicted by the EMEP model from December 2nd to December 5th 2016. The black curve highlights the total concentration. The eight main country contributors are plotted in addition to the natural sources and "Others". "Others" gathers hereafter other European countries, the boundary conditions, the ship traffic, the biogenic sources, the aircraft emission and the lightning.

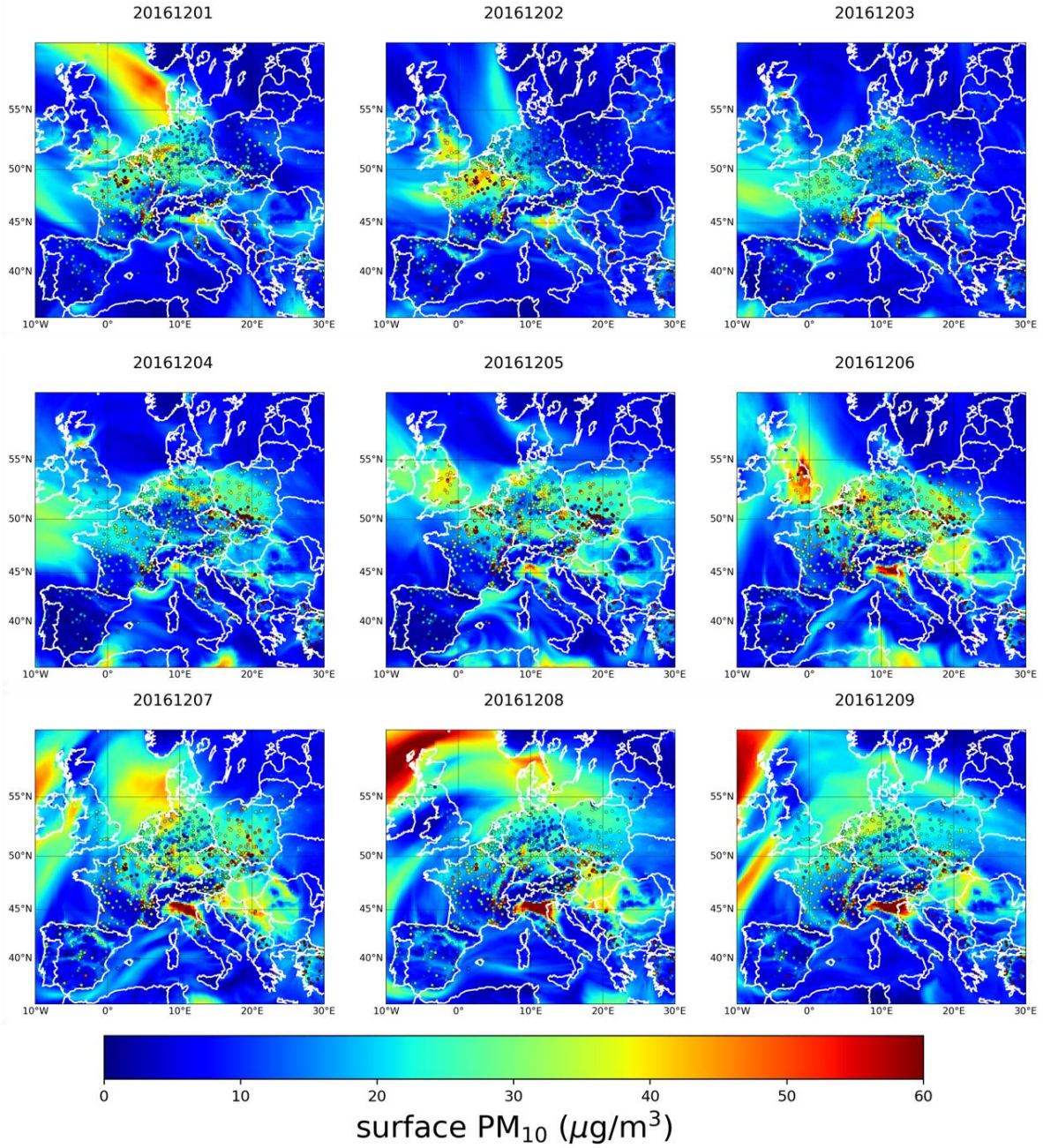


Figure 2: Daily surface PM_{10} concentration in µg/m^3 over Europe predicted by the EMEP model from December 01st to 09th 2016. The colored dots correspond to the daily mean of AirBase stations (rural and urban stations).

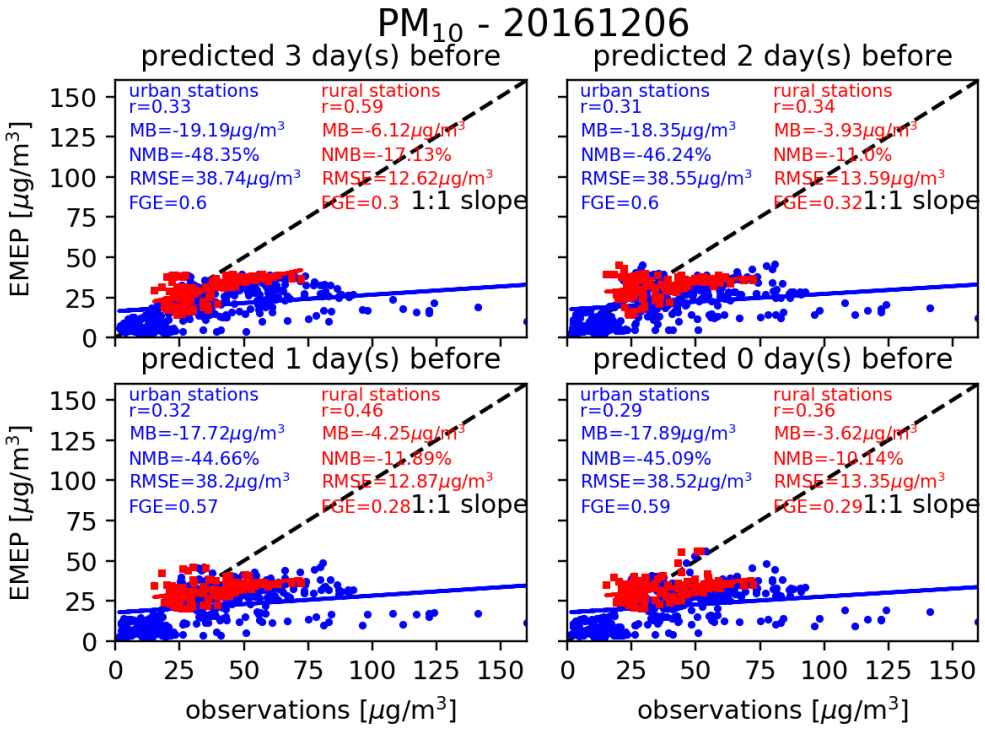


Figure 3: Scatterplot between the hourly PM_{10} concentrations in μg/m³ over all the studied cities using the 9 grid cells definition, predicted by the EMEP model on December 06th 2016 and the observations of the urban sites (blue dot) and rural sites (red square). For this case, there are 19 cities which have urban stations in their domain and 5 cities which have rural stations in their domain. The observations are collocated in time to the EMEP predictions and then averaged within the city edge to match the studied grid. The four panels correspond to the different predictions from 3 days before the December 06th to the actual day, i.e. December 06th. The correlation coefficient (r), the mean bias (MB), the normalized mean bias (NMB), the root-mean-square error (RMSE) and the fractional gross error (FGE) are provided on each panel. The blue and the red lines represent the linear fits.

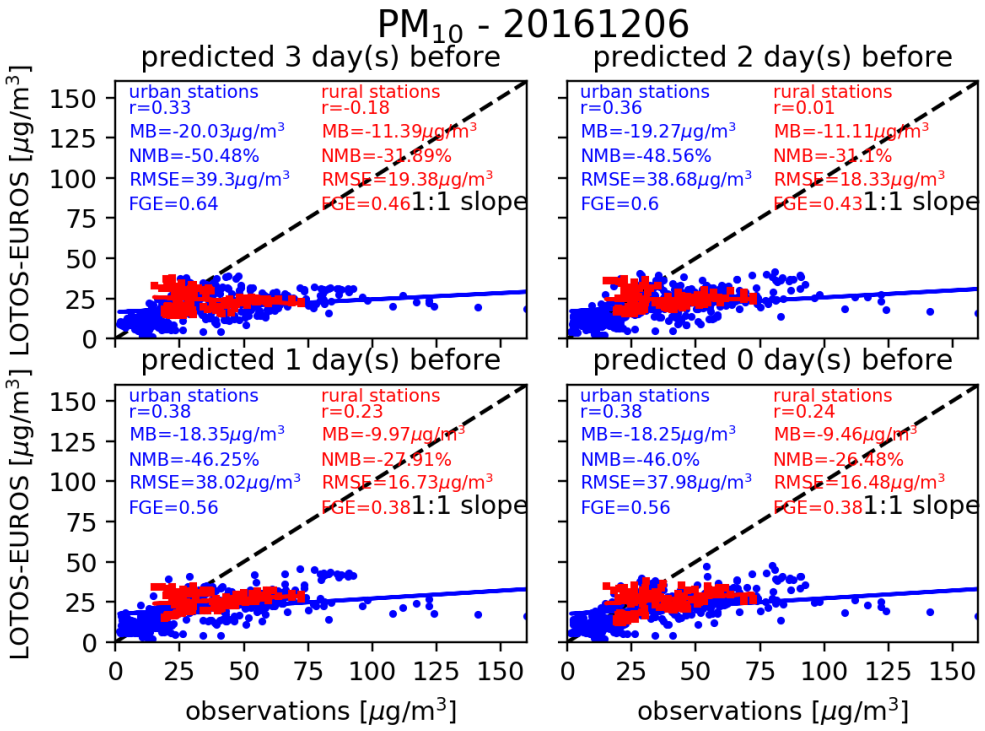


Figure 4: As Fig. 3 for LOTOS-EUROS.

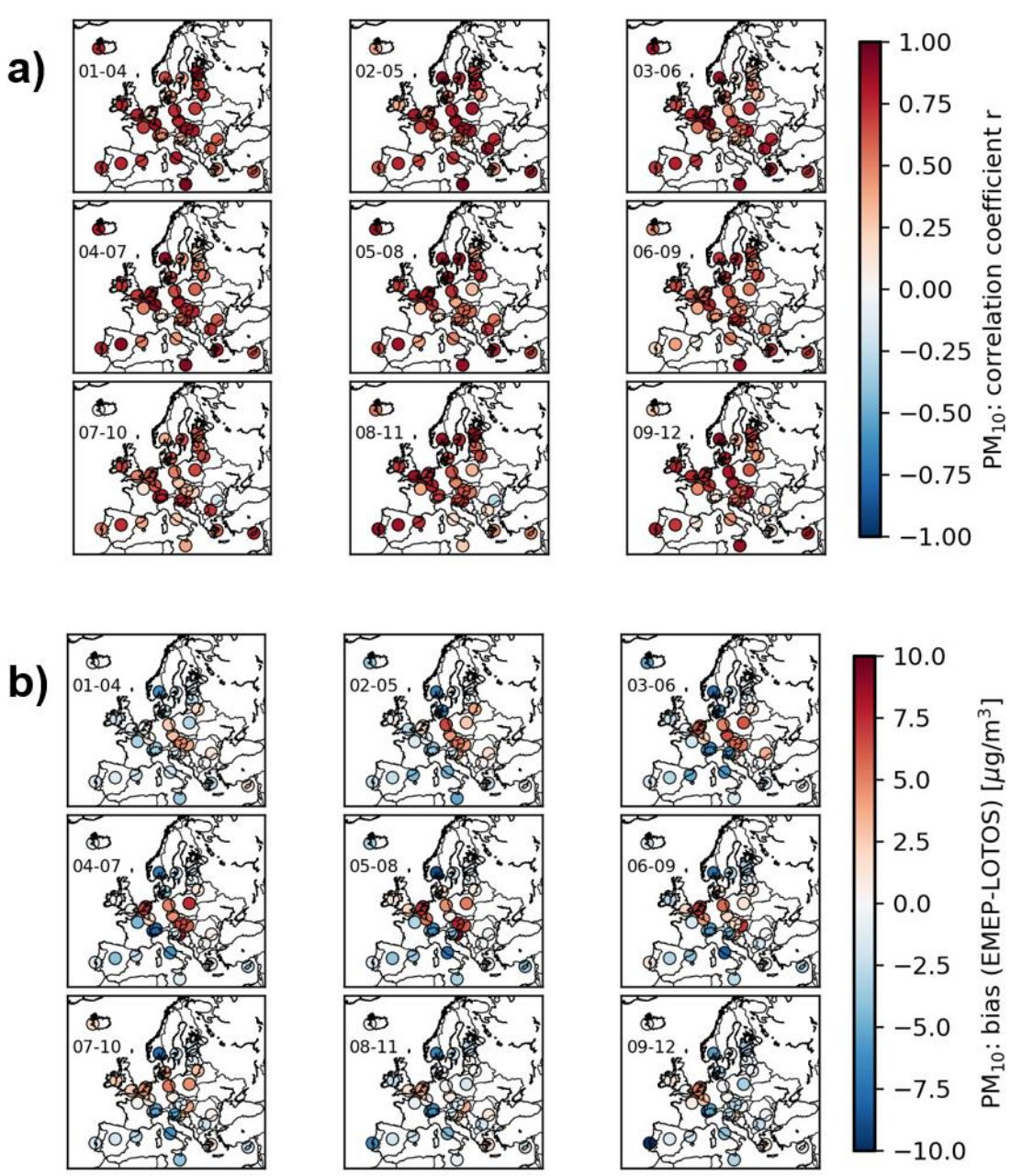


Figure 5: Correlation coefficient (a) and bias (b) in the predicted PM_{10} concentrations between the EMEP model and LOTOS-EUROS over all the studied cities using the 9 grid cells definition for each 4-day forecast (01-04 Dec 2016, 02-05 Dec 2016, 03-06 Dec 2016, 04-07 Dec 2016, 05-08 Dec 2016, 06-09 Dec 2016, 07-10 Dec 2016, 08-11 Dec 2016, 09-12 Dec 2016).

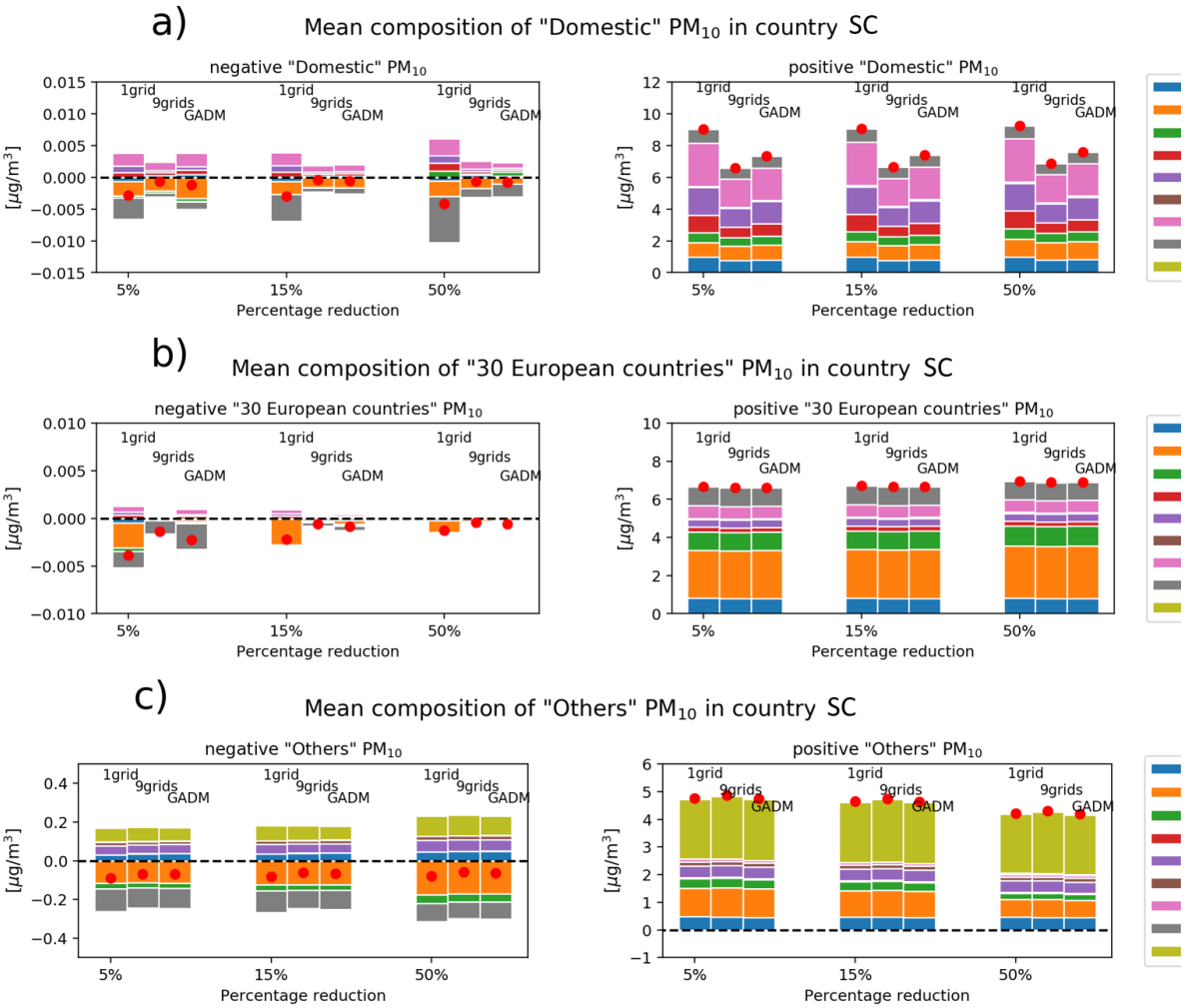


Figure 6: Mean composition of "Domestic" (a), "30 European countries" (b), and "Others" PM_{10} split into a negative concentration (left panel) and a positive concentration (right panel), calculated by the EMEP country SC over the 34 European cities and for each 4-day forecast. The PM_{10} composition is highlighted with the color code. The results for the 3 city definitions (1 grid, 9 grids, GADM) and for the percentage of reduction used in the perturbation EMEP runs (5%, 15%, 50%) are shown. The "Domestic" contribution corresponds to the contribution from the domestic country to the city (e.g. from France to Paris). "30 European countries" corresponds to the other 30 European countries used in the study. "Others" gathers natural sources, the other countries included in the regional domain, the boundary conditions, the ship traffic, the biogenic sources, the aircraft emission and the lightning. The red dot represents the mean PM_{10} concentration.

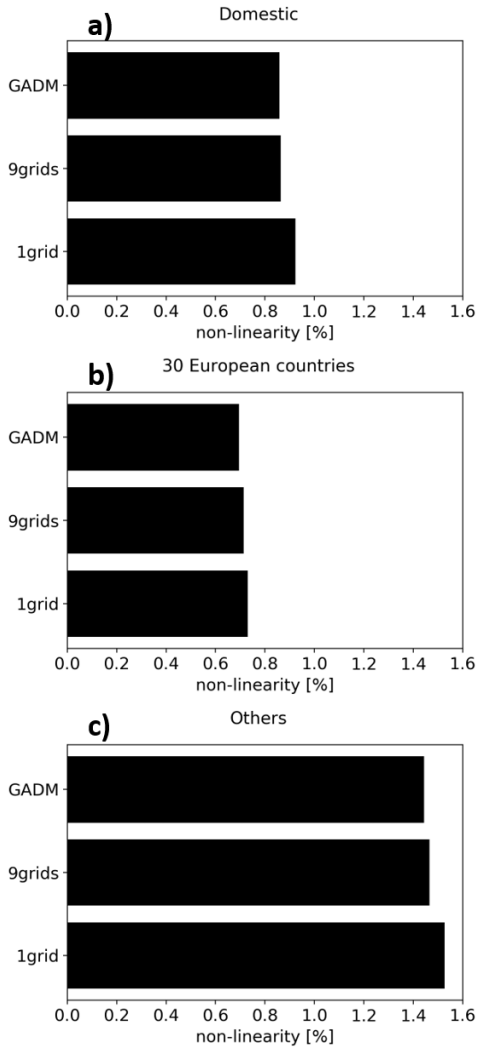


Figure 7: The black horizontal bars show the mean non-linearity calculated for each contribution presented in Figure 6 and for the three city definitions. The non-linearity is calculated for each hourly concentration as the standard deviation of the hourly contribution weighted by the hourly total concentration.

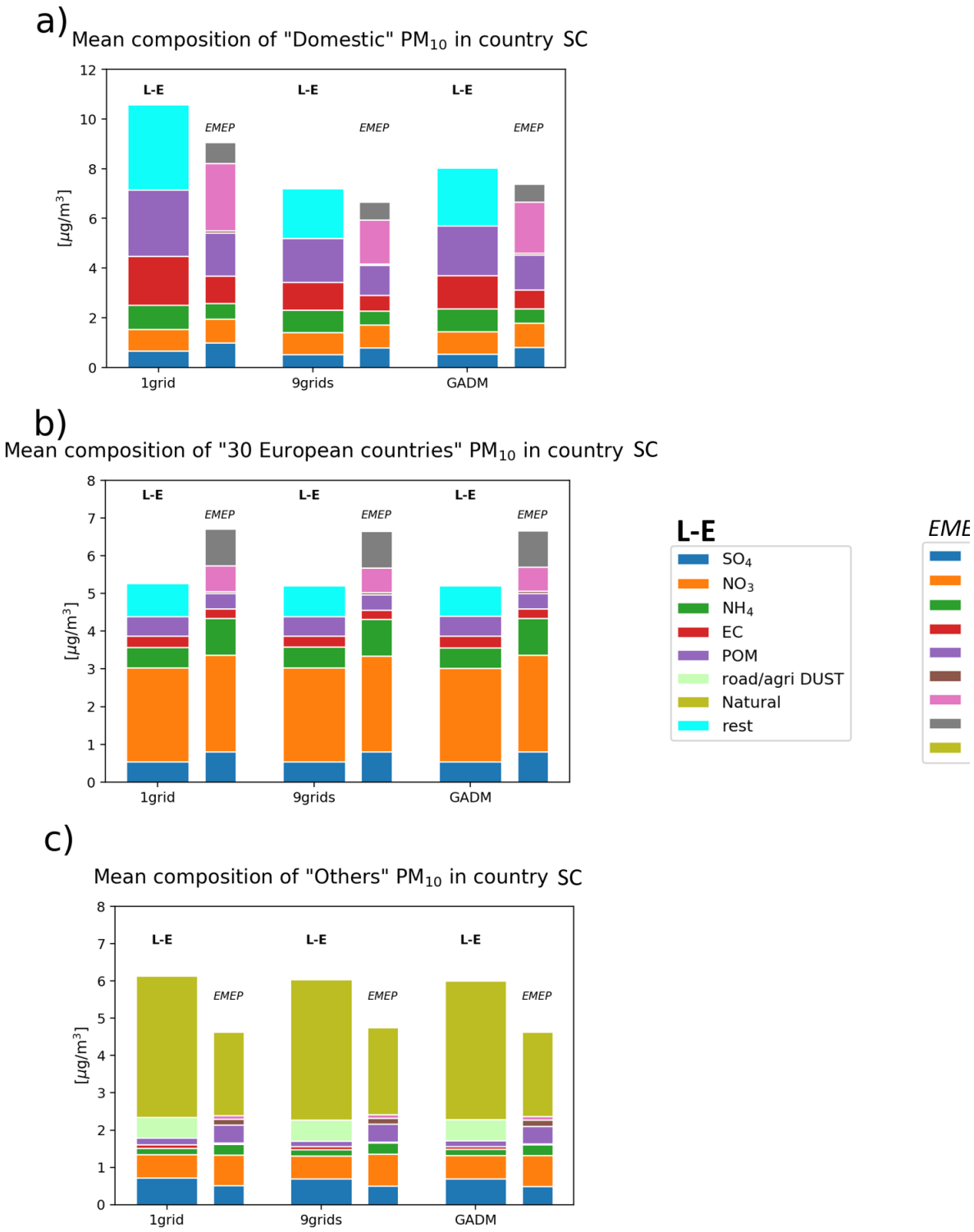


Figure 8: Mean composition of "Domestic" (a), "30 European countries" (b), and "Others" PM_{10} calculated by the LOTOS-EUROS (L-E) country SC over the 34 European cities and for each 4-day forecast. The result from the EMEP country SC, by using a 15% perturbation run has also been added for comparison. The PM_{10} composition is highlighted with the color code. Rest corresponds to the difference between the PM_{10} and the sum of the components listed on the plot. The results for the 3 city definitions (1 grid, 9 grids, GADM) are shown. The "Domestic" contribution corresponds to the contribution from the domestic country to the city (e.g. from France to Paris). "30 European countries" corresponds to the other 30 European countries used in the study. "Others" in the LOTOS-EUROS country SC is slightly different to the EMEP "Others". "Others" in the LOTOS-EUROS country SC gathers natural sources, the other countries included

in the regional domain, the boundary conditions, the dust emitted by the road traffic and agriculture, the ship traffic, the aircraft emission and the lightning.

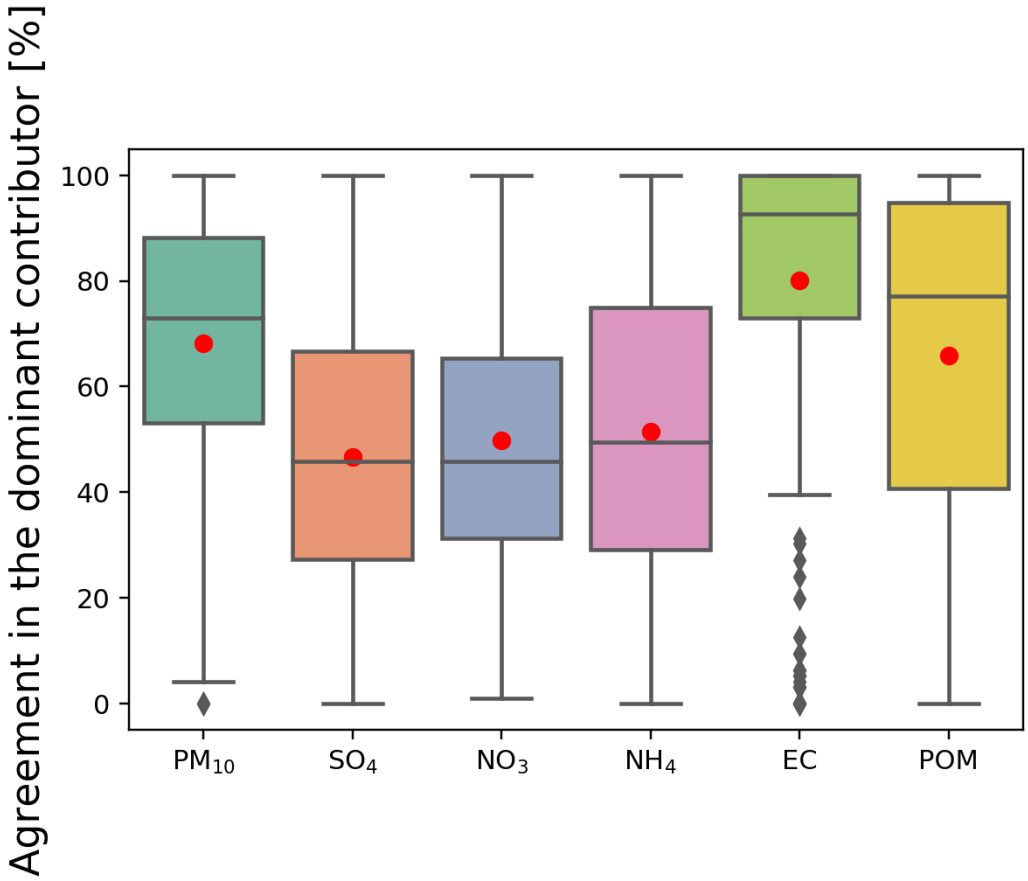


Figure 9: Agreement in the determination of the dominant country contributor for PM_{10}, SO_4, NO_3, NH_4, EC and POM in percent, determined over all the studied cities using the 9 grid cells definition and for all 4-day forecasts. The line that divides the box into two parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers which are represented by grey diamonds. The red dots correspond to the mean of each data set.

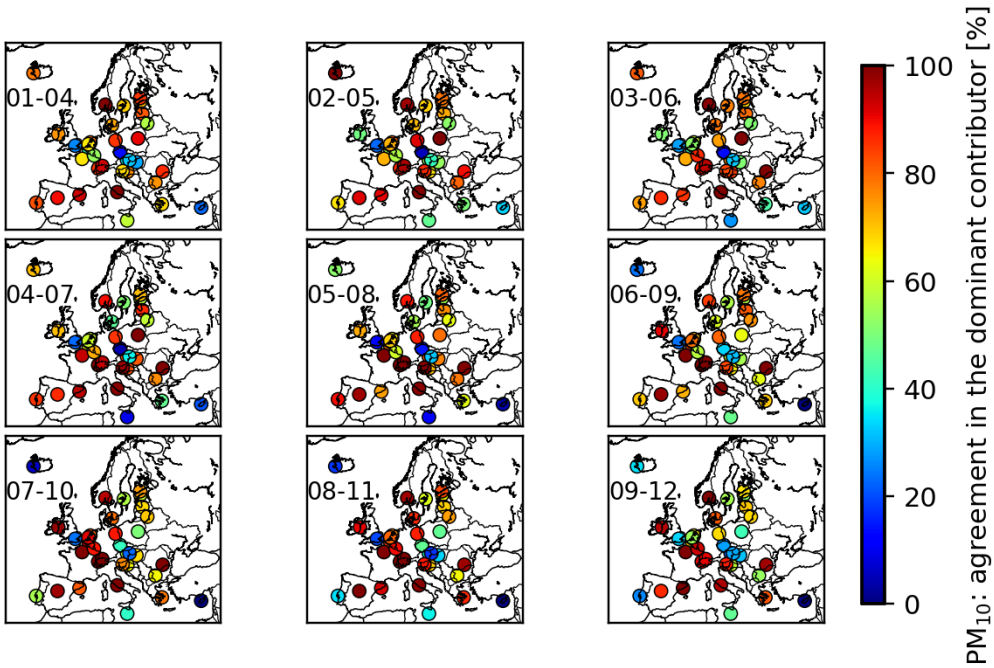


Figure 10: Agreement in the determination of the dominant country contributor for PM_{10} in percent, and for each 4-day forecast (01-04 Dec 2016, 02-05 Dec 2016, 03-06 Dec 2016, 04-07 Dec 2016, 05-08 Dec 2016, 06-09 Dec 2016, 07-10 Dec 2016, 08-11 Dec 2016, 09-12 Dec 2016) over all the cities using the 9 grid cells definition.

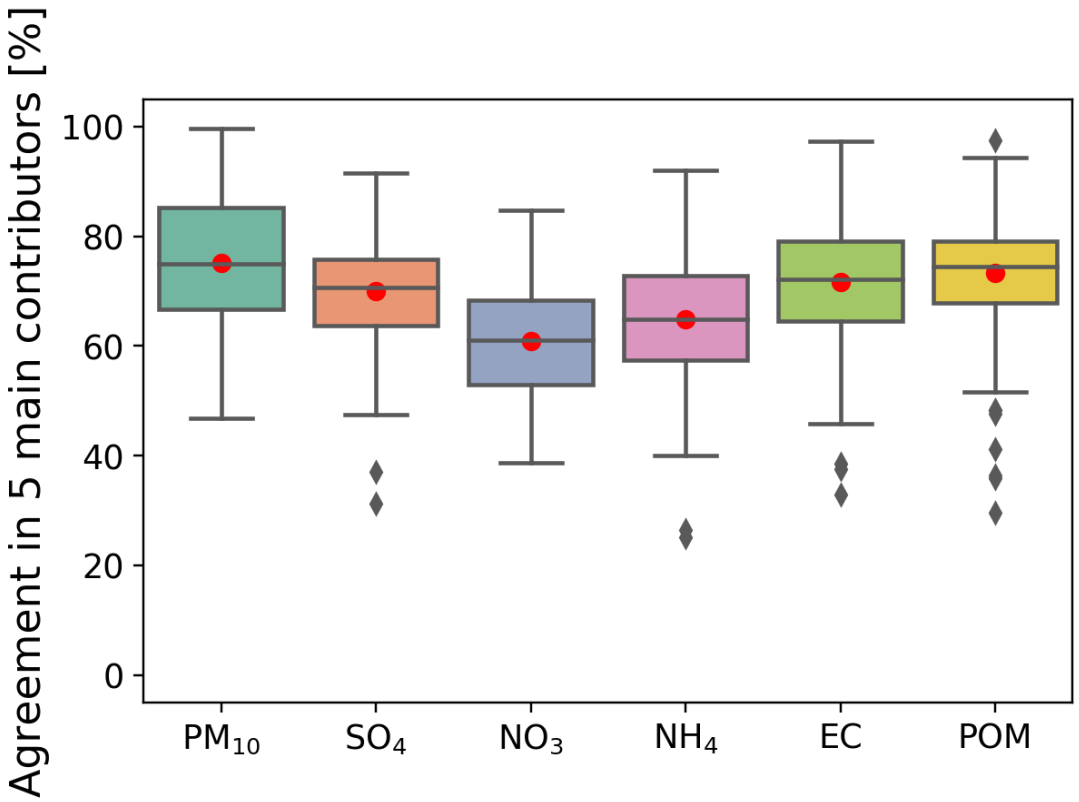


Figure 11: Agreement in the determination of the five main country contributors for PM_{10}, SO_4, NO_3, NH_4, EC and POM in percent, determined over all the studied cities using the 9 grid cells definition and for all 4-day forecasts. The line that divides the box into two parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers which are represented by grey diamonds. The red dots correspond to the mean of each data set.