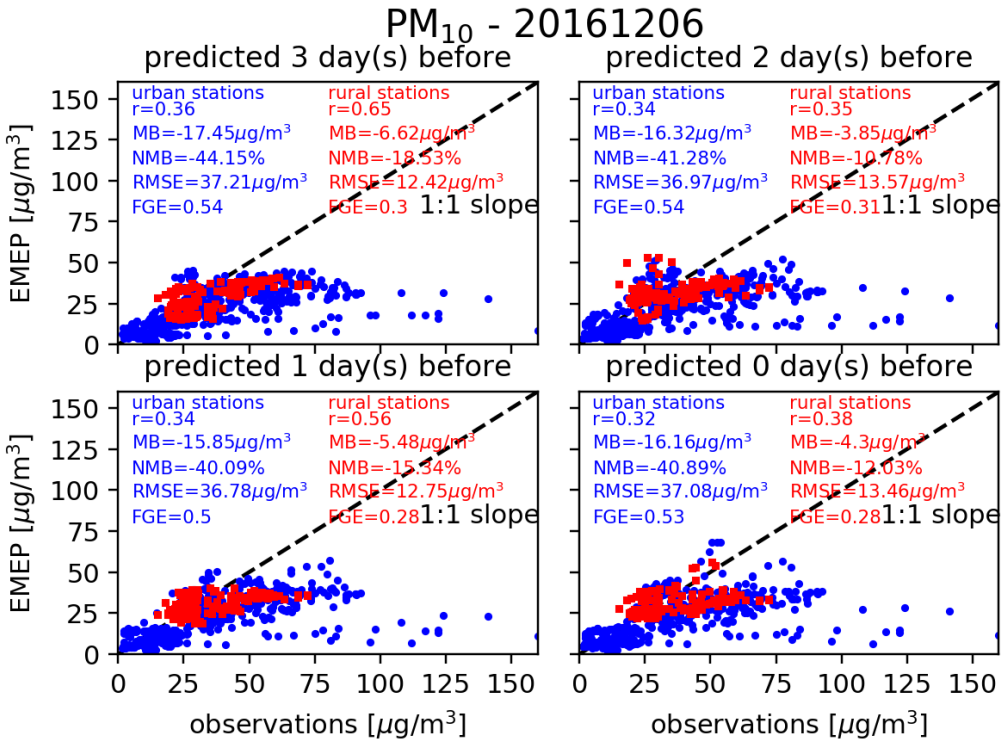
We would like to thank the reviewer 2 for the helpful comments and corrections. We have answered the different points by highlighting our comments in blue.

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

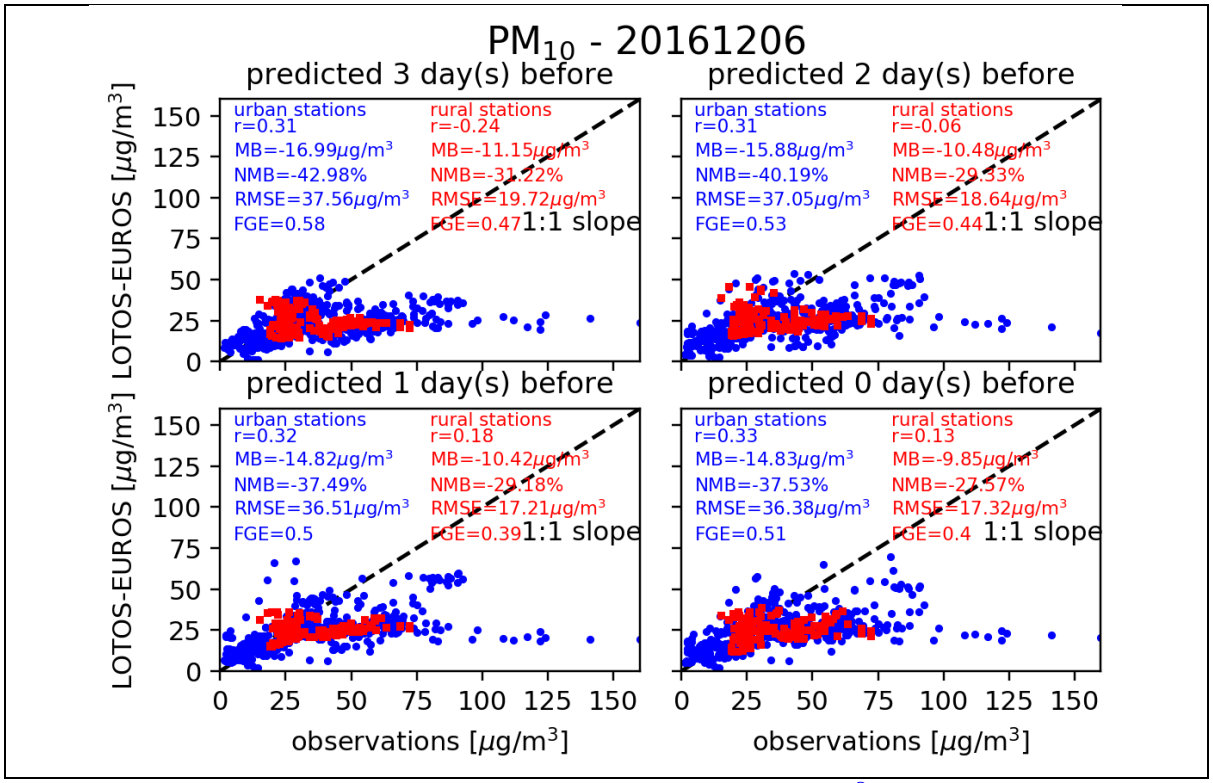


Fig. Scatterplot between the hourly PM_{10} concentrations in µg/m^3 over all the studied cities using the 9 grid cells definition, predicted by the EMEP model (top), LOTOS-EUROS (bottom) on December 06th 2016 and the observations of the urban sites (blue dot) and rural sites (red square). The EMEP predictions 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.

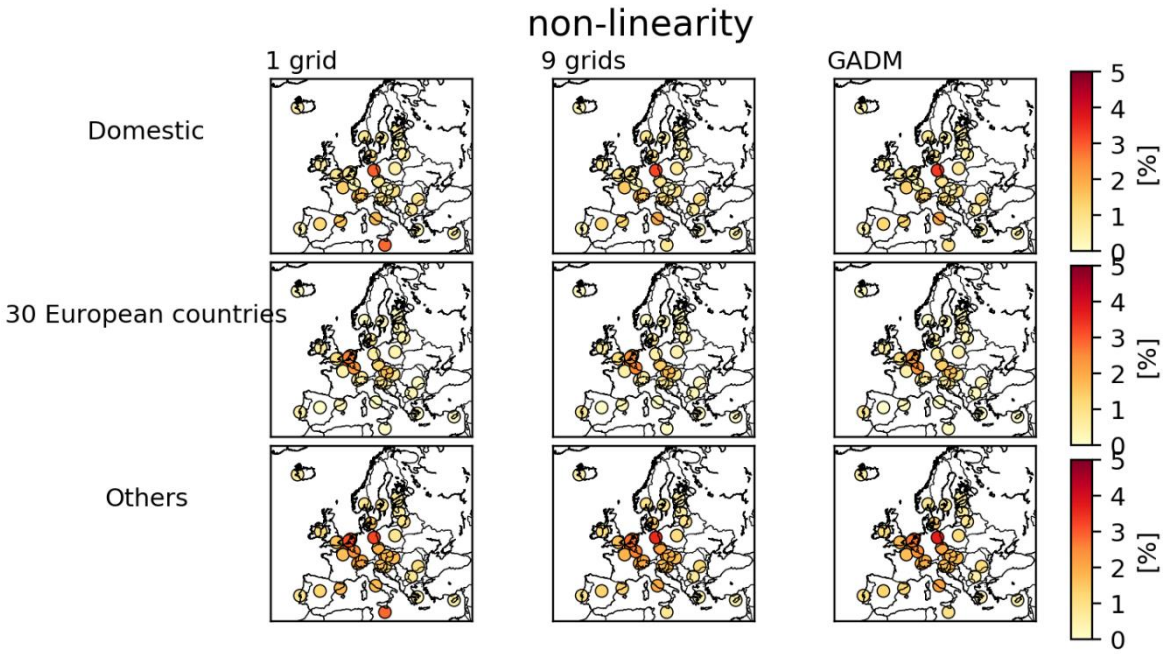


Figure. S10 Mean hourly non-linearity in percent calculated for the "Domestic", "30 European countries" and "Others" contributions, over the 34 European cities and for all 4-day forecasts (i.e. from 01-04 Dec to 09-12 Dec 2016). The non-linearity is presented for the cities defined by 1 grid cell (left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

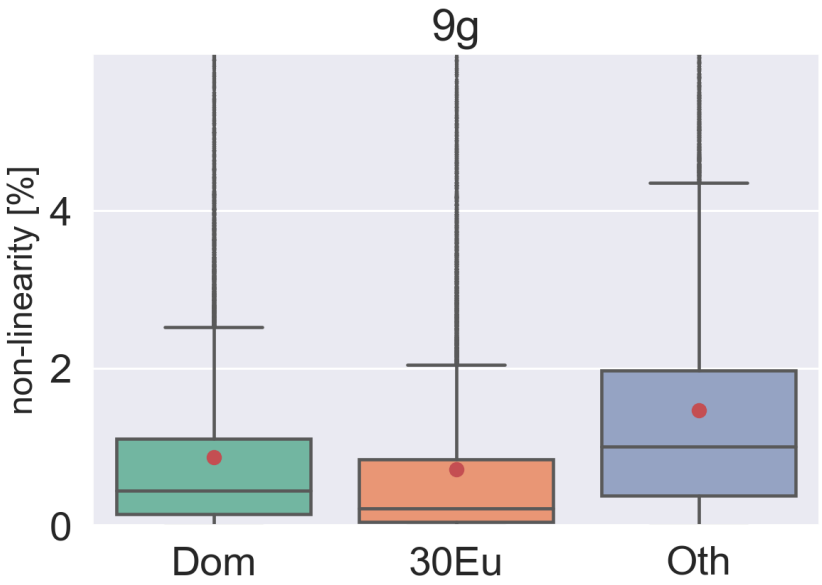


Figure: Distribution of the non-linearity for the "Domestic", "30 European countries" and "Others" contribution. The line that divides the boxes into two parts represents the median of the data. The end of the boxes shows the upper and lower quartiles. The extreme lines show the highest and lowest value 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):

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

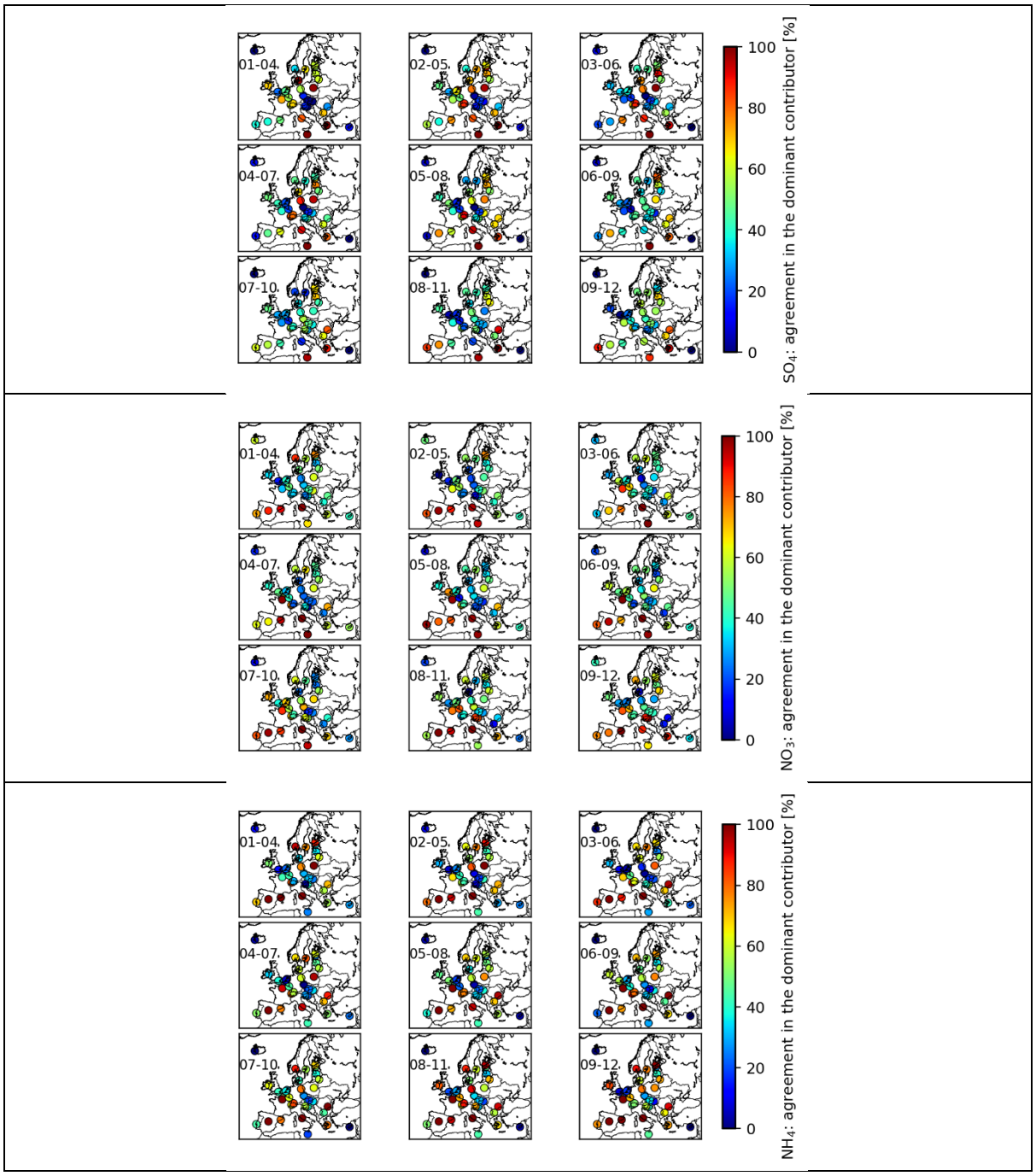
n corresponds to the number of perturbations used (n=3), Ccontrib is the hourly PM_{10} concentration for a specific contribution ("Domestic" or "30 European countries" or "Others") and Ctot is the hourly PM_{10} concentration."

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:



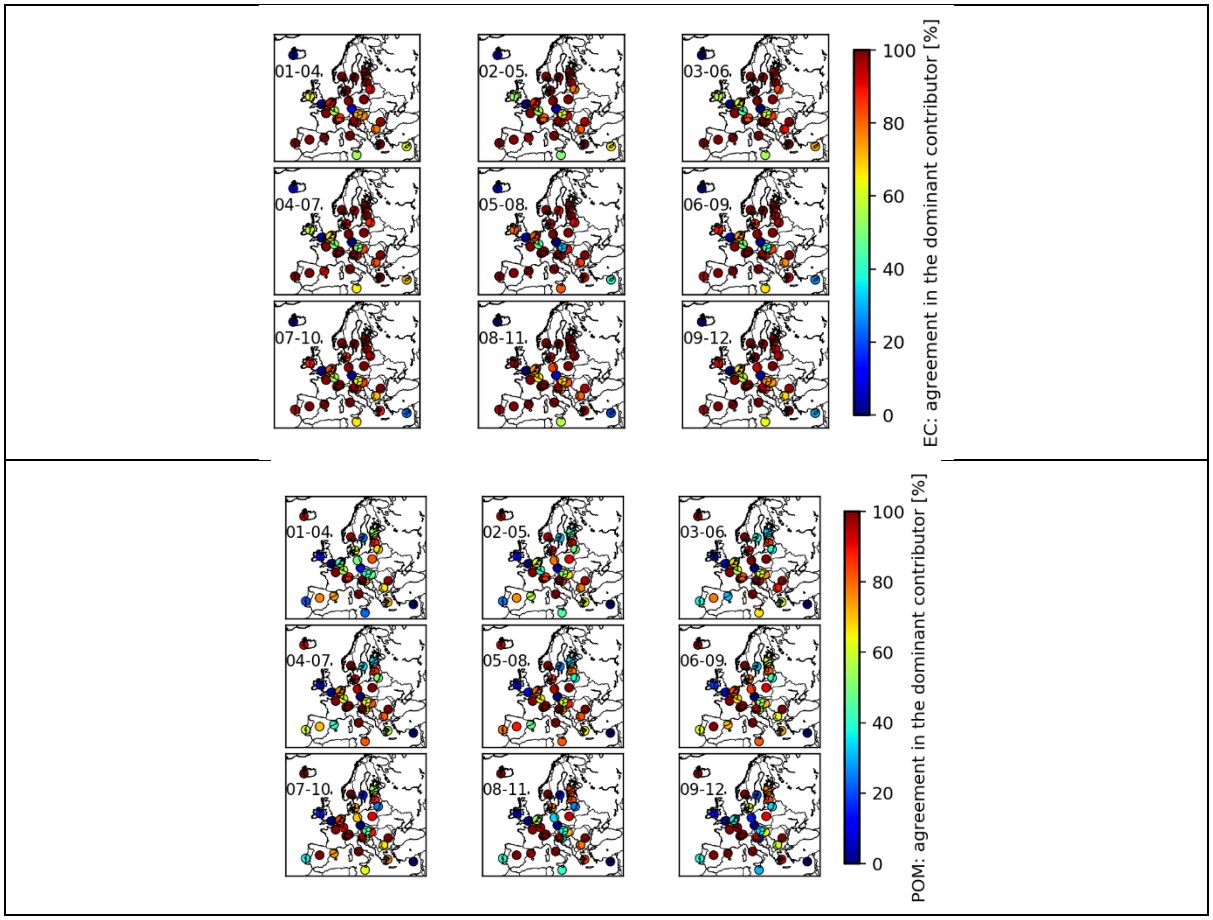


Fig. Agreement in the determination of the dominant country contributor for SO_4, NO_3, NH_4, EC and POM in percent, and for each 4-day forecast (01-04 Dec 2016, 02-05 Dec 2016, 03-06 Dec 2016, 04-07 Dec 2016, 05-08 Dec 2016, 06-09 Dec 2016, 07-10 Dec 2016, 08-11 Dec 2016, 09-12 Dec 2016) over all the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