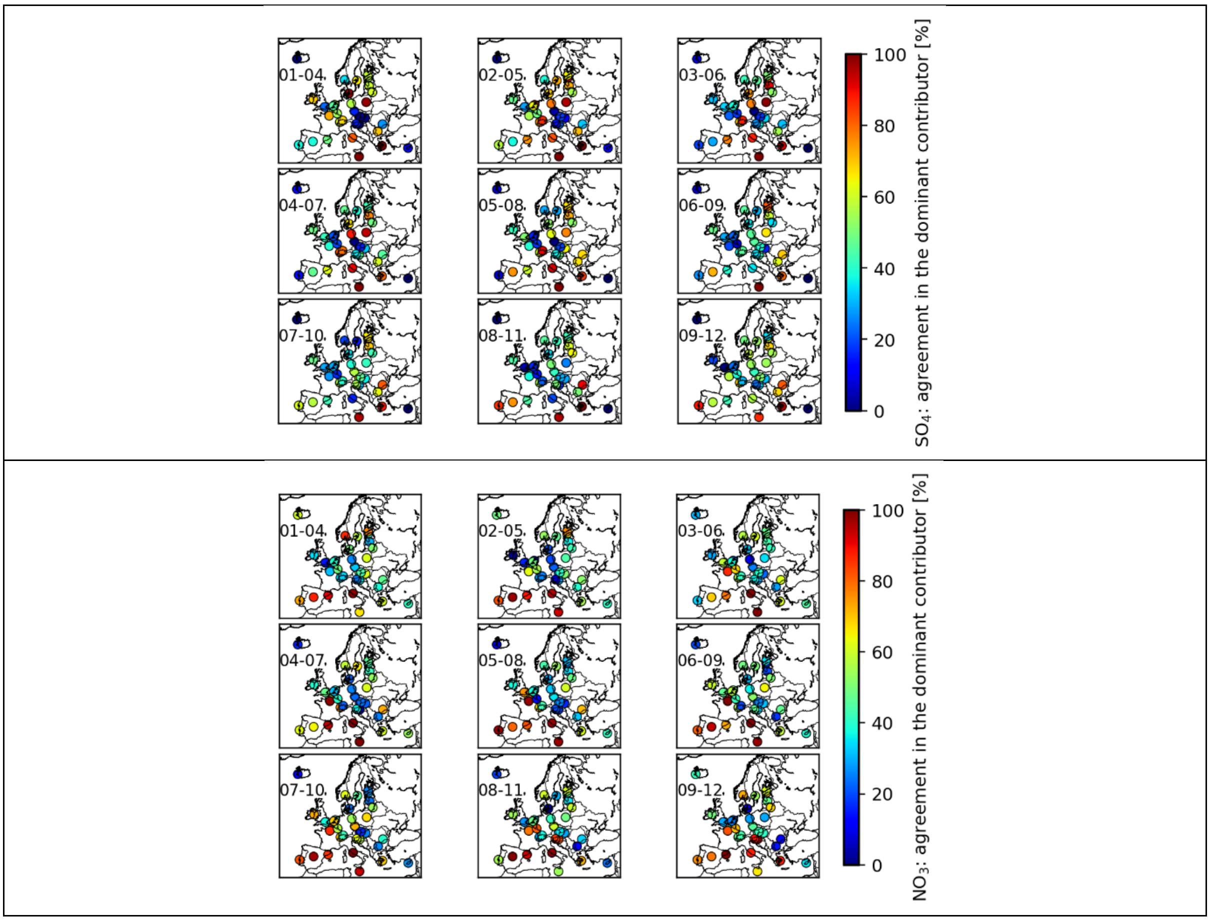
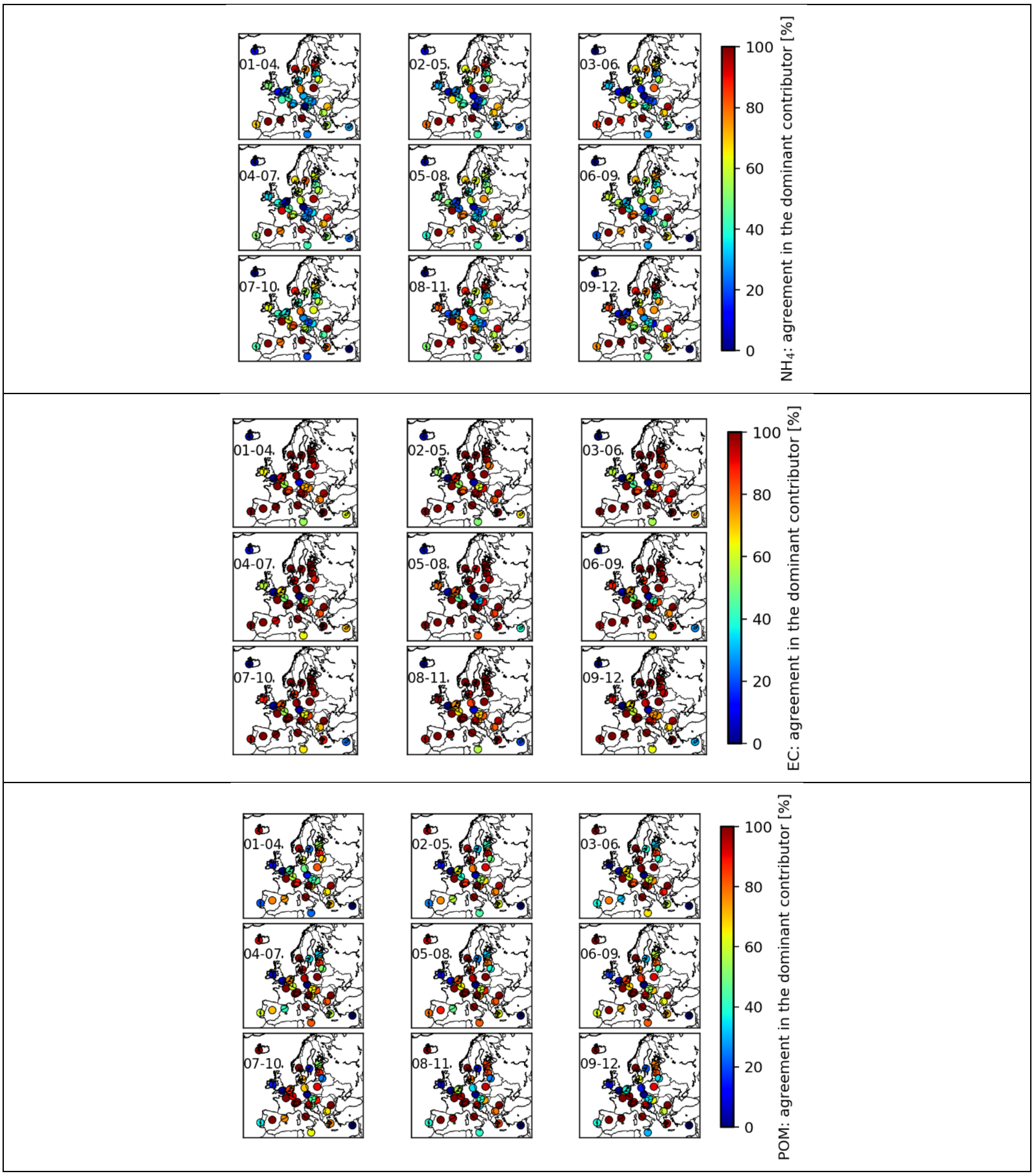
We would like to thank Philippe Thunis for his helpful comments. We have tried to answer all his remarks below.
We also apologize for our missing points in our analysis and we are very grateful for all the details provided by Philippe Thunis.
We also would like to remind that we aimed to provide in this 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):

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

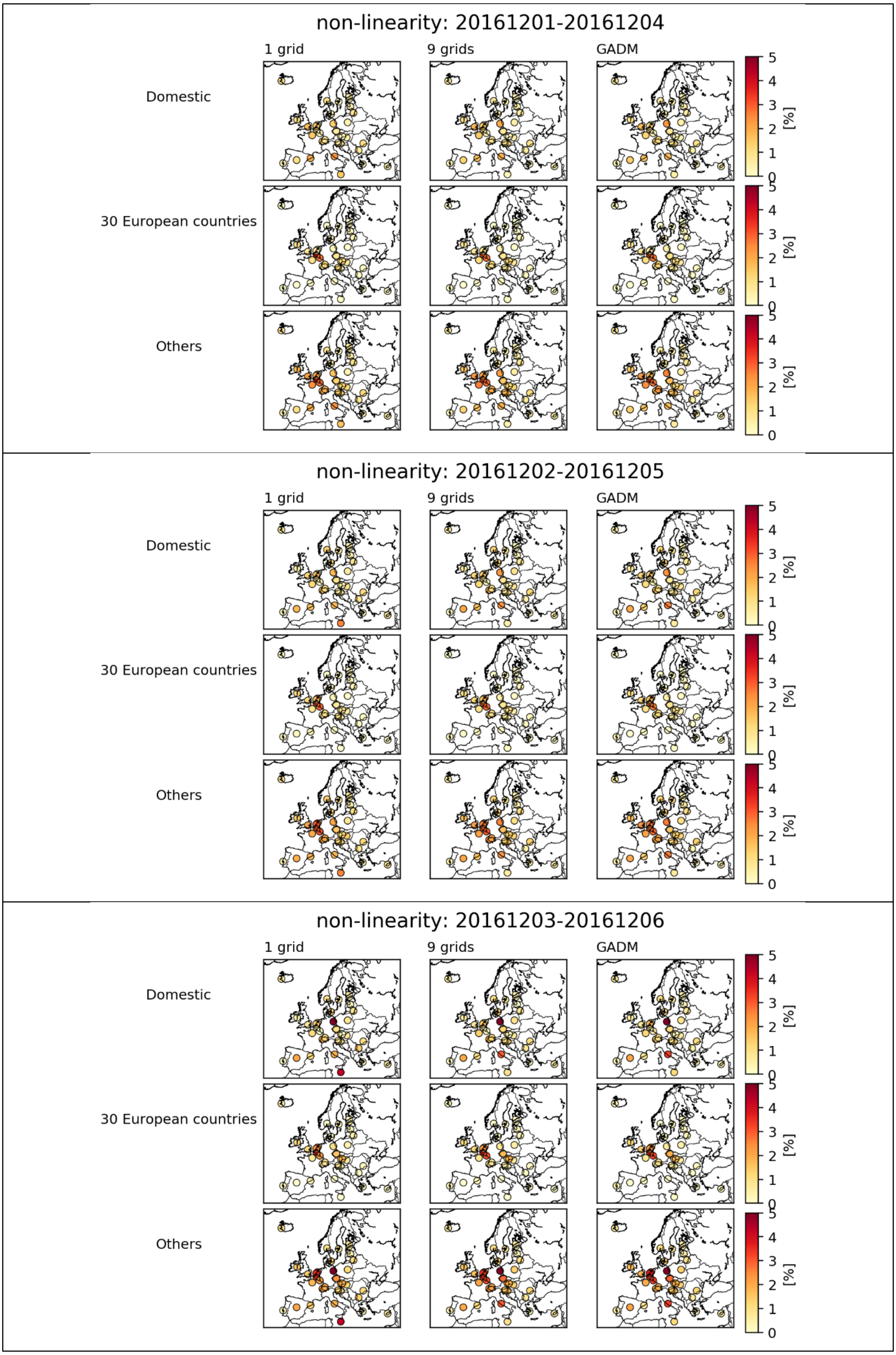
n corresponds to the number of perturbations used (n=3), Ccontrib is the hourly PM_{10} concentration for a specific contribution ("Domestic" or "30 European countries" or "Others") and Ctot is the hourly PM_{10} concentration."

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

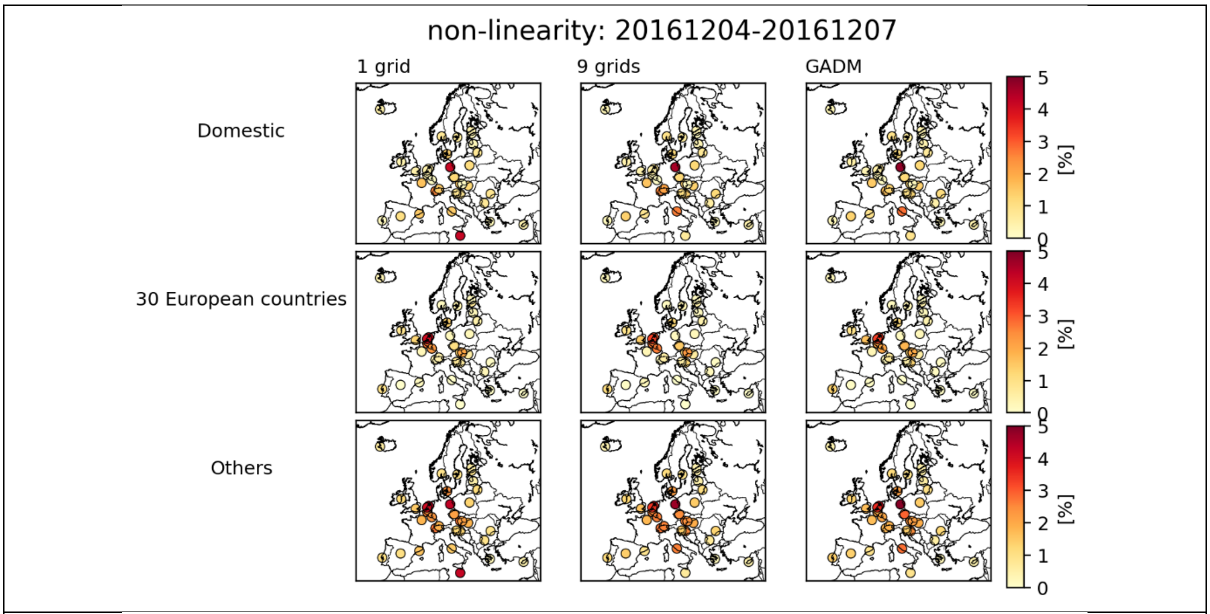
It is also right that we did not study in this work the non-linearity between the different PM_{10} precursors. As explained in the paper, to perform a test related to the non-linearity in the reduction of each individual precursor will be too time consuming.
Without to count the 9 reference runs corresponding 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.

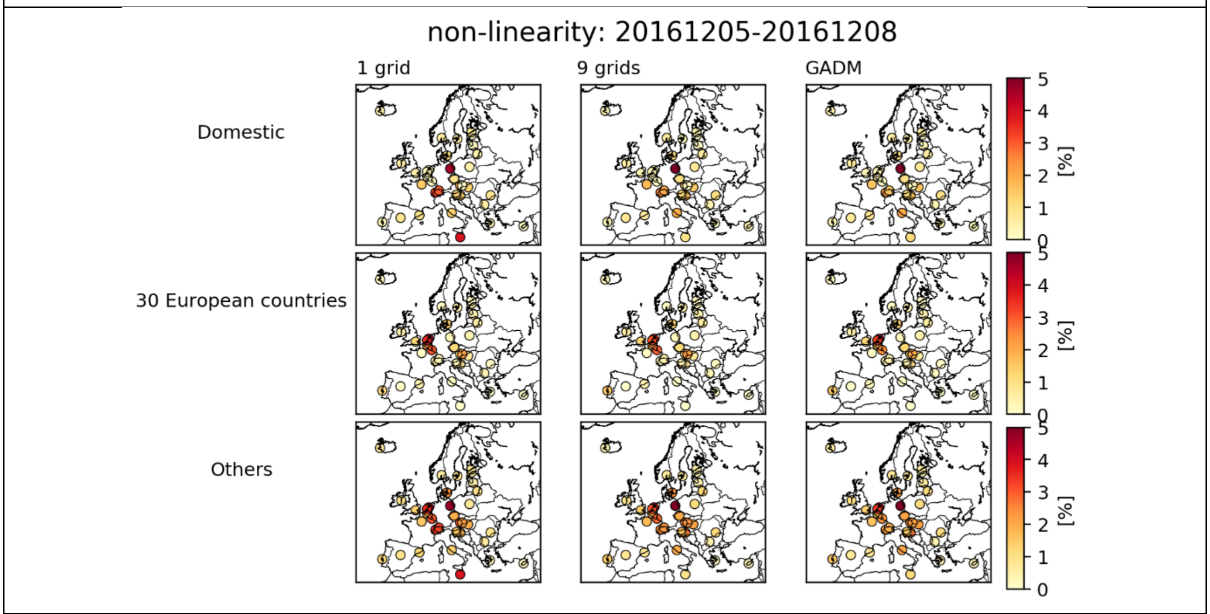
We have added this sentence in section 4.1.2:
"In total, 847 4-day runs have been performed in this work (9 reference runs, and 9 dates × 31 countries × 3 perturbations runs)."
And the following information (in bold) in Section 5.1:
"This also shows that the responses to perturbation runs are robust, **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."**

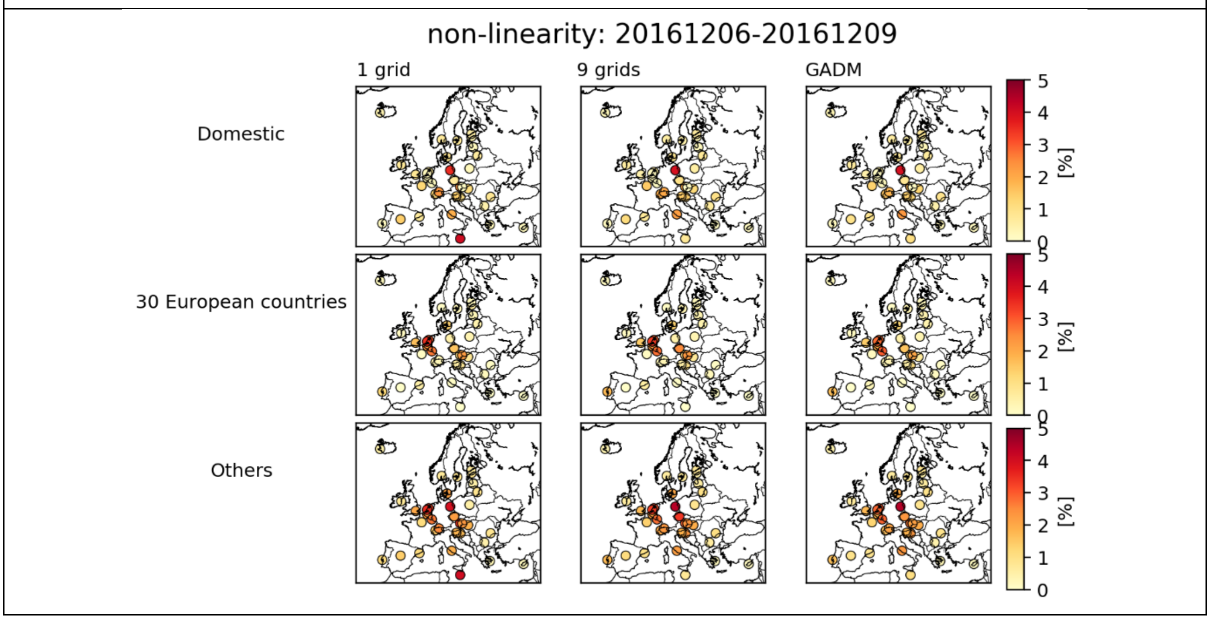
The non-linearity presented below, shows the variability of this non-linearity for each date over the different cities. The non-linearity remains low.

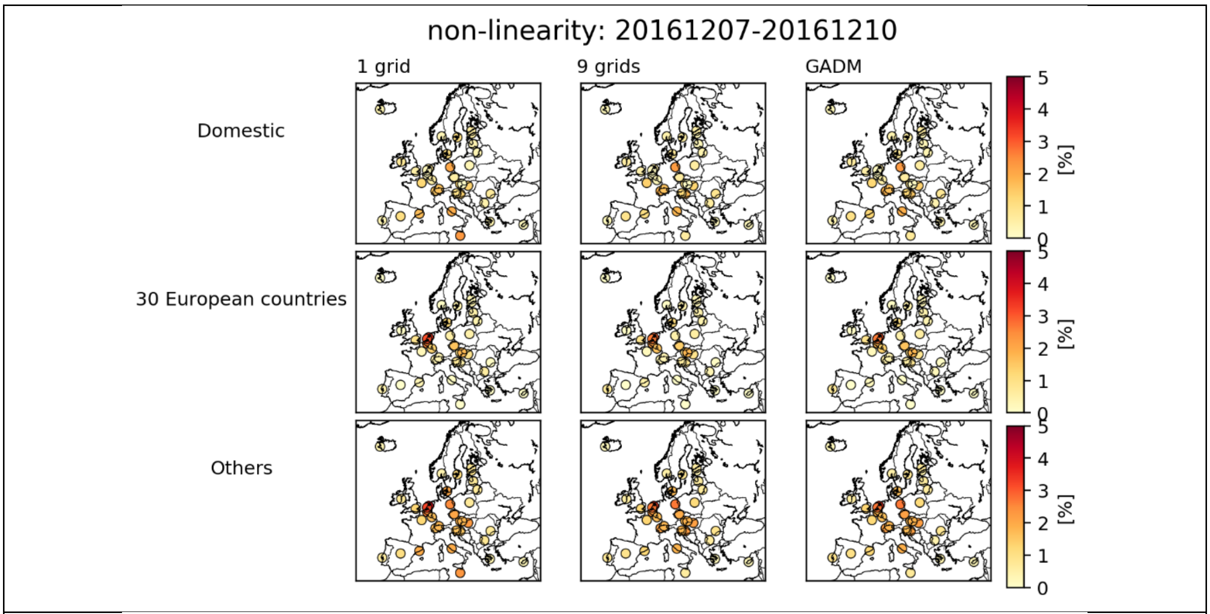


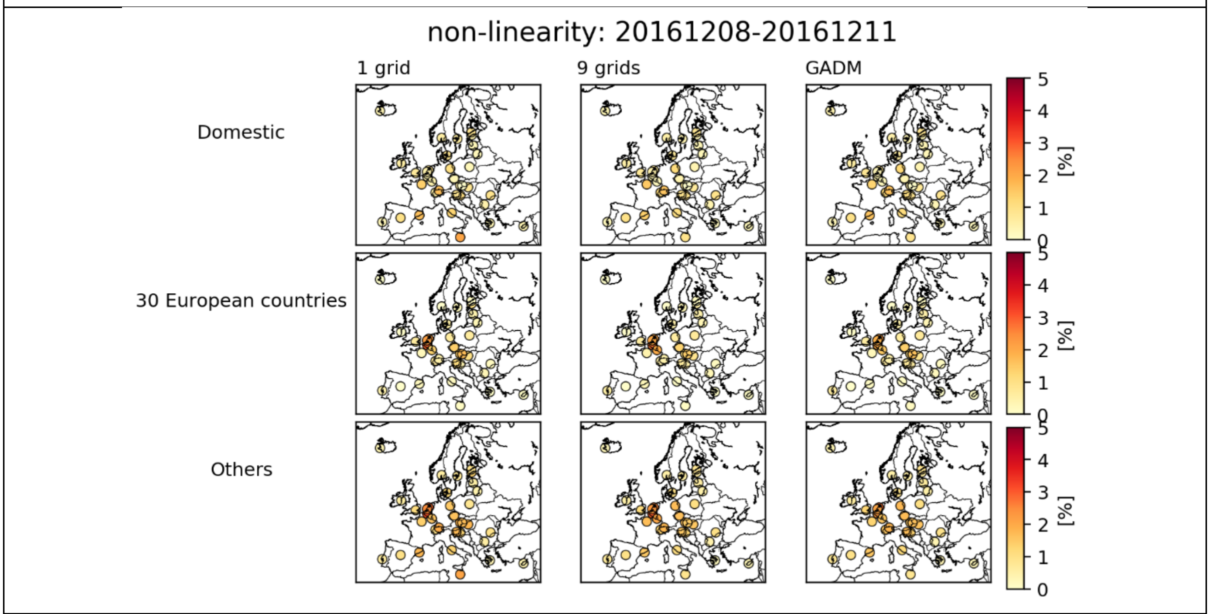
non-linearity: 20161201-20161204
non-linearity: 20161202-20161205
non-linearity: 20161203-20161206
1 grid
9 grids
GADM
Domestic
30 European countries
Others
[%]
0
1
2
3
4
5


non-linearity: 20161204-20161207
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]


non-linearity: 20161205-20161208
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]


non-linearity: 20161206-20161209
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]


non-linearity: 20161207-20161210
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]


non-linearity: 20161208-20161211
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]

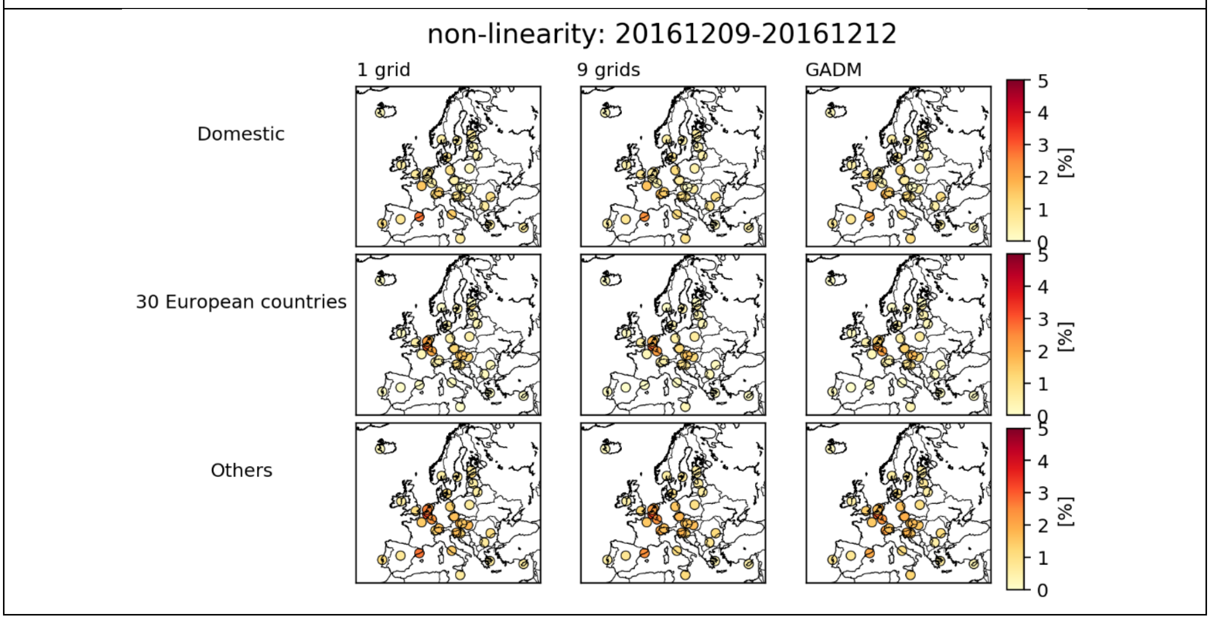

non-linearity: 20161209-20161212
1 grid
9 grids
GADM
Domestic
30 European countries
Others
0 1 2 3 4 5
[%]

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:

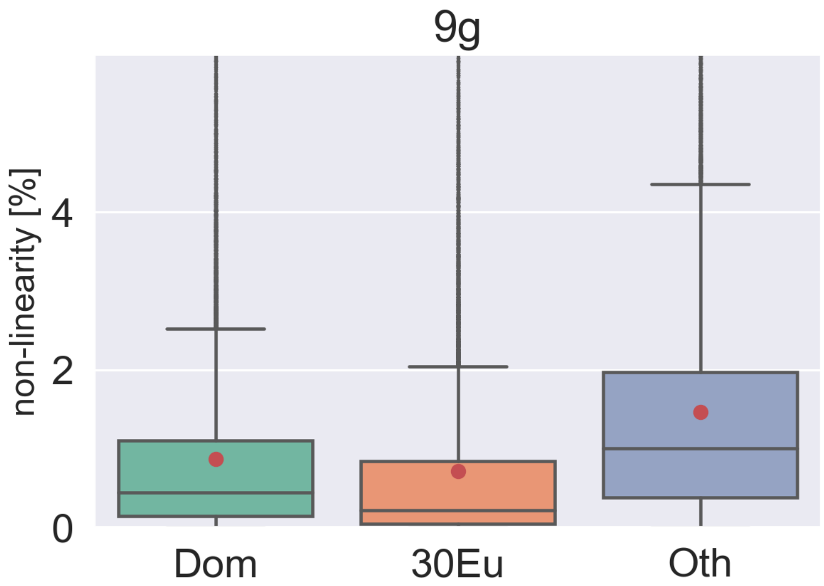


Figure: Distribution of the non-linearity for the "Domestic", "30 European countries" and "Others" contribution. The line that divides the boxes into two parts represents the median of the data. The end of the boxes shows the upper and lower quartiles. The extreme lines show the highest and lowest value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

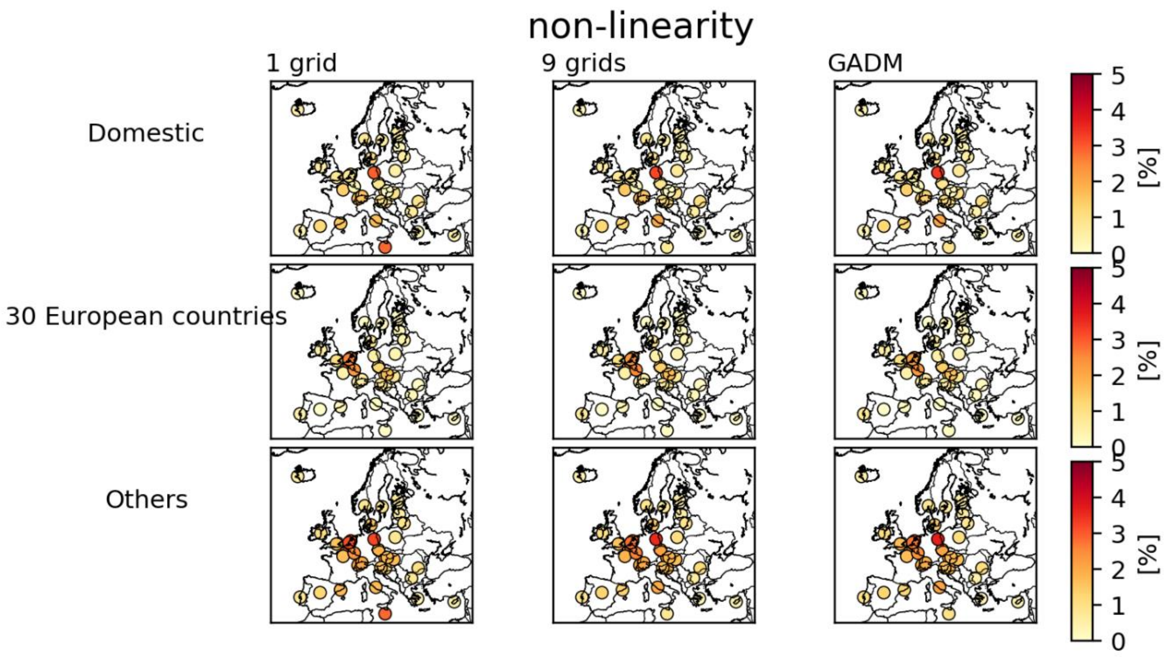


Figure. S10 Mean hourly non-linearity in percent calculated for the "Domestic", "30 European countries" and "Others" contributions, over the 34 European cities and for all 4-day forecasts (i.e. from 01-04 Dec to 09-12 Dec 2016). The non-linearity is presented for the cities defined by 1 grid cell (left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