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.