Author response to the referee comments to the paper by Sofiev et al:

Multi-model ensemble simulations of olive pollen distribution in Europe in 2014

First of all, we would like to thank both reviewers for their efforts and comments, which helped us improving the paper. Below, we address them one by one (the comments themselves are included in *italic*).

A general remark.

We would like to emphasize that this is the first-ever experiment of numerical modelling of the European-scale olive pollen dispersion. It was stated in the introduction of the original paper and stressed more in the revised paper, starting from the slightly modified title. Therefore, revealing the issues with the models performance and reporting them to the scientific community was one of the goals of the paper. They were reported upfront and discussed already in the initial paper and addressed in more details in the revised manuscript. In particular, we added a new section 5.4, which summarizes these challenges.

Responses to individual comments of the referees

Referee 1, Slawomir Potempsky

The ensemble setup is based on 6 models, 5 of them using the same meteorological data (ECMWF IFS). This means that this ensemble concerns, in principle, dispersion models – meteorological variety is practically disregarded. From statistics point of view such number of the models can be not enough in real application, but can be treated as a first step. It seems that adding models driven by different meteo data could be important in case of building operational system of this type modelling.

We agree that the current ensemble is at the lower edge of reasonable sizes of such ensembles. Within CAMS, work is going on to increase the number of models up to 10. Possibility to expand the ensemble using perturbed-meteorology forecasts certainly deserves consideration. A sentence stressing the same-meteo-same-source limitation was added in section 5.2.1 and the summary.

The authors proposed also an "optimized ensemble" model basing on properly chosen linear combination. This model has shown good skills although the choice of some parameters (like alpha, beta) seems to be rather art than to be based on pure mathematical approach.

We have tested several values for regularization strength and the selected values have come out of these tests. The limited size of the datasets and short duration of the season constrained the possibilities for data-driven selection of the regularization strength.

The presented results have shown some capabilities of the constructed ensemble but also indicated problems like the shift of the whole season. This needs further research and is strictly related to meteorological forecast.

We agree, the corresponding sentences are included in section 5.2.1 and newly created sub-section 5.4. Future challenges

1. Paragraph at lines 186-197

This paragraph needs to be revised in order to precisely define the quantities in the formulas (1) and (2) . In the first formula the meaning of t is ambiguous: a_m(t) means coefficients depending on the interval t while c_m(….,t) means concentration at time t. I think it's better to use parameter "tau" for the period (as in the second formula) and t for time point. Thus c_opt should depend on both tau and t and in the second formula A should depend on tau. There are no definitions of a_0 (bias depending on ?) in formula (1) and c_0 (observations I guess) in formula (2). Functional J should be rather a function of time t than the period tau (unless the average over time is considered – but even then the average period time can differ from the period tau used for the analysis).

Corrected

2. Definition of SPI – Seasonal Pollen Index – is this follows the description given at the beginning of section 3.3 ? It would be good to know whether this index is related to exceeding some threshold limit used for allergy risk.

Clarification is added in sections 3.3. What concerns the relation to an allergy-relevant threshold, unfortunately, there is no unified value for it. Extensive discussions have this-far led to conclusion that it is person-specific and chances for generalization are thin. More than that, experience of

HIALINE project showed large variability of the amount of allergen content in the pollen grains, which creates further problems in automatic projection of the pollen forecasts to allergy. References are added to introduction and section 4.1.

3. Relating to the previous comments what seems important is to show the agreement of the model predictions with the observations basing on the exceedance of threshold limit (as used, for example, in allergy forecast). Fig. 4 shows hourly olive pollen concentration – this picture could be accompanied by the other presenting agreement on threshold level with the observation.

Unfortunately, as it follows from the above, such threshold does not exist. We have calculated parameters, such as odds ratio etc in previous papers for birch but the thresholds were each time picked as "expert guess" and criticized in follow-up discussions. Therefore, for the current paper, we paid concentrated on quality of representation of the season propagation.

4. As accumulation heat is one of the crucial parameter the question arises whether it can be somehow taken from observations and data assimilation technique can be applied to include such information into ensemble modelling.

Temperature is among the most-heavily assimilated meteorological parameters, so bringing it here once again looks like a double-counting. The problem definitely exists but may have more dimensions. In the revised paper, we stressed that it is the combination of the heat-sum accumulation procedure, the flowering threshold, and characteristics of meteorological data that comprise the quality of the season forecast.

5. A delicate matter is the calculation of the source term. Thus the question arises whether uncertainty of the source term could be a part of modelling i.e. the whether the models' simulation could be also performed with perturbations of source term. This, with no doubts, would be time consuming, nevertheless having an ensemble system with such capability would be an added value. This can be one of the possibility of further development of the system.

We agree, thank you. The possibility of perturbing also the source term characteristics is now mentioned in the discussion.

Technical remarks

1. Quality of some figures could be improved:

a) On Fig. 3 observation points are not well visible – some example locations could be shown on separate graph – for example the ones with high values.

b) Zooming maps on Figs 5 and 6 can improve quality.

c) parts of Figs 10 and 11 are not well visible.

The figure 3 has been split to two and zooming sub-picture added. The maps were zoomed-in and their color coding improved to better visibility

Anonymous Referee 3

2. The paper presents an existing concept – the MACC pollen dispersion modelling ensemble published by Sofiev et al (2015). As such the idea and the concept has already been published. The ensemble is applied on a new pollen type.

Not quite: this paper, indeed building on the existing CAMS ensemble principles, also significantly develops it in several directions. First of all, this is the first numerical simulations for olives made at European level. Secondly, we expanded the ensemble technology of CAMS by constructing the first fusion model for pollen forecasts – and demonstrated its superiority over the simpler ensemble techniques.

The model manuscript use observations, but there is no station list and no accurate numerical or accurate description of observations.

We have created the full list of stations and reported the basic statistical metrics for all models. Due to its size, the table is put into supplementary material.

3. There are substantial conclusions in the manuscript. However these conclusions do not appear to be founded with data presented in the study. One example is the systematic bias in surface temperatures as a cause of model uncertainties. However the manuscript does not contain data (Figures or tables) concerning surface temperatures.

There seems to be a misunderstanding: our conclusions do not have any sentence regarding the temperature bias of the IFS meteo model. The only reverberation on the quality of this parameter refers to section 5.1 (second paragraph) but that one discussed WRF rather than the core meteo data of IFS. We removed that sentence and reformulated the one in the first paragraph to avoid confusion and to stress that the problem originates from the inconsistency between the features of the temperature forecasts and way they are treated by the source term.

Another example is the conclusion that the model calculations represent large scale transport fairly well. However the results do not seem to agree with previous observations of olive pollen in Europe (Sofiev and Bergmann, 2013). This conclusion there needs better support in the manuscript. The maps in figure 3 show that both the ensemble mean and ensemble median calculates a relevant seasonal pollen index in Northern France, UK, Germany, Poland. For individual models this is extended even to Norway. To my knowledge, olive pollen are rarely or have never been reported from those regions. This shows that the ensemble overestimate the large scale atmospheric transport. The authors clearly write that the sources in France, presented in the source map, are unrealistic low. Adding the missing sources in France must be expected to increase the calculated pollen index in nearby regions such as Germany, UK, Poland, Netherlands etc even more. How much is naturally not known. Nevertheless, the arguments described above, disagrees with the conclusion that the ensemble represent large scale transport fairly well by using the available material in the manuscript.

This is indeed a very relevant discussion and the revised paper has more support for the claim.

However, one limitation has to be kept in mind: The SPI below several tens of pollen day m^{-3} is generally irrelevant from allergy point of view and is very inaccurately observed. The Hirst trap is unreliable up to concentrations as high as 5-10 pollen m^{-3} – such days should be disregarded (Galan et al. 2013; Buters et al. 2012; Buters et al. 2015). In Central- and Northern-European countries, olive pollens are not counted / reported due to their low contribution to the allergy level, high uncertainty and noticeable costs of inflating the list of reported pollens. This is why we could not include their reports and the evaluation of the transport events inevitably rely on the same network of stations. However, looking at northern Spain, it seems (just qualitatively) that the decay of modelled total load is as fast as or even faster than that in the observations. Since the French level is much less than that of Spain, there seems to be no evidence of over-estimation of the transport distance.

4. The scientific methods and assumptions are generally both valid and clearly outlined? However, when they filter the observational record, then they exclude stations with low amounts of pollen. The threshold is 25 pollen day m-3. This is not appropriate. Stations with both high and low numbers should be included in the evaluation of the model simulations.

We respectfully disagree. As stated above, Hirst trap data are very uncertain for low pollen concentrations. The uncertainty of daily values is generally considered to be around 5-10 pollen m^{-3}. Therefore, our current filtering threshold is actually very soft.

5. The results are not sufficient to support the interpretations and conclusions? This is partly caused by the design of figures. As an example, Figure 3 shows simulated (maps) and observed (dots) pollen index. The maps are good for a broad picture, but the accuracy at the sites cannot be assessed with the maps. Ideally this should be combined with a table that demonstrate model results vs observations. Figure 5 shows maps with observed start dates and the corresponding calculations. The chosen color scale makes it almost impossible to see any variations. A difference between 114 and 126 is from on type of green to another, but cover almost 14 days. This means that differences and agreements between observations cannot be assessed to an accuracy of less than 14 days. A table would make this much more clear. Same arguments on accuracy assessments are relevant for Figure 6. This would also have been better in a table. As the manuscript hardly uses the spatial representation of the maps with the dots (Fig 3,5,6), then there is limited argument for showing these comparisons on a map. Using tables instead of figures would therefore improve transparency substantially.

We have added supplementary table repeating the information provided by maps. Also, the color legend was revised to improve the readability of the maps.

The authors claim a strong forecasting skill in the summary. However Figure 7 clearly shows very low correlations (from below 0.1 to less that 0.4) and the RMSE (also figure 7) ranges from about 80 grains/m3 to 120 grains/m3. As far as I know this level is the typical level for severe warnings of tree pollen. If the typical error is of the same level as the warning level and the correlations generally low, then I do not find sufficient support for the statement concerning a strong forecasting skill.

This is again confusing: we do not claim strong forecasting skill in the summary. To the opposite, the paper states “noticeable deviations from both observations and each other” in the abstract and discuss in details the season shift in summary. What concerns “decent level of reproduction of short-term phenomena”, this is confirmed by the sensitivity runs: as soon as the season timing is crudely corrected the correlation nearly doubles, thus suggesting that the short-term features of the time series are fine. Therefore, we respectfully disagree with this criticism.

6. The results cannot be reproduced by fellow scientists. The model simulations can be reproduced, but the accuracy assessment cannot be done as there is no description of the observational record.

Description of the observations is provided in the revised paper as a supplementary table

7. The authors give credit to related work but it is my impression that they have not conducted a sufficient literature review, in particular in relation to other modelling approaches that simulates the start and the strength of the pollen season. This puts limitations to the scientific discussion of the work and the assessment of the quality of the model simulations. A quick search on Google identified studies by Orlandi et al (2006) and Rojo et al (2016). These studies in combination with papers written by the co-authors of this manuscript (e.g. Galan et al, 2005) present methods that seem to have similar or higher accuracy than this study with respect to start of the pollen season as well as the spatial modelling of the season index. This indicates that a deeper literature review is needed in order to position the findings in this manuscript against existing knowledge. The authors claim a strong forecasting skill in the summary. However, all of the models (Figure 7), except SILAM, seem to have a larger error with respect to predicting the start of the season than other multi-site methods presented in scientific literature. It can therefore be questioned if there is a strong forecasting skill of the models and the ensemble with respect to the start of the season.

The topic of comparison of regional numerical transport models with local statistical models (even if the latter ones are based on a few stations) is a large topic, which goes way beyond the current exercise. We discussed it in several previous publications, in particular in recent Ritenberga et al (2016), where showed that even for birch (the most-accurately modelled pollen type) the local model outperforms SILAM. This is how it should be: European-scale models cannot be hard-tuned to a single place in principle, they would inevitably miss rest of Europe. In that sense, for instance, better performance in Perugia of the model made in Perugia (Orlandi et al, 2006) is not surprising. Corresponding discussion is added.

The Rojo et al (2016) paper does not concern the season forecasting, only static dependence of SPI and source map. It is certainly an interesting work and we included it now but its findings are not relevant for the time series analysis.

At the same time, in conclusions we highlight the large season shift as the matter of primary importance, which calls for investigation and, possibly, reparameterization of the source term. And, once again, conclusions do not contain the claim of strong forecasting skills of the ensemble.

8. The title clearly reflects the contents of the paper

9. The abstract provides a concise summary of the study. However as described in the previous sections, then there is not sufficient material in the manuscript to support the findings that are presented in the abstract

See response to items 5, 6 and 7.

10. The overall presentation is well structured and clear.

11. The language is fluent and precise

12. Equation 1 and 2 are defined. However the scaling factors and are only partly described as their values could not be identified.

The omissions are corrected. See response to the Referee 1

13. The figures 3,5,6 are almost impossible to read causing that the findings presented in the conclusion and abstract rely on unclear material

The figure quality has been improved and the table material provided as a supplement.

14. There is about 75 references. Most of these are peer reviewed literature. However a substantial amount of the cited literature appear to be written (as lead or co-author) of the authors to this manuscript. The comments under point number 7 and the large number of references indicates that the authors of this manuscript had put too much weight on own publications and to little weight on publications by other authors.

The current work is a synthetic effort of the many groups playing key roles in the olive pollen aerobiology. We did our best to provide adequate representation of other groups, further expanded the literature review and discussion in the revised paper and included the extra references suggested by the referee. However, one of those papers was also co-authored by the authors of this paper, and we already had two works originated from that same group in the initial paper version. Therefore, we reject the blame of the biased literature review.

15. Supplementary material has not been included