We would like to thank the reviewers for their thorough work and helpful feedback, which led to a significant improvement of the manuscript. In the following, all remarks by the reviewers are listed in black text and our corresponding replies are given in blue.

Reviewer 1

GENERAL COMMENT

The manuscript describes a method to distribute observation time series over clusters, and to use this within the context of data assimilation. The method is applied with the regional air quality model EURAD-IM, which assimilates time series of surface observations to guide the model. Specifically, the clustering is used to sub-divide the observation time series in an "assimilation" subset (~70%, incorporated in the assimilation) and a "validation" subset (~30%, not incorporated). The posterior comparison between analyzed model state and observations should give the same statistics over the "assimilation" and "validation" set, but as shown by the manuscript too, the assimilation usually performs better over the "assimilated" set. The proposed clustering method improves the equality between the statistics over the "assimilation" and "validation" set, and is therefore of interest for all data-assimilation applications.

The clustering method is well described, and easy to follow also for readers without a background in clustering. The application is illustrated for the European air quality network. Especially the maps in Figure 2 and bar plots in Figure 3 are useful here, as they illustrate the result of the clustering and how it was achieved. The 8-clusters obtained with the KSC method shows for example the soft borders between geographical regions, which could not be obtained by simply clustering countries. As described by the authors, the map obtained with KSC shares characteristics with climate zones, which gives trust that the obtained clustering is also related to geographic properties.

The improvement in assimilation/validation (AV) statistics is illustrated based on comparison with CO and NO2 observations (Table 1). In general, RMSE over the assimilation set increases, while the RMSE over the validation set decreases, thus decreasing what is called here the AV-difference. This is a very important result, and shows the usefulness of the method. However, as table C1 shows, the AV-difference increases for most other considered species (SO2, PM2.5, and PM10), and depending on the clustering method, also for O3. These species are rather important for air quality, and one could argue that these are even more important than CO. Therefore, the method seems not immediately applicable yet in for example the CAMS assimilations in which EURAD-IM is included. Could the authors include a discussion on how the clustering could be improved such that the AV-difference is decreased for all chemical species? Are different features of the timeseries needed, for example based on rural/urban locations? Or should for example CO simply be excluded? For the current manuscript it is not necessary to add and evaluate new clustering configurations, but it would be useful to see some guidance for future work.

Thank you for emphasizing the need to be more specific about the results of the clustering method on other species than CO/NO_2. In the presented clustering methodology, all species observations from a specific measurement station are categorized into either the assimilation or the validation

set, called bundling. This is done to avoid introducing artificial advantages in the validation due to the correlation between simulated species. Due to the scarcity of CO observations, they have a large influence on the clustering result. In future work, not applying the bundling or defining a less restrictive bundling could improve the AV-difference for the other chemical species to a similar extent. This is, for example, done in the assimilation/validation split within the CAMS project, where only parts of the species are categorized similarly (NO_2/O_3 and PM_{10}/$PM_{2.5}$).

We have added the following paragraph to the discussion section:

"However, the representativity for the other modelled species is not improved. This is due to the relative scarcity of the CO observations (see Fig. 1) and a resulting strong influence on the clustering. Furthermore, all species observations from a specific measurement site are categorized either into the assimilation or the validation set, called bundling. This is done to prevent advantages in the data assimilation due to the correlation between different species. A less restrictive bundling, as for example in the CAMS observational data set (REF experiment), could mitigate this disadvantage. In the REF experiment NO_2 is bundled with O_3 and PM_{10} is bundled with $PM_{2.5}$."

SPECIFIC COMMENT

Line 184: Could the method used by CAMS to distribute observations in assimilation/validation set be summarized here? At lines 326-327 an essential difference is discussed, it might be useful to mention that earlier too.

We agree that a summary of the CAMS selection process is useful to understand the differences in the experiments. We have added this summary to the manuscript:
"The REF experiment contains the same observation stations as the KSC and CD experiments. The experiments differ only in the assignment of the observation stations to the assimilation/validation set used in the data assimilation runs. Note that the CAMS REF configuration applied a bundling of NO_2 with O_3 and PM_{10} with $PM_{2.5}$ observations (i.e., these species are always assigned to the same data set) before selecting semi-randomly, where validation stations are placed near assimilation stations. Here, spatially isolated observations are used for the assimilation."

The data processing requires many steps, for example outlier removal (lines 144-155), but also removal of stations extreme emission corrections in their vicinity (lines 216-221). It would be useful to summarize all selection criteria in for example a table, including the number (fraction) of removed stations.

Thank you for the comment. We have included the following text passage and table to the main text:

"A summary of the filtering procedures and their effect on the amount of observational data is shown in Table 1.

Table 1. Summary of the filtering procedures and the number of removed time series from the evaluation in percent.

Method	Removed time series [%]
Outlier Removal	0.0
Relative standard deviation	<0.1
Constant intervals	0.4
Annually averaged diurnal cycle outlier	0.3
RMSE outlier (January and CO only)	0.8
Emission factor outlier	2.2

“

The KSC clustering is applied using a location future, which gives a result that collects stations in geographic regions (adjacent countries and/or regions in countries). The map in the right panel of Figure 2 shows that within such cluster there are sometimes small regions with a different classification, for example the Pyrenees are part of cluster 7. Would it make sens to add features based on these "exceptions", for example the altitude of a station?

We thank the reviewer to emphasize that our description was not comprehensive enough, here. Indeed, the altitude is part of the features in the KSC clustering. For example, it clearly contributes to the definition of cluster 7. We have updated our description of the features in the manuscript to make this point clearer.

“In the second experiment, the k-means soft constrained algorithm is applied, named KSC in the following. Here, the features are the geographical coordinates and the altitude for each measurement station as well as the mean and variance of the annual average diurnal cycle.”

SPELL AND GRAMMER

Lines 99 and 106: "k" should be "K" as in Figure 1?

Yes, thank you for the comment. All “k” in the text have been changed to “K”.

Line 126: should be "... some objects, F_m, such that ..."

Done.

Line 225: should be reference to Fig. A1 ?

We have updated the reference. According to another comment by reviewer #2, we have moved the corresponding figure from the appendix to the main body of the text.

Line 240: "the Alps"

Done.

Line 253: remove comma's ?

Done.

Line 304: ".. month .."

Done.