Review of
Identifying source regions of air masses sampled at the tropical high-altitude site of Chacaltaya using WRF-FLEXPART and cluster analysis
by Diego Aliaga et al.

General comments

The work presented in this manuscript may be divided into two major parts:

1. WRF simulations and subsequent FLEXPART-WRF calculations to determine the influence regions for the Chacaltaya station, with hourly resolution at the receptor.

2. Regridding, clustering and interpretation of the results.

I think that the first part is very valuable and mostly well done (for a few issues I noted, see below).

As for the second part, my impression is that it does not fully exploit the information produced in the first part. The complicated postprocessing rather hides features rather than making them more easy to grasp. I cannot see a real benefit from the clustering process, at least in the way it was done and presented. The presentation quality of the clustering is insufficient.

Furthermore, I am afraid that the influence from the La Paz urban agglomeration is not represented as well as the original data would allow. It is obviously an important topic for the Chacaltaya measurements and should be given more attention throughout.

My suggestion would be to either completely replace or to amend the evaluation based on clustering by a direct statistical evaluation of SRR fields, detailling when which regions influence the monitoring station. Rather than a posteriori looking at which kinds of biomes are represented in which cluster areas, one could divide the total domain into regions of interest with similar trace substance emission properties (both natural and anthropogenic!), and then evaluate with respect to these regions, circumventing the complicated clustering process. If the clustering is still kept, the presentation quality both of the method and the results should be improved.

More detailed comments can be found below.

Detailed comments

1. Literature review: While the introduction discusses a number of relevant papers, it could include some more, especially for work in the Alps, including that related to mountain peak stations such as Jungfraujoch and Sonnblick.

2. Line 102: It should be explained why six months are sufficient to capture the climate of the area, and how representative the chosen period is with respect to interannual variability (ENSO).

3. Line 150: It would be better to give more details on the comparison with observations rather than to just claim “reasonable agreement”. At least, some quantitative scores need to be provided to back up this claim.

4. Line 158: Why the choice of 4 d back? How far do you get with that, or in other words, how many of the particle trajectories would end inside the evaluation area? How much of the variability of the atmospheric components of interest can be explained by that?

5. Line 169: Why 500 m for the lowest evaluation layer? The boundary layer height at nighttime could be considerably lower than the 500 m layer. Then, the sensitivity does not represent the influence of surface emissions properly. On the other hand, a resolution of 500 m is much too dense for the upper layers.

6. Line 190: I don’t think that you should neglect the meridian convergence in your domain: the 3% error that you cite (it might be a little more in the extreme case) translates to 30 km for a deviation of 1000 km from your central meridian which is on the order of your grid resolution (38 km). It would not have been such a computational burden to use a more accurate formula. In addition, if the resolution of the two grids is on the same order of magnitude, the regridding is not a trivial process. Therefore, even though a reference for the method is given, its key features should be described.

7. Line 198: It would be useful to provide the actual value of e^{α}.

8. Line 200: Is the resolution of the innermost rings sufficient to properly resolve the influence of the urban agglomeration? I think that it might not be.

9. Line 214: What is the rationale for the smoothing? Unless your FLEXPART output fields are very patchy (in which case one should increase the particle number and/or decrease the output grid resolution - see remark above about vertical resolution), it would not be beneficial. As you are doing this on the polar grid, at larger distances where patchy output is more likely, the grid resolution is already decreased. On the other hand, if you apply smoothing in the near field where strong gradients occur and are important to be represented properly (urban emissions!), you deliberately worsen your data.

10. Line 219: It is not clear to me how you applied the normalisation. What is in the denominator of the normalisation? Note that the total residence time of in each simulated release hour should be the same, except if particles leave the domain, but in that case, I don’t think that normalisation would be proper. The actual residence time is what determines the concentration change in a grid cell, not a normlised one.

11. What is the effect of “filtering out” zero values, and how exactly does it work? Do you mean that you eliminate regions of your domain where you have only rarely nonzero SRR values over your whole period? That might be justified, at least unless there are strong emitters at such locations, which even if it is rarely the case should not be missed.

12. Line 255: First of all, I think that we should primarily be interested in the surface influence and not so much in the boundary-layer influence. It is at the surface where the emissions occur. At times when the layer considered is well mixed, it does not matter how thick you select it (you would normalise with the thickness so that you obtain emission sensitivity with respect to the area source). If it is not well mixed, you are making a mistake if you take too thick a layer, which would be a problem at nighttime, as you mention. The only case where considering the PBL seems to be warranted is PBL-related chemistry, but then a Lagrangian model is probably not a good tool anyway.
    Secondly, I don't think that we are interested in the relative SRR. As explained before, the total SRR should not be highly variable, so this normalisation will not have much influence. But what is more important, the contribution of emissions from some area of interest is independent of whether particles leave the domain (meaning lower SRRtotal) or not.

13. The choice of mixing ratio vs. concentration for source and receptor in FLEXPART is not discussed, even though at the high elevation of the receptor, and as sources (output grid) vary between sea level and >15 km, this is a potentially important topic.

14. Fig 5 and associated discussion: In line with what I explained above, I would think that absolute values of SRR for surface influence would be more interesting than normalised or detrended ones. I do agree that finding out for which fraction of the time (and when!) there was little surface influence, but again, this should be quantified using an absolute threshold, not a fractional one.

15. Line 313: This correlation expresses what I tried to explain in comment 12. If we had well-mixed conditions up to 1.5 km, we would have perfect correlation (actually, a 1:1 relationship if properly normalised). One has to be aware, however, that this correlation will be higher after long transport and much lower very close to the source.
    I think that the regression formula does not serve any practical purposes and should be skipped.

16. Line 325: In addition to Fig. 4b I would like to see a plot of the average diurnal variation of the surface SRR. – Instead of "campaign", you might want to use "investigation period".

17. Figures 6 and 8: Caption illegible, thus I am not able to comment on it.

18. Figures 7, 9, 11: The time series should be presented as filled curves or bar plots to facilitate the interpretation, subfigures should be framed, and vertical lines as time markers be drawn. If my understanding is right that each hour on each day can belong to one cluster only, it would be better to stack all contributions on top of each other (stacked filled curves or stacked bars). This should more quickly show the seasonal variability. Also, it would probably be usefulto separate, also in this figure, in one way or another, the diurnal patterns from the longer-term (seasonal) evolution.

19. Methods and results of clustering: It is too difficult to understand the clustering method in detail and why it was chosen. One reason is that a formal description with symbols and formulae has been avoided and only verbal

explanations are included. I don’t think that the concept of “pathways” is particularly helpful if there is no such thing as natural pathways, and rather a continous spectrum of air movements is found. The results presented don’t appear to be very different from a simple division into sectors, thus one is wondering whether the application of this clustering method has real benefit for the understanding the situation. Also, we are never shown how a single cluster and its members looks like, we only see cluster centroids and homogeneously coloured regions.

20. Conclusions section: While the conclusions appear to make sense, it is difficult to identify them in the figures presented. This might be another hint that the way of evaluation and presentation should be improved.