Comment to *"Revealing dominant patterns of aerosol regimes in the lower troposphere and their evolution from pre-industrial times to the future in global climate models"* by Jingmin Li et al.

This paper uses aerosol concentrations from simulations with one global model as input to machine learning algorithms in order to provide classification of aerosol regimes under pre-industrial, present-day and future emissions. The paper is mostly well-written and structured, and the underlying methodology sound. However, in the current version the statements and conclusions give very limited novel information and added value to the literature and scientific community, and I also question the usefulness of the method in the context of assessing the influence of projected anthropogenic emission changes. Improvements are needed to make the manuscript suitable for publication.

While the use of concept of aerosol regimes does provide a different view on future aerosols than AOD, PM2.5, individual species, etc (of which there have been several studies), the paper at the moment provides many statements, descriptions and conclusions that are well-known and shown in previous literature and/or simply follow directly from the underlying emissions that are used as input to the model. The classification method itself was also well documented in L22. Statements such as "suggesting a general reduction of aerosol and aerosol precursor emissions in line with the underlying assumptions in these scenarios." seems rather circular and given that there is no change in climate in these model runs, really only says that emissions have been properly read into the model and that the classification method works also for different emissions than those used in L22. Another example is "This trend agrees with the temporal development of the corresponding aerosol regimes." – when in fact this trend drives the corresponding aerosol regime change. The "emission analysis" also appears a bit simplistic. At the very least, revisions of the abstract, results and conclusions are needed to emphasize what's important/new knowledge from this paper, incl. e.g. comparing the usefulness of this ML approach and clustering with other model based assessments, and acknowledge how this work supports/complements other published work looking at future aerosol. There is some discussion of previous studies, but it cites only two rather old papers.

One question that arises is how sensitive the classifications, and the subsequent conclusions, are to the model input used? We know that global models have a widespread in simulated aerosol distributions and an extension of the is work that would bring added value is to consider multi-model data, e.g. from CMIP6. E.g. can robust regimes be identified for present-day and future scenarios? The authors should consider adding a multi-model perspective here, either for both present-day and future, or just present-day – or at the very least discuss this.

The classification is also a bit of a black box. While the point of the method is that the classification criteria is not known a priori, understanding the distinctions is important for further application. E.g. what is the criterium for calling something level 1 vs 4. This, combined with the naming, will in my view very much limit the relevance for policy-making/mitigation – which is one of the important applications the authors point out. It should be made clearer here (and at least to a reader who is not a ML expert) how the classification is done, e.g. what is the criteria for transitioning to a lower level continental airmass? What's the role of composition vs. mass concentration vs. number concentration – can level 1 continental air in fact be very differently composed in different time periods if emissions of different species change differently? If so, that would have subsequent implications for the climate effects, which would not be easily extracted. For policy relevance, could the different levels be related to e.g. air quality indices?

Another implication of the coarse aerosol regime approach is that a lot of detail is hidden. For instance, most of Africa is classified as dust-dominated and biogenic/biomass burning. In contrast, scenarios for anthropogenic emissions exhibit a wide range in future evolutions, which has significant implications for climate and air quality, but is not at all captured here. What is the authors' view on this? This draws into question the usefulness of such coarse classifications and I think the authors need to spend some more time justifying why their approach provides what they call a "clear and condensed" picture.

Another limitation is the fixed meteorology. The authors do talk about this, saying that "it would hamper the separation of their respective impacts, further complicating the interpretation of results and the applicability of the proposed method." In my view, this is where ML could be of added value, i.e. application of the algorithm to a more complex data set. While this is likely not possible here, a perhaps more important implication is that not having changing climate has implications for the aerosol load and composition in both future and pre-industrial times and requires more care than currently taken when talking about changes from the pre-industrial. For instance, studies have shown an increase in the dust loading and analysis of CMIP6 data have looked at possible feedbacks on natural aerosols. The authors should include a better discussion of such work and possible implications of these findings for their results.

The authors also discuss differences between 1750 and 1850: given the large uncertainties, I question the robustness of any conclusion drawn for this time period and the authors may want to acknowledge that more clearly.

Specific comments:

Section 3.2: why is transport singled out here from other non-transport emissions when this paper has no transport focus? In many cases there are little or no emissions and hence extracting only this sector does not help inform the changes.

Section 3.2: it would be helpful to have regions named rather than Rx.

Line 461: "transport sector shows the largest contribution to the total emissions in these regions, followed by the contributions from the transport sector" Well of course, you have not separated out any other anthropogenic sector… This statement gives no useful information.

Line 423: "Most of the processes driving the anthropogenic aerosol changes will be addressed by the analysis of these species" – what is meant by the word processes here? The reference is to the emissions documentation paper so I assume it's related to "processes" leading to emission changes – but could be misunderstood to involve also interactions between different species through atm. chemistry when emissions change (e.g. SOA formation changes when OC/POA emissions change). Moreover, this statement is probably true but that's because you have not change in climate, which should be specified. Perhaps rephrase.

Line 450: "The different pathways of emission changes in R2a and R2b can explain why R2a remains in the polluted regimes in 2050, while R2b shifts to a clean aerosol regime under SSP1-1.9" – can explain? What are the other possible explanations in this model study?

Line 457-459: "The emission comparisons for both regions (Fig. 7f and g) show that the emission maxima of NOx, SO2 and BC occur at present-day, while emissions for NH3 increase up to 20% in 2050 under the most pessimistic SSP3-7.0 scenario. However, the maximum aerosol emissions generally peak at present-day." Emissions in North America and Europe declined prior to 2015, so this is not accurate but appears to be the case because you don't show the full time series.

Line 479: "This, however, is less critical in the context of this study, due to the standardization process." I don't understand this statement. If you classify or standardize something that is not representative of the real world, how is that OK or not important?

Line 504: If referring to the dataset used in this study, then "huge and complex" seems a bit of an overstatement…

Line 506: given the list of co-authors I can see why aircraft engines are selected as an example, however, I struggle a bit with this example since the authors point. Given the coarse nature (in space and time) of the classification approach, how would the data be used in engine life cycle modeling? And how would this better come from this study than all the other studies focusing on aerosol composition, with a full 3D spatial distribution?