Review of the manuscript "Statistical downscaling of water vapour satellite measurements from profiles of tropical ice clouds" by Giulia Carella, Mathieu Vrac, Hélène Brogniez, Pascal Yiou, and Hélène Chepfer.

In the manuscript, a novel approach for approximating the relative humidity (RH) in the middle atmosphere from the observations of scattering ratio (SR) of the CALIPSO satellite instrument is presented. Several machine learning models are exploited as the primary approach along with additional iterative corrections aimed to preserve the average estimated RH to be close to observed by SAPHIR instrument. Preprocessing of source data includes spatial co-location of SAPHIR and CALIPSO sounding locations; averaging of characteristics of cloud profiles above 2km; filtering of CALIPSO observations related to ICE and ICE-MIX clouds based on k-means clustering results and additional filtering.

Overall I find the idea and the approach of the study to be very promising. The models exploited in the study are described in a clear manner, without over-complications or obscured details. I find the English of the manuscript a bit knotty, but acceptable.

General comment:

The main concern about the presented study is the unclear general setup of the problem. It is clear that the covariates for a model (one of five regression models) are SR values preprocessed in a way described in Section 2.2. It is also clear that the target values of a model are RH at the points at fine spatial resolution colocated with CALIPSO sounding events. However, it is unclear which values are the ground truth for these target values. Are they just the value of SAPHIR pixel RH? Then it should be mentioned clearly.

As far as I understand the approach of the authors, the general setup of the problem is as follows:

SR (at p=21 height layers) are the covariates for a model. SR are defined at fine spatial (horizontal) resolution. RH_l (at 6 pressure levels) are target values for a model. RH_l are defined at coarse spatial resolution. A model is fitted to estimate RH_l at each point of CALIPSO sounding based on SR values of these points. There is also a step for preserving the so-called "mass balance." And there is no "ground truth" for RH values at the fine spatial resolution other than coarse SAPHIR data.

Following this setup, the authors fitted the models they presented in Section 3.3.1. As a result, the authors obtained RH at a fine (horizontal) resolution with some variability within each SAPHIR pixel. However, there is no reason to claim this variability to be downscaled structure of RH within SAPHIR pixel. The uncertainty of the RH_l value estimated be a model within a SAPHIR pixel is explained by two factors: RH variability itself and errors of a model. The authors did not mention in any way how did they determine the part of the uncertainty that is related to RH variability. And, as far as I understand the setup of the presented study, there is no reason to claim the variability of RH_l value estimated by a model to be a fine structure of RH. It is especially arguable in case of non-parametric models like RF or QRF.

According to the above, the naming of the study is confusing and misleading. I would totally agree with the title mentioning "approximation" or "regression" of RH based on CALIPSO profiles. However, the "downscaling" naming seems incorrect in this case.

Considering those mentioned above, the discussion about the downscaling results is also unconvincing (p.12 L27 - p.13 L12). Taking the design of the study into account, one cannot claim to be able to obtain variations of RH within SAPHIR pixels.

The conclusions related to the downscaling feature of the presented approach are incorrect in this case as well. The results of a model can be considered as an estimate of RH with some uncertainty, but cannot be considered as a downscaling product with estimates of inhomogeneities.

Section 3.2:
The classification of the SR profiles based on clustering techniques does not seem reliable. Since clustering itself is a tool for explorative analysis rather than classification, one cannot claim the clustering result to be the answer for the question "which profiles correspond to ice clouds and which do not." It is not clear how do authors checked if the profiles of the cluster 1 indeed correspond to ICE and ICE-MIX observations (p. 7, L3-9). As it seems at the current state of the manuscript, the similarity of clustering made with k-means with the observations characterized as ICE and ICE-MIX was assessed visually. If that is the case, it does not seem like a convincing way of detection of ice clouds. It is also not clear why the authors did not use cloud phase flags instead of this additional clustering approach.

Section 4:
The results described in Section 4 (p. 12, L1-2) are inconsistent with Figure 5 in terms of the values of the coefficient of determination. R^2 for L6 mentioned to be ~0.3 in the text of the manuscript; meanwhile, it is 0.4 in the Fig.5. Also, for the layers L1-L5 the values of R^2 claimed to be >=0.7, and as presented in the figure, are precisely 0.7. While formally, it is not a mistake; it seems like a mistyping in the figure.

Fig.7:
The description of results shown in Fig.7 is not clear. It is unclear what is the statement to which "Similar results can be found for different choices of the number of clusters…" is related (p.12, L12-15)

Fig.3:
The horizontal axis meaning is not clear enough. It should be labeled with some kind of "cluster number" of another informative label.