Anonymous Referee #1

1. Introduction

Could you please add, why did you choose CALIOP? For me, there is no physical reason to connect these two pieces of information. CALIOP measures SC at 532 and 1064 nm, which has no connection to water vapour. All information in SC is from cloud particles – mainly ice, in your case. You mention a few articles to infer correlations, but you do not really point out or cite the physical reasoning from there. Please explain, why you think CALIOP is a good choice for water vapour information. I know, your regression models do not need a relationship. But you had your reasons to connect a microwave sounder and a lidar, didn't you?! Because for me, this is not an obvious choice. So please reiterate, why you think this would lead to a physically correct downscaling.

As clearly pointed here by the reviewer, from a purely remote sensing point of view, there is indeed no connection between the backscatter at 532 nm and 1064 nm and the water vapour concentration. However, numerous studies have highlighted that there are relationships between the presence of upper tropospheric ice clouds and the surrounding moisture. For instance, Jensen et al. (1996, GRL; 2001, JGR) and Rosenfield et al. (1998, GRL) showed that the formation of thin cirrus is associated to the dehydration of the upper troposphere, essential for their maintenance. Luo & Rossow (2004, J. Clim.) as well as Soden (2004, GRL) and Chung et al (2007, ACP) related the life cycle of cirrus anvils to an increase in the moisture content of their close environment, via detrainment. This was also highlighted by Eguchi & Shiotani (2004, JGR). Moreover, Martins et al (JGR, 2011) examined the link between optically thin ice cloud observed by CALIPSO and the collocated upper troposphere water vapour from the Microwave Limb Sounder (MLS) on the Aura platform and found that cirrus cloud detections in the upper troposphere are correlated with a significant increase in the observed upper tropospheric water vapour concentrations compared to the average. The link between ice clouds and water vapour was also examined in Hoareau et al. (ASL, 2016) using ground base lidar observations collected at la Réunion island in the tropics. On the other hand, as for example explained in Cesana and Chepfer (2013, JGRA), the lidar scattering ratio (SR) also depends on the amount of condensed water (and therefore a mix of concentration, size and shape of ice crystals in the atmosphere) and increases with the amount of condensed ice in the atmosphere (only when the cloud optical depth <3, which is the case for most ice clouds). Therefore, given that both water vapour and the lidar intensity are linked to ice crystals concentration, size and shape, we expect some correlation between the measured RH by SAPHIR and the SR signal from CALIPSO. This being said, the retrieval of RH from the microwave sounder is not biased by the presence of ice particles. Therefore, in the following, we assumed that the retrieved RH from SAPHIR can indeed be reasonably predicted given CALIOP measurements of ice clouds.

We have added some text in the manuscript (page 3, lines 10-28) to clarify this point.

2. Data

Please explain, if the product by Brogniez an official product. Is there an official web-page, source . . ..? Same with CALIOP. Did you use the official product? It seems like, but I want to make sure.

- The RH profiles from SAPHIR used here are depicted in Sivira et al. (2015, AMT) and in Brogniez et al. (2016, JAOT) and this is indeed the official product. It can be downloaded, after registration, from the French ground segment of the Megha-Tropiques data:

  http://www.icare.univ-lille1.fr/mt/

- The scattering ratio profiles from the lidar CALIPSO used in this study are from the GCM-Oriented Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) Cloud Product (CALIPSO-GOCCP, Chepfer et al., 2010)). GOCCP product is part of the CFMIP-OBS database (Cloud Feedback Model Intercompariosn Program). It has been compared to the NASA CALIPSO product in Cesana et al. 2016 (JGR) and Chepfer et al. 2013 (JAOT). Data can be downloaded from:

  http://climserv.ipsl.polytechnique.fr/cfmip-obs/Calipso_goccp.html

3. Methods

3.1. I consider the last sentence in 3.1 is crucial for justification of your technique. You should explain this a little but more, perhaps with more citations from Schroeder's or other papers. It indicates a connection between RH and SR, something which is very important for your approach.

This comment on the connection between RH and SR is similar to what was raised in the comment #1 (Introduction). Please refer to that comment for further justification.

We have also deleted the reference to Schröder et al. (2017), since in that study the authors are not mentioning the fact that the correlation between ice clouds and Upper Tropospheric Humidity is large.

3.2. But I am not quite sure, how to understand chapter 3.2. You have already a cloud classification from CALIOP (right plot), so why did you prefer your own k-mean method? Both tell you, that clouds above 10km are ice, which is not really big news. You could have just used that value or the CALIOP classification, so why do you insist to do your analysis based on this extensive k-mean clustering?

By averaging the SR profiles above the boundary layer to a 1 km resolution with the aim of reducing the noise and the amount of missing data, we also had to apply the same averaging procedure to the cloud phase flag profiles in order to maintain a coherence between the SR profiles used in the regression model and the corresponding cluster. Because of the "mixed" flags resulting from this averaging procedure, the statistically-based clustering method was preferred since it encompasses the problem of giving a physical interpretation to "mixed" profiles (c.f. Fig. 3c). Moreover, in this way, the downscaling method described in this study can be more easily generalized, without having to worry about the physical interpretation of the clusters.

We have added some text in the manuscript (page 6, lines 19-26) to clarify this point.

3.3 Formula 1 in chapter 3.3 is my biggest problem: it assumes RH_l is connected to (SR1, SR2, . . . SR_p) via a function. That seems to be the foundation of your idea. But for me, there is none. At least, no physical connection. Is it enough for this approach to find a correlation without reason? I am ok with that, but the results would be of limited use for research (see conclusion). Please re-iterate more here – or at least in the introduction, perhaps based on the articles by Brogniez or Udelhofen-and-Hartmann.

Once again, this comment discusses the connection between RH and SR is similar to what was raised in previous comments, 1 and 3.1. Considering the expected physical correlation between RH and condensed ice (crystal concentration, shape and size), as well as the published papers that have highlighted the relationships between upper tropospheric moisture and ice clouds, we think that we have sound reasons to link, via a function, RH and SR.

See also the detailed response to comment #1.

3.4. Choice of regression model: I miss some important information: I understand your limitation to ice clouds. How much data do you use for the regression training with respect to the mission time frame? Do you use specific dates? Do you use the same amount for all regression models (RF,QRF,GMRF1,2,3). Did you make tests with different amounts/dates? Where the results always the same?

As stated in the manuscript (c.f. for example Figure 7) we tested the method using different time periods (July and January 2013) and different ocean basins (Indian and Pacific). All methods were trained on the same data, in order to allow for a fair comparison.

3.5. Do you deal with the error of the RH-retrieval in RF and QRF? If you look at Figure 2, there are lots of layers with uncertainties > 30%, especially below 500hPa. (Remark: You might want to choose a different color scale there, it is really hard to understand. Everything above 30% is the same color . . ..). Retrieval tend to get worse closer to the ground. Do you deal with it differently?

In this study, we did not account for errors in the RH retrieval (we used the mean of the RH distribution from the retrieval algorithm) but this point can be further developed in future studies.

Larger uncertainties in the retrieved RH are expected at lower altitude because of the distribution of the sounding channels of the radiometer and because of their bandwidth (Clain et al., 2015 JAOT). The latter is narrow (0.2 GHz) for the central channels of the 183.31 GHz absorption line, which translates into a low uncertainty for the upper tropospheric estimates, and it stretches (2 GHz) for the channels located in the wings of the line, implying a larger uncertainty for the retrieval. Overall, RH measurements with a standard deviation larger than 30% might be considered very uncertain (which explains the chosen colour scale).

We have added some text in the manuscript (page 6, lines 7-14) to highlight this point.

3.6. I try to understand, why you would chose so many variations of the GAM approach. There is no reason to assume a linear connection between RH and SP, so RF and QRF are quite reasonable to me. But here you suddenly force a linear connection. Is it just for comparison? Because it seems to do bad anyway, when I look at later results. Please re-iterate the reasons for this selection and the two derivatives (GMRF and geoadditive). Are there other options? I consider 3 GMF approaches, which have all bad skills, a little bit redundant. I would rather see a third different approach than 3 similar fails. But ok, if you want to keep them, that is fine too.

First, note that, as explained in section 3.3.1, a "Generalized Additive Model" (GAM) is not a linear model. Moreover, compared to tree-based models, GAMs, in addition to offering the advantage of interpretability of the model coefficients, also allow the direct incorporation of a spatial correlation structure. Although this turns out not to be important for this particular downscaling application (as shown in section 4), we believe it is still useful to include the description and the results of these models in the manuscript for potential applications of the method to different data and problems.

3.7. Chapter 4: Actually, p.10, line 18-21 is another short talk about a possible physical connection between RH and SR. If you could extend this a little bit more, especially in the Introduction, then the approach would be much easier to understand.

We have extended the discussion on the link between RH and SR, as suggested also by the previous comments.

3.8. I also have problems to understand Figure 5. Is the predicted from RF? And is the observed RH the one from 10x10km SAPHIR? The description in the text is very short and confusing. Please explain more here: source of predicted, source of observed. Why would you then have such a bad correlation for L6? Please explain this plot in more detail, it seems it is your only source of verification for your approach. Most people would prefer an independent source (radiosonde, airplane observation, ...), but I guess you don't have enough data for this in the Indian ocean. So, you have to convince the reader about the "success" of your approach with this plot. Honestly, I didn't get convinced, you didn't write enough.

Figure 5 shows the median of the distribution of the predicted RH for each vertical layer using the Quantile Random Forest method vs. the RH observed by SAPHIR (at 10x10 km resolution). Here the predictions are the results of the 5-fold cross validation procedure, and are therefore derived from a model trained on an independent part of the data set. Although a comparison with other sources of RH data, as for example those cited by the reviewer, could be interesting, it will not necessarily be a validation of the results of our model. In fact, apart from the difficulty of finding a statistically significant sample of radiosondes or airplane observations co-located in space and time with CALIPSO measurements, these sources are characterized by different spatial resolutions from lidar data, which makes the comparison not straightforward.

We have added some text in the manuscript (page 11, lines 18-31 and page 12, lines 1-6) to clarify what Figure 5 is.

4. Chapter 5

At the moment, I am questioning some bullet points in your conclusion. I am not quite convinced that your data can help “study . . . small scale water processes” or “evaluate . . . water vapour interactions”. You need to convince me, that you have a physical foundation, not just correlated sorting. On the other side, I agree that you can always “evaluate small scale inhomogeneities” in reanalysis or “guide parameterizations”. Models need to know the behaviour of parameterizations on smaller scales, so you might be very helpful to find out scale breaks on scales around 100 m. I also think, you should talk a little bit more about the extension to other clouds. It sounds interesting, but based on your requirements (homogeneity, strong SP signal), you might be in trouble. If you could talk more about future possibilities and obstacles, it would be a better selling point for this article. But that is more my opinion. . .

As stated in comment #1, water vapour and the lidar intensity both are linked to ice crystal concentration and size (and shape). This represents the physical foundation of our method and implies that the correlation observed between SAPHIR RH and CALISPO SR emerges from a physical relationship, which indeed mean that the downscaled profiles can then be used to help to study small scale water cycle processes.

We do not agree that our method requires homogeneity or a strong SR signal. On the other hand, it is true that liquid clouds could present additional challenges in the implementation of the method. In fact, while SAPHIR is not able to retrieve the RH profile in the case of heavy precipitation, which implies that the majority of ice clouds co-located with SAPHIR measurements are non-precipitating, this is not true for light precipitating clouds, which typically correspond to low-level liquid clouds only. Therefore, for liquid clouds, including the radar reflectivity as measured by the radar CloudSat, which is indicative of the intensity of rainfall, might increase the model explanatory power.

We have added some text in the manuscript (page 14, lines 25-28) to clarify this last point.

5. Minor comments, found during reading:

- p. 2, line 3: Should be “state-of-the-art”

  Changed.

- p. 2, line 31: I am not quite sure, what “space clouds” are . . ..

  Changed.

- p. 5, line 10: should be “nadir”

  Changed.

- p. 6, line 1 : is the “1” necessary here?

  Table numbering is required by the journal.

- p. 10, line 2-4: this sentence is hard to read with all the comma and brackets. I would propose to redo it a little bit.

  Changed.

- p. 12, line 16: should be “CALIPSO”

  Changed.

Anonymous Referee #2

1. The need for fine scale observations of the vertical structure of water vapour is clear and well justified. But I probably missed a major thing reading the manuscript: from Figures 8 and 9, it seems to be more of a horizontal downscaling of the SAPHIR RH product than a vertical downscaling of it. Please clarify this either in the introduction of in the results. As I said, I might have missed something but I may not be the only one when reading your work.

The reviewer is correct in pointing out that the focus of this paper is the horizontal downscaling of RH profiles at the resolution of cloud measurements. The main interest of this study is in fact to test a statistical approach to overcome the barrier of the coarse footprint size of the radiometer, which implies that small-scale heterogeneities in the RH field are missed.

The coarse vertical resolution is also critical (not less), especially in cases where there are strong vertical gradients of moisture. For instance, at the top of the atmospheric boundary layer over the oceans in regions of shallow clouds (stratocumulus or cumulus) the boundary layer can be really moist, near saturation, whereas the free troposphere above can be extremely dry. Similarly, at the Upper Troposphere/lower Stratosphere boundary, the moisture is really low and this is critical for the ozone budget. However, these are two different topics that could indeed be tackled with similar approaches, but require different sets of proxies.

We have added some text in the manuscript (page 4, lines 13-20) to clarify this point.

2. The results shown on Figure (8e) indicate that the differences between the predicted RH and the RH estimated from SAPHIR can be quite large. It seems that Figure (8e) is not commented at all in the text but it needs explanations. Can the differences be explained by representativeness errors between the CALIOP lidar and the SAPHIR radiometer?

Differences between the downscaled and the observed RH observations will be larger when the RH field is characterized by finer-scale heterogeneities deriving from finer-scale processes, as for instance Figure 8e seems to suggest for some of the profiles. However, these differences are expected since with the method presented here the predicted relative humidity structure incorporates the higher-resolution variability from cloud profiles. On the other hand, as shown both in Fig. 4, 5 and 7, the downscaling model is able to successfully explain the coarse-scale RH observations from the finer-scale SR measurements and the overall bias is low, which gives us confidence in the predictions.

We have added some text in the manuscript (page 12, lines 29-32 and page 13, lines 1-2) discussing Fig. 8e.

3. The results shown in Figure 9 where RH estimated from SAPHIR and the predicted RH are on the top each other seem to indicate there is a bias between the two, especially in the lower layers. Could you please comment on this? In the paragraph page 11 where these results are presented, there is a comment on the variance of the predicted RH but not on its bias.

The model bias is overall low, as discussed in the previous point. On the other hand, what Fig. 9 is showing is that the variations explained by the spatial smoothing are negligible, and that the SR predictors alone explain the largest component of the variance in the RH field. In other words, once the effect of the SR predictors is taken into account, the residuals (i.e. the difference between the observed and the predicted RH) do not show spatial autocorrelation.

We have added some text in the manuscript (page 13, lines 8-10) discussing Fig. 9.

4. In the Data section on CALIPSO data, it is shortly explained that the noise on the profiles has been reduced using a Principal Component Analysis to keep only 90% of the variance. Why 90%? Would have the results fundamentally changed if you hadn't done this filtering?

Before clustering of the SR profiles and for clustering only, we decided to keep the number of PC components explaining the 90% of the variance (resulting in 19 retained PCs) as little variance was gained by retaining additional components. However, the results of the study did not change fundamentally when no noise reduction was applied prior to clustering (not shown). To improve clarity, we moved the description of the PCA analysis to section 3.2 and refer to the textbook of von Storch and Zwiers (1999).

5. Minor comments:

- Figure 1 shows a case of January 2nd in 2017 but the rest of the examples are for July 2013. Is there a way you could update Figure 1 to show the same meteorological situation all along the manuscript?

  We left Figure 1 as it is because here it is only used as a schematic representation of the method and is not intended to give any physical insight.

- Page 2, line 29 : "These detailed profiles are observed all over the globe" => Isn't SAPHIR observing the Tropics only? Please correct this sentence.

  Changed.

- Page 5, line 8, "3.1 SAPHIR-CALIPSO co-location" => The period of the study is not mentioned here but that we be good to know at this stage and not only later in Section 3.2

  The layer-averaged RH profiles from SAPHIR from Brogniez et al. (2016) are available for the period October 2011 – present, while CALIPSO-GOCCP product is available between June 2006 and December 2018. This has been clarified adding some text in section 2.1 (page 4, line 21-22) and section 2.2 (page 4, line 31 and page 5, line 1) respectively.

- Page 5, line 23 : "recontructed" => reconstructed˘

  Changed.

- Page 10, line 32 : "on the distance from the cost" => coast

  Changed.