Response to Reviewer #2

We thank the Reviewer #2 for the thoughtful and constructive comments concerning our manuscript.
Indeed, the ANN method itself is a well-established method. The novelty consists in its application by training the ANNs over a large statistics of collocated data, though limited in space and time, to develop optimized non-linear regression models to provide a more complete picture in space and time. The provided comments helped us to improve the manuscript for clarity. Where appropriate, we modified the text of the manuscript and the supplement with the changes marked in yellow. This marked text together with the new figures are provided at the end of the response to reviewer #1. Below, we provide point-by-point answers to each of the comments of reviewer #2.

Major comments
1. Section 2.1: This section should be shortened and only the key facts of the CRIS data set relevant for the interpretation of the results should be mentioned. The other part can be moved to the supplement part. Line 90: Is the information about AMSU relevant for CIRS, if yes, this yields to some restrictions for the application of the new ANN-based method.
First, we have added a short paragraph just after the title of section 2, which gives the purpose of the subsections. Section 2.1 describes the CIRS cloud data which are used as input for the machine learning as well as the cloud system data derived from the CIRS data which are used in the analysis of section 4. As both datasets are already published, we have shortened the whole section.
No, we don't use AMSU data. The sentence was there only to explain the grouping of 3 x 3 AIRS measurements. As this information is not relevant for the rest of the article, we have taken out this sentence.

2. Section 2.2: This section can also be shortened or some parts can be moved to the supplement. E.g. the ERA-Interim description, TIGR data set.
We have considerably shortened this section, as the ERA-Interim data are published elsewhere. We describe shortly all variables which are used as input parameters in the ANN models. However, we did not move the removed parts to the supplement, as the supplement is already quite long.

3. Section 2.3: This section can also be shortened or some parts can be moved to the supplement.
This section describes the target data as well as their quality. We also shortened this section, but we kept the description of how the radiative heating rates were determined as well as a summary of their evaluation.
All in all, we restructured and shortened sections 2.1-2.3 by more than 20% and hope that the new description is easier to read.

4. Line 223 ff: the absolute number of pattern (samples for training/test/validation) should be given. These values are important for the interpretation of the results in Tab. 2 as well as for Fig. S1 and S2. For the latter ones, it should be explained why the number of epochs is different for the different data sets and what was the stopping criteria for the training of the ANN.
The four years of collocated data correspond to a very large statistics of more than 16 million data points. We added this information in the first paragraph of section 2.4. When separating by scene type, the samples vary from 94000 Cb samples over land to 4.8 million mid- and lowlevel cloud samples over ocean. These samples contain both, AM and PM data. For the training of the SW heating rates, only half of the data are used (PM), which still leaves very large samples.
The number of epochs to converge towards a minimum loss is relatively small: less than 60 for cloudy scenes (Figure S1) and less than 45 for clear sky scenes (Figure S2). Essentially, the MAE

decreases considerably only within the first 10 (5) epochs for cloudy (clear sky) scenes. The relatively small number of epochs necessary for convergence may be explained by the large statistics we use for the training and the number of relevant variables for the prediction.

5. Line 280 ff.: The different types of models should be given in a bullet list or table with corresponding labels given in Tab. 2 and streamlined with the labels in Fig. S1, S2.

We have summarized the sensitivity experiments with the corresponding variables also in Table 1. In the supplement we added a text for the description and interpretation for Figures S1 and S2, as well as an assignment of the labels to the experiments.

6. Tab. 2 & 3: In these tables as well as in the discussion of it, relative error measures should be given, too. It would be good in order to judge the approximation and generalisation accuracy of the ANN it would be good to have the mean absolute percentage error (MAPE) in addition.

As we discuss in section 3, the average MAE over the vertical HR profiles is only one criterion to choose the best model for the prediction of the vertical HR profiles. In the beginning we also considered percentage errors, but the problem is that all cloud types and clear sky have at some vertical layer a value near 0. In particular for Cb, the lower layers have HRs close to 0. This would make an interpretation of the MAPE quite difficult. MAPE values would be automatically larger for profile types with more small HR values within the troposphere. In order to get reasonable MAPE values one has to introduce a lower absolute limit of the HR values. Considering the new Figures S3 and S4, which present the difference between predicted and observed HR profiles as well as the HR profiles, for Cb for example the difference is close to 0 for layers below an altitude of 800 hPa, while their HRs are also close to 0, leading to an artificially large percentage error. For Cb the maximum LW HR bias of the Cb model (red) is about 0.25 K/day for an average LW HR of -3.5 K/day and the maximum SW HR bias is about 0.5 K/day for an average SW HR of about 4 K/day. This corresponds to a percentage error of 7% and of 13%, respectively. As we use the same metrics in the tables, we can compare the performances of the models using different sets of variables. Figures S3 and S4 make it possible to roughly compute the percentage errors for different layers. MAPE may be a very useful metrics for other applications, but we do not see what the additional computation of MAPE would add to the interpretation of our results.

7. Fig 13 & 14: It is hard to compare the different panels of these figures and the usage e.g. for upcoming climate studies. It would be better to have only one panel of the total net HR. Furthermore in order to judge the influence of the ENSO index, PDO and surface temperature (see Fig. 12) to total net HR over time in a more quantitative way, mean total net HR time series data for different pressure layers (e.g. low, middle, high) should be correlated to the time series data of Fig. 12.

We now only show the 24-hr net radiative heating effects. In addition, we have added time series of the net HRs integrated over three vertical layers (100 - 200 hPa, 200 – 600 hPa, 600 – 900 hPa) and have computed correlation coefficients with the different other variables.

Minor comments

Line 237: Line 255 ff. There are techniques available to deal with partly missing values in the target vector. The target vector can be masked for valid/not valid training value in the target vector. Then only for the valid elements in the target vector, the error in backpropagated during training. For not valid elements (NaN) the error is set to zero. This is a proven concept for training of ANN with incomplete target vectors.

Thank you for this information. The authors were not aware of this, even after having discussed with several AI experts. Therefore we have used another method (replacing invalid values below the surface by mean values classified per month and scene type), which is perhaps less elegant but should give similar results.

Line 423 & 424: ".. is 24% larger, larger than 21% found by Li … " needs some clarification
We have redone the computations by using a LW average over AM and PM (instead of only PM) and added uncertainty estimates from clear sky identification and diurnal cloud amount variability. The final result lies between 20 and 25%, or 22 ± 3 %, which is only slightly larger than 21%. However the shape of the HR profile is different compared to the result of Li et al..

Line 578: data processed for 30N to 30S; but only results of latitude band 15N to 15S are shown (Fig. 10).
We have added maps which show results between 30N and 30S. We have mostly shown results for the deep tropics, as these have been shown by Li et al. 2013.

Fig. 7: the data sources should be mentioned in more detail.
Precipitable water and surface temperature are from ERA-Interim and UT cloud frequency of occurrence from CIRS.

The quality of some figures should be improved e.g. Fig. 6 (use of vector instead of raster graphics is highly recommended)
redone.

Recommendation:
The developed method to derive high resolution 3D HR in the inner tropics uses a lot of different model data: CRIS, ERA-interim, MOIDS AOD. Each of the models mentioned has its own errors and bias which are described well in paper. ANN can handle systematic model error of the input data well, but if one or more models will change over time (which is likely for such kind of long term data sets) the trained ANN model for the generation of 3D HR data will generate most likely biases. ANN are also not able to cope for random errors in the model input data.
This can be omitted if the original satellite data (in this case AIRS spectral radiance data) are used with full spectral resolution as input data. This makes the ANN HR model more applicable and more robust especially in order to transfer this approach to other IR sounder data (e.g. IASI) for further studies. For transfer of a trained ANN model on AIRS data e.g. transfer learning techniques can be used to adapt it for IASI.
We decided to use physical variables instead of radiances for different reasons. We use CIRS cloud data, retrieved from AIRS or IASI, together with ERA-Interim atmospheric and surface data. The latter have been also used as ancillary data in the CIRS retrieval, which gives a certain coherence. We tested two additional variables, vertical velocity from ERA5 and monthly mean AOD from MODIS, but finally we do not use them in the final models, as we could not detect a considerable improvement. We have been careful to only select variables which are also available for the CIRS-IASI data, so that the same models can be applied on IASI data. For the evaluation we need then independent data, as IASI and CALIPSO-CloudSat data do not overlap in the tropical band. We foresee to use ARM data for an independent evaluation, though these also have their issues, as mentioned in sections 2.3 and 4.1.
We agree that we could have used the radiances as input parameters, but it would have been technically much more complicated for us, as we would have needed to download the full AIRS and IASI spectra (2378 channels and 8461 channels, respectively) and then choose the most relevant channels for the training of the ANN models. As we have trained models for different cloud types and clear sky, for the reasons described in our manuscript (for example Cb is very rare,

with less than 10%), we would have needed anyway the CIRS data for the distinction of scene types. Furthermore it would have needed additional care to adapt the IASI spectral channels to those of AIRS in order to use the same models. As our funding is very limited (the three co-authors worked each only 6 months on this project), we decided to use the information which was easily available and which is also used in combination for further studies.