Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-64-AC2, 2017

Atmospheric Measurement Techniques Discussions

Interactive comment on "Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks" *by* Johan Strandgren et al.

Johan Strandgren et al.

johan.strandgren@dlr.de

Received and published: 8 August 2017

Reply to Anonymous Reviewer #3

We are very grateful to the reviewer for reading and reviewing our manuscript. Considering the reviewer's constructive comments clearly improved the quality of the manuscript. Each comment from the reviewer (roman style) is listed below along with the corresponding reply from the authors (in italic font style) as well as possible changes in the manuscript (in blue italic font style).

First we briefly want to address the general question raised by the reviewer on whether this manuscript could be joined with a recently submitted manuscript where the CiPS

algorithm is characterised. Even though the two manuscripts are thematically related we think it is reasonable to split them. The first manuscript (this one) describes the development of a new tool, CiPS, for the passive remote sensing of cirrus clouds from MSG/SEVIRI together with a comprehensive evaluation of its overall performance, a comparison to a similar retrieval, COCS, and an application to the cirrus life cycle. The overall CiPS and COCS performance is assessed against CALIOP, that represents our "truth", and the errors of all CiPS/COCS retrieved variables have been investigated as a function of the corresponding CALIOP quantities. Furthermore, the geographical distribution of the False Alarm Rate (FAR) and the latitudinal distribution of Cloud Top Height (CTH) errors are also discussed. The comparison to COCS enables to quantify the improvements achieved with CiPS. Thus, this paper contains in our opinion all essential information about CiPS and resembles, in its structure, the paper by e.g. Kox et al. (2014) and Holl et al. (2014). Altogether the first paper is more than 20 pages long.

In the second manuscript additional aspects are investigated, including the effect of the underlying surface type and the presence of liquid water clouds and aerosols below the cirrus on the cirrus cloud retrieval, the sensitivity to radiometric noise from SEVIRI, the relative weight of the single input variables, and the characterisation of the errors of the CiPS output quantities IOT (Ice Optical Thickness) and CTH as a function of both CALIOP IOT and CTH simultaneously. These aspects are investigated for CiPS alone since they represent a detailed characterisation of CiPS and are less important and less interesting for the "older and less accurate" COCS. The CiPS user can learn from this second paper additional useful information abut the algorithm and its performance in all these respects. Nevertheless, even though those aspects are investigated for CiPS alone, we consider this knowledge interesting for the broader cirrus cloud remote sensing community and it does not fit the topic of the first paper which, in a broad sense, is the development of CiPS. Altogether the second paper is around 20 pages long as well.

Kox, S., Bugliaro, L., and Ostler, A.: Retrieval of cirrus cloud optical thickness and top altitude from geostationary remote sensing, Atmos. Meas. Tech., 7, 3233-3246, 2014.

Holl, G., Eliasson, S., Mendrok, J., and Buehler, S. A.: SPARE-ICE: Synergistic ice water path from passive operational sensors, J. Geophys. Res. Atmos., 119, 1504–1523, 2014.

General comments

As indicated above, Section 3 could be shortened considerable. In fact, I think the section would become much more clear if the final net topology and training are simply presented "as given". If there is any general experience to draw from the tests performed to reach the final configuration, summarise these separately. The present detailed description of the tests just obscures the final outcome, and it is very hard to extract if there is any experience of general interest.

We are grateful to the reviewer for this good suggestion. We agree that the sections following 3.4.1 can be restructured and written more concisely in order to make the manuscript more clear and separate the training of CiPS from the comparison between different MLP structures. The Anonymous Reviewer #1 requested additional details about the determination of optimal meta-parameters, so in order to meet the requirements of both reviewers we have focused on restructuring and making this section more concise. First of all we have combined Sect. 3.4.2 and 3.4.3. We have also combined and shortened Sect. 3.5, 3.5.1, 3.5.2 and 3.5.3. We think that the comparison between different MLP structures will be interesting for readers that intend to develop ANNs for cloud remote sensing. However, those results are now presented separately at the very end of Sect. 3 ("Evaluation of different MLP structures") as the reviewer suggests. Please see the marked-up manuscript for details.

The preceding subsections in Sect. 3 (3.1–3.4.1) are already considered to be concise and we will keep them as they are unless the reviewer has specific suggestions on where we should shorten.

Some comments on terminology used around the ANN training. In "machine learning" one is usually supposed to work with three datasets: training, validation and test set. The training set should be used to train one or several methods, validation set should be used to select the best one, and finally the test set should be used to evaluate the final system. In the manuscript the authors seem to mix up these sets. They use the validation set to monitor the training of the ANNs and then use the test set to select the best one and evaluate the performance. This is probably not critical for large datasets, nevertheless from a conceptual point of view this is not very nice. The authors refer to those datasets as training data, internal validation data, and (final) validation data, respectively.

Since this is also relevant for the follow-up paper, I suggest the authors create an additional test set that is used exclusively for evaluation.

Again we thank the reviewer for raising an important point that was not described in a standard and clear manner in the manuscript. We do use the training, validation and test datasets that the reviewer refers to, and call them training, internal validation and final validation datasets in the manuscript. The reason for this was that in the meteorological field the term "test" has usually no clear meaning while the term "validation" has usually the meaning of the final evaluation of a dataset using an external independent dataset.

Furthermore, as the reviewer points out we falsely used the test dataset to select the optimal MLP structures after the first stage of the training as well as for the selection of the two classification thresholds. We have redone those steps using the validation *data and there are only marginal differences that don't affected our selection of MLP structures and classification thresholds. Thus, we modified the manuscript in this sense such that the test dataset is now used exclusively for the evaluation of the final ANNs. This makes the manuscript more correct from the conceptual point of view. Moreover, it is now clearly written that the validation datasets are used for the selection of MLP structures and classification thresholds and that the test dataset is exclusively used to evaluate the final performance.*

The authors should reflect upon if using CALIOP alone is the best choice for training data. Why not use some combined CloudSat and CALIOP retrievals, such as DARDAR? As far as I understand, the CALIOP lidar signal is quickly attenuated and I assume that the SEVIRI IR channels have a deeper penetration into the clouds. That is, CALIOP alone does not span a sufficient range of IWP. This problem should vanish if using e.g. DARDAR for training. This comment is a hint, no demand for redoing the work.

Initially, we have had a synergistic CALIOP/CloudSat retrieval (e.g. DARDAR, 2C-ICE) as training reference data in mind. But using IR observations only, SEVIRI does not penetrate significantly deeper into the cloud than CALIOP does. Krebs et al. (2007), show that the radiative signal from the brightness temperature differences between the SEVIRI window channels (commonly used for cirrus remote sensing) peak at an optical thickness around 2–3 and is quickly attenuated for increasing values. Using a synergistic CALIOP/CloudSat retrieval, we would probably be able to retrieve larger optical thickness and IWP. However the point of saturation is not as clearly defined for a mixture of IR channels as for a laser beam. When CALIOP is the limiting factor we can easily identify those clouds where the optical thickness and IWP are untrustworthy (opaque clouds), which is the idea of our opacity classification ANN. If SEVIRI would be the limiting factor, it would be more difficult to identify the transparent clouds where both SEVIRI and CALIOP/CloudSat can retrieve trustworthy results and the retrievals

for thicker cirrus clouds would get ambiguous. Holl et al., (2014) use synergistic CALIOP/CloudSat retrievals to train an ANN, but to avoid too large retrieval errors for thicker clouds they also use additional measurements in the microwave range from MHS (microwave humidity sounder).

The following sentence has been added to the end of the concluding section: "Although this manuscript is limited to CALIOP retrievals, one could investigate the usefulness of synergistic CALIOP/CloudSat retrievals as training reference data."

Krebs, W., Mannstein, H., Bugliaro, L., and Mayer, B.: Technical note: A new day- and night-time Meteosat Second Generation Cirrus Detection Algorithm MeCiDA, Atmos. Chem. Phys., 7, 6145-6159, 2007.

However, this touches upon the comment made above, that the errors of the CALIOP retrievals must be reported. These errors propagate directly into errors in the ANN product. As the test dataset is taken from the same CALIOP retrievals, the inherent CALIOP errors are not revealed. Conservative, quantitative, values on the dynamic range (i.e. coverage of optical thickness and IWP) and accuracy of the CALIOP product used for training shall be given.

We agree with the reviewer that it should be clearly stated that the ANN will inherit all errors of the training data. We have revised the manuscript accordingly and this is now discussed in both Sect. 3.3 (Output data: cirrus properties from CALIOP) as well as in the concluding section. The accuracy of the CALIOP products was briefly summarised in Sect. 2.2 (CALIOP) in the initial manuscript. This part has been extended considerably and moved to Sect. 3.3 in order to make the link to the CiPS/ANN retrievals more clear. The CALIOP V3 optical thickness and IWP products have quality status "provisional", this means that only "limited comparisons with independent sources have been made and obvious artefacts fixed". A full error char-

acterisation of CALIOP is partly available in the literature and has been summarized in the revised manuscript (see below). Regarding the dynamic range of the CALIOP products, we consider it partly covered by Fig. 2, but additional details on the lower detection limit of CALIOP have been added in the revised manuscript. The following text segments have been added to Sect. 3.3 and to the concluding section respectively:

"The CALIOP products are chosen as training reference data for CiPS as they should provide the most accurate estimates of especially CTH but also IOT for thin cirrus clouds from space. It is important to note that an ANN can never be better than its training reference and all deficiencies and/or biases in the training reference data will be inherited by the ANN. Since possibly inherited artefacts of the ANN will not show when validated against independent CALIOP retrievals, one must be aware of the accuracy and limitations of the training data.

Yorks et al. (2011) and Hlavka et al. (2012) validate the spatial and optical properties of cirrus clouds from the V3 CALIOP products using the airborne Cloud Physics Lidar (CPL, Mcgill et al., 2002) during the CALIPSO-CloudSat Validation Experiment (CC-VEX). CPL has a higher signal-to-noise ratio (SNR), higher vertical and horizontal resolution and lower multiple scattering compared to CALIOP, making it the most comprehensive tool for validating the CALIOP retrieved cirrus properties. Ten underpass flights with CALIOP were performed and over 9 500 bins of collocated extinction coefficients were obtained. During the ten flights, extinction coefficients ranging from approx. 0.001–10 km^{-1} and column optical thickness up to approx. 3 were retrieved. CALIOP and CPL agree on 90 % of the scene classifications (cirrus or no cirrus) on average. For all bins classified as cirrus by CPL, CALIOP agrees on 82 % and for the bins classified as cirrus free by CPL, CALIOP agrees on 91 %. For cases where both CALIOP and CPR detect cirrus, the agreement in cirrus top height is excellent (Yorks et al., 2011).

For transparent cirrus layers the agreement in IOT between CALIOP and CPL is

good with on average 15 % higher extinction for CALIOP (0.65 in correlation between CALIOP and CPL). For the unconstrained retrievals where the initial lidar ratio remains unchanged the average difference in extinction is only 7 % (0.80 in correlation between CALIOP and CPL, Hlavka et al., 2012). The latter are the ones used to train CiPS (see above), along with the constrained retrievals. At the time of the CC-VEX campaign (between July 26 and August 14 2006) the laser of CALIOP was pointing just 0.3 degrees from nadir leading to a strong specular reflection by layers of horizontally orientated ice (HOI) (Winker et al., 2009). This lead to disagreements in the extinction retrieval with CPL, whose laser pointed 2 degrees from nadir and therefore only received a very small fraction of specular reflections from the HOI (Hlavka et al., 2011). Since November 2007 the CALIOP lidar points three degrees from nadir in order to overcome this issue for layers with HOI. When the column optical thickness is derived for all cirrus covered bins, the relative difference between CALIOP and CPL is only 2.2 % due to cancellation of opposing CALIOP effects. Holz et al. (2016) recently showed that the single layer IOT derived from unconstrained CALIOP retrievals is low-biased with respect to a single channel thermal/IR IOT retrieval combining CALIOP/MODIS observations and forward radiative transfer modelling. The bias is shown to increase with increasing IOT.

The accuracy of the CALIOP IWC/IWP is directly related to the accuracy of the extinction retrievals as well as the IWC parametrization from Heymsfield et al. (2005). A proper independent validation of the CALIOP IWC/IWP is a difficult task due to the lack of reference data at a comparable spatial and temporal resolution. Protat et al. (2010) evaluate the IWC parametrization used for CALIOP for tropical cirrus using ground based radar-lidar retrievals. The results suggest that the parametrization is quite robust and is shown to work well at most altitudes. Above ∼12 km the IWC is clearly underestimated with respect to the ground based radar-lidar retrieval. Avery et al. (2012) evaluate the CALIOP IWC using coincident data from CloudSat and in situ measurements inside a tropical convective cloud. At the lower altitudes (8–12 km), the CALIOP IWC is underestimated with respect to the in situ measurements, which could *be attributed to a lower penetration depth of CALIOP and the removal of CALIOP layers containing HOI. Between 12–14 km the agreement between the CALIOP IWC and the in situ measurements is good. At all altitudes CALIOP seems to underestimate the IWC with respect to CloudSat. Wu et al. (2014) show that the V3 CALIOP IWC agrees well with airborne in-situ measurements up to approx. 20 mg m^{-3} at an altitude of 12 km. The CALIOP IWC agrees well with the CloudSat IWC within the regions where their sensitivities overlap. This occurs between 5–20 mg m^{-3} at an altitude of 12 km and between 30–200 mg m^{-3} at 15 km."*

"The minimum detectable backscatter of CALIOP depends on the scattering target (the cirrus cloud in this case), the altitude as well as the vertical and horizontal averaging of the data (McGill et al., 2007). Davis et al. (2010) show that CALIOP can detect approx. one third of the sub-visual cirrus clouds with an optical thickness below 0.01."

"The reported errors of CiPS are only with respect to CALIOP. Additionally CiPS, as an ANN, will have inherited any error that the CALIOP products have with respect to the true cirrus properties."

Further, there are also errors originating in the collocation procedure. Probably most important is the fact that CALIOP has a swath smaller than the resolution of SEVIRI. This results in that CALIOP covers only a part of the SEVIRI footprint, and this gives an additional uncertainty in the empirical relationships between CALIOP retrievals and SEVIRI measurements that the ANN is trained to represent. In any case, there are no comments at all of possible errors caused by imperfections in the collocation procedure.

We thank the reviewer for this comment as this important point was indeed under-represented. We have added a paragraph where we discuss errors that should be expected as a result of the different spatial scales and observation techniques. The following text segment has been added to Sect. 3.4.1: "When collocating SEVIRI and

CALIOP observations with the purpose of training an ANN one must consider two aspects. 1) Even though the 5 km average of CALIOP point measurements fits the spatial resolution of SEVIRI (3×3 km^2 at nadir and approx. 4×5 km^2 in mid-latitudes) quite well in the along-track direction, the two observations differ largely in scale in the across-track direction as the footprint of CALIOP is approx. 70 m wide at the Earth's surface. Consequently the 5 km CALIOP orbit segment is representative only for a relatively small fraction of a SEVIRI pixel. This will induce inevitable errors and lead to imperfect information used to train the ANN. This is especially relevant for partial cloud cover, where CALIOP may observe a cloud free area in an otherwise cloud covered SEVIRI pixel. If the error from imperfect collocations is random, this will have a limited effect on the ANN. Only if there is a recurrent systematic difference as a result of the different spatial scales this will be lead to biased retrievals (Holl et al., 2014). 2) Although cirrus clouds leave their mark on both SEVIRI and CALIOP measurements in a similar way, SEVIRI does not share CALIOP's possibility of discerning vertically separated ice clouds, liquid water clouds and aerosols. Consequently SEVIRI should not be expected to discern the signal from liquid water clouds and aerosols when retrieving the IOT as effectively as CALIOP."

The points raised in the last two paragraphs, are they considered in the new manuscript targeting errors?

No, they have instead been added to the revised version of this manuscript.

As ANNs do not provide a direct uncertainty estimate for the retrievals we have considered several approaches to characterise the errors/uncertainties, however we have not found an approach that is more representative than the overall statistics in this manuscript. Below we list some uncertainty estimate approaches together with their limitations. 1) One approach is to train a second set of artificial neural networks (ANNs) to retrieve the uncertainty reported by CALIOP, this would however require that

CiPS retrieves values that are identical or very close to those retrieved by CALIOP. As seen in the manuscript, this is not always the case. 2) A second approach is to run CiPS multiple times with small perturbations in the input data. This is however only representative if 100 % of the retrieval error can be attributed to noise and uncertainties in the input data. This is however not the case as most of the retrieval error of CiPS is likely to stem from the different sensitivities of SEVIRI and CALIOP. This is covered in the second manuscript, but then as a noise sensitivity analysis rather than an uncertainty estimate of CiPS. 3) A third approach is to train a second set of ANNs trained with the absolute difference between CiPS and CALIOP (using the same collocation dataset used to train CiPS). The retrieved accuracy would however be a statistical uncertainty learned using several training points and we consider the added value, in comparison to the characteristics on the overall performance presented here, to be small. So rather than estimating individual uncertainty measures of each retrieval we focus on performing a detailed characterisation of CiPS that we present here and in the second manuscript. Please note that most of this paragraph was taken directly from the reply to Anonymous Reviewer #1.

Specific comments

p 5, l 23: Also the bias neurons need to be assigned the correct values.

Revised.

Sec 3.3: The section fails to clearly report what spatial resolution that is applied. For me this became clear first when reaching p 12, l 31. As 5 km anyhow is used, is the main discussion in Sec 3.3 actually needed? It seems to refer to an older version. That is, this section could be shortened.

This was indeed not very clear. Since the Anonymous Reviewer #1 asked the same

questions we include the same reply here as well: We use the product with a reported spatial resolution of 5 km. But to detect faint cirrus and aerosol layers, the CALIOP team has to average over several consecutive 5 km profiles in order to get a sufficiently high signal-to-noise ratio. This means that in the 5 km cloud layer product, some cirrus were detected using a spatial resolution of 20 or even 80 km. In such a case the 5 km layer product will have 4 or 16 consecutive bins where the cirrus properties are identical. The additional spatial resolutions of 20 and 80 km can be seen as "background resolutions" used by the CALIOP team. This has been clarified by extending the paragraph, which now reads as follows: "Even though the cloud and aerosol layer product are reported with a spatial resolution of 5 km, two additional coarser resolutions of 20 and 80 km are used to detect the cloud and aerosol layers reported in the 5 km products (Vaughan et al., 2009). At a spatial resolution of 5 km, the signal-to-noise ratio of a faint cirrus or aerosol layer is usually too weak to be distinguished from the clear-sky atmospheric signal. By averaging 4 or 16 consecutive 5 km profiles the signal-to-noise ratio is increased, which allows for detection of very thin cirrus and aerosol layers. For example if a thin cirrus cloud with an optical thickness of 0.1 and a top altitude of 10 km is identified only when 16 consecutive 5 km profiles are averaged (80 km spatial resolution), 16 consecutive bins in the L2 5 km cloud layer data will report an optical thickness of 0.1 and a top altitude of 10 km."

p 13, l 24: This is the only place where the authors mention the activation functions of the networks. This information should not be given below "Training data", but as indicated below. Further, why don't the authors use identity activation functions on the output layers of the regression ANNs instead of re-scaling, which would be the more common approach? This could also have an effect on the learning of extreme values since the gradient vanishes at both limits of the output range.

The information about which activation functions are used has been removed from this section and shifted to the training section as proposed by the reviewer.

Our experience shows that using the identity function for output leads to competitive results. The only thing you gain is faster training in some cases, but training instability in others (especially if the range of the output data is not reduced). We do however avoid the extreme values when we scale the output data and use the intervals [0.1,0.9] for the sigmoid activation function and [-0.9,0.9] for the tanh activation function. We also tried more conservative scaling to the more linear part of the activation functions, but with no apparent improvement. This is equivalent to using a linear output as proposed by the reviewer. In our opinion the problem with the learning of extreme values would remain also with the identity activation function as this problem is caused by inevitable inconsistencies within the training data due to the different sensitivities of CALIOP and SEVIRI together with the fact that we minimise the squared error, meaning that the model can be wrong only in one direction at the extreme values.

p 15, l 1-10: When introducing the structure of the networks the authors should mention which activation functions are used in the hidden layers and the output layer.

Revised. The following sentence has been added to Sect. 3.5: "For the classification ANNs (CCF, OPF) the sigmoid activation function is used for both hidden and output layers, whereas the tanh activation function is used for hidden and output layers for the regression ANNs (CTH, IOT & IWP)."

p 15, l 28-29: 1 CPU@3.4 GHz does not describe the computer sufficiently especially since ANN inference is highly parallelizable. The authors should at least give number of cores and processor model.

Revised. This part now reads as follows: "using 1 core à 3.40 GHz, Intel Core i5-3570". We also added the following sentence to clarify that the computation time can be reduced: "ANN computations are highly parallelizable, meaning that the computation

time can be reduced significantly by distributing the computations across multiple cores."

p 16, l 25: Figure 3 does not seem to compare the performance after the final training to the performance before, so the reference here seems pointless.

After the changes made to Sect. 3 (see above), this reference no longer exists.

p 17, l 5-13: The detection threshold for the CCF is a parameter of the classification ANN and is thus prone to overfitting. Its value should not be determined based on the performance on the test set. This touches upon the general comment above. Also,a plot of the POD against FAR for different thresholds would be good as it gives an additional perspective on the performance of the classifier.

We thank the reviewer for pointing this out. As mentioned above we have redone this step using the validation dataset (referred to as internal validation data in the manuscript). The results are nearly identical to those we had with the test dataset (referred to as final validation data in the manuscript). Consequently, we choose the same thresholds as initially done with the test dataset i.e. 0.62 for the cirrus cloud detection and 0.86 for the opacity classification. The results are so similar that the numbers for the overall POD and FAR mentioned in this section remains the same.

A figure showing the POD and FAR as a function of the classification threshold (also know as the receiver operating characteristic (ROC) curve) has been added, as proposed by the reviewer.

p 18, l 4: "the values corresponds" should be "the values correspond"

Revised.

p 18, l 9: Also here the authors claim the test data was excluded from the training but it seems that it has been used for the selection of the network structure and meta parameters.

As explained above, the test dataset is now excluded from the training/development of CiPS and exclusively used for the final validation presented in Sect. 4.2.

p 19, l 3 - 6: See comment p 17, l 5 - 13

A reference has been added to the new figure showing the FAR and POD as a function of the classification threshold.

p 21, l 5: "might seems high" should be "seem"

Revised.

p 21, l 7 - 14: The authors should explain in more detail what they mean with uncertainty and solution and/or provide a reference for their claims on the behaviour of ANNs.

*With **uncertainty** we mean the situation where a set of similar x-values, in the simplified ANN function f(x)=y, corresponds to quite different y-values. In such a situation it is not possible for the ANN to accurately model a relationship between x and y. Instead the ANN will output a mean value over the distribution of the most likely y-values that this x-value represented during the training, weighted by their probability. We call those y-values **solutions**. This segment has been rewritten and clarified and those terms are no longer used in the revised manuscript. Please see the response to p21, l 15 - 21 below.*

p21, l 15 - 21: Isn't that the reason for the "uncertainty" mentioned in the paragraph (see comment p21, l 7 - 14) in the training data? (I.e. low signal to noise ratio for thin clouds?)

Yes, the reviewer is correct. This (the different sensitivities of CALIOP and SEVIRI) is indeed the reason behind the "uncertainty" mentioned above. We clarified this part accordingly and extended it with additional discussion and comments on the retrieval errors of CiPS, as suggested by the reviewer. Furthermore this part has been detached from the general discussion about the results and is now located at the very end of Sect. 4. Consequently, redundant parts, previously located earlier in that section, have been removed. The segment where the retrieval errors and the reason for systematic over- and underestimations are discussed now reads as follows: "As expected and as seen in Fig. 9, 11 and 12, CiPS is not able to perfectly model the CALIOP cirrus properties using the SEVIRI, ECMWF and auxiliary data. There are several sources of error that add to the final performance of CiPS. Most importantly CALIOP and SEVIRI have different sensitivities to cirrus clouds. This is especially clear for thin to sub-visual cirrus clouds where CALIOP is able to accurately retrieve the top height and optical properties. Such faint cirrus leave a considerably weaker or no mark on the SEVIRI observations though, making it difficult to inversely determine the cirrus properties. Similarly the CTH is not necessarily defined equally by CALIOP and SEVIRI, as CALIOP is able to discern thinner icy layers at the cloud top, that may appear as "invisible" to SEVIRI. Also for thicker cirrus clouds where both CALIOP and SEVIRI (thermal observations) approaches the point of saturation, the different sensitivities lead to ambiguous collocations. When an ANN is trained with a set of different output values that correspond to approximately the same input data as a result of the lower sensitivity, the ANN will not be able to model an accurate relationship. The reason for this is that the input vector contains no information on how the difference in sensitivity affects the target values. This can be regarded as an unknown hidden variable. This is not an ANN *specific weakness, but applies to all regression models minimising the squared error. When such a set of incomplete input data (in the sense that there is a strong hidden variable) is given to the final ANN, it will output a conservative mean value that can be understood as an average over the distribution of the most likely solutions weighted by their probability. The larger the difference in sensitivity the higher will the variance within the distribution of the most likely solutions be, leading to larger retrieval errors. Throughout most of the output data range this error will be random. But obviously, the distribution of the most likely solutions cannot be centred around the extreme values leading to systematic over- and underestimations of low and high output values when a conservative mean value is calculated. This effect increases towards the extreme values as the desired output value is skewed towards the edge of the distribution of the most likely solutions. This effect is clearly seen in Fig. 11c and 12c where low and high IOT_{CALIOP}/IWP_{CALIOP} are over- and underestimated respectively. This is to some extent also seen for the CTH_{CiPS} retrieval in Fig. 9c, especially for low CTH_{CALIOP}. Due to the randomness of the effects a lower sensitivity introduces, adding information about the magnitude of the sensitivity to the input vector is not likely to improve this situation. The larger CTH_{CiPS} retrieval errors observed for low clouds can also be attributed to the smaller temperature contrast with respect to the surface temperature and thus the weaker radiative signal that those clouds have compared to higher cirrus clouds. Another source of error that amplifies the effect discussed above, is the risk that there are additional variables relevant for finding an accurate relationship that are not represented by the vector of input data.*

As discussed in Sect. 3.4.1 imperfect collocations as a result of the different spatial scales of CALIOP and SEVIRI together with partial cloud cover or spatially inhomogeneous clouds will further add to the retrieval errors. In a situation where CALIOP observed a small optically thin area of an otherwise optically thick cirrus inside a SEVIRI pixel, CiPS is likely to overestimate IOT_{CALIOP} and IWP_{CALIOP}. Similarly if CALIOP observed a small optically thick area of an otherwise optically thin cirrus inside a SEVIRI pixel, CiPS is likely to underestimate IOT_{CALIOP} and IWP_{CALIOP}."