Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks

— RESPONSE TO REFEREE 2 —

contact: hendrik.andersen@kit.edu

We would like to thank referee 2 for her/his review of the manuscript and her/his constructive criticism. Comments by the referee are colored in blue, our replies are colored in black.

This paper addresses a topic of significant current research, namely quantifying the effect of aerosols on cloud properties. The authors note the importance of local meteorology in determining the properties of clouds and that as meteorological factors are also correlated to aerosol properties, this can obscure the influence of aerosols on cloud properties. To explore the role of meteorology and aerosols, they make use of an artificial neural network (ANN) to examine the sensitivity of cloud properties to different predictors. Similar to previous studies, they show that meteorology is a strong control on the cloud properties, such that the cloud properties can be accurately predicted on a monthly timescale using reanalysis data and observed aerosol properties.

I think that this paper is a good addition to the literature on this topic, presenting a new way to investigate the drivers of cloud properties. However, there are a couple of points, listed below, that I think should be clarified before publication. I particular, I think that using monthly data rather than daily/instantaneous data must be better justified. It would also make the paper stronger if the ANN method was compared to a more comparable statistical technique, such as a multiple linear regression across meteorological parameters. This might help to highlight the benefits of using an ANN, especially if it results in a different sensitivity of cloud properties to aerosol. Following these changes, I feel that this article would be suitable for publication in Atmospheric Chemistry and Physics.

We respond to each point individually below.

Main points

1) While some previous studies have used monthly data for investigations into aerosol-cloud interactions, this disguises a lot of the variability in the cloud field and focuses on very large scale changes in cloud properties. The effect of seasonal variations can generate non-causal relationships between cloud properties and meteorological factors that might be accounted for if the study was done on a sub-seasonal scale using higher temporal resolution data. Can the authors explain why monthly data is used in this case and why daily data is unsuitable?

With this study, we specifically aim at analyzing the aerosol-cloud-climate system at a very large scale ('system scale'). The monthly time scale is used here, as a) this enables a focus on the large-scale patterns and relationships and b) GCM output is also at a monthly time scale, so that future comparisons between our observationally-based results and GCMs can be conducted. We acknowledge the 'non-causal relationship' argument by referee 2 by using only a very limited number predictors in ANNs that have previously been shown to be the main drivers of liquid-water clouds. The results of the ANNs are physically plausible (signs, magnitudes and regional patterns of the sensitivities) and give another line of independent evidence that strengthens the confidence in our current system understanding. That being said, we cannot exclude the possibility that some of the observed relationships might be in part non-causal (which is true for other averaging time scales as well).

2) The use of an ANN seems to give a large improvement over just using AOD as a predictive variable for cloud properties. However, I am not sure this is a suitable comparison, as AOD is rarely assumed to be a good predictive variable for cloud properties on its own. As better comparison would be the predictive

ability of (log) AOD on its own using a linear regression and from the ANN. Alternatively a comparison of a multiple linear regression and an ANN for predicting the cloud properties could show the added utility of using an ANN over existing methods. This might then highlight further useful properties of the ANN - for example, does it show a stronger (or weaker) sensitivity of cloud properties to aerosols when compared to current methods?

We probably did not communicate the intention of this figure with sufficient clarity: This figure was simply intended to show how well a combination of aerosol and meteorological conditions can explain the variance of cloud properties (multi-variate statistics) as opposed to a simple bivariate approach. We have added results of a multiple linear regression using all the ANN predictors to the figure as suggested to illustrate the skill of the ANN vs. another multivariate method. The comparison of the results of the multiple linear regression and the ANN suggests that the ANN is an appropriate method to be used in this context. As suggested, we have switched from using the AOD to the AI and used log(AI) for this figure.

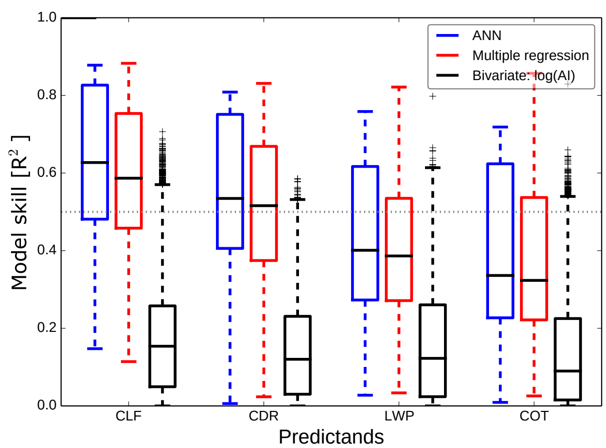

Figure 1: Predictand correlation with ANN (multivariate) test output, multiple linear regression (multivariate) and log(AI) (bivariate).

3) How do regional ANNs compare to a single global model? Presumably if enough meteorological parameters can be included, a single global model should be able to predict cloud properties everywhere. Requiring different models in different locations would then indicate that some meteorological parameter is missing from the ANN. A global pattern of the accuracy of the ANN might then give an indicator as to which parameters should be included. The ANN might be expected to differ as a function of cloud type, but perhaps a separate model for each cloud type (e.g. Gryspeerdt and Stier, 2012 or Oreopoulos et al., 2016) might be useful.

If one trains a single global model to predict CLF, using the same predictors and model setup as for the regional ANNs, it cannot predict CLF as well as most regional ANNs (R^2 of global model $\approx$ 0.45; median of regional ANNs $>$ 0.60). While adding additional predictors to the global ANN could still improve the skill of the model, it is unrealistic to think that a single model could represent clouds as well as regional models can (it would also increase the probability of non-causal relationships). Regional ANNs are superior, as they are able to reproduce the regionally varying predictor-predictand relationships (c.f. fig. 6 in the manuscript). These regional differences would be blurred or missed completely when using a single global ANN. Regional ANNs also have the advantage that knowledge on typical regional characteristics (e.g. aerosol species composition) can be included in the interpretation of the results (as in Andersen et al., 2016). That being said, cloud type-specific ANNs seem to be an interesting idea for future work.

Minor points

P2L9: Perhaps only e.g. is necessary

We agree and have changed the manuscript accordingly.

P2L24: Why is the 2.1um effective radius used with the 3.7um LWP retrieval?

We have changed the cloud products used (see our response below).

P2L29: Is the liquid fraction a suitable measure of cloud fraction, as it depends on the overlying ice cloud fraction? The authors could consider using cases where only liquid cloud exists in a gridbox, as this would remove this source of uncertainty.

After internal and peer discussions, we have decided to run the ANN with monthly means of single layer clouds only. While the results are nearly identical, the argument is valid, so that we only use single layer cloud products the current version of the manuscript.

P3L4: AOD is proportional to CCN (at least at some scales, see Andreae, 2009), it is just not a direct measurement (the same as with mass, as it also depends on aerosol optical properties)

Yes, we agree. We have corrected this in the revised manuscript.

P3L7: Many recent studies have used aerosol index (AOD times angstrom exponent) or a reanalysis aerosol parameter (e.g. Lebsock et al., 2008; McCoy et al., 2016). As these have been shown to more accurately predict cloud properties, they might further improve the skill of the ANN. Although MODIS AI is not necessarily accurate over land (Levy et al., 2013), it could be used over ocean in this study.

For this study, we used the newest version of MODIS products available, collection 6 (C6). In C6, the MODIS Ångström exponent (needed for the computation of the aerosol index as it is the product of AOD and the Ångström exponent) has been discontinued in level 3 (L3) data (p. 3018 Levy et al., 2013). We believe that for this and for other reasons, other recent studies also use the AOD as a proxy for CCN (e.g. Chakraborty et al., 2016; Stathopoulos et al., 2017; Patel et al., 2017). We agree with the referee though that the aerosol index might be a more appropriate measure for CCN and have thus chosen to compute the Ångström exponent (550 and 867nm) ourselves to use aerosol index instead of AOD in the ANN. The following figures 2 and 3 are the new results of the ANN when using AI instead of AOD. The spatial patterns in ANN skill, as well as the mean global sensitivities are nearly identical (cf. figures 3 and 5 in the original ACPD manuscript).

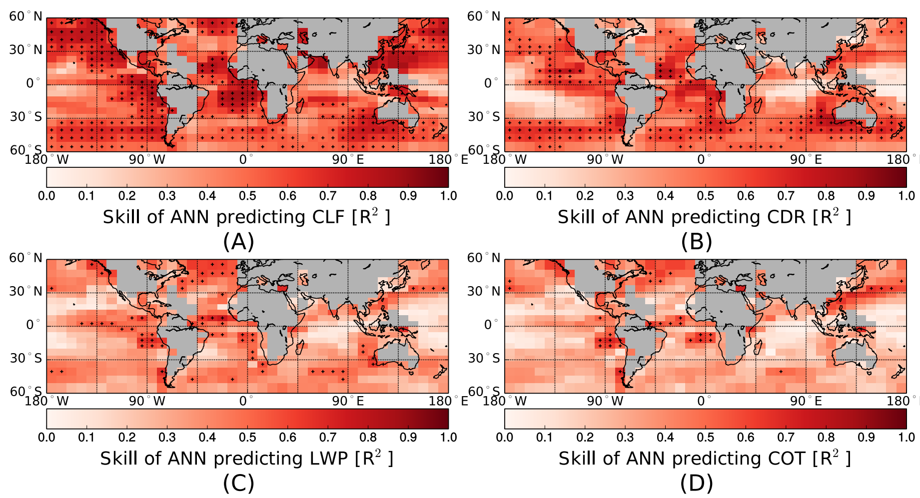

Figure 2: Global patterns of ANN skill as in the manuscript; AI has been used instead of AOD.

Small differences can be observed in the regional patterns of ANN sensitivities (fig. 4) to AI vs. AOD. The CLF sensitivity to AI is higher in the Southeast Atlantic than its sensitivity to AOD in that specific region. The Southeast Atlantic is of course dominated by biomass-burning aerosols, which are mostly fine mode and thus feature a relatively larger AI than AOD. The sensitivity of CDR

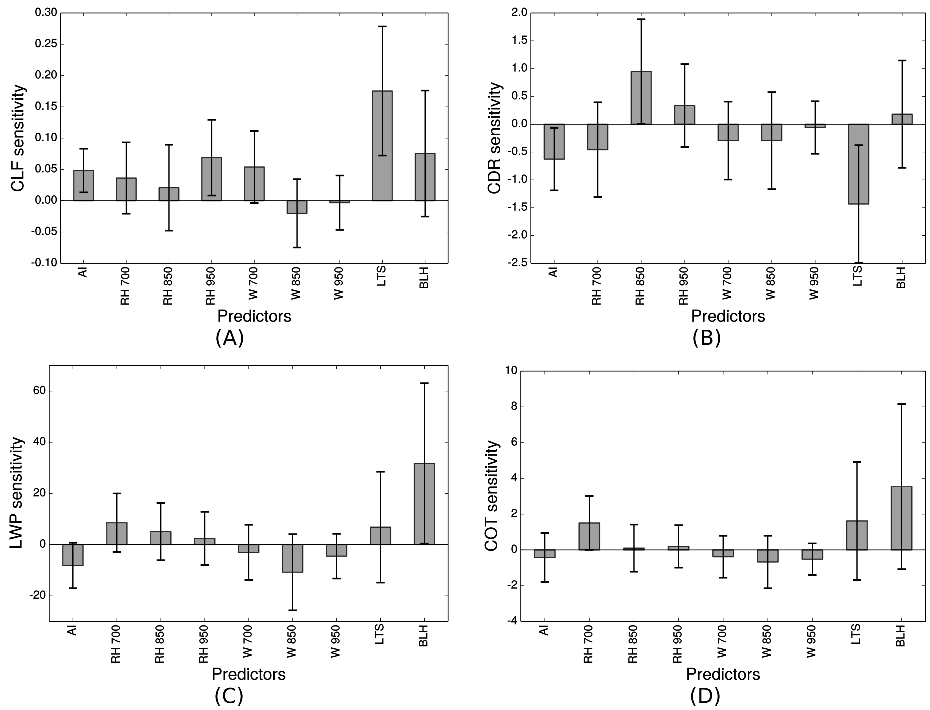

Figure 3: Global mean relative sensitivities as in the manuscript; AI has been used instead of AOD.

to AI differs from its sensitivity to AOD in regions that are dominated by desert dust. Dust is relatively coarse, so that the AI would be disproportionally lower than the AOD in these regions, which might explain the differences between the sensitivities of the two.

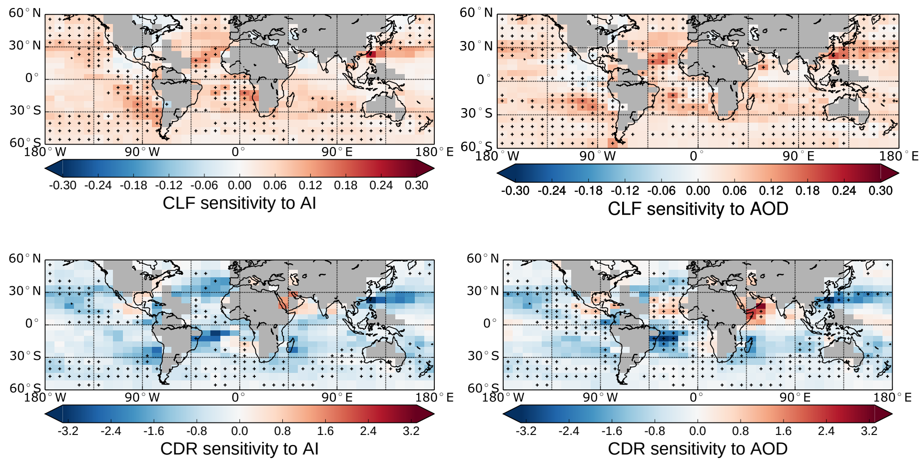

Figure 4: Differences in sensitivities of CLF and CDR to AI (left-hand column) vs. AOD (right-hand column).

P3L13: It is definitely a good idea to investigate variables that have been previously used in aerosol-cloud studies. Koren et al., (2010) might also provide some useful guidance here. Although it was focussed on looking at convective clouds, some of the results (e.g. Figs. 8,9) might help decide which variables should be included in the ANN).

We agree that additional variables (e.g. geopotential height, horizontal winds) might improve the ANN performance in some regions. Our goal in predictor selection was to minimize the number of predictors to a few key variables, in order to prevent covaration between the predictors. Also, additional predictors increase the probability of highlighting non-causal relationships.

P4L33: Is there any significance behind using five hidden nodes?

After thorough testing, five hidden nodes appeared to be a good global number. In general, the optimum number of nodes is dependent on the problem at hand. The number of nodes needed is connected to the complexity of the relationships, the amount of noise in the data and the amount of training data available. Too many nodes can lead to overfitting and poor generalization, whereas the ANN may not converge to a global minimum when too few nodes are used (Gardner and Dorling, 1998). We found that while regional ANNs may differ, five nodes where a reasonable choice, as additional nodes typically only marginally, if at all, increased model skill. To illustrate this, figure 5 is an example of the effect of the number of hidden nodes on ANN skill in the Southeast Atlantic region. This figure is obviously not the basis for our decision to use 5 nodes, but is intended to illustrate a typical example for the dependence of a regional ANN skill on the number of hidden nodes.

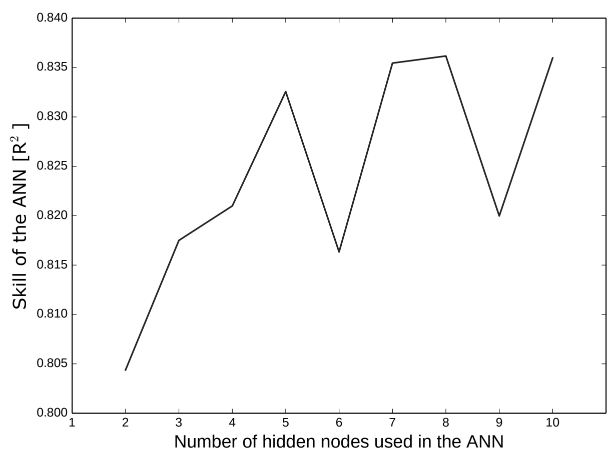

Figure 5: Example (Southeast Atlantic) for the effect of the number of hidden nodes in the ANN.

P5L7: Are the sensitivities calculated using the local variation of meteorological values, or the same artificial values globally? If the relationship is non-linear and the mean values of the meteorological variables vary across the globe, this could strongly affect the calculated sensitivity.

This sentence was intended to describe how sensitivities can generally be computed with an ANN. In the text passage further down (P5L14), we describe how sensitivities are computed in this study. To answer your question: Yes, the sensitivities are calculated using the local variation of meteorological values ('grid cell specific mean values'). In the revised version of the manuscript, we will attempt to describe both text passages more clearly.

P5L14: I am not sure I understand this sentence (which might explain my previous query?)

We compute ANN-predicted outputs for two groups of input data:

- All grid-cell specific retrievals of a specific predictor smaller than its 25th percentile.
- All grid-cell specific retrievals of a specific predictor greater than its 75th percentile.

In all cases, all other predictors are held constant at their grid-cell specific mean values. We then compute the average of both groups of ANN-predicted outputs. The difference between the two averages is defined as the sensitivity of the predictand to the specific predictor that was varied. We will try to more clearly describe this in the revised version of the manuscript.

P5L20: If the other meteorological factors in the ANN are held constant, does this produce a different result for the simple sensitivity? (see main point)

We have tested this for the sensitivity of CLF to AI. As above, we have also used data from the the Southeast Atlantic for this example. We found that the sensitivity (linear slope of AI-CLF relationship) of CLF to AI is $\approx 40\,\%$ lower in the ANN than in the observations. This is, of course, because in the sensitivity of the ANN, the other predictors are held constant, constraining their effect on CLF. This corresponds rather well to Gryspeerdt et al. (2016) who found that the sensitivity of CLF to AOD is reduced even further ($80\,\%$) when including information on CDNC along the causal pathway of the AOD-CLF relationship.

P6L7: As I understand it previous work focusses on the sensitivity as this is related to the strength of the cloud response to aerosol. It is not often assumed that aerosols can explain much of the variability in cloud properties which might explain the low skill here.

Yes, we agree. This figure is not intended to illustrate sensitivities, but that we are in a space of large uncertainty when we derive sensitivities using bivariate methods. Using a multivariate approach (also the case for multiple regression, as outlined above) we are capturing more of the aerosol-cloud climate system. The derived sensitivities might thus be more reliable.

P7L1: Perhaps another measure of skill might be useful in addition to the R^2? It could be argued that the skill in the shallow cumulus regions is quite good, in that the ANN (presumably) gets the cloud properties roughly right (the rms error might be small)?

Yes, indeed, we also looked at the relative RMSE. Actually, the a combination of relative RMSE and R^2 thresholds (P7L4) are used to select the regions that are used for the computation of sensitivities (marked with a '+' in the maps). The relative RMSE and R^2 are basically invertly related.

P7L4: Does this removal of the poor skill models bias the results, perhaps as a function of meteorology (as would appear to be the case from the maps in Fig. 3)

The computed sensitivities are only valid for the regions and are not intended to be "global" in that sense.

P7L9: How does these sensitivities compare to previous results? Several studies have calculated AOD-CF or AOD-droplet number concentration sensitivities which could be compared here (e.g. Quaas et al (2008), Grandey et al. (2012), Gryspeerdt et al. (2016))

We compute the sensitivity a slightly different way, so a straight-forward comparison is not possible. However, in a similar way that Gryspeerdt et al. (2016) constrain the aerosol-CLF relationship with CDNC, the ANN constrains the aerosol-cloud relationships by meteorology. In the updated version of the manuscript, we will include comparisons to sensitivities found by other recent studies.

P12L3: Are the covariations really spurious? The argument here is not that the covariations don't exist, but that they are not representative of the causal relationship. I would suggest that if 'direct physical relationship' was replaced with 'causal relationship', this could instead mention the issue of confounding variables, similar to Gryspeerdt et al.,(2016).

We will restructure this text passage in the updated version of the manuscript.

P12L4: To what extent has using RH in the ANN accounted for this effect?

As shown in figure 6 within this document, the sensitivity of CLF to AI is weakened in the ANN, probably due to the meteorological constrains of the model. These are hard to track down to a single predictor, though (e.g. RH). It is likely that the main confounding factor for this relationship is RH and that most of the change in AI-CLF sensitivity is due to constraining RH.

References

Andersen, H., Cermak, J., Fuchs, J., and Schwarz, K. (2016). Global observations of cloud-sensitive aerosol loadings in low-level marine clouds. *Journal of Geophysical Research: Atmospheres*, 121(21):12936–12946.

Chakraborty, S., Fu, R., Massie, S. T., and Stephens, G. (2016). Relative influence of meteorological conditions and aerosols on the lifetime of mesoscale convective systems. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7426–7431.

Gardner, M. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14):2627–2636.

Gryspeerdt, E., Quaas, J., and Bellouin, N. (2016). Constraining the aerosol influence on cloud fraction. *Journal of Geophysical Research: Atmospheres*, 121:3566–3583.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C. (2013). The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, 6(11):2989–3034.

Patel, P. N., Quaas, J., and Kumar, R. (2017). A new statistical approach to improve the satellite based estimation of the radiative forcing by aerosol–cloud interactions. *Atmospheric Chemistry and Physics Discussions*, 17:3687–3698.

Stathopoulos, S., Georgoulias, A., and Kourtidis, K. (2017). Space-borne observations of aerosol - cloud relations for cloud systems of different heights. *Atmospheric Research*, 183:191–201.