



Interactive  
Comment

## ***Interactive comment on “A refined statistical cloud closure using double-Gaussian probability density functions” by A. K. Naumann et al.***

**A. K. Naumann et al.**

ann-kristin.naumann@mpimet.mpg.de

Received and published: 12 August 2013

We thank the referee for his/her comments on the revised version of the manuscript. Because we think that it is a useful discussion, we would like to make some more general remarks on model selection, a priori and a posteriori testing before we address the reviewer’s comments individually.

Concerning model selection, we fear that the discussion among ourselves, the editor, and the reviewers has become derailed by a secondary point. We do not claim that we can, based on our study, decide which parameterization is in general better (although we can probably rule out the single-Gaussian scheme for cumulus convection), and we do not mean to suggest that rigorous statistical model (or parameterization) selection

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



can or should be done based on the error measure presented in our study. The decision which parameterization is better (for a certain model application) can only be made after a posteriori testing with independent data, see e.g. Pope, 2000, p. 601, who writes

“It is natural and appropriate to perform a priori test to assess directly the validity and accuracy of approximations being made. However, for the LES approach to be useful, it is success in a posteriori tests that is needed”.

A priori testing such as we perform is useful to understand the behavior on the individual parameterizations in simple idealized conditions and with perfect input data. In our example this testing gives only some indication that the old closure equations and coefficients of the Larson et al. (2001) scheme are not optimal for our data. We give some physical reasons why we think that the closure relations should not be anti-symmetric (a property that the original Larson scheme has). We also provide some analysis on the scale-dependency of the scheme, quantify its range of applicability and give some insight into the challenges of sub-grid autoconversion.

Such qualitative properties of the physical behavior is probably what Pope means when he talks about a priori testing being useful to “asses the validity and accuracy of approximations”. But this should not be confused with model selection. Quite often this step of a priori testing is not done as thoroughly as we do it here, e.g. including an attempt to characterize the scale-dependency of the parameterization. For example, Tompkins et al. (2007, QJRMS) do not do any a priori testing and go immediately to a posteriori testing. This is quite common in operational numerical weather prediction (NWP), but can be prone to compensating errors. Still, a posteriori testing is essential for an understanding of the parameterization when it is coupled to the full model and exposed to more complicated real-case meteorological situations. For NWP or climate models this should be done on several levels, e.g., in a single-column model as well as in full 3D real-case simulations (in NWP including data assimilation etc.). Model (aka parameterization) selection can only be made by taking all results into account.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

It may be that by providing so much detail we have over-emphasized the a priori testing, but we feel these results are generally interesting. But to perform statistical model selection (i.e., testing on independent data) and/or to recalibrate older parameterizations to our LES data, as our reviewers suggest, would compound this problem and further emphasize a priori testing. We note, too, that this path is rare in the atmospheric sciences.

That all said, if there are counter arguments to this point of view we would welcome learning from them, and if there are not relevant counter arguments in the literature, but the reviewer nonetheless feels strongly about the topic, perhaps the debate would be better served if this point (which we do not believe is central to the manuscript) is aired in a comment/reply, which the literature anyway does not have enough of.

In the following we address the comments of the referee individually. The referee's comments are in *italics*, authors answers are in normal font.

*p. 1: One reason for using cross-validation in this case is that the manuscript does model selection. That is, the manuscript attempts to compare the errors in several parametrisations. Therefore, the errors reported should be comparable between parametrisations, i.e. they both should be out-of-sample (generalisation) errors.*

For the topic of model selection, please see our comment at the beginning of our reply. In a revised version of the manuscript, we put more emphasis on the discussion on model selection and stress that we do a priori testing only. For example, we add in Sect. 4 “Note that the usefulness of a priori testing is in the assessment of validity and accuracy of the parameterizations assumptions (e.g., Pope, 2000). To decide which parameterization is most useful in a certain NWP model or GCM a comparison based on a posteriori testing has to be done.” Also, we agree that data separation is reasonable here. We therefore do a more rigorous separation in training and testing

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

data in a revised version of the manuscript as shown, e.g., in our answers to the last three comments of the referee.

*p. 2: When the authors tune the closure equations of Larson et al. (2001, L01, Eq. 3 in the GMDD manuscript), they have tuned only one of two parameters. Namely, they have tuned gamma but they have kept alpha=2. Furthermore, negative variances can be avoided by insisting that gamma<1 and alpha>0, as can be seen by inspection of Eq. 3. The authors have not attempted to tune the parametrisation of Cuijpers and Bechtold (1995), even though a parameter could be introduced as a prefactor to the exponent and another within the argument of exponent.*

We have tried to make it clear that our contribution is a re-calibration of the Larson et al. (2001) parameterization based on our LES data. We are not suggesting a completely new scheme, but just a slight modification of an existing parameterization. In one regime (positive skewness) we replace one of the closure relations by a very simple linear function which provides a good fit to our data. This is where the “refined” in the title of our manuscript originates from. In this context we do not see the point of testing more complicated relations.

*p. 2: Table 2 of the revised manuscript lumps together the errors from the training datasets (RICO and DYCOMS) and the generalisation datasets (ASTEX and ARM) for the new parametrisation. It also includes the data from L01, CB95, and the new parametrisation. This is misleading, because the errors from the training datasets are not comparable with generalisation error. Either the RICO and DYCOMS errors should be omitted from the table, or else the data from those two cases should be presented separately in a way that does not compare training and generalisation errors.*

We split the table, to separate the different datasets. Also we tried to point to the differences more clearly in the text. Please see, e.g., our answer to the last comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

of the referee.

*p. 5: The authors' response here conflates inputs and tunable parameters. The inputs are the mean, variance, and skewness, and those cannot be tuned. The tunable parameters are coefficients like gamma and alpha that can be tuned. The new parametrization has more tunable parameters than the older ones. More tunable parameters can sometimes lead to less robust behaviour when tested in very different data (e.g. congestus clouds). However, there are statistical methods to fairly compare formulas with different numbers of tunable parameters, such as the Akaike Information Criterion and the Bayesian Information Criterion.*

The additional tuneable parameters in the closure equations for both the parameterization of the cloud cover and the average liquid water ( $\gamma_n$ ) and the parameterization of the buoyancy flux ( $a$  and  $b$ ) result from the introduction of a functional form that is (slightly) different from the functional form suggested in literature (here, Larson et al., 2001 and Cuijpers and Bechtold, 1995). The different and slightly more complicated relations that we use have a clear physical motivation and are well supported by our data. As discussed above, model selection can only be done by taking a priori and a posteriori testing into account. In this context, we think that the suggested statistical methods should be considered in a posteriori testing.

*p. 7: The revised manuscript should list the number of iterations required for convergence of the equation involving the relative weight,  $a$ .*

In a revised version of the manuscript, we added "... where Eq. (5) may be solved numerically for  $a$ . Alternatively, to avoid an iterative solution for a more computational efficient implementation in a GCM or an NWP model, an (e.g. polynomial or matched asymptotics) approximation of  $a$  as a function of  $sk$  can be used. For the present analysis however, we solve for  $a$  numerically using a simple bisection method with an

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



accuracy of  $10^{-6}$  which typically took about 30 iterations.”

*In the revised manuscript, the following passages make comparisons based on training data, whereas the comparisons should be made for the generalisation data:*

*“In Fig. 6, the new parameterization and the parameterization of Larson et al. (2001a) are shown compared to the LES data of the RICO case. We focus on the RICO case because the main differences between these two parameterizations are found for the cumulus regime. For stratocumulus the two parameterizations differ only marginally.”*

In the revised manuscript, we use ASTEX (test data) instead of RICO (training data) for Fig. 6 and changed the text accordingly.

*“For comparison the parameterization by Cuijpers and Bechtold (1995) using an exponential fit of  $F$  that only depends on  $Q_1$  is also shown in Fig. 7c.” (where Fig. 7c shows the RICO case).*

In the revised manuscript, we discuss only ASTEX data in this context.

*“Comparing the two parameterizations based on double-Gaussian distributions, the new parameterization is superior to the parameterization by Larson et al. (2001a) for RICO and ASTEX, but not for ARM and DYCOMS. For the latter two cases the new parameterization and the parameterization by Larson et al. (2001a) seem to have comparable error magnitudes. This is reasonable, because the closure equations have most notably been changed for high positive skewness which correspond to the cumulus cloud regime. Because the new parameterization is better able to reproduce the highly skewed distributions occurring mostly in RICO and ASTEX compared to the parameterization by Larson et al. (2001a), the new parameterization is superior for these cases but not remarkably different for small positive or negative skewness.”*

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



We changed this paragraph to “Comparing the two parameterizations based on double-Gaussian distributions, the new parameterization matches the LES data better than the parameterization by Larson et al. (2001) for ASTEX whereas for ARM the two parameterizations have similar error magnitudes (Table 2). This is reasonable, because the closure equations have most notably been changed for high positive skewness which frequently occurs in ASTEX but is rather scarce for ARM. The same effect can also be found in the training error (Table 3). While a lower error of the new parameterization compared to the error of the parameterization of Larson et al. (2001) is found for RICO (where high positive skewness occurs frequently), similar error magnitudes are found for DYCOMS (where the skewness is small).”

---

Interactive comment on Geosci. Model Dev. Discuss., 6, 1085, 2013.

## GMDD

6, C1220–C1226, 2013

---

Interactive  
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

