Reply to referees:
A refined statistical cloud closure using double-Gaussian probability density functions

Ann Kristin Naumann[1,2], Axel Seifert[3] and Juan Pedro Mellado [1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany

[2]International Max Planck Research School on Earth System Modelling (IMPRS), Hamburg, Germany

[3]Hans-Ertel Centre for Weather Research, Deutscher Wetterdienst, Hamburg, Germany

We thank the two referees for their valuable and constructive comments on the manuscript. We understand that the main concern of both referees is related to the use of the same data sets for both training and testing/validation of the parameterization, i.e. that we are not using a more rigorous statistical cross-validation. There are obviously cultural differences in different areas of geophysics. A more rigorous testing with a separation of training and test data is usually not done for parameterizations in atmospheric models. This may have several reasons. One may be that parameterization developers see their schemes more as physically-based theories with a few constant but unknown coefficients. These coefficients are then estimated based on the available observations. In physics a separation of training and test data is usually not done, because the parameters are interpreted as (fundamental) physical constants. Another reason is that in atmospheric science, especially for the development of physical parameterizations for NWP or climate models, the final testing/validation has to be done in the full model where all feedbacks are present (see, e.g., Jakob 2010, Bull. Am. Met. Soc., 91, 869–875, for a review of parameterization development in atmospheric science). This 'process level' and 'full model' testing is similar to the 'a priori' and 'a posteriori' testing in turbulence modeling and computational fluid dynamics (see, e.g., Pope, 2000).

In the present paper we only present an analysis on the process level and no feedbacks with the large-scale are taken into account, i.e., we assume that we have perfect input data (in our case of the first three moments of the PDF of s). Being on the process level we have used all available data, e.g., to reduce the statistical error of the fitting procedure. Because this is not the final testing and validation anyway and process-level data is usually rare, this may or may not be a reasonable approach.

But we very much agree with the reviewers that especially in our case a separation of training and test data might be helpful. First, because our parameterization relies strongly on statistical arguments and the data is an essential part of the scheme, i.e., it is as much a statistical scheme as it is based on the underlying physical behavior. Second, because we actually do have enough data to make a useful separation in test and training data. In the following we first would like to discuss an additional analysis which we performed using the suggested separation of data and cross-validation. In the second part we will then address the remaining comments of the referees individually.

1 Cross-validation

For performing the suggested cross-validation, we choose to use the LES cases of DYCOMS and RICO as training data and the LES cases of ASTEX and ARM as testing data. A suggested by the reviewers, we attempt to recalibrate the closure equations of Larson et al. (2001, L01, Eq. 3 in the GMDD manuscript) to our training data. When keeping the full symmetry of their approach and fitting the tuneable parameter γ with a least square fit to the LES data of DYCOMS and RICO, we get a best fit for $\gamma = 1.18$. However, this fit is unphysical because then $\sigma_1/\sigma < 0$ for $sk < -2.3$ and $\sigma_2/\sigma < 0$ for $sk > 2.3$.

We therefore relax the symmetric restriction for γ, i.e. allow for non-symmetric closure equations, but keep the functional form suggested by L01. Summarizing Eq. (3) and (4) from the manuscript, we use:

$$
\begin{aligned}
\frac{\sigma_1}{\sigma} &= \begin{cases} 1+\gamma_1^{\mathrm{L01}}\frac{sk}{\sqrt{2.0+sk^2}} \quad \text{or} \quad 1+\gamma_1^{\mathrm{new}}\frac{sk}{\sqrt{2.0}} & \text{if} \quad sk > 0 \\ 1+\gamma_3\frac{sk}{\sqrt{2.0+sk^2}} & \text{if} \quad sk \le 0 \end{cases} \\
\frac{\sigma_2}{\sigma} &= \begin{cases} 1-\gamma_2\frac{sk}{\sqrt{2.0+sk^2}} & \text{if} \quad sk > 0 \\ 1-\gamma_4\frac{sk}{\sqrt{2.0+sk^2}} & \text{if} \quad sk \le 0 \end{cases}
\end{aligned} \tag{1}
$$

Note the labeling of the tuneable parameters, γ_n^{p}, which is different from the labeling in the manuscript. With a least square fit to Eq. 1 for each segment of the training data, we find $\gamma_1^{\mathrm{L01}} = 2.15$, $\gamma_1^{\mathrm{new}} = 0.73$, $\gamma_2 = 0.46$, $\gamma_3 = 0.78$ and $\gamma_4 = 0.73$ (Tab. 1 and Fig. 1). L01 originally were using $\gamma = \gamma_1^{\mathrm{L01}} = \gamma_2 = \gamma_3 = \gamma_4$. To keep parts of the symmetry of this original approach, we here add a semi-symmetric approach restraining $\gamma_1^{\mathrm{L01}} = \gamma_4$ and $\gamma_2 = \gamma_3$. Fitted values for γ_n^{p} to the training data, DYCOMS and RICO, are given in Tab. 1.

The error of the L01 parameterizations (semi-sym. and non-sym.) indeed decreases for the ASTEX case when using the retuned parameters instead of parameters originally used from L01. At

Table 1: Values of the tunable parameters γ_n^{p} in Eq. 1 for the different parameterizations. The values are obtained using a least square fit to the training data (DYCOMS and RICO).

	semi-sym. L01	non-sym. L01	fitted new para.
γ_1^{p}	1.82	2.15	0.73
γ_2	0.54	0.46	0.46
γ_3	0.54	0.78	0.78
γ_4	1.82	0.73	0.73

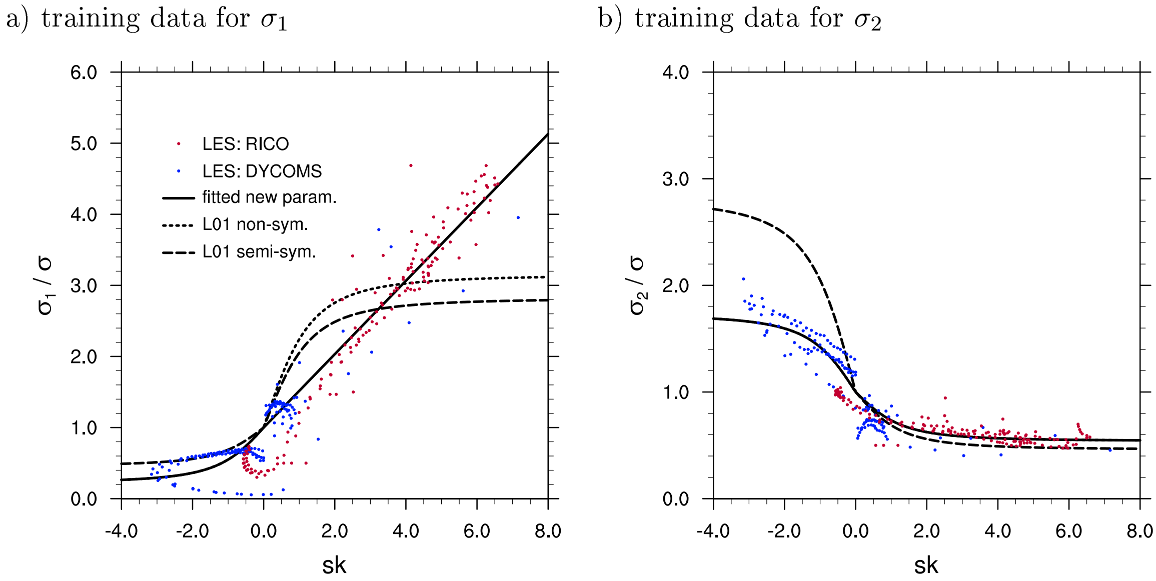

Figure 1: Closure equation with the fitted parameters in Eq. 1 along with the training data. For σ_1/σ with $sk < 0$ and σ_2/σ the non-symmetric L01 closure equation coincides with the fitted new closure equation. The legend in (a) also applies in (b).

the same time the error of the new parameterization remains almost the same for the fitted parameters compared to the parameters suggested in the manuscript (Tab. 2 here and Tab. 2 in the manuscript). Nevertheless, it can be seen from Fig. 1 that the functional form of the parameterization from L01 for σ_1/σ ($sk > 0$) is not adequate to the problem even if the symmetric restrictions are relaxed which motivates us to study a different functional form for the closure equations. Supporting this, the error for the testing data is smaller for the introduced new parameterization than for the retuned L01 parameterization (Tab. 2). The non-symmetric approach mostly gives a smaller error than the semi-symmetric approach. For ARM the conclusions are not so clear in terms of a comparison between the original and the retuned parameterizations but the error of the fitted new parameterization is smaller than the errors of the retuned L01 parameterizations for all error metrics.

Considering the above analysis, we conclude that a retuning of the approach of L01 is somewhat inappropriate. First, because the original L01 scheme cannot be fitted to our data. Second, because the modified formulations are inconsistent with our data for the semi-symmetric formulation (see L01 semi-sym. in Fig. 1), and they already introduce a major change to L01 by giving up the symmetry. We therefore find it a more reasonable approach to use the L01 scheme as it was published. Nevertheless, we agree with the reviewers that separation in test and training data is useful. We will therefore apply such separation in a revised manuscript.

Table 2: Errors of the different parameterizations with the fitted parameters for the testing data of the LES cases of ASTEX and ARM.

		C [%]					$\overline{q}_l$ [g/kg ·10e-3]				
		SG	orig. L01	semi-s. L01	non-s. L01	fitted new para.	SG	orig. L01	semi-s. L01	non-s. L01	fitted new para.
ASTEX	l_1	`1.10`	0.66	0.51	0.45	**0.41**	`2.37`	1.21	0.95	0.80	**0.72**
	RMSE	`2.67`	1.26	1.01	0.89	**0.88**	`3.97`	2.00	1.84	1.74	**1.51**
	l_∞	`19.70`	9.31	8.75	**7.05**	**7.05**	`23.12`	**10.73**	12.96	14.06	10.98
	bias	**-0.16**	`0.35`	0.30	0.18	0.19	`-1.22`	0.82	0.71	0.51	**0.42**
ARM	l_1	`1.35`	0.61	0.57	0.68	**0.53**	`4.60`	0.97	0.86	0.94	**0.57**
	RMSE	`1.85`	**0.84**	0.98	1.17	0.87	`6.67`	1.42	1.48	1.53	**1.15**
	l_∞	`5.33`	**2.83**	4.43	4.91	3.58	`16.00`	**6.10**	6.91	6.70	6.30
	bias	`-1.21`	0.30	-0.07	-0.27	**-0.02**	`-4.43`	**-0.29**	0.70	0.69	0.32

The parameterization based on the single-Gaussian distribution (SG) and the original parameterization from L01 are not changed but are repeated here for convenience. Smallest errors are printed in bold, largest in typewriter. For the formulas of the different error metrics and further details please see Tab. 2 in the discussion paper.

2 Individual comments of referee 1

The referee's comments are in *italics*, authors answers are in normal font.

I. The method of evaluation of the new parametrisation does not follow standard statistical practice. In particular, the same LES dataset is (apparently) used both to tune and to evaluate the new parametrisation. Instead, in standard statistical cross-validation, the data are divided into independent training and testing parts; a parametrisation is tuned using the training data but then evaluated using the testing data. It is risky to rely on estimates of training error, such as those shown in Table 2 for the new parametrisation, because they typically underestimate the test or generalisation error, i.e. the performance of the model for other data sets. Relatedly, the comparison between the new parametrisation and the older parametrisations is not entirely meaningful for 3 related reasons:
1. The older parametrisations (L01 and CB95) were trained on different observational or LES datasets than the LES used for evaluation in the manuscript, and hence a (higher) generalisation error is being calculated for the older parametrisations. However, because the error estimate of the new parametrisation is a (lower) training error, the two types of error listed in Table 2 are not comparable quantities.
Please see our above analysis (Sect. 1).

2. The new parametrisation adds extra tunable parameters that are not present in the older parametrisations. For instance, the CB95 parametrisation for $\overline{w'q'_l}$ has two parameters (1.4 and 1.0), whereas the new parametrisation has four parameters (a, b, sk, and 1.0). (Although skewness is readily available from LES data, it is not easy to parametrise accurately in a large-scale model.) Similar, the new parametrisation adds parameters (γ_2 and γ_3) in the prediction of cloud fraction and liquid water. The addition of extra parameters might degrade the generalisation error of the new parametrisation.
The Larson et al. (2001) as well as our parameterization needs three input parameters: $\bar{s}$, σ and the skewness of s. As the reviewer points out, while the first two parameters are available in many large-scale atmospheric models, the skewness is not. Nevertheless our scheme does aim at boundary layer schemes which would provide an estimate of the skewness, and in some sense the analysis of L01 and our paper show that, especially for shallow cumulus clouds, the prediction of the skewness is crucial. There are ongoing efforts to develop higher-order schemes which include some estimate of the third moments, i.e. skewness. See Mironov (2009: Turbulence in the Lower Troposphere: Second-Order Closure and Mass Flux Modelling Frameworks, in: Interdisciplinary Aspects of Turbulence, Springer.) for an extensive discussion of this topic. We extend the discussion of this issue in the introduction of the revised manuscript.
The additional parameters (γ_n) in the closure equations for the parameterization of the cloud cover and the average liquid water result from a relaxation of the strictly symmetric behaviour of the closure equation as suggested by L01. The introduced non-symmetric behaviour is physically based on the different dynamics of the stratocumulus and the cumulus regime which is confirmed by both LES data and the observational dataset. For the liquid water flux, Cuijpers and Bechtold (1995)

also found a dependence of F on the skewness of s. Presumably because the skewness was not available from turbulence closures at that time, they suggested a parameterization solely depending on the saturation deficit. Nevertheless a dependence on the skewness is physically based, clearly found from our datasets (Fig. 5 in the manuscript) and apparent as the skewness becomes available.

3. The parameters were tuned for the new parametrisation but not re-tuned for the older ones. I recommend that the authors divide the LES data into training and test parts. Then tune all the parametrisations, old and new, to the training data. Finally, compute error statistics and generate plots for the test data. That is, I recommend that the evaluation use cross validation.
Please see our above analysis (Sect. 1).

II. Although the PDF works well for the cases selected, it is unclear how well the PDF will generalise to practical applications of interest.
1. The new parametrisation of $\overline{w'q'_l}$ does not work for the full range of relative humidities. The manuscript states 'Because this new parametrisation is designed to fit the LES data with $Q_1 > -4.0$, we limit the range of application for this parametrisation to $Q_1 > -4.0$.' (p. 1099, lines 21-22) The manuscript also states 'Note again that for layers with $Q_1 < -4.0$ the parametrisations of the liquid water flux are not valid.' (p. 1101, lines 11-12). This might hinder use in weather forecast or climate models, which need to behave reasonably in all conditions. I suggest that the authors either test the new parametrisation over the full range of Q1 or else modify the parametrisation so that it does work for the full range.
Using the suggested parameterization in the full range of Q_1 would give unreasonable values for the parameterized liquid water flux at a thin layer at cloud top where the liquid water flux is very small (Fig. 7 in the manuscript) and the suggested parameterization as well as the parameterization suggested by Cuijpers and Bechtold (1995) are too sensitive to high sk at low Q_1. In a full model (as opposed to the process level study we conduct here) the cloud top behaviour is very sensitive to the interplay of the cloud parameterization and the boundary layer scheme. Therefore a meaningful validation of the cloud top behaviour should be done in the full model with all feedbacks present. However, as a first attempt $\overline{w'q'_l} = 0$ for $Q_1 < -4.0$ might be sufficient. We add this discussion in a revised version of the manuscript.

2. For large values of skewness (6), the new parameterisation leads to large values of σ_1/σ (4!). This will lead a long tail on the right side of the distribution and large values of kurtosis. This is apparently a good assumption for the PDF of s for the chosen LES of cumulus, but it is unclear how such a spiky PDF will behave when generalised to other cases. Further discussion in the manuscript is welcome.
Positive skewness of the PDF of s is a feature of a cumulus type cloud layer while the skewness is negative for stratiform cloud layers. Therefore large positive values for the skewness and hence large σ_1/σ occur only for cumulus type cloud layers.

Minor comments: I. The computational cost of the method might be larger than expected because *it involves iterating to find the relative weight (a), and at each iteration, a square root must be computed. Could the authors comment on the number of iterations used to obtain the results shown in the figures and tables?*

To solve for a, we used a simple bisection method with an accuracy of 10^{-6} which typically took about 30 iterations, i.e. our implementation is not optimized for computational efficiency. Using the closure equations, a is a function of the skewness only. To avoid an iterative solution for the use in a weather forecast or climate model, one might want to use an (e.g. ploynomial) approximation of a as a function of sk.

II. p. 1094, lines 18-19: In a list of references, it is customary to list earlier references before the later ones.

We change that.

III. Fig. 7, panel c: Is the blue dashed line computed from Larson et al. 2001a or Cuijpers and Bechtold 1995? There is no legend in panel c. It is unclear whether to use the legend in panel a or panel f.

The legend in panel f is meant to apply also in panel c. For more clarity, we shift the legend to panel c and state its validity in the caption in a revised manuscript.

3 Individual comments of referee 2

The referee's comments are in *italics*, authors answers are in normal font.

A few points for improvements (major comments):

1) It is not quite clear whether and how the 'training data', in particular data from the RICO simulations, differ from the test data, which seem to be from RICO simulations as well. Of course, if training and testing data are the same, the model derived from the training data will produce good results, but without guarantee that it works with independent test data as well. This is a critical point and some of the following comments are due to a related problem.

Please see our above analysis (Sect. 1).

2) (page 1094): The double Gaussian has 5 free parameters (here a, s_1, s_2, σ_1, σ_2). I do not understand that the number of free parameters should change when s_1 is expressed as a function of the remaining 4 parameters plus the 3 parameters $\overline{s}$, σ, sk, which are then 7 free parameters. Evidently they cannot all be independent. Please clarify.

For fitting a double-Gaussian distribution to a given PDF the five free parameters of the double-Gaussian (a, s_1, s_2, σ_1, σ_2) can in principle be chosen freely. Because the skewness is such a crucial parameter in our closure, we make sure that the skewness of the given PDF equals the skewness of the fitted double-Gaussian distribution by adding an additional constraint. This constraint is determining s_1 as a function of the other parameters (as given in the equation on p. 1094 in the

manuscript) and therefore reduces the number of free parameters from five to four. The moments of a full double-Gaussian ($\overline{s}$, σ, sk) are in general of course not independent from the parameters of the two individual Gaussians (a, s_1, s_2, σ_1, σ_1) and exactly such a relationship is used for the constraining equation. We clarify that in a revised manuscript.
In terms of the suggested parameterization, the three moments of the subgrid PDF are supposed to be obtained from a higher order closure boundary layer model. Using the closure equations (Eq. 4 in the manuscript), σ_1 and σ_2 are determined as a function of σ and sk. Then the dependencies of the parameters of the two individual Gaussians on the moments of the full double-Gaussian (Eq. 5, 6 and 7 in the manuscript) are used to determine the remaining three parameters (a, s_1, s_2).

3) (page 1096 and Figure 3): In Figures 3a and 3b it is shown how the new parameterisation (Eq. 4) is derived by fitting parameters to data from LES model runs. Figure 3c shows how this new parameterisation has been tested. But there is a problem. Instead of using an observable quantity like LWC, optical thickness, radar reflectivity, the quantity σ_1/σ is shown. It seems to me that here the test data are treated in the same way as the training data and thus they should show similar behaviour, isn't it? Further, as σ_1/σ is not an observable quantity (probably), it must have been derived from simulations and thus it seems again that training and testing data are closely related which would render the results useless.
First we would like to clarify that σ_1/σ is in some sense an 'observable' quantity. In-situ aircraft measurements, for example, do provide high-resolution time series of temperature, water vapor mixing ratio and liquid water content. From those three variables s can be calculated and the PDF of s is estimated based on the time series data. By fitting a double-Gaussian distribution to the PDF of s, σ_1/σ is obtained. This procedure has also been used by L01 in their paper. In our paper we use both, observations and LES data, to estimate the PDF of s. Please note that we are not using the observations to validate either the LES or the parameterization, instead LES data and observations are both used to support the choices made for the parameterization and to estimate the remaining coefficients. We clarify that in a revised manuscript.
There seems to be yet another misunderstanding concerning the different RICO datasets which we use. For RICO we use both a) the 'observational dataset' from the RICO field campaign and here in particular the airborne measurements as described in Sect. 2.2 in the manuscript and b) the 'LES datasets' which are the output data of a LES case based on the RICO field campaign as described in Sect. 2.1.4 in the manuscript. In Fig. 3 c the observational data (blue dots) are obtained using the 'observational dataset' described above. It is therefore not obvious to us that the PDFs from the observational dataset behave similar to the PDFs from the LES dataset.

4) (figures 5 and 6): these figures are not very useful as again non-observable quantities (i.e. model quantities) are plotted against other non-observables. For testing I expect to see plots with quantity y modelled against quantity x observed, usually giving a cloud of data points scattered more or less around $y = x$ or $y = a + bx$. The scatter around the $y = x$ line then allows statements about the quality of the model. Here I have problems to estimate a quality. In figure 6, all lines look very different to the concentrated patch of data points, but I dont know what it means. In contrast,

figure 7 looks much more useful, showing measured (?) profiles against modelled ones with different parameterisations. This is understandable, but the testing in the former figures is not. This and the corresponding text should be improved.
Unlike the referee we think that Fig. 5 and 6 are essential to the manuscript because they show the actual parameterization relations, i.e. C and $\overline{q}_l/\sigma$ as a function of Q_1 and sk. Fig. 7 surely is a more intuitively accessible depiction to the reader but is only showing the exemplary usage of the parameterization. We therefore would like to keep Fig. 5 and 6 in a revised manuscript. Regarding Fig. 7, we think that there is a misunderstanding concerning the observational dataset and the LES dataset, similar to the issue raised in the referees comment 3. To clarify, Fig. 7 does not show measured profiles from a field campaign but LES data against different parameterisations.

Minor points: 1) (page 1092, line 21): 'surface fluxes', please say what is flowing (heat, vapour?).
We refer to a heat flux and add that in a revised manuscript.

2) (Equation 8, 2nd line): what is the difference between $\overline{q_l}$ and $\overline{s}$? According to the equation they should be identical? If so, please state it.
Because s has negative values in subsaturated air while q_l obviously cannot be negative, there is a difference in their respective mean values. In terms of the integral the difference is: $\overline{q}_l = \int_0^\infty P(s)sds$ while $\overline{s} = \int_{-\infty}^\infty P(s)sds$.

3) (page 1103, lines 7-9): first "deterministic PDF" sounds strange, but contrasting it to a "stochastic approach" sound even more so. Perhaps you can find better expressions.
'Deterministic' was meant to refer to 'scheme' in the phrase 'deterministic PDF scheme'. To avoid this misunderstanding, we change it to 'PDF-based, deterministic scheme'. By 'stochastic approach' we refer to parameterizations based on stochastic processes or Monte-Carlo simulations.

4) (page 1103, line 12): 'In these moist cases...'.
We change that.

5) (page 1106, line 11): I know what a joint pdf is, but what is a two-point pdf?
While a joint PDF is normally relating the PDFs of different quantities (in this case, s and τ) at the same point in space and time, a two-point PDF is relating the PDFs of quantities at different points in time or space (e.g., $s(t_1)$ and $s(t_2)$; cf. Pope, 2000). For clarity, we avoid the term 'two-point PDF' and change the corresponding phrase to '... which would require the use of a joint PDF or even the introduction of time correlations to the problem.'

6) (table2, footnotes): check brackets in the definition of RMSE.
We checked and found agreement of our definition with Eq. 7.28 in D. S. Wilks (2006): Statistical methods in the atmospheric science, Elsevier, p.279.