

Vincent E. Larson^{1*}, Jean-Christophe Golaz², James A. Hansen³, David P. Schanen¹, and Brian M. Griffin¹

¹ Dept. of Mathematical Sciences, University of Wisconsin — Milwaukee, Milwaukee, WI

² NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, NJ

³ Naval Research Laboratory, Monterey, CA

1. ABSTRACT

It is often easy to see when an atmospheric model disagrees with data. It is usually much harder to locate the ultimate sources of model error.

It is particularly difficult to diagnose errors in a model's structure, that is, errors in the functional form of the model equations. One technique that may help is parameter estimation or calibration, that is, the optimization of model parameter values. Typically, calibration is used solely to improve the fit between a model and observational data. In the process, however, calibration may cover up structural model errors.

In a quite opposite application, calibration may be used to uncover the ways in which a model is wrong. The basic idea is to separately optimize model parameters to two different data sets, and then identify parameter values that differ between the two optimizations. When no single value of a particular parameter fits both datasets, then there must exist a related structural error.

The calibration method that we use produces an entire multi-variate distribution of parameter values. It may prove useful for a wide range of parameterizations. We apply the method to a parameterization of boundary layer clouds, uncover the presence of a structural model error, revise the model structure, and obtain improved results.

2. INTRODUCTION

Parameterization packages for shallow clouds unavoidably contain undetermined parameters. They control, at a minimum, eddy diffusivity and microphysics. Furthermore, these parameters cannot be derived theoretically from first principles. Rather, they must be estimated or calibrated, that is, fitted directly or indirectly to data (Jackson et al. 2003, 2004; Carrió et al. 2006).

This process of calibration has a somewhat sordid reputation in the parameterization community. Although every cloud parameterization is calibrated at least informally as a stand-alone single-column model, the calibration of cloud parameterizations is little discussed in the literature (see, however, Emanuel and Žiković-Rothman 1999). The reputation of calibration suffers

because one often suspects that calibration has been used to mask structural model errors. A structural error is a type of model deficiency in which there is a mis-specification of a term's functional form, not merely a mis-specification of a parameter value. This paper argues that the evil here is not calibration per se, but rather model structural error; calibration should not be marginalized, but rather exploited to detect structural error.

Two common symptoms of structural error are underfitting and overfitting (Geman et al. 1992; Moody 1994; Wilks 1995).

Underfitting occurs when a model's structure is not rich enough to capture true variability in a dataset. In such situations, calibration can distinguish true structural errors from merely poor parameter values. For instance, it may occur that no single set of parameter values yields a good fit for all cases (e.g. stratocumulus, cumulus, etc.) in the dataset, even though good parameter values can be obtained for each case separately. Differences in the parameter values in the separate calibrations can provide clues about the source of structural error. In these situations, calibration does not hide errors, but exposes them.

Overfitting occurs when too many parameters are fitted using too few data. Overfitting may hide structural errors because it may introduce compensating errors between terms. This occurs when, in the process of fitting a model to a limited dataset, parameter values are inadvertently chosen such that one term cancels structural errors in another. If the structural errors persist undetected, then the overfitted model is unlikely to match other, different datasets. However, a means to mitigate overfitting is cross-validation against independent datasets.

This paper applies an ensemble parameter estimation technique to a single-column model (SCM) for boundary layer clouds and turbulence. Our two main goals are to (1) detect structural model errors in the SCM; and (2) improve the SCM's fit over a broad range of cloud regimes.

The structure of this paper is as follows. In section 3, we outline the SCM that we calibrate. In section 4, we describe our ensemble-based parameter estimation framework. In section 5, we discuss the initial parameter estimation experiments and the model deficiencies revealed by them. In section 6, we propose empirical model modifications and test them with reference large-eddy simulation (LES) datasets. In section 7, we cross-validate these modifications using indepen-

*Corresponding author address: Vincent E. Larson, Department of Mathematical Sciences, University of Wisconsin — Milwaukee, P. O. Box 413, Milwaukee, WI 53201-0413, vlarsan at uwm dot edu, <http://www.uwm.edu/~vlarsan>.

dent datasets. A fuller description of this research is contained in Golaz et al. (2007).

3. DESCRIPTION OF SCM

Our SCM simulates boundary layer clouds and is fully described in Golaz et al. (2002a). Briefly, the SCM is a higher-order turbulence closure model that uses a multi-variate probability density function (PDF) to close higher-order turbulence and buoyancy terms. The multi-variate PDF represents the horizontal subgrid-scale variability of vertical velocity, temperature, and total moisture. A functional form of the PDF is specified, and for each vertical level and time step, moments for that functional form are predicted (such as mean, standard deviation, etc.), thus allowing the PDF to vary with height and time. The underlying functional form of the PDF is a mixture of two trivariate Gaussians. The shape was determined empirically from both aircraft measurements and LES data by Larson et al. (2002) with further modifications by Larson and Golaz (2005).

4. ENSEMBLE-BASED PARAMETER ESTIMATION FRAMEWORK

A number of factors guided our choice of parameter estimation algorithm:

1. Our single-column model is computationally inexpensive compared to three-dimensional models, and the number of parameters we want to estimate is moderate. Therefore, the efficiency of the parameter estimation algorithm is not an urgent concern.
2. We wish to use a uniform prior parameter distribution, so as to allow the algorithm the freedom to yield the aberrant parameter values that signal model error.
3. We desire a parameter estimation algorithm that is easy to use, even for individual scientists who have expertise in cloud parameterization but not in parameter estimation.
4. Our source of “data” is LES output that is based on observed cases. Using LES output as data has two advantages: the LES model can be set up identically to the SCM, and the LES generates difficult-to-observe fields such as higher-order moments, liquid water, and cloud fraction. Our goal is limited to emulating LES output, not observational data. Therefore, we treat the LES data as perfect input. Improving the agreement between LES and observations is a separate project that is beyond our scope.

Our parameter estimation algorithm for a single ensemble member is as follows. Before beginning the parameter estimation procedure, we select the SCM

and LES output fields (e.g. liquid water) that we wish to match. Then we decide which SCM parameters to calibrate, and we choose initial values for these parameters. Then we perform the following steps (see Fig. 1 for a flowchart): we run the SCM and evaluate the mismatch between the SCM and LES using a cost function. If the mismatch falls below a predetermined threshold, the algorithm stops. Otherwise, the optimizer chooses a new set of parameter values and the procedure is repeated until convergence. The optimization algorithm that we use is the downhill simplex method (Press et al. 1992). The same procedure is repeated for each ensemble member but with different initial parameter values. Our parameter estimation algorithm allows complete flexibility in the choice of field(s) to be optimized and parameter(s) to be estimated.

Central to the parameter estimation algorithm is the choice of the cost function, J . When errors in the data are assumed to be Gaussian, the cost function takes the generic form:

$$J(\mathbf{m}) = \sum_{i=1}^N \frac{1}{2N} \left\{ [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}]^T \mathbf{C}^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}] \right\}_i, \quad (1)$$

where N is the number of observations sets. For each set, there are M observations represented by the vector \mathbf{d}_{obs} . $\mathbf{g}(\mathbf{m})$ is the corresponding vector of model predictions obtained with the model parameter set \mathbf{m} . \mathbf{C}^{-1} is the inverse of the covariance matrix. We simplify it by keeping only its diagonal elements.

In our application, $\mathbf{g}(\mathbf{m})$ is obtained from the SCM output (denoted by $\mathbf{g}_{\text{SCM}}(\mathbf{m})$) and \mathbf{d}_{obs} from the LES “observations” (\mathbf{d}_{LES}). $\mathbf{g}_{\text{SCM}}(\mathbf{m})$ and \mathbf{d}_{LES} can comprise any combination of variables produced by both the SCM and the LES. They could include mean profiles, such as liquid water potential temperature θ_l , total water mixing ratio \bar{q}_t , cloud fraction, or cloud water mixing ratio \bar{q}_c . They could also include any one of the vertical turbulence moments, such as $\overline{w'^2}$ or $\overline{w'^3}$. The observations can include an arbitrary number of LES cases (for example cumulus or stratocumulus).

The minimization of J proceeds on a N -dimensional surface, where N is the number of parameters that we wish to estimate. Because of the complexity and dimensionality of J , the topology likely consists of a large number of local valleys and floors where the minimization may stop. A given function J may possess many comparably good minima. For this reason, we choose to perform an *ensemble* of minimizations.

Each ensemble member starts from a slightly different initial condition and therefore yields a different optimized parameter set. This ensemble approach would be wasteful if the model structure were perfect and the topology of cost function simple: then each ensemble member would produce identical results. Instead, the complexity of the SCM allows different parameter sets to yield similar cost function values.

Our approach to parameter estimation can be cast as an approximation to a Bayesian stochastic inversion with a uniform prior parameter distribution (Jack-

son et al. 2004). Each ensemble member of optimized parameter values does not represent a random draw from the correct posterior distribution, but rather needs to be weighted by its (possibly scaled) likelihood given the LES data. The scaling is necessary to account for data uncertainty. We approximate this scaled weighting below by selecting only the 20 ensemble members with the highest likelihood (lowest cost function value). This represents a sub-optimal weighting that will produce biases in the estimates of the posterior distribution, but we feel that the inaccuracy is unimportant for our qualitative application.

See Golaz et al. (2007) for more details of the parameter estimation algorithm.

5. INITIAL PARAMETER ESTIMATION EXPERIMENTS

5.1 Configuration

A total of 10 parameters from the SCM have been selected for the initial calibration: C_1 , C_2 , C_5 , C_6 , C_7 , C_8 , C_{11} , β , $\bar{\sigma}_w^2$, and μ . The actual model equations in which all these parameters occur can be found in Golaz et al. (2002a) and Larson and Golaz (2005). C_1 controls the dissipation rate of the vertical velocity variance $\overline{w'^2}$. C_2 controls the dissipation rates of the scalar variances and covariance $\overline{q_t'^2}$, $\overline{\theta_t'^2}$, $\overline{q_t'\theta_t'}$. C_5 appears in the parameterization of the pressure correlation term in the $\overline{w'^2}$ equation. C_6 , C_7 appear in the pressure correlation terms of the scalar flux equations $\overline{w'q_t'}$ and $\overline{w'\theta_t'}$. C_8 and C_{11} are part of the pressure term parameterization in the third moment vertical velocity $\overline{w'^3}$. The parameters β and $\bar{\sigma}_w^2$ arise from the PDF functional form. β appears in the diagnostic relationship linking the skewness of θ_t and q_t to the predicted skewness of w . $\bar{\sigma}_w^2$ controls the width of the individual Gaussians in the PDF. Finally, μ is a mixing timescale used in the computation of the mixing length.

We estimate parameter values for two boundary layer cloud regimes separately. The differences in parameter values help reveal model structural errors.

The set-up of both cases is based on observations. The first case is a trade-wind cumulus regime based on the Barbados Oceanographic and Meteorological Experiment (BOMEX) (Siebesma and Coauthors 2003). The second is a marine stratocumulus case, DYCOMS-II RF01, hereafter referred to as RF01 (Stevens and Coauthors 2005). BOMEX and RF01 are selected because they represent different ends of the boundary layer cloud regime spectrum. For each case, the SCM is calibrated against LES results obtained with a version of COAMPS[®] that is suitably modified for LES scales, which we call ‘‘COAMPS-LES’’ (Golaz et al. 2005).

The variables appearing in the cost function (1) are chosen to be cloud fraction and cloud water mixing ratio. The initial experiments we present consist of three ensembles: one that uses BOMEX data exclusively (B1) in the optimization, a second that uses only RF01 data

(D1), and a third that combines both BOMEX and RF01 data (BD1). Each experiment consists of an ensemble of 400 members.

5.2 Results

Results of the initial parameter estimation experiments are shown as scatter plots in Fig. 2. In the scatter plots, each dot represents one ensemble member. The dots are color coded by experiment: green for BOMEX (B1), red for RF01 (D1), and blue for the combined experiment (BD1). The horizontal axes represent the final parameter value, and the vertical axes represent the normalized cost function end value: $\hat{J} = J/J_{\min}$ where J_{\min} is the lowest cost function value of the ensemble. J_{\min} is computed separately for each ensemble. Therefore, the best fitting members (as measured by J) within a given ensemble reside on the lower portion of each panel, and the worst in the upper portion.

For each of the 10 parameters, the scatter plots reveal a surprisingly large spread in the final parameter values compared to the initial range (gray shaded area). The plots clearly illustrate the implausibility of finding a global minimum that is substantially better than other local minima and justifies the use of an *ensemble*-based optimization approach. The product of the optimization is an ensemble of *parameter sets* drawn from a single 10-dimensional distribution and not independent parameters drawn from 10 separate one-dimensional distributions. Scatter plots can only depict the marginal projections of this multidimensional distribution and cannot reveal how parameters co-vary. Therefore, changing one and only one parameter value to another arbitrary value within the range of the scatter plot is likely to worsen the fit, because it would neglect the covariation with other parameters. Also, because of this covariation between parameters, it would not be justifiable to select the mean of each marginal parameter distribution as an optimum parameter value.

The optimized parameter distribution reveals some unexpected features. For some parameters, the distributions for BOMEX (green dots) and RF01 (red dots) overlap considerably, whereas other parameter distributions overlap only slightly. In particular, note the small overlap between green and red dots for C_7 and C_{11} . This small overlap indicates underfitting, which is symptomatic of model structural error.

Profiles from the SCM simulations using the 20 best parameter sets are depicted in Figs 3 and 4. The profiles shown are mean liquid water potential temperature ($\bar{\theta}_l$), mean total and cloud water mixing ratios (\bar{q}_t , \bar{q}_c), cloud fraction, second and third central moments of the vertical velocity ($\overline{w'^2}$, $\overline{w'^3}$). For BOMEX (Fig. 3), the calibrated SCM runs are able to adequately reproduce the LES profiles. Note that only the cloud fraction and \bar{q}_c enter the definition of the cost function J . $\bar{\theta}_l$, \bar{q}_t , $\overline{w'^2}$, $\overline{w'^3}$ are not directly driven to match the corresponding LES profiles. This shows that reasonable physical constraints are embedded in the SCM. However, none of the simulations accurately reproduces the cloud fraction

near cloud base.

The RF01 profiles paint a different picture (Fig. 4). Even though cloud water for the DYCOMS RF01 ensemble (D1, red) and the combined ensemble (BD1, blue) appear comparable, some significant differences are present in other fields. In particular, the SCM is unable to produce a well-mixed total water profile, an indication of a poor representation of boundary layer mixing processes. Furthermore, $\overline{w'^3}$ is unrealistically negative in the lower portion of the domain.

5.3 Summary

The initial calibration experiments have revealed that the SCM simulations agree relatively well with the reference LES for both BOMEX and RF01, if the SCM uses separately calibrated parameter values. A reasonable time evolution of RF01 can also be simulated without having to change the mixing length formulation, a finding that was not at all obvious before the experiments were performed.

However, the experiments clearly demonstrate that improved BOMEX and RF01 require different values of the parameters C_7 and C_{11} . This is undesirable since the SCM is intended to serve as a general boundary layer parameterization. We address this issue further in the next section.

6. REVISED PARAMETER ESTIMATION EXPERIMENTS

6.1 Proposed model modifications

One major difference between BOMEX and RF01, or more generally between shallow cumulus and stratocumulus clouds, is the vertical velocity skewness, $Sk_w = \overline{w'^3} / \overline{w'^2}^{3/2}$. The skewness of w measures the asymmetry between updrafts and downdrafts. In shallow convection, updrafts tend to be narrow and strong and the compensating downdrafts are broad and weak, giving rise to a large positive skewness. For stratocumulus, areas and vertical velocities of updrafts and downdrafts tend to be comparable, which translates into small positive or negative skewness values. Based on the findings of the previous section, we propose to reformulate the parameters C_7 and C_{11} so as to convert them into skewness-dependent functions:

$$C_7(Sk_w) = C_{7b} + (C_{7a} - C_{7b}) e^{-\frac{1}{2} \left(\frac{Sk_w}{C_{7c}} \right)^2} \quad (2)$$

$$C_{11}(Sk_w) = C_{11b} + (C_{11a} - C_{11b}) e^{-\frac{1}{2} \left(\frac{Sk_w}{C_{11c}} \right)^2} \quad (3)$$

Equation 2 implies that in the limit of small skewness magnitudes, $C_7 \rightarrow C_{7a}$, and in the limit of large skewness, $C_7 \rightarrow C_{7b}$. The sharpness of the transition between small and large values is controlled by C_{7c} . Equation 3 has a similar structure. Equations (2)-(3) are purely empirical and we make no attempt to justify them theoretically.

A new set of ensemble-based parameter estimation experiments is performed using the new formulations for C_7 and C_{11} . The methodology is the same as for the initial experiments, except that the dimensionality of the optimization problem is now 14.

6.2 Results

The final parameter values of all the members for the revised experiments B2, D2 and BD2 are shown as scatter plots in Fig. 5. The overlap between BOMEX (green points) and RF01 (red points) ensembles for C_{7x} and C_{11x} is now improved compared to Fig. 2. As a caveat, we note that because the SCM inevitably still contains structural errors and because we have optimized simultaneously for all parameter values, the optimized values have undoubtedly been influenced by compensating errors between terms.

We now focus on the SCM profiles of the best 20 parameter sets of each ensemble. The profiles from the BOMEX ensemble (B2, green lines; Fig. 6) show little change compared to the initial experiment (B1, green lines; Fig. 3). Cloud water profile is improved in the revised combined experiment (BD2, blue lines), but cloud fraction is still underestimated near cloud base.

The impact of the modified pressure terms C_7 and C_{11} is more significant for RF01 (Fig. 7). The RF01 ensemble (D2) and the combined ensemble (BD2) now both yield SCM results that agree better with COAMPS-LES. This is in contrast to the initial experiments (Fig. 4) in which the RF01 ensemble (D1) produced total water mixing ratio profiles that were not sufficiently well-mixed, and had erroneous $\overline{w'^3}$ profiles.

The results from the BD2 ensemble demonstrate that the modifications made to C_7 and C_{11} in Eqns. (2)-(3) allow for the existence of parameter sets that produce reasonable results for BOMEX and RF01 *simultaneously*. This was not the case with the unmodified SCM. Furthermore, before this work was performed, it would have been difficult to identify modifications to the SCM that would have been likely to faithfully simulate both BOMEX and RF01.

7. EVALUATION WITH INDEPENDENT DATA

We have intentionally calibrated only two LES cases and reserved other cases for cross-validation in order to avoid overfitting. To verify that we have indeed avoided overfitting, we simulate four additional test cases using the best 20 parameter sets from the BD1 and BD2 ensembles. The additional test cases are all set up according to the specifications of GCSS intercomparisons, which are based loosely on observations.

The first case is shallow cumulus over land from the Southern Great Plain (SGP) Atmospheric and Radiation Measurement (ARM) site (Brown and Coauthors 2002). The second case involves cumulus clouds rising under a broken stratocumulus deck that were observed during the Atlantic Trade Wind Experiment (ATEX) (Stevens

and Coauthors 2001). The last two cases are both nocturnal stratocumulus-topped layers. One is based on observations from the FIRE [First ISCCP (International Satellite Cloud Climatology Project) Regional Experiment] (Moeng and Coauthors 1996). The second stratocumulus case is based on the second research flight (RF02) of the DYCOMS-II field experiment¹.

Figure 8 shows the SCM cloud properties obtained with the 20 best parameter sets from the ensembles BD1 (blue) and BD2 (orange) and compares them with COAMPS-LES (black). The main interest is in comparing the BD1 and BD2 SCM profiles, which are the “before” and “after” pictures showing the effects of our empirical modifications.

The biggest difference occurs for RF02, where the BD2 SCM (“after”) cloud water profiles are markedly superior to the BD1 (“before”) profiles. Most BD1 ensemble members predict liquid water amount near cloud top that underestimates the LES value by nearly 50%. In contrast, the BD2 ensemble members almost exactly match the LES. For FIRE, ARM, and ATEX, the changes between BD1 and BD2 are modest. Given that these changes are solely based on BOMEX and RF01 datasets, we can safely state that we have avoided overfitting and hence can have some confidence in the generality of the SCM modifications, despite their empirical nature.

8. CONCLUSION

We have presented an ensemble method of parameter estimation. It has three chief advantages:

1. It allows complete flexibility in the choice of parameters to be estimated and fields to be optimized. For instance, we may simultaneously estimate any combination of the parameters C_1 , C_2 , and so forth. Furthermore, we may optimize any combination of fields (e.g. cloud fraction and liquid water) that are produced by the SCM and contained in the LES data.
2. The method is easy to implement, because it does not require writing an adjoint of the model code.
3. The method produces an *ensemble* of sets of best-fit parameter values. This is useful in cases in which the cost function contains many comparable local minima. The ensemble methodology provides, among other things, the range of acceptable values of parameters.

We have used the ensemble parameter estimation method to calibrate a single-column model (SCM) of boundary layer clouds. The “data” used is output from six large-eddy simulations (LES). These consist of three stratocumulus cases, a trade-wind cumu-

lus case, a continental cumulus case, and a cumulus-under-stratocumulus case. We calibrate 10 SCM parameters simultaneously against profiles of cloud fraction and liquid water.

In calibrating the SCM, we sought to avoid the opposing problems of overfitting and underfitting.

To avoid overfitting, we fit only two fields, cloud fraction and liquid water, and two cases, the BOMEX trade-wind cumulus case and the DYCOMS-II RF01 marine stratocumulus case. Other fields and cases were used for cross-validation. That is, they were used to verify that the chosen parameter values fit well generally, not merely for the two fields and cases used in the calibration.

To diagnose the cause of underfitting, we calibrated BOMEX and RF01 *separately*, thereby obtaining two sets of parameter values. The separate calibrations revealed differences in the values of the parameters C_7 and C_{11} , which indicates underfitting due to structural model error. Assessing the significance of these differences was made possible by the ensemble methodology, which clearly showed the lack of overlap in the acceptable parameter values. This demonstrates that calibration need not obscure model structural error; quite oppositely, if used strategically, calibration may *reveal* structural errors. We then replaced the parameters C_7 and C_{11} by empirical functions of skewness. This structural modification ameliorated the underfitting and permitted the SCM to model all six cloud cases more accurately without case-specific adjustments.

Although the parameter estimation technique can help identify the existence of model structural errors, it cannot propose new ideas to fix those errors. Nevertheless, automated parameter estimation does speed up the process of model development because it allows rapid re-calibration when a new model improvement is introduced. This is useful because the introduction of a true model improvement often produces a worse fit to data, since errors in other parts of the model are no longer compensated.

9. ACKNOWLEDGMENTS

COAMPS[®] is a registered trademark of the Naval Research Laboratory. J.-C. Golaz was supported by the Visiting Scientist Program at the NOAA Geophysical Fluid Dynamics Laboratory, administered by the University Corporation for Atmospheric Research (UCAR). V. E. Larson, D. P. Schanen, and B. M. Griffin are grateful for financial support provided by Grant ATM-0442605 from the National Science Foundation, and by subaward G-7424-1 from the DoD Center for Geosciences/Atmospheric Research at Colorado State University via Cooperative Agreement DAAD19-02-2-0005 with the Army Research Laboratory. J. A. Hansen acknowledges support from ONR YIP N00014-02-1-0473. C. Jackson is acknowledged for a useful discussion about this work.

¹For a description of the intercomparison, see <http://sky.arc.nasa.gov:6996/ack/gcss9/index.html>.

References

- Brown, A. R. and Coauthors, 2002: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quart. J. Roy. Meteor. Soc.*, **128**, 1075–1093.
- Carrió, G., W. R. Cotton, and D. Zupanski, 2006: Data assimilation into a LES model: Retrieval of IFN and CCN concentrations. Preprints, *12th Conf. on Cloud Physics*, Madison, WI, Amer. Meteor. Soc.
- Emanuel, K. A. and M. Žiković-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56**, 1766–1782.
- Geman, S., E. Bienenstock, and R. Doursat, 1992: Neural networks and the bias/variance dilemma. *Neural computation*, **4**, 1–58.
- Golaz, J.-C., V. E. Larson, and W. R. Cotton, 2002a: A PDF-based model for boundary layer clouds. Part I: Method and model description. *J. Atmos. Sci.*, **59**, 3540–3551.
- Golaz, J.-C., V. E. Larson, and W. R. Cotton, 2002b: A PDF-based model for boundary layer clouds. Part II: Model results. *J. Atmos. Sci.*, **59**, 3552–3571.
- Golaz, J.-C., S. Wang, J. D. Doyle, and J. M. Schmidt, 2005: COAMPS[®]-LES: Model evaluation and analysis of second and third moment vertical velocity budgets. *Bound.-Layer Meteor.*, **116**, 487–517.
- Golaz, J.-C., V. E. Larson, J. A. Hansen, D. P. Schanen, and B. M. Griffin, 2007: Elucidating model inadequacies in a cloud parameterization by use of an ensemble-based calibration framework. *Mon. Wea. Rev.*, **135**, 4077–4096.
- Jackson, C., Y. Xia, K. Sen, and P. L. Stoffa, 2003: Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands. *J. Geophys. Res.*, **108**, 4853, doi:10.1029/2002JD002991.
- Jackson, C., M. K. Sen, and P. L. Stoffa, 2004: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Climate*, **17**, 2828–2841.
- Larson, V. E. and J.-C. Golaz, 2005: Using probability density functions to derive consistent closure relationships among higher-order moments. *Mon. Wea. Rev.*, **133**, 1023–1042.
- Larson, V. E., J.-C. Golaz, and W. R. Cotton, 2002: Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. *J. Atmos. Sci.*, **59**, 3519–3539.
- Moeng, C.-H. and Coauthors, 1996: Simulation of a stratocumulus-topped planetary boundary layer: Intercomparison among different numerical codes. *Bull. Amer. Meteor. Soc.*, **77**, 261–278.
- Moody, J., 1994: Prediction risk and architecture selection for neural networks. *From statistics to neural networks: Theory and pattern recognition applications*, V. Cherkassky, J. H. Friedman, and H. Wechsler, Eds., NATO ASI Series F, Springer-Verlag, 143–156.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992: *Numerical Recipes in FORTRAN. The art of scientific computing. Second edition*. Cambridge University Press, 965 pp.
- Siebesma, A. P. and Coauthors, 2003: A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.*, **60**, 1201–1219.
- Stevens, B. and Coauthors, 2001: Simulations of trade wind cumuli under a strong inversion. *J. Atmos. Sci.*, **58**, 1870–1891.
- Stevens, B. and Coauthors, 2005: Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Mon. Wea. Rev.*, **133**, 1443–1462.
- Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press, 467 pp.

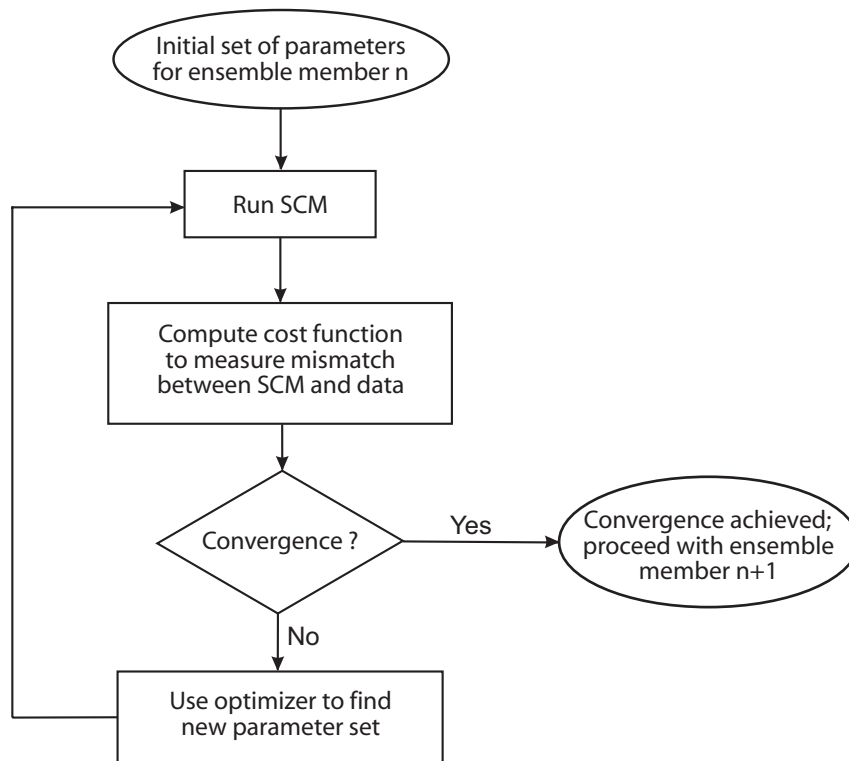


Figure 1: Flowchart illustrating the optimization algorithm for a single ensemble member.

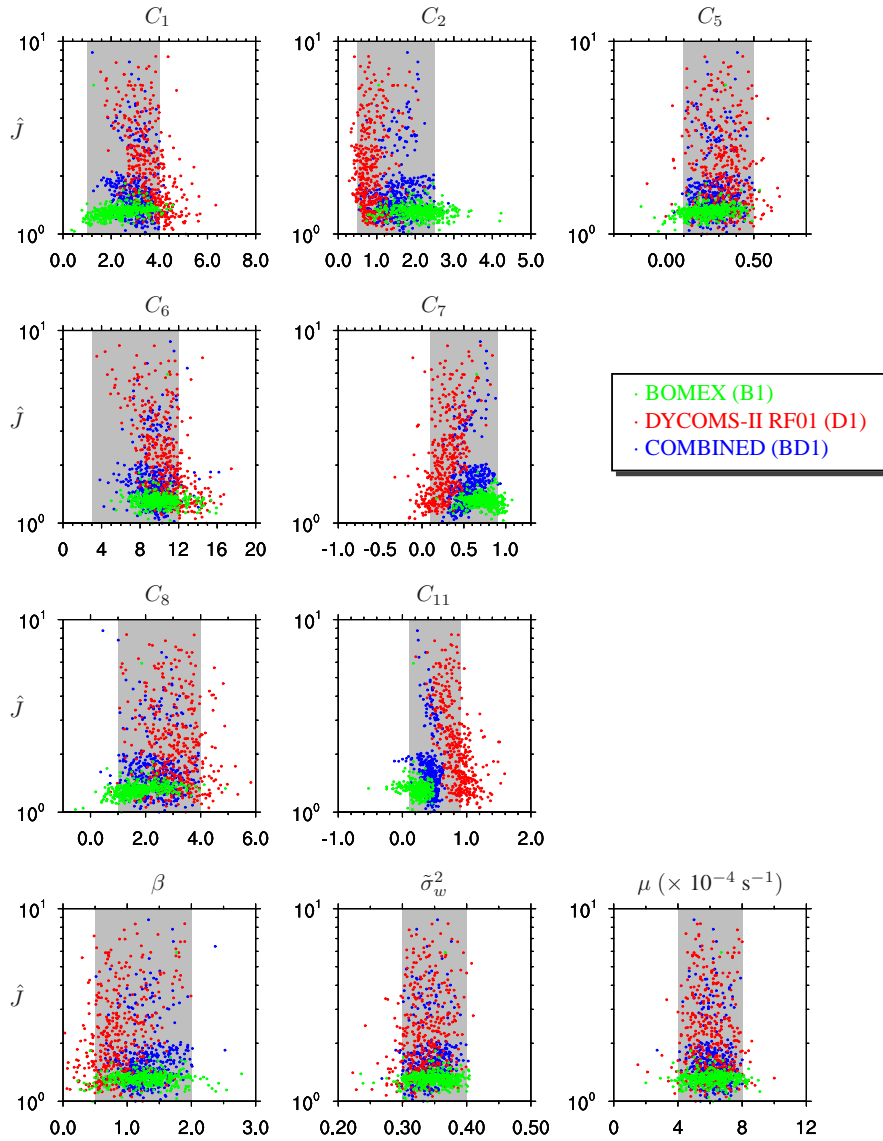


Figure 2: Results of the initial parameter estimation experiments (B1, D1, BD1). Each panel represent one of the 10 parameters. The horizontal axis is the final parameter value, and the vertical axis the normalized error, \hat{J} , of a given optimization with respect to the best member of the ensemble. Each optimization is represented by a single dot. Green dots are for the BOMEX ensemble (B1), red dots for the RF01 ensemble (D1), and blue for the combined ensemble (BD1). The gray shaded areas indicate the initial allowable parameter ranges.

BOMEX Hours 4-6

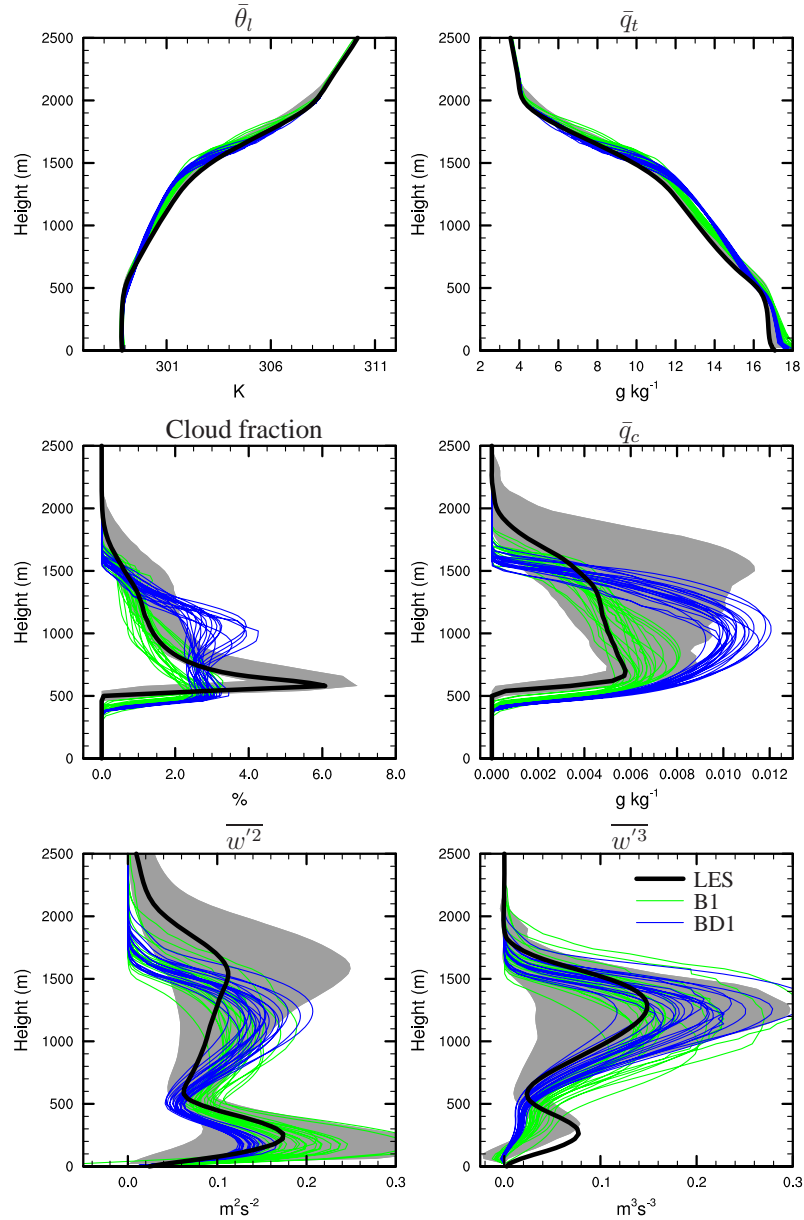


Figure 3: Comparison of the COAMPS-LES profiles (black) with the 20 lowest J value SCM simulations for the BOMEX ensemble (B1, green) and combined ensemble (BD1, blue) of the initial parameter estimation experiments. Profiles shown are liquid water potential temperature ($\bar{\theta}_l$), total and cloud water mixing ratios (\bar{q}_t , \bar{q}_c), cloud fraction, second and third moments of the vertical velocity ($\overline{w'^2}$, $\overline{w'^3}$). They are averaged over the last three hours of the simulation. The gray shaded areas indicate the range (minimum and maximum bounds) of other LES models from the intercomparison.

DYCOMS-II RF01 Hour 4

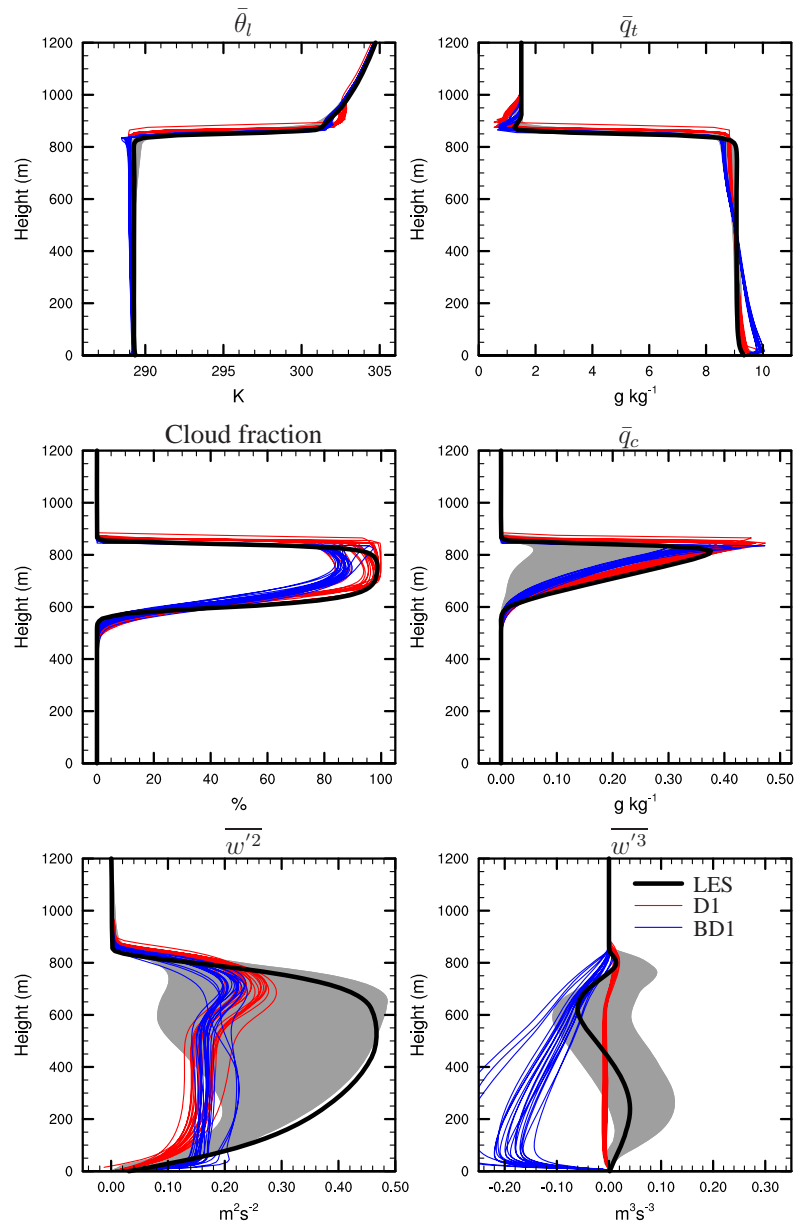


Figure 4: Same as Fig. 3 but for RF01. Red lines are SCM results from the RF01 ensemble (D1) and blue lines from the combined ensemble (BD1). Profiles are averaged over the last simulation hour.

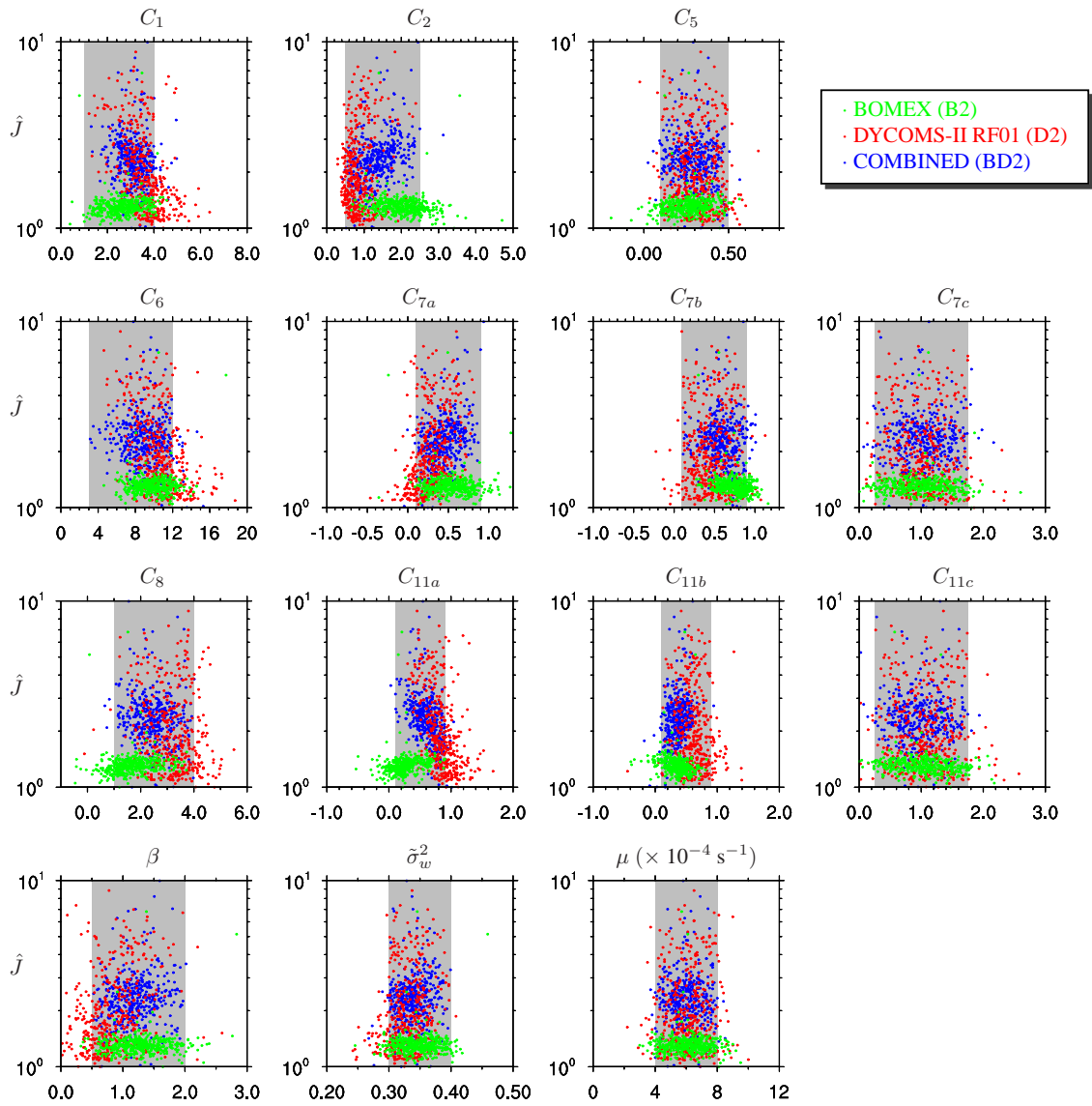


Figure 5: Same as Fig. 2 but for the results of the revised parameter estimation experiments (B2, D2, BD2) with 14 parameters. Green dots are for the BOMEX ensemble (B2), red dots for the RF01 ensemble (D2), and blue for the combined ensemble (BD2).

BOMEX Hours 4-6

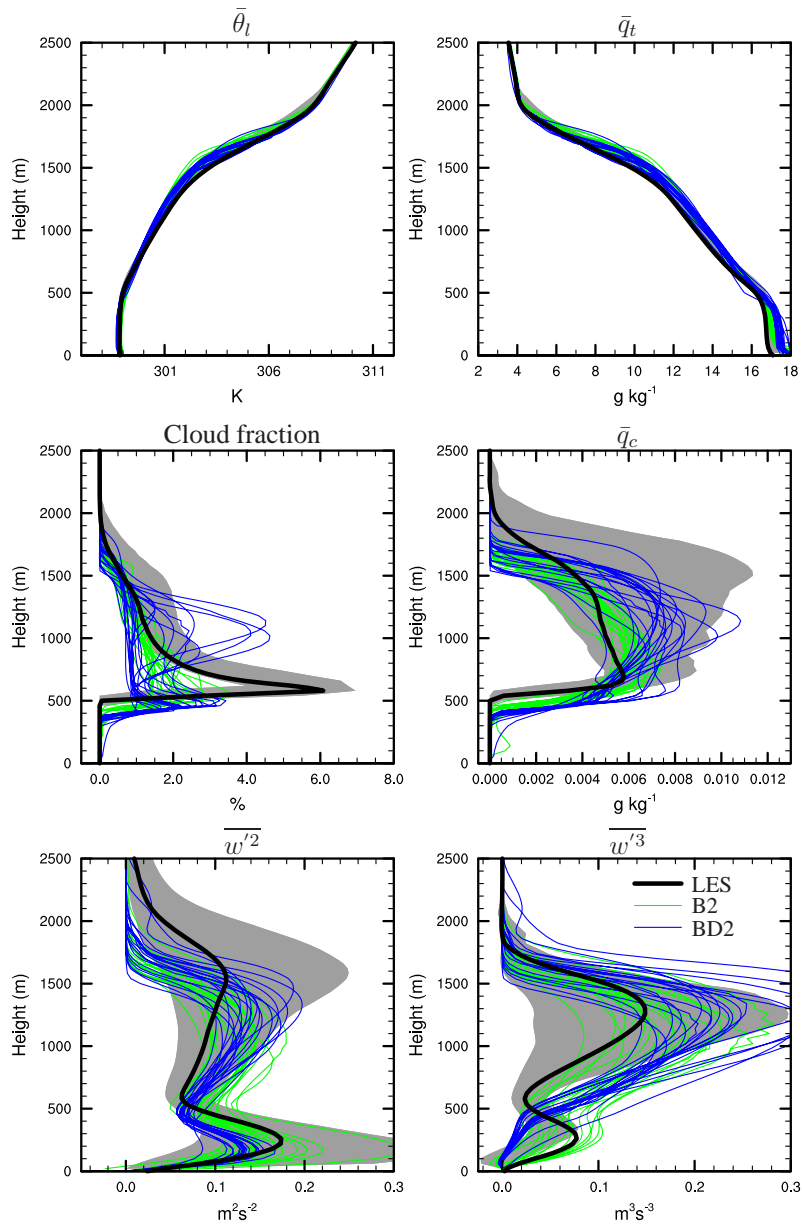


Figure 6: Same as Fig. 3 but for the revised parameter estimation experiments (B2, BD2). Green lines are SCM results from the BOMEX ensemble (B2) and blue lines from the combined ensemble (BD2).

DYCOMS-II RF01 Hour 4

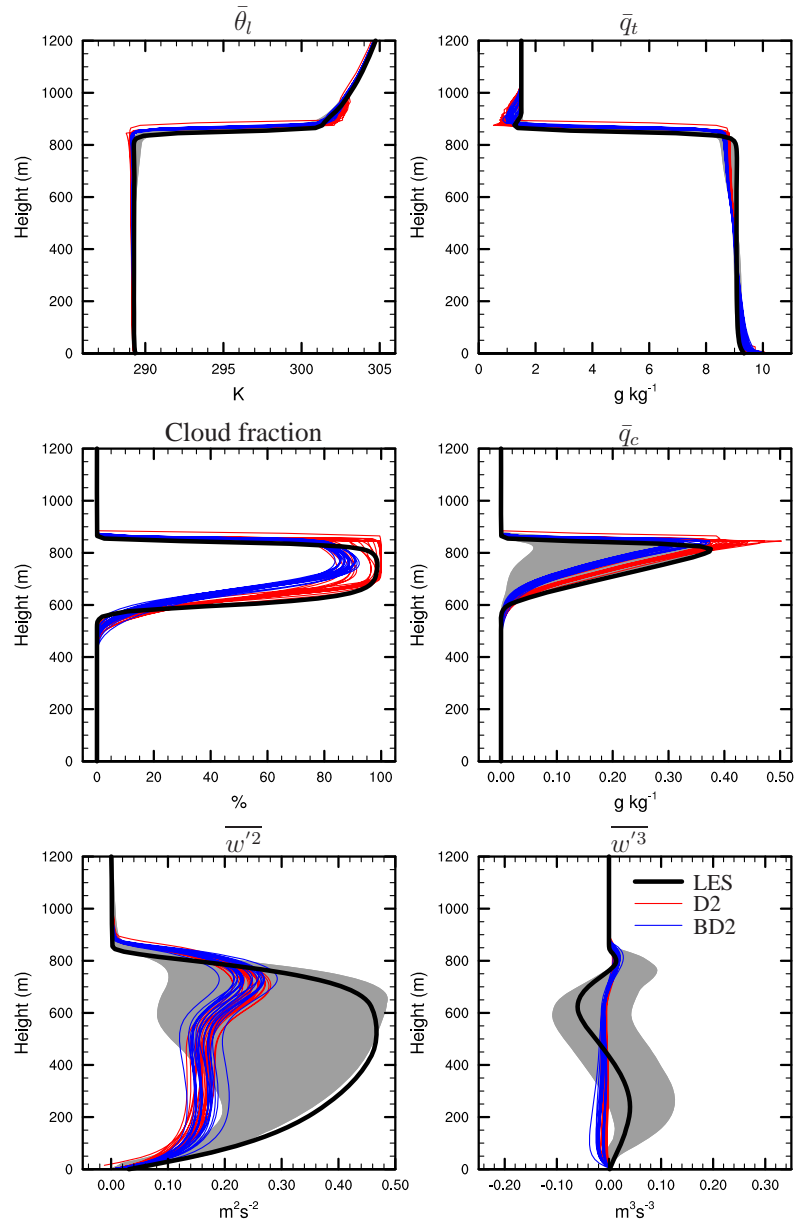


Figure 7: Same as Fig. 4 but for the revised parameter estimation experiments (D2, BD2). Red lines are SCM results from the RF01 ensemble (D2) and blue lines from the combined ensemble (BD2).

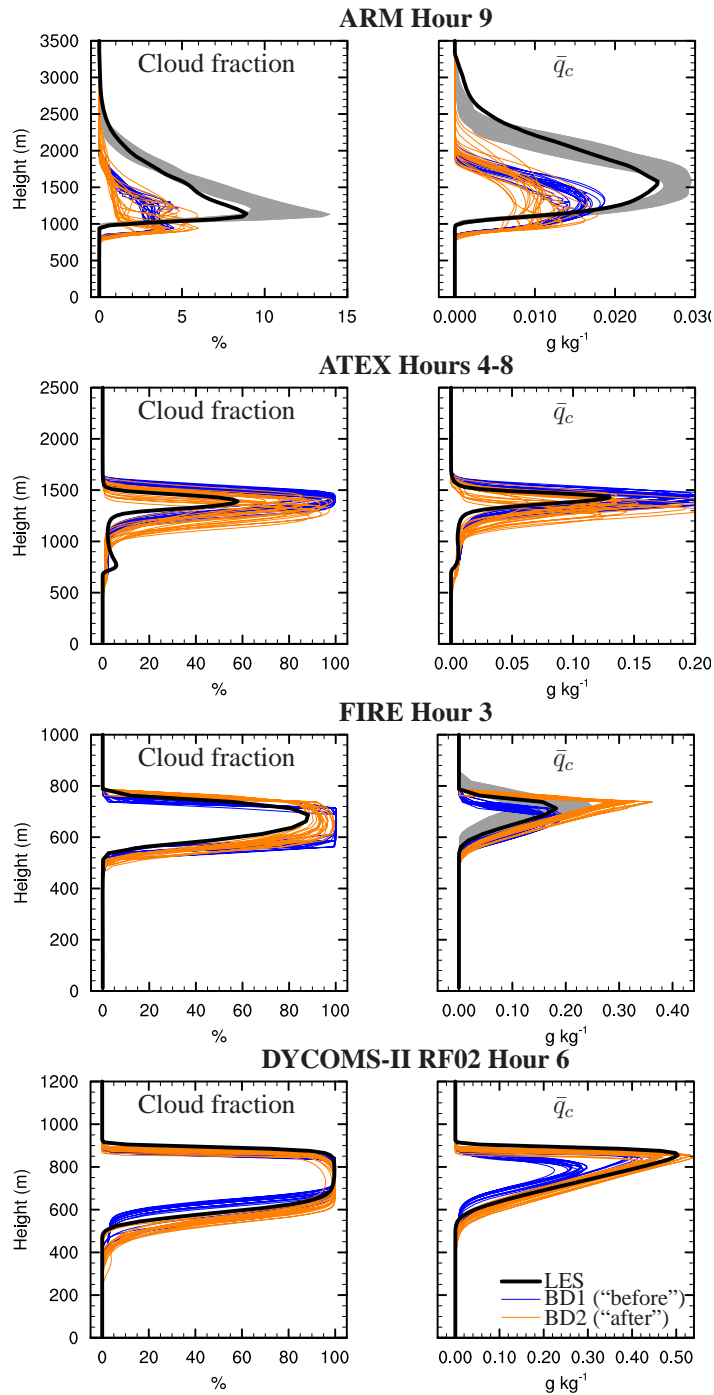


Figure 8: Evaluation of the best 20 parameter sets obtained from the combined experiments BD1 and BD2 using independent datasets. Red lines are COAMPS-LES results and the gray shaded areas represent LES ranges from model intercomparisons. The SCM results from the 20 best parameter sets from BD1 (i.e. before empirical modifications) are plotted in blue. The BD2 (i.e. after empirical modifications) results are in orange.