

# Constraining a System of Interacting Parameterizations through Multiple-Parameter Evaluation: Tracing a Compensating Error between Cloud Vertical Structure and Cloud Overlap

R. A. J. NEGGERS

*Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*

A. P. SIEBESMA

*Royal Netherlands Meteorological Institute (KNMI), De Bilt, and Delft University of Technology, Delft, Netherlands*

(Manuscript received 29 October 2012, in final form 7 March 2013)

## ABSTRACT

This study explores the opportunities created by subjecting a system of interacting fast-acting parameterizations to long-term single-column model evaluation against multiple independent measurements at a permanent meteorological site. It is argued that constraining the system at multiple key points facilitates the tracing and identification of compensating errors between individual parametric components. The extended time range of the evaluation helps to enhance the statistical significance and representativeness of the single-column model result, which facilitates the attribution of model behavior as diagnosed in a general circulation model to its subgrid parameterizations. At the same time, the high model transparency and computational efficiency typical of single-column modeling is preserved.

The method is illustrated by investigating the impact of a model change in the Regional Atmospheric Climate Model (RACMO) on the representation of the coupled boundary layer–soil system at the Cabauw meteorological site in the Netherlands. A set of 12 relevant variables is defined that covers all involved processes, including cloud structure and amplitude, radiative transfer, the surface energy budget, and the thermodynamic state of the soil and various heights of the lower atmosphere. These variables are either routinely measured at the Cabauw site or are obtained from continuous large-eddy simulation at that site. This 12-point check proves effective in revealing the existence of a compensating error between cloud structure and radiative transfer, residing in the cloud overlap assumption. In this exercise, the application of conditional sampling proves a valuable tool in establishing which cloud regime exhibits the biggest impact.

## 1. Introduction

Clouds significantly affect the earth's climate, for a large part because of their impact on the transfer of solar and thermal radiation (Ramanathan 1987). Clouds are often generated by processes that act on spatial and temporal scales that are much smaller than the scales of discretization in general circulation models (GCMs), and as a consequence their impact has to be represented through parameterization. Great variety exists among the suites of subgrid parameterizations in the various present-day operational GCMs. On the one hand, this

reflects the long history of the scientific research behind their formulation, going back decades. On the other hand, this variety reflects the significant complexity of such parameterization schemes, which typically consist of many individual parametric functions, each representing an observed statistical relation between one quantity and another. This complexity brings some considerable risks. The first is in transparency, that is, the interaction between the many parametric components is often not fully understood, which might result in unexpected behavior or instability in the model. Another risk is that of introducing so-called compensating errors between parametric components. These are situations in which a structural error by one component is erroneously compensated by another. In a shifting future climate, when each process might act differently, it is not guaranteed that such an artificial correction will

---

*Corresponding author address:* Roel Neggers, Institute for Geophysics and Meteorology, University of Cologne, Zùlpicher Str. 49a, 50674 Cologne, Germany.  
E-mail: neggers@meteo.uni-koeln.de

still hold. Another potential side effect of compensating errors is that the improvement of one parametric component does not guarantee an improvement in the overall GCM performance.

The complexity of parameterization schemes has hampered progress in recent decades (e.g., Randall et al. 2003a). To address this issue, the evaluation of parameterization schemes against relevant measurements has been an active field of research (GEWEX Cloud System Science Team 1993; Randall et al. 2003b), as testified by the numerous model intercomparison studies at process level (e.g., Stevens et al. 2001; Brown et al. 2002; Siebesma et al. 2003; van Zanten et al. 2011). These initiatives have been successful in providing the modeling community with benchmark results for certain controlled situations or regimes that have to be reproduced by any model, thus acting as a “testing ground” for new or existing parameterizations. However, some problems remain. It can be argued that the available idealized cases are too few to guarantee representativeness. Also, to properly constrain a complex interacting system with measurements (and thus identify any existing compensating errors), the representation of *each* process should be individually confronted with relevant, independent measurements. In reality this has proven hard to realize because either the required multitude of independent measurements was not available or the measurements did not cover sufficiently long time periods to ensure statistical significance in the evaluation result.

How should we move forward? In the evaluation of GCMs at process level, there has been a recent drive toward a more comprehensive approach, attempting to involve many more observational datasets (e.g., Sengupta et al. 2004; Paquin-Ricard et al. 2010; Morcrette et al. 2012; Ahlgrim and Forbes 2012) and to ensure better guidance by the GCM (e.g., Jakob 2003). In a companion paper to this study (Neggers et al. 2012, hereafter NSH12), a new strategy is proposed that consists of the continuous, long-term simulation and evaluation of single-column models (SCMs) against a multitude of independently measured parameters at meteorological supersites. The purpose of this strategy is twofold:

- To constrain the system of interacting parameterization at multiple key points with independent measurements
- To make a statistically significant assessment of model performance

By adopting this strategy, we hope to improve the detection of compensating errors, and to make the SCM result more representative of its native GCM while at the same time maintaining the proven benefits of single-case studies (such as model transparency).

This study is part of a series of three companion papers that all deal with the Royal Netherlands Meteorological

Institute (KNMI) Parameterization Testbed (KPT). Each of these papers has a different purpose. The first paper (NSH12) gives a general introduction to the KPT, describing the adopted methodology in detail and discussing its motivation. The second study, described in this paper, illustrates the opportunities for model improvement created by adopting the multiple-parameter approach as proposed by NSH12 by means of an example evaluation study. The third companion paper (Neggers et al. 2011) is purely a spin-off study on the phenomenon of boundary layer (BL) cloud overlap using large-eddy simulation (LES) results that were inspired by results obtained in the study described here.

The model evaluation described in this paper concerns the implementation of a new, integrated scheme for boundary layer transport and clouds into the Regional Atmospheric Climate Model (RACMO). Both the new and control version will be evaluated at process level against multiple measurements at the Cabauw meteorological site in the Netherlands, to investigate if the new scheme improves the representation of the local cloud-radiative model climate. Results of continuous LES at Cabauw will be used to supplement the observational datasets with information on boundary layer cloud structure for which measurements are unavailable. We will demonstrate how multiple parameter evaluation at 12 key points allows the tracing of the impact of a change in cloud representation throughout the coupled boundary layer–soil system, and how this technique in the end proves effective in revealing the existence of a compensating error in the interaction between clouds and radiation.

Section 2 contains a brief introduction to the Cabauw site and its measurements, as well as detailed descriptions of the model setup and the applied research method. The results of the multiyear evaluation are presented in section 3, including among others a summary of model performance using simple statistical metrics, a conditional sampling exercise to determine the cloud regime of interest, an evaluation against LES results, and the impact of a model improvement that was inspired by these results. Finally, in section 4 the main conclusions will be summarized and their implications will be briefly discussed.

## 2. Method

### a. The site

The Cabauw Experimental Site for Atmospheric Research (CESAR) is situated in a flat grassland area in the vicinity of the small village of Cabauw in the Netherlands. The site has been operated by the Royal Netherlands Meteorological Institute since 1973. Its

TABLE 1. Set of observed and LES-diagnosed parameters at Cabauw used in the SCM evaluation. sfc: surface; BSRN: Baseline Surface Radiation Network.

	Parameter	Abbreviation	Units	Instrument
1	Total cloud cover	TCC	%	CloudNet
2	Downward shortwave radiative flux at sfc	$SW_d$	$W m^{-2}$	BSRN
3	Downward longwave radiative flux at sfc	$LW_d$	$W m^{-2}$	BSRN
4	Sensible heat flux at 5 m	SHF	$W m^{-2}$	Sonic anemometer and thermometer
5	Latent heat flux at 5 m	LHF	$W m^{-2}$	Sonic anemometer and optical open-path sensor
6	Soil temperature at 0 cm	$T_{soil}$	K	KNMI nickel-wired needles
7	Air temperature at 2 m	$T_{2m}$	K	KNMI Pt500
8	Air temperature at 200 m	$T_{200m}$	K	KNMI Pt500
9	Lowest cloud-base height	$z_{base}$	m	LD40 ceilometer and LES
10	Maximum cloud fraction within BL	$cf_{max}$	%	LES
11	Height of maximum cloud fraction	$z_{cfmax}$	m	LES
12	Cloud overlap ratio	$r_{overlap}$	—	LES

main asset is the 213-m tower, equipped at regular intervals with sensors for the purpose of atmospheric boundary layer research, air pollution studies, and climate monitoring (e.g., Driedonks et al. 1978; Ulden and Wieringa 1996). In addition, an array of continuously operational instruments is installed at the site that includes both in situ and remote sensing equipment [described in detail by Russchenberg et al. (2005)]. The Cabauw site participates in the CloudNet project (Illingworth et al. 2007). All Cabauw data used in this study are publicly accessible online (at <http://www.cesar-database.nl/>).

### b. Model setup

Two versions of the RACMO SCM are evaluated in this study, each representing a different boundary layer scheme. The first is the scheme as implemented in the operational RACMO, and is identical to that of the model cycle 31R1 of the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). This scheme will be referred to as the “control” version. The second version is the eddy diffusivity mass flux (EDMF) scheme, including the dual mass flux framework (DualM; Negggers et al. 2009; Negggers 2009). The two model versions will be run and evaluated at Cabauw for the period covering 2007–10.

The setup of the RACMO SCM is already described in detail by NSH12, but will be summarized here. On each day within the selected time period, a simulation is performed of 72-h duration, initialized from the ECMWF analysis at 1200 UTC at Cabauw. The simulations are performed on the L91 grid, using an integration time step of 900 s. As usual in SCM experiments, the tendencies due to advection by the larger-scale circulation are prescribed. These are obtained from short-term regional forecasts by the RACMO 3D model that occupies the same time window. RACMO 3D is also run at L91 vertical grid, has a 25-km horizontal grid spacing, and is

initialized and forced (at the boundaries) by fields obtained from global forecasts with the ECMWF model. In addition, continuous relaxation is applied toward the state variables ( $U$ ,  $V$ ,  $T$ ,  $q$ ) as sampled from the same RACMO 3D forecasts, using a time scale of 6 h. This synoptic time scale is chosen such that the relaxation prevents excessive model drift in time, while still allowing the faster-acting parameterized physics to create their own unique state or “fingerprint.” The soil scheme is initialized using the RACMO 3D state but is free running during the simulation, allowing it to interact with the atmosphere through flexible surface fluxes (including radiation, sensible and latent heat, and precipitation). The evaluation is limited to the first complete day in the staggered simulations, which corresponds to hours 12–36 after initialization. This time window gives the models a 12-h spin-up time, but it is still sufficiently close to the initialization time, and ECMWF analysis, to minimize uncertainties in the applied forcing and relaxation.

The LES results used in this study are generated using the code of the Dutch Atmospheric Large-Eddy Simulation (DALES) model (Heus et al. 2010). The LES is initialized and forced in exactly the same way as the SCMs, so that their intercomparison remains meaningful. The spatial resolution of the LES is 100 m in the horizontal and 40 m in the vertical, in a domain of 6.4 km  $\times$  6.4 km wide and 6 km high, with a time step of 2 s. In this domain and at these resolutions, the turbulence, convection, and associated clouds in the atmospheric boundary layer can be expected to be reasonably well resolved. As part of the KPT (NSH12), the DALES model is run on a daily basis at the Cabauw site, enabled by the significant enhancement in simulation speed brought by making use of a graphical processing unit [GPU; the GPU-Based Atmospheric LES (GALES); see Schalkwijk et al. 2011]. As a result, the LES-derived datasets available at Cabauw now cover multiple years.

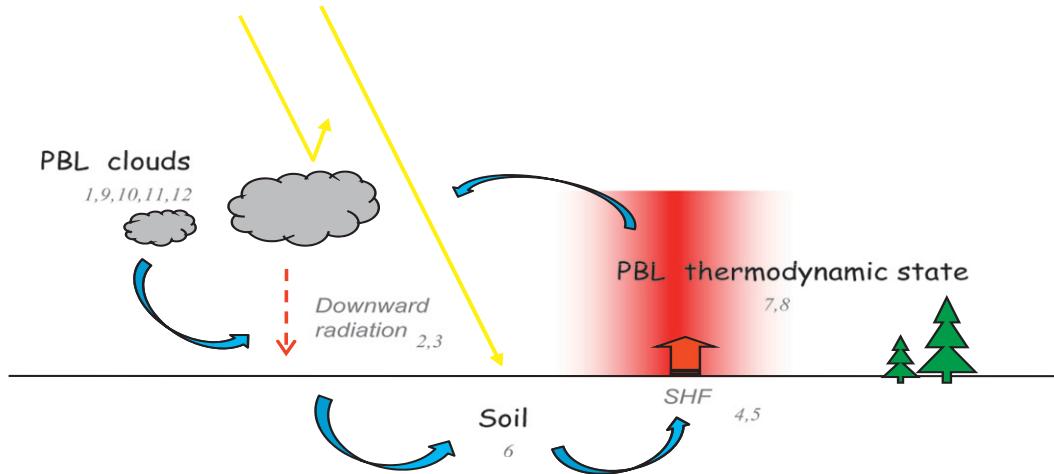


FIG. 1. Schematic illustration of the chain of interacting processes in the coupled soil–BL system that is investigated in this study. Numbers refer to the 12 measurements and LES diagnostics as listed in Table 1 that are used to constrain this system.

*c. A 12-point check*

The set of parameters used in this study for model evaluation is chosen to reflect the cloud, radiative, and thermodynamic state of the atmospheric boundary layer, as well as the surface heat budget and the soil temperature. The 12 data streams as listed in Table 1 consist of two types, namely, (i) measurements by instrumentation at Cabauw and (ii) LES results at Cabauw. The observational parameters are routinely measured at Cabauw

and are therefore available for long and continuous periods of time. The LES data supplement this set with information on key aspects of cloud structure for which no measurements are available.

The chosen set of 12 parameters is designed to constrain the following impact mechanism in the coupled boundary layer–soil system, which mainly involves fast physics and thus acts on very short time scales. This mechanism is schematically illustrated in Fig. 1. We suspect from preliminary tests for various idealized

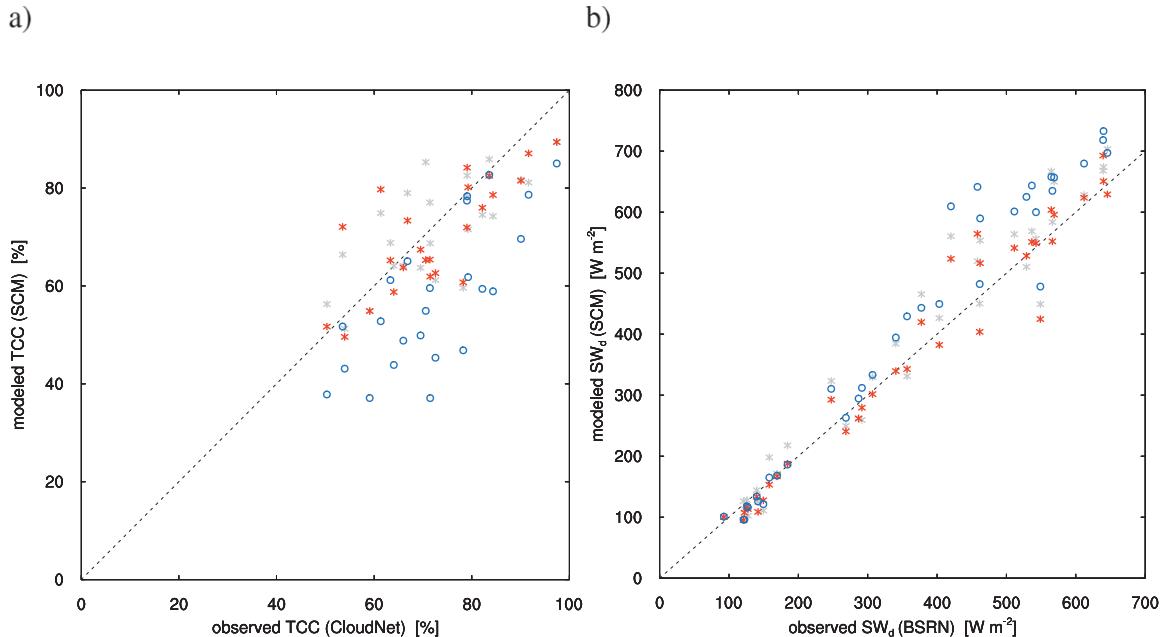


FIG. 2. Scatterplot of the observed (abscissa) vs simulated (ordinate) monthly-mean (a) TCC and (b)  $SW_d$  at Cabauw at 1200 UTC for the period 2007–10. Host model state (RACMO) is shown in gray, its native SCM in red, and the SCM including the new BL scheme in blue.

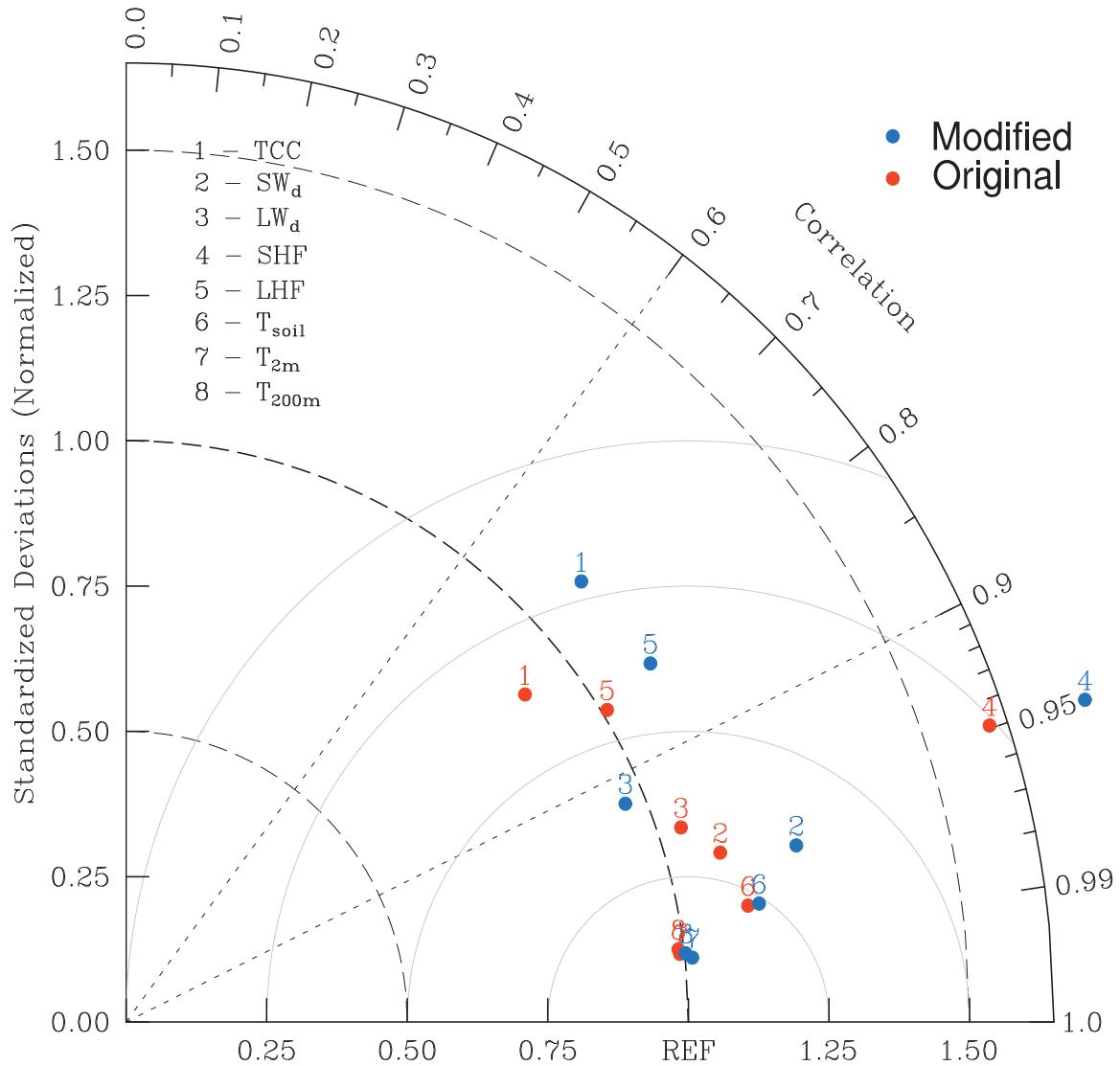


FIG. 3. Taylor diagram quantifying the monthly-mean model performance at Cabauw at 1200 UTC for the period 2007–10 for the eight observed variables in Table 1. Legend and interpretation are explained in the text.

cases that the representation of boundary layer clouds by the new scheme will differ considerably from the control model, both in amount and in vertical structure. Suppose such a difference will also materialize in multiyear simulations with the SCM at Cabauw. This difference in clouds will affect the radiative transfer through the atmosphere, which should affect the surface downward radiative fluxes. These are part of the surface energy budget, which will affect both the surface temperature and the surface sensible heat flux. Last, this will impact the low-level temperature in the atmospheric boundary layer. All main processes in this chain of fast interactions thus react to a change in cloud representation; at the same time, the state of these processes is also routinely measured at Cabauw or can be estimated from

LES simulations. This allows constraining the representation of this interactive chain at multiple points (as indicated in Fig. 1), and thus identifying possible compensating errors between them.

One could argue that many more parameterizations are involved in this chain of interactions than there are measurements. Accordingly, some compensating errors might still remain undetected. However, using 12 independent measurements is already an improvement from evaluating against a single measurement, such as, say, total cloud cover—many examples of the latter type of model evaluation already exist in the literature. The first improvement is that confronting a model with multiple independent measurements should give the investigator more confidence in the result compared

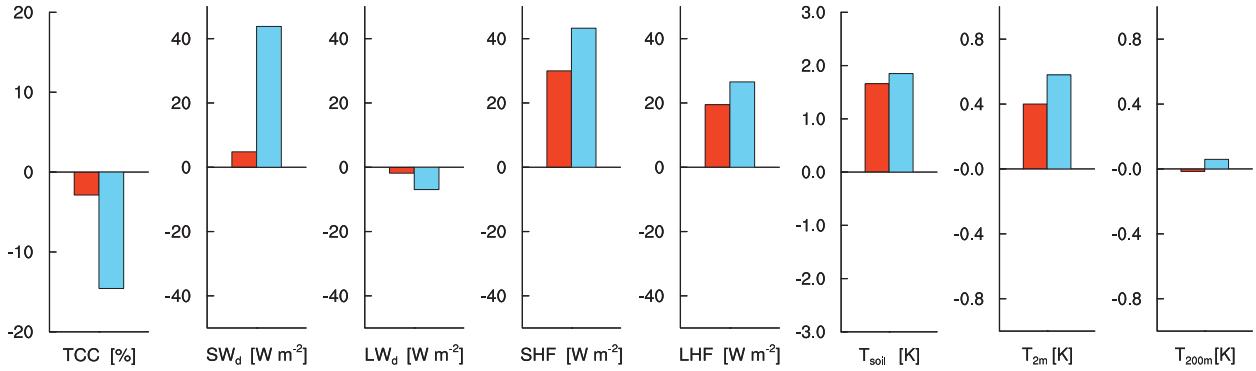


FIG. 4. Bar chart showing the biases of the control SCM (red) and the new SCM (blue) for the eight observed variables in Table 1 as calculated from the monthly means in the period 2007–10.

to a single-variable evaluation. Second, as will be demonstrated, evaluations for a limited number of parameters can already be successful at revealing compensating errors at the first-order level, that is, between the main components in the interacting system. More generally speaking, this study should be interpreted as a *first* attempt at constraining a system of interacting parameterizations more comprehensively—somewhat limited (but not complicated) by the measurements that are currently available.

In our experience, a key problem with improving the representation of an interacting, complex system of parameterizations in existing operational models is that a priori the model evaluation one does not know where exactly the problem lies in the chain of interactions. Compensating errors could sit anywhere, most of the time unconsciously introduced through system tuning after a change in one of the individual model components. To be comprehensive from the start, one should therefore constrain the whole system with as many independent measurements as possible (i.e., available). That this might result in evaluating parts of the system that in the end do not cause the problem is not important; what is important is that this exercise can be effective in narrowing down the search area, by identifying variables for which large biases exist that also correlate well.

The model performance for each parameter will be assessed through simple statistical metrics, such as the bias and the root-mean-square error. The basic time unit of these statistical analyses is the monthly mean, which in our experience results in clear signals. It should be noted that some errors will be shared by both model versions, for which various reasons can exist. For example, there is the question of how representative a point measurement is of an area mean as represented by a model grid box. A second cause can be errors in the prescribed large-scale forcings, for example, due to

small-scale effects that are not resolved by the 3D model. However, the added value of a multiple-parameter approach as opposed to a single-parameter approach is that it makes the evaluation less fragile and more robust, by reducing the sensitivity to the possible peculiarities of a single measurement. Accordingly, model differences will only be taken seriously when they are significant and when they materialize for *multiple* parameters.

### 3. Results

#### a. Quantifying performance

We commence by evaluating the cloud-radiative climate over multiple years (2007–10), as generated by SCM simulations with the two versions of the CY31R1 physics package as described in the previous section. Figure 2 shows the monthly means of the total cloud cover (TCC) and the surface downward shortwave radiation ( $SW_d$ ) at 1200 UTC, plotting the observed (abscissa) against the modeled (ordinate) values. While the control SCM reproduces to a degree the cloud-radiative climate of its host model, the new SCM differs considerably. This shows that the boundary layer (the only physics component that was changed) significantly affects the cloud-radiative climate at Cabauw. In general, the new scheme underestimates the TCC, while it overestimates the  $SW_d$ .

Scatterplots of model results against measurements like Fig. 2 can be interpreted in terms of the bias and the centered root-mean-square (CRMS) error, defined as

$$\text{bias} = \frac{1}{N} \sum_{n=1}^N (\phi_n^m - \phi_n^o) \quad (1)$$

and

$$\text{CRMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N [(\phi_n^m - \overline{\phi^m}) - (\phi_n^o - \overline{\phi^o})]^2}, \quad (2)$$

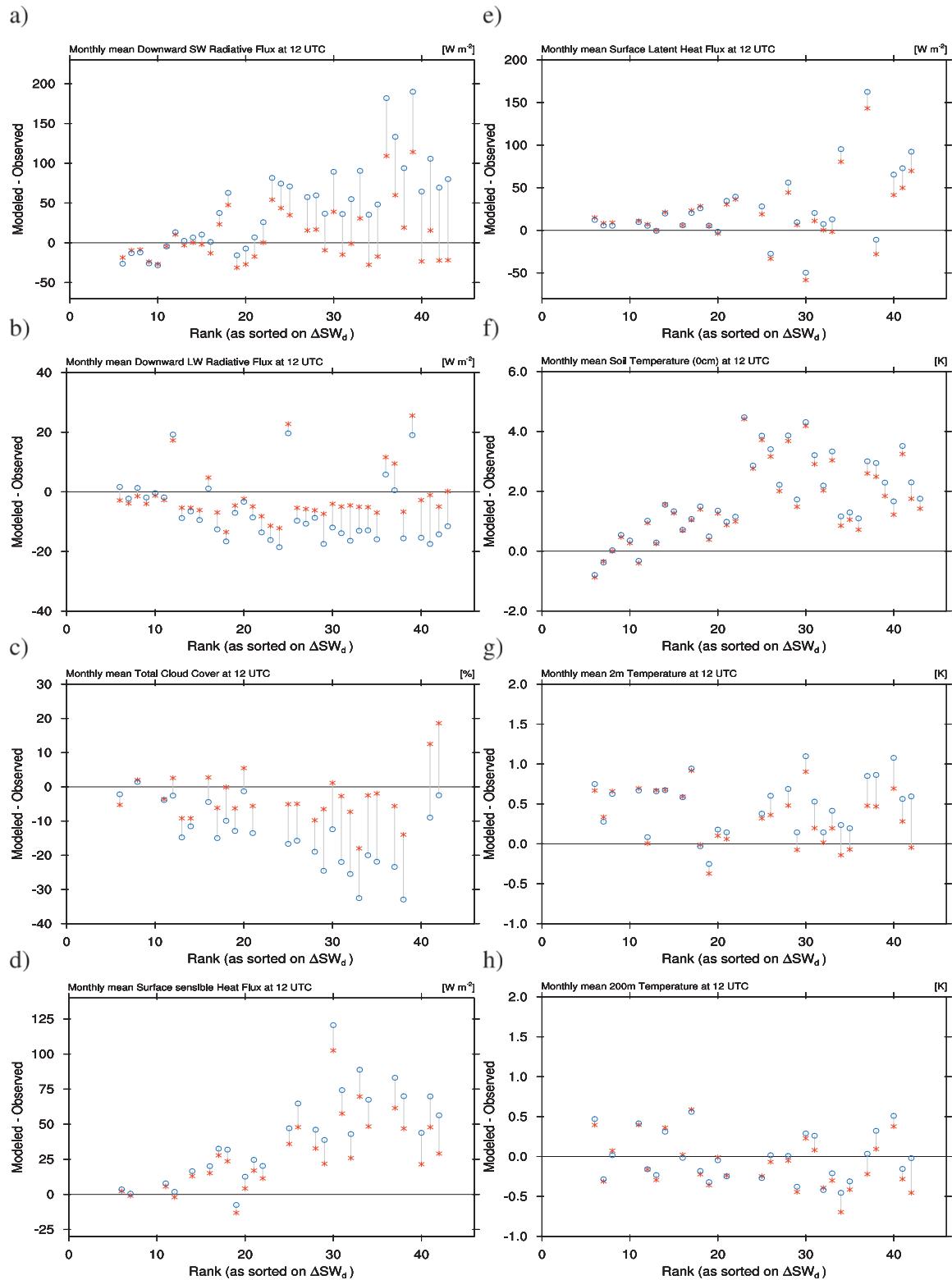


FIG. 5. Modeled – observed monthly-mean values of (a)–(h) the eight chosen variables plotted as a function of the rank in the difference in  $\Delta SW_d$  between the new (blue) and control (red) SCM for the period 2007–10. Gray line connecting the red and blue symbols indicates the change in the bias as a result of the implementation of the new BL scheme.

respectively. Here,  $\phi$  is the monthly mean of a time-dependent variable;  $N$  is the number of samples in the time series; the superscripts  $m$  and  $o$  indicate the modeled and observed data, respectively; and the overbar indicates the time mean over the whole period of evaluation. The bias and the CRMS error are metrics often used to measure model performance; the latter here reflects how well a model reproduces the observed variability among the monthly means. Using these metrics, we now expand the model evaluation to multiple parameters by including six more independently measured variables, as listed in Table 1. A convenient way of visualizing model performance for multiple variables is the so-called Taylor (2001) diagram, as shown in Fig. 3. These two-dimensional diagrams plot the normalized standard deviation against the correlation coefficient. In this diagram, the distance from the point labeled “REF” represents the CRMS difference, and is a measure for how well a model reproduces the observed pattern. The results show that the old model best reproduces the observed pattern in monthly-mean variability for all selected variables, thus confirming the results obtained for only two variables.

### b. Tracing impacts

Differences in biases between models for multiple variables can reflect how a change in the representation of one parameter (say, cloud cover) impacts another, and so works its way through a chain of interacting fast processes (such as the boundary layer system). For  $SW_d$  and TCC, Fig. 2 already shows that the new model has a larger bias in both variables. The biases have opposite sign, reflecting the known impact of the one parameter on the other (i.e., more cloud cover enhances the reflection of solar radiation). The largest biases in shortwave radiation occur at large SW values, reflecting that the error is largest during the summer period. For the total cloud cover the largest biases occur at medium values, which are typical for summertime climate at Cabauw.

The evaluation of biases is expanded to multiple parameters in Fig. 4, which summarizes and quantifies the impact of the model change on all components in the feedback mechanism, as illustrated in Fig. 1. A 12% reduction in total cloud cover (reflecting a multiyear average at 1200 UTC) enhances the downward shortwave radiative flux by about  $36 \text{ W m}^{-2}$ , but it reduces the downward longwave radiative flux by  $4.5 \text{ W m}^{-2}$ . The change in the shortwave dominates, so that the net downward radiative flux at the surface is also larger. This leads to a warmer soil, and through the surface energy budget this also boosts the sensible and latent heat fluxes into the atmosphere. In turn, this heats the atmosphere near the surface by about 0.2 K.

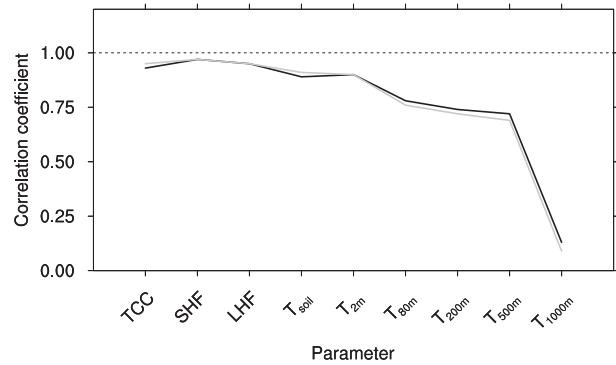


FIG. 6. Correlation coefficients between the monthly-mean model differences in  $SW_d$  and various other variables for the period 2007–2010. Black line represents Pearson's correlation coefficient, while gray line represents Spearman's rank correlation coefficient.

The nature of this impact mechanism is investigated in more detail in Fig. 5, by adopting a special plotting method that differs from Fig. 2 in two ways. First, on the vertical axis the modeled-minus-observed value is now shown. This highlights the difference between models, but it simultaneously maintains information about the measurement. Second, on the horizontal axis the monthly means are now sorted on the associated difference in  $SW_d$  between the two models. As a result, each data point (representing the mean over a specific month) has the same position on the horizontal axis in all panels. What this reveals is that the difference in bias between the two models increases with rank for all parameters in this set—with the largest differences on the right-hand side. This suggests that the impacts on all parameters are related to the change in  $SW_d$ , and that clear correlations should exist between model differences in various variables.

Figure 6 lists the correlation coefficients between the model difference in  $SW_d$  and all other variables. Comparing the degree of correlation among variables provides further insight into where this impact mechanism starts, and how deeply it works its way into the coupled boundary layer–soil system. For example, the model difference in the surface downward shortwave radiation is highly correlated to the model differences in cloud cover. The correlations between  $SW_d$  and the surface heat fluxes are similarly high, reflecting a substantial impact on the surface energy budget. However, the correlation between  $SW_d$  and the soil temperature, as well as the air temperature at 2 m, is already somewhat weaker; and then it further weakens with height above the surface. At 1-km height the correlation coefficient has reduced to only 0.13. In general, the correlation weakens when a process is further down the chain of interacting processes, reflecting that other processes also start to play a role; in the case of air temperature, this

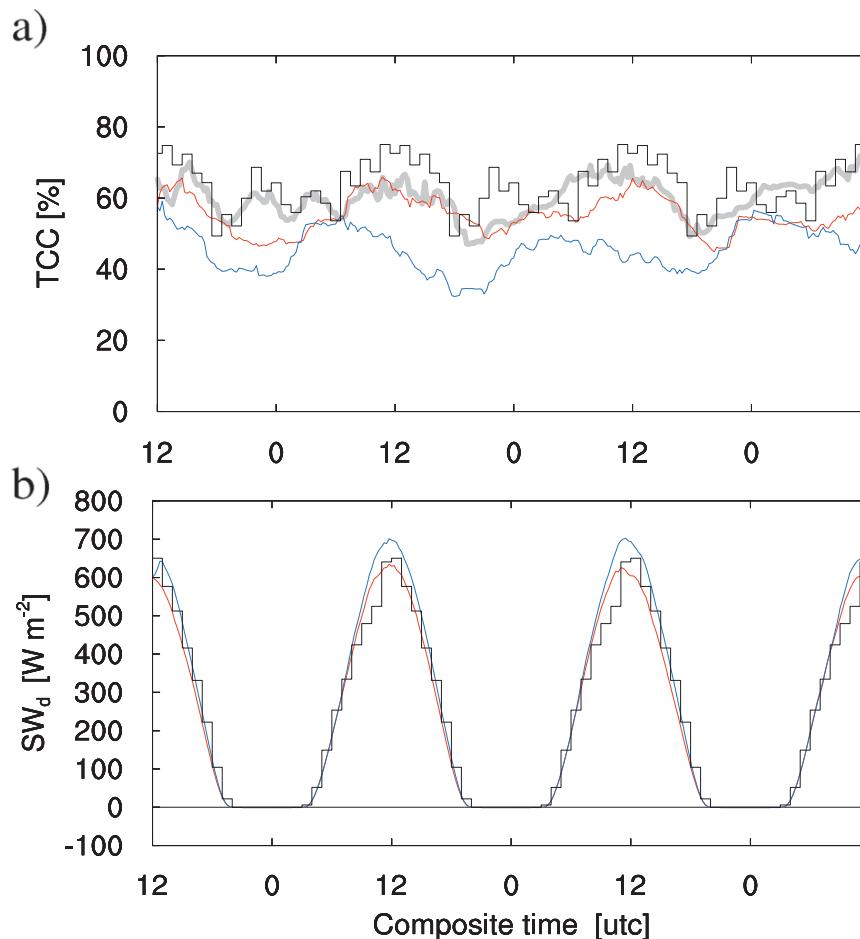


FIG. 7. Monthly-mean composite time series of (a) TCC and (b)  $SW_d$  as measured (solid black) and modeled. Host model state (RACMO) is shown in gray [(a) only], its native SCM in red, and the SCM including the new BL scheme in blue. While the model results represent monthly-mean composites of simulations lasting 72 h, the observations are plotted as repetitions of the monthly-mean diurnal cycle lasting 24 h.

probably reflects a difference in the vertical structure of the thermodynamic state, directly resulting from the use of a different model for boundary layer transport and cloud.

One should always be cautious when interpreting correlations, as correlation is not necessarily equivalent to causality. So, the results by themselves do not prove that the change in cloud cover is the single cause for the change in radiative transfer; changes in other model components could also be responsible. However, it should be noted that both models use the same microphysics and cloud overlap schemes—only the representation of the macrophysical cloud structure was changed. So, the results presented so far do strongly suggest that the change in cloud fraction drives the changes in radiation and other parameters; we now adopt this as our working hypothesis.

### c. Regime sampling

The next step is to gain more insight into the nature of this change in cloud representation, by studying the model behavior at process level. The aim of this exercise is to establish if the change in cloud cover can be attributed at all to fast-acting parameterized processes. Such parameterizations can act on very small time scales, sometimes even close to the integration time step. This should be reflected in the evaluation strategy.

The method adopted here follows the strategy proposed by NSH12. First, the day or days are identified that contribute most to the long-term mean model difference; this choice ensures that the most relevant cases are selected for further study. These cases are then investigated in great detail; for example, we ask if those selected days have anything in common, or define a specific regime. If

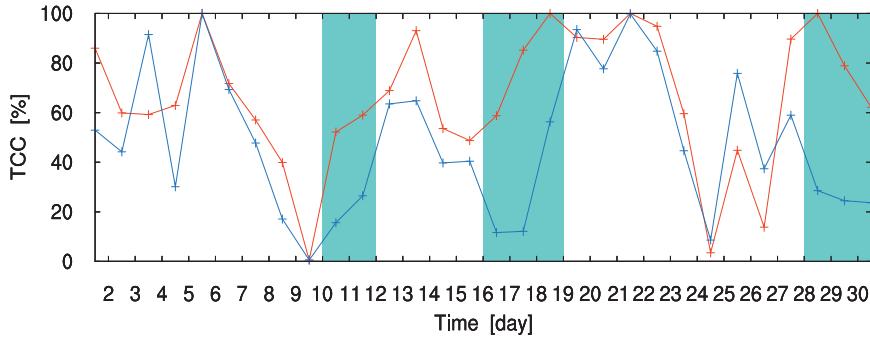


FIG. 8. Distribution by individual days of the monthly-mean TCC at 1200 UTC for June 2008 as shown in Fig. 7a. Light-blue shading highlights the 8 days with the largest intermodel difference.

so, a criterion is defined that reflects this regime. This criterion is subsequently used to calculate conditionally sampled long-term means. If the signal of interest (i.e., the model difference) amplifies for these conditional means, then the corresponding regime is responsible for most of the bias. The benefit of such conditional sampling over long time periods is that it can improve the statistical significance and representativeness of the choice of the “most relevant case” or “golden day.”

First, a single month is selected—for example, using Fig. 5—for which the difference in TCC and  $SW_d$  between the models is large. In Fig. 7 the composite-mean time development of these parameters is shown for June 2008. A well-defined diurnal cycle exists in the model difference for both parameters, with the maximum difference

occurring at about 1200 UTC. The breakdown of this monthly-mean signal into individual days is shown in Fig. 8. The days in this month that contribute most to the monthly-mean model difference are highlighted. To give the reader an idea of actual cloud field in nature on these days, snapshots by the Cabauw webcam at 1200 UTC on each of these days are shown in Fig. 9. All days featured fair-weather cumulus. This would suggest that this boundary layer regime is responsible for most of the bias.

To improve our confidence in this conclusion, we now enhance its statistical significance by assessing the model difference for the subset of days within the period 2007–10 on which fair-weather cumulus occurred. A day is labeled as fair-weather cumulus when the following three criteria apply at 1200 UTC:

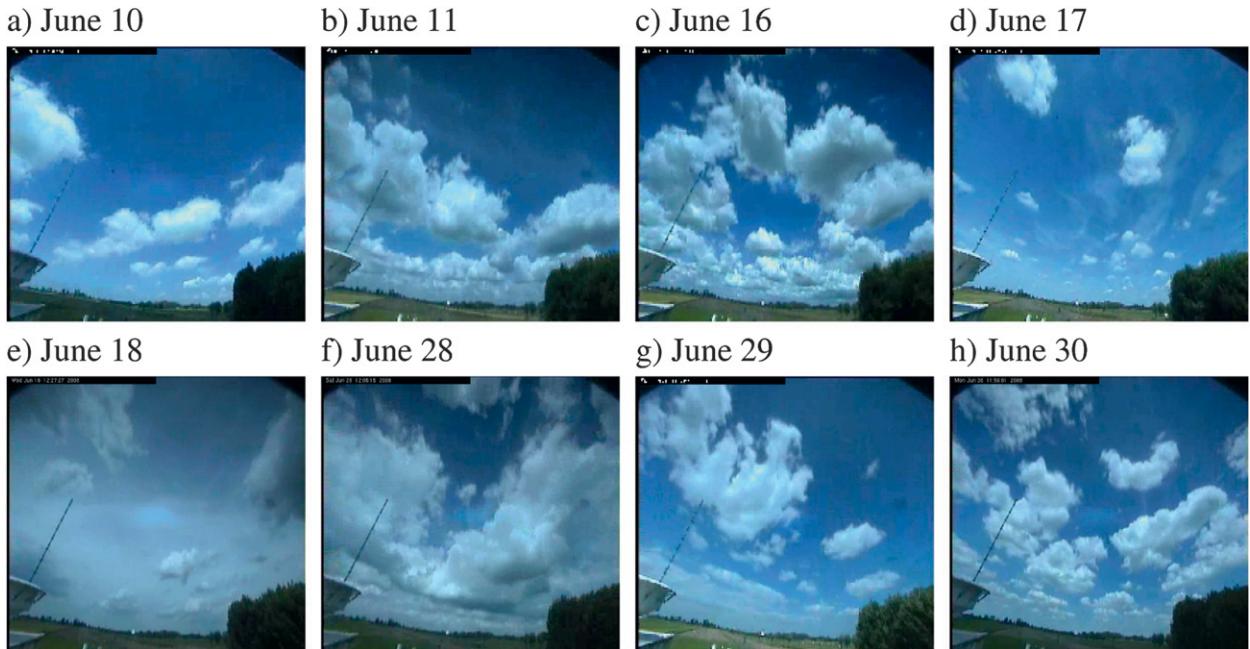


FIG. 9. Snapshots by the Cabauw webcam of the cloud field at about 1200 UTC on the selected days indicated.

- (i) a positive surface buoyancy flux
- (ii) a lifting condensation level below boundary layer top
- (iii) a total cloud cover smaller than 50%

Figure 10a shows the percentage of days per month that were labeled as fair-weather cumulus, using the SCM output of the new scheme. The percentage peaks at about 50% in Northern Hemisphere summer. Figures 10b and 10c show the corresponding model differences in TCC and  $SW_d$ , respectively, for both the full monthly mean and the conditionally sampled monthly mean. For both variables the model differences are consistently larger for the conditional mean. This confirms that the fair-weather-cumulus-topped boundary layer is the regime in which the impact of the model change is strongest.

#### d. Process-level evaluation

We next investigate in detail the representation of boundary layer clouds on the shallow cumulus days selected in the previous section. Figure 11 shows the corresponding time–height cross sections of the cloud fraction in the boundary layer. For reference the LES results are also shown. One difference in cloud structure that immediately stands out is the tendency by the control scheme to create a single-model-level “anvil” at the top of the cloud layer. In contrast, with the new scheme the cloud fraction is at a maximum at the cloud base and decreases monotonically with height above. In addition, the maximum cloud fraction is larger in the original scheme. In these two aspects (i.e., the vertical structure and amplitude), the new scheme better resembles the LES results.

To judge which vertical structure is most realistic, the model results are now confronted with two relevant observational data products available at the Cabauw site; see Fig. 12. The CloudNet dataset (Illingworth et al. 2007) includes profiles of cloud fraction based on independent measurements by multiple instruments (cloud radar, microwave radiometer, and ceilometer). A complicating factor is that at 300 m, its vertical discretization is too coarse to properly resolve the vertical structure of most cases of boundary layer clouds at Cabauw, which are typically less than 1 km deep. The few remaining cases with a deep enough shallow cumulus cloud layer therefore only allow anecdotal (or by case by case) evaluation. For example, Fig. 12a shows that at 1200 UTC 16 June 2008, the shallow cumulus cloud layer was about 2 km deep. In this case the “bottom heavy” profile as created by the LES as well as the new scheme is supported by the CloudNet observation.

The location of the cloud layer boundaries are a first-order aspect of the vertical structure of the cloudy

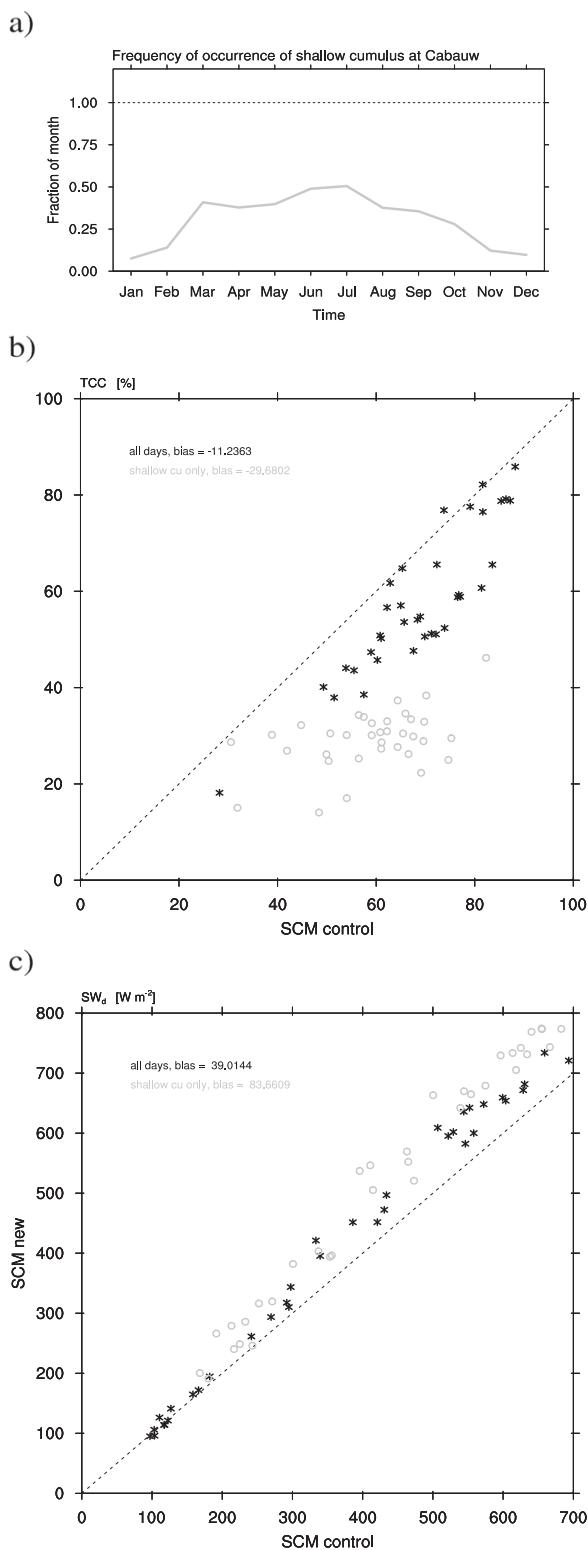


FIG. 10. Impact of conditional sampling on the SCM monthly means for the period 2007–10. (a) Frequency of occurrence of days at Cabauw labeled shallow cumulus as a function of the month in the year. Unsampled (black) and sampled (gray) monthly means of (b) TCC and (c)  $SW_d$  with the control model (abscissa) plotted against the modified model (ordinate).

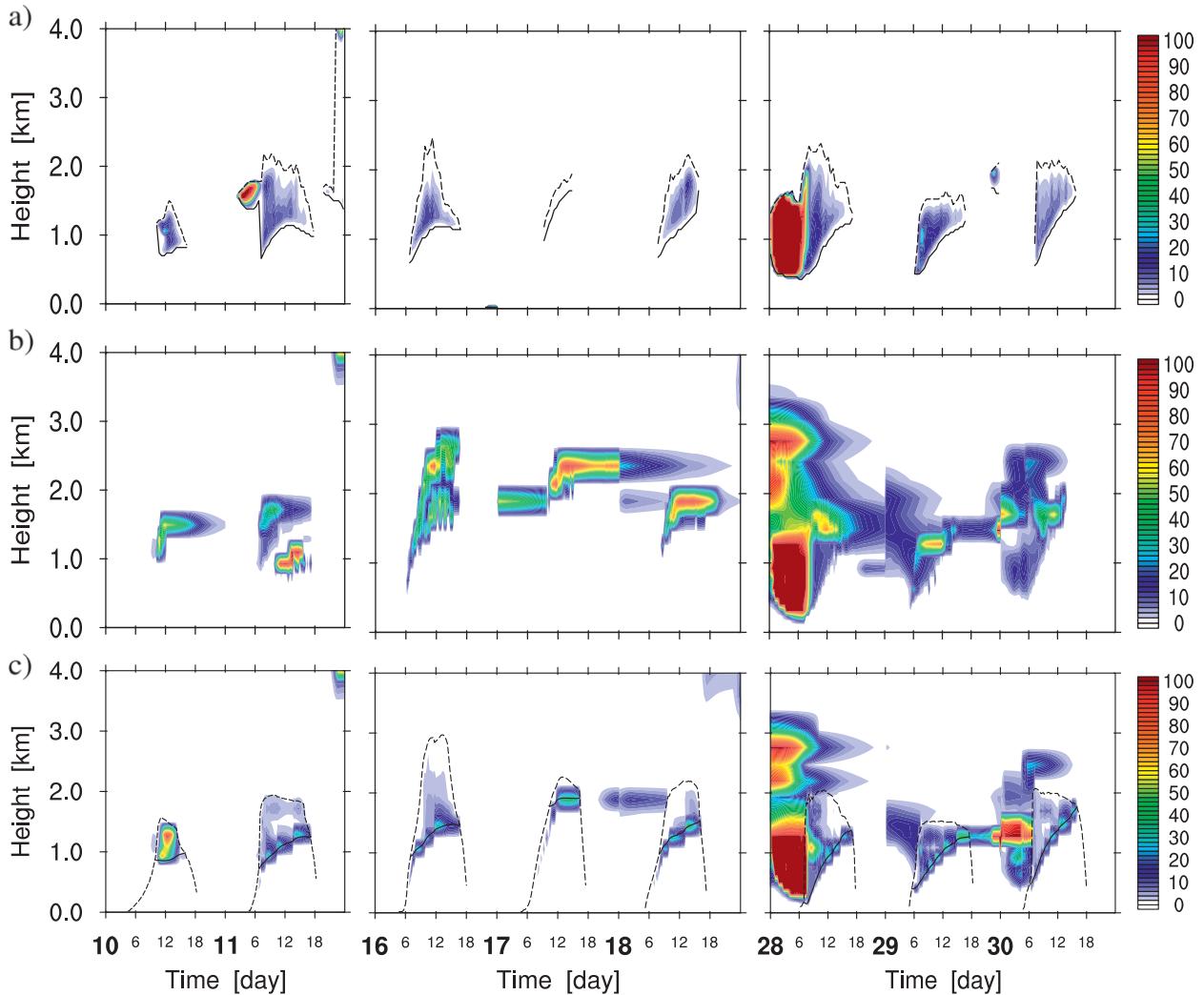


FIG. 11. Time–height contour plots of the cloud fraction (%) as produced by (a) DALES, (b) the control SCM, and (c) the new SCM on (left to right) the days highlighted in Fig. 8. For reference, the base (solid black) and top (dashed black) of the moist convective cloud layer are also shown when available. For the SCMs, these heights are defined as the LCL and the termination height of the model updraft as applied in the BL scheme.

boundary layer. To assess representativeness the LES results are now confronted with high-frequency measurements of the lowest cloud-base height by the LD40 ceilometer at Cabauw. To this purpose the lowest cloud-base height as observed by the LD40 within a 1-h time bin is used, in order to minimize the chance that samples are included that actually represent reflections from the sides of cumulus clouds. The dataset thus obtained should reflect the lifting condensation level (LCL), which is generally used as a definition of the base of the cumulus cloud layer. The evaluation covers the whole month of June 2008, including a daily time window of 4 h around 1200 UTC in order to cover some of the diurnal variation in cloud-base height, as can be observed in Fig. 11. This period also roughly corresponds to the

period of the largest intermodel difference, as visible in Fig. 8. The results are plotted as a probability density function (pdf) in Fig. 12b, suggesting that the LES satisfactorily reproduces the observed LCL with a relatively small bias of 105.6 m, a CRMS of 311.7 m, and a correlation coefficient of 0.65. This correlation is reasonable, given that many factors exist that still prevent a perfect correlation between LES and the measurements, such as (i) uncertainties in the prescribed forcing, (ii) low-level fog due to local terrain features that are not captured by the model setup, and (iii) the sampling method still failing to exclude all samples above the LCL.

Having obtained some confidence in the LES results, the final step is to apply the same method of multiple

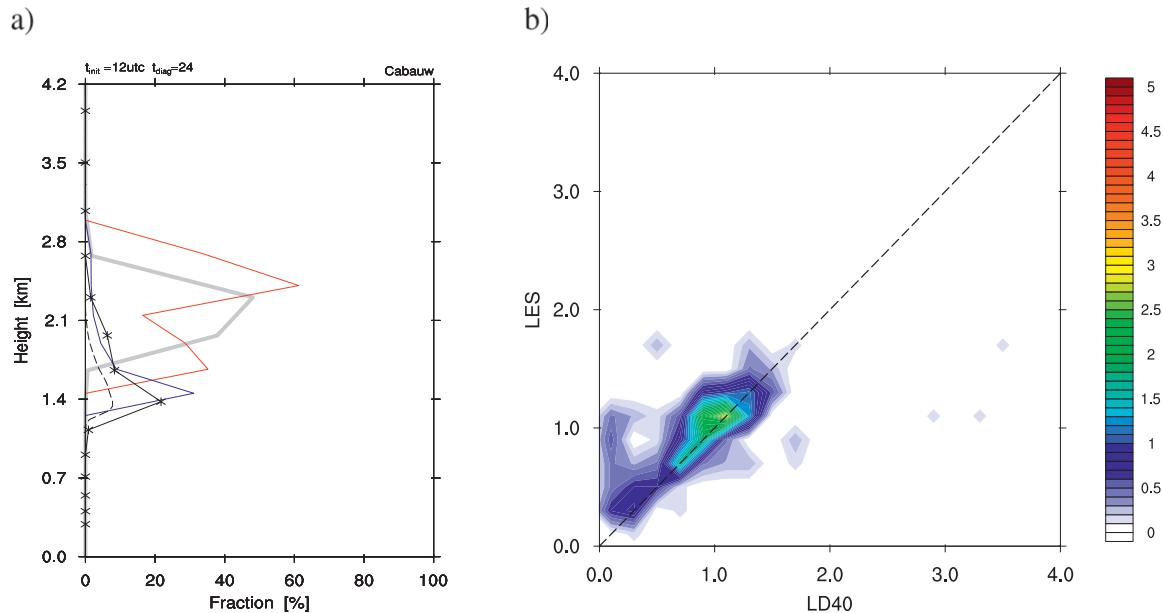


FIG. 12. Comparison of the vertical structure of BL clouds in models and in measurements. (a) Profiles of cloud fraction at 1200 UTC 16 Jun 2008. Solid black line (including asterisks) indicates the CloudNet dataset, while the black dashed line represents the LES result. RACMO is shown in gray, its native SCM in red, and the SCM including the new BL scheme in blue. (b) The pdf of the lowest cloud-base height as diagnosed in the LES (ordinate) vs measurements by the LD40 ceilometer (abscissa). Latter is calculated as the lowest cloud-base height observed within a 1-h time period. Histograms consist of 20 bins within the shown range on each axis. Plotted is  $P/dx dy = 1$ . Period covers June 2008 for the daily time window 1000–1400 UTC.

parameter analysis as previously used in section 3a in the evaluation of the vertical structure of the boundary layer cloud field in the SCMs against the LES. To this purpose an additional set of four relevant variables is defined that reflects the key aspects of the cloud vertical structure in which the two SCM versions differ, as already established in the discussion of Fig. 11:

- the cloud-base height
- the maximum cloud fraction in the boundary layer
- the height of maximum cloud fraction in the boundary layer
- the boundary layer cloud overlap ratio

The latter is defined as the maximum cloud fraction over total cloud cover, both diagnosed over the lowest 4 km. The overlap ratio is included in this set because of its potentially important impact on radiative transfer. The set of four parameters is evaluated using the same setup, time coverage, and visualization method as was applied in Fig. 12b. The results are shown in Fig. 13 and are statistically summarized in Fig. 14. The new model performs significantly better for the cloud-base height, the maximum cloud fraction, and the height of the maximum cloud fraction. The maximum cloud fraction, in particular, seems to be overpredicted by the control model; it is unable to reproduce the small amplitudes typical of fair-weather cumulus as diagnosed in the LES. Interestingly,

both models perform poorly for the cloud overlap ratio, as expressed by the shared large bias for this parameter. On average, the SCMs give  $r_{\text{overlap}} = 1$ , while the LES gives  $r_{\text{overlap}} = 0.5$ ; in other words, the SCMs in effect apply the maximum overlap limit (i.e., total cover equals maximum fraction), while in the LES the overlap is much less efficient.

The better performance by the new model on boundary layer cloud structure, in combination with the worse performance for all other variables, might seem paradoxical at first. However, this apparent contradiction is explained by the shared error on cloud overlap as found in Fig. 13d. This is schematically illustrated in Fig. 15. Similar to most operational GCMs, the maximum-random overlap function (e.g., Geleyn and Hollingsworth 1979; Räisänen 1998) is applied in the radiation scheme in the RACMO. This overlap function was not affected by the implementation of the new boundary layer scheme, so that both model versions use the same overlap function. In the case of the control scheme, the overestimation of the maximum cloud fraction in the boundary layer is compensated by the assumption of too efficient vertical overlap, resulting in a still reasonable estimate of the projected cloud cover (but for the wrong reason). In contrast, the new scheme better reproduces the smaller cloud fractions, as seen in the LES; but in the radiation scheme, this is still combined with the too efficient

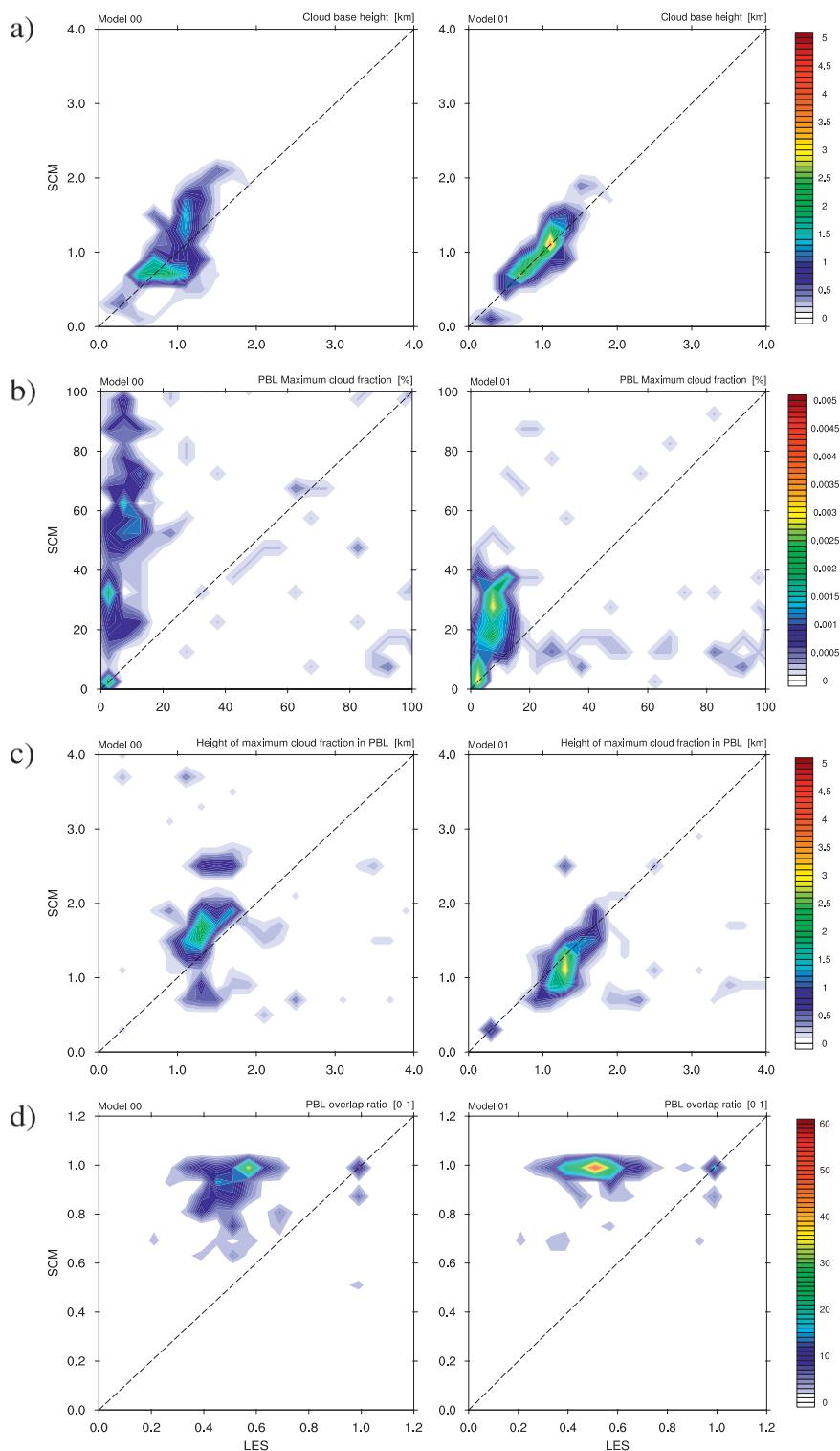


FIG. 13. As in Fig. 12b, except showing LES results (abscissa) vs SCM results (ordinate) for four parameters expressing BL cloud structure: (a) cloud-base height, (b) maximum cloud fraction, (c) height of the maximum cloud fraction, and (d) cloud overlap ratio (as defined in the text). (left) Control model and (right) new model.

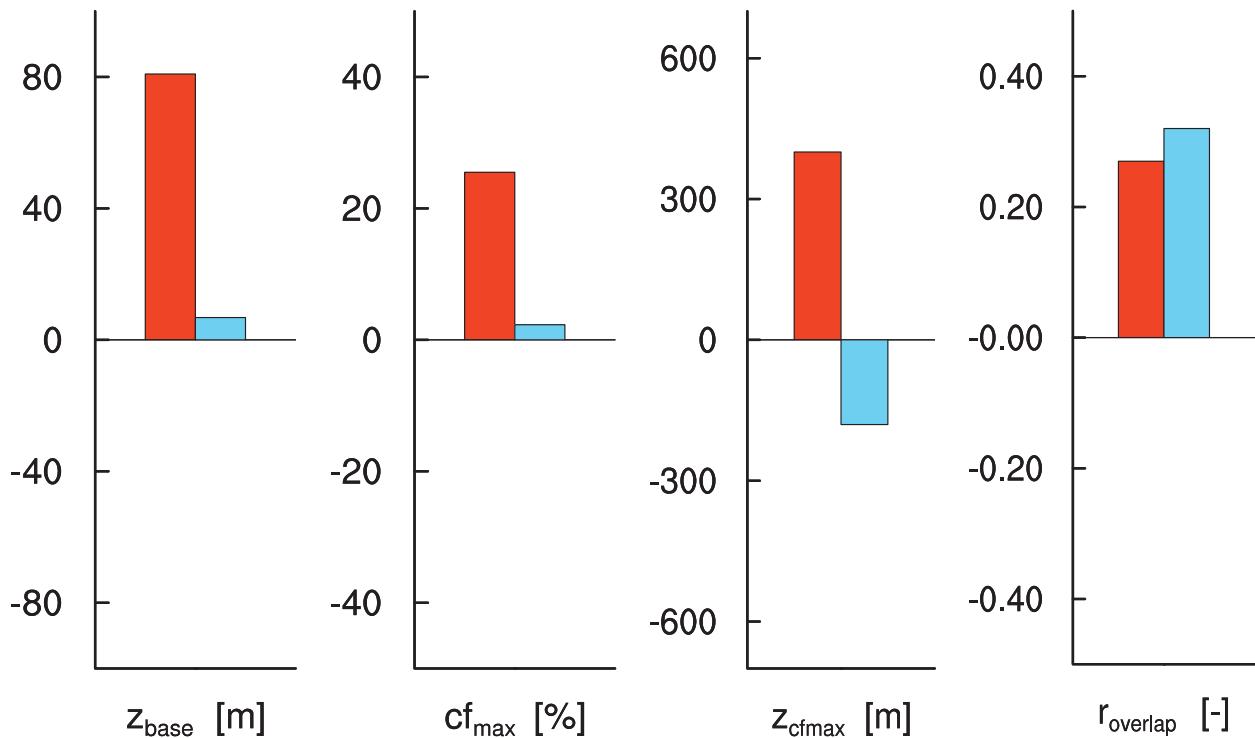


FIG. 14. Bar chart showing the biases of the control SCM (red) and the new SCM (blue) against LES as calculated from the pdfs shown in Fig. 13.

vertical overlap. This results in a structural underestimation of the projected cloud cover, which then leads to an overestimated downward net radiative flux at the surface, with all its further impacts on the state of the coupled boundary layer–soil system, as already documented in the previous sections.

#### e. Model improvement

The insight that errors in the representation of the vertical overlap in cumiform boundary layer cloud fields are responsible for the degraded overall performance by the new scheme has motivated the authors to investigate this phenomenon more closely. The results of that study were published by Neggers et al. (2011), reporting that much of the inefficiency in the vertical overlap already takes place at depth scales that are subgrid scale (SGS) at the vertical discretizations typical of most present-day GCMs. As a consequence, most GCMs do not account for this important phenomenon. They also proposed an inverse linear function to describe the observed behavior of overlap in cumiform cloud fields as a function of layer depth.

The new scheme was rerun for the whole period 2007–10, this time accounting for the overlap on subgrid scales. To this purpose the overlap function as proposed by Neggers et al. (2011) was implemented, which effectively

results in an increase of cloud fraction per level within the boundary layer (as illustrated in Fig. 16a). Figure 17 gives a statistical summary of its performance, in terms of the main characteristics of Fig. 4 (the multiyear bias) and Fig. 3 (the centered RMS difference). Compared to the simulation without the SGS overlap function, the mean bias has reduced significantly for all variables, almost to the level of the control model. The same is true for the centered RMS difference. Note that both errors still remain somewhat larger than those of the control model; this is probably because only the overlap on subgrid scales was altered. A further reduction in the biases can be expected when also the supergrid-scale (SuperGS) overlap (i.e., the overlap between model levels; see Fig. 16b) is improved.

Some operational models have by now adopted the Hogan and Illingworth (2000) parameterization of cloud overlap, which in principle allows for less efficient overlap within the boundary layer compared to the maximum-random overlap approach. Note, however, that the constants of proportionality proposed by Hogan and Illingworth (2000) were derived from analyses of observational datasets using a discretization of 300 m in the vertical, which as a consequence does not capture the inefficient overlap at much smaller, cumiform length scales, as documented by Neggers et al. (2011).

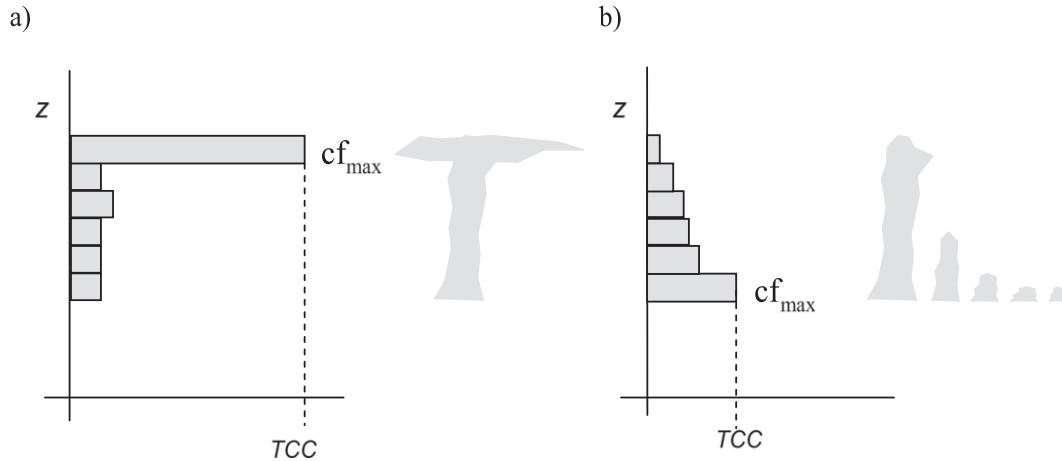


FIG. 15. Schematic illustration of the typical representation of the vertical structure of moist convective BL clouds by the two model versions: (a) control and (b) new. Label  $cf_{max}$  indicates the level of the maximum cloud fraction, while label TCC indicates the corresponding magnitude of the TCC assuming maximum-random overlap. Gray sketches are free-hand expressions of the corresponding cloud field.

**4. Summary and conclusions**

The results obtained in this study illustrate that the multiple parameter evaluation of continuous, long-term SCM simulations can be an efficient method to (i) constrain a system of interacting parameterizations, (ii) trace impacts of model changes throughout this system, and (iii) reveal the existence of compensating errors between parametric components. In the example documented in this study, we applied this method to evaluate the impact of the implementation of a new boundary layer scheme in the RACMO on the cloud-radiative model climate at Cabauw. The “12-point check” revealed the existence of a compensating error in the interaction

of boundary layer clouds and the radiative transfer, residing in the cloud overlap function. Equipping the new scheme with an improved cloud overlap function resulted in much-improved model performance for all parameters.

The results obtained in this study suggest and emphasize that the important phenomenon of vertical overlap in boundary layer cloud fields is still poorly represented in GCMs, and deserves more scientific research. The functionality found by Neggers et al. (2011) captures the inefficient overlap as diagnosed in LES to the first order, a finding used in this study to justify its application in a simple experiment for the purpose of illustrating sensitivity. However, they also documented

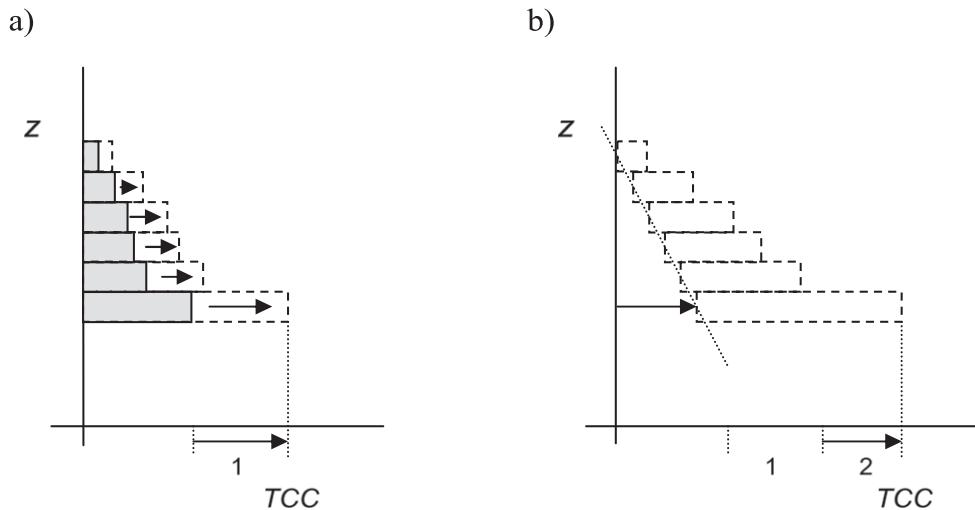


FIG. 16. Schematic illustrations of the impact on cloud fraction per level and TCC by the implementation of (a) a SGS overlap function, and (b) a SuperGS overlap function. Arrows indicate impact.

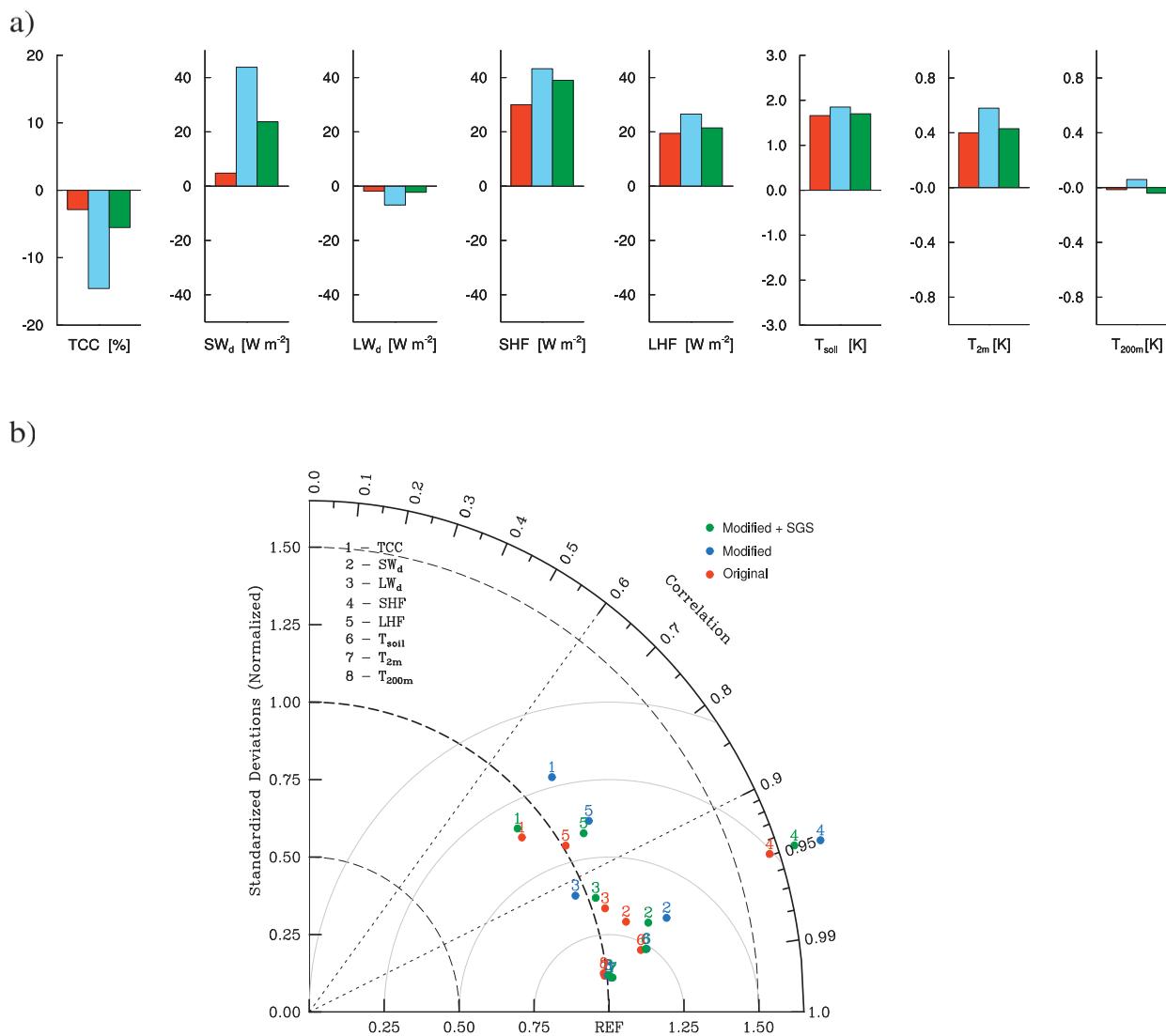


FIG. 17. (a) As in Fig. 4 and (b) as in Fig. 3, except now also showing the new SCM, including the SGS overlap function (dark green).

considerable case dependence in the associated constant of proportionality, which was speculated to be linked to the size statistics of the cumuloform cloud ensemble. More research is needed to fully understand this dependence.

This study focuses on the interaction between boundary layer clouds and radiative transfer, and their impact on the coupled boundary layer–soil system. To this purpose a specific set of measurements was defined for model evaluation. The selection of such sets of multiple parameters really depends on the problem of interest. First, a hypothesis should be formed detailing which of the parameterized processes involved in the problem need to be constrained by measurements. The choice of the site should also depend on the problem of interest; at some locations the associated weather regime occurs

more frequently than at others, making these sites a more logical choice.

Other topics can be studied using the method followed in this study, although some terms and conditions apply. The process of interest should act on time and length scales small enough so that (i) the phenomenon acts much faster than the atmospheric circulation in which it is embedded and (ii) it is “locally forced” enough to allow its study in the absence of interaction with the larger scales. Only then can the problem be addressed with single-column modeling using prescribed large-scale forcings. Examples of topics that could be studied are (i) the representation of momentum transport in the boundary layer, (ii) the humidity budget of the boundary layer (left out of this study for the sake of simplicity and unity of topic), and (iii) impacts of soil moisture on evaporation. An example

of a process that is less appropriate to study is mature deep convection, as this often involves mesoscale effects that might be partially resolved in the associated GCM.

In practice, another limiting factor in multiple-parameter evaluation at process level often proves to be the availability of instrumentation at a site, or the insufficient time coverage of the relevant measurements. The approach described here advocates the long-term, continuous measurement of a range of relevant variables at supersites, and promotes their availability to the scientific community. In this study LES-generated datasets were used to supplement the observational datasets on parameters required to solve the problem. However, one should realize that LES is still a model. It should itself be evaluated against measurements, to increase confidence in its use as a virtual laboratory. This is an ongoing activity at the Cabauw site.

The evaluation of a system of interacting fast-acting parameterizations in isolated mode from the larger-scale circulation against long-term measurements at permanent meteorological sites can facilitate the attribution of GCM behavior to specific parameterizations. For example, it could be of assistance in reducing the uncertainty in numerical predictions of future climate that is known to be caused by the representation of subtropical marine boundary layer clouds. This opportunity is explored in the ongoing European Cloud Intercomparison, Process-Study and Evaluation (EUCLIPSE) project.

*Acknowledgments.* The research presented in this paper has received funding from the European Union, Seventh Framework Programme (FP7/2007-2013) under Grant Agreement 244067.

#### REFERENCES

- Ahlgrim, M., and R. Forbes, 2012: The impact of low clouds on surface shortwave radiation in the ECMWF model. *Mon. Wea. Rev.*, **140**, 3783–3794.
- 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**, 1075–1094.
- Driedonks, A. G. M., H. van Dop, and W. Kohsiek, 1978: Meteorological observations on the 213 m mast at Cabauw, in the Netherlands. Preprints, *Fourth Symp. on Meteorological Observations and Instrumentation*, Denver, CO, Amer. Meteor. Soc., 41–46.
- Geleyn, J. F., and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Contrib. Atmos. Phys.*, **52**, 1–16.
- GEWEX Cloud System Science Team, 1993: The GEWEX Cloud System Study (GCSS). *Bull. Amer. Meteor. Soc.*, **74**, 387–399.
- Heus, T., and Coauthors, 2010: Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geosci. Model. Dev.*, **3**, 415–444, doi:10.5194/gmd-3-415-2010.
- Hogan, R. J., and A. J. Illingworth, 2000: Deriving cloud overlap statistics from radar. *Quart. J. Roy. Meteor. Soc.*, **126**, 2903–2909.
- Illingworth, A. J., and Coauthors, 2007: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bull. Amer. Meteor. Soc.*, **88**, 883–898.
- Jakob, C., 2003: An improved strategy for the evaluation of cloud parameterizations in GCMs. *Bull. Amer. Meteor. Soc.*, **84**, 1387–1401.
- Morecette, C. J., E. J. O'Connor, and J. C. Petch, 2012: Evaluation of two cloud parametrization schemes using ARM and Cloud-Net observations. *Quart. J. Roy. Meteor. Soc.*, **138**, 964–979, doi:10.1002/qj.969.
- Neggers, R. A. J., 2009: A dual mass flux framework for boundary layer convection. Part II: Clouds. *J. Atmos. Sci.*, **66**, 1489–1506.
- , M. Koehler, and A. A. M. Beljaars, 2009: A dual mass flux framework for boundary layer convection. Part I: Transport. *J. Atmos. Sci.*, **66**, 1465–1487.
- , T. Heus, and A. P. Siebesma, 2011: Overlap statistics of cumiform boundary-layer cloud fields in large-eddy simulations. *J. Geophys. Res.*, **116**, D21202, doi:10.1029/2011JD015650.
- , A. P. Siebesma, and T. Heus, 2012: Continuous single-column model evaluation at a permanent meteorological supersite. *Bull. Amer. Meteor. Soc.*, **93**, 1389–1400.
- Paquin-Ricard, D., C. Jones, and P. A. Vaillancourt, 2010: Using ARM observations to evaluate cloud and clear-sky radiation processes as simulated by the Canadian regional climate model GEM. *Mon. Wea. Rev.*, **138**, 818–838.
- Räsänen, P., 1998: Effective longwave cloud fraction and maximum-random overlap of clouds: A problem and a solution. *Mon. Wea. Rev.*, **126**, 3336–3340.
- Ramanathan, V., 1987: The role of earth radiation budget studies in climate and general circulation research. *J. Geophys. Res.*, **92** (D4), 4075–4094.
- Randall, D. A., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003a: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.*, **84**, 1547–1564.
- , and Coauthors, 2003b: Confronting models with data: The GEWEX Cloud Systems Study. *Bull. Amer. Meteor. Soc.*, **84**, 455–469.
- Russchenberg, H., and Coauthors, 2005: Ground-based atmospheric remote sensing in the Netherlands: European outlook. *IEICE Trans. Commun.*, **E88-B**, 2252–2258, doi:10.1093/ietcom/e88-b.6.2252.
- Schalkwijk, J., E. J. Griffith, F. H. Post, and H. J. J. Jonker, 2011: High-performance simulations of turbulent clouds on a desktop PC: Exploiting the GPU. *Bull. Amer. Meteor. Soc.*, **93**, 307–314.
- Sengupta, M., E. E. Clothiaux, and T. P. Ackerman, 2004: Climatology of warm boundary layer clouds at the ARM SGP site and their comparison to models. *J. Climate*, **17**, 4760–4782.
- 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.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106** (D7), 7183–7192.
- Ulden, A. V., and J. Wieringa, 1996: Atmospheric boundary layer research at Cabauw. *Bound.-Layer Meteor.*, **78**, 39–69.
- van Zanten, M., and Coauthors, 2011: Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.*, **3**, M06001, doi:10.3894/JAMES.2011.3.5.