

Single-Column Model Simulations of Subtropical Marine Boundary-Layer Cloud Transitions Under Weakening Inversions

R. A. J. Neggers, A. S. Ackerman, W. M. Angevine, E. Bazile, I. Beau, P. N. Blossey, I. A. Boutle, C. de Bruijn, A. Cheng, J. van Der Dussen, et al.

▶ **To cite this version:**

R. A. J. Neggers, A. S. Ackerman, W. M. Angevine, E. Bazile, I. Beau, et al.. Single-Column Model Simulations of Subtropical Marine Boundary-Layer Cloud Transitions Under Weakening Inversions. Journal of Advances in Modeling Earth Systems, 2017, 9, pp.2385-2412. 10.1002/2017MS001064 . insu-03668335

HAL Id: insu-03668335

https://insu.hal.science/insu-03668335

Submitted on 16 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2017MS001064

Single-Column Model Simulations of Subtropical Marine Boundary-Layer Cloud Transitions Under Weakening Inversions

R. A. J. Neggers[1,2], A. S. Ackerman[3], W. M. Angevine[4,5], E. Bazile[6], I. Beau[7], P. N. Blossey[8], I. A. Boutle[9], C. de Bruijn[2], A. Cheng[10], J. van der Dussen[11], J. Fletcher[8,12], S. Dal Gesso[1,2], A. Jam[13], H. Kawai[14], S. K. Cheedela[15], V. E. Larson[16], M.-P. Lefebvre[13], A. P. Lock[9], N. R. Meyer[16], S. R. de Roode[11], W. de Rooy[2], I. Sandu[17], H. Xiao[18,19], and K.-M. Xu[20]

[1]Institute for Geophysics and Meteorology, Department of Geosciences, University of Cologne, Cologne, Germany, [2]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands, [3]NASA Goddard Institute for Space Studies, New York, NY, USA, [4]CIRES, University of Colorado, Boulder, CO, USA, [5]NOAA Earth System Research Laboratory, Boulder, CO, USA, [6]Météo France/CNRM, Toulouse, France, [7]Météo France/ENM, Toulouse, France, [8]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [9]Met Office, Exeter, UK, [10]NOAA Center for Weather and Climate Prediction, Environmental Modeling Center, College Park, MD, USA, [11]Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, The Netherlands, [12]University of Leeds, Leeds, UK, [13]Météo-France/CNRM & CNRS/IPSL/LMD, Toulouse, France, [14]Meteorological Research Institute, Climate Research Department, Japan Meteorological Agency, Tsukuba, Japan, [15]Department of Atmosphere in the Earth System, Max-Planck Institut für Meteorologie, Hamburg, Germany, [16]Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA, [17]Section of Physical Aspects, European Centre for Medium-Range Weather Forecasts, Reading, UK, [18]University of California at Los Angeles, Los Angeles, CA, USA, [19]Pacific Northwest National Laboratory, Richland, WA, USA, [20]NASA Langley Research Centre, Hampton, VI, USA

Key Points:
- SCM simulations of low-level cloud transitions are confronted with LES results
- Longstanding problems are identified along with encouraging progress
- The model-ensemble median outperforms the individual models

Correspondence to:
R. A. J. Neggers,
neggers@meteo.uni-koeln.de

Citation:
Neggers, R. A. J., Ackerman, A. S., Angevine, W. M., Bazile, E., Beau, I., Blossey, P. N., . . . Xu, K.-M. (2017). Single-column model simulations of subtropical marine boundary-layer cloud transitions under weakening inversions. *Journal of Advances in Modeling Earth Systems*, *9*, 2385–2412. https://doi.org/10.1002/2017MS001064

Received 1 JUN 2017
Accepted 16 SEP 2017
Accepted article online 20 SEP 2017
Published online 27 OCT 2017

Abstract Results are presented of the GASS/EUCLIPSE single-column model intercomparison study on the subtropical marine low-level cloud transition. A central goal is to establish the performance of state-of-the-art boundary-layer schemes for weather and climate models for this cloud regime, using large-eddy simulations of the same scenes as a reference. A novelty is that the comparison covers four different cases instead of one, in order to broaden the covered parameter space. Three cases are situated in the North-Eastern Pacific, while one reflects conditions in the North-Eastern Atlantic. A set of variables is considered that reflects key aspects of the transition process, making use of simple metrics to establish the model performance. Using this method, some longstanding problems in low-level cloud representation are identified. Considerable spread exists among models concerning the cloud amount, its vertical structure, and the associated impact on radiative transfer. The sign and amplitude of these biases differ somewhat per case, depending on how far the transition has progressed. After cloud breakup the ensemble median exhibits the well-known "too few too bright" problem. The boundary-layer deepening rate and its state of decoupling are both underestimated, while the representation of the thin capping cloud layer appears complicated by a lack of vertical resolution. Encouragingly, some models are successful in representing the full set of variables, in particular, the vertical structure and diurnal cycle of the cloud layer in transition. An intriguing result is that the median of the model ensemble performs best, inspiring a new approach in subgrid parameterization.

1. Introduction

Low boundary-layer clouds occur frequently and persistently over the subtropical oceans. Two dominant regimes can be distinguished as embedded in the low-level Trade-wind flow, with stratocumulus situated in the upstream subsidence areas and fair-weather cumulus emerging more downstream (Klein & Hartmann, 1993; Norris, 1998). At some point in the trajectory, a transition from the one cloud type into the other takes place, associated with a significant impact on the transfer of radiative energy. For these reasons, it is important for General Circulation Models (GCMs) to correctly predict the properties of both cloud regimes as well as their spatial distribution, in both present and future climate. A realistic representation of transitions from the one cloud regime to the other is arguably an intrinsic part of this capability.

© 2017. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

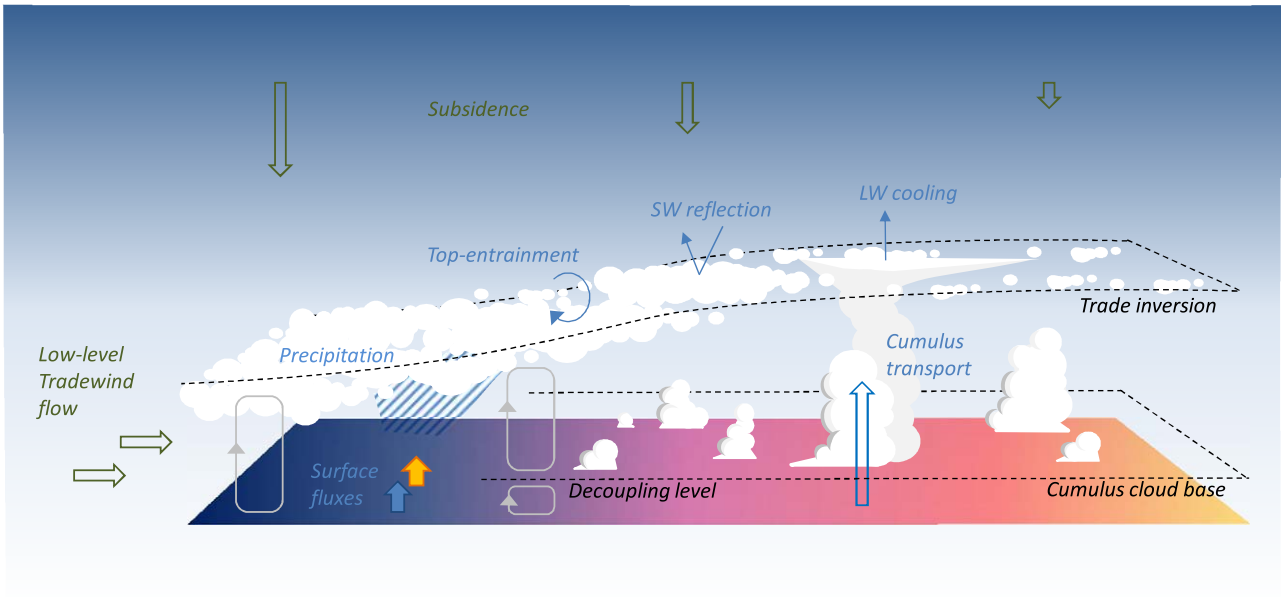


Figure 1. A schematic illustration of the stratocumulus to shallow cumulus cloud transition in the subtropical marine Trade-wind flow. Small-scale physical processes are indicated in blue, while large-scale processes are indicated in green. Various heights within the transitioning boundary layer are indicated in black. The grey lines represent layer-internal circulations. Figure inspired by Albrecht et al. (1995).

These arguments have motivated intense scientific research into low-level cloud transitions in the past (Bretherton et al., 1995; Bretherton & Pincus, 1995; de Roode & Duynkerke, 1997). From these studies, a conceptual picture has emerged that consists of the following sequence of events (Albrecht et al., 1995, see Figure 1). After an initial period of gradual deepening, a thermodynamic decoupling takes place within the originally well-mixed boundary layer, after which a shallow cumulus cloud base emerges below the capping StCu layer. Subsequently the boundary-layer deepening continues, with the capping cloud layer thinning and eventually breaking up. The transitional situation, consisting of shallow cumuli rising into a capping cloud layer, is sometimes recognized as a separate cloud regime (e.g., Norris, 1998; Stevens et al., 2001).

Because of the small timescale and length-scale of the physical processes behind the cloud transition its representation in GCMs is for a large part carried by parameterizations. The first intercomparison study for Single-Column Models (SCM) on the StCu-ShCu cloud transition by the boundary-layer cloud working group of the GEWEX Cloud System Studies (GCSS, Browning, 1993; Randall et al., 2003a) revealed that while most models reproduced a general deepening of the boundary layer, the performance was much worse concerning the cloud amount and vertical structure during the transition (Bretherton et al., 1999). Considerable time has elapsed since this first intercomparison project, during which most boundary-layer schemes in operational weather and climate models have seen significant development. Simultaneously our scientific understanding of the transition process has progressed, due to results from new field-campaigns and studies relying on large-eddy simulation (LES). Examples of recent new insights are the deepening-warming mechanism (Bretherton & Wyant, 1997), the role of the decoupling process (Park et al., 2004; Wood & Bretherton, 2004), and the role of precipitation in cloud breakup (DYCOMS, Stevens et al., 2003). One wonders if our improved understanding has also resulted in a demonstrable improvement in the representation of clouds during the transition in GCMs. This question motivates revisiting this case.

This study presents results of a recent intercomparison project for SCMs on the subtropical marine low-level cloud transition that is designed to address these questions. The project is a joint activity of the Global Atmospheric System Studies (GASS) and the European Union Cloud Intercomparison, Process Study and Evaluation Project (EUCLIPSE). Extensive use is made of LES results of the same cases by multiple codes, the intercomparison of which was recently reported on in great detail by de Roode et al. (2016). The LES and SCM intercomparison studies are integral parts of the same research project, and the experiments have been designed to optimally support each other. While similarities exist between this SCM intercomparison study and its predecessor (Bretherton et al., 1999), there are also important novelties. Simulations are

performed for four cases instead of one, creating a broader parameter space for testing parameterizations. New variables are considered that recent research has suggested reflect key aspects of the transition process. These include the cloud vertical structure, the decoupling process, and the relative stability of the capping inversion. A third novelty is the use of (combinations of) simple metrics to quantify and summarize the model performance for the selected set of variables. Metrics also allow the objective identification of parameterization schemes that yield promising results, both for single variables and for all of them.

Section 2 gives a detailed description of the experimental setup, the participating codes and the method of validation. Key aspects of the low-level cloud transition on which the model evaluation focuses will be introduced in section 3. The results are presented in detail in section 4, making use of various techniques to visualize and quantify model behavior. Further interpretation and discussion are provided in section 5, and the main conclusions are summarized in section 6.

2. Experimental Setup

2.1. Modeling Strategy

Various modeling methodologies have been developed over the years to simulate and understand the transition from stratocumulus to cumulus, and to evaluate their representation in GCMs. One option is to purely use GCM output, by identifying areas of frequent occurrence of these transitions and then confronting the model output with maps of observational data, for example, obtained from cloud detecting satellites (Stephens et al., 2002). More process-oriented options that also allow the use of SCM and LES include the Eulerian approach at fixed locations of interest (Siebesma et al., 2004; Teixeira et al., 2011; Zhang et al., 2013) and the Lagrangian approach, in which the low-level air mass is followed during the transition (Bretherton et al., 1999). A useful advantage of the Lagrangian approach is that the large-scale advective forcings can be considerably simplified. This reduces the potential uncertainty in these forcings, which improves comparability to measurements but also simplifies the simulations. This enhances the transparency of the experiment, which can facilitate the investigation of the potentially complex behavior of small-scale parameterized processes and their interaction. A potential downside of the Lagrangian approach is that the air mass can change shape during its trajectory, for example, through wind shear across the Trade-inversion or large-scale horizontal divergence. Nevertheless, it has been become a commonly used method to study the low-level cloud transition, which also allows interpreting the results of this study in the context of previous ones.

2.2. Case Descriptions

Four Lagrangian cases are defined in this study that describe low-level trajectories during which the marine boundary-layer transitions from a well-mixed stratocumulus-topped state into a decoupled state featuring shallow cumulus convection rising into a capping cloud layer. A detailed description of all four cases was recently provided by de Roode et al. (2016), but will be briefly summarized here.

The first case is based on observations made during the ASTEX field campaign in the North-Eastern Atlantic (Albrecht et al., 1995). The basic setup and the LES results are described in great detail in a companion paper (van der Dussen et al., 2013). The case covers 40 h, starting on 13 June 1992 at 00:00 UTC (23:00 LT) at location (34°N, 25°W) just south of the Azores islands. The other three cases represent composites based on subselections of an ensemble close to 500 trajectories in the northeast Pacific in the period JJA 2007 that were obtained from ERA-INTERIM reanalysis data and MODIS satellite observations by Sandu et al. (2010). From this ensemble, slow, reference, and fast composite cases were constructed by Sandu and Stevens (2011), differentiated by the speed of the low-level cloud transition. These three composites will be referred to as the SLOW, REF, and FAST cases, respectively, in the remainder of this paper. All composite cases have a duration of 72 h, starting on 15 July 2007 at 18:00 UTC (10:00 LT) at location (25°N, 125°W), which is roughly at about 1,000 km off the coast of Baja California toward Hawaii.

What all cases share is their Lagrangian setup; the properties of the low-level air mass are considered as it moves with the Trade-wind flow. This means that in the budget equations for energy, humidity, and momentum the advective tendencies disappear. Model simulations are performed with interactive surface fluxes and radiative transfer. To achieve this, the downward shortwave radiative flux at the top of the

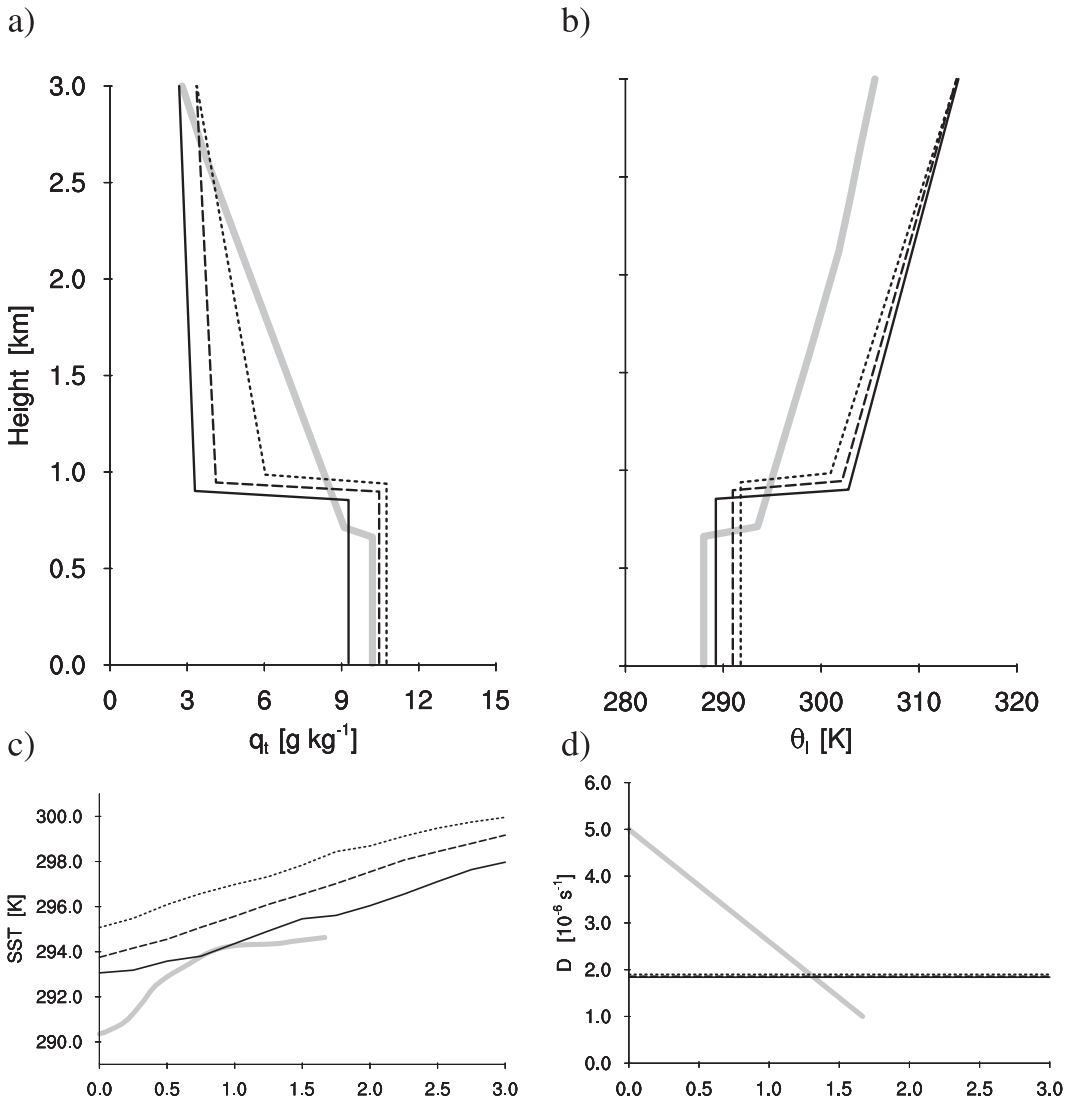


Figure 2. Defining aspects of the setup of the four transition cases. Figures 2a and 2b show the initial profiles of total specific humidity q_t and the liquid water potential temperature θ_l, while Figures 2c and 2d show the time series of the prescribed SST and the large-scale divergence. The solid grey line represents the ASTEX case, while the SLOW, REF, and FAST composite cases are plotted as solid, dashed, and dotted black lines, respectively.

atmosphere is prescribed, following a diurnal cycle that depends on the location on the globe and the date. In all cases, the large-scale subsidence w_s is computed from the prescribed large-scale divergence D, using

$$w_s = -D\ z, \tag{1}$$

from $w_s = 0$ at the surface up to a fixed height with a constant subsidence rate above; this height is 1,600 m in the ASTEX case and 2,000 m in the composite cases. The time-dependent sea-surface temperature is also prescribed. The Coriolis and pressure-gradient forces are retained in the momentum budget, the latter expressed in terms of a prescribed geostrophic wind profile.

Considerable differences exist among the four cases, as shown in Figure 2. Compared to the composite cases, the initial temperature and humidity jumps across the boundary-layer inversion are significantly smaller in the ASTEX case, featuring relatively high free-tropospheric humidity values. Another difference in the ASTEX case is that the prescribed large-scale divergence and geostrophic wind change with time, while they are constant in all three composite cases. In the composite cases, the horizontally averaged thermodynamic and kinematic state variables are continuously nudged toward their initial state, at an adjustment

timescale of 3 h. For temperature and humidity, this nudging is applied above 3 km height, while for momentum it is applied throughout the column. The motivation for the continuous nudging is to prevent excessive drift in the free troposphere.

The NetCDF files containing the configuration of the four transition cases are provided on the EUCLIPSE project website http://www.euclipse.eu/ and in the intercomparison data repository at http://gop.meteo.uni-koeln.de/~neggers/transitions/.

2.3. Participating Models

Boundary-layer cloud parameterizations have evolved significantly since the Bretherton et al. (1999) study. For example, this period has seen the breakthrough of higher-order closure modeling and statistical cloud schemes, of higher-moment microphysics schemes, and of more sophisticated mass flux schemes for the vertical turbulent-convective transport of thermodynamic and kinematic properties. In addition, progress has been made in the research of entrainment. A central research goal in many modeling efforts has been to unify the representation of processes that were previously modeled separately, such as the combined representation of mixing by boundary-layer turbulence and transport by convective plumes in eddy-diffusivity mass flux closures (Siebesma et al., 2007). A relatively new phenomenon is the introduction of stochastic effects as a result of undersampling of the cloud population due to increased GCM resolutions. One of the goals of this intercomparison project is to reflect these developments in boundary-layer cloud parameterization as much as possible, as well as the increased diversity in parameterization approaches that this has created. A large number of participating models is required to this purpose.

Table 1 lists the LES and SCM codes that participate in this intercomparison study. For a detailed description of the LES codes and their configuration, we refer to de Roode et al. (2016). The number of SCM codes is much larger compared to the first intercomparison by Bretherton et al. (1999). These codes represent a wide range of larger-scale models, covering different purposes, discretizations, and domain sizes. Included are both weather prediction and climate models, with domain sizes ranging from global to more limited areas. Operational versions as well as experimental research versions are included, for which various reasons exist. Typically nonoperational schemes have not yet been subjected to the constraints applied to operational schemes, such as skills reflecting the global energy budget in climate simulation and the geopotential position of certain isobars in numerical weather forecasting. As a result, their design and calibration often still reflects a much more process-oriented view of what makes up a good model. Also, some promising new approaches could thus be objectively identified that perhaps deserve to become operational. A more detailed overview of the parameterizations active in each SCM is provided in Table A1. The SCM simulations are performed at the "native" resolution of the associated larger-scale model, so that they represent the behavior of the associated larger-scale model as much as possible. For the ASTEX case, some SCMs performed an additional simulation at a specified L80 resolution. The dependence of the vertical gridspacing with height for all SCMs is visualized in Figure Figure B1.

2.4. LES Results

For reference, the cloud structures as diagnosed from the LES realizations are shown in Figure 3. In all cases, a cloud transition occurs, featuring a capping cloud layer that thins with time and under which at some point shallow cumulus clouds emerge. This transition is accompanied by a continuous deepening of the boundary layer, featuring a diurnal cycle in cloud amount, deepening rate, and cloud layer depth. A diurnal cycle is clearly visible, and has been studied extensively in the past (Duynkerke et al., 2004; Wood et al., 2002). The speed of the transition varies significantly among the four cases, accompanied by different degrees of cloud breakup at the end. The ASTEX case describes the swiftest transition. For a more thorough discussion of these LES results for these cases, we refer to van der Dussen et al. (2013), de Roode et al. (2016), and Sandu and Stevens (2011). van der Dussen et al. (2013) reported a satisfactory skill of the LES codes in reproducing observations of clouds and turbulence made during the ASTEX campaign, with relatively small intermodel spread. This encouraging result is used here to justify the application of LES results as a reference state that the participating SCM codes have to reproduce. For a more general and philosophical discussion on the use of LES, we refer to the discussion published by Stevens et al. (1999) and Stevens and Lenschow (2001).

Table 1
Details of the LES (Upper Plot) and SCM (Lower Plot) Codes Participating in This Study

LES code	References	Investigator	Institute
DALES 3.2	Heus et al. (2010)	J. van der Dussen	TU Delft
DHARMA	Stevens et al. (2002) and Morrison et al. (2005)	A. Ackerman	NASA Goddard
MOLEM	Shutts and Gray (1994) and Abel and Shipway (2007)	A. Lock	Met Office, UK
SAM 6.8.2	Khairoutdinov and Randall (2003)	P. Blossey	UW
UCLA LES	Stevens and Seifert (2013)	I. Sandu	ECMWF

	SCM code	n	ASRF	Investigator	Institute
a	AROME CBR EDKF	24	1111	E. Bazile	Météo France
b	ARPEGE-NWP	24	1111	E. Bazile	"
c	ARPEGE-CLIMAT L31	7	1111	I. Beau	"
d	ARPEGE-CLIMAT L80	24	1111	I. Beau	"
e	ARPEGE-CLIMAT L80 no deep	24	1111	I. Beau	"
f	CLUBB	21	1000	V. E. Larson and N. R. Meyer	UWM
g	ECHAM6	7	0111	S. Kumar	MPI Hamburg
h	EC-Earth	17	1111	S. dal Gesso	KNMI
i	EC-Earth L80	24	1111	S. dal Gesso	"
j	EC-Earth DualM	17	1111	R. Neggers	KNMI & UKöln
k	EC-Earth DualM L80	24	1111	R. Neggers	"
l	HARMONIE[a] CBR noMF	24	1000	W. de Rooy and C. de Bruijn	KNMI
m	HARMONIE[a] CBR EDMFM	24	1000	W. de Rooy and C. de Bruijn	"
n	HARMONIE[a] HARATU EDMFM	24	1000	W. de Rooy and C. de Bruijn	"
o	IFS cy36r1	17	1111	I. Sandu	ECMWF
p	IFS cy36r1 DualM	17	1111	I. Sandu	"
q	IFS cy36r4	17	0111	I. Sandu	"
r	JMA-GSM v1 L60	13	1000	H. Kawai	MRI, JMA
s	JMA-GSM v1 L80	24	1000	H. Kawai	"
t	JMA-GSM v2 L60	13	1000	H. Kawai	"
u	JMA-GSM v2 L80	24	1000	H. Kawai	"
v	JMA-GSM 300s	13	0111	H. Kawai	"
w	JMA-GSM 600s	13	0111	H. Kawai	"
x	LaRC	50	1111	A. Cheng and K.-M. Xu	NASA LaRC
y	LMDZ-AR4	10	0111	M.-P. Lefebvre	LMD
z	LMDZ-THPL	26	0111	M.-P. Lefebvre and A. Jam	"
α	Met Office L63 GA3.0	16	1111	I. A. Boutle	Met Office, UK
β	Met Office L70 GA3.0	23	1111	I. A. Boutle	"
γ	Met Office L38 HadGEM2	10	1111	I. A. Boutle	"
δ	NCEP GFS	15	0111	J. Fletcher	UW
ϵ	RACMO	24	1111	S. Dal Gesso	KNMI
ζ	RACMO L80	24	1000	S. Dal Gesso	"
η	UCLA-AGCM	20	1111	H. Xiao	UCLA
ι	WRF-TEMF	23	1000	W. Angevine	CIRES & NOAA

Note. For the SCMs, the columns give (from left to right) the reference symbol (as used in subsequent figures), the name, the low-level vertical discretization $L_l\langle n\rangle$ (with n the number of levels below 2 km), the availability of model output per case, the associated scientist and the institute at which the model is operated. The simulations with a code at an atmosphere-deep resolution close to L80 are indicated in grey. The availability of SCM results is indicated using a single-row matrix, with the acronym ASRF standing for the ASTEX, SLOW, REF, and FAST cases, respectively, and positive availability indicated with a 1.

[a]HARMONIE here refers to the HARMONIE-AROME model.

3. Evaluation Method

3.1. Metrics

What most SCM intercomparison studies have in common is their double goal of (i) documenting the performance of the various realizations and (ii) comparing them on key aspects of the situation to be represented. More precisely, the SCMs are hoped to reproduce a certain "true" state, in this case as derived from LES, as closely as possible. This study is no different in its aims. However, the presentation of the results is complicated somewhat because of the unusually high number of cases (four instead of one) and participants (34 SCM simulations). An additional goal is to explore methods to summarize the presentation of

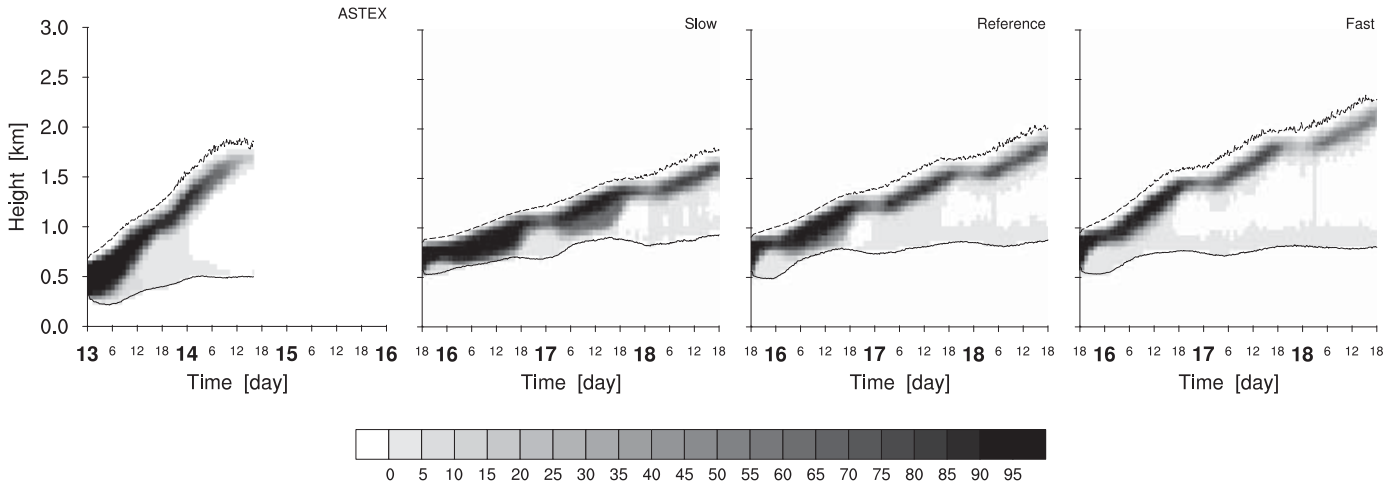


Figure 3. Time-height raster plots of the cloud area fraction a_c (%) during the four transition cases as diagnosed from the ensemble of LES codes. Shown are the ensemble-mean time-height fields. The bold labels on the time-axis indicate the day in the month, while the small labels indicate the UTC time. Local time lags UTC by 1 h in the ASTEX case and 8 h in the composite cases. The grid resolution reflects the bin sizes used for averaging over the model ensemble.

model results, keeping it transparent while still conveying the main message, and without losing important information about individual model performance.

The first step is to focus on the model ensemble instead of individual models. This can be achieved by plotting the distribution of results using a limited number of percentiles, including the median and two outlying values on each side to reflect the intermodel spread. This technique has previously been applied in intercomparison studies for LES models (e.g., Siebesma et al., 2003; Stevens et al., 2001) but also SCMs (Zhu et al., 2005). In case of LES results, the width of the ensemble is a measure for the disagreement between different codes. In case of SCM results, the width can be interpreted as the uncertainty among parameterization schemes in predicting certain aspects of the cloud transition. Note that the median of the model ensemble might perform better than any individual members; ensemble prediction systems as routinely used in present-day numerical weather forecasting rely on this behavior (Epstein, 1969; Leith, 1974).

The second step is to introduce metrics that summarize and express the success of a single SCM in reproducing key aspects of the transition. Two basic, commonly used metrics are applied in this study, mainly in the evaluation of time series. The first metric is the bias, defined as

$$\text{bias} = \frac{1}{N}\sum_{n=1}^{N}\left(\phi_n^m - \phi_n^o\right), \tag{2}$$

where ϕ is the variable of interest, N is the number of samples, superscript m the SCM result, and superscript o the reference result. The bias is a measure for the structural offset relative to the reference signal, in this case the median of the LES results. The second metric is the centered root-mean-square error (crms), defined as

$$\text{crms} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left[\left(\phi_n^m - \overline{\phi^m}\right) - \left(\phi_n^o - \overline{\phi^o}\right)\right]^2}, \tag{3}$$

where the overbar denotes the time-mean over the evaluation period. The crms expresses how well the SCM follows the time-development as diagnosed in the LES. The misprediction of a long-term trend as well as intermittency at short timescales can both contribute to a high crms value. One could combine both metrics into a single "score" S, defined as

$$S^2 = \text{bias}^2 + \text{crms}^2. \tag{4}$$

Score S is an inverse measure of performance that accounts for the capability of the SCM to reproduce both the amplitude (bias) and the time-development (crms) as diagnosed in the LES. The smaller the S, the better

the model performance. This score can be better understood by considering a coordinate system setup by the bias on the horizontal axis and the crms on the vertical. In this score diagram, which somewhat resembles the Taylor diagram (Taylor, 2001) and the diagrams used by Jolliff et al. (2009), score S is the distance to the origin. Such diagrams could be an efficient way to summarize the performance of many individual models in one diagram.

In principle, the ensemble evaluation technique could also be applied to subensembles. These could, for example, be defined to reflect certain model approaches, perhaps yielding insight into the typical behavior of a certain type of parameterization. Indeed the variety in approaches is considerable, as is illustrated by Table A1. However, the behavior of a subgrid scheme as a whole is affected by many individual parameterizations, which might counteract each other. Also, different approaches might effectively yield the same behavior. For these reasons, the evaluation in this study is consciously limited to the model ensemble as a whole. Impact studies of individual parameterizations are for now considered as research efforts that are best conducted within the controlled environment of a single model code.

3.2. A Selection of Key Variables

To evaluate the models, a set of key variables is defined that reflect important aspects of the transition. Table 2 lists all 14 variables, as schematically illustrated in Figure 4. The variables are required to be well-resolved in the LES. Given the importance of cloud representation in climate modeling the set of variables should at least cover cloud amount and vertical structure. Also considered are the state of internal decoupling of the boundary layer and the relative strength of the capping inversion. Finally, the surface energy budget is included, given its important role in (i) the global energy budget, (ii) surface-atmosphere interactions, and (iii) bulk mixed layer interpretations.

The first three variables are the Total Cloud Cover (TCC), the vertically integrated Liquid Water Path (LWP), and the surface precipitation flux P. These are standard measures of cloud amount and condensed water which (i) are relatively easy to calculate in models, (ii) are part of the standard diagnostics package in most operational SCM codes, and (iii) are relatively robust variables as they result from integration over multiple model levels. In the SCMs, TCC is defined as the vertically projected cloud cover in the full atmospheric column, depending on both cloud fraction at each model level and the overlap function. In the LES, TCC is defined as the number of columns that contain cloud condensate divided by the total number of columns in the domain. Another reason for including these variables is to allow comparison to previous evaluation studies that relied on these variables.

The next five variables reflect the vertical structure of the cloud layer. In this cloud regime, the cloud top height z_{top} can well be used as a proxy for boundary-layer depth, which evolves strongly during the transition. Its rate of increase also exhibits a diurnal cycle, best visible in the 3 day composite cases. The cloud base height z_{base} is typically observed to be more or less constant at about 500 m in the Trade-wind regions

Table 2
Overview of the Set of Key Variables Used in This Study to Assess the Performance of the SCMs

Acronym	Description	Units
TCC	Total cloud cover	%
LWP	Liquid water path	g m^{-2}
P	Surface precipitation rate	W m^{-2}
z_{top}	Cloud top height	m
z_{base}	Cloud base height	m
z^1_{max}	Height of first maximum in cloud fraction	m
z^2_{max}	Height of second maximum in cloud fraction	m
h_{cap}	Depth of capping cloud layer	m
κ	Cloud top entrainment instability parameter	
α	Decoupling parameter	
SHF	Surface sensible heat flux	W m^{-2}
LHF	Surface latent heat flux	W m^{-2}
$\mathrm{SW}^s_{\mathrm{net}}$	Net shortwave radiative flux at the surface (positive downward)	W m^{-2}
$\mathrm{LW}^s_{\mathrm{net}}$	Net longwave radiative flux at the surface (positive upward)	W m^{-2}

Note. The variables are defined in the text.

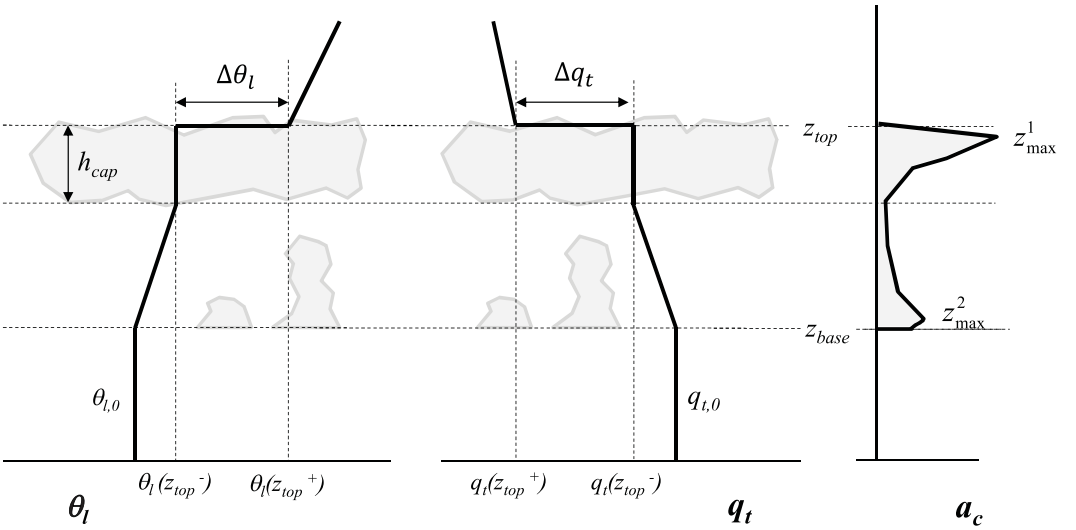


Figure 4. Schematic illustration of the vertical structure of the subtropical marine boundary layer in transition. Various variables as used in Table 2 are shown, including vertical profiles of (left) liquid water potential temperature θ_l, (middle) total water specific humidity q_t, and (right) cloud fraction a_c. Figure inspired by Wood and Bretherton (2004).

(Nuijens et al., 2015; Riehl et al., 1951). The heights of the first and second maximum in the cloud fraction, $z^1_{\max}$ and $z^2_{\max}$, respectively, are other key aspects of the cloud structure that will be evaluated, given their importance in cloud-climate feedbacks (Brient et al., 2015; Vial et al., 2016). Figure 3 shows that the first (largest) maximum is always located near BL top, associated with the capping cloud layer. The second (smaller) maximum is always located near cloud base, representing shallow cumulus clouds rising from the decoupled subcloud layer. The final variable reflecting cloud structure is the depth of the capping cloud layer h_{cap}, which gradually reduces during the transition. In practice, these variables are calculated as follows:

1. The first (i.e., largest) maximum $z^1_{\max}$ has to be located in the top half of the cloud layer.
2. The second maximum only exists when the first maximum exists.
3. $z^2_{\max} < z^1_{\max}$.
4. The capping cloud layer is defined as the layer of depth h_{cap} around $z^1_{\max}$ in which the cloud fraction is larger than 75%.

The fourth condition is adopted to match the definition of h_{cap} as applied by van der Dussen et al. (2013), for consistency.

Two variables are included that reflect important aspects of the vertical thermodynamic structure of the marine boundary layer. The first is the so-called "decoupling parameter," defined by Wood and Bretherton (2004) and Park et al. (2004) as

$$\alpha = \frac{q_t(z_{top}^-) - q_{t,0}}{q_t(z_{top}^+) - q_{t,0}}. \tag{5}$$

The superscripts + and − reflect properties of the air immediately above and below the inversion, respectively (see Figure 4). The larger the α, the more decoupled the cloud layer, with $\alpha = 0$ describing full coupling and the existence of a well-mixed boundary layer. The second variable is the Cloud Top Entrainment Instability parameter (CTEI, Kuo & Schubert, 1988; Lock, 2009; van der Dussen et al., 2014), and is a function of the jumps in both temperature and specific humidity across the inversion

$$\kappa = 1 + \frac{\Delta\theta_l}{(L_v/c_p)\Delta q_t}, \tag{6}$$

where Δ indicates the change in value across the inversion, between z_{top}^+ and z_{top}^-, and L_v and c_p are the latent heat release due to condensation and the specific heat capacity of air at constant pressure,

respectively. In this study, κ is used to evaluate the relative strength of the inversion in temperature compared to that in humidity. When interpreting this ratio, it should be noted that the humidity jump is always negative while the potential temperature jump is always positive (see also Figure 4). The exact method used to diagnose inversion properties from the vertical grid of the SCM and LES results is described in detail by Neggers (2015a).

The final four variables on the list are all energy fluxes at the surface. SHF and LHF are the sensible and latent heat fluxes, respectively. These variables are affected by the thermodynamic state of the surface and the near-surface air, by the near-surface wind and the net radiative flux (through the surface energy budget). The net radiative flux can be further partitioned into contributions by the net shortwave (SW) and longwave (LW) radiation. All fluxes can be expected to be strongly affected by the presence of clouds.

4. Results

4.1. Time-Height Contour Plots

A collection of time-height contour plots of a selection of relevant variables for all cases and for all SCMs are available in the online repository. Those of cloud fraction in the ASTEX case are shown in Figure 5. Similar to Figure 3 raster plotting is used to reflect the vertical and temporal discretizations of the simulations. Almost all models show a deepening of the boundary layer, a result that is consistent with previous intercomparison studies (Bretherton et al., 1999). However, significant differences exist among models in the time-development and amplitude of this deepening. Comparison to Figure 3 suggests that most SCMs underestimate the deepening rate. In addition, the diurnal dependence of the deepening rate that can be distinguished in the LES results does not clearly materialize in most SCM results, with the exception of the LaRC model; in most SCMs, the vertical grid-spacing is simply too coarse to resolve this signal.

The cloud vertical structure differs greatly among models. This is also true for the SLOW, REF, and FAST cases (not shown), and can thus be considered typical behavior. Some models reproduce the gradual time-development of the cloud structure as diagnosed in the LES to a reasonable degree. Other realizations exhibit significant scatter, in that their cloud structure changes significantly at small timescales. This might point at model instability. A more detailed analysis reveals that in particular the gradual thinning of the capping cloud layer as diagnosed in LES appears difficult to reproduce. The emergence of second maximum seems to be captured by a significant fraction of the participating models, which is encouraging.

4.2. Vertical Profiles

The vertical profiles of liquid water potential temperature θ_l and total specific humidity q_t are shown in Figures 6a and 6b. The general structure is captured by the SCM ensemble, featuring a well-mixed subcloud layer below 500 m and a conditionally unstable cloud layer above. However, the inversion height is significantly underpredicted, reflecting the lack of boundary-layer deepening among SCMs as noticed earlier. As a consequence, the boundary layer remains too cold, which is consistent with bulk mixed layer interpretation. In general, the ABL is too well mixed in humidity in the SCM ensemble, with the top of the cloud layer being too humid and the subcloud layer slightly too dry. Note that the impact of salinity on the surface latent heat flux is not accounted for in the LES simulations, which could make their LHF too large and consequently cause a slight moist bias in their subcloud layer (van der Dussen et al., 2013).

Figures 6c and 6d show the vertical profiles of cloud fraction a_c and liquid cloud water q_l. The generally underestimated inversion height is also apparent in these variables. However, the width of the shaded area is much larger, reflecting a large spread among the SCMs. The spread is maximum at about 1.3 km height, which probably also reflects the spread in inversion height among the SCMs as visible in Figure 6. These results emphasize that cloud representation is still problematic in large-scale models, which is in line with the study by Nuijens et al. (2015).

For reference, the vertical profile of the best performing SCM, as identified by its score S, is indicated by the dashed red line. The LaRC model successfully reproduces the vertical structure of both the thermodynamic and cloud state for this period in the ASTEX case, perhaps aided by its relatively high vertical resolution.

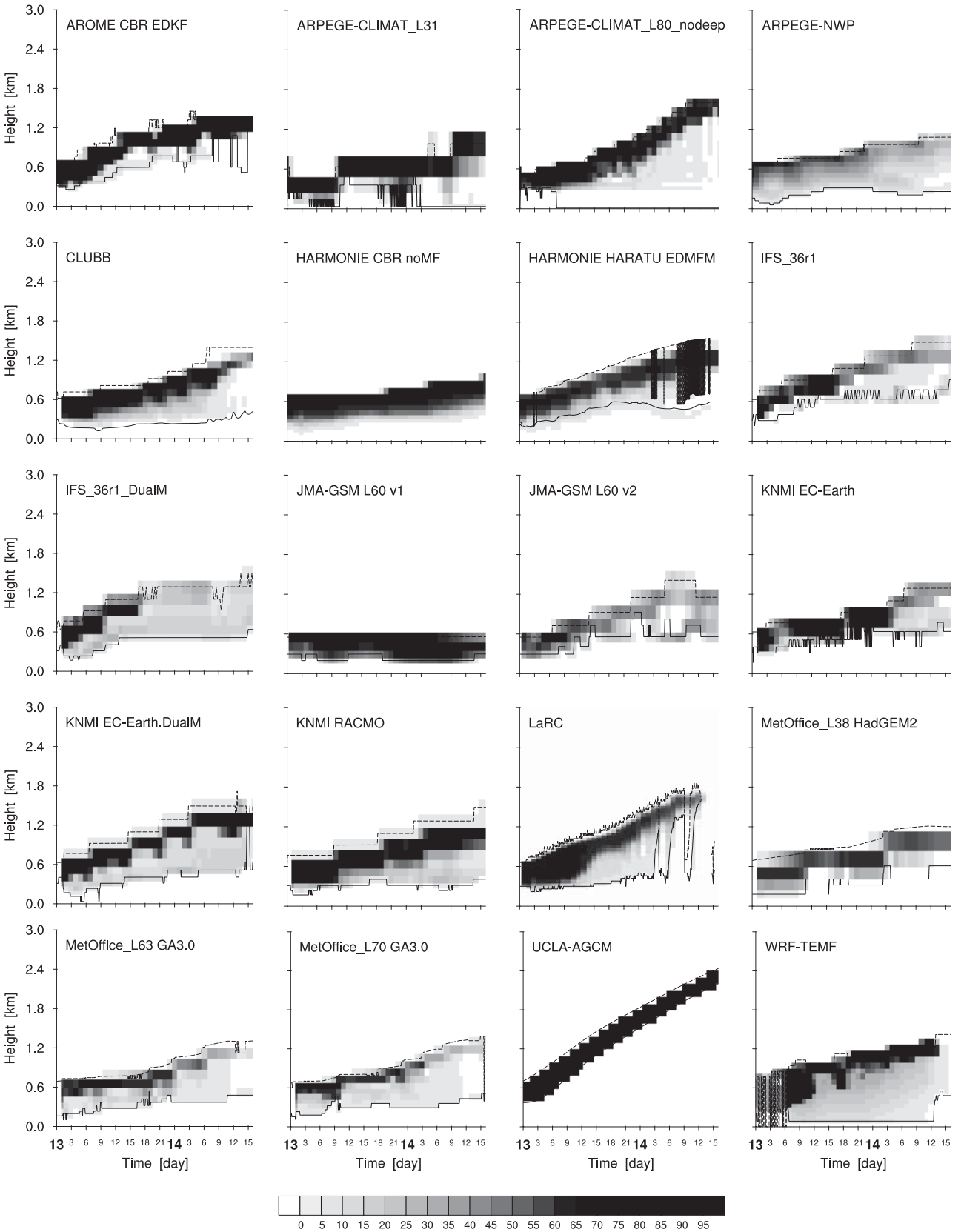


Figure 5. Time-height raster plots of various SCM realizations of the cloud fraction (%) for the ASTEX case. The solid and dashed black lines indicate z_{base} and z_{top}, respectively. The grid resolution reflects the spatial and temporal discretization of each SCM simulation.

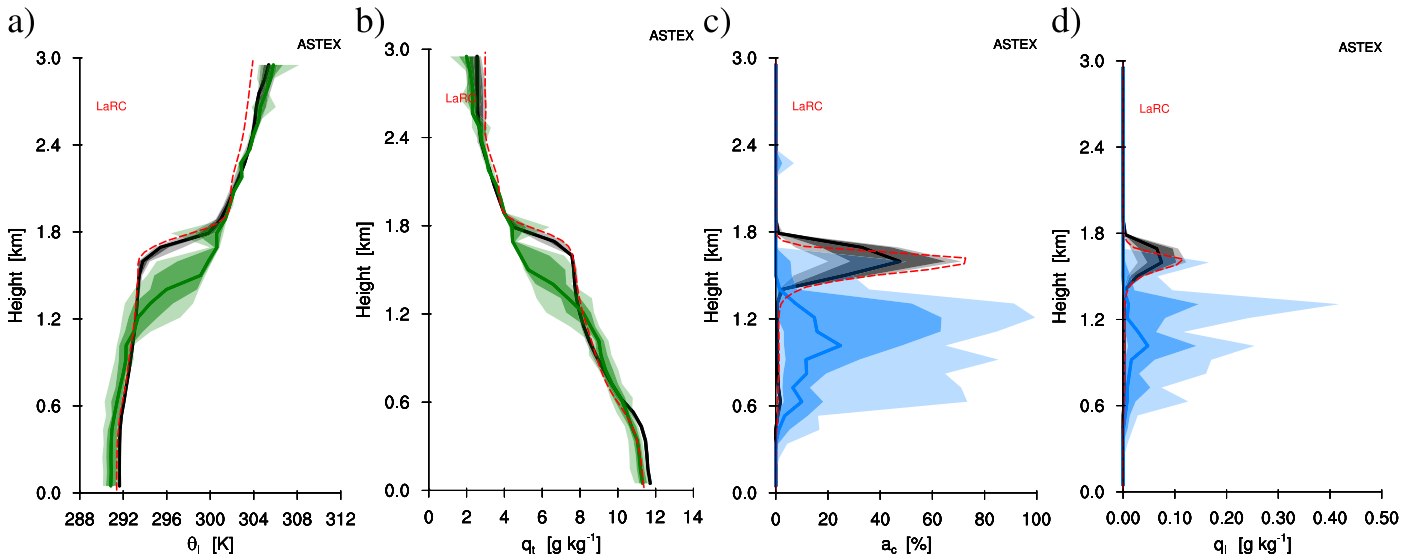


Figure 6. Profiles of the mean thermodynamic state in the ASTEX case. (a) Liquid water potential temperature θ_l. (b) Total water specific humidity q_t. (c) Cloud fraction a_c. (d) Liquid cloud condensate q_l. Model results are time averaged for the period 32–36 h after initialization. The distribution in these variables is obtained using binning in the vertical at 100 m resolution. The LES ensemble is shown in grey, while the SCM ensemble is shown in color. The light shading encloses the 10–90 percentile range, while the darker shading indicates the 25–75 percentile range. The median is indicated by the solid thick line. The profile of the best performing individual model is indicated by the dashed red line.

4.3. Time Series

The representation of the set of key variables as listed in Table 2 is now investigated by means of a time series analysis. For efficiency reasons, only one plot will be shown per variable, focusing on the case in which it shows the largest time-evolution. Score diagrams will be included, allowing an objective identification of the best performing model, with the score defined as the distance to the origin in bias-CRMS space.

4.3.1. Bulk Cloud Properties

Figure 7 shows bulk characteristics of the boundary-layer clouds in transition. These include total cloud cover (*TCC*), the liquid water path (*LWP*), and the surface precipitation (*P*). The FAST case is selected to highlight the first two variables, motivated by their strong diurnal variation. For the surface precipitation, the ASTEX case is chosen, as this is the only case that features significant values. After an initial period of significant values, the surface precipitation reduces substantially in the LES, disappearing completely during the second half of the simulated period. Note that precipitation in LES is parameterized and should be treated with caution, for example, showing dependence on domain size (e.g., Vogel et al., 2016).

For all three variables, the median of the SCM ensemble reproduces some key aspects of the time evolution of the LES median. This includes a gradual decrease with time, as well as a strong diurnal cycle. For TCC and LWP, the diurnal variation takes the form of minima in the afternoons (local time, lagging UTC by 8 h in the composite cases, and 1 h in the ASTEX case) and maxima at night. The SCM ensemble overestimates the LWP throughout the transition in the FAST case. This is an interesting difference with previous SCM intercomparison studies of marine stratocumulus (Duynkerke et al., 2004) and its transition (Bretherton et al., 1999), in which underestimations were reported. As will be shown later, some case-dependence exists in this signal. The SCM ensemble tends to lose cloud cover too quickly, in particular at a later stage in the transition, and has a tendency to overestimate the surface precipitation. However, the spread in TCC is large in the final hours, with some models maintaining full sky coverage until the end.

For all variables, the spread among the SCM realizations is considerable; the distance between the 25 and 75 percentiles is often as large as the mean. This reflects that many individual models have trouble reproducing both the time-development and amplitude of these variables. This spread is also visible in the vertical profiles shown in Figures 6c and 6d, and is consistent with previous model evaluation studies (e.g., van Meijgaard & Crewell, 2005). A positive result is that the best performing model in the SCM ensemble, as identified using the score diagram, reproduces the LES median to a satisfactory degree.

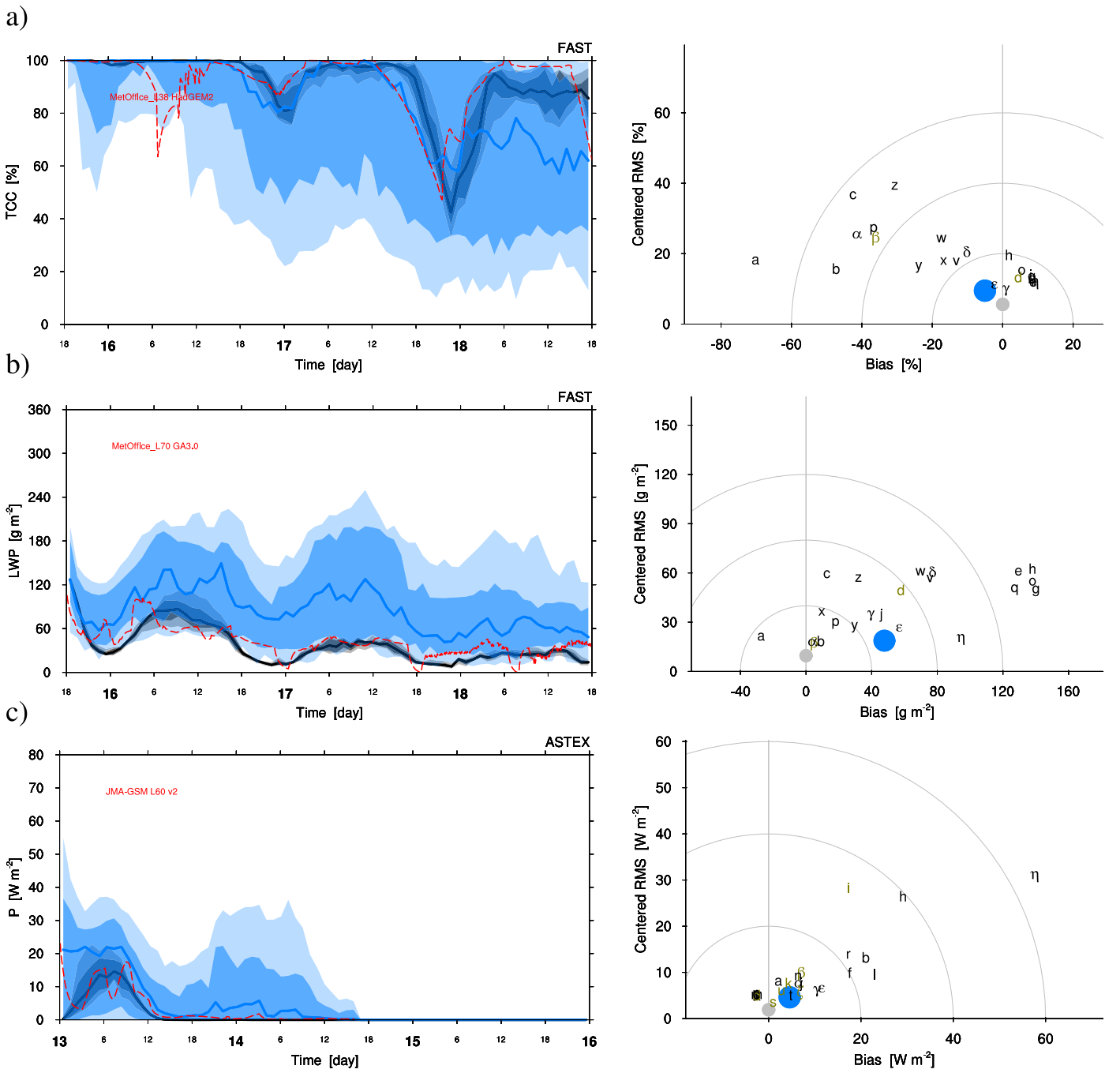


Figure 7. Time series of bulk hydrometeor properties. (a) The total cloud cover (FAST case), (b) the liquid water path (FAST case), and (c) the surface precipitation flux (ASTEX). On the right are the associated score diagrams, plotted with the same units. The grey circles represent isolines of equal score, defined as the distance to the origin, thus giving equal weight to the bias and the crms. The characters identify the model, referring to the first column of Table 1. The SCM ensemble median is plotted as a thick colored dot, while the crms among the LES models is indicated by the grey dot. The L80 simulations are indicated in dark green. The same plotting method will be applied in Figures 8–11.

The associated score diagrams give more insight into the performance of individual SCMs. Many models have a negative bias in TCC (the position on the *x* axis), reflecting that cloud breakup happens too soon. In contrast, many SCMs maintain too much cloud water. This combination of a low TCC-high LWP bias is a well-known compensating error in climate models introduced to calibrate the Earth's energy budget, and is sometimes referred to as the "too few too bright" problem (e.g., Nam et al., 2012; Webb et al., 2001). Finally,

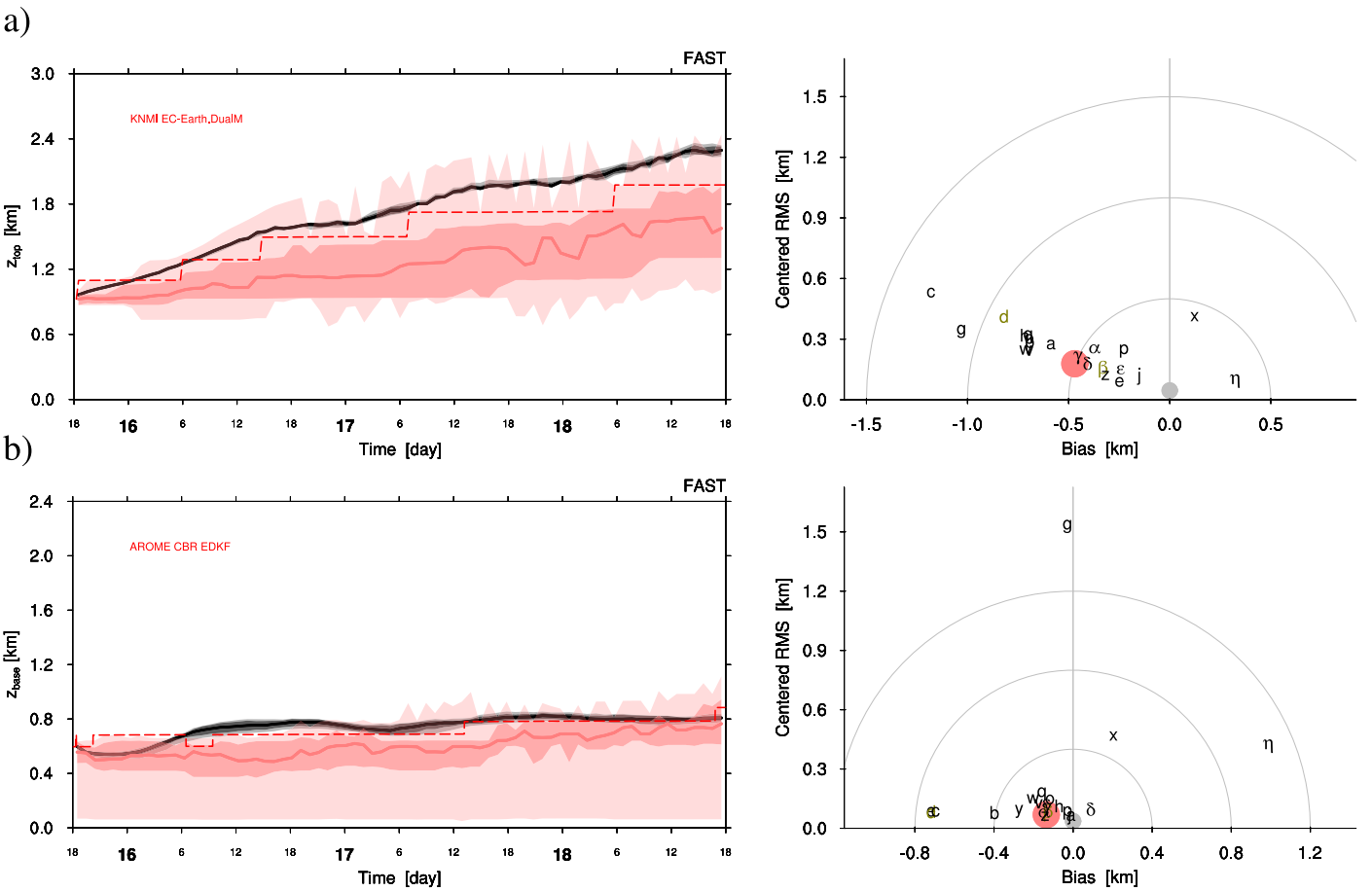


Figure 8. Time series of cloud layer boundaries in the FAST case, including (a) the cloud top height z_{top} and (b) the cloud base height z_{base}.

the SCM ensemble median tends to perform best, as expressed by its relatively close proximity to the origin. Its low crms value contributes much to this skill.

4.3.2. Cloud Boundaries

Figure 8 shows the time series of the cloud boundaries. A slight diurnal variation in z_{top} can be distinguished in the LES results, reflecting the reduction of the entrainment rate during daytime. This feature cannot clearly be distinguished in the SCM ensemble median. In general, the SCM ensemble underestimates the overall deepening rate, as already established earlier. Such lack of deepening was also reported by previous SCM intercomparison studies (Bretherton et al., 1999; Duynkerke et al., 2004), and is apparently still present in many present-day models. Figure 8b shows that the cloud base height z_{base} is captured reasonably well by most SCMs, as reflected by the tightness of the SCM ensemble in the time series plot, as well as the clustering of the points near the origin in the score diagram. However, a few positive and negative outliers do exist, as reflected by the remoteness of the higher percentiles from the median.

4.3.3. Cloud Vertical Structure

Figure 9 shows the three heights reflecting the cloud vertical structure. The height of the first maximum in cloud fraction z^1_{max} is underestimated, and thus reflects the common lack of deepening already established earlier. However, most SCMs correctly locate this height near their boundary layer top (compare to Figure 8a). The height of the second maximum near cloud base, associated with cumulus clouds in a decoupled boundary layer, is captured reasonably well by those SCMs that manage to create one. Note that these include many simulations at L80 resolution, suggesting that a higher resolution is helpful in this respect. What is disappointing is the general lack of performance concerning the thickness of the capping cloud layer h_{cap}, as shown in the last plot. Even the best performing SCMs do not really capture the diurnal cycle evident in the LES results. We speculate that this is due to the lack of vertical resolution near the inversion (see Table 1 and Figure Figure B1). During the advanced phase of the cloud transition the capping

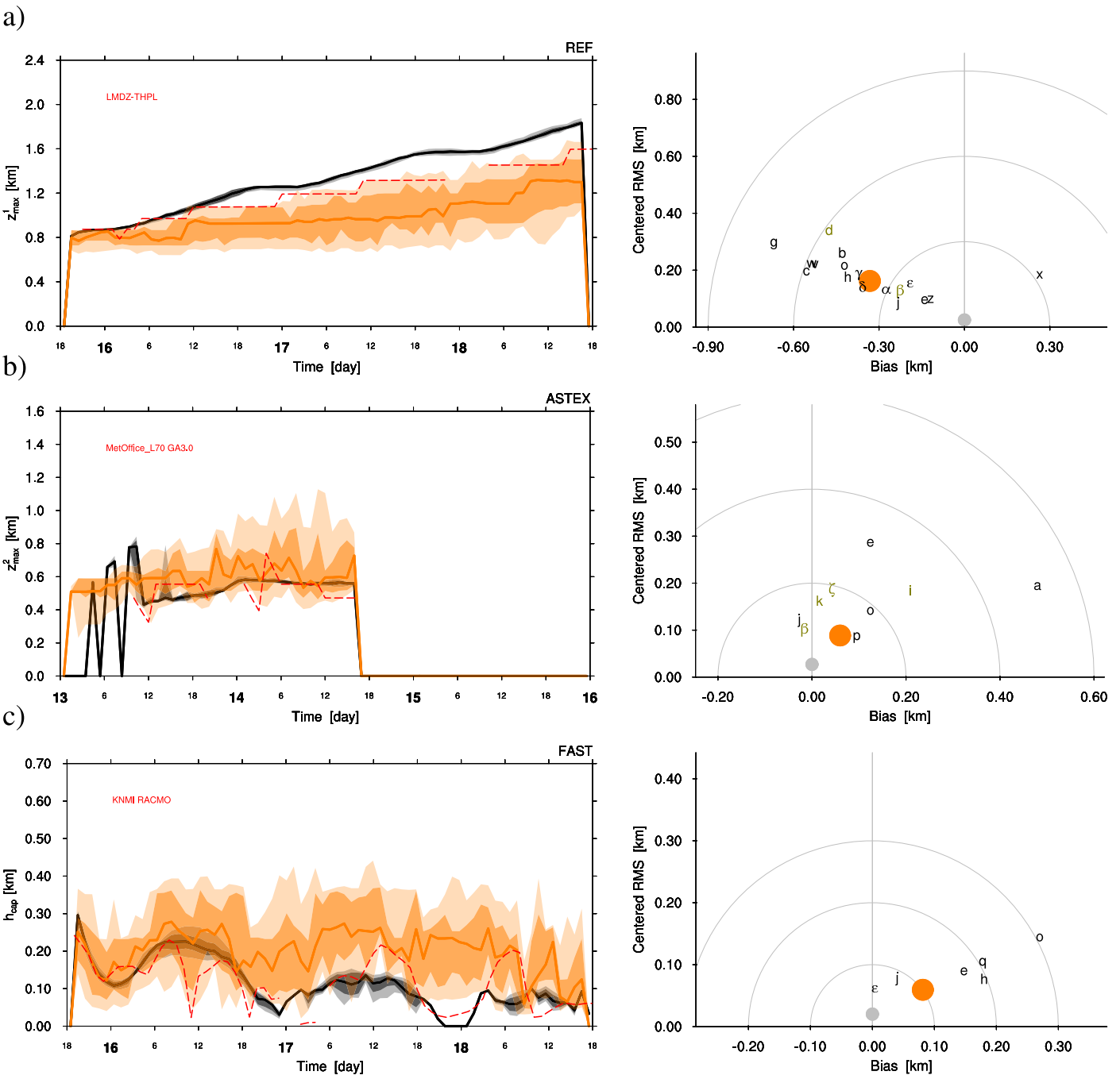


Figure 9. Time series of variables reflecting the cloud vertical structure. (a) The height of the first maximum in cloud fraction z^1_{max} (REF case). (b) The height of the second maximum z^2_{max} (ASTEX case). (c) The depth of the capping cloud layer h_{cap} (FAST case). In contrast to previous figure these variables might not be present at every time step. Accordingly, models are only included when their hit-rate over the time period exceeds 75%.

cloud layer, which typically "carries" most of the total cloud cover, can then become shallower than the vertical gridspacing, introducing dependence on properties at single model levels.

4.3.4. Thermodynamic State

Figure 10 focuses on the decoupling and the relative inversion strength. Most SCMs capture the evolution of the α parameter but underestimate its magnitude, implying that the cloud layer remains too well coupled to the subcloud mixed layer. This behavior is also apparent in the vertical profiles shown in Figure 6. The decoupling process has been associated by some studies with the buoyancy flux becoming negative at this

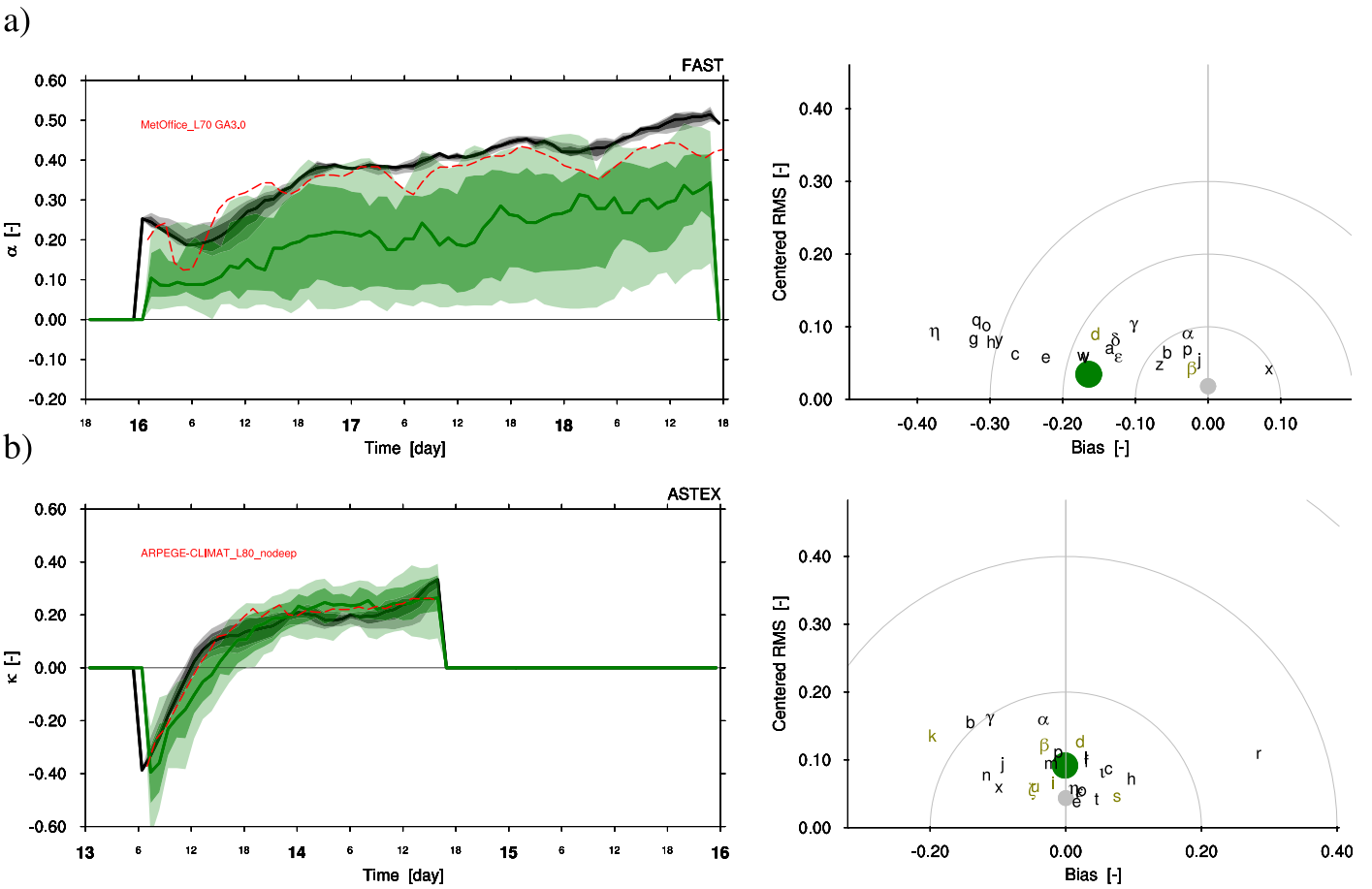


Figure 10. Time series of aspects of the thermodynamic state. (a) The decoupling parameter α (FAST case). (b) The cloud top entrainment instability parameter κ (ASTEX case), as defined by (5) and (6), respectively.

level, which happens when the boundary layer has sufficiently deepened (Bretherton & Wyant, 1997). Both the deepening and the transport through cloud base are fully parameterized in a model; accordingly, it is not guaranteed that a lack of deepening in a GCM is automatically associated with a more coupled state, and vice versa. While the lack of deepening hints at an underestimated effective top-entrainment, the lack of decoupling implies that the mixing through cloud base remains too intense during the transition period.

Figure 10b shows that the ensemble of SCM models reproduces the general time-development of the relative inversion strength as expressed by the κ parameter to a remarkable degree, changing from negative to positive values during the simulated period. In the first phase in which $\kappa < 0.2$, most SCMs slightly underestimate the LES median, which implies that in this phase of the transition the jump in q_t across the inversion is too small compared to the jump in θ_l. The reverse occurs in the second phase. Interestingly, the spread in κ among the LES results is almost comparable to that of the SCM population, much more so than for other variables.

4.3.5. Energy Fluxes

Figures 11a and 11b show results for the sensible and latent heat fluxes. The SHF is overestimated, while the LHF is underestimated (which might be related to the representation of salinity, as already discussed in section 4.2). The SCM median shows a diurnal signal in both heat fluxes, but is overestimated in the SHF but underestimated in the LHF. The gradual increase of LHF during the transition, associated with its deepening, is well-captured. This interaction between deepening and the surface fluxes is probably also responsible for the over(under)estimation of the SHF(LHF); when depth is underestimated, a certain amount of heat and moisture becomes distributed over a shallower layer, which affects the bulk fluxes at the lower boundary.

Figures 11c and 11d show the surface net SW and LW radiative fluxes, respectively. The median of the SCM population underestimates the net SW flux by about 30 W m^{-2}, featuring significant spread. This bias

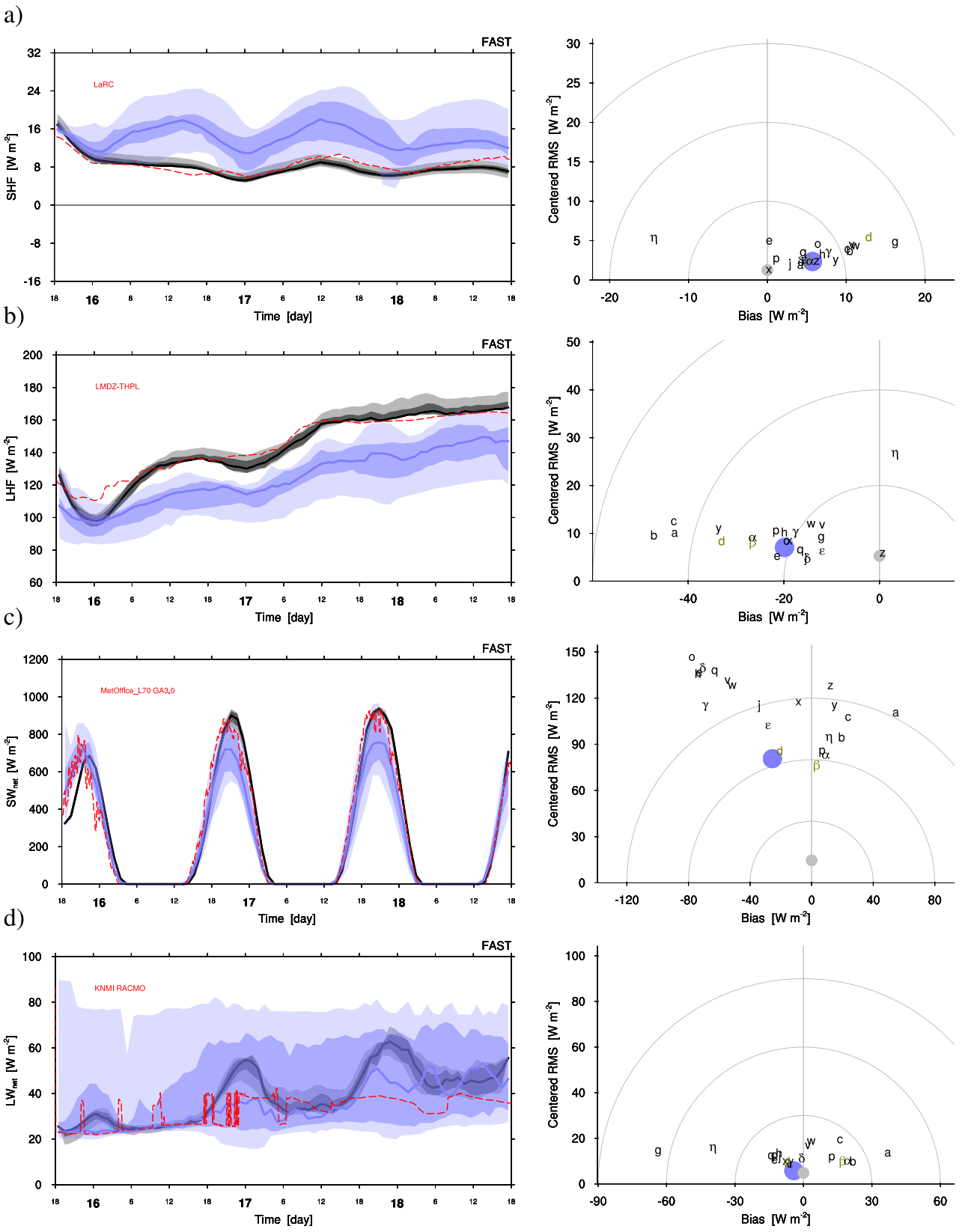


Figure 11. Time series of various surface energy fluxes for the FAST case. (a) The sensible heat flux SHF, (b) the latent heat flux LHF, (c) the net shortwave radiative flux SW_{net} (positive downward), and (d) the net longwave radiative flux LW_{net} (positive upward).

Table 3
Median Bias (Upper Plot) and crms (Lower Plot) of the SCM Population for All Cases Individually (First Four Columns) and for All Cases Combined (Fifth Column)

Variable	Units	ASTEX	SLOW	REF	FAST	All
TCC	%	4.9	−4.4	−11.6	−10.5	−4.4
LWP	g m^{-2}	−2.9	18.5	31.1	47.8	27.2
P	W m^{-2}	6.1	0.3	0.6	0.6	1.9
z_{top}	m	−330	−372	−398	−459	−372
z_{base}	m	−39	−215	−188	−137	−154
z^1_{max}	m	−225	−298	−362	−466	−294
z^2_{max}	m	76	−222	−108	−63	−83
h_{cap}	m	−29	−9	42	49	10
α		−0.132	−0.088	−0.131	−0.157	−0.115
κ		−0.007	−0.015	0.029	0.119	0.014
SHF	W m^{-2}	2.8	7.1	7.6	5.7	5.3
LHF	W m^{-2}	−12.9	−15.1	−15.5	−19.6	−15.6
SW_{net}	W m^{-2}	22.6	4.1	−5.0	−25.5	4.0
LW_{net}	W m^{-2}	−4.8	0.1	−1.6	−4.3	−2.1

Variable	Units	ASTEX	SLOW	REF	FAST	All	All (LES)
TCC	%	18.3	14.2	16.7	16.1	17.7	4.57
LWP	g m^{-2}	62.1	35.4	42.0	36.3	46.3	13.55
P	W m^{-2}	6.5	0.7	2.5	1.7	5.0	0.1
z_{top}	m	189	201	212	246	204	40
z_{base}	m	102	118	109	85	109	34
z^1_{max}	m	176	156	162	217	167	27
z^2_{max}	m	136	84	84	92	96	25
h_{cap}	m	90	65	79	76	77	28
α		0.073	0.075	0.085	0.065	0.075	0.022
κ		0.089	0.099	0.093	0.085	0.091	0.034
SHF	W m^{-2}	3.1	3.1	3.5	3.3	3.2	1.2
LHF	W m^{-2}	8.0	6.4	8.2	8.9	7.9	5.7
SW_{net}	W m^{-2}	131.6	102.6	113.3	114.5	114.7	17.9
LW_{net}	W m^{-2}	13.0	11.0	11.2	11.3	11.8	4.3

Note. The last column in the lower plot shows the median crms among the LES codes, for reference.

represents a diurnal average; the peak differences are much larger, occurring at around noon local time. In the first 6 h of the simulation, the SCMs overestimate the net SW flux, but on subsequent days the flux is always underestimated. This probably reflects the time-development of the LWP as shown in Figure 7b. For the LW flux, the picture is less clear; in general, the magnitude is reproduced reasonably well, but the spread is large, and the diurnal cycle is not well reproduced. Comparison of the score diagrams shows that the bias and crms are much larger in the shortwave compared to the longwave.

4.4. Ensemble Statistics

Table 3 summarizes the overall performance of the SCM population. First, the median of the distribution of SCM values is calculated for each variable and each case, shown in the first four columns. Then the median over all four cases is calculated, shown in the fifth column. For reference, the crms of the LES ensemble is also given.

For some variables, the sign of the bias is identical for all cases. This set includes z_{top}, z_{base}, z^1_{max}, SHF, LHF, α, and *P*. The cloud properties TCC and LWP have an opposite sign in the ASTEX case compared to the composite cases. Figure 3 shows that the early phase of the ASTEX features a relatively thick cloud layer with large LWP amounts, which might contribute to this difference. For many variables, the bias and the crms are smallest for the SLOW case. The smallness of the crms probably reflects that the boundary layer and its clouds do not change that much compared to the other cases. Apparently, the point at which the cloud breakup has advanced significantly is also associated with the worst skill. This result is in line with the study by Nuijens et al. (2015), who also report a large spread in the behavior of GCM parameterizations in regions of shallow cumulus convection.

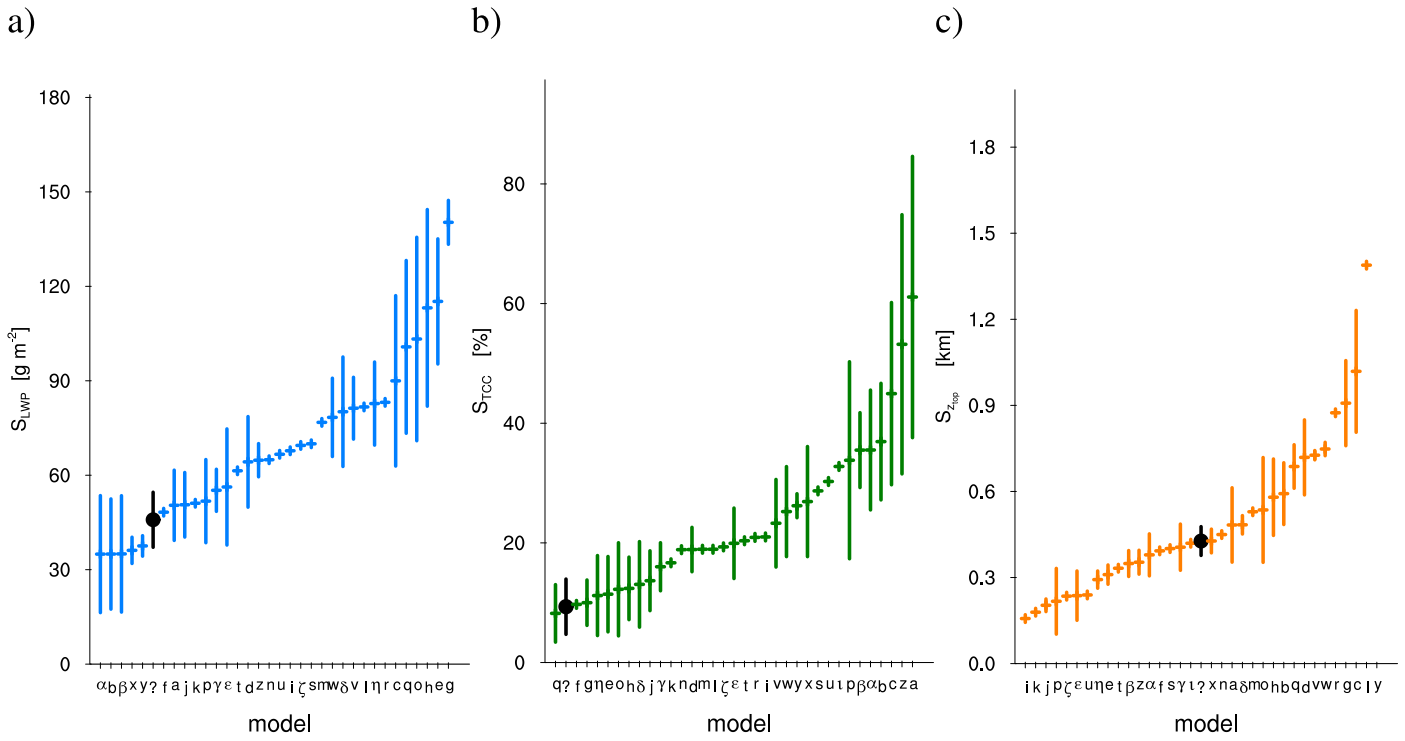


Figure 12. Model scores for (a) LWP, (b) TCC, and (c) z_{top} as averaged over all cases. The models are sorted on their average score. The vertical bar indicates twice the standard deviation. The ensemble median is indicated by the thick black dot and line and by the question mark on the abscissa.

4.5. Persistence and Case-Dependence

The results discussed so far suggest that (i) no SCM consistently outperforms the others, and that (ii) significant case-dependence exists in the ensemble performance. More insight into this behavior is provided in Figure 12, showing the case-average score S for all SCMs individually, for the variables LWP, TCC, and z_{top}. The variation of this score over all cases (the standard deviation) is also indicated. Although the standard deviation is nonnegligible it is always much smaller than the mean. In addition, the difference in mean score between the best and worst performing model is significant. This suggests that the rank of a model is reasonably robust, implying that the skill for one case is reasonably indicative for the others. This result seems to support the use of single case studies (if well defined) for the general improvement of parameterizations. Figure 12 also shows that some models consistently perform as good as or better than the SCM ensemble median. Apparently, despite the large spread among models as reported earlier, some SCMs can actually be identified that structurally show promising results.

4.6. A Multivariable Assessment

The overall skill of a model in reproducing the cloud transition can be expressed by somehow combining its performance for the full set of selected variables and for all four cases. This allows identifying approaches in parameterization that perform well overall, all of the time. This is not trivial; compensating errors introduced by tuning, as well as discrete modeling of separate cloud regimes, can introduce inconsistent scoring across the set of variables. In addition, while a scheme might be "lucky" to score well for one case, doing so for four cases is arguably a lot more difficult. Accordingly, a multivariable approach in evaluation is a much more robust assessment of model performance compared to just considering a single aspect of the transition.

To this purpose a "cumulative rank score," R_i is defined for each model i,

$$R_i = \sum_{v=1}^{V} f_{i,v} \tag{7}$$

where $f_{i,v}$ is the case-averaged fractional rank,

$$f_{i,v} = \frac{1}{C} \sum_{c=1}^{C} \frac{r_{i,v}}{I} \tag{8}$$

First, the models are sorted by their score S for a certain variable v, yielding a rank $r_{i,v}$. Their rank is then normalized by the number of models I, so that a value between {0, 1} is obtained. This fraction is then averaged over C cases, yielding the fractional rank $f_{i,v}$. Then summing over V variables yields the final cumulative rank score R_i. The larger this score, the worse the model performance relative to the other models in the ensemble. The general purpose of the cumulative rank score is not to name bad or good models; instead, the purpose is much more methodological. The main idea is to make use of informative metrics to achieve a comparison of multiple aspects of cloud-topped boundary layers in transition as simulated by a large number of SCMs that is still compact; a multitude of overly dense figures can thus be avoided.

Figure 13a shows the cumulative rank score R_i of all models, with the contribution by individual variables indicated by the coloring. What first catches the eye is that the ensemble median is located at the very top; apparently it is hard for a single model to beat the ensemble effect. However, some models closely match the ensemble performance, scoring well for all key variables. What actually makes them do so deserves further investigation. This is considered beyond the scope of this study, and is left to the associated investigators.

The cumulative rank score R_i for a model i as defined by (7) depends on the choice of variables included in the evaluation. One way to assess the impact of this choice is to apply a bootstrap method with replacement. The first step is to calculate a bootstrap resample R_i^b by randomly selecting the variable to include in the summation

$$R_i^b = \sum_{v=1}^{V} f_{i,w(v)} \tag{9}$$

where $w(v)$ is a random integer between 1 and V. This means that variables can be included multiple times in the summation. Repeating this procedure $B = 100$ times yields a distribution of bootstrap resamples $P(R_i^b)$. The mean and the width of this distribution are shown in Figure 13b, and express how dependent the cumulative rank score R_i is on the variable selection. R_i resembles the mean of the bootstrap distribution to a reasonable degree. However, the width of the bootstrap distribution is still significant; for example, the score range spanned by twice the standard deviation overlaps for the top 9 models. One concludes that, although the cumulative rank score R_i is indicative of the general model performance, it should be interpreted as a loose reflection of relative skill.

4.7. Effective Relations

Some of the key variables listed in Table 2 are not independent, such as the impact of clouds on radiation and those controlled by bulk mixed layer constraints. In GCMs, the representation of such effective relations is at least partially carried by parameterizations. The aim of this section is to interpret the ensemble of SCM results in the context of such theoretical and empirical relations. In doing so a more physical interpretation can be given to the differences among models. The relatively large number of participating SCMs in this study facilitates this effort. Only averages over the final 6 h of each case are considered, because then the differences among models are typically largest, which maximizes the parameter space for fitting functions.

4.7.1. Decoupling

The first effective relation studied is the link between internal decoupling in the boundary layer and the depth of the cumulus cloud layer. The combined process of deepening and warming of the boundary layer during the transition will at some point trigger an internal decoupling between two layers, one dominated by cloud-top cooling driven turbulence and the other by surface-driven turbulence. This process, named the "deepening-warming" mechanism by Bretherton and Wyant (1997), is associated with the emergence of shallow cumulus clouds below the capping cloud deck, expressing the difference in cloud base height between top-driven downdrafts and surface-driven updrafts.

Figure 14a shows a scatter-plot of cloud depth $(z_{\text{top}} - z_{\text{base}})$ versus the decoupling parameter α as calculated from the q_t profile. This definition of cloud depth includes the shallow cumulus clouds that rise into it (as opposed to h_{cap}. de Roode et al. (2016) reported that the LES results follow the functional fit proposed by Park et al. (2004) and Wood and Bretherton (2004) reasonably well for these four transition cases

$$\alpha = \left(\frac{z_{top} - z_b}{\Delta z_s} \right)^{\gamma} \tag{10}$$

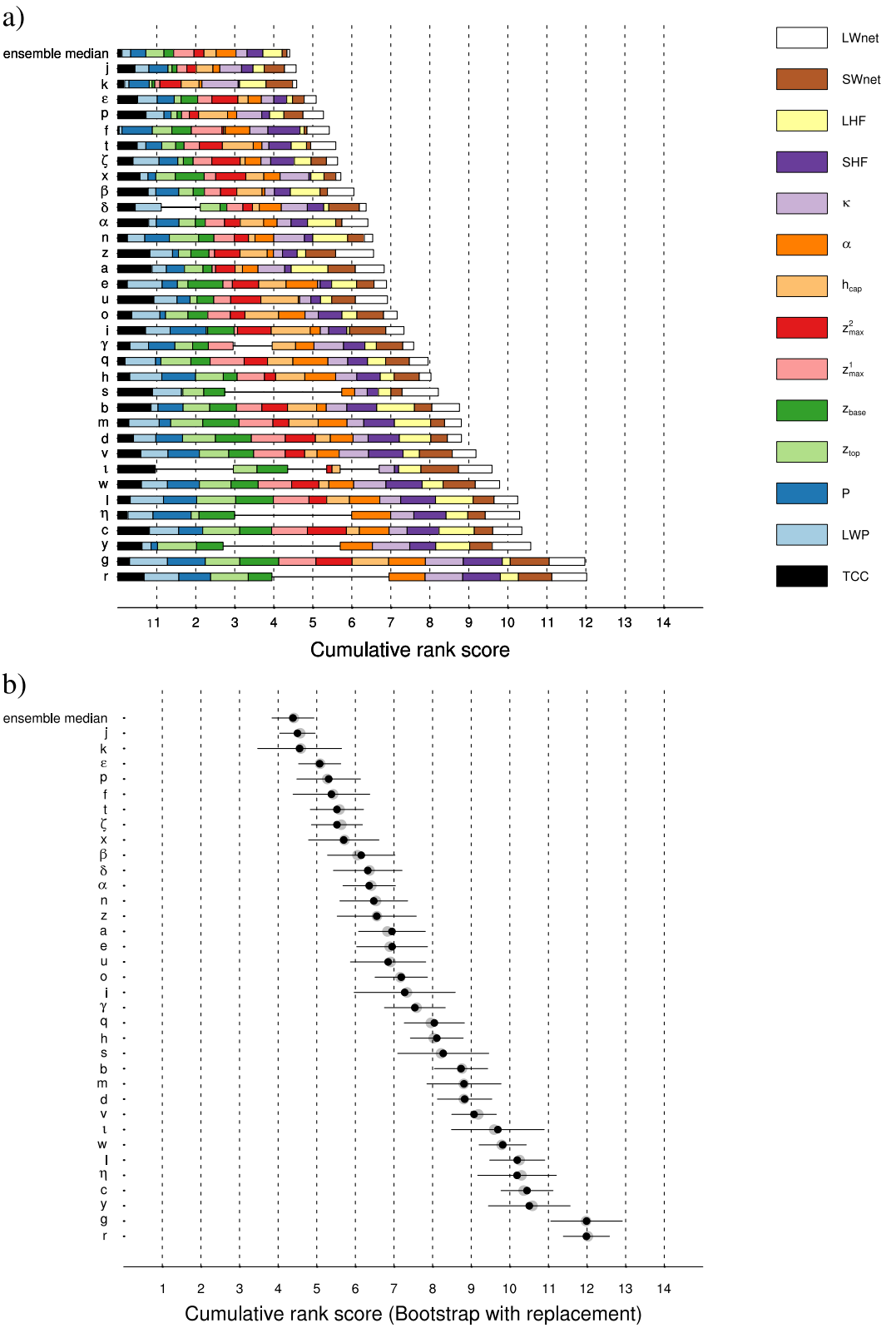


Figure 13. The cumulative rank score R_i of all SCM codes, sorted to place the best performing model on top. (a) The color indicates the contribution by a specific variable. Variables for which no model output is available are given a contribution of 1, and are plotted with a zero width. (b) Compares these results (grey dots) with the mean and width of the distribution of bootstrap resamples $P(R_i^b)$ as obtained from $B = 100$ calculations of (9). The black dots here represent the mean, while the horizontal black lines indicate twice the standard deviation.

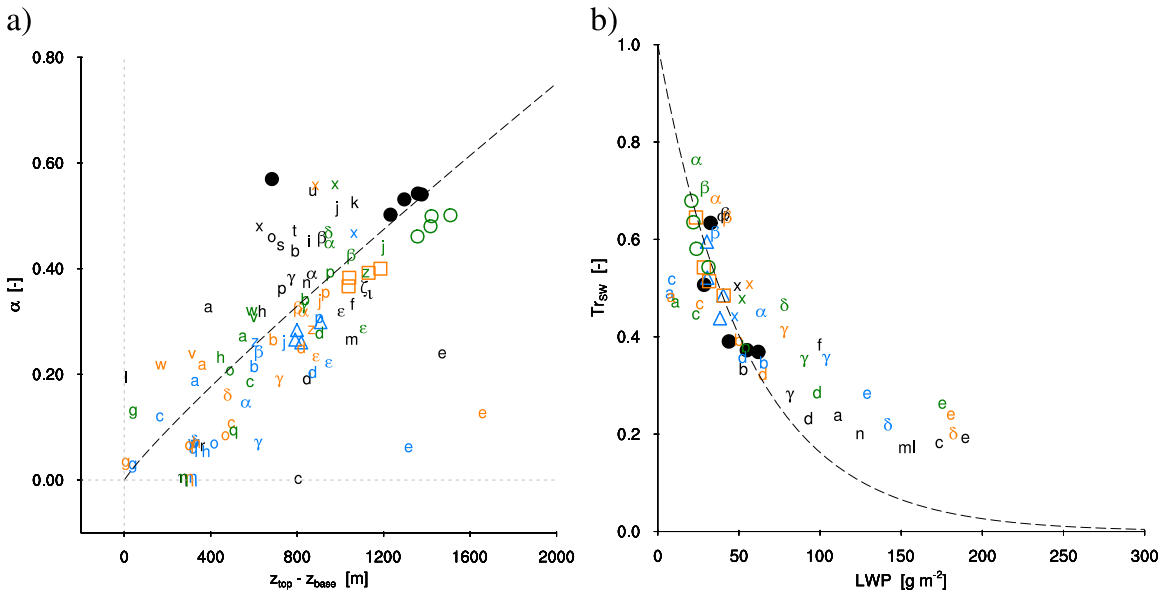


Figure 14. Effective relations between variables representing boundary layer vertical structure. (a) Cloud layer depth versus decoupling. (b) SW transmissivity (over the lowest 2.5 km) versus LWP. The color indicates the case, with ASTEX in black, SLOW in blue, REF in orange and FAST in green. Each symbol represents a time average over the last 5 h of the simulation, with the big symbols representing LES results and the small characters the SCMs as listed in Table 1. As explained in the text, the dashed lines represent either theoretical models or fits proposed in the previous literature.

where scaling parameters $\Delta z_s = 2750$ m and $\gamma = 0.9$ were estimated from sounding data analyses. Comparing the LES results for the SLOW, REF, and FAST cases highlights, the progressed state of the decoupling in the latter. The SCM population also reproduces the relation to a certain degree, although significant scatter exists. In all four cases, most SCMs have not achieved sufficient decoupling and deepening in the final stage of the transitions, which reflects results discussed earlier. But the fact that the SCM population captures the relation between decoupling and cloud depth is remarkable, and encouraging, suggesting that the SCMs produce a realistic boundary-layer structure for a given boundary-layer depth and that errors in decoupling could be related to those in boundary-layer deepening.

4.7.2. Transmissivity

Figure 14b shows the transmissivity of the lowest 2.5 km of the atmosphere for downwelling SW radiation plotted against LWP. The transmissivity is defined here as the ratio of the downwelling SW flux at the surface versus that at 2.5 km height

$$\mathrm{Tr_{sw}} = \frac{\mathrm{SW}_d^0}{\mathrm{SW}_d^{2.5}} \tag{11}$$

The flux at 2.5 km is used because this height corresponds to the top of the LES domains. Because all cloudiness occurs below this height, this definition captures the full impact of low clouds on the downwelling SW radiation. Previous model intercomparisons have made use of transmissivities (Bretherton et al., 1999; Siebesma et al., 2004), the results of which can be used here as a reference. Note that their transmissivities made use of the TOA flux and can therefore be expected to be somewhat smaller.

The LES models show a sharp drop-off in transmissivity at relatively low LWP values, by about 50% within the first 40 g m^{-2}. According to the Beer's law, the transmissivity should be an exponential function of the optical thickness, which depends linearly on the LWP (Stephens, 1978),

$$\mathrm{Tr} = \exp(-a\mathrm{LWP}), \tag{12}$$

Fitting this function to the LES data yields $a = 1/55$. W^{-1} m^2. The SCMs reproduce this relation, which is encouraging, but the spread in transmissivities at a given LWP is significant. This spread is similar to that reported by Siebesma et al. (2004) and Bretherton et al. (1999), and is related in part to variations in cloud cover. However, the SCM ensemble also includes some excessive LWP values (see Figure 7). In this area of the diagram, the exponential fit should be interpreted with caution, because (i) it assumes full cloud

coverage and (ii) the fit is a pure extrapolation from LES results in the low LWP range. Even though LES might resolve cloud inhomogeneity, its microphysics are still fully parameterized. To obtain a more reliable fit, observational transmissivity data covering the full LWP range would be needed.

5. Discussion

One of the aims of this SCM intercomparison study is to explore the use of simple metrics to objectively establish the performance of models and to make an assessment of their overall skill in reproducing a set of key aspects of the marine boundary layer in transition. Apart from condensing a surplus of information into a limited number of metrics, the technique has provided additional insights, for example, into the persistence and representativity of model behavior. In particular, the combination of metrics for a range of variables has proven useful in making a statement about general performance. The use of effective relations between variables proved helpful in explaining some of the common behavior among models, by putting model differences in the context of reference relations from theory or observations and by providing links between model biases in different variables that would otherwise remain unnoticed.

Have models improved since the first intercomparisons on the stratocumulus-to-cumulus transition? Although some differences exist between this study and Bretherton et al. (1999), for example, in the number of cases and participants and in the variables considered, we still can draw some comparative conclusions. Indeed some participating models now show promising capacity in reproducing multiple key aspects of the transition, which is encouraging. One example is the typical bimodal structure in the cloud fraction profile that arises during the transition, with the lower maximum near cloud base reflecting the presence of shallow cumulus. Also represented well by some models is the diurnal cycle typical in this low-level cloud regime. Common deficiencies are also identified, some of which are longstanding. The boundary-layer deepening rate is often underestimated. The degree of internal decoupling is underestimated as well, though its relationship to the cloud layer depth is well-captured by the SCMs. Models tend to transition too quickly toward a low cloud cover regime but maintain too much cloud condensate in doing so, which is likely a manifestation of the well-known too few too bright problem. Another problem is the significant spread in cloud properties that exists across the model ensemble. Apparently some of the major current issues concerning the representation of low clouds in weather and climate models come to the surface in this set of four cases, which suggests that these cases can well be used for process studies to address these problems.

Some reported problems can be attributed to the parameterization of vertical transport in the boundary layer. In particular, the ventilation of heat and moisture through the two interface levels within the transitioning BL, at ABL top and at the decoupling level, appears difficult to parameterize. First, ABL deepening has been linked by Duynkerke et al. (2004) to the top entrainment closure, for which in most models a diffusion model is used. However, advective (mass) flux parameterizations also contribute to the net deepening, if present. This is illustrated by intercomparing two HARMONIE simulations; its noMF run is identically configured to its EDMFM run except that the mass flux scheme is switched off. Including mass flux transport results in a much improved deepening rate in this model. Second, the ventilation at the decoupling level plays an important role in establishing and maintaining the coupled or decoupled state. While the SCMs do apparently reproduce the rough link between ABL deepening and internal decoupling, it remains unclear why this happens, given that the associated transport at both levels is fully parameterized. The cumulus valve mechanism has been found to control transport through cloud base after decoupling (Bretherton et al., 2004; Neggers, 2015b; Neggers et al., 2004). While this mechanism is explicitly represented in some SCMs, in others it might materialize effectively across more time steps, through intermittency in vertical transport. This is a topic for future research.

The results obtained in this study suggest that the vertical resolutions of GCMs are still too low to allow the representation of some subtle but important characteristics of the cloud transition. These include the diurnal cycle in the boundary-layer deepening rate, and the gradually decreasing thickness of the capping cloud layer. These two problems in turn harm the representation of the total cloud cover, both in its diurnal cycle and in the eventual breakup point. Previous studies have shown that situations with strong inversions remain a challenge even for LES models (e.g., Stevens et al., 1999, 2001), so perhaps this result is not too surprising. But an important implication for numerical weather and climate prediction is still that, even when the key (internal) aspects of a transitioning boundary layer are well represented by a parameterization scheme, the total cloud cover is still difficult to get right. In that sense, progress in the conceptual

representation of subgrid physics is still hampered by the numerical reality. Increasing the vertical resolution at inversion height could theoretically solve this problem; but the question remains what would be sufficient.

The good performance of the SCM ensemble median is intriguing; judged over all variables combined it beats all individual models. In particular, the representation of the diurnal time variation contributes to this skill. An interesting wider implication of this result is that performing an ensemble of subgrid scale integrations within a GCM gridbox could be a way forward in improving the representation of low clouds in weather and climate models. Such ensemble subgrid physics modeling could be achieved in two ways: (i) multiple simulations with one physics package, each initialized slightly differently and (ii) multiple simulations with many different physics packages, each initialized identically. This study is in essence an example of the second option: how the ensemble approach would affect the transparency of the link between cloud parameterization and climate feedbacks is also a future research topic.

A result relevant for the evaluation and development of subgrid physics in general is that the SCM performance for one transition case seems representative, to a certain degree, of the performance for the other

Table A1
Summary of the Subgrid Physics Operating in the SCMs

Model name	References	Turb	Entr	Conv-D	Conv-S	Cld-μ	Cld-M	RadSW	RadLW
AROME CBR EDKF	Cuxart et al. (2000) Pergaud et al. (2009)	K1(TKE)			M1	B1	S1		RRTM
ARPEGE-NWP	Bazile et al. (2011)								
ARPEGE-CLIMAT	Bazile et al. (2011)								
ARPEGE-CLIMAT no deep	Bazile et al. (2011)								
CLUBB	Larson et al. (2012)	HOC		HOC	HOC	KK	S2	BugsRad	BugsRad
ECHAM6	Stevens et al. (2008)								
EC-Earth	Hazeleger et al. (2012)	K1(P)	Ri,LW	M1	M1	B1	P		RRTM
EC-Earth DualM	Neggers et al. (2009) Neggers (2009)	K1(P)	Ri,LW		M2	B1	S2		RRTM
HARMONIE CBR noMF	Cuxart et al. (2000)	K1(TKE)				B1	S1		RRTM
HARMONIE CBR EDMFM	Cuxart et al. (2000) De Rooy and Siebesma (2008)	K1(TKE)			M2	B1	S1		RRTM
HARMONIE HARATU EDMFM	De Rooy and Siebesma (2008) Bengtsson et al. (2017)	K1(TKE)			M2	B1	S1		RRTM
IFS cy36r1	www.ecmwf.int	K1		M1	M1		P	SRTM	RRTM
IFS cy36r1 DualM	Neggers et al. (2004)	K1		M1	M2		P	SRTM	RRTM
IFS cy36r4	www.ecmwf.int	K1		M1	M1		P	SRTM	RRTM
JMA-GSM v1	Kawai and Inoue (2006) Kawai (2012b)	K1		Sp		B1	S1	d-E	bE
JMA-GSM v2	Kawai (2012a) Kawai (2012b)	K1		Sp		B1	S1	d-E	bE
LaRC	Cheng and Xu (2006)	HOC		HOC	HOC	B1	S2	RRTMG	RRTMG
LMDZ-AR4	Hourdin et al. (2013)	K1		M1		B1	B1		
LMDZ-THPL	Hourdin et al. (2006)	K2.5(TKE)		M1	M1	B1	S2		
Met Office GA3.0	Walters et al. (2011)	K1(P)	Ri,LW,SW,Ev	M1	M1	B1	P	SOCR	SOCR
Met Office HadGEM2	Martin et al. (2011)	K1(P)	Ri,LW,SW,Ev	M1	M1	B1	D	SOCR	SOCR
NCEP GFS	Fletcher et al. (2014)	K1(P)	Ri,LW	M1	M1	B1	P	NASA	RRTM
RACMO	Lenderink and Holtslag (2004)	K1(TKE)	Ri	M1	M2	B1	P		RRTM
UCLA-AGCM	Konor et al. (2009) Xiao et al. (2012)	Bulk ML	Ri	Sp		B1		RR	RR
WRF-TEMF	Skamarock et al. (2008) Angevine et al. (2010)	K1(TTE)			M1	B1(Eta)	AoN	RRTMG	RRTMG

Note. Turb: turbulence scheme, Entr: Entrainment closure at mixed layer top, Conv-D: deep convective transport, Conv-S: shallow convective transport (both dry and moist), Cld-M and Cld-μ indicate the cloud macrophysics and microphysics scheme, Rad: radiation scheme, both SW and LW. In all columns, a non-entry indicates that the scheme is inactive or absent. The other entries refer to the following. K*i*: K-diffusion, with *i* indicating the order of the closure (P: profile method, TKE: prognostic TKE). HOC: higher-order closure. Ri, LW, SW, Ev: dependence on Richardson number, LW flux, SW flux, evaporative cooling. M*i*: mass flux, with *i* indicating the number of transporting updrafts. Sp: spectral. B*i*: bulk microphysics, with *i* indicating nr of moments). S*i*: statistical cloud scheme, with *i* indicating the number of PDFs. P: prognostic. D: diagnostic. AoN: All-or-nothing condensation scheme. d-E: delta-Edington. bE: band emissivity. For all other acronyms, we refer to the specified references.

three. While this result in itself encourages the use of case studies, one still wonders what is the extent of this representativeness. For example, are SCM results actually representative of GCM behavior? Can persistent biases concerning low clouds as diagnosed in GCM simulations be attributed to subgrid scale parameterizations? These are key questions for improving our understanding of uncertainty in climate predictions and the behavior of cloud-climate feedbacks.

6. Summary and Outlook

This study confronts an ensemble of SCM simulations of the low-level cloud transition in the marine Trade-wind subsidence regime with LES results. The SCM ensemble represents the subgrid schemes of many state-of-the-art weather and climate models, and includes both operational and research versions. Use is made of multiple cases, each slightly different, in order to broaden the parameter space. Combinations of simple metrics are applied to objectively assess the model performance, both of individual codes and of the ensemble as a whole. The main conclusions about the model performance can be summarized as follows:

1. Some longstanding biases are identified as established in previous intercomparison studies, including (i) a lack of BL deepening, (ii) a lack of decoupling, and (iii) the emergence of the too few too bright bias during the advanced stage of the transition to fair-weather cumulus.
2. On the other hand, some individual models show encouraging overall skill in reproducing the full set of variables.
3. When judged over the complete set of variables the median of the model ensemble robustly outperforms any individual model.
4. Model results obtained for single cases of low-level cloud transitions are found to be representative of the other cases, to a reasonable degree.
5. The comparison of model results to theoretical and observational relations facilitates the interpretation of model biases.

The participation of many subgrid schemes from present-day GCMs, combined with the use of state-of-the-art LES results, gives some hope that the outcome of this intercomparison study can guide the model development process in the years to come. The results obtained also raise some new questions, inspiring new research efforts. Some of these initiatives are already ongoing. For example, the merits of ensemble parameterization within one gridbox is currently explored in an ongoing follow-up study, making use of the multiple transition cases that form the basis of the three composite cases. Various recent studies attempt to link future climate uncertainty due to feedbacks involving low-level clouds to parameterizations. For example, Dal Gesso et al. (2015) compare SCM simulations of stratocumulus to results with their host GCM, making use of a phase-space approach. Alternatively Neggers (2015a) make use of the fingerprints of models in κ-TCC space to compare SCM and GCM simulations at selected gridpoints in the Eastern Pacific, reporting that GCM behavior can to some extent be attributed to subgrid physics as apparent in SCM simulations.

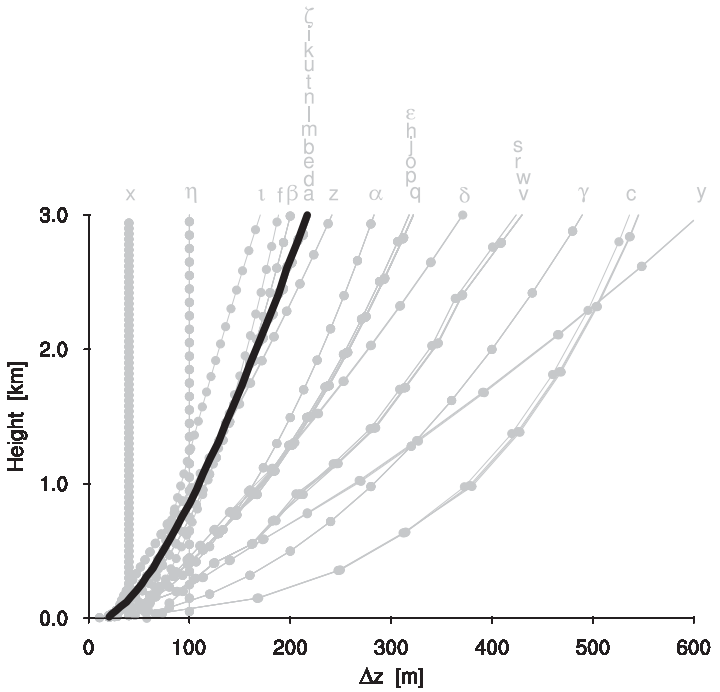


Figure B1. Layer depth versus midlevel height of the various vertical SCM grids as used in this study. The L80 grid is plotted in black, while the grids of the individual SCMs are plotted in grey. The symbols refer to Table 1.

Appendix A: Summary of Model Physics

Table A1 gives detailed information about the participating SCMs, including the associated publications as well as an overview of the type of various parameterizations.

Appendix B: Vertical Grids

Figure B1 shows the thickness of the model levels in the SCMs as a function of height in the lowest 3 km.

Acknowledgments
The research presented in this paper has received funding from the European Union, Seventh Framework Programme (FP7/2007–2013) under grant agreement 244067. P.N. Blossey acknowledges support from NOAA MAPP grant GC10–670a as a part of the NOAA Sc-Cu Transition Climate Process Team. Computational support was provided to A. Ackerman by the U.S. DOE National Energy Research Scientific Computing Center and the NASA Advanced Supercomputing Division for the DHARMA simulations. V. Larson and N. Meyer acknowledge financial support under grant AGS-0968640 from the U.S. National Science Foundation. S. Dal Gesso has received funding from the "Energy Transitions and Climate Change" project supported by the Excellence Initiative of the University of Cologne. The LES and SCM data at the basis of the results presented in this study can be accessed through the intercomparison data repository at http://gop.meteo.uni-koeln.de/~neggers/transitions/.

References

Abel, S. J., & Shipway, B. J. (2007). A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. *Quarterly Journal of the Royal Meteorological Society*, *133*(624), 781–794.

Albrecht, B. A., Bretherton, C. S., Johnson, D., Schubert, W. H., & Frisch, A. S. (1995). The Atlantic Stratocumulus transition experiment—ASTEX. *Bulletin of the American Meteorological Society*, *76*, 889–904. https://doi.org/10.1175/1520-0477(1995)076<0889:TASTE>2.0.CO;2

Angevine, W. M., Jiang, H., & Mauritsen, T. (2010). Performance of an eddy diffusivity-mass flux scheme for shallow cumulus boundary layers. *Monthly Weather Review*, *138*, 2895–2912. https://doi.org/10.1175/2010MWR3142.2891

Bazile, E., Marquet, P., Bouteloup, Y., & Bouyssel, F. (2011). *The Turbulent Kinetic Energy (TKE) scheme in the NWP models at Météo-France*. Paper presented at Workshop on Diurnal Cycles and the Stable Boundary Layer, 7–10 November 2011. Reading, UK: ECMWF.

Bengtsson, L., Andrae, U., Aspelien, T., Batrak, Y., Calvo, J., de Rooy, W. . . . Køltzow, M. Ø. (2017). The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Monthly Weather Review*, *145*, 1919–1935. https://doi.org/10.1175/MWR-D-16-0417.1

Bretherton, C. S., Austin, P., & Siems, S. T. (1995). Cloudiness and marine boundary layer dynamics in the ASTEX Lagrangian experiments. Part II: Cloudiness, drizzle, surface fluxes, and entrainment. *Journal of the Atmospheric Sciences*, *52*(16), 2724–2735. https://doi.org/10.1175/1520-0469(1995)052<2724:CAMBLD>2.0.CO;2

Bretherton, C. S., Krueger, S. K., Wyant, M. C., Bechtold, P., Van Meijgaard, E., Stevens, B., & Teixeira, J. (1999). A GCSS boundary-layer cloud model intercomparison study of the first ASTEX Lagrangian experiment. *Boundary-Layer Meteorology*, *93*(3), 341–380. https://doi.org/10.1023/A:1002005429969

Bretherton, C. S., McCaa, J., & Grenier, H. (2004). A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. Part I: Description and 1D results. *Monthly Weather Review*, *132*, 864–882.

Bretherton, C. S., & Pincus, R. (1995). Cloudiness and marine boundary layer dynamics in the ASTEX Lagrangian experiments. Part I: Synoptic setting and vertical structure. *Journal of the Atmospheric Sciences*, *52*(16), 2707–2723. https://doi.org/10.1175/1520-0469(1995)052<2707:CAMBLD>2.0.CO;2

Bretherton, C. S., & Wyant, M. C. (1997). Moisture transport, lower-tropospheric stability, and decoupling of cloud-topped boundary layers. *Journal of the Atmospheric Sciences*, *54*, 148–167. https://doi.org/10.1175/1520-0469(1997)054<0148:MTLTSA>2.0.CO;2

Brient, F., Schneider, T., Tan, Z., Bony, S., Qu, X., & Hall, A. (2015). Shallowness of tropical low clouds as a predictor of climate models' response to warming. *Climate Dynamics*, *47*, 433–449. https://doi.org/10.1007/s00382-015-2846-0

Browning, K. A. (1993). The GEWEX Cloud System Study (GCSS). *Bulletin of the American Meteorological Society*, *74*, 387–399.

Cheng, A., & Xu, K.-M. (2006). Simulation of shallow cumuli and their transition to deep convective clouds by cloud-resolving models with different third-order turbulence closures. *Quarterly Journal of the Royal Meteorological Society*, *132*, 359–382.

Cuxart, J., Bougeault, P., & Redelsperger, J.-L. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, *126*(562), 1–30.

Dal Gesso, S., van der Dussen, J. J., Siebesma, A. P., de Roode, S. R., Boutle, I. A., Kamae, Y. . . .Vial, J. (2015). A single-column model intercomparison on the stratocumulus representation in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *7*, 617–647. https://doi.org/10.1002/2014MS000377

de Roode, S. R., & Duynkerke, P. G. (1997). Observed Lagrangian transition of stratocumulus into cumulus during ASTEX: Mean state and turbulence structure. *Journal of the Atmospheric Sciences*, *54*(17), 2157–2173. https://doi.org/10.1175/1520-0469(1997)054<2157:OLTOSI>2.0.CO;2

de Roode, S. R., Sandu, I., van der Dussen, J. J., Ackerman, A. S., Blossey, P., Jarecka, D. . . .Stevens, B. (2016). Large-eddy simulations of EUCLIPSE-GASS Lagrangian stratocumulus-to-cumulus transitions: Mean state, turbulence, and decoupling. *Journal of the Atmospheric Sciences*, *73*(6), 2485–2508. https://doi.org/10.1175/JAS-D-15-0215.1

De Rooy, W. C., & Siebesma, A. P. (2008). A simple parameterization for detrainment in shallow cumulus. *Monthly Weather Review*, *136*, 560–576.

Duynkerke, P. G., de Roode, S. R., van Zanten, M. C., Calvo, J., Cuxart, J., Cheinet, S. . . . Sednev, I. (2004). Observations and numerical simulations of the diurnal cycle of the eurocs stratocumulus case. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3269–3296. https://doi.org/10.1256/qj.03.139

Epstein, E. S. (1969). Stochastic dynamic prediction1. *Tellus*, *21*(6), 739–759. https://doi.org/10.1111/j. 2153–3490.1969.tb00483.x

Fletcher, J., Bretherton, C., Xiao, H., Sun, R., & Han, J. (2014). Improving subtropical boundary layer cloudiness in the 2011 NCEP GFS. *Geoscientific Model Development*, *7*(5), 2107–2120.

Hazeleger, W., Wang, X., Severijns, C., Stefănescu, S., Bintanja, R., Sterl, A. . . . van der Wiel, K. (2012). EC-Earth V2.2: Description and validation of a new seamless earth system prediction model. *Climate Dynamics*, *39*(11), 2611–2629. https://doi.org/10.1007/s00382-011-1228-5

Heus, T., Heerwaarden, C. C., Jonker, H. J. J., Siebesma, A. P., Axelsen, S., van den Dries, K. . . .Vilà-Guerau de Arellano, J. (2010). Formulation of the Dutch Atmospheric Large-Eddy Simulation (DALES) and overview of its applications. *Geoscientific Model Development*, *3*, 415–444. https://doi.org/10.5194/gmd-3-415-2010

Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F. . . . Roehrig, R. (2013). LMDZ5B: The atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, *40*, 2193–2222.

Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J.-L. . . . Lott, F. (2006). The LMDZ4 general circulation model: Climate performance and sensitivity to parameterized physics with emphasis on tropical convection. *Climate Dynamics*, *27*, 787–813.

Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A., Helber, R., & Arnone, R. A. (2009). Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems*, *76*(1), 64–82. https://doi.org/10.1016/j.jmarsys.2008.05.014

Kawai, H. (2012a). Examples of mechanisms for negative cloud feedback of stratocumulus and stratus in cloud parameterizations. *SOLA*, *8*, 150–154.

Kawai, H. (2012b). Results of ASTEX and composite model intercomparison cases using two versions of JMA-GSM SCM. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling/WMO*, *42*, 0411–0412.

Kawai, H., & Inoue, T. (2006). A simple parameterization scheme for subtropical marine stratocumulus. *SOLA*, *2*, 17–20.

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*, 607–625.

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*, 1587–1606.

Konor, C. S., Boezio, G. C., Mechoso, C. R., & Arakawa, A. (2009). Parameterization of PBL processes in an atmospheric general circulation model: Description and preliminary assessment. *Monthly Weather Review*, *137*, 1061–1082.

Kuo, H., & Schubert, W. T. (1988). Stability of cloud-topped boundary layers. *Quarterly Journal of the Royal Meteorological Society*, *114*, 887–916.

Larson, V. E., Schanen, D. P., Wang, M., Ovchinnikov, M., & Ghan, S. (2012). PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km. *Monthly Weather Review*, *140*, 285–306. https://doi.org/10.1175/MWR-D-10-05059.1

Leith, C. E. (1974). Theoretical skill of Monte Carlo forecasts. *Monthly Weather Review*, *102*(6), 409–418. https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2

Lenderink, G., & Holtslag, A. A. (2004). An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. *Quarterly Journal of the Royal Meteorological Society*, *130*, 3405–3427.

Lock, A. P. (2009). Factors influencing cloud area at the capping inversion for shallow cumulus clouds. *Quarterly Journal of the Royal Meteorological Society*, *135*(641), 941–952. https://doi.org/10.1002/qj.424

Martin, T. H. D. T. G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C. . . . Wiltshire, A. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, *4*(3), 723–757. https://doi.org/10.5194/gmd-4–723-2011

Morrison, H., Curry, J. A., & Khvorostyanov, V. I. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *Journal of the Atmospheric Sciences*, *62*, 1665–1677.

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The "too few, too bright" tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*, L21801. https://doi.org/10.1029/2012GL053421

Neggers, R. A. J. (2009). A dual mass flux framework for boundary-layer convection. Part II: Clouds. *Journal of the Atmospheric Sciences*, *66*, 1489–1506. https://doi.org/10.1175/2008JAS2636.1

Neggers, R. A. J. (2015a). Attributing the behavior of low-level clouds in large-scale models to subgrid-scale parameterizations. *Journal of Advances in Modeling Earth Systems*, *7*. *7*, 2029–2043. https://doi.org/10.1002/2015MS000503

Neggers, R. A. J. (2015b). Exploring bin-macrophysics models for moist convective transport and clouds. *Journal of Advances in Modeling Earth Systems*, *7*, 2079–2104. https://doi.org/10.1002/2015MS000502

Neggers, R. A. J., Köhler, M., & Beljaars, A. A. M. (2009). A dual mass flux framework for boundary-layer convection. Part I: Transport. *Journal of the Atmospheric Sciences*, *66*, 1465–1487. https://doi.org/10.1175/2008JAS2635.1

Neggers, R. A. J., Siebesma, A., Lenderink, G., & Holtslag, A. (2004). An evaluation of mass flux closures for diurnal cycles of shallow cumulus. *Monthly Weather Review*, *132*, 2525–2538.

Norris, J. R. (1998). Low cloud type over the ocean from surface observations. Part II: Geographical and seasonal variations. *Journal of Climate*, *11*, 383–403. https://doi.org/10.1175/1520-0442(1998)011<0383:LCTOTO>2.0.CO;2

Nuijens, L., Medeiros, B., Sandu, I., & Ahlgrimm, M. (2015). The behavior of trade-wind cloudiness in observations and models: The major cloud components and their variability. *Journal of Advances in Modeling Earth Systems*, *7*, 600–616. https://doi.org/10.1002/2014MS000390

Park, S., Leovy, C. B., & Rozendaal, M. A. (2004). A new heuristic Lagrangian marine boundary layer cloud model. *Journal of the Atmospheric Sciences*, *61*, 3002–3024. https://doi.org/10.1175/JAS-3344.1

Pergaud, J., Masson, V., Malardel, S., & Couvreux, F. (2009). A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-Layer Meteorology*, *132*, 83. https://doi.org/10.1007/s10546-009-9388-0

Randall, D. A., Krueger, S., Bretherton, C., Curry, J., Duynkerke, P., Moncrieff, M. . . . Wielicki, B. (2003a). Confronting models with data: The GEWEX Cloud Systems Study. *Bulletin of the American Meteorological Society*, *84*, 455–469.

Riehl, H., Yeh, C., Malkus, J. S., & LaSeur, N. E. (1951). The northeast Trade of the Pacific Ocean. *Quarterly Journal of the Royal Meteorological Society*, *77*, 598–626.

Sandu, I., & Stevens, B. (2011). On the factors modulating the stratocumulus to cumulus transitions. *Journal of the Atmospheric Sciences*, *68*, 1865–1881. https://doi.org/10.1175/2011JAS3614.1

Sandu, I., Stevens, B., & Pincus, R. (2010). On the transitions in marine boundary layer cloudiness. *Atmospheric Chemistry and Physics*, *10*(5), 2377–2391. https://doi.org/10.5194/acp-10-2377-2010

Shutts, G. J., & Gray, M. E. B. (1994). A numerical modelling study of the geostrophic adjustment process following deep convection. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1145–1178.

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G. . . . Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*, 1201–1219.

Siebesma, A. P., Jakob, C., Lenderink, G., Neggers, R. A. J., Teixeira, J., Van Meijgaard, E. . . . Severijns, C. (2004). Cloud representation in general-circulation models over the northern Pacific Ocean: A EUROCS intercomparison study. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3245–3267. https://doi.org/10.1256/qj.03.146

Siebesma, A. P., Soares, P. M. M., & Teixeira, J. (2007). A combined eddy diffusivity mass-flux approach for the convective boundary layer. *Journal of the Atmospheric Sciences*, *64*, 1230–1248.

Skamarock, W. C., Klemp, J., Dudhia, B. J., Gill, D. O., Barker, D. M., Duda, M. G. . . . Powers, J. G. (2008). *A description of the advanced research WRF version 3* (NCAR Tech. Note TN-475+STR, p. 113). https://dx.doi.org/10.5065/D68S4MVH

Stephens, G. L. (1978). Radiation profiles in extended water clouds. II: Parameterization schemes. *Journal of the Atmospheric Sciences*, *35*, 2123–2132.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z. . . . Team, T. C. S. (2002). The cloudsat mission and the a-train. *Bulletin of the American Meteorological Society*, *83*(12), 1771–1790. https://doi.org/10.1175/BAMS-83-12-1771

Stevens, B., Ackerman, A. S., Albrecht, B. A., Brown, A. R., Chlond, A., Cuxart, J. . . . Stevens, D. E. (2001). Simulations of Trade-wind cumuli under a strong inversion. *Journal of the Atmospheric Sciences*, *58*, 1870–1891.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S. . . . Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*, 146–172. https://doi.org/10.1002/jame.20015

Stevens, B., & Lenschow, D. H. (2001). Observations, experiments, and large-eddy simulation. *Bulletin of the American Meteorological Society*, *82*, 283–294. https://doi.org/10.1175/1520-0477(2001)082<0283:OEALES>2.3.CO;2

Stevens, B., Lenschow, D. H., Vali, G., Gerber, H., Bandy, A., Blomquist, B. . . . van Zanten, M. C. (2003). Dynamics and chemistry of marine stratocumulus-DYCOMS-II. *Bulletin of the American Meteorological Society*, *84*, 579–593. https://doi.org/10.1175/BAMS-84-5-57 9

Stevens, B., Moeng, C.-H., & Sullivan, P. P. (1999). Large-eddy simulations of radiatively driven convection: Sensitivities to the representation of small scales. *Journal of the Atmospheric Sciences*, *56*, 3963–3984. https://doi.org/10.1175/1520-0469(1999)056<3963:LESORD>2.0.CO;2

Stevens, B., & Seifert, A. (2008). Understanding macrophysical outcomes of microphysical choices in simulations of shallow cumulus convection. *Journal of the Meteorological Society of Japan*, *86A*, 143–162.

Stevens, D. E., Ackerman, A. S., & Bretherton, C. S. (2002). Effects of domain size and numerical resolution on the simulation of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *59*, 3285–3301.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*, 7183–7192.

Teixeira, J., Cardoso, S., Bonazzola, M., Cole, J., DelGenio, A., DeMott, C. . . . Zhao, M. (2011). Tropical and sub-tropical cloud transitions in weather and climate prediction models: The GCSS/WGNE Pacific Cross-section Intercomparison (GPCI). *Journal of Climate*, *24*, 5223–5256. https://doi.org/10.1175/2011JCLI3672.1

van der Dussen, J. J., de Roode, S. R., Ackerman, A. S., Blossey, P. N., Bretherton, C. S., Kurowski, M. J. . . . Siebesma, A. P. (2013). The GASS/EUCLIPSE model intercomparison of the stratocumulus transition as observed during ASTEX: LES results. *Journal of Advances in Modeling Earth Systems*, *5*, 483–499. https://doi.org/10.1002/10.1002/jame.20033

van der Dussen, J. J., de Roode, S. R., & Siebesma, A. P. (2014). Factors controlling rapid stratocumulus cloud thinning. *Journal of the Atmospheric Sciences*, *71*(2), 655–664. https://doi.org/10. 1175/JAS-D-13–0114.1

van Meijgaard, E., & Crewell, S. (2005). Comparison of model predicted liquid water path with ground-based measurements during cliwa-net. *Atmospheric Research*, *75*(3), 201–226. https://doi.org/10.1016/j.atmosres.2004.12.006

Vial, J., Bony, S., Dufresne, J.-L., & Roehrig, R. (2016). Coupling between lower-tropospheric convective mixing and low-level clouds: Physical mechanisms and dependence on convection scheme. *Journal of Advances in Modeling Earth Systems*, *8*, 1892–1911. https://doi.org/10.1002/2016MS000740

Vogel, R., Nuijens, L., & Stevens, B. (2016). The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *Journal of Advances in Modeling Earth Systems*, *8*, 843–862. https://doi.org/10.1002/2015MS000568

Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D. . . . Williams, K. D. (2011). The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations. *Geoscientific Model Development*, *4*(4), 919–941. https://doi.org/10.5194/gmd-4-919-2011

Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dynamics*, *17*(12), 905–922. https://doi.org/10.1007/s003820100157

Wood, R., & Bretherton, C. S. (2004). Boundary layer depth, entrainment, and decoupling in the cloud-capped subtropical and tropical marine boundary layer. *Journal of Climate*, *17*, 3576–3588. https://doi.org/10.1175/1520-0442(2004)017<3576:BLDEAD>2.0.CO;2

Wood, R., Bretherton, C. S., & Hartmann, D. L. (2002). Diurnal cycle of liquid water path over the subtropical and tropical oceans. *Geophysical Research Letters*, *29*(23), 2092. https://doi.org/10.1029/2002GL015371

Xiao, H., Wu, C.-M., Mechoso, C. R., & Ma, H. Y. (2012). A treatment for the stratocumulus-to-cumulus transition in GCMs. *Climate Dynamics*, *39*, 3075–3089. https://doi.org/10.1007/s00382-012-1342-z

Zhang, M., Bretherton, C. S., Blossey, P. N., Austin, P. H., Bacmeister, J. T., Bony, S. . . . Zhao, M. (2013). CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *Journal of Advances in Modeling Earth Systems*, *5*, 826–842. https://doi.org/10.1002/2013MS000246

Zhu, P., Bretherton, C. S., Köhler, M., Cheng, A., Chlond, A., Geng, Q. . . . Stevens, B. (2005). Intercomparison and interpretation of single-column model simulations of a nocturnal stratocumulus-topped marine boundary layer. *Monthly Weather Review*, *133*(9), 2741–2758. https://doi.org/10.1175/MWR2997.1