Response to the comments on the manuscript: Parameterising cloud base updraft velocity of marine stratocumuli, Ref. ACP-2021-757.

Dear Editor, dear Reviewers, we would like to thank the Editorial Board for considering our paper for publication in ACP and the reviewers for their constructive comments. We have addressed all of them and modified the paper accordingly. Our detailed answers follow. Text from the original manuscript that has been removed in the revised manuscript is marked in red. New text in the revised manuscript is marked in blue.

Answers to Reviewer 1

Overview R1.1 This article presents an LES assessment of 3 types of parametrisations used to derive cloud base vertical velocity in large scale models.

The methods tested include one "traditional" linear parametrisation where updraft velocity is derived from cloud top radiative cooling and two parametrisations derived from different machine learning techniques (Gaussian process emulation and random forest). The authors demonstrate that when compared to LES simulations, which are viewed as truth, the machine learning techniques produce a more accurate representation of the cloud base vertical velocity than the "traditional" method.

The authors do a nice job of explaining machine learning methods and defining a workflow. In particular, the authors have structured the methods around the workflow design, which makes the method easy to read and reference. Using the workflow, the authors clearly show the application of the machine learning techniques can improve on more "traditional" parametrisation methods when all methods are compared to LES data. On face value this is a nice result, however, it is very difficult to understand if either parametrisation is performing well because the article fails to present enough information about simulations to understand the validity of the training data.
Answer to R1.1 We thank the referee for these comments and do our best to improve the manuscript according to the suggestions.

General comment R1.2 The authors fail to present any cloud evolutions over time, e.g. cloud base and height plots, LWP and surface/cloud base precipitation. The authors do present statistics, but it is important to present a mean and a range of the time evolution of the simulated marine Sc, so that a reader can be confident the machine learning is training on sensible pseudo-data.
Answer to R1.2 Cloud evolution is not shown in the manuscript, because we have 1190 simulations in total. However, based on the referee comment we have now added additional statistics about cloud development including changes in cloud base and top heights and LWP in the supplements to maintain the readability and focus of the manuscript (see the specific comment R1.11). We also show statistics about accumulated surface and cloud base precipitation.

We have further elaborated the explanation in the answer to question R1.11.

General comment R1.3 There are a lot of thresholds defined in the paper, which are used to include or ignore data, but many of these thresholds are not justified and it is not demonstrated how sensitive the techniques are to these thresholds.
Answer to R1.3 We used thresholds first to sample stratocumulus-like clouds from ECHAM for the LES simulations (line 88), and then different threshold we used to exclude a fraction of the LES simulations that clearly deviated from the linear updraft velocity vs. radiative cooling trend expected for marine stratocumulus (line 245).

We used ECHAM as source data keeping the goal in mind that we could improve the global model updraft parameterisations. However, ECHAM is somewhat problematic in representing stratocumulus clouds and hence, the LES cannot forcefully be made to maintain the cloud in the problematic initial state. With these thresholds, we aim to keep the applicability of the methods as large as possible by excluding the most problematic cases.

Concerning the threshold given in line 88 is further discussed in the answer to the question R1.8.

Thresholds for filtering LES simulations (line 245) are discussed in the answers to the question R1.15 and R2.2.

General comment R1.4 It is unclear why the authors use a fixed solar zenith angle for daytime and separate daytime and night-time, rather than run diurnal simulations. This choice, without further justification, casts doubt on the simulations, i.e. a diurnal simulation was not possible.

Hence, while I think there is good explanation of the methods, the minimal presentation of the simulation results to demonstrate the simulations produce a good representation of marine Sc mean that I recommend major revisions.

Answer to R1.4 Our parameterisations were specifically designed to be used in ECHAM simulations. For this reason, we sampled the LES inputs from ECHAM simulations so that the sample is truly representing ECHAM (sampling based on the BSP method). Then we made short LES simulations which are representative of the inputs (this is why solar zenith angle is fixed), yet long enough to develop turbulence and other processes. With these we could use ECHAM inputs and LES outputs in developing the parameterisations. If we would have taken both the inputs and outputs from the longer (and fewer) LES runs, which change with time, the sample would not be representative of ECHAM anymore. We have clarified our strategy in the revised manuscript.

We have further clarified our approach in the revised manuscript. Specifically, our machine learning methods take the inputs from the global model ECHAM instead of using those calculated from the LES simulations. This ensures that the parameter space is representative of ECHAM clouds, for which the updraft parameterisations are designed for, although the LES would produce more realistic stratocumulus cases. Similarly, diurnal simulations would provide more realistic clouds. However, that would limit the applicability of data within global modelling framework where also different microphysical process representations (e.g precipitation formation, lack of subgrid processes) might lead to different cloud properties than those simulated by the LES in the same conditions. This could be avoided only through emulating all cloud properties with the output from the same LES, which naturally should be the goal of future studies, but is out of the scope here.

Changes in the manuscript, first paragraph of Sect. 2 *Methods*:

The parameterisations are specifically designed for marine clouds seen in ECHAM. This means that each LES run is initialised using an ECHAM cloud state as an input and then the LES predicts the updraft velocity related to this cloud state. The process for creating a LES-based updraft velocity parameterisation for marine boundary layer clouds is shown as a pipeline in Fig. 1. Here the aim is to create a parameterisation that would represent detailed LES runs with low computational cost. The pipeline has three main stages. First, we need source data that describes boundary layer conditions in a large number of marine boundary layer cloud cases (Sect. 2.2.1). Second, a representative subset of this data is selected (Sect. 2.2.2) and the corresponding simulations are run with the LES model (Sect. 2.3). Third, specific LES model outputs are selected from these simulations and corresponding parameterisations are created (Sect. 2.4).

Changes in the manuscript, third paragraph of Sect. 2.3 *Setting up the LES runs*:

"The remaining design variables are related to cloud microphysics and solar radiation. Daytime simulations have a fixed solar zenith angle, i.e. it is not changing with time. We did not include diurnal cycle as that would disconnect the end state of the simulations from the input values, which would violate the underlying assumptions of the machine-learning-based approach presented in this study. For SB microphysics, the only cloud micro-physical parameter is cloud droplet number concentration. The SALSA simulations are initialised with the specified tri-modal aerosol size distribution in each grid cell (Sect. 2.2.1)."

Specific comment R1.5 Line 70 – I was surprised to read that there is an autoconversion scheme in the bin model. It is important to note that such a configuration for a bin model is non-standard, possibly unique(?) since bin models tend to represent the (semi)-continuous growth of activated droplets to rain through condensation and collision-coalescence processes. From the description in this article, it is not clear whether SALSA includes collision-coalescence for cloud bins or do the cloud bins just experience condensation? This should be clarified in the text.
Answer to R1.5 UCLALES-SALSA is indeed a unique bin model. Most bin models track cloud droplet wet size, but UCLALES-SALSA cloud (and aerosol) bins are based on the dry size. Tracking the dry size means that the wet size may become inaccurate when the droplets grow larger through collisions. For this reason, we have implemented rain bins that are based on wet size, which allow modelling their continuous growth. Similar to bulk schemes, when cloud droplets become larger, the autoconversion scheme moves the largest droplets to rain bins with a minimum size of 50 micrometers, where their size-dependent dynamics can be modelled accurately. Precipitation forms when the initially small rain drops grow mainly with the collision-coalescence process. Both rain and cloud droplets grow by condensation and collision-coalescence. In the recent model release there is also an option not to use autoconversion scheme. We have clarified this in the revised manuscript.

Changes in the manuscript first paragraph of Sect. 2.1 *LES*:

"The parameterisations developed in this study are based on simulations with UCLALES-SALSA (Tonttila et al., 2017; Ahola et al., 2020), which models atmospheric dynamics with a LES and includes a four-stream radiative transfer solver (Fu and Liou, 1993). We used two different cloud microphysical modules with the LES. First, the default UCLALES (Stevens et al., 1999, 2005) includes Seifert & Beheng microphysics (SB) with diagnostic clouds (saturation adjustment for cloud water, constant cloud droplet number concentration as an input) and a two-moment bulk scheme with parameterised autoconversion for rain (Seifert and Beheng, 2006). The second set-up employs SALSA, which is a sectional scheme where aerosol, cloud droplet and raindrop size distributions and chemical composition are described with several size bins (Kokkola et al., 2008; Tonttila et al., 2017). Cloud activation is calculated by solving equations for condensation of water on aerosol particles and then counting the number of droplets reaching the critical droplet size (prognostic supersaturation scheme). However, rain formation uses the same autoconversion parameterisation as used in the SB microphysics. This scheme is employed because UCLALES-SALSA cloud and aerosol bins are based on the dry size. Tracking the dry size means that the wet size may become inaccurate when the droplets grow larger. For this reason, we have implemented rain bins that are based on wet size. Both rain and cloud droplets grow by condensation and collision-coalescence. The autoconversion scheme moves the largest cloud droplets to rain bins with a minimum size of 50 μm, where their size-dependent dynamics can be modelled accurately. Due to the high number of prognostics variables, SALSA simulations are about ten times computationally heavier compared to the SB simulations."

Specific comment R1.6 Line 70 – The authors explicitly state that SB uses saturation adjustment but they do not state what SALSA uses. Does SALSA use a prognostic supersaturation

scheme? This needs to be clarified and stated for completeness.
Answer to R1.6 SALSA uses a prognostic supersaturation scheme. This has been clarified in the revised manuscript.

See the changes made in the manuscript in the answer to question R1.5.

Specific comment R1.7 Line 85 – The filtering applied to the ECHAM data isolates single layer clouds, which are below 3000 m. What do the authors mean by single layer cloud? Could they provide more information since it is not clear whether the authors are filtering on clouds that are one vertical grid in thickness or clouds that are multiple grids but continuous, i.e. not multi-layer clouds between the surface and 3000 m. Without more information, I have assumed that the filtering is on multi grid clouds and I am concerned that the filtering of this data will thus result in the inclusion of Marine Sc and Cu. If this is the case then the clouds could develop from different dynamical forcing, which could be a problem because the "traditional" parametrisation (Zheng 2016) is only applicable to marine Sc. Hence, could the authors confirm that Cu is filtered out of the data and if this can not be confirmed, could they discuss the impact of the inclusion of Cu in the data?
Answer to R1.7 We assume a single layer cloud for cases when there are one or more consecutive cloudy grid cells in the vertical. Cloud top has to be below or at 3000 m altitude. It is true that with this definition, clouds can contain different stages of development. The relaxed initial sampling is part of our sampling strategy, and it is based on the fact that the low resolution ECHAM (the AMIP setup) has difficulties in producing Sc. The LES runs based on the initial sample showed that some cases do not represent stratocumulus, so we applied filtering based on cloud fraction and changes in cloud top height. Especially the minimum cloud fraction limit 0.61 should remove most cumulus cases. Although the filtered sample contains some non-ideal cloud cases, the vast majority of the data points represent stable stratocumulus clouds.

Specific comment R1.8 Line 89 – How sensitive is the filtering to the cloud water threshold?
Answer to R1.8 This limit sets the minimum LWP (i.e., a single cloudy layer with this cloud water content) and also determines the separation between clear and cloudy conditions. For the latter purpose, this limit is fixed to a value commonly used at least in LES modelling, so it should not be adjusted. Also, the threshold should be low so that the training data for the parameterisation would include only single layer clouds. An additional LWP threshold could have been set, but it was not found necessary, because the minimum LWP for a steady Sc depends on other conditions like the inversion strength.

Changes in the manuscript second paragraph of Sect. 2.2.1 *Source data from ECHAM*:

"The source data was collected from standard model outputs of a one-year ECHAM-HAMMOZ AMIP (Atmospheric Model Intercomparison Project) type run (Fig. 1 point A). Filtered source data was sampled from open ocean columns that represent single-layer low clouds (Fig. 1 point 1.). In practise, first, continental or sea ice covered columns were excluded. Next, columns without a single-layer cloud above the sea surface and below 700 hPa (about 3000 m) level were screened out. Stratocumulus clouds rarely extend beyond 1500 m, but ECHAM produced such liquid clouds at higher altitudes. The current 3000 m cut-off limit is based on the computational requirements of the LES model, as with this limit we can still undertake high resolution simulations with reasonable computational cost. The threshold cloud water content for a cloudy grid cell was 0.01 g kg^{-1}, which is the limit used in UCLALES-SALSA. Liquid (LWP) and ice (IWP) water paths were calculated for the low cloud and for the whole column. The column was accepted if the column IWP was less than 10 % of the low cloud's LWP and the low cloud's LWP was more than half of the total cloud water path (LWP+IWP). These ~~conditions~~ 10% and 50% thresholds should ensure that the selected columns contain mostly liquid single-layer clouds below 700 hPa level ~~(about 3000 m)~~and eliminate radiative impacts from other clouds above the stratocumulus deck."

Specific comment R1.9 Line 90-91 – Is there a justification for the IWP and LWP percentage thresholds, beyond the maintenance of the single-layer cloud.
Answer to R1.9 These limits were decided in order to exclude cases with mixed-phase Sc and to avoid radiative impacts from other clouds above the Sc deck. The exact numbers were decided so that these filters would not be too restrictive.

Changes in the manuscript second paragraph of Sect. 2.2.1 *Source data from ECHAM*, see answer to the question R1.8.

Specific comment R1.10 Line 92 – Why has 3000 m been selected as the cut-off for the clouds?
Answer to R1.10 Stratocumulus clouds rarely extend beyond 1500 m, but ECHAM produced such liquid clouds at higher altitudes. The current 3000 m cut-off is based on the computational requirements of the LES. With this limit we can still do high resolution simulations with reasonable computational cost. A lower cut-off could have been used, but this would have limited the use of the developed updraft parameterisations in ECHAM.

Changes in the manuscript second paragraph of Sect. 2.2.1 *Source data from ECHAM*, see answer to the question R1.8.

Specific comment R1.11 Line 145 – the authors present examples of the initial profiles, which is good and they look very sensible for an idealised marine stratocumulus. However, the authors do not present any demonstration of the simulation beyond the statistics of wpos versus cloud top longwave cooling. This means that as a reader and a reviewer, I can not make any assessment of the validity of the training data, i.e. how does the daytime or nightime cloud evolve? Is the turbulence spun up? Is the simulated marine Sc coupled or decoupled from the surface, are they precipitating, etc? This is really important information, which the reader needs to understand the presented plots. Could the authors provide example plots of the evolution of cloud top and base, LWP and RWP and surface precipitation? Ideally, the plots should include some representation of the range of values.
Answer to R1.11 As explained above, showing cloud development in all 1190 simulations is not practical. However, we can show statistics about cloud development based on the difference between the initial and final states. We provide these in the supplements to maintain the readability of the manuscript. The evolution of individual clouds is publicly available in the raw data. Ideally, cloud top (Fig. R1) and base (Fig. R2) height and LWP (Fig. R3) would not be changing from the initial state but RWP and precipitation occur only after spin-up (Fig. R4). Cloud top and base hardly show any changes, but the LWC inside the stratocumulus decreases due to entrainment mixing at the cloud top, leading to sub-adiabatic LWC profile and lower LWP than the initial value. As ideal profiles are assumed based on ECHAM LWC output, we have still selected to use the initial simulation state as an input for the parameterisations.

Figure R5 shows the histogram of the tendency of updraft velocity during the last simulation hour. The mean is close to zero, which means that updraft velocities are not changing much during the last hour, so the turbulence is fully developed.

In Figs. R6 and R7 we have calculated decoupling analysis in the same way as in Jones et al. (2011). The lower left corner of Figs. R6a,b,c,d (where most of our simulations lie) present cases that are not decoupled. Decoupling would show as a large difference between boundary layer bottom and top liquid water potential temperatures (Fig. R7). We used temperature nudging to avoid such situations.

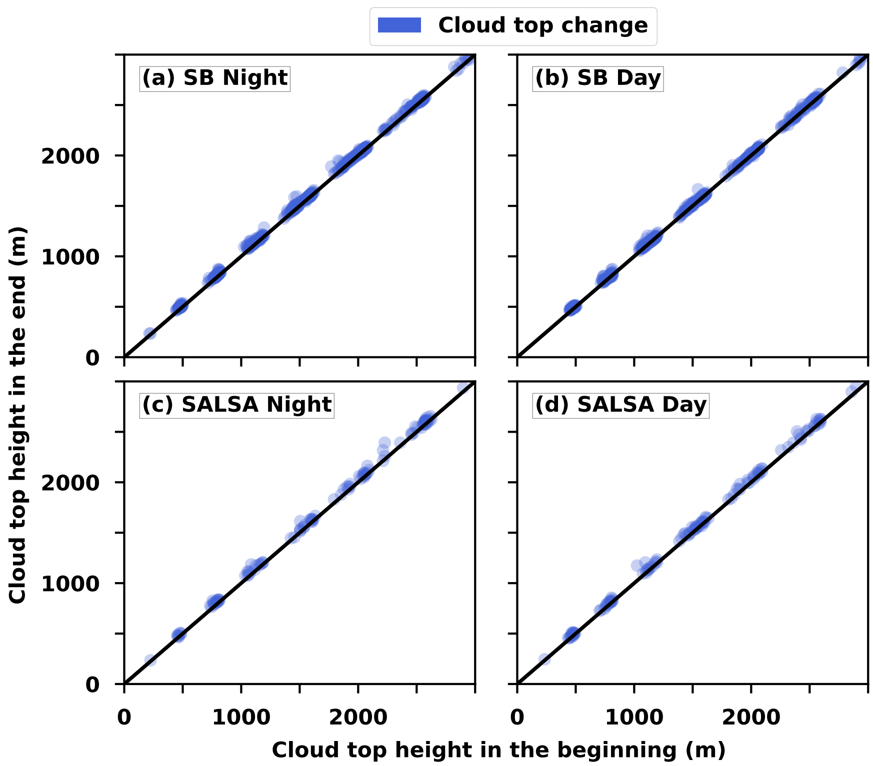

Figure R1: Cloud top change

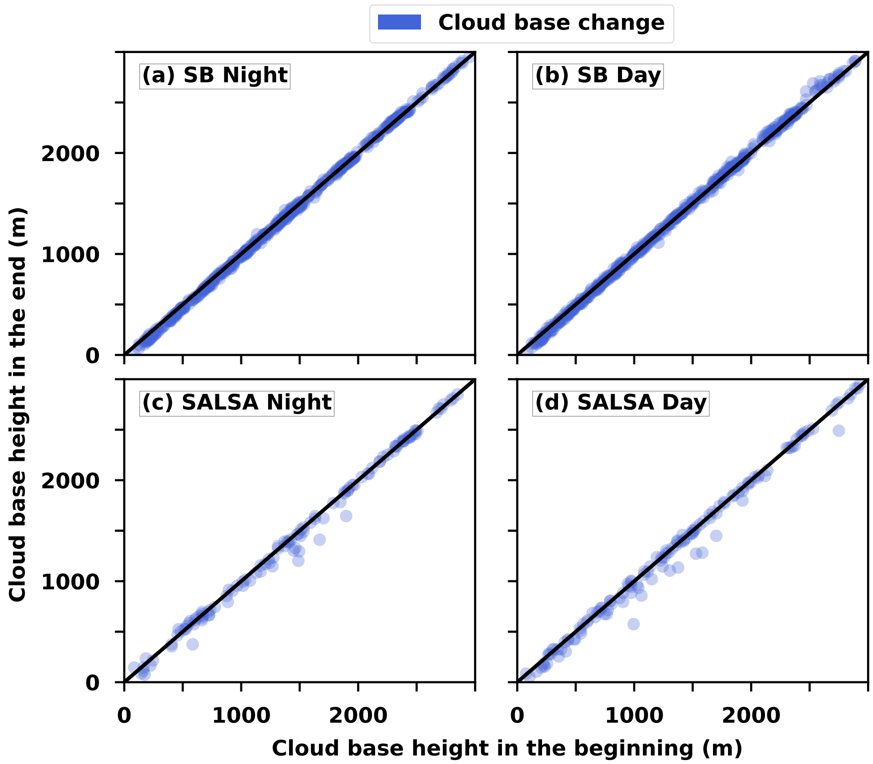

Figure R2: Cloud base change

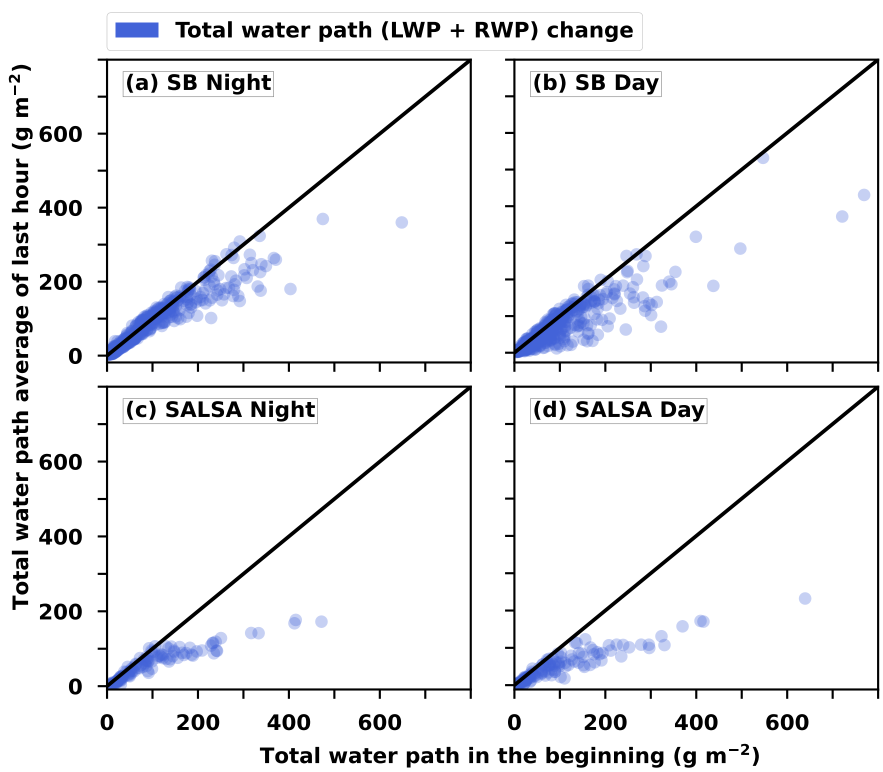

Figure R3: Total water path

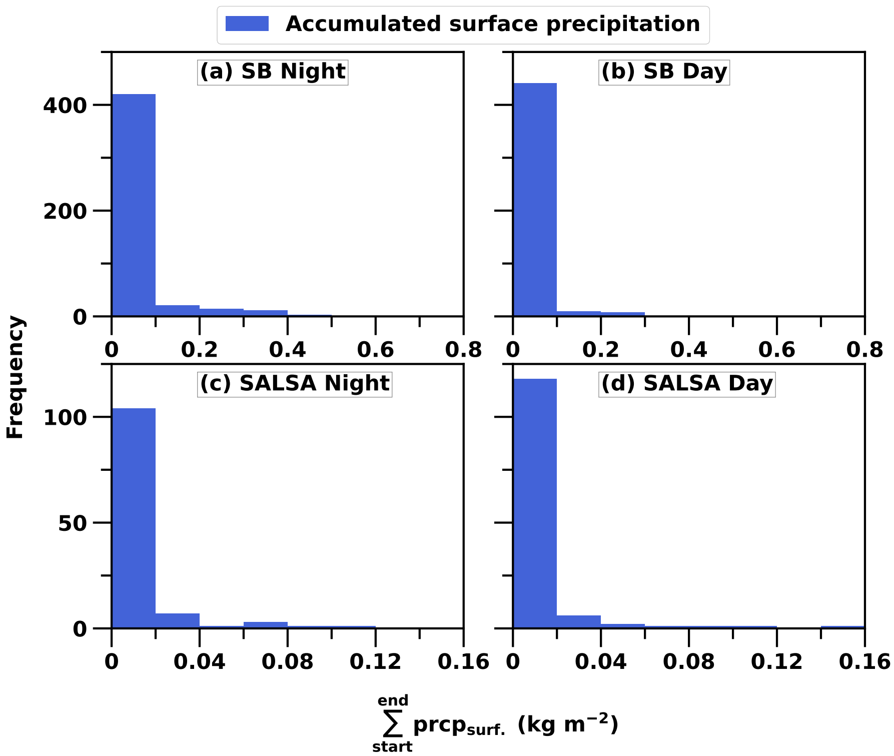

Figure R4: Accumulated surface precipitation.

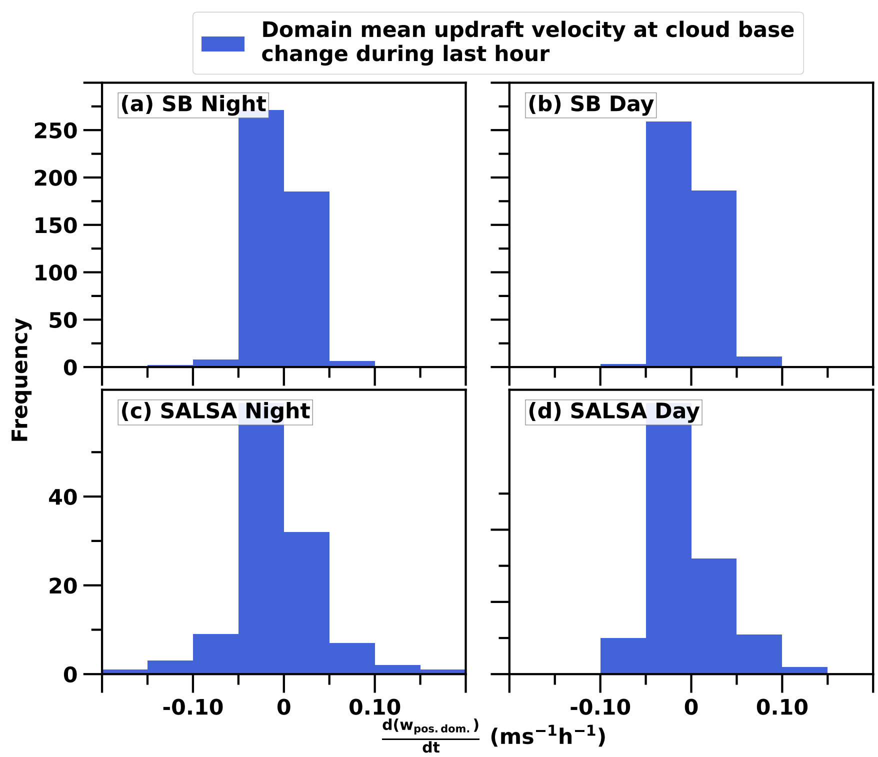

Figure R5: Tendency of updraft velocity during last hour

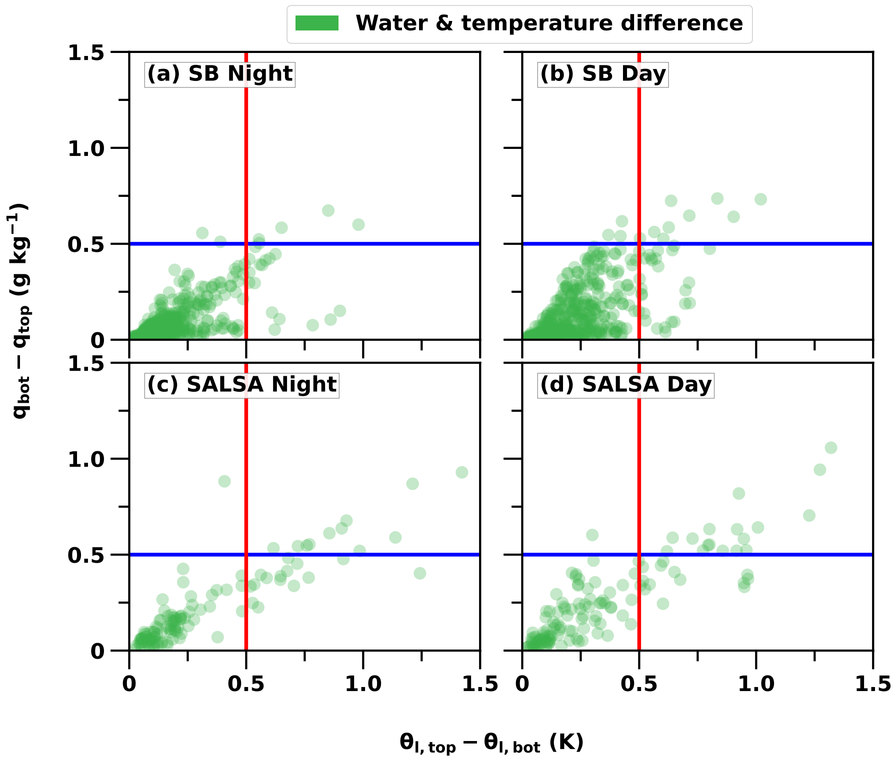

Figure R6: Scatter plot of decoupling measures

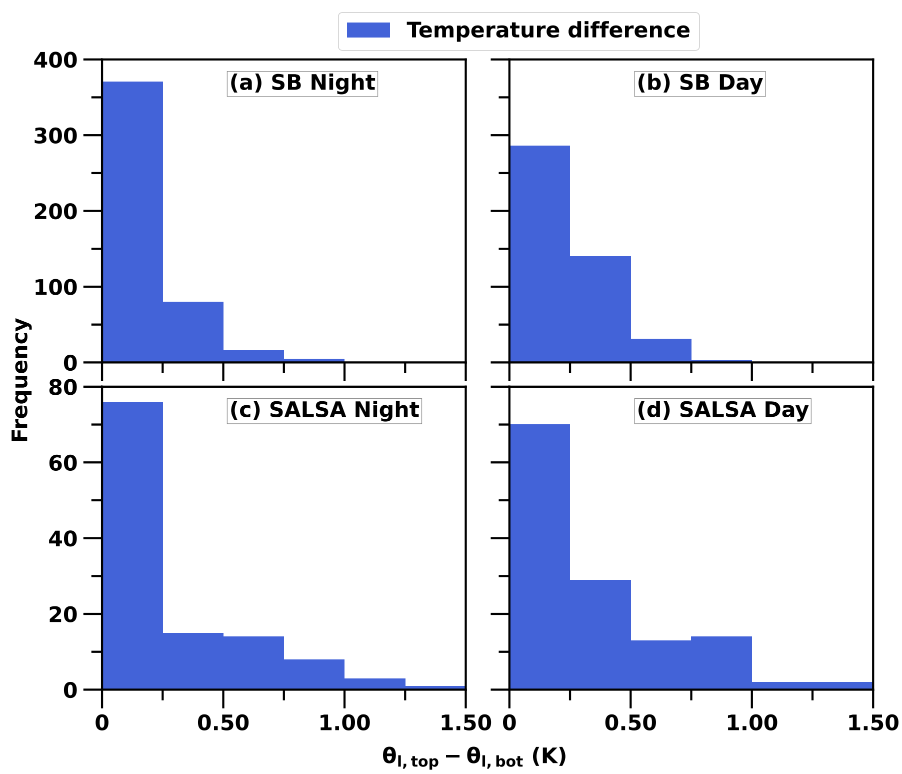

Figure R7: Difference between surface and cloud top potential temperatures

Specific comment R1.12 Line 150 – The article presents 2 sets of simulations – Daytime with fixed solar zenith and Nighttime. I am puzzled why the authors did not do a diurnal cycle and sample the data in day and night? Could the authors justify the choice of fixed solar zenith angle over a diurnal cycle? Also, are the results sensitive to the choice of solar zenith angle?
Answer to R1.12 We have elaborated our choice of fixed solar zenith angle in the answer to question R1.4.

Changes in the manuscript, third paragraph of Sect. 2.3 *Setting up the LES runs*, see the answer to the question R1.4.

Specific comment R1.13 Line 175 – the simple, non-machine learning parametrisation is based on Zheng et al (2016), why was this the only non-machine learning parametrisation tested?
Answer to R1.13 The Zheng et al (2016) parametrisation has only one input and it is likely the simplest parametrisation available, and still seems to be performing well. Our focus is in introducing machine learning methods instead of comparing different updraft parameterisations. This is now explained in the revised manuscript.

Changes in the manuscript second paragraph of Sect. 2.4 *Updraft parameterisations*:

"We used only one non-machine learning method (LF, Zheng et al. (2016)) as it has only one input variable and hence is possibly the simplest parameterisation available. Yet, as we will show below, it still performs quite well. Here, we focus on introducing machine learning based parameterisations for which the LF parameterisation serves as a baseline."

Specific comment R1.14 Line 230 – 235 – Are the authors filtering for columns that include evaporating rain? For example, the cloud base is defined as the lowest grid where LWC > 0.01g kg^{-1}. If rain is evaporating then this threshold could be very low, since the evaporation will return drops to cloud sized bins in SALSA (this will not be an issue in SB), hence there is the risk the updraft velocity at that grid does not represent cloud base where activation occurs. Could the authors clarify that evaporating rains does not bias the cloud base vertical velocity?
Answer to R1.14 Precipitating clouds were not filtered for the reasons explained in the manuscript (second paragraph of Sect. 2.4.1). In the definition of cloud base, the precipitating droplet are not accounted as liquid water and therefore precipitation has no direct impact on the determined cloud base.

Specific comment R1.15 Line 245 – Data was thrown away if the cloud fraction was smaller than 0.61 or cloud top rose more than 10%, could the authors justify these thresholds and add some discussion about the sensitivity of the parametrisation comparison to these thresholds? (I note the discussion of Feingold et al but that does not justify the 0.61.

Answer to R1.15 The initial ECHAM sampling strategy and the filters applied after that do have an impact on the parametrisation comparison, but with filtering we made parametrisation comparison fairer for the different methods. Although the Linear Fit improved with a Random Forest (LFRF) and the Gaussian Process Emulator (GPE) can predict updraft velocities for a wide variety of conditions and cloud cases, the Linear Fit (LF) parametrisation is originally for non-precipitating stratocumulus (Zheng et al., 2016). The initial ECHAM data contains also less ideal cloud cases, and some of these resulted in clear outliers for the linear fit. The above-mentioned thresholds were specifically designed to remove these outliers (Fig. 4 in the manuscript), which seem to represent other cloud types than stratocumulus.

Figure R8 shows all data points, and in SB simulations there is a group of points near -80 W m^{-2} that clearly deviate from the linear trend. For SALSA, there are also points near 0 W m^{-2} which are essentially cloud free. Initial visual analysis showed that the outliers are related to rising cloud tops and fractional cloudiness. Based on this visual analysis, we chose the limits from our data so that only the worst outliers (5.6–13.3% of the data) are removed. With this filter the data set is still representative of the original data set (and ECHAM) while excluding cases that deviate from the linear updraft velocity vs radiative cooling trend expected for stratocumulus.

Changing these thresholds would have an impact for the accuracy of the Linear Fit parametrisation simply by changing the number of outliers, but not so for the two machine learning parameterisations that can represent complex dependencies. In other words, the LF is sensitive on the thresholds while the two machine learning parameterisations are not.

Changes in the manuscript, last paragraph of Sect. 2.5 *Post-processing the LES runs*:

"Training data for the parameterisations includes the design variables and the corresponding LES updraft velocity (Eq. 2) and CTRC (Eq. 3) outputs. Some simulations that diverged significantly from the initial conditions produced outliers, which reduced the accuracy of the parameterisations in representing the rest of the cases. Therefore, in all following results, before creating a parameterisation, we filtered out simulations where cloud fraction was smaller than 0.61 or cloud top rose more than 10 % (see Table 2). ~~These cases were spotted in~~ In an initial analysis conducted when calculating the LF, these cases were found to produce most of the outliers. All filtering parameter values were the last retrievable values from the simulations. The cases where the cloud top rose more than 10 % were mostly related to weak temperature inversions at the cloud top (Sect. 2.2.1). For example, our temperature inversions start from 0.78 K (Fig. 2b) while Feingold et al. (2016) excluded values lower than 6 K. Weak temperature inversions are less effective in reducing entrainment mixing, which causes the deepening of the boundary layer (Wood, 2012). Nudging the model fields towards the initial conditions was used to suppress the issue, but it was not entirely eliminated. Using more restrictive initial conditions would have produced a more idealised stratocumulus sample, but at the same time we would have lost the ability to predict updraft velocities for less ideal cases commonly present in ECHAM simulation. In short, the limits for the filter (Table 2) were chosen so that they eliminate clear outliers from the linear fit. Changing these thresholds would have an impact for the accuracy of the Linear Fit parametrisation simply by changing the number of outliers, but not so for the two machine learning parameterisations that can represent complex dependencies. In other words, the LF is sensitive on the thresholds while the two machine learning methods are less sensitive. "

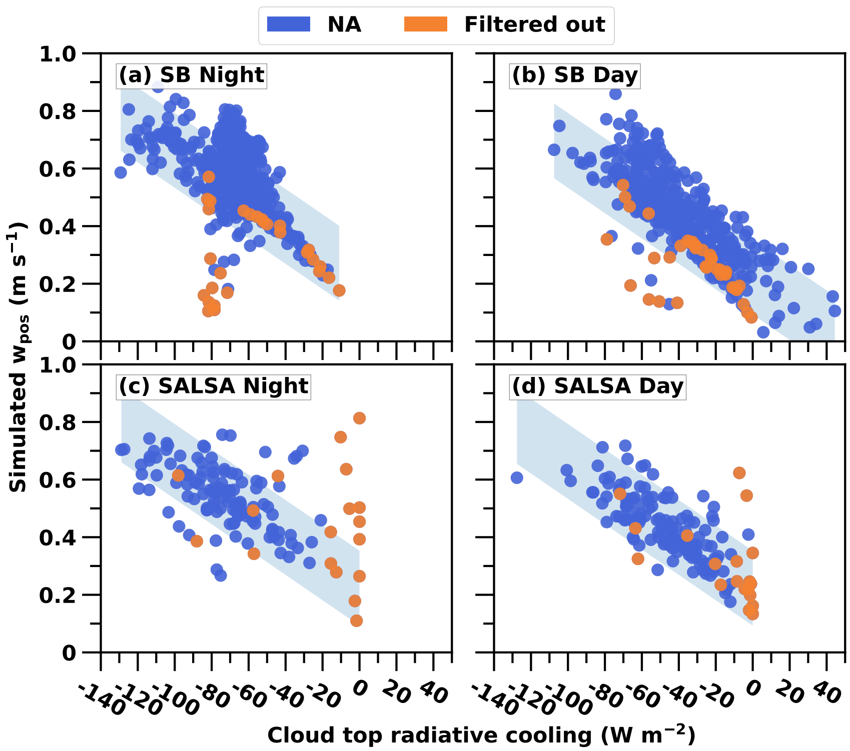

Figure R8: Calculating the Linear Fit (LF) without any filters (NA = No simulations excluded) or with LES filter used in the manuscript (Filtered out = orange points excluded).

Specific comment R1.16 Line 281 – Updraft velocity parameterisations from SB are shown to be better than those from SALSA. I agree that this is mainly due to the difference in sample set number but, could the difference also be due to the use of a prognostic supersaturation in SALSA vs "all-or-nothing" scheme in SB.
Answer to R1.16 Using a prognostic supersaturation scheme means that there are more degrees of freedom. Because we cannot increase the number of simulations when increasing the number of variables (we have even fewer simulations due to the significantly increased computational costs), prediction errors are likely to increase. Arguably, it is easier to develop parameterisations for the simple "all-or-nothing" scheme than for the more detailed SALSA scheme where for example the evaporation of cloud droplets and related latent heat change at cloud edges is slower than when the saturation adjustment is employed. We have adjusted the manuscript according to the referee's notion.

Changes in the manuscript last paragraph of Sect. 3.1 *Parameterisation intercomparison*:

"Parameterisations perform slightly better for simulation sets with SB than the sets with SALSA microphysics. One reason is that accounting for aerosol-cloud interactions increases the variability of model predictions ~~, which~~ as prognostic supersaturation scheme increases degrees of freedom. Hence, the variability can be difficult to capture in a parameterisation based on a relatively small set of training data. This additional variability can be seen as lower R^2 values in Fig. 4. The other main reason is that the number of computationally heavy SALSA simulations had to be limited to the lowest possible. Limited learning data set will have an impact on the accuracy of the predictions. ”

Answers to Reviewer 2

Overview R2.1 This is an interesting study that attempts to improve the parameterization of cloud-base updraft speed for marine stratocumulus using a large ensemble of large-eddy simulations (LES) driven by a GCM. Three methods are examined: (1) a linear physics-based scheme developed by a previous study, (2) a machine learning method that incorporates the (1), developed by the authors, and (3) a purely machine-learning-based method without incorporating physics, also developed by the authors. By comparing these three methods, it is found that (2) and (3) markedly outperform the (1), suggesting their potential usage for improving cloud-base updraft simulation in more coarse-resolution models. Some analyses about which predictor is more dominant were conducted.

Overall, the manuscript is clearly written and straightforward to understand. The conclusion that the new parameterizations improve the previous one is robust and convincing. Also, the data of the large ensemble of LES, which are released publically, should be very useful datasets. However, there are two major gaps that undermine the scientific quality, with one regarding the sampling issue and one regarding a lack of physical interpretation of the results. If these two issues can be well addressed, the manuscript should be a good fit for the ACP. See my detailed comments below.

Answer to R2.1 We thank the referee for these comments and do our best to improve the manuscript according to the suggestions.

Major comment R2.2 Marine stratocumulus only occurs under certain large-scale environments (e.g. see Wood (2012)'s review article). Your samples include many cases with $\Delta\theta_L$ of only several K degrees (Fig. 2b). Under this condition, an overcast stratocumulus deck can hardly sustain because of the strong cloud-top entrainment under a weak temperature inversion. Moreover, under weak inversions, the strong entrainment of warm air from above into the boundary layer, stably stratifying (or decoupling) the boundary layer so that your assumption of well-mixedness becomes invalid. Although you removed some of these cases (in Lines 244-251), the standard seems arbitrary to me and it cannot guarantee the remaining samples are physically reasonable since they are sampled based on a purely statistical method. I would suggest plotting a map showing where those samples are, so that the readers can have a sense of what kinds of stratocumulus we are looking at: eastern subtropical stratocumulus deck? or stratocumulus in the postfrontal region of midlatitude cyclones? or polar stratocumulus under an unperturbed environment? The point is to build a physical context for understanding those samples. Actually, Zheng et al. (2016)'s study is limited to only two regions, which does not necessarily guarantee a universal relationship.

Answer to R2.2 The initial sample contains a wide range of ECHAM-based cloud states, which means that the developed parameterisations can also be applicable to non-ideal stratocumulus cloud cases, which are common in our low-resolution ECHAM simulations. For example, low model vertical resolution is probably the main reason for the low $\Delta\theta_L$ values. Although these cloud states are physically reasonable, they may differ from the cloud states seen in the nature. Another reason for sampling a wide range of cloud states from ECHAM is that the LES is the best tool to show which ones of the initially sampled cases are representing reasonable stratocumulus clouds. This is because cloud stability and other cloud properties depend on the combination of the input parameters (e.g., humidity inversion), and not just the absolute value of temperature inversion. The final filtering criterion is based on identifying cases which deviate from the Zheng et al. (2016) parametrisation, which was developed for stratocumulus clouds. The chosen filtering limits exclude quite well unstable cases (rapidly decreasing cloud fraction) and those representing marine cumulus (rising cloud top). More strict filtering could have been applied (e.g., precipitating cases, which are not included in Zheng et al. (2016) study), but this seemed unnecessary from the parametrisation point of view; the linear fit performs well considering its simplicity.

More analysis on the LES filter, including a sensitivity analysis on the filter limits, is presented in the answer to question R1.15.

Unfortunately, geographic coordinates were not stored in the input data set for this study that was originally pulled from ECHAM, so we cannot draw a map from this data. Based on maps from our initial analysis, the known stratocumulus regions have the highest frequency of occurrences, but otherwise the clouds are fairly evenly distributed over ice-free marine regions. The specific region in ours or the Zheng et al. (2016) study are not important as updraft velocity depends on cloud state and not on location.

Major comment R2.3 I feel the discussion of the results is somewhat superficial. For example, entrainment is a crucial process governing the boundary layer dynamics. Stronger entrainment of warm air above can stabilize the boundary layer, reducing the turbulence. Such an important process, however, is not even mentioned in the discussion. Another important process is evaporative cooling that can also promote instability and enhance turbulence. Discussions should include these fundamental processes that contribute to the TKE budget of the boundary layer.
Answer to R2.3 Entrainment and evaporative cooling are known to be important processes (see, e.g., Wood, 2012; Zheng et al., 2016), and our simulations fully account for these and several other processes not explicitly mentioned in the manuscript. We now mention some of these processes in the discussion. Because we are not able to provide any new insight into these processes (they are not included in the current LES output statistics) and our focus is elsewhere, we keep this discussion brief.

Changes in the manuscript

Last paragraph of Sect. 2.5. *Post-processing the LES runs*, see answer to the question R1.15.

Second paragraph of Sect. 4 *Conclusions*:

"As can be expected, the simple LF works well for cases where radiative cooling is the main driver for turbulence. The other machine learning techniques perform better, because they account for additional variables such as cloud thickness and inversion strength, which have an additional influence on turbulence via processes like cloud top entrainment mixing and evaporative cooling, and drizzle formation (Wood, 2012). Overall, LFRF performs slightly better than GPE."

Minor comment R2.4 L335: "developed three xxx". I think the first method is developed by previous work. The wording should be revised.
Answer to R2.4 This is now corrected in the revised manuscript.

Changes in the manuscript first paragraph of Sect. 4 *Conclusions*:

"In this study we ~~developed~~ present three cloud base updraft velocity parameterisations which are based on detailed cloud simulations and can be used in global atmospheric models. The parameterisations represent the predictions of the large-eddy simulation model UCLALES-SALSA (Tonttila et al., 2017) for a wide range of marine boundary layer clouds described by the global climate model ECHAM. One parameterisation is a linear fit (LF) depending on cloud top radiative cooling only. The LF was first presented in Zheng et al. (2016) and is based on cloud observations. Another is based on the linear fit which is improved with a random forest model (LFRF). The random forest model was trained to predict the error of the linear fit as function of parameters describing marine boundary layer clouds. The third is a stand-alone Gaussian Process Emulator (GPE) for predicting updraft velocities based on the cloud parameters."