Authors' response to the comments of Referee #I on the manuscript

Evaluation of large-eddy simulations forced with mesoscale output for a multi-week period during a measurement campaign

by R. Heinze et al.

submitted to ACP

27th April, 2017

Dear Referee I,

Thank you very much for your comments on the manuscript. In what follows we present our replies (plain text) to your comments and questions (*italics*). The line numbers and the page numbers refer to the revised manuscript (without mark-up) which is attached to this response.

- **) Presumably, the reason to do these LES simulations in the first place is to retrieve better quality data than from the COSMO host model. The current results (e.g., fig 6) mainly suggest that the LESs are just close to COSMO. There also seems to be little added value in using 2 different LES codes, at least for the results that are presented here. Does that suggest that the host model strongly dictates the LES result? I understand that it is not the point of this paper to make any of the models (or indeed, observations) look bad, but it would be nice to highlight the benefit of your method by pointing out areas where LES outperforms COSMO, and other areas (such as the nocturnal BL, most likely), where COSMO keeps LES in check.*

  The primary reason to do these simulations is to receive turbulence-resolving modeling output localized at a special region with a fairly simple (semi-idealized) setup. It is highly desirable to get added value from the LES in variables that are determined by small-scale processes where turbulence plays a role, such as higher-order moments or microphysics. It is true that Fig. 6 (now Figs. 6-8) shows that the LES output is just close to COSMO which shows that the larger-scale forcing has a decent impact on the results. Anyhow, it should be kept in mind that COSMO is an operational NWP-model highly tuned compared to LES models running in a semi-idealized setup where only warm microphysics are used. A comment on the closeness of the LES to COSMO was made in the conclusion section (p. 22, l. 32). Another point where the LES are slightly better than COSMO is the specific rain water. COSMO shows much more rain than observed (p. 12, l. 21).

  Looking closer at Fig. 2, it can be seen that larger-scale forcing and nudging is as important as the fast LES physics during night-times where our 50m-LES is much too coarse to resolve the stable boundary

layer. Thus, here COSMO keeps the LES in check as the referee proposed. However, this should not affect the daytime development as previous work (van Stratum and Stevens, 2015) showed that the influence of biases in the representation of the nocturnal boundary layer do not substantially influence the subsequent daytime development. A small paragraph was added to section 2.4 (p. 8, l. 19-22).

Furthermore, on days with strong impact on the larger-scales, also the larger-scale tendencies dominate slaving the LES output to COSMO by construction of the forcing. A real benefit from LES is that it can provide variance (turbulence) profiles that are in good agreement with lidar observations (see Fig. 9). Unfortunately, the variances are not available from the COSMO-output so that a direct comparison to COSMO is not possible at this stage.

To conclude, the LES has the advantage in representing the structure of clouds and microphyscial changes in shallow convection and its coupling to the boundary layer. Still, in the way the LES is forced it is also in someways held back by COSMO. Thus, where there are improvements they are modest.

- *) Fig 1: As you mention in the text, the chosen site shows some pretty strong orography and heterogeneity. Is there a way that you could assess the influence of this on your simulations? For instance, a run with interactive land surface with one of the models would be nice.

  Considering the influence of the orography and the heterogeneous surface would require some major changes in the setup. As we are using horizontal cyclic boundary conditions, a sufficient buffer zone around the analysis region would be needed to overcome the influence of the cyclicity of the flow. Thus, the modeling domain would have to become larger in dependence on wind-speed and wind-direction. Considerably more effort would be needed to study the influence of the real-world heterogeneities on the flow.

  A run with an interactive land surface model instead of prescribing surface temperature and humidity could be a first step towards assessing the influence of a more realistic surface. But still, the land surface model as it is implemented in PALM and UCLA-LES needs input like soil and vegetation type which has to be prescribed horizontally homogeneous. In general we refrained from using the land surface model in the two LES models to facilitate the comparison between the two models.

  Nonetheless, there have been comparisons between simulations around Jülich in a semi-idealized setup with PALM, UCLA-LES and the Dutch Atmospheric Large-Eddy Simulation model DALES. In the simulations with DALES, a land-surface model based on the IFS-scheme was used. The following figures show comparison of the models with observations for a 3-day run (24-26 April 2013). Note that the DALES simulation differs from the PALM and UCLA simulations also in other aspects (different radiation scheme and large-scale forcing data set) which prohibits to assess how large the influence of

the land-surface scheme is.
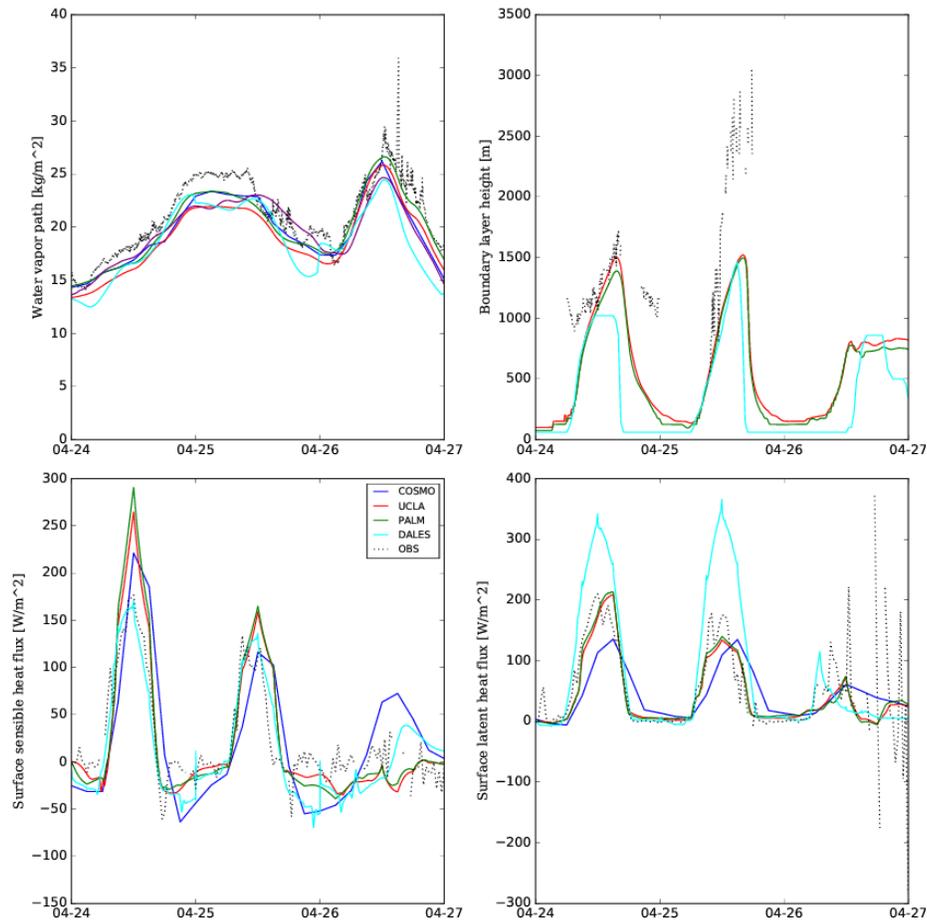
DALES is able to reproduce the surface sensible heat flux much better than PALM and UCLA-LES. The surface latent heat flux is overestimated strongly on 24 and 25 April and is much higher than in the simulations with PALM and UCLA-LES. The boundary layer depth in DALES is comparable to PALM and UCLA-LES. Thus, even though an interactive land-surface model is used (and radiation and another forcing data set) in DALES, the difference to the observations cannot be significantly reduced.

To asses the influence of surface heterogeneity and orography around Jülich the limited-area LES over Germany presented in Heinze et al. (2017) can be used (see also the statement on p. 23 l. ??). The present manuscript can be seen as basis for a further study were we try to understand which importance the mesoscale forcing coming from orography and surface heterogeneity has. In this ongoing work comparisons are made between the simulations with ICON in the semi-idealized setup, limited-area LES with ICON on smaller domains centered around Jülich and the limited-area LES over Germany.

Please also see our answer to point 8 of referee #II (p. 3).

- *) p8, l5: Could this lagging (that you also note later in the paper) be caused by the way the nudging is set up? Since the nudging is always towards the current COSMO state, and with a finite time scale, I'd expect some lag to be present. An experiment with nudging towards a future state (may be an e-folding time away?) could be interesting to test this.*

We agree that the way the nudging is set up, a lag is to be expected. The following figure shows how the boundary layer depth evolves in case that different nudging time-scales are used. The first 5 days of the simulation-period are shown. REF denotes the simulation with a nudging time-scale of 6 h, N1 has a time-scale of 1 h and N12 a time-scale of 12 h. No considerable differences occur between REF and N12. In run N1, however, z_i agrees more with COSMO than with REF, which can be clearly observed for Days 4 and 5 and which indicates a large impact of the nudging terms in this case. The stronger (tighter) the nudging the smaller the lag.
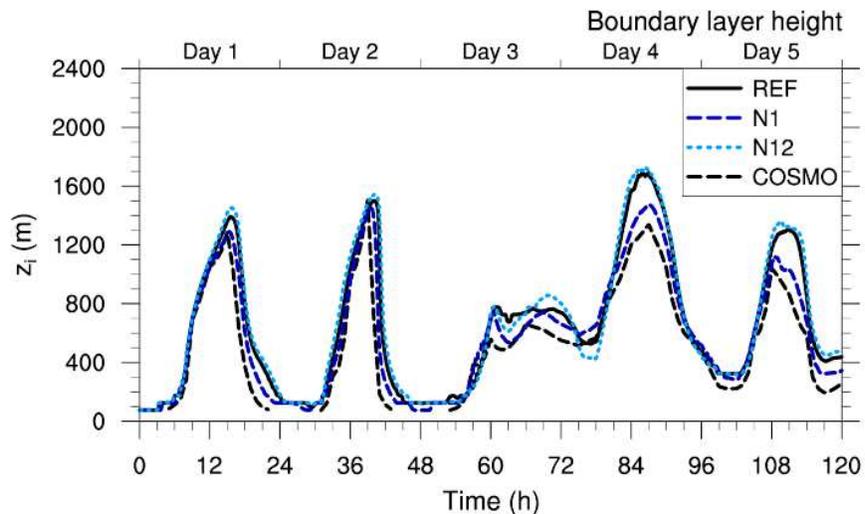
Nudging towards a future state is a very interesting idea. Unfortunately, it would go beyond the scope of the present paper.

- *) p8, l9: Could you quantify the magnitude of the nudging tendencies? For instance with some normalized scatter plots of each of the tendencies for several cases? Figure 2 is not perfect to distinguish what happens for instance during the night, or under specific circumstances. Another way of presenting this could be to look at the aggregate diurnal cycle of the tendency terms.*

We followed your suggestion and calculated the aggregated diurnal cycle of the tendency terms by averaging the tendency terms shown in Fig. 2 over the 19 days. The results are provided in the following figure.

4
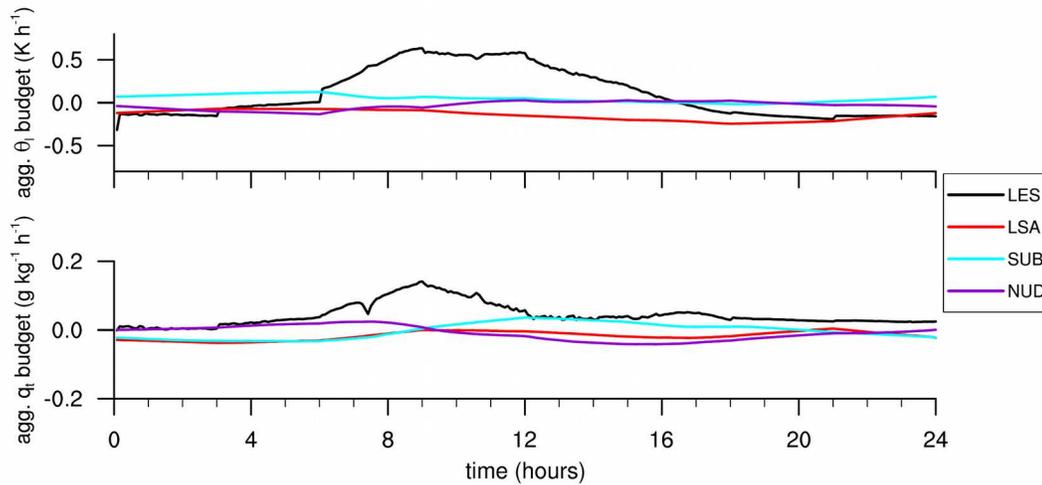
On average the fast physics (LES tendencies) are of the same order of magnitude as larger-scale advection (LSA), subsidence (SUB) and nudging (NUD) during the night. For the liquid water potential temperature budget, advection and nudging lead both to cooling between 0h and about 9h. The effect of nudging on the total water specific humidity budget is rather small between 20h and 7h. During this time, larger-scale horizontal and vertical advection act simultaneously to dry the atmosphere.

- *) Table 2: I notice that there are no false positive cumulus days in your simulations. Is this a coincidence, or is this a sign of some bias, for instance in the surface energy balance?*

This is a good point. A false positive cumulus day means that shallow clouds were simulated although they have not been observed. There are indeed no such days in our simulations. Looking also at Fig. 6b we can see that there are several days where the peak surface latent heat flux in the LES is lower than the observed weighted average (e.g., 27-30 April, 7-12 May). This could hint to a dry bias where the simulated boundary layer is not moist enough to form capping clouds.

- *) p12, l 5: Is there a way you could prove that the advection schemes are to blame for the difference between the LESs? Would the difference for instance be mitigated at a higher resolution, or is there a way to change the advection scheme in one of the models? A single day run would be sufficient to make the point that numerics matter.*

We did an additional simulation with PALM on the small horizontal domain (setup RPS, see Tab. 5) where the advection scheme for momentum and scalars was changed from a fifth-order scheme based on Wicker and Skamarock (2002) [WS] to a second-order scheme based on Piacsek and Willimas (1970) [PW]. The following figure shows cloud and rain water in the lower 5 km of the modeling domain for WS (upper figure) and PW (lower figure). The same contours as in Fig. 4 are used.

5
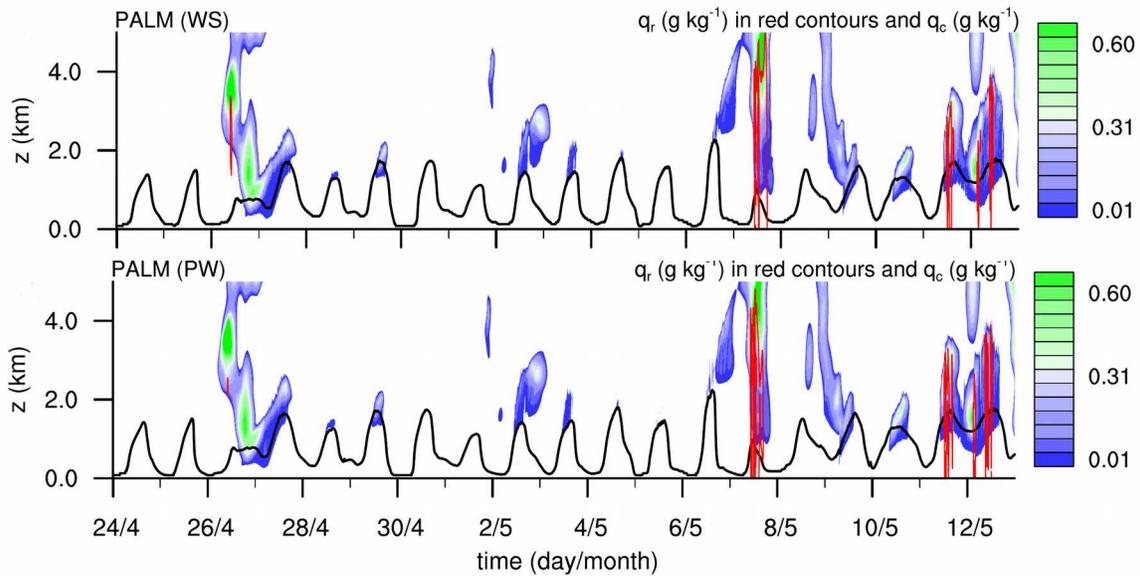
It can be seen that the cloud water is slightly sensitive to the advection scheme. On days with capping cloud layers (28/4, 29/4, 3/5), the WS-scheme produces slightly larger qc than the PW-scheme. The sensitivity of rain water is stronger leading to higher rain water on 7/5, 11/5 and 12/5 in case the PW-scheme is used. Overall, the influence of the advection scheme is small but numerics do mater nonetheless.

- *) p12, l 21: I'm not entirely convinced about the need for your Richardson based BL depth assessment. I see that this is an attempt to come up with something that applies to all BL types, but then you never discuss the stable boundary layer anyway. Why not stick with something as close as possible to the observed quantities (e.g., max gradient in some profile)? That would make the comparisons you make with those observations a lot more fair, and you could produce a more realistic error/bias assessment.*

We decided to use a more universal criterion like the height of a critical Richardson number after carefully comparing several methods. The following figure shows a comparison between the chosen method (bulk Richardson number, black), the vertical location of the average minimum sensible heat flux (standard flux method, blue) and the vertical location of the largest increase of potential temperature (gradient method, red) for the simulation of 5 days (24-28 April 2013).
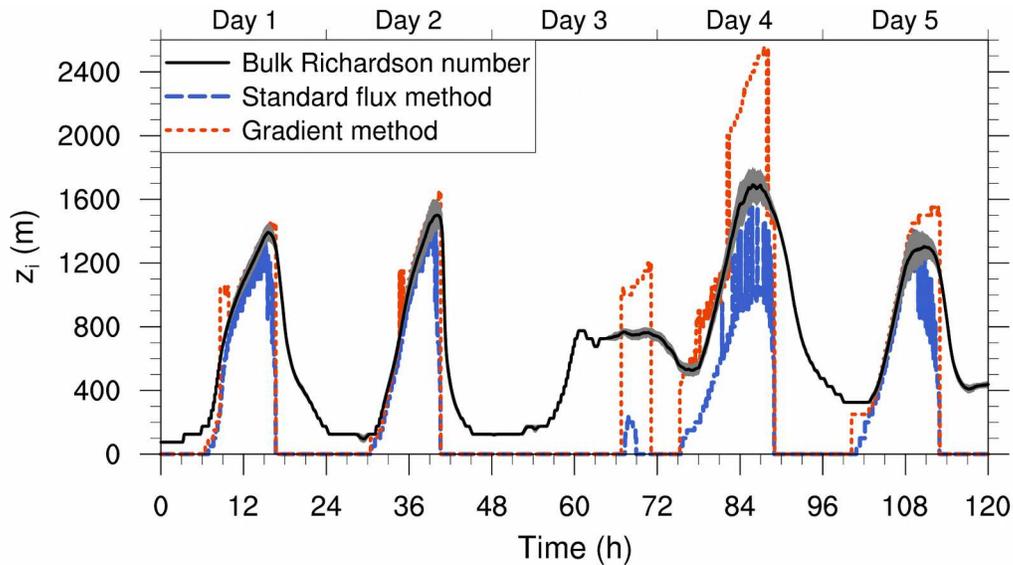
It can be seen that the standard flux and th gradient methods do not yield satisfying solutions for this simulation. They show unrealistic sudden jumps of z_i and fail at night-time (i.e., no boundary layer height is diagnosed at all). Furthermore, the bulk Richardson number method, however, produces a relatively smooth solution and is able to determine a $z_i > 0$ m at all times for this simulation. During the daytime of Days 1 and 2 typical CBLs form, and general agreement of the results of all three methods is found. To a lesser extend, this also applies for Day 5. During the frontal passage on Day 3 and the following post-frontal situation on Day 4, the flux and gradient method yield no trustworthy solutions.

- *) p18, l 12: I'm not sure why May 5 was chosen as a highlight, at least not as a cumulus day. As the figure shows, it mostly displays forced clouds, with little buoyancy of their own. I would expect this BL to behave like a dry CBL, as it does. If you want to show case a cumulus day, it may be nice to pick another one (if available).

You are absolutely right, May 5 is not the perfect show-case for a shallow cumulus topped boundary layers. Unfortunately, other days with a prototype shallow cumulus layer are not available.

- *Other comments:*
  - *) p1, l 20: There is quite some newer literature on LES of the stable BL. For instance the GABLS 3 cases (Basu, 2012 or Edwards, BLM 2014)

    A reference to Edwards et al. (2014) and Ansorge and Mellado (2014, 2016) was added (p. 2, l. 1)

  - *) p4, l7: both; momentum <- extra semicolon

    Comma was replaced by semicolon (p. 4, l. 15)

  - *) p5, l20: SUB is not homogeneous because phi=phi(x,y,z), right? so uLS=uLS(z)

It is correct that the SUB tendency is not homogeneous which is already reflected in the text (p 5, l. 32)

○ *) p6, eq 2: Does this mean that your nudging time scale is constant over height? Have you tried playing with this?*

Correct, the nudging time scale is constant over height. We played with the profile of the nudging time-scale. The following figures show the tested profiles and the effect on the boundary layer depth.

PBL height and standard deviation

IOP 6 (Clear Sky) z_i: Bulk Richardson Number
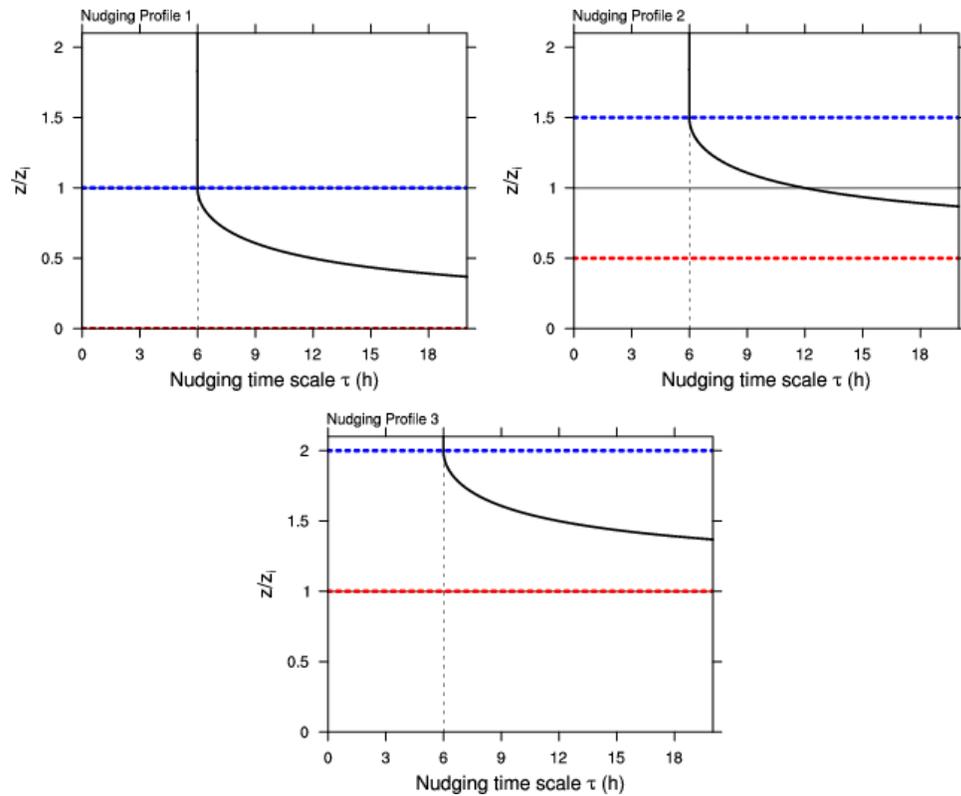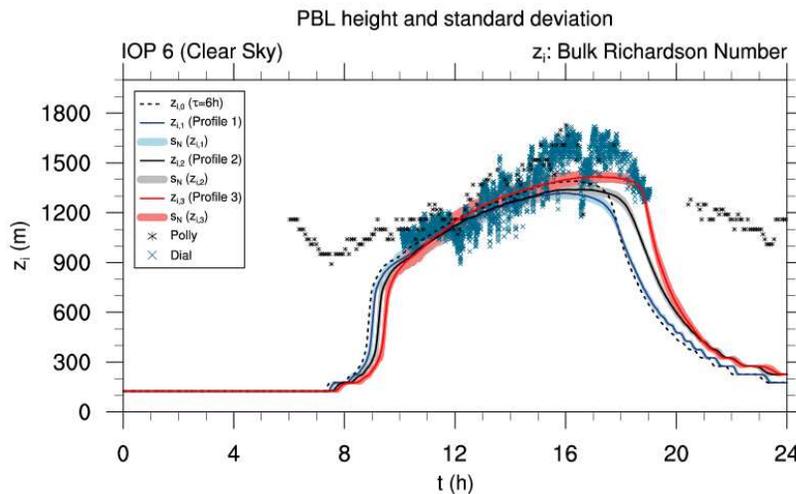
For the Nudging Profile 1, the nudging time scale is set to 10^{10} s between the surface and 0.5 z_i (which results in having no nudging at all). Between 0.5 z_i and z_i, the time scale is then brought to 6 h via a cosine-function. For profile 2 and 3 the same procedure is applied between 0.5 z_i and z_i (Profile 2), and z_i and 2 z_i (Profile 3), respectively. Test simulations of 24 April 2013 (IOP 6) show that in terms of the boundary layer depth, the effect is minor during the convective hours (10 UTC to 16 UTC). All realizations lie in the spread of the measurements.

○ *) p7, l 11: depth of the boundary layer, not height. This may be present a few more times in the manuscript

   The word *height* was replaced by *depth* where appropriate.

○ *) p11, l 31: "Anyhow...." very colloquial. Anyhow's happen a few more times in the manuscript

   The word *anyhow* was replaced by *however* where appropriate.

○ *) p11, l 31 "the more challenging..." not entirely clear what you mean with this sentence.

   We wanted to make the point that simulating proper shallow cumulus layers on 25 April and 1 May is difficult for the LES. The sentence was rephrased to reflect this (p. 12, l. 12).

○ *) p13, l9: How does the BL depth perform in SCu topped BLs? I assume that the cloud base is then the desired BL depth, but I can imagine that a strong variability can be observed in those cases.

   In the 19-day period which was simulated there was unfortunately no real stratocumulus-day. The cloud structure on 10 May resembled stratocumulus as close as possible around noon. In Fig. 3l (p. 32), it can be seen that the cloud top in PALM is around 1.3 km (maximum of qc) which is comparable to the value of z_i during noon (see Fig. 5c, p. 34). Thus, cloud base and z_i based on the bulk Richardson number criterion resemble.

- *) p13, l 21: Why this threshold? It seems arbitrary to me. Can't this be normalized with the maximum variance or so?

  In Schween et al. (2014) a detailed discussion on the choice of the threshold and its sensitivity to the derived mixing layer height can be found.

- *) p15, l9: I printed the paper in Black and White (yes, my department is old fashioned). While I can follow the rest of the manuscript without problems, I could not see the highlighted yellow sections. A better highlight, or just a table with a description of all the days (clear/cumulus/frontal/strong forcing/...) would be nice.

  We changed the highlighting from light yellow to black stippled in Figs. 5-8.

- *) p20, l 12: What does RPS mean?

  RPS is the name of a case listed in Table 5 (p. 20), S denotes the usage of a small horizontal modeling domain. A hint is given in the text now. (p 20, l. 6)

- *) p20, l 33: "the question arose" may be a bit more formal?

  The sentence was changed to "the question came up ..." . (p 21, l. 7)

- *) P24,l 9: The URL is not yet complete. I assume that will happen later?

  Correct.

Authors' response to the comments of Referee #II on the manuscript

Evaluation of large-eddy simulations forced with mesoscale output for a multi-week period during a measurement campaign
by R. Heinze et al.
submitted to ACP

27th April, 2017

Dear Referee II,

Thank you very much for your comments on the manuscript. In what follows we present our replies (plain text) to your comments and questions (*italics*). The line numbers and the page numbers refer to the revised manuscript (without mark-up) which is attached to this response.

- *1.P1 L20. The key problem with stably stratified BL is the turbulence intermittency and wave-turbulence coupling.*

  This information was added (p. 1, l. 20).

- *2. P2 L17. I do not think the discussion following the statement "in various single-column and cloud-resolving modeling studies "has anything to do with CRM studies. I suggest revising it to read "in various single-column and LES studies".*

  In this sentence, "cloud resolving modeling" was replaced by "LES" (p. 2, l. 18).

- *3. Section 2.1 and in other places in the text. The terminology applied is not correct. The Boussinesq approximation refers to applying density perturbations only in the gravity term. Thus, most (if not all) nonhydrostatic atmospheric models are Boussinesq (both anelastic and compressible). A more appropriate distinction would be shallow convection approximation (in which case the incompressible assumption is valid and density is constant)) versus deep convection approximation (in which case anelastic or compressible equations are needed). So PALM applies shallow convection approximation (and obviously cannot work properly for deep convection), whereas UCLA-LES is anelastic.*

  The terminology "Boussinesq approximation" was replaced by "shallow-convection approximation" (p. 4, l. 10 and p. 6, l. 29).

- *4. A more general comment on model variables (no need to respond, just think about this for the future).*

1

Liquid water potential temperature is a perfect variable for shallow nonprecipitating convection. For deep convection, two problems arise. One, the exact formulation is rather cumbersome and I doubt it is used in the two models. Second, when precipitation is considered, then it is not conserved and its sources need to be considered. In contrast, equivalent potential temperature (or moist static energy, MSE) is conserved as long as ice processes are excluded. Should then MSE be used?

Thanks for pointing this out. There is also a problem with MSE as it is not conserved for isobaric ascent, i.e., when isobars are not aligned with iso-heights. It also has difficulties with the ice phase and is only approximately conserved in the face of precipitation.

- *5. Another comment on the model. I do not think the water variable is the "total water specific humidity". I think it is (or should be) the mixing ratio. Please do a simple math exercise with the density of total water, dry air density, and total air density to show that mixing ratio is an appropriate variable when sources (e.g., of water vapor) are considered. The continuity equation for the specific humidity has an extra multiplier for the source term. The equation for the mixing ratio does not, and thus is preferable. Of course in practice the differences are minuscule and can be neglected.*

Indeed, your are correct. Both models use mixing ratios as mass fraction and not specific humidities. The text was changed accordingly (e.g., p. 4, l. 24).

- *6. P. 4. I think the fundamental differences between Eulerian thermodynamics in UCLA LES and Lagrangian approach in PALM should be better exposed in point 4. These are more significant than the authors realize I think (see below). Also, is PALM really applying saturation adjustment (the paragraph starting on P4L15) and Seifert/Beheng 2-moment microphysics? This is the Lagrangian Cloud Model, so it uses superdroplets, correct?*

Point 4 on page 4 was thought to show some different fields of study in which PALM and UCLA-LES are generally used. Whereas PALM is well known to have a Lagrangian cloud model focusing on shallow convection, UCLA-LES is widely used to further develop and test microphysical schemes in an Eulerian approach. Point 4 was not meant to reflect specifically what was used in the present simulations. As described from page 4, line 24 onwards, PALM and UCLA-LES both used saturation adjustment and the warm 2-moment microphysics scheme. Thus, PALM does not make use of the Lagrangian cloud model in the present simulations.

The text was changed to point out more strongly that both models use saturation adjustment in the present simulations (p. 4, l. 25-26).

- *7. This is perhaps the most significant comment. Formulation of the forcing methodology in section 2.2 comes out of nowhere. I realize that similar methodology has been used by others, but I still think that the methodology can be better explained (perhaps with the exception of the nudging). Please have a look at section 2 in Grabowski et al. (JAS 1996, p. 3684-3709) that formally derived the forcing terms in case of evolving large-scale conditions and later applied it in cloud-resolving model simulations. Perhaps some of the forcing terms derived in that paper are missing in the way LES is forced. Just a thought...*

We agree that the methodology on how the larger-scale forcing components are derived is missing. We added a small paragraph to point out that the forcing terms have a sound physical basis and that they are derived in detail for example in Grabowski et al. (1996) (p. 5, l. 16-19). We refrain from presenting the methodology anew as Grabowski et al. (1996) give a precise description on how to derive the larger-scale forcing terms.

- *8. The domain shown in Fig. 1 features quite a significant topography. Is then the periodic domain without topography justified? Can one use COSMO output to show if the topography has some effect (e.g., by comparing COSMO results north and south from the observation sites). Does the hill NE to the observation sites affect the observations? Are any observation made from the top of the hill?*

Considering the influence of the topography would require some major changes in the present setup. As we are using horizontal cyclic boundary conditions, a sufficient buffer zone around the analysis region would be needed to overcome the influence of the cyclicity of the flow. Thus, the modeling domain would have to become larger in dependence on the wind-speed and wind-direction. Considerably more effort would be needed to study the influence of the real-world topography on the flow.

The COSMO output which is used here for driving the LES cannot be used to study the effect of the hill as the output is already horizontally averaged over squares where Jülich is in the centre (see Fig. 11 (p. 40)). Thus, the original COSMO fields would have to be obtained from the archive of the German Weather Service to answer this question.

The measurements during HOPE were taken between the debris hills and unfortunately not on top of the hill Sophienhöhe (see also the ACP HOPE overview paper of Macke et al. (2017)). Figure 1b shows the three principal measurement sites which are located in the rather flat region between the two debris hills. Thus, we cannot assess the influence of the hill on the measurements.

To asses the influence of surface heterogeneity and orography around Jülich the limited-area LES over Germany presented in Heinze et al. (2017) can be used (see also the statement on p. 23 l. 30). The present manuscript can be seen as basis for a further study were we try to understand which importance

the mesoscale forcing coming from orography and surface heterogeneity has. In this ongoing work comparisons are made between the simulations with ICON in the semi-idealized setup, limited-area LES with ICON on smaller domains centered around Jülich and the limited-area LES over Germany.

Please also see our answer to point 2 of referee #I (p. 2).

- *9. Section 2.3 does not mention how surface fluxes are calculated. Is the same surface temperature and humidity assumed throughout the LES domain? How variable are surface conditions (e.g., soil type, soil moisture) over the LES domain. Do such considerations matter?*

At the surface, temperature and humidity are prescribed horizontally homogeneous (see p. 7, l. 8). Then Monin-Obukhov similarity theory is used to calculate surface sensible and latent heat flux (see p. 14, l. 23). A statement was added in section 2.3 to point out how the surface fluxes are derived (p 7, l. 10).

The surface conditions are horizontally homogeneous but change in time as surface temperature and humidity are taken from the larger-scale forcing data set. In the LES models, no land-surface model is used to facilitate the comparison between the two LES models. Thus, parameters like soil type, vegetation type or soil moisture do no enter our simulations.

- *10. I like section 2.4. I think it is important to realize the magnitude of various terms and their evolutions. A small technical comment: the red and orange lines are not easy to distinguish. One of them can be green.*

The color line of the nudging tendency in Fig. 2 was changed from orange to violet. We refrain from using red and green in one figure as red-green color blindness is rather widespread.

- *11. Fig. 3 is very nice. It also shows that BL tends to be systematically colder in LES than in observations. Is that a coincidence, or is this true for the entire simulation?*

Looking at Fig. 7c where the potential temperature at a height of 25 m is shown, it can be seen that the LES are usually colder during day times (with some exceptions on 26 April, 27 April,11 May) in the boundary layer. Thus, there seems to be a tendency that the simulations are a bit too cold during daytimes compared to radio soundings. This information was added in the manuscript (p. 15, l. 12).

- *12. P. 12, paragraph below Table 2. I think the origin of the difference between clouds simulated by the two LES models goes beyond just the advection scheme. See comment 6 above. For instance, PALM predicts supersaturation, correct?*

As already stated in the answer to comment 6, the Lagrangian cloud model was not used in PALM. Thus, PALM and UCLA-LES both use saturation-adjustment and the same microphysics scheme. Additional simulations with another advection scheme in PALM are presented in the answer to referee I and show that cloud and rain water are slightly dependent on the advection scheme.

- *13. P13L35. I think this is 5 cm/s, not 50 (this would agree with captions to Fig. 5 and 6).*

The typo was corrected (p. 14, l. 12).

- *14. Fig. 6. Despite taking entire page, details of the figure are difficult to see. I suggest breaking the figure into 3 separate figures.*

Figure 6 was broken into 3 separate figures (Fig. 6, 7, 8).

- *15. The integrated water vapor is commonly referred to as the precipitable water (PW).*

We decided to retain the formulation "integrated water vapor" as it can synonymously be used with precipitable water.

- *16. P20L12-16. I am not sure about the discussion of mesoscale circulations. Since the model excludes topography and (I assume) applies homogeneous surface conditions, the simulated mesoscale circulations may be significantly weaker than in nature.*

We totally agree that the simulated mesoscale circulation may be weaker than in nature as orography and surface heterogeneities are neglected. We changed the description by naming these circulations internally generated (p. 19 l. 32; p. 23 l. 7). Furthermore, the semi-idealized simulations presented in the manuscript can be seen as first step towards assessing the role of the mesoscale as there is ongoing work to explore the effect of orography and surface heterogeneity by comparing simulations with ICON in semi-idealized setup and limited-area setup (p 3., l. 3).

- *17. Figure 8 shows that impact of some of the changes is smaller than standard deviations which suggests that the differences may not be statistically significant. How is the standard deviation defined in those plots?*

The metric *mean peak difference to PALM* including the standard deviation is calculated as follows and possibly best followed while also looking at Fig. 5 where the temporal evolution of the boundary layer depth of different sources are shown. For the metric, the absolute difference PALM – COSMO or observation is taken for every available data point. (We refrained from using the norm of the difference

in order to obtain an unbiased result.) Then only those values are further processed which occur between 12 and 14 UTC for each day. This data subset is then averaged and the standard deviation is calculated as the square root of the variance of the data subset.

The large standard deviation in some cases can be attributed to the large scatter in the observations. Figure 5 shows that for example on 24 April Polly and Halo are close to PALM whereas they deviate noticeably on 25 and 26 April. Thus, a large scatter in the peak difference to PALM can be expected. As for some sensitivity studies only 3 days instead of 19 days are simulated (see Tab. 5) the number of values entering the statistics is also quite small in comparison to the statistics over 19 day runs.

Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign

Rieke Heinze[1,2], Christopher Moseley[1], Lennart Nils Böske[2], Shravan Kumar Muppa[3], Vera Maurer[4,5], Siegfried Raasch[2], and Bjorn Stevens[1]

[1]Atmosphäre im Erdsystem, Max-Planck-Institut für Meteorologie, Hamburg, Germany
[2]Institut für Meteorologie und Klimatologie, Leibniz Universität Hannover, Hannover, Germany
[3]Institut für Physik und Meteorologie, Universität Hohenheim, Stuttgart, Germany
[4]Institut für Meteorologie und Klimaforschung - Department Troposphärenforschung, Karlsruhe Institut für Technologie, Karlsruhe, Germany
[5]Klima- und Umweltberatung, Deutscher Wetterdienst, Offenbach, Germany

Correspondence to: R. Heinze (rieke.heinze@mpimet.mpg.de)

Abstract. Large-eddy simulations (LES) of a multi-week period during the HD(CP)[2] (High-Definition Clouds and Precipitation for advancing Climate Prediction) Observational Prototype Experiment (HOPE) conducted in Germany are evaluated with respect to mean boundary layer quantities and turbulence statistics. Two LES models are used in a semi-idealized setup through forcing with mescoscale model output to account for the synoptic-scale conditions. Evaluation is performed based on the HOPE observations. The mean boundary layer characteristics like the boundary layer depth are in a principal agreement with observations. Simulating shallow-cumulus layers in agreement with the measurements poses a challenge for both LES models. Variance profiles agree satisfactorily with lidar measurements. The results depend on how the forcing data stemming from mesoscale model output is constructed. The mean boundary layer characteristics become less sensitive if the averaging domain for the forcing is large enough to filter out mesoscale fluctuations.

1 Introduction

Large-eddy simulation (LES) studies have usually focused on a specific atmospheric boundary layer type often with the purpose of addressing a specific theoretical question. Many early atmospheric LES initially focused on cloud-free, convective boundary layers (e.g., Deardorff, 1970b, 1972; Moeng, 1984). Later various studies additionally investigated the effects of wind shear in the convective boundary layer (e.g. Mason, 1992; Moeng and Sullivan, 1994). The role of clouds in the dynamics of the boundary layer has motivated more sophisticated LES of cloud-topped boundary layers. Stratus and stratocumulus clouds have been considered in numerous works (e.g., Deardorff, 1976, 1980; Moeng, 1986; Stevens et al., 1998) and shallow cumulus clouds have also been successfully simulated (e.g. Sommeria, 1976; Cuijpers and Duynkerke, 1993; Brown et al., 2002; Siebesma et al., 2003). Less attention has been paid to stably stratified boundary layers because their simulation requires even higher resolutions and computer resources (compared to LES of convective situations) as the stable boundary layers are usually very shallow~~and exhibit small turbulence intensity~~. Furthermore, turbulence in the stable layers is usually intermittend and also coupled to waves. Nonetheless, there are several LES and direct numerical simulation studies of the stable boundary layer

(e.g. Mason and Derbyshire, 1990; Saiki et al., 2000; Beare et al.,
There are other studies investigating the diurnal transition between different boundary layer types (e.g., Nieuwstadt and Brost,
1986; Sorbjan, 1997; van Stratum and Stevens, 2015).

In reality, however, different types of the atmospheric boundary layer occur consecutively if longer time periods spanning
weeks to months and even years are considered. LES of these longer time periods (in the following called *long-term LES*) be-
came computationally tractable through massively parallel codes and advances in computing (Schalkwijk et al., 2012, 2015).
What benefits and new insights can be gained from the long-term LES approach compared to previous studies? First of all,
LES models can be regarded as virtual laboratories, in which the characteristics of atmospheric micro-scale flows can be
studied and understood under controlled conditions (Neggers et al., 2012). One major practical benefit from LES is the de-
velopment and improvement of boundary layer parameterization schemes (e.g., Noh et al., 2003). By testing parameterization
schemes with a multitude of different boundary layer situations (including transitions), the tuning towards special atmospheric
conditions, which might even not be representative, can be avoided (Neggers et al., 2012). Furthermore, realistic long-term
turbulence datasets are also of great interest in other fields of study, especially those with a high practical orientation (e.g.,
studies concerning wind energy (Vollmer et al., 2015) or air quality and ventilation effects in urban environments).

When focusing on LES longer than several hours the importance to include synoptic-scale meteorological conditions in LES
increases. Larger-scale forcing in terms of time-varying horizontal and vertical advective tendencies as well as larger-scale
pressure gradients (geostrophic wind) should be prescribed to account for the overall larger-scale conditions. The strategy
to prescribe larger-scale forcing terms has been applied in various single-column and ~~cloud-resolving modeling~~ LES studies
(e.g., Randall and Cripe, 1999). Even early LES case studies (e.g., Sommeria, 1976) included synoptic-scale forcing. For
idealized LES case studies focusing on a specific boundary layer type, the larger-scale forcing is usually constructed based on
observations from measurement campaigns (e.g., Siebesma et al., 2003; Stevens et al., 2005). Synoptic-scale forcing can also
be obtained from larger-scale models (e.g., Neggers and Siebesma, 2013) or a combination of observations and models (e.g.,
Baas et al., 2010; vanZanten et al., 2011; Pietersen et al., 2015). Regarding long-term simulations in the semi-idealized setup,
relaxation towards a reference state given by a larger-scale model or observations can be used in combination with advective
forcing to prevent model drift in time (Neggers et al., 2012).

In this study LES covering almost three weeks (19 days) of the HD(CP)^2 (High-Definition Clouds and Precipitation for
advancing Climate Prediction) Observational Prototype Experiment ~~(HOPE)~~ (HOPE, Macke et al., 2017) are evaluated by
comparing the results with the multi-sensor HOPE dataset specifically designed with this purpose in mind. Particularly, the
operation of several new lidar systems simultaneously during HOPE provided the unique opportunity to study planetary
boundary layer characteristics with unprecedented detail. The importance of high-resolution thermodynamic profiling for
model evaluation is also outlined in Wulfmeyer et al. (2015). Results of a year-long LES centered at a meteorological ob-
servational supersite were presented by Schalkwijk et al. (2015). They followed a statistical approach to assess the quality of
their long-term LES by comparing yearly-averaged diurnal cycles and climatologies with those from observations and con-
cluded that the semi-idealized approach is stable enough to simulate a whole year of varying conditions. The present study
focuses on a day-to-day comparison with observations from a measurement campaign which also accounts for spatial vari-

ability by providing measurements at three different principal measurement sites. Here, we want to tackle the question if the long-term LES approach is able to deliver a realistic boundary layer representation. Within this regard, the study is one of the first approaches to allow for a direct comparison of LES to measurements for a period longer than several days. Furthermore, the study can serve as a basis for understanding the role of the mesoscale by comparing results of LES in a limited-area setup
5   (where, e.g., orography and surface heterogeneity are considered as in Heinze et al. (2017)) with LES in the semi-idealized setup presented here.

To asses how representative and robust the results of the present study are, two strategies are followed. One strategy is to use two well-established LES models instead of just one. Applying two LES models provides a measure for the variability among the LES - comparable to assessing observations of one quantity from multiple sensors. The other strategy is to study
10   how the results depend on details of the setup. As the long-term LES approach relies on prescribing larger-scale forcing it is ~~for example~~ important to know how sensitive the LES results are with respect to details of the forcing like the calculation of the larger-scale advective tendencies from the mesoscale model or the relaxation (nudging) to the mesoscale model. Furthermore, this gives us the opportunity to assess the extent to which mesoscale variability plays a role in determining boundary layer characteristics.

15   This study also has some relevance for using 3D LES in the form of a *superparameterization* in large-scale (global) models as proposed by Grabowski (2016). In this approach, an LES model is embedded in each column of the large-scale model with horizontal grid lengths in the order of 10-50 km to account for an improved representation of small-scale processes in global models. In each global model grid box, one LES runs on a separate core of a massive parallel computer and communicates with the global model by exchanging only mean profiles during the simulation. The long-term LES approach under investigation
20   would be representative for the superparameterization of one global model grid box.

Note that the LES statistics in this semi-idealized setup with prescribed forcing, can only provide a mean over a certain representative area. The measurements on the other hand are exposed to spatial variability and a point measurement at a certain location is only a local sample. As measurements from all available HOPE-sites are used for the comparison between LES and observations, a certain degree of variability can be expected in the measurements and compared to LES. In this sense we expect
25   that a fair comparison between observations and LES is possible and that the measurements inform what should be expected to be seen in the representative LES.

The paper is organized as follows. Section 2 provides a description of the LES models applied and of the setup. The relative importance of the larger-scale forcing terms is also assessed. Section 3 gives an overview of the measurement campaign HOPE and of the observations used in this study. In Sect. 4 the 19-day reference simulation is analyzed. First (Sect. 4.1), the temporal
30   evolution of key boundary layer quantities is discussed. Next (Sect. 4.2), vertical profiles of second-moment turbulent quantities for a cloud-free and a shallow cumulus case are compared with profiles obtained from lidar. In Sect. 5, the results of various sensitivity runs are presented. Summary and conclusions are presented in Sect. 6.

2 Large-eddy simulations

2.1 Large-eddy models

Two well-established LES models PALM (PArallelized Large-eddy simulation Model 4.0, revision 1574, https://palm.muk.
uni-hannover.de, Maronga et al., 2015) and the UCLA-LES (University of California, Los Angeles Large-Eddy Simulation
model, Stevens et al., 2005) are used in the present study. Both finite-difference models solve the same set of implicitly filtered,
incompressible, nonhydrostatic Navier-Stokes equations including the three velocity components u, v, w and the perturbation
pressure p as well as the transport equations for liquid water potential temperature θ_l, ~~the total water specific humidity~~ total
water mixing ratio q_t, rain water ~~specific humidity~~ mixing ratio q_r and number concentration N_r on a staggered, C-type (Harlow
and Welch, 1965; Arakawa and Lamb, 1977) Cartesian grid. The major differences between the two models are listed below.

1. In PALM the ~~Boussinesq-approximation~~ shallow-convection approximation (Dutton and Fichtl, 1969) is used where the reference density is constant. UCLA-LES solves the equations in the less constrained anelastic approximation (Ogura and Phillips, 1962) allowing for a varying reference density with height.

2. Sub-grid scale (SGS) turbulence closure is prognostic in PALM by solving the equation for the SGS turbulence kinetic energy according to Deardorff (1980) and diagnostic in UCLA-LES using a classical Smagorinsky (1963) scheme.

3. PALM uses a fifth-order advection scheme based on Wicker and Skamarock (2002) for both~~,~~ momentum and scalars. In UCLA-LES a fourth-order central advection scheme is applied for momentum and a monotone second-order scheme with flux-limiter for scalars.

4. PALM includes a Lagrangian cloud model and was often used in studies discussing shallow-convection (e.g., Riechelmann et al., 2015; Hoffmann et al., 2015; Hoffmann, 2016). UCLA-LES incorporates a hierarchy of microphysical models and representations of radiative transfer and was applied in studies more focusing on deep-convection (e.g., Rieck et al., 2015; Schlemmer and Hohenegger, 2016).

PALM and UCLA-LES both apply the fractional-step method to ensure incompressibility of the flow and the resulting Poisson-equation for the perturbation pressure is solved by a fast Fourier transform. ~~The cloud water specific humidity~~ In the simulations presented here, the cloud water mixing ratio q_c is obtained via the simple saturation-adjustment scheme in both models. The warm microphysics scheme of Seifert and Beheng (2001, 2006) and Seifert (2008) is applied and Monin-Obukhov similiarity theory is used at the surface. A no-slip condition is applied to the horizontal velocity components at the surface. The horizontal boundaries are cyclic and both models use a third-order Runge-Kutta method with a variable time step to advance in time. The parallelization method follows a 2D domain-decomposition using Message Passing Interface for inter-process communication.

2.2 Forcing with mesoscale model output

To account for synoptic-scale forcing, the effects of larger-scale pressure gradients, horizontal advection and vertical motions have to be prescribed in LES. However, the usage of lateral periodic boundary conditions constrains how the synoptic scales

can be represented. As horizontal LES domain-scale gradients cannot be represented, larger-scale advection, pressure gradients and vertical motions are assumed to be horizontally homogeneous, but they may vary in time and height. This approach has direct implications on how larger-scale phenomena can be represented in the LES, for instance frontal passages. In the presence of a front, the flow field exhibits strong local gradients perpendicular to the front. However in the LES, a front would arrive and depart from the entire domain simultaneously at a specific height due to the periodic boundary conditions. Thus, the evolution of frontal passages is represented in time rather than in space (Schalkwijk et al., 2015).

Time-dependent surface conditions, which are representative for the entire LES domain, are required. To facilitate the comparison between the two LES models, surface values are prescribed instead of using a land-surface model.

The larger-scale forcing can be generated from 3D output of a larger-scale (global or limited area) climate or numerical weather prediction model (e.g., Neggers and Siebesma, 2013). Creating larger-scale forcing solely from measurements is also possible (e.g., Grabowski et al., 1996), or a combination (blending) of larger-scale model and observations can be applied (e.g. Baas et al., 2010; Bosveld et al., 2014). Here, the forcing ~~if~~ is calculated from analysis output of the operational mesoscale numerical weather prediction model COSMO-DE (Baldauf et al., 2011, denoted as COSMO hereinafter). The COSMO analysis is thought to provide a good estimate of a current state as it is a combination of model output and assimilated measurements. COSMO is also denoted as the host model in the following.

The larger-scale (LS) tendencies for the governing equations ~~are calculated as follows~~ can be derived formally by decomposing the variables into larger-scale and turbulence-scale components where the larger-scale component is further decomposed into a horizontally averaged and a space-dependent part. The latter is then neglected which can be justified by a scale analysis. A detailed description of this methodology is given by Grabowski et al. (1996) for using larger-scale forcing in a cloud-resolving model.

The calculation of the larger-scale tendencies in the LES is provided in the following. The effect of the larger-scale pressure gradient (LSP) enters the horizontal momentum equations

$$\left.\frac{\partial u_i}{\partial t}\right|_{\mathrm{LSP}} = \varepsilon_{i3j} f_3 u_{\mathrm{g},j}, \tag{1}$$

where $u_{\mathrm{g},i} = (u_{\mathrm{g},1}, u_{\mathrm{g},2}, 0)$ denotes the geostrophic wind vector which is calculated by means of the larger-scale pressure (p_{LS}) gradients and density (ρ_{LS}) as $u_{\mathrm{g},1} = -(\rho_{\mathrm{LS}} f_3)^{-1} \partial p_{\mathrm{LS}}/\partial x_2$, $u_{\mathrm{g},2} = (\rho_{\mathrm{LS}} f_3)^{-1} \partial p_{\mathrm{LS}}/\partial x_1$ and $u_{\mathrm{g},3} = 0$. Einstein summation convention for repeated indices is used, $f_i = (0, 2\Omega \cos(\phi), 2\Omega \sin(\phi))$ denotes the Coriolis parameter, where Ω is the angular speed of the Earth and ϕ the geographical latitude.

The contributions due to LS horizontal advection (LSA) and vertical advection (*subsidence*, SUB) enter the scalar prognostic equations only:

$$\left.\frac{\partial \varphi}{\partial t}\right|_{\mathrm{LSA}} = -\left(u_{\mathrm{LS},1} \frac{\partial \varphi_{\mathrm{LS}}}{\partial x_1} + u_{\mathrm{LS},2} \frac{\partial \varphi_{\mathrm{LS}}}{\partial x_2} \right), \tag{2}$$

$$\left.\frac{\partial \varphi}{\partial t}\right|_{\mathrm{SUB}} = -u_{\mathrm{LS},3} \frac{\partial \varphi}{\partial x_3} \quad \text{with} \quad \varphi \in \{\theta_\mathrm{l}, q_\mathrm{t}\}. \tag{3}$$

All three larger-scale velocity components $u_{LS,i}$ and scalar components φ_{LS} are needed. Note that the LSA contribution (2) is horizontally homogeneous whereas the SUB contribution (3) is not. Here, the horizontal homogeneous subsidence velocity $u_{LS,3} = w_{SUB}$ is combined with the local gradient of the LES scalar φ. This ensures that the tendencies are strongest where the local scalar gradients are largest and w_{SUB} is not negligible (which is usually at the top of the boundary layer).

5 The simulations presented in this study use Newtonian relaxation (nudging) additionally to the previously discussed larger-scale components. The main function of nudging in the larger-scale forcing framework is to prevent excessive model drift in time (Neggers et al., 2012). This drift may be introduced by errors in the LES or from systematic errors in the larger-scale forcing terms. By means of nudging the simulated flow is adjusted to the flow situation of the host model (Anthes, 1974; Stauffer and Bao, 1993). This is an additional possibility to account for larger-scale processes in an LES. However, relaxation

10 has to be handled with care, since it represents no real physical process (Randall and Cripe, 1999). To preserve turbulent structures the applied nudging tendency is horizontally homogeneous in analogue to the LSP and LSA tendencies and it is given by

$$\left.\frac{\partial \varphi}{\partial t}\right|_{NUD} = -\frac{\langle \varphi \rangle - \varphi_{LS}}{\tau} \quad \text{with} \quad \varphi \in \{u, v, \theta_l, q_t\}, \tag{4}$$

where the angle brackets ($\langle ... \rangle$) denote the horizontal average of the LES variable and τ is the relaxation time-scale which

15 defines the strength of the nudging. With a small τ, the horizontal averages of the prognostic variables are adjusted relatively fast towards the corresponding state of the host model. A nudging time scale of $\tau = 6$ h is used which is long enough for the fast boundary layer physics to develop their own unique state and short enough, so that larger-scale disturbances, such as weather fronts, can be represented in the LES (Neggers et al., 2012; Schalkwijk et al., 2015).

 The larger-scale tendency terms (1)-(4) are calculated from the operational COSMO analysis data which have a horizontal

20 and temporal resolution of 2.8 km and 3 h, respectively. Thus, the larger-scale forcing terms used in this study do not stem from pure model output as the analysis is composed of a combination of model output and assimilated measurements. It should be noted that the larger-scale tendencies should not contain any impacts of small scale phenomena which are explicitly resolved by the LES. Thus, the COSMO data is averaged spatially to filter out these scales. The averaging procedure is further described in Appendix A. The resulting larger-scale forcing profiles are linearly interpolated in time between every three hours to obtain

25 a forcing at every time-step in the LES.

2.3 Setup

The reference simulation performed with both models (denoted as *RP* and *RU* for PALM and UCLA-LES, respectively) consists of a continuous 19-day simulation covering 24 April to 12 May 2013 over the HOPE region. An isotropic grid spacing of $\Delta = 50$ m is used up to a height of 5 km above ground. Above, vertical grid stretching is applied resulting in a model top

30 of about 13 km. Note that due to the underlying assumption of incompressibility in the set of model equations, the results above a height of approx. 5 km should be interpreted with care especially for PALM due to the ~~Boussinesq~~ shallow-convection approximation used. A model top of 13 km is chosen nonetheless, as then the evolution of the prognostic variables above a certain height can be almost entirely ascribed to the larger-scale and deep-convective events in the forcing may find some

6
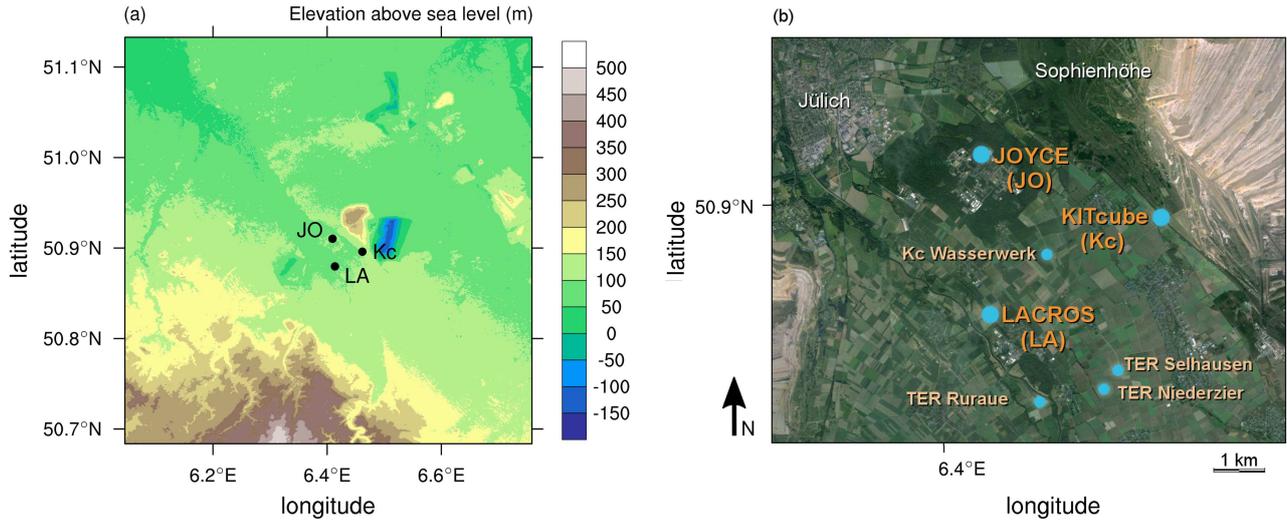
Figure 1. Location of different measurement sites during the HOPE campaign. The abbreviations JOYCE (JO), LACROS (LA) and KITcube (Kc) denote the three principal measurement sites Jülich ObservatorY for Cloud Evolution, Leipzig Aerosol and Cloud Remote Observations System and the Karlsruhe Institute of Technology cube, respectively. Panel a shows the topography in a 50 km × 50 km domain centered around JOYCE (source: ASTER GDEM Validation Team (2011)) and panel b provides a closer view on the HOPE measurement sites (source: Google Maps). Additional surface flux measurements were taken at the Kc site Wasserwerk (Kc Was) and the TERENO (TER) sites Ruraue (TER Rur), Selhausen (TER Sel) and Niederzier (TER Nied).

representation in the LES. The horizontal extension of the modeling domain is 48 km×48 km. In total, the model domain is resolved by $960 \times 960 \times 144$ grid cells. Figure 1a shows the topography in a 50 km×50 km domain around the central HOPE region. Apart from the Eifel mountain range in the south west of the region, the domain is rather flat which is reflected by using a flat homogeneous surface in the LES.

5   As explained in Sect. 2.4 and appendix A, the larger-scale forcing data is constructed by averaging COSMO analysis data. The center of the averaging domain is located at 6.375°E and 50.875°N which is centered in the HOPE-region (see Sect. 3). The larger-scale forcing data is averaged over a domain with the size of 2.0°×2.0° on the geographical grid (80×80 COSMO grid points) to eliminate small scale fluctuations. This corresponds to a zonal and meridional extension of the averaging domain of 140 km and 222 km, respectively. The latitude is set to $\phi = 50.92°$ to define the Coriolis parameter for the HOPE region. At

10   the surface, temperature and humidity are prescribed horizontally homogeneous (Dirichlet conditions). The roughness length z_0 for momentum is adopted from the averaged COSMO data and, thus, depends on the chosen averaging domain. It results in a value of $z_0 = 0.4493$ m for the chosen 2.0°×2.0° averaging domain. The roughness length for scalars is usually smaller than that for momentum (Brutsaert, 1975) and chosen to be $0.1 \cdot z_0$. The surface sensible and latent heat flux are then calculated locally by means of Monin-Obukhov similarity theory. By constructing the forcing data-set as described, it is assumed to be

15   representative for the HOPE area.

Note that the LES are run without radiation (neither interactive nor prescribed). Radiation is neglected as the radiative cooling rates are usually an order of magnitude smaller than the heating rates from the surface heat flux in the mixed layer (Stull, 1988). However, through the use of nudging the effect of radiation can be regarded as indirectly accounted for.

2.4 Relative importance of larger-scale forcing terms

The impact of the larger-scale forcing terms on the numerical solution is evaluated and quantified. For that purpose the budget terms of the prognostic equations for liquid water potential temperature θ_l and total water ~~specific humidity~~ mixing ratio q_t of the simulation *RP* are compared. Figure 2 shows the single tendency terms which were horizontally and also vertically averaged. The vertical average is taken between the surface and the ~~height~~ depth of the boundary layer z_i (in case $z_i < 500\,\mathrm{m}$, the upper limit for the averaging is 500 m to obtain also meaningful information during night times (robust statistics) where the boundary layer is resolved by a few grid points only).

It is apparent, that during daytime the fast physics have the largest impact on the numerical solution on most of the days. The impact of the different larger-scale forcing terms are comparably small. Sometimes (e.g., 26 April, 5 May, 11 May) the larger-scale forcing terms are of opposite sign also. A clear exception is 26 April, on which a frontal passage occurs (see Sect. 3). Here, the fast physics have almost no impact on the numerical solution, and the larger-scale forcing terms dominate the change of θ_l and q_t inside the boundary layer. Before noon on 26 April the LSA and SUB tendencies heat the boundary layer, and then the LSA tendencies cause a rapid and strong cooling. However, judging from the nudging tendencies for θ_l, this cooling should begin some hours earlier. This circumstance may be caused by the low temporal resolution of the forcing data (3 h intervals). As the nudging tendencies are corrective tendencies, they are also a measure of the deviation between the states of COSMO and the LES. Since the nudging tendencies are generally smaller than the LSA and SUB tendencies, the latter are a sufficient representation of larger-scale physics. However, days with strong larger-scale forcing usually show slightly larger nudging tendencies.

During nighttimes all the tendencies are equally important. The grid-spacing of 50 m used in the LES is much too coarse to resolve processes in the stable, nocturnal boundary layer. Thus, the larger-scale forcing terms from COSMO keep the LES in check in the night. However, van Stratum and Stevens (2015) suggest that the influence of biases in the representation of the nocturnal boundary layer do not substantially influence the subsequent daytime development.

3 HOPE

HOPE took place near Jülich (located in the western part of Germany) in April and May 2013. The agricultural area around the permanent observational site JOYCE (Jülich ObservatorY for Cloud Evolution (JO) at 50.907° N / 6.414° E / 111 m AMSL, Löhnert et al., 2015) was chosen to employ various in-situ and remote sensing instruments to capture a most complete set of atmospheric parameters at a high temporal and spatial resolution. JOYCE was complemented by two additional measurement sites, LACROS (Leipzig Aerosol and Cloud Remote Observations System (LA) at 50.880° N / 6.415° E / 99 m AMSL, Bühl et al., 2013) and the KITcube (Karlsruhe advanced mobile observation platform (Kc) at 50.897° N / 6.464° N / 110 m AMSL,
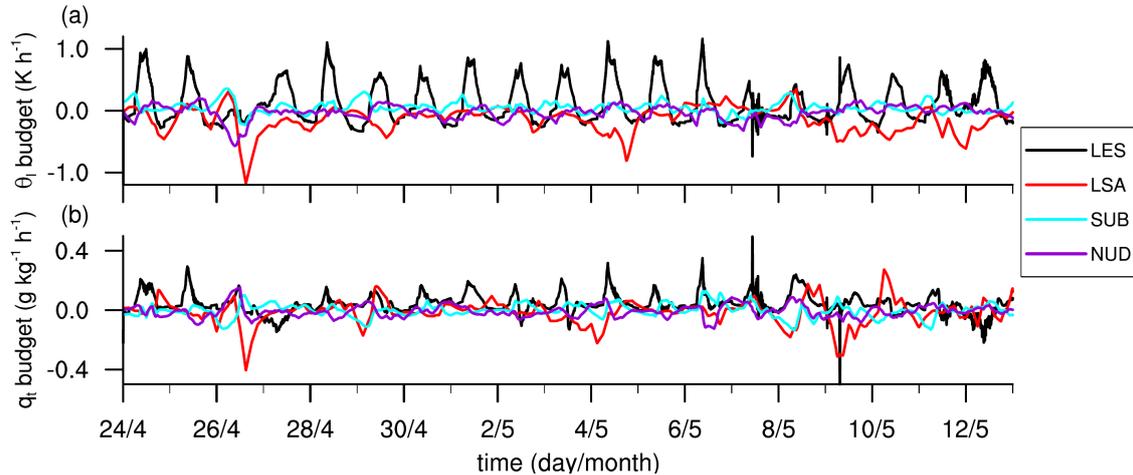
Figure 2. Temporal evolution of the vertically averaged budget terms of liquid water potential temperature (panel a) and total water ~~specific humidity~~ mixing ratio (panel b) of case RP. The black lines (LES) show the sum of fast LES physics (advective, subgrid diffusive and microphysical) tendencies, the red lines (LSA) denote the tendencies due to larger-scale horizontal advection, the cyan lines (SUB) show the larger-scale subsidence tendencies and the ~~orange~~ violet lines (NUD) denote the nudging tendencies. The vertical average is taken between the surface and the ~~height~~ depth of the boundary layer z_i (in case $z_i < 500\,\mathrm{m}$, the upper limit for the averaging is 500 m).

Kalthoff et al., 2013) during the HOPE period. The locations of the three sites and Jülich are shown in Fig. 1. Additional surface flux measurements used in this study were obtained at KITcube (Kc) site Wasserwerk and the three TERENO (TERrestrial Network of Observations (TER), Zacharias et al., 2011) sites Selhausen, Ruraue and Niederzier (see Fig. 1b), where energy balance stations were located. Within a 50 km × 50 km domain centered around JOYCE, the Eifel mountain range is located
5   south-west of the HOPE domain (see Fig. 1a). The most significant orographic element in the area around the HOPE sites is the Sophienhöhe (which can be seen in the upper right part of Fig. 1b) with a maximum altitude of 301 m AMSL. This element results from an open-pit mine located east of Sophienhöhe. The measurements during HOPE were taken by a multitude of instruments, such as Doppler lidars, Raman lidars, differential absorption lidar, ceilometers, microwave radiometers, cloud Doppler radars, meteorological towers, eddy-covariance stations and radiosondes. However, only a selection of these
10   measurements are actually used in this study, as the main emphasis is put on boundary layer characteristics and turbulence.

    The 19-day period from 24 April to 12 May 2013 was chosen for the following reasons. This period contains different weather regimes (clear-sky, convective, cloudy, frontal and post-frontal situations). Furthermore, during this time-span 7 of the 18 conducted intensive observation periods (IOPs) took place in which the temporal coverage of measurements was higher (e.g., radiosondes were launched every 2 h during day time). Moreover, it covers the passage of a frontal system, i.e., an event
15   which is strongly controlled by fast changing larger-scale flow conditions. The frontal passage allows to study how the LES models react to such forcings. The selected period is too short for a feasible statistical analysis as conducted by Schalkwijk et al. (2015), but it is long enough to showcase and analyze the general capability to perform long-term LES.

The synoptic conditions during the 19-day period can be grouped into four different periods. During the first two days (24-25 April) high-pressure was dominating the HOPE-area resulting in a calm clear-sky (24 April) and a shallow cumulus (25 April) day. On 26 April the situation changed noticeable as a frontal system passed from north-westerly directions over the HOPE-domain accompanied by an overcast, rainy situation and followed by three days (27-29 April) under post-frontal, overcast conditions where temperatures were significantly lower than before. The third period covering 30 April to 6 May was characterized by a calm, high-pressure period with mostly low- to mid-level convective clouds (where 3 May and 4 May were even clear-sky days). The last period began on 7 May with strong convective events (local thunderstorms). The following days were determined by local troughs of low-pressure-systems forming over England resulting in a rough and predominantly wet period with westerly gusts up to 14 m s^{-1}. In terms of clouds this evolution is also apparent in Fig. 4a, which displays the Cloudnet target classifications (Illingworth et al., 2007) at LACROS site.

4 Reference simulation

To obtain a first visual impression of the LES data sets, snapshots of four different days (one out of each of the four weather periods previously described) of the PALM reference simulation *RP* are compared with images from the total sky imager TSI-880 (Löhnert et al., 2015) at JOYCE site. Additionally, horizontally averaged mean profiles of potential temperature θ, ~~specific humidity~~ mixing ratio q_v from PALM and radiosondes launched at 11 UTC at KITcube site together with simulated cloud (q_c) and rain (q_r) water ~~specific humidity~~ mixing ratio (if present), are shown in Fig. 3. The snapshots were taken at 11 UTC on each day corresponding to the launching time of the radiosondes. The visualization of simulated cloud fields, which was performed with the Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers (VAPOR, Clyne et al., 2007), allows for a first impression about the diversity of weather conditions encountered in the simulations.

Comparing sky imager and volume rendered cloud fields of the four days visually (left and middle columns of Fig. 3), it can be noted that the simulated cloud types agree qualitatively with the observed ones. The three-layer vertical structure in the boundary layer on 24 April is principally reproduced by PALM (Fig. 3c). However, the potential temperature is about 2 K lower than measured in the well-mixed layer and up to 1 K lower above. On 26 April, the day where the front passes the HOPE region, a significant amount of clouds and precipitation is simulated at 11 UTC (Fig. 3f). The temperature profile of PALM is reproduced very well. However, PALM simulates a well-mixed humidity layer below 1.5 km which is not seen in the sounding. Similar to 24 April, the boundary layer and lower tropospheric layer are about 1 to 2 K colder than observed on 5 May (Fig. 3i) and also on 10 May (Fig. 3l). The vertical structure is reproduced well on both latter days.

4.1 Temporal evolution

4.1.1 Principal character of the simulated days

To provide an overview we first show how well the principal character of the day in terms of clouds and precipitation is represented in the LES over the course of the 19-day period. A qualitative comparison of cloud water and cloud rain produced

Table 1. Overview of the HOPE measurements used in this study.

Variables	Explanation	Device	Location	Time span	Referen
z_i	boundary layer ~~height~~ depth	Doppler lidar HALO	JOYCE	04/24-05/12	Schweer
		Raman lidar Polly^{XT}	LACROS	04/24-05/12	~~Althaus~~
					Althaus
					Engelma
		radiosondes (Graw DFM-09)	KITcube	04/24-05/12	Kalthoff
cc	cloud cover	Total Sky Imager TSI-880	JOYCE	04/24, 04/26, 05/05, 05/10	Löhnert
IWV, LWP	integrated water vapor, liquid water path	micro wave radiometer HATPRO	JOYCE	04/24-05/12	Löhnert Steinke
shf, lhf	surface sensible heat flux, surface latent heat flux	energy balance stations	KITcube Kc Wasserwerk TER Selhausen TER Niederzier TER Ruraue	04/24-05/12	Kalthoff Maurer Graf et a Zacharia
$\overline{w'^2}$	vertical velocity variance	Doppler lidar WLS7-V2 (~~$z < 400\,\mathrm{m}$~~ $z < 400$ m), Doppler lidar WindTracer WTX (~~$z \geq 400\,\mathrm{m}$~~ $z > 400$ m)	KITcube	04/24, 05/05	Maurer
$\overline{T'^2}$	temperature variance	rotational Raman lidar RRL	KITcube	04/24, 05/05	Behrend
$\overline{\rho_v'^2}$	absolute humidity variance	water vapor differential absorption lidar WVDIAL	KITcube	04/24, 05/05	Muppa e
z_{cb}, z_{ct}, d_c	cloud base height, cloud top height, cloud layer depth	Cloudnet Cloudnet ceilometer CHM15k	JOYCE LACROS JOYCE	04/24-05/12	Illingwo Illingwo Löhnert
T_{2m}, ~~$\|v\|_{h,120\,\mathrm{m}}$~~ $\|v\|_{h,120m}$, ~~wdir$_{120\,\mathrm{m}}$~~ wdir$_{120m}$	temperature at 2 m, wind speed at 120 m, wind direction at 120 m	~~120 m meterological~~ 120 m meteorological tower	JOYCE	04/24-05/12	Löhnert

Most of the data sets are available via the HD(CP)^2 data portal at https://icdc.zmaw.de/hdcp2.html.

by the LES with the Cloudnet product (Illingworth et al., 2007) at LACROS site complemented with the weather overview archive produced during HOPE, is presented. Note that Cloudnet is a composite measurement product, which is derived from ceilometer, cloud radar, microwave radiometer and output from the COSMO model (e.g., Löhnert et al., 2015).

Figure 4a shows the Cloudnet target classification at LACROS. Roughly, the period consists of two clear-sky days (24 April and 4 May), nine predominantly cloudy days (25, 28, 29 and 30 April, 1, 2, 3, 5 and 6 May) and eight days where precipitation occurred (26 and 27 April, 7, 8, 9, 10, 11 and 12 May). Applying the same qualitative criteria (clear-sky, cloudy, rainy) to the PALM and UCLA-LES representation of clouds and precipitation in terms of cloud and rain water ~~specific humidity~~ mixing ratio (Fig. 4c,d), the following summary can be given (see also Tab. 2). On two days (25 April, 1 May), both LES models were not able to simulate shallow cumuli during the day although shallow cumuli were observed. Precipitation was simulated on too few days. UCLA-LES did not simulate precipitation on three days and PALM on one day. This sums up to a qualitative agreement in the principal character of the day on 16 days for PALM and 14 days for UCLA-LES, which is an agreement of 84 % and 74 %, respectively.

Comparing specific cloud and rain water mixing ratios of the two LES models with the COSMO forcing (Fig. 4b) we want to stress that only a warm-rain microphysics scheme has been applied in both models. This clearly restricts the possibility to realistically form upper-level clouds and precipitation in the LES as these processes usually require the ice-phase in mid-latitudes. Nonetheless, PALM and UCLA-LES both find a representation of higher level clouds, especially on days with strong impact of larger-scale forcing like the frontal day of 26 April. The shallow cloud layers usually form on top of the boundary layer as can be seen in Fig. 4c,d. These cloud layers usually find a good representation when using a warm-microphysics scheme only. ~~Anyhow, the more challenging situations for PALM and UCLA-LES seem to be simulations~~ However, the simulation of proper shallow cumulus layers on some days (25 April and 1 May) is a challenge for PALM and UCLA-LES.

The cloud and precipitation structure over the 19 days is very similar in both models, but UCLA-LES produces a lesser amount of cloud and rain water (the latter leading to two more days of qualitative misrepresentation in UCLA-LES as compared to PALM, see Tab. 2). This difference roots in the usage of different advection schemes for scalars as PALM uses a fifth-order scheme whereas UCLA-LES applies a monotone second-order scheme with flux-limiter (see Sect. 2.1). Monotone schemes show a rather diffusive character (Durran, 1999). Thus, the horizontal and vertical gradients are smoothed more strongly in UCLA-LES than in PALM which could even lead to a complete damping of small amplitudes of humidity and updrafts prohibiting formation of weak clouds and precipitation. Furthermore, the specific rain water is slightly better represented in the LES than in COSMO. COSMO shows much more rain than observed.

4.1.2 Boundary layer depth

The boundary layer depth z_i is one of the major defining characteristics of the boundary layer. In this study, z_i has to be determined for different types of boundary layers (stable, convective, cloud-topped) and the respective transitional phases, because several diurnal cycles are simulated. Therefore, a robust criterion, that works well for the different boundary layer types, has to be chosen for an adequate determination of z_i. However, most established methods are closely tied to one boundary layer type (e.g., the height of the minimum buoyancy flux for the convective boundary layer). Thinking of a broader definition

Table 2. Summary on qualitative agreement in the principal character of the simulated days compared to Cloudnet and HOPE weather overview archive.

Criteria	PALM days	no. days	UCLA-LES days	no. days
LES days without shallow cumuli, when shallow cumuli were observed	25/04, 01/05	2	25/04, 01/05	2
LES days without rain, when rain was observed	08/05	1	26/04, 08/05, 09/05	3
LES days with qualitative agreement to observations	remaining	16	remaining	14

of the boundary layer, it can be identified as the layer in which turbulent mixing occurs due to the presence of the surface. The dimensionless Richardson number Ri is defined as the ratio of buoyancy to shear production of turbulence kinetic energy. The boundary layer depth can also be defined as the height where Ri exceeds a critical value as Ri provides a measure of the dynamic stability of the flow. Criteria based on Ri have been frequently used in a number of studies over the last decades (e.g., Richardson et al., 2013, and references therein). The bulk Richardson number Ri_b is derived from the gradient Richardson number by approximating local gradients to a finite difference across a layer and it is defined as

$$Ri_b = \left(\frac{g}{\theta_{v,s}} \right) \frac{\theta_v - \theta_{v,s}}{u_1^2 + u_2^2} \cdot z, \tag{5}$$

where g is the acceleration due to gravity, θ_v denotes the virtual potential temperature and $\theta_{v,s}$ its value close to the surface. Following classical theory (Taylor, 1931), turbulence of a homogeneous stably stratified sheared flow in steady state decays, if the gradient Richardson number exceeds a value of 0.25. In the definition (Eq. (5)), Ri_b is defined from the surface upwards. If z is replaced by the boundary layer ~~height~~ depth z_i, Ri_b becomes the critical bulk Richardson number whose value depends on stability (e.g., Richardson et al., 2013; Basu et al., 2014). However, this dependence is neglected in this study and a value of $Ri_{b,c} = 0.25$ is assumed to be valid for all stability regimes. This applied value also lies in the interval for the critical bulk Richardson number $0.2 < Ri_{b,c} < 0.5$, proposed by Zilitinkevich and Baklanov (2002).

In PALM and UCLA-LES, z_i is determined locally (at each grid point in the horizontal domain). Starting at the lowest prognostic level and continuing upwards, Ri_b is calculated using Eq. (5) until $Ri_b > Ri_{b,c} = 0.25$. The height of the grid point at which the critical value is exceeded is then assumed to coincide with z_i. For $\theta_{v,s}$ the second prognostic level above the surface is used. The resulting 2D field of z_i is then averaged horizontally and the horizontal variability is quantified by means of retaining the standard deviation.

Figure 5 shows the temporal evolution of the boundary layer depth. The aforementioned spatial variability is depicted as twice the standard-deviation in light gray shading for PALM and light green shading for UCLA-LES in Fig. 5. It is strongest during day-time. The bulk Richardson number criterion is also applied to the mean COSMO profiles and the resulting z_i is shown as blue dashed line. The LES models produce a very similar boundary layer depth. Both models are lagging behind COSMO. As the LES are tied to the COSMO forcing, they also show peak heights close to COSMO. This behavior can partly be attributed to the Newtonian relaxation which pulls the LES back towards the mean state given by the forcing.

Measurements from the three different major HOPE sites are taken into account for evaluating the performance of the LES models in terms of the boundary layer depth. The aerosol Raman lidar Polly^{XT} (Althausen et al., 2009, Polly hereafter) at LACROS site provides an estimate for the boundary layer depth based on the heights where the detected aerosols show a strong back-scatter signal (Baars et al., 2008). The Doppler wind lidar HALO ~~(Schween et al., 2014)~~ provides profiles of vertical velocity variance at the JOYCE-site from which the boundary layer depth is deduced as the lowest height from the surface onwards where the vertical velocity variance is smaller than a threshold of 0.4 m^2 s^{-2} (Schween et al., 2014). As third data source 78 radio-soundings from KITcube site were used. The bulk Richardson number method was applied to the available soundings. In analyzing the soundings erroneous values near the surface were detected so the critical Ri$_b$ is calculated from 100 m onwards. The applied criteria to the lidar data (vertical velocity variance and aerosol layer) are not boundary layer regime independent and usually work best for convective boundary layer situations. ~~Anyhow~~However, they are a standard measurement product and the independent measurements of z_i used in this study provide a general corridor for a representative boundary layer depth observed in the HOPE domain. Due to the different methods used to deduce the boundary layer depth, the aerosol lidar typically shows larger depths than the wind lidar (see Fig. 5) as the detected aerosol layers are a passive tracer for the boundary layer depth as compared to the dynamic criterion based on vertical velocity variance.

On most days, PALM, UCLA-LES and COSMO are able to reproduce the development of the boundary layer as the models lie inside the spread of the measurements resulting from surface heterogeneity and spatial variability of the boundary layer depth between the three sites. On days with strong vertical forcing, which are ~~marked in light yellow colors~~ stippled in Fig. 5, the simulated peak depths agree less well with the observations. A day is characterized as a day with strong vertical forcing in case the prescribed larger-scale subsidence velocity averaged between 4 km and 8 km, denoted as $\widetilde{w_{SUB}}$ in the following, is larger than ~~50~~5 cm s^{-1}.

On 25 April, a day where shallow cumulus was observed but not simulated (see Tab. 2), the peak height is strongly underestimated by the LES, but also by the host model COSMO. Overall, the daily development of the boundary layer depth can be qualitatively reproduced by both LES models.

4.1.3 Further boundary layer quantities

~~Figure ?? gives~~ Figures 6-8 give a further overview about the performance of the LES for boundary layer quantities like the surface sensible and latent heat fluxes, near surface wind direction, wind speed and potential temperature and the integrated water vapor IWV and liquid water path LWP. For the LES, the horizontal mean of the quantities is shown. At first glance, general agreement with observations is given. PALM and UCLA-LES are nearly indistinguishable apart from LWP. They are also rather close to COSMO

In the reference setup, potential temperature and humidity from the COSMO averaging box are prescribed homogeneously at the surface and the sensible and latent heat fluxes (shf and lhf) are calculated locally via Monin-Obukhov similiarity theory (see Sect. 2.3). These surface fluxes are very important as they directly determine the amount of energy input into the boundary layer. In Fig. ~~??a,b~~ 6 the surface fluxes from PALM and UCLA-LES are compared with the fluxes from the COSMO forcing and measurements from different energy balance stations. A total of five different stations located over different land-use

14

classes in close vicinity to the principal HOPE sites are taken into account (see Fig. 1b). From these measurements spatially representative values of the surface fluxes are derived and provided by Maurer et al. (2016). A weighted average (w.av.) of the five stations with the fraction of the respective land-use class in an area of 30 km × 30 km centered around KITcube site is calculated (see Maurer et al., 2016, for further details). ~~In panels a and b, the~~ The weighted average is marked by purple

5 stars in Fig. 6. The fluxes at the individual stations show a considerable spread reflecting the large spatial variability for surface fluxes (heterogeneity) in the HOPE region. By construction, shf and lhf in the LES are closely tied to the surfaces fluxes in the forcing and are also slightly lagging behind like the boundary layer depth. They agree roughly with the weighted average on most days. The peak shf in the LES and COSMO tends to be overestimated compared to the weighted average, whereas the lhf tends to be underestimated, especially for the last six days of the simulation period. Overall, the simulated surface fluxes can

10 be seen as representative for the HOPE region.

For wind-engineering purposes, surface layer winds are very important. Measurements from the 120 m-meteorological tower at JOYCE site (Löhnert et al., 2015) and radio-soundings are compared to the LES and COSMO in Fig. 7a,b. The wind components u and v were linearly interpolated between the second and third prognostic level to obtain values for 120 m. All major changes in wind direction at a height of 120 m can be reproduced very well by the two LES models and COSMO (~~panel~~

15 ~~e~~Fig. 7a). For the wind speed at 120 m height, the tower measurements and soundings show larger fluctuations than the models as the point measurements contain turbulent signals which are smoothed out in the shown horizontal mean of the LES output. Taking these differences into account, the LES agree rather well with the wind speed observations.

The near-surface potential temperature at a height of 25 m from the JOYCE tower, the radio-soundings and the LES is depicted in ~~panel e~~Fig. 7c. For the LES the output at the first prognostic level is taken. PALM, UCLA-LES and COSMO are

20 systematically too warm during the night. During day-times, the LES are usually colder (with some exceptions on 26 April, 27 April,11 May). Overall, there is good agreement with observations although the amplitudes of the observations are slightly larger.

Observations of the column integrated quantities integrated water vapor IWV and liquid water path LWP, shown in ~~panels f and g~~Fig. 8, are provided by the microwave radiometer HATPRO (Löhnert et al., 2015; Steinke et al., 2015) at JOYCE

25 site. There is good agreement for IWV between the LES, COSMO and HATPRO. Hence, the total amount of water vapor is accurately included into the LES by means of the larger-scale forcing method. The LWP (panel ~~g~~b) of the LES matches the observations better than COSMO despite the deficiency in terms of used warm microphysics. ~~Anyhow~~However, modeling LWP (which can be seen as proxy for clouds) with the long-term LES approach correctly is rather challenging.

The six days with strong vertical forcing (~~highlighted in yellow~~stippled) show all rather high values of LWP in rough

30 accordance with HATPRO. As already discussed in section 4.1.1, there are 25 April and 1 May where shallow clouds could not be simulated although they have been observed which is also apparent in ~~panel g~~Fig. 8b. Furthermore, both LES differ more strongly compared to the previously discussed quantities as microphysics and numerics are closely tied and they are very important for allowing cloud formation in the LES.

4.2 Vertical structure

The main strength of LES is to resolve turbulence. To assess whether the long-term LES approach is able to produce realistic turbulence statistics, variance profiles for two distinct situations are discussed. The variances of vertical velocity $\overline{w'}^2$, potential temperature $\overline{\theta'}^2$ and ~~specific humidity~~ mixing ratio $\overline{q_v'}^2$ from PALM and UCLA-LES are compared to variance profiles from

5   lidars located at KITcube site for a one hour period (11-12 UTC) for the clear-sky situation of 24 April and the shallow cumulus situation of 5 May. Three different lidars, namely the Doppler lidar WindTracer WTX combined with the Doppler lidar WLS7 (Maurer et al., 2016) from KIT, the rotational Raman lidar RRL ~~(Behrendt et al., 2015)~~ (Hammann et al., 2015; Behrendt et al., 2015) and the water vapor differential absorption lidar WVDIAL (Muppa et al., 2016) from University of Hohenheim were operated simultaneously during IOPs of HOPE, allows us to compare different lidar-based higher order moments with the LES to discuss

10   the turbulence structure of the boundary layer on these two days.

Figure 9 shows the vertical velocity, potential temperature and ~~specific humidity~~ mixing ratio variances for 24 April and 5 May (11-12 UTC). For 24 April, the lidar-based variances (solid purple lines) of the vertical velocity, the actual temperature and the absolute humidity were each recently published by Maurer et al. (2016); Behrendt et al. (2015) and Muppa et al. (2016). They also provide data for the cumulus-topped boundary layer of 5 May, which is analyzed for the first time in the present

15   paper. The lidar turbulence signal at each height is calculated by subtracting the linear fit of the recorded time-series between 11 and 12 UTC from the original time-series. Based on this turbulence time-series, the variance for each record is calculated (see, e.g., Behrendt et al., 2015). Note that the actual temperature variance as given by RRL was converted to potential temperature variance assuming a constant Exner function which was taken from the radio-sounding profile at 11 UTC of the respective day. The absolute humidity variance was converted similarly by means of the air density taken from the same sounding. In

20   the cumulus case (5 May), the data points inside cloudy regions are not taken into account for the estimation of higher-order moments with RRL and WVDIAL. Furthermore, the potential temperature variance of RRL is only shown up to a height of $0.7z_i$ which is near cloud base (see panel e) as the cloud layer is affected by saturation of the detector. In this case more noise is found in the data and overlaps the true data thoroughly making the measurements less reliable.

Typically, higher-order moments from LES are deduced from a spatial (horizontal) average (e.g., Heinze et al., 2015) as

25   opposed to lidar measurements which define turbulence as departure from a temporal mean. To account for this difference, variances from LES are shown in two different ways in Fig. 9. The solid black and green line denote the one-hour average of the variances as defined by the departure from the horizontal mean (*hom*). Solid gray and light green areas show twice the standard deviation resulting from the one-hour average of the slab-averaged variance profiles. Furthermore, virtual measurements were conducted in the LES at four distinct locations which are equally spaced in the modeling domain. Grid-point data for four

30   independent columns (*colX*) with a high temporal resolution (30 s and 5 min for PALM and UCLA-LES, respectively) have been saved. These time-series were used to calculate variances exactly as for the lidar data (detrending and temporal average over one hour). These variance profiles are representative for a single-measurement inside the LES and are thus directly comparable to the variances deduced from lidar. They are depicted as thin dashed black and green lines in Fig. 9.

Table 3. Scaling values for 24 April 2013 and 5 May 2013 for 11-12 UTC (used in Fig. 9).

		PALM	UCLA-LES	Lidar (Kc)
24 April 2013	z_i (m)	1033	1091	1312
11-12 UTC	$\overline{w'\theta'_{v_s}}$ (W m^{-2})	285.3	292.1	163.1
	$\overline{w'q'_{v_s}}$ (W m^{-2})	168.3	156.0	129.6
	w^* (m s^{-1})	2.022	2.075	1.810
	θ^* (K)	0.117	0.117	0.075
	q^* (g kg^{-1})	0.028	0.025	0.060
5 May 2013	z_i (m)	1641	1465	1723
11-12 UTC	$\overline{w'\theta'_{v_s}}$ (W m^{-2})	181.3	202.1	185.2
	$\overline{w'q'_{v_s}}$ (W m^{-2})	160.9	140.6	127.5
	w^* (m s^{-1})	2.037	2.031	2.053
	θ^* (K)	0.076	0.084	0.077
	q^* (g kg^{-1})	0.027	0.023	0.052

Values are averaged over one hour (11-12 UTC) on both days. Boundary layer depth z_i in the LES is determined based on the bulk-Richardson number criterion. For the lidar, z_i is the top of the aerosol layer based on backscatter signal. Surface buoyancy and latent heat fluxes, $\overline{w'\theta'_{v_s}}$ and $\overline{w'q'_{v_s}}$ respectively, are horizontally averaged values in the LES and weighted averaged values from the energy balance stations in case of lidar.

To account for a better comparison between observed and simulated variances, all profiles in Fig. 9 are scaled (non-dimensionalized) by means of the free convective Deardorff (1970a) scales. These are the convective velocity scale $w^* = \left(\frac{g}{\theta_{v,s}} \overline{w'\theta'_{v_s}} z_i \right)^{\frac{1}{3}}$, the convective temperature scale $\theta^* = \frac{\overline{w'\theta'_{v_s}}}{w^*}$ and the convective humidity scale $q^* = \frac{\overline{w'q'_{v_s}}}{w^*}$, where $\overline{w'\theta'_{v_s}}$ denotes the kinematic surface buoyancy flux and $\overline{w'q'_{v_s}}$ is the kinematic surface latent heat flux (see Tab. 3). The vertical axis (height) is normalized

5   by means of the boundary layer depth. For all lidar-derived profiles, the boundary layer depth is determined by estimating the top of the aerosol layer from lidar backscatter data (method 2 in Maurer et al., 2016). The required surface fluxes are taken from the weighted average of five different energy balance stations (see also Sect. 4.1) which is, based on Maurer et al. (2016), representative for a larger area. The LES-based scaling values are derived from the bulk-Richardson number-based z_i and the one-hour average of the horizontal mean surface buoyancy and latent heat flux. All values are summarized in Table 3.

10   Comparing the horizontal mean variances of PALM and UCLA-LES in Fig. 9 generally, we note that they both show a very similar vertical structure. In all six cases variances from PALM are slightly larger than variances from UCLA-LES which becomes most prominent for the peak values of the scalar variances at the top of the boundary layer (panels b, c, e and f). The differences in variances between PALM and UCLA-LES are in the same order as discussed in several LES intercomparison studies (e.g. Stevens et al., 2001; Siebesma et al., 2003; Stevens et al., 2005). It can be attributed to different numerics like the

15   advection scheme. As UCLA-LES uses a monontone scheme for the scalars and PALM not (see also Sect. 2.1), fluctuations are damped more strongly resulting in slightly less variances (turbulence).

Table 4. Simulated and observed cloud boundaries on 5 May 2013 for 11-12 UTC.

	PALM	UCLA-LES	Cloudnet (LA)	Cloudnet (JO)	~~ceilometer~~ Ceilometer (JO)
z_{cb} (m)	1333 ± 40	1294 ± 43	1464 ± 112	1546 ± 178	1365 ± 49
z_{ct} (m)	1721 ± 83	1713 ± 62	1594 ± 118	1735 ± 196	1526 ± 56
d_c (m)	388 ± 58	419 ± 33	133 ± 60	189 ± 146	171 ± 53
N	13	13	33	21	125

Values include mean and standard deviation over 11-12 UTC. Cloud boundaries in LES are determined based on horizontally averaged profiles of cloud liquid water. Cloud base height, cloud top height and cloud layer depth are denoted by z_{cb}, z_{ct} and d_c respectively. The number of samples entering the averaging period is N. See Tab. 1 for an overview of the observations used.

On 24 April around noon, the boundary layer is cloud-free, well mixed and topped by a capping inversion as seen by radiosounding profiles in Fig. 3c. The LES reproduce this structure which manifests also in the variance profiles (Fig. 9a-c). The LES-based vertical velocity variances reveal the typical peak around $0.3z_i$ and decrease monotonically above. The vertical velocity variance from Doppler lidar exhibits a maximum at around $0.5z_i$ and shows a second smaller peak around $0.9z_i$. A
5   longer averaging period of about 3 h would lead to a decrease of the height of the lower maximum to about $0.3z_i$ (Maurer et al., 2016) which emphasizes that the chosen averaging time might be too small to receive robust $\overline{w'}^2$ statistics comparable to LES. This is confirmed by the large differences between the given virtual measurements. The horizontal mean profiles and to a larger extent also the virtual measurements are inside the uncertainty range of the Doppler lidar. Nonetheless, it should be kept in mind that a departure of the horizontally averaged LES variances from the lidar variances does not necessarily mean
10   that the LES variances are not representative as the statistical error based on Lenschow et al. (1994) does not always show how large the uncertainties really are - especially in case of heterogeneous surfaces (Sühring and Raasch, 2013).

The LES-based scalar variances show their distinct maxima on 24 April at the top of the boundary layer (Fig. 9b,c) where warmer and less humid tropospheric air is entrained producing large turbulent fluctuations. This is principally in accordance with the lidar measurements. The peak values of the lidar-based scalar variances are significantly higher than the ones of the
15   LES - even when taking the virtual measurements in the LES models into account. Here, it becomes apparent that the vertical grid-spacing of 50 m used in LES is much too coarse to sufficiently resolve the strong vertical gradients at the boundary layer top. Recently, it was demonstrated that entrainment processes have an important influence on the structure of variance profiles and should be accounted for (Wulfmeyer et al., 2016). Another reason for the underestimation of scalar variance peak values might also be the usage of homogeneous surface forcing which allows only to prescribe surface forcing representative for
20   the larger area which might not necessarily be similar to the forcing actually present at the measurement site. The ~~specific humidity~~ mixing ratio variance from WVDIAL shows a rather unusual lower peak at around $0.85z_i$ (panel c) which Muppa et al. (2016) associate with entrainment of an elevated humidity layer into the convective boundary layer. The second peak in vertical velocity variance at around $0.9z_i$ might be also associated with this event.

On 5 May a shallow cumulus layer was observed at JOYCE site and simulated around noon (see Fig. 3g,h). The mean profiles
25   of potential temperature and ~~specific humidity~~ mixing ratio of PALM and the radiosoundings barely show the existence of the

cloud layer as it is rather shallow. Table 4 provides an overview of the observed and simulated cloud boundaries between 11 and 12 UTC. An average of Cloudnet observations at LACROS and JOYCE and a ceilometer at JOYCE site results in a 156 m deep layer. The cloud layer in both simulations is about 2.5 times deeper with about 388 m for PALM and 419 m for UCLA-LES. The LES are expected to show deeper cloud layers as the maximum height of a sampled cloud in the domain determines

5 the depth whereas the measurements sample at one point only. Both LES simulate a total cloud cover during noon that is not higher than 5 % (not shown) and also the LWP does not show a significant signal (see Fig. ??g8b) supporting the finding of a very weak shallow cumulus layer in the models. The cloud boundaries are also depicted in Fig. 9d,e,f as gray and green dashed layer for the LES. The cloud boundaries from observations at KITcube are not shown as it was not possible to reliably estimate them from the lidars at KITcube site. There have been only 4 tiny clouds passing the lidars during the one-hour period (not

10 shown). Note that the cloud layers are also scaled which might lead to a different impression while comparing the thicknesses.

The variances on 5 May also show no distinct feature of a well developed cumulus layer on top of a well-mixed sub-cloud layer in the LES as well as in the observations. Their shapes resemble strongly those of the variances in the cloud-free convective boundary layer discussed before. For the vertical velocity variance, the LES horizontal mean as well as most of the virtual measurements are close to the uncertainty range of the lidar showing also a similar shape as the lidar. The potential

15 temperature variance can only be compared below $0.7z_i$ as it is not available from RRL higher above. LES and lidar both show low variances in the well-mixed part of the boundary layer. The maximum of specific humidity mixing ratio variance is located slightly higher than those of the LES.

Overall, the long-term LES approach is able to deliver variance (turbulence) profiles that are in a satisfactory agreement with lidar observations.

20 5   Sensitivities

To study how robust the previously discussed results are with respect to the chosen setup, the reference simulations *RP* and *RU* were complemented by 14 additional simulations with PALM. Table 5 lists the simulations with their differences in the setups relative to the setup *RP/RU*, which was described in Sect. 2.3. Most of these additional simulations were run on a smaller horizontal domain (4.8×4.8 km^2 instead of 48×48 km^2, denoted with a capital S in Tab. 5) and for the first three days only

25 (24-26 April) for the sake of computational resources. (Note that *RP* and *RU* ran on 2000 cores for around 7 and 10 days, respectively).The period 24-26 April.) The period 24-26 April was chosen as it contains three different boundary layer states (clear-sky, shallow clouds, frontal passage) in a row, being a condensed representative of the longer period.

To compare all the experiments a metric based on the boundary layer depth (see Fig. 5) is constructed. As z_i is a central quantity for evaluating mean boundary layer characteristics, it is chosen as basis for the metric. For each available value, the

30 absolute difference in boundary layer depth of PALM between the host model COSMO, the aerosol lidar Polly and the wind lidar HALO, respectively, are calculated. Then, an average over the number of available daily time-spans from 12-14 UTC (either 19 or 3 depending on the case) is taken and the standard deviation is provided accordingly. This metric is called *mean peak difference to PALM* in the following. A daily averaging time-span of two hours (12-14 UTC) was chosen to consider the

Table 5. Parameters of the simulated cases.

Case	LES model	Δ	$L_1 \times L_2$	$N_1 \times N_2 \times N_3$	z_0	t_{sim}	τ	L_{COSMO}	$\mathcal{T}_{\text{COSMO}}$	~~surface~~ Surface BC
		(m)	(km)		(m)	(d)	(h)	(°)	(h)	
RP	PALM	50	48×48	$960 \times 960 \times 144$	0.45	19	6	2.0	3	prescr. θ and q_v
RU	UCLA	50	48×48	$960 \times 960 \times 144$	0.45	19	6	2.0	3	prescr. θ and q_v
RPS	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	6	2.0	3	prescr. θ and q_v
RUS	UCLA	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	6	2.0	3	prescr. θ and q_v
F0.25	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.27	3	6	0.25	3	prescr. θ and q_v
F0.5	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.31	3	6	0.5	3	prescr. θ and q_v
F1.0	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.41	3	6	1.0	3	prescr. θ and q_v
F3.0	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.44	3	6	3.0	3	prescr. θ and q_v
F4.0	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.40	3	6	4.0	3	prescr. θ and q_v
TR1	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	3	6	2.0	1	prescr. θ and q_v
Nno	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	∞	2.0	3	prescr. θ and q_v
N1	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	1	2.0	3	prescr. θ and q_v
N12	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	12	2.0	3	prescr. θ and q_v
FLX	PALM	50	4.8×4.8	$96 \times 96 \times 144$	0.45	19	6	2.0	3	prescr. fluxes
RPS12.5	PALM	12.5	4.8×4.8	$384 \times 384 \times 480$	0.45	3	6	2.0	3	prescr. θ and q_v
RPS25	PALM	25	4.8×4.8	$192 \times 192 \times 266$	0.45	3	6	2.0	3	prescr. θ and q_v
RPS100	PALM	100	4.8×4.8	$48 \times 48 \times 84$	0.45	3	6	2.0	3	prescr. θ and q_v

Δ denotes the grid spacing, L_1, L_2 are the model domain sizes in x_1 and x_2 directions, respectively, N_1, N_2 and N_3 are the number of grid points in x_1, x_2 and x_3 directions, respectively, z_0 is the roughness length for momentum, t_{sim} is the simulation time, τ is the relaxation time-scale, L_{COSMO} is the averaging domain size of the larger-scale forcing data (given in degrees on the geographical grid), $\mathcal{T}_{\text{COSMO}}$ is the temporal resolution of the larger-scale forcing data. The abbreviations *surface BC* and *prescr.* stand for for surface boundary conditions and prescribed, respectively.

state of a well developed boundary layer in a quasi-steady period. Figure 10 shows the mean peak difference in boundary layer depth to PALM for all the additional simulations. At a first glance it can be noted that the mean peak differences to PALM of COSMO, Polly and HALO in most cases show the same behavior. The metric based on wind lidar HALO usually shows the highest and positive values meaning that the peak boundary layer depth of PALM is usually higher than the one measured by HALO.

Comparing the 19-days reference simulation *RP* with the 19-days simulation RPS, which was conducted on the small horizontal domain, we note that the domain size has virtually no effect on the mean peak difference to PALM (Fig. 10a, comparing cases *RP* and *RPS (19d)*). Thus, robust first-order statistics are gained even in case the domain size is significantly smaller than in the reference case. This finding suggests that the ~~internally generated (non-forced) mesoscale circulations ,~~

mescoscale circulations which can develop internally on a 50 km × 50 km domain without orography and surface heterogeneity, are not particularly important.

A fundamental parameter of the larger-scale forcing method is the averaging domain size for the applied forcing data L_{COSMO}, specified in degrees on the geographical grid (see also Appendix A). For the reference runs *RP* and *RU*, a size of $L_{COSMO} = 2.0°$ was used. To evaluate whether the size of the averaging box is appropriate to represent larger-scale processes, the simulations *F0.25*, *F0.5*, *F1.0*, *F3.0* and *F4.0* (see Tab. 5) were conducted, where the COSMO averaging domain sizes were varying from 0.25° to 4.0° which corresponds to horizontal extensions $\mathcal{D}_x \times \mathcal{D}_y$ of 17.5×27.8 km^2 to 280×444 km^2 being equivalent to averaging over 10×10 to 160×160 COSMO grid points. The averaging domain size of the COSMO forcing has a large impact on the boundary layer depth as can be seen in Fig. 10b. Especially the two smallest averaging domain sizes produce large discrepancies in peak boundary layer depth to the estimates of z_i stemming from Polly and HALO lidar. Thus, as the averaging area gets small, more mesoscale flows, which COSMO does not necessarily well represent, are sampled. Nonetheless, mean boundary layer characteristics become less sensitive if a 2.0° averaging domain size or larger is used. Cloud structures and precipitation depend more strongly on the averaging domain size of the forcing (not shown). Overall, the averaging domain should have a size which is large enough to not include mesoscale fluctutations on the one side and which is small enough to still account for a localized, representative area like the HOPE region.

The temporal resolution of the forcing data is 3 h which also includes the prescribed surface temperature and humidity and via Monin-Obukhov similiarity theory the surface fluxes. However, boundary layer time-scales are usually much shorter (the turnover time-scale is about 10 minutes around noon for the presented period). As the simulations are strongly determined by the imposed surface fluxes, the question ~~arose~~ came up whether prescribing new surface values every 3 h is too infrequent to impose the signal of a proper diurnal cycle. Thus, the simulation *TR1* was performed, where forcing data with a temporal resolution of 1 h was used. As the larger-scale horizontal and vertical advective forcing act on larger time-scales than the surface forcing, a higher temporal resolution should affect the surface fluxes most. Comparing the cases *RPS (3d)* and *TR1* shown in Fig. 10a, it can be noticed that the metrics are nearly identical. The higher temporal resolution seems to bring no additional value. Hence, it is concluded that a 3-hourly forcing data set is sufficient to impose a proper diurnal cycle in the simulations.

As nudging (Newtonian relaxation) does not represent a real physical process (Randall and Cripe, 1999), it was analyzed how crucial the results depend on the nudging time-scale and on the nudging itself. Three additional simulations were performed where a stronger nudging with $\tau = 1$ h (case *N1*), a weaker nudging with $\tau = 12$ h (case *N12*) and no nudging at all ~~($\tau \rightarrow \infty$~~($\tau \rightarrow \infty$, case *Nno*) compared to the reference nudging time-scale of 6 h were used. The simulation without nudging can also be interpreted as a simulation were the radiative forcing is completely switched off as the effect of radiation is indirectly mimicked via the relaxation (see Sect. 2.3). The mean peak difference to PALM (Fig. 10c) shows only a weak dependence for $\tau \leq 12$ h. In case Newtonian relaxation is completely turned off, the mean peak difference to PALM increases strongly. In this case PALM strongly overestimates the boundary layer depth compared to the forcing and the observations. The overall performance of the simulation becomes worse. This analysis shows that using nudging with reasonable nudging time-

scales of several hours is beneficial for the long-term LES framework. Furthermore, the mean boundary layer characteristics barely depend on the actual choice of the nudging time-scale supporting the robustness of the setup.

To test the impact of the individual larger-scale forcing components, several tests were made in which the forcing components were mutually switched off and then added one after the other (not shown). These tests suggested that all components should be used in combination for obtaining the best results with respect to the observations. This is in agreement with the single-column model study of Sterk et al. (2015) where they studied the realistic simulation of clear-sky stable boundary layers over snow-covered surfaces.

In the reference setup, Dirichlet conditions are used at the surface meaning that potential temperature and ~~specific humidity~~ mixing ratio are prescribed at the surface. The alternative is to prescribe surface fluxes directly (using Neumann boundary conditions). The latter was used in case *FLX*. Overall, the prescribed surface fluxes are slightly smaller and show a time lag in respect to the fluxes which are calculated in case *RPS (19d)* (not shown). Comparing the cases *RPS (19d)* and *FLX* concerning the mean peak difference to PALM (Fig. 10a) it can be seen that the metric for COSMO changes only marginally and that the metric for Polly deteriorates whereas the metric for HALO improves. Taking also the arguments of Basu et al. (2008) into account that for modeling stable boundary layers prescribing surface fluxes should be avoided, we think prescribing surface values is the better option, as during the conducted multiple-day LES stable regimes are simulated to a considerable fraction.

To evaluate the influence of the numerical grid spacing, the three-day simulation *RPS (3d)* with an isotropic grid spacing $\Delta = 50$ m was rerun using two finer grid spacings ($\Delta = 25$ m called *RPS25* and $\Delta = 12.5$ m called *RPS12.5*) and one coarser grid spacing $\Delta = 100$ m called *RPS100*). Only minor differences were observed between the runs in the time-series of the boundary layer depth which mainly occur during nighttime. This indicates, that the differences between the runs are closely linked to their different capabilities of resolving the shallow stable boundary layer at night. The influence on the better resolved nighttime stable boundary layer on the following convective day is rather small as van Stratum and Stevens (2015) already showed. The simulated clouds also do not show any dependence on the grid-spacing. Figure 10d shows in terms of the mean peak difference metric that the influence of the grid spacing on mean boundary layer characteristics is negligible.

6   Summary and conclusions

In this study long-term LES with PALM and UCLA-LES are evaluated to ~~answer the question if~~ assess the ability of LES in a semi-idealized ~~setup are able to simulate boundary layer characteristics and turbulencein a realistic manner~~set up to simulate observed characteristics of boundary layer turbulence. The semi-idealized approach consists of using periodic lateral boundary conditions and a homogeneous surface ~~forcing~~ together with prescribing time-dependent larger-scale forcing and nudging deduced from the mesoscale numerical weather prediction model COSMO to account for the synoptic conditions at a specific location. A continuous period of 19 days of the HOPE measurement campaign is chosen and the simulation results are compared to the multi-sensor HOPE data set. The three principal measurement sites of HOPE ~~allow to obtain~~ enable a more representative view on the larger observational area. This circumstance facilitates the comparison to the LES which, by construction, can only deliver a flow which is representative for the HOPE region. The analysis focuses on key boundary layer

quantities like the boundary layer depth, near surface temperatures and winds, integrated quantities like IWV and LWP and turbulence statistics in terms of variance profiles. A metric based on the peak boundary layer depth is used to compare several sensitivity runs. With these additional simulations the robustness of the reference setup is investigated.

The (unphysical) nudging tendency, which prevents model drift in time, is generally less important compared to larger-scale horizontal and vertical advective tendencies. ~~In~~ The exceptions are cases with strong larger-scale forcing, ~~also~~ then the nudging tendencies ~~are~~ can be significant.

The reference simulation shows reasonable agreement with the HOPE measurements. The principal character of the day (weather situation) can be reproduced by the LES in about 80 % of the cases. Simulating cloud-topped boundary layers correctly is a challenge for the long-term LES. The daily development of the boundary layer depth is in principal agreement with lidar measurements. The LES surfaces fluxes are in a rough agreement with the weighted averaged surface fluxes in the HOPE area showing that the surface forcing is representative for the HOPE area. Both LES models used produce very similar results.

The LES models seem to track COSMO closely and deviate from the observations in a similar fashion as COSMO does. This can be interpreted in two ways. Either deviations from the observations are inherited from the host model or they represent the signature of mesoscale forcing that the present approach is incapable of capturing. By using LES in a more realistic setup with open boundary conditions these hypotheses might be tested.

LES turbulence statistics in terms of variance profiles are in a satisfactory agreement with lidar measurements during HOPE. The peak in scalar variances at the top of the boundary layer is underestimated by LES indicating that presumably the resolution used in the LES is rather coarse for correctly representing strong gradients, and that heterogeneity is missing.

The chosen semi-idealized setup is ~~rather robust and~~ insensitive to the horizontal domain size, the grid spacing, the temporal resolution of the forcing data and the surface boundary condition in terms of mean boundary layer characteristics. Thus, the internally generated mesoscale circulation on a larger domain are not particularly important and the character of the biases is not strongly dependent on the model or how the forcing is applied. There is a dependence on the averaging size of the forcing data. If the averaging domain is large enough and mesoscale fluctuations are sufficiently filtered out, the results converge. Using nudging itself to prevent model drift in time is important. The actual value for the relaxation time-scale is of minor importance provided that it is in the order of several hours.

As the semi-idealized setup stably represents a wide range of observed weather situations, it is also applicable as superparameterization (Grabowski, 2016) in a global model. It would be interesting to study how the overall performance of a global model with superparameterization depends on the chosen grid size which is tied to the horizontal domain size of the imbedded LES. As the LES obtain mean forcing profiles from the global model, the overall domain size from which the forcing is constructed might play a role as the semi-idealized setup depends on the averaging size of the forcing data.

The long-term LES approach cannot only be used to simulate periods at meteorological super-sites like in Schalkwijk et al. (2015) but also for simulating periods of (or even whole) measurement campaigns to support the interpretation of measurement results. This approach ~~is for example also~~ has been adopted for the Next Generation Aircraft Remote Sensing for Validation (NARVAL) series of flight campaigns over the tropical Atlantic (Klepp et al., 2014; Stevens et al., 2016) and is being followed

in the Large-Eddy Simulation (LES) ARM Symbiotic Simulation and Observation (LASSO) project (http://www.arm.gov/science/themes/lasso) where continuous LES of the Southern Great Plains atmospheric radiation measurement (ARM) supersite are under development.

One strength of the semi-idealized approach is that it is able to deliver robust turbulence statistics and a good representation of clouds, as typical for LES, and that it accounts for a localized area responding to every-day weather. However, a certain variability coming from the heterogeneous surface which usually surrounds any real observational site is neglected in the LES. The semi-idealized long-term LES approach can also be seen as an intermediate step towards LES in an limited-area setup, where, for example, a land-surface model and interactive radiation are used. In the framework of HD(CP)^2, these kind of simulations are performed over Germany. They are compared to the semi-idealized simulations presented here and the HOPE data set in Heinze et al. (2017). Comparing LES in semi-idealized and limited-area setups also allows to quantify the role of the mesoscale.

Appendix A: Construction of forcing data

To filter out any impact of small scale phenomena in the forcing data, the used COSMO (Baldauf et al., 2011) analysis data (with a spatial and temporal resolution of 2.8 km and 3 h, respectively) is averaged spatially. Note that the semi-idealized LES approach requires vertical profiles of geostrophic wind components $u_{g,i}$, of larger-scale velocity vector $u_{i,LS}$, of liquid water potential temperature $\theta_{l,LS}$, of total water ~~specific humidity~~ mixing ratio $q_{t,LS}$ and of larger-scale gradients (horizontal and vertical) of $\theta_{l,LS}$ and $q_{t,LS}$ (see Eqn. (1)-(4)). Moreover, corresponding surface conditions of temperature, humidity (or the respective sensible and latent heat fluxes) and hydrostatic pressure (which is important for cloud microphysics) are needed.

First, a spatial averaging domain with side lengths $\mathcal{D}_x$ (zonal) and $\mathcal{D}_y$ (meridional) is defined. These side lengths should be large enough to filter the small scales (see Sect. 5 for a discussion of adequate averaging domain sizes). For determining the entire set of larger-scale quantities required for the long-term LES approach, five averaging domains are needed, as shown in Fig. 11:

– one centered domain (black square) for the determination of surface conditions and vertical profiles of $u_{g,i}$, $u_{i,LS}$, $\theta_{l,LS}$ and $q_{t,LS}$ and

– four shifted domains (red and blue squares) for the determination of larger-scale horizontal gradients of $\theta_{l,LS}$ and $q_{t,LS}$.

The averaged quantities of the centered domain are then assumed to represent the large scale quantities in the LES. The centers of the shifted domains are located one-half $\mathcal{D}_x$ in the east-west direction and one-half $\mathcal{D}_y$ in the north-south direction, respectively. Hence, the larger-scale gradients used in Eq. (2) are approximated as follows:

$$\frac{\partial \varphi_{LS}}{\partial x_1} = \frac{\varphi_{LS,east} - \varphi_{LS,west}}{\mathcal{D}_x} \tag{A1}$$

$$\frac{\partial \varphi_{LS}}{\partial x_2} = \frac{\varphi_{LS,north} - \varphi_{LS,south}}{\mathcal{D}_y}, \tag{A2}$$

since the averaged quantities are assumed to represent the larger-scale conditions at the center of each domain.

Data availability

Primary data and scripts used in the analysis and other supplementary information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained at http://hdl.handle.net/.... (Heinze, 2017).

References

Althausen, D., Engelmann, R., Baars, H., Heese, B., Ansmann, A., Müller, D., and Komppula, M.: Portable Raman lidar Polly^{XT} for automated profiling of aerosol backscatter, extinction, and depolarization, J. Atmos. Oceanic Technol., 26, 2366–2378, doi:10.1175/2009JTECHA1304.1, 2009.

5   Ansorge, C. and Mellado, J. P.: Global intermittency and collapsing turbulence in the stratified planetary boundary layer, 2014.

Ansorge, C. and Mellado, J. P.: Analyses of external and global intermittency in the logarithmic layer of Ekman flow, J. Fluid Mech., 805, 611–635, doi:10.1017/jfm.2016.534, 2016.

Anthes, R. A.: Data assimilation and initialization of hurricane prediction models, J. Atmos. Sci., 32, 702–719, 1974.

Arakawa, A. and Lamb, V. R.: Computational design of the basic dynamical processes of the UCLA general cirulation model, in: General

10   circulation models of the atmosphere, edited by Chang, J., vol. 17, pp. 173–265, Academic Press, 1977.

ASTER GDEM Validation Team: ASTER global digital elevation model version 2 – summary of validation results, Tech. rep., NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA, http://www.jspacesystems.or.jp/ersdac/GDEM/ver2Validation/Summary_GDEM2_validation_report_final.pdf, 2011.

15   Baars, H., Ansmann, A., Engelmann, R., and Althausen, D.: Continuous monitoring of the boundary-layer top with lidar, Atmospheric Chemistry and Physics, 8, 7281–7296, doi:10.5194/acp-8-7281-2008, 2008.

Baas, P., Bosveld, F., Lenderink, G., van Meijgaard, E., and Holts, A. A. M.: How to design single-column model experiments for comparison with observed nocturnal low-level jets, Q.J.R. Meteorol. Soc., 136, 671–684, doi:10.1002/qj.592, 2010.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational convective-scale numerical weather

20   prediction with the COSMO model: description and sensitivities, Mon. Wea. Rev., 139, 3887–3905, doi:10.1175/MWR-D-10-05013.1, 2011.

Basu, S., Holtslag, A. A. M., van de Wiel, B. J. H., Moene, A., and Steeneveld, G.-J.: An inconvenient 'truth' about using sensible heat flux as a surface boundary condition in models under stably stratified regimes, Acta Geophysica, 56, 88–99, doi:10.2478/s11600-007-0038-y, 2008.

25   Basu, S., Holtslag, A., Caporaso, L., Riccio, A., and Steeneveld, G.-J.: Observational support for the stability dependence of the bulk Richardson number across the stable boundary layer, Boundary-Layer Meteorol., 150, 515–523, doi:10.1007/s10546-013-9878-y, 2014.

Beare, R. J., Macvean, M. K., Holtslag, A. A. M., Cuxart, J., Esau, I., Golaz, J.-C., Jimenez, M. A., Khairoutdinov, M., Kosovic, B., Lewellen, D., Lund, T. S., Lundquist, J. K., Mccabe, A., Moene, A. F., Noh, Y., Raasch, S., and Sullivan, P. P.: An intercomparison of large-eddy simulations of the stable boundary layer, Boundary-Layer Meteorol., 118, 247–272, 2006.

30   Behrendt, A., Wulfmeyer, V., Hammann, E., Muppa, S. K., and Pal, S.: Profiles of second- to fourth-order moments of turbulent temperature fluctuations in the convective boundary layer: first measurements with rotational Raman lidar, Atmos. Chem. Phys., 15, 5485–5500, doi:10.5194/acp-15-5485-2015, 2015.

Bosveld, F. C., Baas, P., van Meijgaard, E., de Bruijn, E. I. F., Steeneveld, G.-J., and Holtslag, A. A. M.: The third GABLS intercomparison case for evaluation studies of boundary-layer models. Part A: Case selection and set-Up, Boundary-Layer Meteorol., 142, 133–156,

35   doi:10.1007/s10546-014-9917-3, 2014.

Brown, A. R., Cederwall, R. T., Chlond, A., abd J.-C. Golaz, P. G. D., Khairoutdinov, M., Lewellen, D. C., Lock, A. P., MacVean, M. K., Moeng, C.-H., Neggers, R. A. J., Siebesma, A. P., and Stevens, B.: Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land, Q.J.R. Meteorol. Soc., 128, 1075–1094, doi:10.1256/003590002320373210, 2002.

Brutsaert, W.: The roughness length for water vapor sensible heat, and other scalars, J. Atmos. Sci., 32, 2029–2031, 1975.

5   Bühl, J., Seifert, P., Wandinger, U., Baars, H., Kanitz, T., Schmidt, J., Myagkov, A., Engelmann, R., Skupin, A., Heese, B., Klepel, A., Althausen, D., and Ansmann, A.: LACROS: the Leipzig Aerosol and Cloud Remote Observations System, Proc. SPIE 8890, Remote Sensing of Clouds and the Atmosphere XVIII; and Optics in Atmospheric Propagation and Adaptive Systems XVI, 889002, doi:10.1117/12.2030911, 2013.

Clyne, J., Mininni, P., Norton, A., and Rast, M.: Interactive desktop analysis of high resolution simulations: application to turbulent plume
10   dynamics and current sheet formation, New Journal of Physics, 9, 301, 2007.

Cuijpers, J. W. M. and Duynkerke, P. G.: Large eddy simulation of trade wind cumulus clouds, J. Atmos. Sci., 50, 3894–3908, doi:10.1175/1520-0469(1993)050<3894:LESOTW>2.0.CO;2, 1993.

Deardorff, J. W.: Convective velocity and temperature scales for the unstable planetary boundary layer and for Rayleigh convection, J. Atmos. Sci., 27, 1211–1213, doi:10.1175/1520-0469(1970)027<1211:CVATSF>2.0.CO;2, 1970a.

15 Deardorff, J. W.: Preliminary results from numerical integrations of the unstable planetary boundary layer, J. Atmos. Sci., 27, 1209–1211, doi:10.1175/1520-0469(1970)027<1209:PRFNIO>2.0.CO;2, 1970b.

Deardorff, J. W.: Numerical Investigation of Neutral and Unstable Planetary Boundary Layers, J. Atmos. Sci., 29, 91–115, doi:10.1175/1520-0469(1972)029<0091:NIONAU>2.0.CO;2, 1972.

Deardorff, J. W.: On the entrainment rate of a stratocumulus-topped mixed layer, Quart. J. Roy. Meteor. Soc., 102, 563–582,
20   doi:10.1002/qj.49710243306, 1976.

Deardorff, J. W.: Stratocumulus-capped mixed layers derived from a three-dimensional model, Boundary-Layer Meteorol., 18, 495–527, doi:10.1007/BF00119502, 1980.

Durran, D. R.: Numerical methods for wave equations in geophysical fluid dynamics, Springer-Verlag New York, 1999.

Dutton, J. A. and Fichtl, G. H.: Approximate equations of motion for gases and liquids, J. Atmos. Sci., 26, 241–254, doi:10.1175/1520-
25   0469(1969)026<0241:AEOMFG>2.0.CO;2, 1969.

Edwards, J. M., Basu, S., Bosveld, F. C., and Holtslag, A. A. M.: The impact of radiation on the GABLS3 large-eddy simulation through the night and during the morning transition, Boundary-Layer Meteorology, 152, 189–211, doi:10.1007/s10546-013-9895-x, 2014.

Engelmann, R., Kanitz, T., Baars, H., Heese, B., Althausen, D., Skupin, A., Wandinger, U., Komppula, M., Stachlewska, I. S., Marinou, V. A. E., Mattis, I., Linné, H., A., and Ansmann: The automated multiwavelength Raman polarization and water-vapor lidar PollyXT: the
30   neXT generation, Atmos. Meas. Tech., 9, 1767–1784, doi:10.5194/amt-9-1767-2016, 2016.

Grabowski, W. W.: Towards global large eddy simulation: super-parameterization revisited, J. Meteor. Soc. Japan, 94, in press, doi:10.2151/jmsj.2016-017, 2016.

Grabowski, W. W., Wu, X., and Moncrieff, M. W.: Cloud-resolving modeling of tropical cloud systems during phase III of GATE. Part I: Two-dimensional experiments, J. Atmos. Sci., 53, 3684–3709, doi:10.1175/1520-0469(1996)053<3684:CRMOTC>2.0.CO;2, 1996.

35 Graf, A., Schüttemeyer, D., Geiß, H., Knaps, A., Möllmann-Coers, M., Schween, J. H., Kollet, S., Neininger, B., Herbst, M., and Vereecken, H.: Boundedness of turbulent temperature probability distributions, and their relation to the vertical profile in the convective boundary layer, Boundary-Layer Meteorol., 134, 459–486, doi:10.1007/s10546-009-9444-9, 2010.

Hammann, E., Behrend, A., Mounier, F. L., and Wulfmeye, V.: Temperature profiling of the atmospheric boundary layer with rotational Raman lidar during the HD(CP)2 Observational Prototype Experiment, Atmos. Chem. Phys., 15, 2867–2881, doi:10.5194/acp-15-2867-2015, 2015.

Harlow, F. H. and Welch, J. E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface, Phys. Fluids, 8, 2182–2189, 1965.

Heinze, R.: Primary data and statistics used for the paper "Evaluation of large-eddy simulations forced with mesoscale model output for a multi-week period during a measurement campaign", available at:, http://hdl.handle.net/...., 2017.

Heinze, R., Mironov, D., and Raasch, S.: Second-moment budgets in cloud-topped boundary layers: A large-eddy simulation study, J. Adv. Model. Earth Syst., 07, doi:10.1002/2014MS000376, 2015.

Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., Xie, X., Adamidis, P., Ament, F., Baars, H., Barthlott, C., Behrendt, A., Blahak, U., Bley, S., Brdar, S., Brueck, M., Crewell, S., Deneke, H., Di Girolamo, P., Evaristo, R., Fischer, J., Frank, C., Friederichs, P., Göcke, T., Gorges, K., Hande, L., Hanke, M., Hansen, A., Hege, H.-C., Hoose, C., Jahns, T., Kalthoff, N., Klocke, D., Kneifel, S., Knippertz, P., Kuhn, A., van Laar, T., Macke, A., Maurer, V., Mayer, B., Meyer, C. I., Muppa, S. K., Neggers, R. A. J., Orlandi, E., Pantillon, F., Pospichal, B., Röber, N., Scheck, L., Seifert, A., Seifert, P., Senf, F., Siligam, P., Simmer, C., Steinke, S., Stevens, B., Wapler, K., Weniger, M., Wulfmeyer, V., Zängl, G., Zhang, D., and Quaas, J.: Large-eddy simulations over Germany using ICON: a comprehensive evaluation, Q.J.R. Meteorol. Soc., 143, 69–100, doi:10.1002/qj.2947, 2017.

Hoffmann, F.: The effect of spurious cloud edge supersaturations in Lagrangian cloud models: An analytical and numerical study,, Mon. Wea. Rev., 144, 107–118, doi:10.1175/MWR-D-15-0234.1, 2016.

Hoffmann, F., Raasch, S., and Noh, Y.: Entrainment of aerosols and their activation in a shallow cumulus cloud studied with a coupled LCM-LES approach, Atmos. Res., 156, 43–57, doi:10.1016/j.atmosres.2014.12.008, 2015.

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Delanoë, J., Pelon, J., Protat, A., Brooks, M. E., Gaussiat, N., Wilson, D. R., Donovan, D. P., Baltink, H. K., van Zadelhoff, G.-J., Eastment, J. D., Goddard, J. W. F., Wrench, C. L., Haeffelin, M., Krasnov, O. A., Russchenberg, H. W. J., Piriou, J.-M., Vinit, F., Seifert, A., Tompkins, A. M., and Willén, U.: Cloudnet, Bull. Amer. Meteor. Soc., 88, 883–998, doi:10.1175/BAMS-88-6-883, 2007.

Kalthoff, N., Adler, B., Wieser, A., Kohler, M., Träumner, K., Handwerker, J., Corsmeier, U., Khodayar, S., Lambert, D., Kopmann, A., Kunka, N., Dick, G., Ramatschi, M., Wickert, J., and Kottmeier, C.: KITcube – a mobile observation platform for convection studies deployed during HyMeX, Meteorol. Z., 22, 633–647, doi:10.1127/0941-2948/2013/0542, 2013.

Klepp, C., Ament, F., Bakan, S., Hirsch, L., and Stevens, B.: The NARVAL Campaign Report, Tech. rep., Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie, http://hdl.handle.net/11858/00-001M-0000-0026-A6F3-7, 2014.

Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics, J. Atmos. Oceanic Technol., 11, 661–673., doi:10.1175/1520-0426(1994)011<0661:HLILEW>2.0.CO;2, 1994.

Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring second- through fourth-order moments in noisy data, J. Atmos. Oceanic Technol., 17, 1330–1347, doi:10.1175/1520-0426(2000)017<1330:MSTFOM>2.0.CO;2, 2000.

Löhnert, U., Schween, J. H., Acquistapace, C., Ebell, K., Maahn, M., Barreraverdejo, M., Hirsikko, A., Bohn, B., Knaps, A., O'connor, E., Simmer, C., Wahner, A., and Crewell, S.: JOYCE: Jülich observatory for cloud evolution, Bull. Amer. Meteor. Soc., 96, 1157–1174, doi:10.1175/BAMS-D-14-00105.1, 2015.

Macke, A., Seifert, P., Baars, H., Barthlott, C., Beekmans, C., Behrendt, A., Bohn, B., Brück, M., Bühl, J., Crewell, S., Damian, T., Deneke, H., Düsing, S., Foth, A., Di Girolamo, P., Hammann, E., Heinze, R., Hirsikko, A., Kalisch, J., Kalthoff, N., Kinne, S., Kohler, M., Löhnert,

U., Madhavan, B. L., Maurer, V., Muppa, S. K., Schween, J., Serikov, I., Siebert, H., Simmer, C., Späth, F., Steinke, S., Träumner, K., Trömel, S., Wehner, B., Wieser, A., Wulfmeyer, V., and Xie, X.: The HD(CP)2 Observational Prototype Experiment HOPE – an overview, Atmos. Chem. Phys., pp. 4887–4914, doi:10.5194/acp-17-4887-2017, 2017.

Maronga, B., Gryschka, M., Heinze, R., Hoffmann, F., Kanani-Sühring, F., Keck, M., Ketelsen, K., Letzel, M.-O., Sühring, M., and Raasch, S.: The Parallelized Large-Eddy Simulation Model (PALM) version 4.0 for atmospheric and oceanic flows: model formulation, recent developments, and future perspectives, Geosci. Model Dev. Discuss., 8, 1539–1637, doi:10.5194/gmdd-8-1539-2015, 2015.

Mason, P. J.: Large-eddy simulation of dispersion in convective boundary layers with wind shear, Atm. Env., 26A, 15611 571, doi:10.1016/0960-1686(92)90056-Q, 1992.

Mason, P. J. and Derbyshire, S. H.: Large-eddy simulation of the stably-stratified atmospheric boundary layer, Boundary-Layer Meteorol., 53, 117–162, doi:10.1007/BF00122467, 1990.

Maurer, V., Kalthoff, N., Wieser, A., M. Kohler, M. M., and Gantner, L.: Observed spatial variability of boundary-layer turbulence over flat, heterogeneous terrain, Atmos. Chem. Phys., 16, 1377–1400, doi:10.5194/acp-16-1377-2016, 2016.

Moeng, C.-H.: A large-eddy-simulation model for the study of planetary boundary-layer turbulence, J. Atmos. Sci., 41, 2052–2062, doi:10.1175/1520-0469(1984)041<2052:ALESMF>2.0.CO;2, 1984.

Moeng, C.-H.: Large-eddy simulation of a stratus topped boundary layer. Part I: structure und budgets, J. Atmos. Sci., 43, 2880–2900, doi:10.1175/1520-0469(1986)043<2886:LESOAS>2.0.CO;2, 1986.

Moeng, C.-H. and Sullivan, P.: A comparison of shear- and buoyancy-driven planetary boundary layer flows., J. Atmos. Sci., 51, 999–1022, doi:10.1175/1520-0469(1994)051<0999:ACOSAB>2.0.CO;2, 1994.

Muppa, S. K., Behrendt, A., Späth, F., Wulfmeyer, V., Metzendorf, S., and Riede, A.: Turbulent humidity fluctuations in the convective boundary layer: Case studies using water vapor differential absorption lidar measurements, Boundary-Layer Meteorol., 16, 43–66, doi:10.1007/s10546-015-0078-9, 2016.

Neggers, R. A. J. and Siebesma, A. P.: Constraining a system of interacting parameterizations through multiple-parameter evaluation: tracing a compensating error between cloud vertical structure and cloud overlap, J. Climate, 26, 6698–6715, doi:10.1175/JCLI-D-12-00779.1, 2013.

Neggers, R. A. J., Siebesma, A. P., and Heus, T.: Continuous single-column model evaluation at a permanent meteorological supersite, Bull. Amer. Meteor. Soc., 93, 1389–1400, doi:10.1175/BAMS-D-11-00162.1, 2012.

Nieuwstadt, F. T. M. and Brost, R. A.: The decay of convective turbulence, J. Atmos. Sci., 45, 532–546, doi:10.1175/1520-0469(1986)043<0532:TDOCT>2.0.CO;2, 1986.

Noh, Y. W., Cheon, W. G., Hong, S. Y., and Raasch, S.: Improvement of the k-profile model for the planetary boundary layer based on large eddy simulation data, Boundary-Layer Meteorol., 107, 401–427, doi:10.1023/A:1022146015946, 2003.

Ogura, Y. and Phillips, N. A.: Scale analysis of deep and shallow convection in the attmosphere, J. Atmos. Sci., 19, 1962.

Pietersen, H. P., de Arellano, J. V.-G., Augustin, P., van de Boer, A., de Coster, O., Delbarre, H., Durand, P., Fourmentin, M., Gioli, B., abd F. Lohou, O. H., Lothon, M., Ouwersloot, H. G., Pino, D., and Reuder, J.: Study of a prototypical convective boundary layer observed during BLLAST: contributions by large-scale forcings , Atmos. Chem. Phys., 15, 4241–4257, doi:10.5194/acp-15-4241-2015, 2015.

Randall, D. A. and Cripe, D. G.: Alternative methods for specification of observed forcing in single-column models and cloud system models, J. Geophys. Res., 104, D20, 24 527–24 545, doi:10.1029/1999JD900765, 1999.

Richardson, H., Basu, S., and Holtslag, A.: Improving stable boundary-layer height estimation using a stability-dependent critical bulk richardson number, Boundary-Layer Meteorol., 148, 93–109, doi:10.1007/s10546-013-9812-3, 2013.

Riechelmann, T., Wacker, U., Beheng, K., Etling, D., and Raasch, S.: Influence of turbulence on the drop growth in warm clouds, Part II: Sensitivity studies with a spectral bin microphysics and a Lagrangian cloud model, Meteorol. Z., 24, 293–311, doi:0.1127/metz/2015/0608, 2015.

Rieck, M., Hohenegger, C., and Gentine, P.: The effect of moist convection on thermally induced mesoscale circulations, Q.J.R. Meteorol. Soc., 141, 2418–2428, doi:10.1002/qj.2532, http://dx.doi.org/10.1002/qj.2532, 2015.

Saiki, E. M., Moeng, C.-H., and Sullivan, P. P.: Large-eddy simulation of the stably stratified planetary boundary layer, Boundary-Layer Meteorol., 95, 1–30, doi:10.1023/A:1002428223156, 2000.

Schalkwijk, J., Griffith, E. J., Post, F. H., and Jonker, H. J. J.: High-performance simulations of turbulent clouds on a desktop PC: Exploiting the GPU, Bull. Amer. Meteor. Soc., 93, 307–314, doi:10.1175/BAMS-D-11-00059.1, 2012.

Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., and Bosveld, F. C.: A year-long large-eddy simulation of the weather over Cabauw: An overview, Mon. Wea. Rev., 143, 828–844, doi:10.1175/MWR-D-14-00293.1, 2015.

Schlemmer, L. and Hohenegger, C.: Modifications of the atmospheric moisture field as a result of cold-pool dynamics, Q.J.R. Meteorol. Soc., 142, 30–42, doi:10.1002/qj.2625, http://dx.doi.org/10.1002/qj.2625, 2016.

Schween, J. H., Hirsikko, A., Löhnert, U., and Crewell, S.: Mixing-layer height retrieval with ceilometer and Doppler lidar: from case studies to long-term assessment , Atmos. Meas. Tech., 7, Atmos. Meas. Tech., doi:10.5194/amt-7-3685-2014, 2014.

Seifert, A.: On the parameterization of evaporation of raindrops as simulated by a one-dimensional rainshaft model, J. Atmos. Sci., 65, 3608–3619, doi:10.1175/2008JAS2586.1, 2008.

Seifert, A. and Beheng, K. D.: A double-moment parameterization for simulating autoconversion, accretion and selfcollection, Atmospheric Re, 59-60, 265–281, doi:10.1016/S0169-8095(01)00126-0, 2001.

Seifert, A. and Beheng, K. D.: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 1: Model description, Meteorol. Atmos. Phys., 92, 45–66, doi:10.1007/s00703-005-0112-4, 2006.

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., Jiang, H., Khairoutdinov, M., Lewellen, D., Moeng, C.-H., Sanchez, E., Stevens, B., and Stevens, D. E.: A large eddy simulation intercomparison study of shallow cumulus convection, J. Atmos. Sci., 60, 1201–1219, doi:10.1175/1520-0469(2003)60<1201:ALESIS>2.0.CO;2, 2003.

Smagorinsky, J.: General circulation experiments with the primitive equations I. The basic experiment, Mon. Wea. Rev., 91(3), ., 1963.

Sommeria, G.: Three-dimensional simulation of turbulent processes in an undisturbed trade wind boundary layer, J. Atmos. Sci., 33, 216–241, doi:10.1175/1520-0469(1976)033<0216:TDSOTP>2.0.CO;2, 1976.

Sorbjan, Z.: Decay of convective turbulence revisited. Boundary-Layer Meteor, Boundary-Layer Meteorol., 82, 501–515, doi:10.1023/A:1000231524314, 1997.

Stauffer, D. R. and Bao, J.-W.: Optimal determination of nudging coefficients using adjoint equations, Tellus, 45A, 358–369, 1993.

Steinke, S., Eikenberg, S., Löhnert, U., Dick, G., Klocke, D., Girolamo, P. D., and Crewell, S.: Assessment of small-scale integrated water vapour variability during HOPE, Atmos. Chem. Phys., 15, 2675–2692, doi:10.5194/acp-15-2675-2015, 2015.

Sterk, H. A. M., Steeneveld, G. J., Vihma, T., Anderson, P. S., Bosveld, F. C., and Holtslag, A. A. M.: Clear-sky stable boundary layers with low winds over snow-covered surfaces Part I: A WRF model evaluation, Quart. J. Roy. Meteor. Soc., 141, 2165–2184, doi:10.1002/qj.2513, 2015.

Stevens, B., Cotton, W. R., Feingold, G., and Moeng, C.-H.: Large-eddy simulations ofstrongly precipitating, shallow, stratocumulus-topped boundary layers, J. Atmos. Sci., 55, 3616–3638, doi:10.1175/1520-0469(1998)055<3616:LESOSP>2.0.CO;2, 1998.

Stevens, B., Ackermann, A. S., Albrecht, B. A., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., Lewellen, D. C., Macvean, M. K., Neggers, R. A. J., Sanchez, E., Siebesma, A. P., and Stevens, D. E.: Simulations of trade wind cumuli under a strong inversion, J. Atmos. Sci., 58, 1870–1891, 2001.

Stevens, B., Moeng, C.-H., Ackermann, A. S., Bretherton, C. S., Chlond, A., de Roode, S., Edwards, J., Golaz, J.-C., Jiang, H., Khairoutdinov, A., Kirkpatrick, M. P., Lewellen, D. C., Lock, A., Müller, F., Stevens, D. E., Whelan, E., and Zhu, P.: Evaluation of large-eddy simulations via observations of nocturnal marine stratocumlus, Mon. Wea. Rev., 133, 1443–1462, doi:10.1175/MWR2930.1, 2005.

Stevens, B. S., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., and. B. Brügmann, I. S., Forde, M., Linné, H., Lonitz, K., and Prospero, J. M.: The Barbados Cloud Observatory: Anchoring Investigations of Clouds and Circulation on the Edge of the ITCZ, Bull. Amer. Meteor. Soc., 97, 787–801, doi:10.1175/BAMS-D-14-00247.1, 2016.

Stull, R. B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, 1988.

Sühring, M. and Raasch, S.: Heterogeneity-induced heat flux patterns in the convective boundary layer: Can they be detected from observations and is there a blending height? - A large-eddy simulation study for the LITFASS-2003 experiment, Boundary-Layer Meteorol., 148, 309–331, doi:10.1007/s10546-013-9822-1, 2013.

Taylor, G. I.: Effects of variation in density on the stability of superimposed streams of fluids, Proc. Roy. Soc. (Lond.), Ser. A, 132, 499–523, 1931.

van Stratum, B. J. H. and Stevens, B.: The influence of misrepresenting the nocturnal boundary layer on idealized daytime convection in large-eddy simulation, J. Adv. Model. Earth Syst., p. doi:10.1002/2014MS000370, 2015.

vanZanten, A. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackermann, A. S., Burnet, F., Cheng, A., Couvreux, F., Jiang, H., Khairoutdinov, M., Kogan, Y., Lewellen, D. C., Mechem, D., Nakamura, K., Noda, A., Shipway, B. J., Slawinska, J., Wang, S., and Wyszogrodzki, A.: Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during Rico, J. Adv. Model. Earth Syst., 3, 20 pp, doi:10.1029/2011MS000056, 2011.

Vollmer, L., van Dooren, M., Trabucchi, D., Schneemann, J., Steinfeld, G., Witha, B., Trujillo, J., and Kühn, M.: First comparison of LES of an offshore wind turbine wake with dual-Doppler lidar measurements in a German offshore wind farm, Journal of Physics: Conference Series, 625, 012 001, http://stacks.iop.org/1742-6596/625/i=1/a=012001, 2015.

Wicker, L. J. and Skamarock, W. S.: Time-splitting methods for elastic models using forward time schemes, J. Atmos. Sci., 130, 2088–2097, doi:10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2, 2002.

Wulfmeyer, D., Muppa, S., Behrendt, A., Hammann, E., Späth, F., Sorbjan, Z., Turner, D., and Hardesty, R.: Determination of the convective boundary layer entrianment fluxes, dissipation rates and the molecular destruction of variances: Theoretical descritpion and a strategy for its confirmation with a novel lidar system synergy, J. Atmos. Sci., 73, 667–692, doi:10.1175/JAS-D-14-0392.1, 2016.

Wulfmeyer, V., Hardesty, R. M., Turner, D. D., Behrendt, A., Cadeddu, M. P., Girolamo, P. D., Schlüssel, P., Baelen, J. V., and Zus, F.: A review of the remote sensing of lower-tropospheric thermodynamic profiles and its indispensable role for the understanding and the simulation of water and energy cycles, Rev. Geophys., 53, 819–895, doi:10.1002/2014RG000476, 2015.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A network of terrestrial environmental observatories in Germany, Vadose Zone J., 10, 955–973, doi:10.2136/vzj2010.0139, 2011.

Zilitinkevich, S. and Baklanov, A.: Calculation of the height of the stable boundary layer in practical applications, Boundary-Layer Meteorol., 105, 389–409, doi:10.1023/A:1020376832738, 2002.
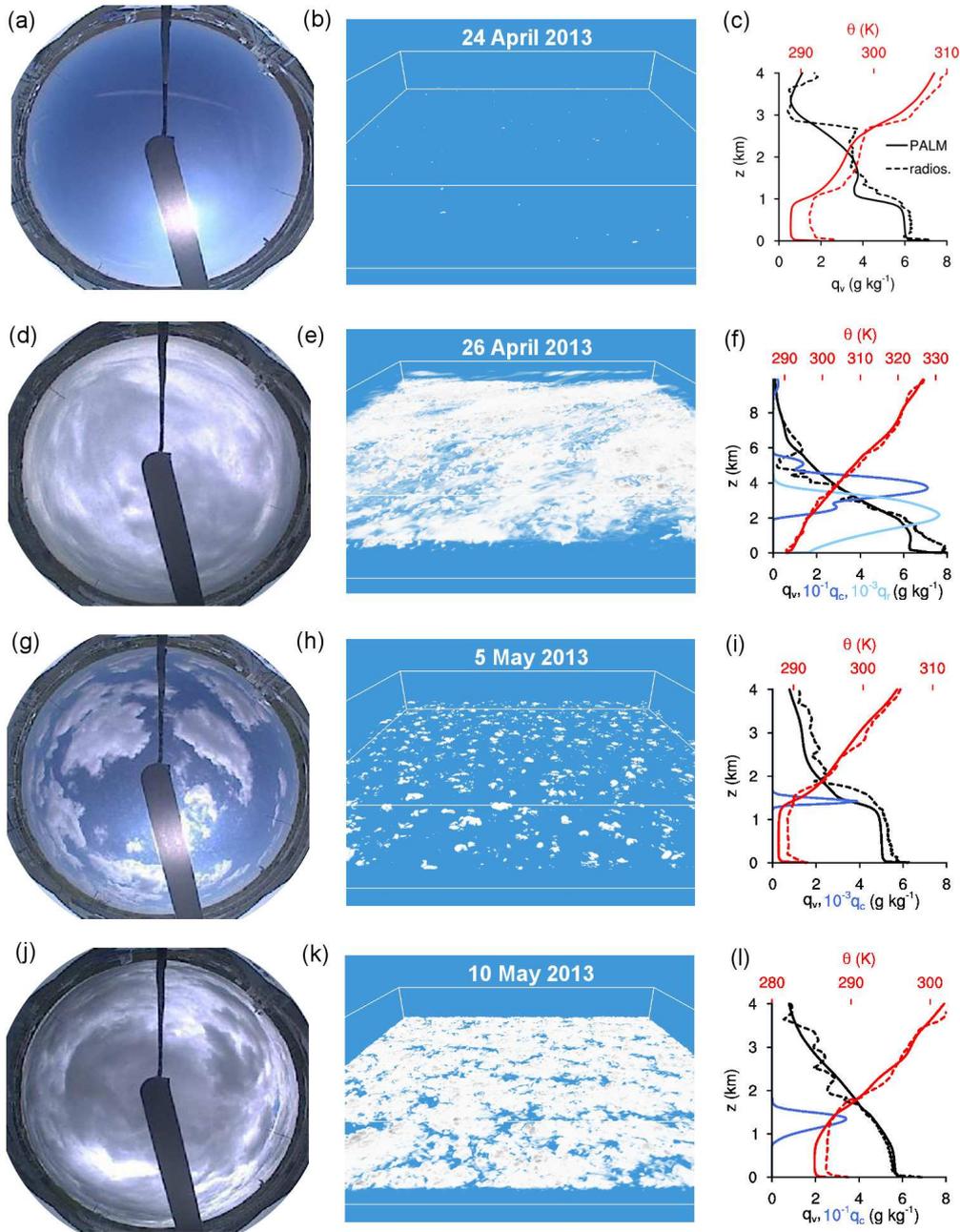
Figure 3. Snapshots and mean profiles of four different days (24 April, 26 April, 5 May, 10 May) taken at 11 UTC. The left column (panels a, d, g, j) shows images taken with the total sky imager TSI-880 at JOYCE site, the middle column (panels b, e, h, k) shows volume-rendered cloud water ~~specific humidity~~ mixing ratio q_c of the PALM reference simulation RP and the right column (panels c, f, i, l) shows mean profiles of ~~specific humidity~~ mixing ratio q_v (black)), potential temperature θ (red), cloud and rain water ~~specific humidity~~ mixing ratio q_c (blue) and q_r (light blue), respectively. The solid lines are horizontally averaged profiles of RP and the dashed lines are profiles from radio soundings (radios.) launched at KITcube site. Note that the vertical axis in panel f extends up to 10 km.
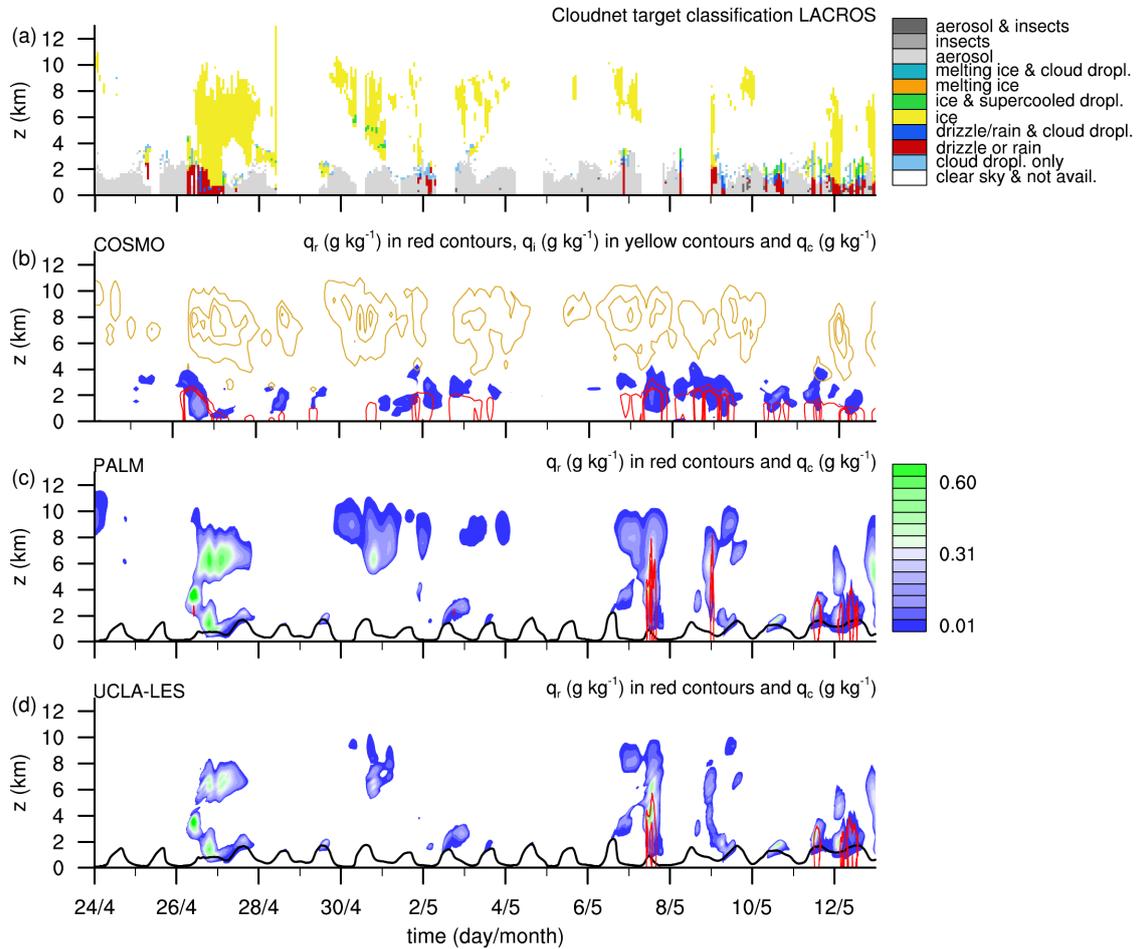
Figure 4. Time-height cross-sections of Cloudnet target classification (in panel a), specific cloud ice q_i (in yellow contours ranging from $0.001 - 0.21$ g kg^{-1} by 0.1 g kg^{-1}), specific rain water q_r (in red contours ranging from $0.001 - 0.01$ g kg^{-1} by 0.05 g kg^{-1}) and specific cloud water q_c (colored contours) of the COSMO forcing in panel b, specific cloud and rain water of PALM (run *RP*) in panel c and specific cloud and rain water of UCLA-LES (run *RU*) in panel d. The red contours in panels c and d have the same values as in panel b. The black lines in panels c and d denote the boundary layer depth according to the bulk-Richardson number criterion (see also Fig. 5). The same colorbar is used in panels b, c and d. Note, that in panels b, c and d the time series of horizontally averaged profiles are shown.
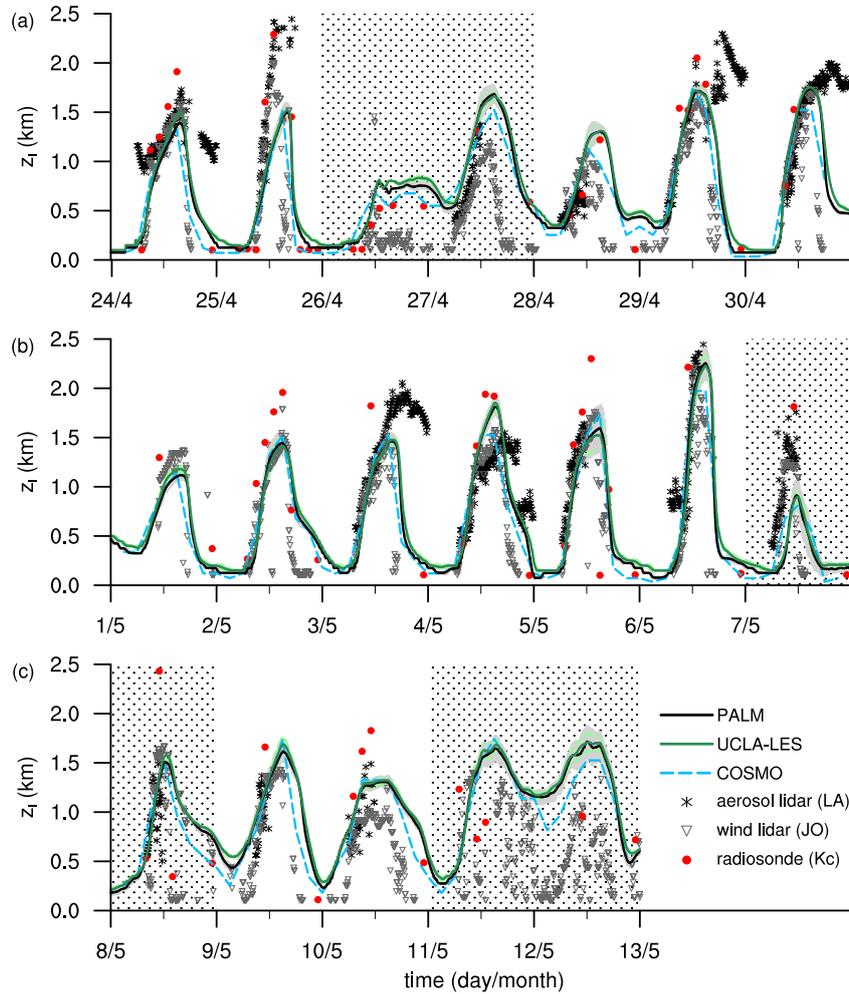
Figure 5. Temporal evolution of the boundary layer depth z_i for the total 19-day period (grouped in weeks). z_i is determined by means of the bulk-Richardson number criterion in all three models (PALM, UCLA-LES and COSMO) and in the radiosonde data. A criterion based on the vertical velocity variance and detected aerosol layers is used for the wind lidar and aerosol lidar, respectively. Radiosondes were launched at the KIT-cube site, the wind lidar and aerosol lidar took measurements at JOYCE and LACROS site, respectively. Gray and green shading denote twice the standard deviation of z_i in PALM and UCLA-LES, respectively. ~~Yellow~~ Stippled highlighting marks days with strong vertical forcing ($\widetilde{w_{SUB}} > 0.05$ m s^{-1}).
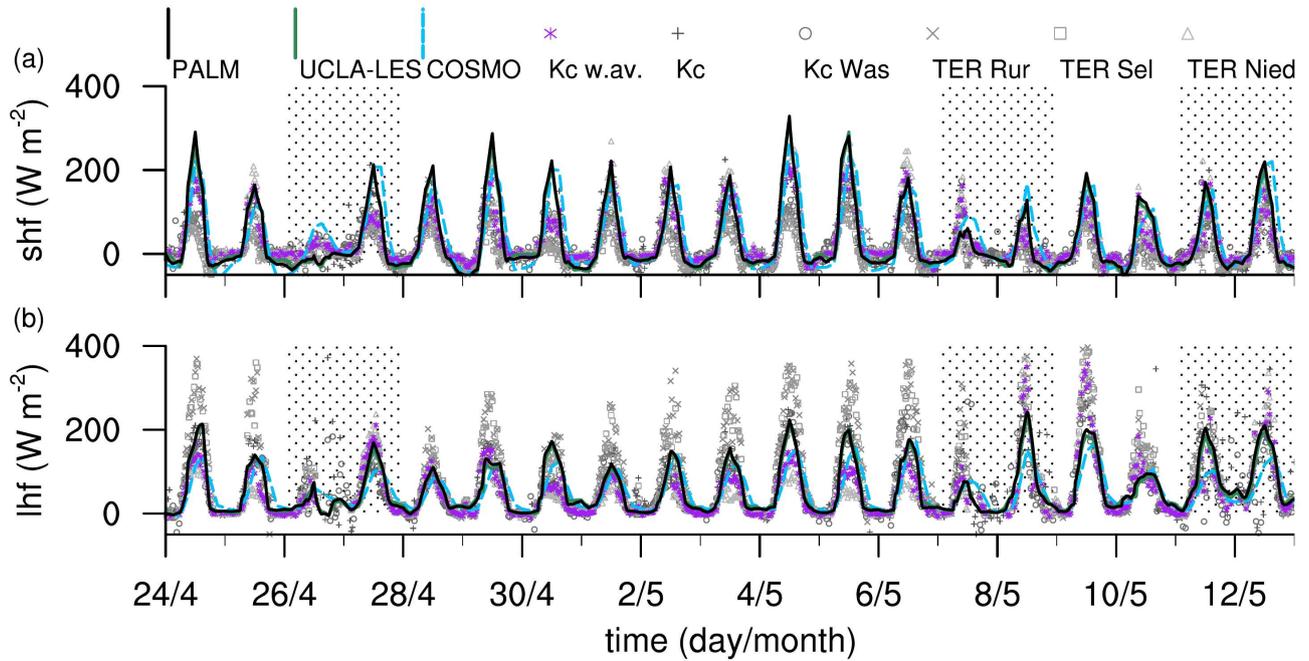
Figure 6. Temporal evolution of surface sensible heat flux ~~(~~shf <u>in</u> panel a ~~), and~~ surface latent heat flux ~~(~~lhf <u>in</u> panel b~~), wind direction at 120 m height (wdir$_{120m}$ panel c), wind speed at 120 m height (|v|$_{h,120m}$ panel d), potential temperature at 25 m height (θ_{25m} panel e), integrated water vapor (IWV panel f), and liquid water path (LWP panel g)~~. An overview of the measurements and abbreviations is given in Tab. 1. ~~The top legend refers to panels a and b, the middle panel refers to panels c, d and e and the lower legend refers to panels f and g. Gray and green shading in panel g denote twice the standard deviation of LWP in PALM and UCLA-LES, respectively. Yellow~~ <u>Stippled</u> highlighting marks days with strong vertical forcing ($\widetilde{w_{SUB}} > 0.05$ m s^{-1}).
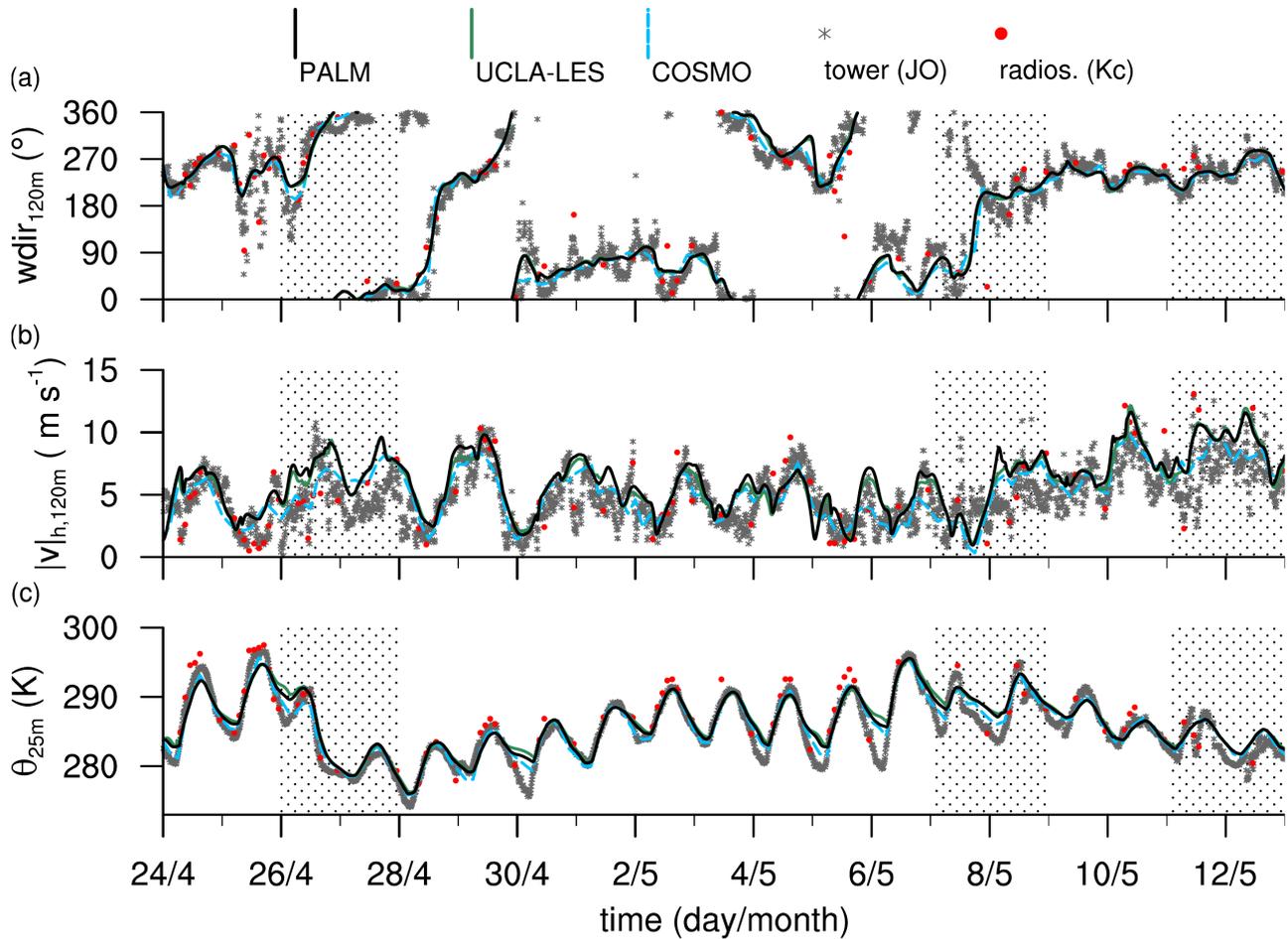
Figure 7. Temporal evolution of wind direction at 120 m height $wdir_{120m}$ in panel a, wind speed at 120 m height $|v|_{h,120m}$ in panel b and potential temperature at 25 m height θ_{25m} in panel c. An overview of the measurements and abbreviations is given in Tab. 1. Stippled highlighting as in Fig. 6.

36
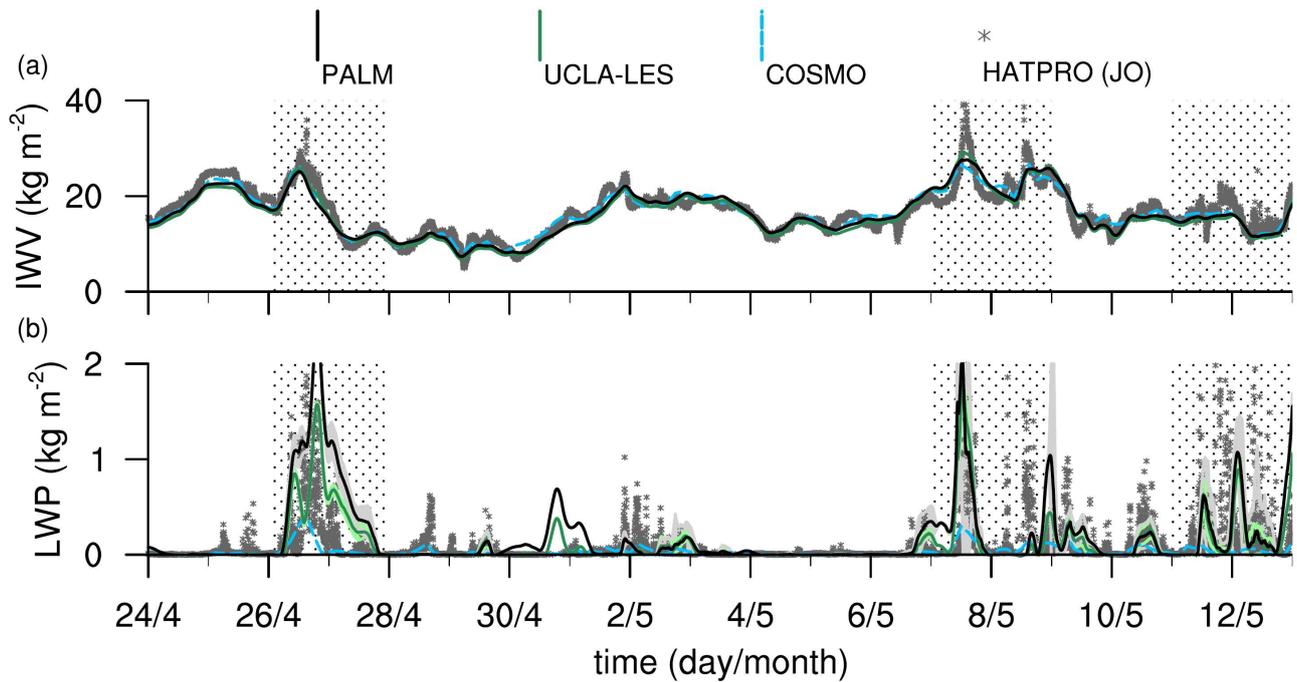
Figure 8. Temporal evolution of integrated water vapor IWV in panel a and liquid water path LWP in panel b. An overview of the measurements and abbreviations is given in Tab. 1. Gray and green shading in panel b denote twice the standard deviation of LWP in PALM and UCLA-LES, respectively. Stippled highlighting as in Fig. 6.
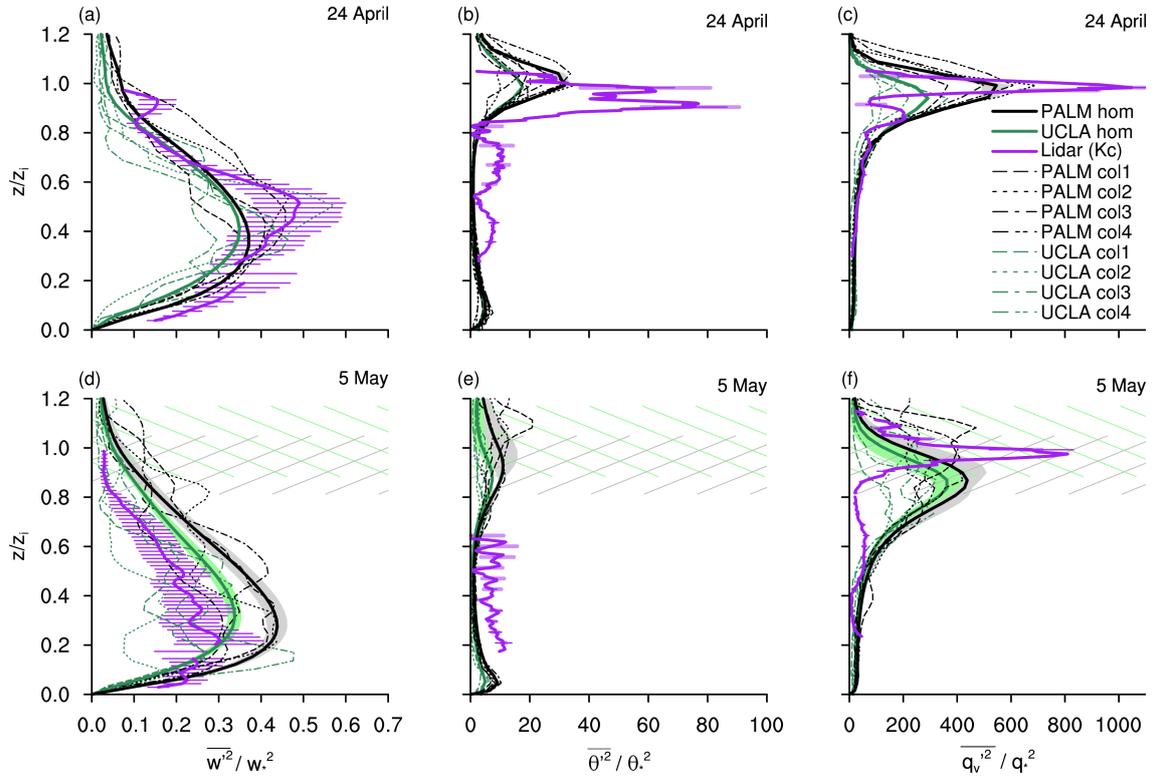
Figure 9. Normalized vertical profiles of vertical velocity variance, panels a and d, potential temperature variance, panels b and e, and ~~specific humidity~~ mixing ratio variance, panels c and f, for a 1 h-period between 11 and 12 UTC for 24 April and 5 May 2013, respectively. Solid black and green lines show variances of PALM and UCLA-LES determined as departure from the horizontal mean (*hom*) and averaged over 1 h including standard deviations denoted as solid gray and light green areas. Thin dashed black and green lines show variances from single-column output (*colX*) at four different grid points determined as departure from a one hour temporal mean. Solid purple lines denote variances from the KIT Doppler lidar including the statistical error according to Lenschow et al. (1994) as error bars, panels a and d, the rotational Raman lidar, panels b and e, and the water vapor differential absorption lidar from University of Hohenheim, panels c and f, (see Tab. 1 for further details). The thin purple (thick light purple) error bars in panels b, c, e and f show the noise (sampling) error according to Lenschow et al. (2000). Gray and light green shaded regions in panels d-f denote the cloud boundaries of PALM and UCLA-LES respectively. See Tab. 3 for the scaling values used.
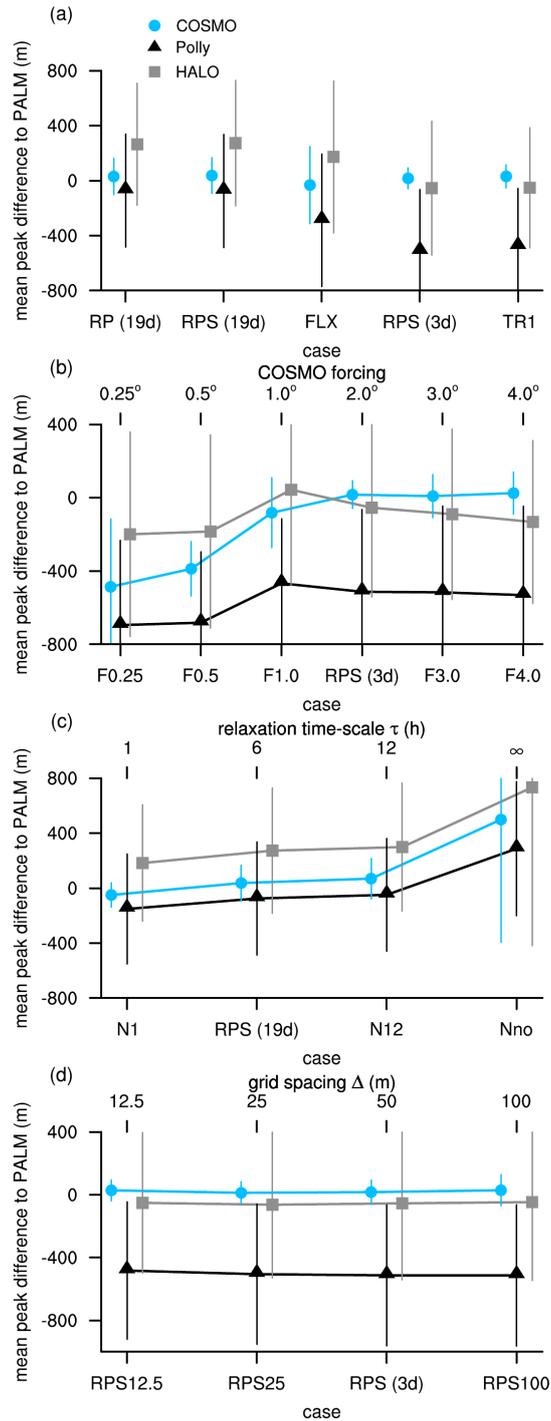
Figure 10. Mean peak difference in boundary layer ~~height~~ depth to PALM between 12 and 14 UTC for the simulated cases listed in Tab. 5. Standard deviations are provided along with the means. Panels b,c and d include the mean peak difference for the sensitivity experiments about the averaging size of the COSMO forcing, the nudging time-scale and the grid spacing, respectively. Panel a lists the remaining cases. Note that the mean peak difference of the PALM reference run on the small domain (*RPS*) is calculated over the whole 19 days (*RPS (19d)*) and the three day testing period (*RPS (3d)*), respectively. The number of values entering the average are (38,147,464) for 19 day runs and (6,38,78) for 3 day runs, respectively. The tuples denote the number of values entering the mean of the difference in boundary layer depth to (COSMO,Polly,HALO).

39
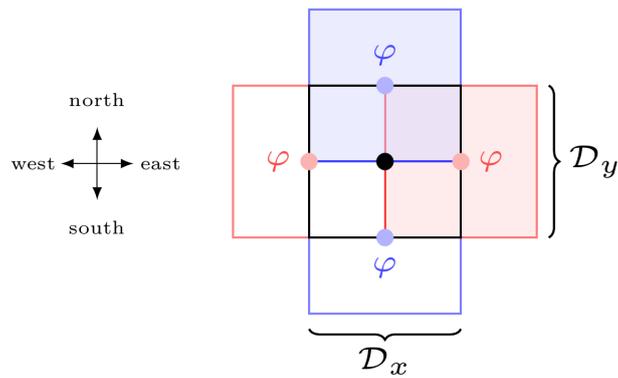
Figure 11. Averaging concept for the determination of larger-scale forcing terms from COSMO model output. The shifted domains (red/blue) are used for the calculation of larger-scale gradients of $\phi \in \{\theta_{l,LS}, q_{t,LS}\}$. The centered averaging domain (black) is used for the calculation of all other larger-scale quantities.