Comments on: Boundary-layer turbulent processes and mesoscale variability represented by numerical weather prediction models during the BLLAST campaign, by Couvreux et al.

General comments

I already made several in my overall quick review (attached at the end of this document). However, I think it is very important to focus more on the impacts of the different grids, in terms of what model TKE is most comparable to observations, in terms of shadowing (and its impact on surface fluxes, especially in the evening and early morning), and in terms of resolved boundary layer structures, which can account for an important part of the TKE (w of order of 1 m/s in Ching et al 2014 MWR and LeMone et al. 2013 MWR).

Also, the impact of the different model terrain, particularly on heterogeneity at night. Acevedo and Fitzjarrald and LeMone et al. (2003, JAS) both show terrain plays a role in nighttime horizontal heterogeneity.

I spend a lot of time writing what model variables might be directly comparable to the TKE measured in the atmosphere. This would be unambiguous if all PBL transport were proportional to the local gradient (i.e., don't need mass flux in the PBL schemes) and there are no resolved PBL eddies (possible with large horizontal grid spacing). It starts to get ambiguous when you have the resolved eddies (I'd just add their TKE to the subgrid TKE), and when you have mass flux in your EDMF schemes. What I don't know is whether the "MF" in the mass flux scheme is by TKE is completely separate from that in the "TKE" part of the scheme. In my comments, I assumed that it was, i.e., that the model TKE was the sum of the subgrid TKE + MF TKE + resolved-eddy TKE.

Peggy LeMone

Specific Comments:

P1 L26. Should be 24-h forecasts

P2 L1-2: Not sure what you mean here. Do you mean that there were more forests in the model or in reality? You could clarify by being more specific, for example, "related to identifying mixed forest and meadows as "forest" in two grid scales in the model." (If there is too much forest in the model).

P2 L10-11. How about "The model reproduced the range of variables to within an order of magnitude." (This is more compact; you don't need to write that it was analyzed).

P2 L29. Don't need "the" before "Europe"

P2 L29. It is interesting that this model has a warm bias in cold and stable conditions. Don't most models show cold biases under such circumstances?

P3, L3-4. LeMone et al. evaluated PBL schemes and their diagnositics.

L25, work of Acevedo and Fitzjarrald. LeMone et al. (2003, JAS) showed from CASES-97 data and evaluation of results of earlier field programs that the timing of maximum horizontal variability depends on the scale of the terrain. This is because of how long it takes for drainage currents to flow from high points to low points. This timing has to be also affected by "frost hollows" that are sheltered from the wind. This makes LES of limited value unless it has a very large domain with very fine mesh. Just a comment; doesn't need a response. And it also implies an important role of model resolution.

P4, L11-14. I don't understand this sentence, especially the use of the word "punctual," which means "on time", as in "She was punctual" -- she arrived just when we expected her to.

Maybe you should just write that observations of TKE profiles, being made only during field campaigns, are quite rare.

P5, L14, L16. Could you describe "moor" in more detail? Is that a specific kind of vegetation or mix of vegetation?

Figure 1. This figure is extremely hard to read. Need bigger range of color or lighter colors. And maybe larger size.

P 6, L7. Suggest "unique aspect" rather than "specificity."

Section 2.2. Suggest details regarding numerical models in a brief table. This helps the reader refer back to model physics (especially the PBL schemes), grid spacing, run length, beginning of runs, etc.

Table 2. Why not include vegetation type, rather than a lot of the detail here, since readers will know, for example, that "forest" has a larger roughness length and LAI than "grassland," and "forest" has a lower albedo than "grassland." And then you could include a column describing what you consider to be the land cover. Or, if not, at least you could refer back to the table when noting result mismatches due to mis-characterization of vegetation.

Also, it would be instructive to including a four-frame figure showing the terrain contours for the three models and what it really looks like.

A fussy comment: should be "grid points" not "grid-points."

Section 2.2. Also, did you run the ECMWF model or download output?

As to vertical grid points, you could put them in your profile figure to give the reader an idea of where they are.

Section 2.3
P 8, L8-10. Are you referring to Lothon et al? If so, refer to it.

P8 L11. "Clues as to" rather than "inferences on"?

Need to give conversion from UTC to local time, which is what drives PBL development.

P8, L26. Replace "an" with "a vertical interpolation"

P8 L30. Which model? All three?

Also – why don't you try using some of your observational criteria on the model profiles? In some sense, you are often comparing apples and oranges rather than apples to apples, since different criteria can give different PBL heights. (See LeMone et al. 2013 – we very rapidly abandoned the diagnosed values because they were often inconsistent with the model theta profiles).

P9, L4. Can delete "previous"

As noted in earlier general comments, a look at the paper by Lindsey Bennett et al. (MWR, 2010) might be helpful.

P9, L20-21. See "also" comment above.

P9, L23-5. Why not apply the different criteria to the model profiles to see how they relate within the model? (I.e., different criteria give different PBL depths).

P10, L3-13. Evaluation of TKE in the PBL is hard; and comparing it to TKE in the model is even more challenging.

Averages are probably too short; and aircraft high-pass filtering eliminates important scales. See Grossman et al. (1992) and Kelly et al. (1992), both in J. Geophys. Research for flux profiles. In the CBL, you should expect to see large eddies of scale of the order of 1.5-3 times the depth of the CBL; a 5-km cutoff will diminish these eddies significantly. In fact, use of such a short averaging time (and cutoff) is not consistent with the 30-min averages for surface fluxes, which are designed to capture all the fluxes.

Of course this is for capturing the total TKE. For 16-km horizontal grid spacing, this would represent the TKE from the PBL scheme, plus the TKE associated with the parameterized mass flux. For 2.5-km grid spacing, this would be the TKE from the PBL scheme plus the TKE associated with mass flux, plus the TKE associated with partially-resolved eddies.

It's not surprising that you get larger measured TKE than model TKE simply because you don't include the "MF" part (which is only w). Perhaps one meaningful comparison could be made at mid-PBL when the horizontal TKE is smallest and vertical TKE the largest.

Or you could derive a rough representation of mass flux in the PBL scheme by developing an empirical relationship between mass flux and TKE from the aircraft data (for larger scales). Not sure this would work – the relationship between w in TKE and w in mass flux in EDMF schemes is not clear to me.

A really tough but useful test (but more doable than TKE from the model point of view) would be to compare the moisture flux from the model and from the observations, since the model uses total flux divergences (at least for the two coarse-grid models; you would need to add flux by the resolved eddies in the 2.5-km grid spacing model). Heat flux also – but that is so tightly constrained that it doesn't give you as much information. (We tried this in Tastula et al. QJRMS 2015 or 2016).

It is reassuring that your TKE is typically larger than the model TKE – that is what one would expect, given the above discussion. I'd expect the discrepancy to be even larger if you used averages that included the larger scales.

P11, L 19-20. Ambiguous. I thought these four days had no clouds or a few clouds (and by implication, the other days had more clouds). Suggest rewording, perhaps like this.

Those days correspond to mainly high-pressure fair-weather conditions with no cloud cover, or, for 14, 15, 24, and 30 June, a small amount.

In Figure 2 caption, don't use "range" since range means maximum minus minimum. Suggest "black curve with horizontal standard deviations indicated by error bars" instead of "black curves … shaded in grey" since you already have nighttime shaded in gray and this avoids the use of the word "range"). I put in "horizontal" since I think that is what you mean.

Also, you should replace "variability" with "range," which is the correct label – and you have room for a bigger font, which is important. Print is very marginal in size for readability.

Finally, you need to explain the dashed lines in the figure (they are explained in Figure 3). Also, if you label your points in Table 1 with land use, it would help interpretation here as well as in the text.

P12 L1. Similarly, you don't need the "gray shading that indicates the envelope containing the different surface sites," since (a) it's described in the caption, (b) the "gray shading" is confusing, since the error bars look black on the graph and the gray shows night, and (c) an "envelope" typically describes the range (maximum – minimum).

P12 L3. What does "for a given type" mean? Don't you mean for a given day?

P12 L3. This is correct use of "range." So why not use that instead of "variability" on the right side of figure 2. This word is also shorter, so you can make the letters bigger. (I can't read it easily unless I enlarge the electronic version)

P12 L4. Suggest (no C in Figure) after "cloudy days" since you do not explain what the C means (I thought it meant "cloudy"!). Maybe – if possible, you could include a circle with cloud fraction instead of the C. That way it would be less ambiguous (since both "cloudy" and "clear" start with C.

P12 L7. Either "clear" or "cloud-free" but not "clear-free"

Figure 3.

You should repeat the labels on the plot that you put in Fig. 2; also replace "variability" with a "range" in a larger font. Also label the "hot" days, since you discuss them.

P12, L23-4 "which has similar range as observations above the forest". I am not sure what you mean by this. When you say range, are you referring to range in time, since there is only one curve? If you are referring to the difference between two forest sites in model and observations, should point out that they are not shown in the graph. Again, a vegetation type label would be useful.

P12. L29-30. I THINK you are saying that the model assumes more trees in the grid box than there actually is. That is not captured by "much more surface heterogeneities at this size." Also, reference to Fig. 1 doesn't help since you really can't see much (it might if you improve the figure). If you put surface type in the table, this would help. And perhaps label the points in the figure that you discuss in the text. (I.e., you don't have to label all of them).

P12 L31. The only gray shading I see is the nighttime.

P13 L3. Again, please label the hot days somehow on Figure 3.

P13 L17. Have you looked into the "coupling constant"?   I.e., the coefficient in the bulk formula used to calculate flux? We have found it sometimes to be off in the model when compared to the observed value.   This could account for both latent and sensible heat flux being too high, since the solar radiation doesn't look that far off.

P13 L18. What is "high vegetation"?

P13 L24. This is a new thought, so should start a new paragraph. I noted in my earlier set of comments the citation to LeMone et al., which you appear to have in the references but not obviously in the text.   As noted previously, a negative slope in the plot means a constant available energy, not a constant Bowen ratio.   For a constant Bowen ratio:

Bowen ratio = B = SH/LH. If it were constant, LH = SH/B, which would mean that the slope would be positive, not negative.

P14, L17-19. We found that wind reduced horizontal variability during the night for CASES-97 in LeMone et al. (2003). I would guess Acevedo and Fitzjarrald did as well for their data; because the BL remains coupled to the ground.   In strong winds, we found theta almost constant at night. Curious that the model didn't – but then you wouldn't get as much terrain-induced variability with the coarser-grid models. (Again, would be nice to see what the terrain looks like with the coarse-grid models).

Figure 4 caption: should note what the double vertical lines are. Did the rain occur at the same time every day, as the figure implies?

Regarding diurnal cycles for mixing ratio (bottom, P 14). It does look as though you get the morning and evening maxima at least at some sites (associated with large latent heat flux into a shallow BL). This is a good marker for the creation of the shallow PBL in the evening locally. If the terrain is complex, perhaps this happens at different times at different sites.

This feature is strongest for weak winds and strong LH.

End, section 3.2 – yes, mixing ratio is the most difficult!

Figure 5. Regarding warm and cold biases in the lowest 500 m for the models. Have you factored in differences in PBL depth? For example, if the PBL depth were underestimated by the models, the mixing ratio would be greater. (Of course, horizontal advection – and initial conditions – could also have an effect).

Figure 6. Suggest taking advantage of this figure to show where the lowest grid points are. One could do this by putting points on one profile for each of the models,

or you could mark grid points in a three columns within one of the frames – (top right figure would be excellent for this).

For 27 June, I am intrigued by the large horizontal variability even though the skies are clear. Do you have resolved PBL eddies? These can affect surface fluxes, and especially humidity and wind (also temperature, depending on PBL scheme). On strong-wind days you could be getting model rolls as well as observed rolls, which are associated with strong horizontal changes (see e.g., Weckwerth et al. MWR, 1996). Also Ching et al. (2014, MWR) and references therein.

P16L11. "Mesoscale circulation" very vague. It would be good to give a scale and perhaps a likely cause. Do you mean terrain-induced circulations? Or something larger in scale?

In Fig. 7, label the hot days.

It would be more meaningful to compare similarly-diagnosed PBL depths from observations and model. And to compare differently diagnosed PBLs internal to the model and internal to the observations.

P16, bottom. Have you an idea what causes different types of morning PBL growth? Subsidence? Strength of inversion? And as you note, different criteria can give you different PBL depths. One you might mention is RH max, which will give you the top of the PBL in the absence of cumulus clouds, but will give you cloud base in the presence of cumulus clouds. (Curiously, we have found that PBL turbulence statistics scale well with cloud base, but it can be argued that the true PBL depth is somewhere in the cumulus cloud layer.)

P17, L1. Isn't the top of the stable layer and the top of the inversion layer the same thing?

P17, L7. Better prediction for AROME makes sense to me if you include shading, which could be a factor in decreasing surface buoyancy flux, especially near sunset. Do you?

P17 L16. "unique feature" rather than "specificity."

Figure 8. Bigger font on right side. I can barely read the labels in my printed version – I'm working off my computer screen.

P17L30-31. Measured ON the evening … and IS reproduced.

P17, bottom to P 18, top. I hadn't even thought of slope winds – it's very hard to expect a 1:1 correspondence of measured to model TKE even for a horizontally-homogeneous area – but in complex terrain, it's even more unlikely to expect "agreement" except very roughly.

P 18. How do model and observed TKE compare if you count the TKE associated with resolved PBL eddies in AROME? And including some representation of the mass flux associated with the PBL scheme?

P18 L7, L14-15 While the difference in height might be a factor (at the lowest level), you would expect more parameterized TKE in ARPEGE compared to AROME because the PBL scheme in ARPEGE has to do almost all the transport (because the resolved eddies grow much more slowly for 10-km grid than for 2.5-km grid), see Ching et al. 2014, also LeMone et al. 2013). If it's an EDMF scheme, it should account for all the TKE. (Again one has to include somehow the mass flux in the TKE estimate).

P18, L19-20. Resolved PBL eddies grow during the day until saturation is reached. It could be that horizontally-averaged model TKE starts to go down as the resolved eddies grow.

Though I am obviously skeptical that you will even achieve exact agreement, it is encouraging that the trends are similar.

P19L5. "physical processes … are small ".. You mean terms in the TKE equation are small? If you don't want to write out the equation, you could write something like

"Most of the terms in the TKE equation, -- buoyancy production, shear production, dissipation – are small." ?

Or are all the terms small?

P19L15. "where the height of the reflectivity gradient decreases with time …"?

P19L25-6. This makes sense, since dissipation and TKE are closely related.

P20 paragraph 1. The earlier time at which sensible heat flux goes negative at the surface is consistent with large latent heat fluxes. This makes sense both from the point of view that more of the total energy is going into LH. But it also means that the buoyancy flux remains positive after the sensible heat flux goes negative. I would guess that the time when the buoyancy flux goes negative is also earlier for AROME, and this would be more directly related to turbulence generation than sensible heat flux. It would be good to see what a plot similar to Figure 10 for buoyancy flux looks like.

P20 paragraph 2.Because of the large latent heat fluxes, it might be useful to normalize thing in terms of buoyancy flux rather than sensible heat flux.

P20L32-P21L1, suggest ..

Models and observations produce lower sensible heat fluxes, higher temperatures, stronger winds, and weaker TKE than (what? For the other days?)

P21 L22-3. From P7, L12-14, I thought that ARPEGE had the same PBL scheme as AROME. This sentence implies there is no "MF" in the ARPEGE PBL scheme. This could be clarified by listing the PBL physics schemes in a table and describing them more carefully. If there is no "MF" in the ARPEGE scheme, then the TKE should be pretty comparable to the total (no high-pass filtering) TKE. (Though I would expect some discrepancy since pure TKE schemes don't really represent what is going on in the CBL). Please clarify.

As noted earlier, the parameterized TKE in an EDMF scheme should be smaller than measured, particularly for fine-mesh model runs (because of the contribution of resolved eddies to the TKE).

P21, end of 2nd paragraph. Estimation of some terms in the TKE budget might be simpler than the estimation of the TKE, at least in terms of direct comparison of model with observations, for reasons discussed earlier.

Supplementary Figure 2. Is this discussed?

P10, bottom. You refer to an Appendix here (which isn't part of the paper). Perhaps you should just refer to supplementary figure 2? Was there an appendix?

Earlier Review:

General comments

The paper is interesting and I think will be in a publishable form with some modifications. I include here only some major thoughts (in no particular order).

1. The SH-vs-LH graphs having a slope close to -1 doesn't indicate a constant Bowen ratio – quite the opposite, it shows horizontal variation in Bowen ratio. Rather, it shows a constant available energy. (I.e., + SH = constant) This is discussed for CASES-97 in LeMone et al. 2003 (J. Hydromet, choosing the averaging interval.) and discussed as a function of soil moisture using both observations and a land-surface model in LeMone et al. (2007). It is nice to see someone exploring this.
2. When discussing horizontal heterogeneity, terrain plays a big role. This it would be good for the authors to show maps of the terrain used in the three NWP models.
    a. This is true, as the authors recognize, because of the presence of mountain-valley circulations of tall types. The different terrains will produce different circulations.

   b. This is also true for horizontal variability. Although one gets downslope drainage winds even with gentle terrain, more extreme terrain probably has more cold-air pooling. So there might be less horizontal variability smoothed terrain.
3. When discussing TKE, the measurements will inevitably include the impact of large eddies (horizontal wavelength between 1.5 and 3 times the depth of the PBL, roughly). These are likely partially resolved in AROME, and they tend to grow in models under convective conditions, faster with smaller grid spacing. Comparison to TKE in the other two models is in some sense more realistic from this point of view, since nearly all TKE will be parameterized. For fine grid spacing (< ~4 km), the interaction between this resolved eddies and PBL schemes can exaggerate local variations in TKE (see Ching et al. Monthly Weather Review 2014 and references therein). Other issues:
   a. TKE is mostly horizontal near the surface, especially for eddies extending through the PBL (for which w is very small; from mass-continuity equation).
   b. Large eddies travel roughly at the mean speed of the wind through their depth (i.e., the boundary layer). Thus if one filters according to scale, the scale should be defined not be the wind at the level of the measurement, but by the mean PBL wind.
   c. A philosophical point (discussed in Ching et al.) is the in the "gray zone" or "terra incognita" the PBL scheme should account for all the TKE in the PBL, which for fine-grid models means several grid points horizontally, and there should be no large PBL eddies (convective rolls or cells). (This is the purists' view; the semi-resolved eddies have been useful in storm initiation or propagation – because large eddies, especially rolls, have been shown to play a role in storm propagation and evolution.). One way to look at this is by considering the buoyancy-flux profile. It should be continuous from the surface (where its value is determined by a land-surface model) up through the PBL. If one does time- or space-filtering that is too fine, the fluxes above the surface are too small. I gather from the discussion that the authors were wrestling with this.
4. The authors should look at the paper by Lindsey Bennett et al. (MWR, 2010) regarding estimates from different instrumentation, in addition to the LeMone et al. and Grimsdell and Angevine work cited.
5. Figure 10. Should look at the time when the virtual-temperature flux goes negative in the afternoon; or, similarly, how this time relates to latent heat flux (and hence vegetation type and soil moisture). If I recall correctly, the sensible heat flux went negative earlier where there was large latent heat flux, based on CASES-97 data. Which meant more variability in this time both spatially and from day to day for sensible heat flux than for virtual temperature flux.

General editorial comments (more detail later)

1. The figures are impossible to study in printed form. I am reviewing the paper with the figures enlarged on the screen. The labels on the right hand side need to be larger, and “range” might be a better label than “variability.” Also it would be helpful to the reader to label the “hot” days referred to in the text. And have labels on all the figures to make it easier for the reader.
2. It would help in the profile figure (6) to figures to have grid points on the curves for each model – for one curve for each model. Also, might consider plotting the average profile for each model and time. And finally, might consider offsetting the soundings by adding a few degrees for each time interval. The last might not be practical. (You could stretch the horizontal axis and only have one altitude label).
3. It would be useful to have a table describing the properties of each model (horizontal and vertical grid spacing, PBL scheme, etc.) as well as model-run length and initiation time and data used to initialize the model. Also, how the land-surface properties were initialized (often there is a long-spinup).

Peggy LeMone 25 January 2016