Interactive comment on "Validation of the PALM model system 6.0 in real urban environment; case study of Prague-Dejvice, Czech Republic" by Jaroslav Resler et al.

Anonymous Referee #1

Received and published: 30 October 2020

General considerations

In this contribution, the authors use the recently developed PALM-4U urban LES model at very fine grid spacing (2 m) to model details of atmospheric properties, wall/surface temperatures etc., and compare the results to a detailed (dedicated) data set from the city of Prague. This is, first of all, a tremendous achievement – with respect to the modeling effort, as well as the dedicated observational campaign. It constitutes a major challenge for a numerical model to 'reproduce' various point measurements in an environment as complicated as an urban canopy layer. Overall, the observations and also the modeling approach are well and comprehensively described (there are a number of editorial issues, though. See major comment 1). With respect to the description of the obtained results, however, the paper looses its conciseness. By showing too much material (major comment 2), the 35 (!) figures are spent on just showing and describing the outcome of the simulation, rather than trying to identifying the true potential of PALM-4U – and also its limitations. According to my judgement, one may indeed conclude from the present results that, provided the correct wall and surface, etc. parameters are available, one can reproduce even the details of the wall temperatures, etc. If it comes to the flow characteristics (within the canopy!), temperature is pretty well reproduced (rms's likely below 1 K even for the bad cases). The magnitude of wind speed is largely improved over the driving (WRF) diagnosed wind speed, but the details (timing of local maxima etc.) are essentially not captured. As for air pollution (which of course is the greatest challenge), we get no more than what we can expect from a 'good dispersion model' (minor comment to l. 817). If the major problems were addressed (major comments 3-5) rather than just describing the output, one could further assess the potential of these fascinating simulations. For example, one of the relevant questions would be to decide what has more potential: to further improve the model (parameterizations) or to assess all the necessary surface parameters at the best possible detail.

First of all, we would like to thank very much the reviewer for this very valuable detailed review with a lot of helpful suggestions.
We understand that the presentation of our results may have suffered from the fact that our analysis was a summary of two years of model development and improvement in which focusing on the details was often the key to identification of model issues and subsequent model improvement. In this revision, we tried to restructure the content of the manuscript to focus more

on condensed summarised outputs according to the suggestions of both reviewers. We also tried to extend and strengthen the discussions of our findings. We hope the revised version is much more attractive for an “interested researcher”. We updated all parts of the manuscript, and chapter 4 was fully restructured and rewritten. We tried to address all suggestions of both reviewers as described in the answers to the particular comments.

Considering the question of interest given as an example, the question of the potential of assessing the surface parameters in the greatest possible detail is very complex and we attempt to tackle this problem in a companion paper (gmd-2020-126: Sensitivity analysis of the PALM model system 6.0 in the urban environment). To address it in this manuscript, we added a short discussion to the end of section 6.2.

Major comments

1) Editorial issues. The material is not very well described. Most of the figure captions are not complete (I have added a number of specifically missing explanations in the detailed comments – but certainly not all). The authors should carefully go through all the figure captions and assess whether all the lines, symbols etc. are explained. The most important missing information for most of the figures is, which simulation (parent / child domain) is shown. Furthermore, essentially all the figures having a color code, this is not large enough to allow to distinguish the numbers (e.g., Fig, 17, 26, 28, ..). Finally, a lot of material is provided in the supplementary material (which is good in principle) but should be better identified (see detailed comment to l. 144)

We went thoroughly through all figures and their descriptions and we considered again the graphical issues (fonts, lines, colors) as well as the completeness and comprehensibility of the description and fixed all issues we found. The supplements were restructured to provide better orientation in their content (see detailed answer after comment l. 144).
To improve the quality of the presentation, the revised manuscript also underwent language revision by a native speaker with a good knowledge of the topic.

2) In fact, the authors are showing much too much material. Even if they have already put quite substantial part of the examples to the supplemental material, it remains much too much what is being presented. The problem with this is that it produces lengthy descriptions of each example (where and when which variable shows under/overestimation etc.). Even an interested reader will get lost with all the unnecessary information (e.g., on the date, site, etc.), which of course has to be provided when showing all the individual examples. But as an ‘interested researcher’ I am not interested in learning whether PALM underestimated the daily cycle on July 15 (and not on July 16) in the afternoon, at site 11 while the underestimation was not so great at site 07 on the same day, but with slight differences on the next). I am interested in learning i) the typical (average) behavior and ii) the exceptions. Thus, rather than showing so many different sites (e.g., for wall surfaces it is Fig 13-15 and associated text, and Fig 18-21 plus associated text), one should seek a form to best characterize the mean performance and show one example and then provide statistics on the other sites (the reader could then see what is presented in Fig. xx *and judged 'average performance', and learn that this figure is associated with a mean bias of yy and an rms of zz (whatever the statistics are to best characterize the behavior). This would leave more room (and attention) for the exceptions – e.g., for the wall temperatures there is this example with 'old buildings' having been insulated in the meantime (since the data base has been produced) and thus producing wrong wall heat fluxes and consequently temperatures.*

We are aware that there is too much material for one paper. It represents an important part of more than two years of work of a sizable group of researchers. We considered splitting the material into two papers at the beginning of our work on this manuscript but we finally decided to publish it all together as we considered all this information interconnected, and we felt that the publication as one body of work allows the relationships to come into view..

To address this comment, we completely restructured former chapter 4. We split the chapter into two chapters "Evaluation of model simulation setup" and "Results". The evaluation of the correct model setup was improved and extended by analysis of spatial development of the urban boundary layer. In the results, we removed most of the information about individual locations, with temporal graphs of quantities in individual points from the manuscript and we moved them to the supplements, as we still consider them valuable information, especially for the model developers. We also removed most of the corresponding text in the manuscript and we kept only carefully selected examples. Instead, in order to show the average performance, we supplemented the statistical metrics of these quantities in the form of tables and scatter plots. We also added corresponding discussion of the findings following from these metrics. The examples were selected as cases which require either special analysis or which highlight specific model issues, also taking into account the detailed suggestions in both reviewers comments and we have rewritten the discussion of those cases. In the case of surface temperatures, the analysis and discussion include modelling of grass surfaces, modelling of walls of contemporary office buildings and the issue of glass walls, the analysis of the "strange peaks" of modelled surface temperature, and discussion of the surfaces influenced by the plant canopy. We restructured, condensed, and clarified the section about the effects of the discretization. Similar changes were done for other evaluations (wall heat fluxes, street canyon observation, roof observation) with the goal of providing a more condensed and focused text. This way we hope to show both the average performance as well as some interesting individual cases.

As for the issue of insulation in old buildings, this was an improper formulation on our part. The buildings were insulated at some point in history before the data collection. The problem in the results was probably that the insulating efficiency of this layer was underestimated in the model input data. We reformulated the text to express it clearly.

3) To some degree, the authors suffer from a quite common weakness in 'model validation studies', i.e. that they find everything 'well reproduced' (in the conclusions, for example, 'all is well represented' (lines, 840, 843, 848, 856, 862). Even if some 'exceptions' are usually mentioned, this is not really helpful. Most often, one learns much more from the cases, in which the model fails. Examples of issues that could/should be addressed are: the 'opposite wind direction' (major comment 5); the strange peaks in wall temperatures (e.g. Figs 13, 14, 15, 18) –

which might be related to the wall heat flux (Fig. 30); the wrong attribution to surface characteristics (see the above insulations issue); and the reflection (non-Lambertian reflection missing in the parameterization). In all cases, the problem is 'mentioned' (sometimes with a hypothesis why it occurs, sometimes not). If possible the origin of the failure should be discussed – and also some statistics should be provided on how often (how strongly etc.) this occurs.

We agree with the reviewer that this formulation isn't too helpful to the reader. While not wishing to excuse this error, in our defence we would mention that we were working hard on model development and improvement, along with finding the proper configuration and obtaining input data and parameters for nearly two years, and many of the "well reproduced" quantities were not well reproduced at the start of the project; but this is looking at things from the perspective of a model developer. To address this comment, we went through the text and we tried to restructure it and to remove unnecessary elements and statements. As outlined in the answer to the previous comment, we added parts where the mean performance of the model is discussed and averaged statistics indicating the bias or the scatter are provided. The presentations of special cases (grass, walls with complex structure, plant canopy effects) were improved and condensed. We also refined the text about non-Lambertian reflection and improved it by discussion of its practical influence on the results. This particular observation campaign does not provide sufficient data for "hard" results supported by corresponding statistics in this particular regard, so we added only an estimation based on the available observation data. We also added an analysis and discussion of the "strange peaks" in section 5.1.5 to address this part of the reviewer's comments. Our revisions address the presentation of the wind on FSv and its discussion, as detailed in our answer to comment 5). We added statistics measures for the FSv rooftop observations, as well as for street canyon measurement vehicle observations.

4) The authors provide vertical profiles in comparison to a radio sounding – but the sounding is not within the modeling domain. This, in principle, is a 'no go' in a model validation study. Either I have data that is appropriate for a comparison, then I compare it - or I leave it. In the present case, I have no clue, whether the differences are due to a model deficiency or just due to different locations being compared (and the differences are substantial, e.g. Fig. 9 for wind speed 20.7. 00:00). Thus the options are: only show the model in order to characterize the situation. Alternatively, the coarser model results (WRF) could be used to estimate the height, above which the grid point-to-grid point variability 'vanishes' (i.e., the model results are more or less spatially representative over the larger 'urban' area) – and then make the comparison only down to that height.

We agree with the reviewer that the different location of the sounding and modelling domain (even just a few km) is a serious problem. On the other hand, we still consider the information provided by the soundings valuable to some extent, as the distance between the locations is small and the surface conditions are similar because both locations are situated in a similar urban complex. In the revised manuscript we emphasize this fact and discuss possible implications so that the reader is aware of this issue. Moreover, we connected this comment with suggestion 2.1) of RC2 and we supplemented the graphs with the WRF profiles in both locations, that is the sounding location as well as the location of the modelling domain. Further, we unified the different sets of comparison graphs - sounding vs. WRF profiles (Fig. 6, 7) and sounding vs. PALM profiles (Fig. 9, 10, S13, S14, S15) - into one series of graphs (FIg. 5 and 6 in the revised manuscript) which provides complex information about vertical profiles; we hope they give the reader valuable and clearly arranged information. We updated the text description accordingly.

5) Comparison to wind speed at the highest building (referring to l. 510: 'This confirms the disagreement of wind speed'): if I understand Section 3.5 correctly, this analysis uses the nearest grid point in the PALM model (and we would be interested to learn which domain, see minor comments). When I check the wind direction at the times when PALM has what the authors call an extraneous (but probably mean extraordinary) peak (e.g., afternoon of 21.7., morning of 4.12., etc.) it is always when the wind direction is plus/minus opposite (or at least largely deviating). This brings us back to the (minor) comment to l. 232: even if wind speed was measured on the tallest building, if the observation is close to roof level, the flow will be influenced by the building itself. The flow impinging on the building will detach, form a wake, re-attach etc. – the details of which depend on the building structure, the flow of course – and the direction (see, e.g. Oke et al. / Climates of Cities / Chapter 4, e.g. Fig, 4.3). If the grid spacing is 2 m, all this is resolved by the model. If the overall (synoptic) wind direction is correct, the model might still not capture the details of the near-roof flow structure. I consequently think that i) this type of situation deserves a more in-depth analysis and ii) the reason for the wrong wind direction needs to be found. If we take, e.g., Fig S17, 27.11.: the wind direction is almost constantly 'opposite'. If this is not wrong already in the WRF simulation (what I cannot imagine), it probably has to be a local (very local) type of reversed flow or similar close to the roof. A first step therefore would be to compare the WRF wind direction (and compare to the synoptic) over these periods of 'wrong wind direction' in PALM.

As in the original version we considered that the wind observation on FSv building is only a supporting information and not part of the main focus of the study, we did not discuss it in detail; the focus of the study is the street canyon level comparison. To attempt to address this issue, we have dived deeper into the problem. At first, we tried to decide whether this effect (which appears only during winter episodes) is a technical problem or some microscale dynamic effect. We went through the entire process of the measurement and data collection with the technicians, and we checked all our postprocessing tools. According to the technicians, the placement of the sensor and the settings of the equipment were identical for summer and winter episodes. However, we found no photo documentation of the sensor for the winter episode similarly as for the summer. We thus were not able to conclusively check and prove the actual orientation of the sensor in the winter study. We also tried to compare wind direction data from nearby synoptic stations as well as from WRF and PALM 3D fields. Despite all this effort, we came to no solid conclusions above some speculations. As we cannot exclude the possibility of a technical issue (the wrong orientation of the sensor) during the winter episode, which would reverse the wind direction, we decided to exclude the wind direction from the presented FSv

wind observation and we updated the description. The potential misorientation of the sensor would not influence the observed wind speed so we consider this approach as the correct one.
We are aware that there are recirculation zones present behind sharp corners on bluff bodies. The grid resolution in the inner domain was too coarse to simulate them accurately but the sensor location at 2 m above the roof in the centre of the building should be well outside of them. Still, the flow is certainly strongly affected by the building and an accurate simulation of the details of the local flow would require higher grid resolution.
We consider this issue a very interesting and important problem and in new projects we are starting, we intend to repeat observations of similar type using an improved design.
To improve the presentation of the wind on FSv and its discussion, we first reworked the wind graphs themselves; the wind magnitude graphs are taller and the wind values from WRF and nearby synoptic station Ruzyne were added. We added statistics for this observation and we also improved the description of this part of the study to express clearly that we do not consider these observations as the observations of the free flow but as the observations of a flow influenced by surface effects, but at a higher level than the street canyon observations. To emphasize this, we moved this part from section 4.2 to section 5.3.3 “Wind speed on the roof” in the revised version.

Minor comments

l. 11 for certain. . .

Accepted.

l. 36 the UHI effect. . ..

Accepted.

l. 83 The Prague Dejvice. . ..

Accepted.

l. 116 The majority. . .

Accepted.

l. 117 The lancover map. . .

Accepted.

l. 143 . . .by the infrared camera: only one? Or rather ‘by infrared cameras’? (if only one: by an infrared camera. . ..)

In fact, measurements were done with two IR cameras as a precaution, but we utilized the results only from one camera in this study. We adjusted the text in the l.143 accordingly.

l. 144 Tables S2 and S3: after having found those tables in the myriad of files provided in the supplemental material (it would probably be helpful to somewhere state that all the supplemental figures and tables can be assessed by following the 'index link '), the tables contain a number of unexplained acronyms (what is MV, HF, AQ, etc., also LCZ is not explained [can be guessed when reading the caption. . ..])

We looked for the optimal way to organize the supplements and we found the html approach better than all other approaches considered for this particular case. To ensure that the reader does not overlook the index.html file, we restructured the supplements in such a way that the main directory contains only this index.html file and the directory ".content". We also refined the structure of the sections and we improved the captions. We added explanations of MV, HF, AQ, LCZ, and other abbreviations in supplement tables. For LCZ, the explanation cites a reference to a paper paper with a more detailed description (Stewart and Oke, 2012).

l. 167 what is the TYRSY01 system (not introduced)

The previous sentence referred to the next subsection, where heat flux measurements and TRSYS01 were described. We reformulated the first sentence to make this more clear.

l. 229 proposed? Rather 'positioned'

The text of the paragraph was reformulated.

l. 232 60 m high: more important is the height above roof. Even if the building where the sensor is mounted is high (the highest), it will by itself impact the flow. Also, in Fig. 1 the location is not labelled 'FSv' (but 'anemometer'). See major comment 5

The text was reformulated - we no longer claim that the measurement represented above roof wind flow. Information on anemometer height above the rooftop (2 m) was added. The label of anemometer in Fig. 1 was changed to "rooftop wind FSv." Corresponding changes were made in Table S1: "FSv (above-roof wind)" -> "FSv rooftop wind", table S2 and S3 "Wind (above roof)" -> "FSv rooftop wind.

l. 235 of the inner domain

Accepted.

l. 236 in Fig 1 the 'drone site' is labelled 'autumn only'

Accepted. It was supposed to inform the reader that drone measurements were done in autumn only. The same remark is made for some mobile measurement locations e.g. “18 (12 July)”. But we renamed the drone location in the map from “drone (autumn only)” to “drone”.

l. 246 The radiosonde. . .

Accepted.

l. 254 The PALM. . .

Accepted

l. 272 the stand-alone. . .

Accepted.

l. 297 in Fig. S1

Accepted.

l. 318 wherever possible

Accepted.

l. 322 The Prague. . .

Accepted.

l. 326 using data from the terrain mapping campaign

Accepted.

l. 336 described in Tab S6

Accepted.

l. 345 initial and boundary conditions. The authors only provide the source (WRF) for the BC, but not the type of BC (also: at the domain top). Furthermore: how are the 49 WRF levels mapped to the PALM levels (especially within the canopy)? For example, the lowest WRF level (probably at 10 m [?]), is in the NOAH LSM meant to be above the canopy (and a bulk canopy approach is used), while in the lowest PALM level it is ‘between the buildings’. The same probably is true for the second lowest level. Finally: how are the BCs interpolated in time?

While this review process has been ongoing, our scripts for the transformation of the WRF and CAMx outputs to the PALM initial and boundary conditions ("PALM dynamic driver") were included in the PALM official distribution and they are now available in the PALM SVN repository under the UTIL/WRF_interface directory together with a brief description of the principle and utilization. We added this reference to the text together with a brief description and we added a more detailed description of the transformation process to the supplements. We also were able to add a reference to the companion GMD discussion paper about the PALM mesoscale nesting (Kadasch et al.(2020), in review) which has meanwhile been published. Moreover, we added further details regarding temporal interpolation in the mesoscale nesting, as well as which quantities are nested.

Fig. 3 caption: what are the green and red rectangles?

Figure 3 and its caption were updated to provide all needed information.

l. 423 the 'weather balloons' are usually referred to as 'radiosondes' (later occurrences)

Accepted.

l. 462 (i.e., caption Fig. S16): first, it should be mentioned that 'modelled' refers to the CAMx simulation (that's at least what I assume). The reference to Fig. Supp13 should likely read Fig S10. How is the 95% confidence interval determined (from the distribution of all the sites and days?)

The caption was corrected. The 95% confidence interval of the mean is calculated by the bootstrap technique as described in the timeVariation function from the openAir package (https://davidcarslaw.github.io/openair/reference/timeVariation.html). Data from all the stations corresponding to the particular hour are used for its calculation (the same holds for model data). The main text was reformulated slightly to make it clear that also graphs of diurnal variation were made by the openair package.

l. 465 morning and evening peaks are. . . and appear. . ..

Corrected.

l. 466 than in observations. . . .. But the CAMx model. . .

Corrected.

l. 468 play a more. . .

Corrected.

l. 473 at the southern edge. . .. . . with the model simulation

Corrected.

l. 474 necessarily to match

Reformulated.

Fig. 9 apparently, the modelled results are averages from the parent domain (caption). When explaining (in the text) that the comparison is 'not fair' (because the sounding is from a different location) this should be mentioned there, too.

Information was added to the introduction of section 4.1 and to the figure captions (FIg. 5 and 6 in the revised manuscript). The figures were reorganized and reworked using both WRF and PALM profiles (see answer to major comment 4) and the corresponding discussion was updated.

l. 482 is much smaller: indeed, not only much smaller, but also has a completely different shape. For the midnight sounding, the maximum wind speed is 10 ms-1 at about 800 m height, while PALM has very low wind speeds. This must be detectable in the WRF simulation (PALM can likely increase the drag, but I wouldn't expect this to be so dramatic. Can the authors comment on this?

This problem was due to a technical issue in processing PALM output profile files and was corrected in the revised version. The profile from the WRF simulation was added to the figures for comparison (original Fig. 9 and 10 correspond to Fig. 5 and 6 in revised manuscript) and the text was restructured. In the corrected version the PALM simulation shows better agreement with WRF higher above the surface.

l. 486 these discrepancies (or this discrepancy)

Accepted.

Fig 11 caption: state, which side is the summer and which the winter episode. Also, it should be stated which simulation (parent, child) is used here.

Caption updated accordingly.

l. 504 'extraneous': the authors probably mean extraordinary? (same on l. 508, 509, ...)

Removed.

l. 509 is somewhat overpredicted

Missing word added.

Fig. 12 caption: south-west (not west-south). Also, the colour scale is hard to even detect (nothing to say about being able to read the numbers)

Text fixed, scale and numbers enlarged.

L 516 of about 290 K

Accepted.

Fig. 13ff what is the grey shading? (day/night?)

“The grey areas denote the night time.” was added to the figure caption.

l. 551 when the surface temperature. . .

This text disappeared during the restructuring of the chapter.

l. 554 two sharp peaks: these are 10 to 15 K (which is quite substantial) – and it also occurs on the day before. Can the authors offer an explanation for these peaks? (if we look at Fig 8, it does not seem to be a radiation effect. . ..)

The analysis and discussion of these peaks were added to the chapter 5.1.5 “Rapid changes of surface temperature”

l. 555 it is striking to note. . .. Here the horizontal surfaces are probably referred to.

This text disappeared during the restructuring of the chapter.

Fig. 17 soil moisture panel, inlet: this is definitely too small to be distinguished (in fact, all the inlets are too small)

All graphs of the surface temperature were reworked, the inlets were enlarged and the corresponding figures were recreated.

l. 595 was adjusted: based on what?

At this point we must confess that we were quite unspecific. Soil moisture values at this level of detail were not available. Hence, we categorized grass surfaces into natural grass surfaces and urban-like grass surfaces with and without irrigation. Adjustment factors are only based on a best guess resulting from the survey of the locations and our personal experience. In the revised manuscript we explicitly note this.

l. 666 the first two examples: in the caption of the corresponding figures (also Fig 24) it would be helpful to explain what influence is being shown (the identification of the site number is not extremely helpful for the reader who does not know all the sites).

Figure 22 disappeared during revision, figures 23 and 24 were combined into one figure and the caption of this new figure was enhanced by the description of the location.

l. 690 The terrain. . .

This text disappeared during the restructuring of the chapter.

l. 712 Two further consequences. . .. are an altered. . .

Text of section 4.3.5 was restructured, the comment was taken into account.

l. 720 due to their local nature

Accepted.

l. 731 with the modelling episode

Accepted.

l. 736 the wall temperature: shows again – as in Fig. 14 for site 12 – these strange peaks. These are also visible in the heat flux. Could this be a starting point to evaluate the origin of those peaks?

The analysis and discussion of these peaks were added to the new subsection 5.1.5 “Rapid changes of the surface temperature” (compare also the answer to RC2 p. 2.2). This analysis also explains the peaks in wall heat fluxes.

Fig 32 caption: the yellow band. . .: this is only visible in a few of the sub-plots. Also, there seems to be a black dashed line associated with the yellow band, which is not explained (all the same for Fig. 33)

The thin black dashed line shows the original 10-minutes averages of the value in the gridbox corresponding to the point of the observation. This was explained in the figure caption but the word “dashed” was omitted by mistake. This omission was fixed.
The yellow band represents the spatial variability of the 10-minutes averages over adjacent grid boxes. The problem of its absence in some graphs lies in the fact that the local spatial variability of the variable in some places is very low and its low spread can lead to the situation that the corresponding band is overlaid by the black dashed line or by the thick solid line representing one-hour moving average of the values. This seems to be also valid information. The band is

only supporting information which gives the reader an estimation how much the value can be influenced by the spatial error caused e.g. by imperfection of the discretized position. To make the band morevisible, we reworked the graphs selecting another colour scheme and thickness for graph variables. We also mentioned the small spread of temperature values in the corresponding text to give the reader relevant information.

l. 791 in accordance. . ..: overall, the most relevant observation, i.e. that (diagnosed) wind speed from WRF largely overestimates magnitude and amplitudes in the time series' (at all sites, all episodes) is not even made.

This fact is connected with the configuration of WRF without special urban parameterization. It is noted and discussed in the revised manuscript.

l. 800 when concentrations. . .

Accepted.

l. 806 . . .aerological soundings at 06:00 UTC 21 July (Fig. 33): I don't think Fig. 33 shows results from the aerological soundings – nor does it show a (particularly strong) underprediction.

Correctly should refer to Figure 5. The reference was fixed.

l. 817 . . .fulfil the criteria for dispersion models. . .: it must be noted, however, that the 'dispersion models' of Chang and Hanna are simple 'Gaussian' or hybrid plume dispersion models. Thus with an immense increase in computing time and effort, the results are not convincingly better. Can the authors comment on that?

There are many sources of uncertainties in a real urban atmospheric dispersion scenario. COST ES1006 Model evaluation case studies: approach and results (Baumann-Stanzer, Trini-Castelli, Stenzel (eds.), 2015) thoroughly examined the differences in the accuracy of dispersion modelling (for accidental releases) for well-controlled (wind-tunnel) dispersion scenarios and showed that statistical metrics consistently improve with more complex models (Gaussian, Langrangian with diagnostic flow, CFD = RANS and LES). However, for real outdoor dispersion problems other sources of uncertainty can dominate as demonstrated by other scenarios in the same document. In such cases the advantage of accurate flow simulation by a more complex model can be diminished.
We should stress that the analysis in our paper compares spatio-temporally paired values in geometrically-complex locations. The original requirements on sufficient models by Chang and Hanna mostly originate in an intercomparison by Hanna (2000) which considers several scenarios, where however either arc maxima, or one-hour mean concentrations are compared from releases with well-known emissions, to lessen the effects of natural variability in turbulent dispersion. In our study the comparison is much more strict as 10-minute averages in the original submission and now 1-hour averages in the revised manuscript are paired in time and space and there are significant uncertainties in the emissions.

More complex models, such as LES, can reveal more information about the contaminant dispersion, as also shown by, e.g., COST ES1006 Model evaluation case studies, which also considered other concentration percentiles and peak concentrations or the statistical distribution in puff dispersion. These advanced quantities were not evaluated in our study, but they could be obtained from the simulation results.
In short: although similar criteria of success were used, the quantities compared are more difficult to simulate and make larger demands on the more complex model. The real outdoor dispersion without a single well-controlled emission source contributes to the uncertainty that has to be expected.

We added this short statement to the manuscript: “Although these criteria were developed for simpler models, they are applied to a more complex problem here and are good indicators of usefulness for purpose.”

Fig. 35 what are the numbers in the inlet (identifying the ‘+’ signs) referring to?

Numbers refer to mobile measurement locations in Fig. 1. This description was added to the caption of Fig. 35.