I went over the complete manuscript today.
1.) I appreciate the systematic nature of sensitivity tests.

Thank you for the thorough review and many helpful suggestions. The systematic approach was in fact our main goal here. As commented on in various places, the aim was to provide an extensive sensitivity study and therefore some aspects of general model evaluation are not discussed here. For that we refer to either a previous paper by Resler et al. (2017) or a companion paper by Resler et al. (2020) also in review for this special issue.

2.) I miss a discussion on how general your results are. Do you expect similar results for other neighborhoods?Which of the observed phenomena in your specific test case would expect to occur in other cases as well.

We added discussion in the first paragraph of the Conclusions section. In brief, the transferability to other locations largely depends on the type of analysis required. Neighborhood-wide response will most likely be comparable, however, when analysing very local influences, the sensitivity may vary considerably, not only in values, but at places even in the sign of the response.

3.) In the model limitation section you talk about missing model components (like variable building temperatures). In my opinion, the model system already includes many processes. What is your general strategy in demonstrating that the model produces plausible results.

We realize that calling the section "Model limitations" is a bit misleading, as some components are present in the model but not used for the study (discussed for each point, most of them can be neglected in the study settings). We renamed the subsection Study limitations and changed the formulation a bit (we also reorganized the Discussion and conclusions section to have a more logical structure).
We added a paragraph at the end of the limitations section clarifying the issue of validation, which is beyond the scope of this manuscript, however, extensive validation has been performed in another study by Resler et al. (2020) currently also under revision for this special issue to which we refer the readers.

About the presentation style:
My biggest problem is that you show many plots, but they are often not explained in the text. They are often only referenced, but not explained in detail.
This also implies that it is not made clear what you want to demonstrate with each figure.
This gives the reader a bit the impression that the figures are randomly selected.

We went thoroughly through the text and reorganized many of the figures to better capture the message of the paper (see below for individual details) and added discussion where appropriate. Among other changes (described below as replies to specific comments) we:

- Joined Fig.1 and 2 into one
- Moved Fig.3 into the supplement
- Moved Fig.10 into the supplement
- Joined Figs. 11 and 12

Would it make sense to create some summary plots where the sensitivity of a few selected variables (like average temperature difference) is shown for all scenarios. Fig. 17 in doi:10.5194/acp-14-2713-2014 hopefully makes clearer what I mean and can serve as an inspiration.

We added a summary boxplot for air temperature (Fig. 3) in the beginning of the Results section (added other variables in the supplement in order not to make the manuscript even longer).

Minor comments (copied from the annotated PDF)
L. 20: redundant stop
Corrected

L. 43: missing references to SOLWEIG and RayMan models
Added

L. 102&104: missing Sect. in cross reference
Corrected

L. 120: window model not mentioned in previous text
Reformulated (window treatment in the model is a part of the Building Surface Model - BSM, which is described in previous text).

L. 127-130: What is the motivation behind using two overlapping runs? Do mean partially overlap or do they simulate the identical period? How are the two horizontal resolutions applied?
Reformulated.

Boundary conditions for PALM are generated from the inner 3km WRF domain, the 9km outer domain serves as a standard way to deal with a large resolution jump from the global analysis to the regional model (double-nesting).

L. 135: misplaced parentheses in citation
Corrected

L. 219: missing citation

Corrected

L. 219: stop instead of comma
Corrected

L. 227: In the results section, you use PM2.5 as a proxy for air quality. But without chemical modelling and treating only the dispersion as a passive tracer, I am not sure if a link to air quality should be made. At least it should be stressed in the text, that only dispersion of the pollutant is considered.
We added a short clarification of this and referred the reader to the complete discussion of this issue in the "Limitations" section.
Fig 3.: What does EPSG: 32633 mean?
Clarified (Projection: WGS 84/UTM zone 33N)

L. 245: is SA used for the baseline simulation or the 21 sensitivity studies? I assume for the latter. Then move the text inside the brackets one line up.
Corrected

L. 288: Why do you show this figure as it treats a rather unimportant parameter variation? Why not choose instead to show the differences in T and MRT for SA1,SA2 in a style similar to Fig. 5.
Moved the figure to the section discussing surface temperature. Added more discussion of the relevance of the figure (briefly: it may not have a huge influence on average, but is important locally in green areas or their vicinity).

L. 294: at around
Corrected

L. 301: not clear what "higher absolute values" refers to
Reformulated

L. 307: It is not clear if this holds for all SA scenarios or only SA1 and SA", which were discussed in the subsequent paragraph.
Corrected

Fig.6: What is the information gain of the second panel?. Are there any characteristic differences between those two panels that are explained in the text? If not, you could remove one panel.
The main point here is that the responses are highly heterogeneous and non-symmetric for symmetric scenarios, mainly in the spatial distribution. We added some discussion on that to the text.

L. 324: Do not leave it to the reader to find out which row of Fig.9 shows which case. Again if you show four panels, then a minimum amount of description is needed in the text.
Added description to text.

L. 329: This figure shows 6 different lines. Again, the description in the text is cursory.

Added description to text, figure moved to supplement.

L. 330: If you increase resolution, variables are allowed to vary on smaller scales. Thats agreed on. But what is the added value? Are those more refined patterns really better? Do you know from a comparison with observations, that the higher resolution really leads to an increased prediction skill?

We replaced the "added value" with "importance" here, as we understand that talking about added value is misleading as it has quite a firmly established meaning to which we cannot attest in this study. The design of this experiment was semi-synthetic and was not intended for performance evaluation and comparison of the results to observations. For validation experiments, we refer to our previous validation study (Resler et al., 2017) and the one in review for this special issue (Resler et al., 2020).

Fig. 7: If you intend to show this figure in the manuscript, its content should be explained with more than just a few words.

More discussion was added in the text for this figure (now Fig. 5) and the relevant Fig. 4. We also replotted the figure to show differences of the scenario to basecase which allowed the figure to be shorter yet (hopefully) more intelligible.

Fig. 8: the y axis label says PM10 not PM2.5
Moreover, the caption does not explain the solid and dotted line styles.
Again: If you intend to show this figure in the manuscript, its content should be explained with more than just a few words. I can only see random time series and do not understand what you want to convey with this figure. You have to explain it in the text.

(Now Fig.6) We understand that the choice of the figure was not ideal, we replaced it with a better example of spatial and temporal heterogeneity of the response in the albedo changing scenarios. Also we added a description in the text.

Fig. 9: it would be better include this information (*point designations*) as a legend in each panel. Otherwise you have to scroll down. Or use a 2x2 placement of panels.

(Now Fig.7) Added point designations to the panels. We discussed using the 2x2 panel, but unless the image was rotated, the 2x2 placement would lose a lot of detail.

L. 455: Is it model limitation or a limitation of the study design?

That's a great point, we changed the section to "Study limitations" to cover both aspects, changed the formulation slightly in the intro and added a paragraph at the end of this subsection discussing the plausibility of the results given all the limitations mentioned.

L. 460: That's the kind of discussion that I missed and tried express in a previous comment.

We added the short mention in the previously mentioned paragraph and referred the reader to this discussion as we feel it is better to have the detailed discussion in this section than disrupt the flow of the text in the experiment setup section.

L. 483: Unnecessary appendix
Removed

Second round reviewer's comments

Although some improvements have been made to the mansucript, I would further encourage the authors to
a) better explain the rationale for selecting the scenarios analysed

We clarified the motivation behind the scenario selection for both types of scenarios, although we feel that this comment was perhaps more aimed at the B-type scenarios (the A-type scenarios are simply selected to cover all important material and surface parameters present in the model, so the choice there was obvious). The B-type scenarios (renamed *urbanistic*) were designed together with urban planners from the City of Prague in the framework of the Urbi Pragensi project, not necessarily to be realistic in all cases, but to provide them with some sort of "envelope" response, or upper and lower limit of the possible developments in the neighborhood (both "in" and "contrary" to the direction of UHI mitigation). We added the clarification to the introduction of the sect. 2.7.2.

b) reduce the number of figures and make the analysis more accessible to the reader

- Joined Fig.1 and 2 into one
- Moved Fig.3 into the supplement
- Moved Fig.10 into the supplement
- Joined Figs. 11 and 12
- Added more discussion to the figures, that were perhaps a bit neglected in the previous version

c) better highlight the added value of LES modelling and put findings in relation to existing studies evaluation urban planning scenarios.

The added value, as far as the established meaning in the modeling community goes, cannot be really corroborated by our study as we didn't do any extensive evaluation of the model results against observations, except preliminary checking that the baseline simulation produces realistic results in line with the previous paper of Resler et al. (2017). However, in the discussion we added references to other studies that validated LES and RANS against wind-tunnel

measurements and all conclude that LES shows much better performance than RANS.

L323: not clear. What are "materials used in reality"?
    clarified

Section 2.7.1 and 2.72 – the description and rational for the design of scenarios is still not entirely clear. While "synthetic" scenarios are designed to understand the impact of parameter uncertainty, the real life scenarios are evaluating planning strategies. However, more details should be provided explaining why the specific scenarios are selected. For example, why would you test a scenario without any vegetation? Is this considered a "worst-case example"? Maybe the term "real life" scenario is not suitable if the setups are actually not all realistic such as when removing roads but not the cars.
    As mentioned previously in the reply to point *a)*, we discussed these scenarios with the Prague City authorities and the aim was not always to analyze super-realistic scenarios, but to provide some sort of upper/lower limit assessment (e.g. removing all trees, changing all surfaces to asphalt etc.). We agree that calling these "real-life" was a bit of a stretch, so we changed it to "Urbanistic scenarios".
    The other point about removing roads but not car emissions was in fact motivated by the fact that changing both at once could potentially mask the influence of vegetation on dynamical drivers on air quality, when the car emissions serve more as a tracer as we do not consider any chemical reactions.