1. **'I have a major concern regarding the use of only one LST image for model evaluation. Former model evaluations that used LST tried to acquire as many images as possible for one period. The argument that Landsat 8 images is more representative of the heterogeneity of the surface urban climate is receivable but not sufficient. These heterogeneities are also captured by MODIS data for instance, although smoothed by the lower resolution. Bechtel et al. 2019 who did a somparative study of LST in different cities using LCZ, MODIS and Landsat data found out that there was a great spatial correlation between the two. I believe they should be used jointly. Hu et al. (2014), Wouters et al. (2016) and Brousse et al. (2020) also used MODIS LST for evaluation of their model simulations. I would therefore like the authors to include such additional evaluation before acceptance of the manuscript.**

We understand the concerns regarding the limitations of using only one satellite image for this study. As explained in the manuscript, no further high resolution Landsat 8 images were available during the summer campaign period. Given the timescales involved in acquiring and analysing LSTs from alternative sources such as MODIS, and that I have now completed my PhD, this would be outside the scope of feasibility for this study. Furthermore, the information lost by smoothing the high resolution model output for comparisons with coarse resolution satellite data would considerably detract from the novelty of this particular study; an investigation of Beijing's neighbourhood-scale UHI characteristics. Additionally, as explained throughout the manuscript, given the inherent differences between LSTs and air temperatures, we are not expecting a high correlation in their spatial variabilities. The comparison is rather merely used as a guide and to highlight both the successful use of LCZ and OSM data to capture fine cool spaces and to help understand the air temperature-LST differences through the sensitivity studies. Given all of these reasons, we will not be adding further to this section of the manuscript.

2. **I still did not understand the rationale with having air temperature measured at the Pinggu station and upwind from the airport. The reasoning with the wind measurements being WMO standard makes sense. But then why not simply use the air temperature from the airport itself as it can be considered as rural too?**

We thank the editor for this comment. As explained in detail in the manuscript, rural air temperature measurements are perturbed in the model due to spatially varying urban land use characteristics and anthropogenic heat emissions. These rural measurements are ideally recorded at a site located upwind of the modelled urban area so that the air temperature sensor is not strongly influenced by the advection of urban heat; the urban heat island phenomenon is best expressed under stable conditions that inhibit the mixing of urban and rural air. The airport is located within urban Beijing (within the Sixth Ring Road); thus, the local environment is likely to be heavily affected by the surrounding urban surface thermal properties and nearby AHEs. Therefore, the use of air temperatures measured at the airport was not considered appropriate to drive the model. To clarify this, the manuscript is updated on Page 5 Line 197:

"…an upwind measurement is desirable so that the air temperature sensor is not strongly influenced by the advection of urban heat. Air temperatures measured at the airport site were not deemed appropriate for this study given the local environment is likely heavily affected by the warming effects of surrounding urban surface thermal properties and nearby AHEs"

3. **The implications on the results from adapting the surface resistance to evaporation for LCZ 4, 6 and 9 need to be discussed.**

We thank the editor for this comment. We lower the surface resistance to evaporation values of LCZ classes 4-6 and 9, relative to the 'fully' urban LCZ classes, as they are described by Stewart and Oke (2012) as consisting of significant pervious land cover. Assigning a surface resistance to evaporation value of 200 s m^{-1} to all urban LCZ classes would lack physical sense. In Sect. 3.2 (Lines 455-465) we describe how the spatial heterogeneity in urban surface resistance to evaporation values produces greater spatial variability in modelled urban near-surface air temperatures and thus results in a stronger correlation with the spatial variability of LSTs. The high urban moisture case reduces the spatial correlation between modelled air temperatures and LSTs but increases the measured-modelled air temperature agreement at the IAP site. As explained in detail in the manuscript, this likely reflects the inherent differences between LSTs and air temperatures. It is unclear what further discussion the editor is requesting here.