Response to Anonymous Referee #2

(Note: Reviewer comments are listed in grey, and responses to reviewer comments are in black. Pasted text from the new version of the paper is in italics.)

Dear authors, the paper is well written and clearly structured, however I would recommend a number of major changes in order to be suitable for publication. Please find my comments below:

We thank the reviewer for his/her thoughtful and valuable comments. These comments substantially help to improve our manuscript by addressing these issues.

General: I see a general problem in the definition of the scope of the study. A 'before human settlement' scenario should not consider emissions at all and further describes a period about 100-150 years ago which means that you would also have to consider a different climate period, land use etc.. I definitely would recommend to redefine the scope of the study, because in the current state, just distinguishing between 100% urban vs. 0% urban is not sufficient to analyze the above mentioned scenario.

We thank the reviewer for bringing this up. The "Nonurban" simulation in this study is a hypothetical scenario in which we assume current anthropogenic emissions and climate, but natural land cover prior to human perturbation. By doing so, we focus our study on the relative importance of land cover changes via urbanization on regional meteorology and air quality. To make it clearer in the paper that the "Nonurban" simulation is a hypothetical rather than realistic scenario, we added the following sentences to the introduction (section 1), and modified relevant sentences in section 2.5 and the conclusion (section 4).

Section 1

"In this paper, we aim to quantify the importance of historical land cover change on air pollutant concentrations, and thus the "Nonurban" scenario assumes current anthropogenic pollutant emissions. This hypothetical scenario cannot exist in reality, since current anthropogenic emissions would not exist without the city, but our intent is to tease out the relative importance of land cover change through urbanization (assuming constant emissions) on air pollutant concentrations."

Section 2.5

"Note that all three aforementioned scenarios adopt identical anthropogenic emission inventories described in Section 2.3. Using current anthropogenic emissions for "Nonurban" is a hypothetical scenario that cannot exist in reality, but allows us to tease out the effects of land surface changes via urbanization on meteorology and air pollutant concentrations. "

Section 4

"The two main simulations of focus in this study are the real-world "Present-day" and the hypothetical "Nonurban" scenarios"

I am further not fully convinced about the added benefit of this study for sustainable urban planning recommendations. I am aware that these model systems are not suitable for applied urban planning, but however the currently existing urban canopy models in WRF-Chem (and other models), together with high resolution datasets for both emission and urban morphology do offer a framework for a number of different scenarios in the context of climate change/UHI mitigation. Recent studies have been analyzing the impact of highly reflecting building materials, urban greening or varying building density for a number urban areas. These aspects should also be possible with this model system and worth being discussed in order to increase the scientific substance of that work and highlight the new contribution to the field. In light of the scope of the journal, it should also be worked out more detailed what are the implications for atmospheric science in general rather than purely investigating local/regional aspects.

We agree with the reviewer that our description of the implications of this study were somewhat ambiguous in the submitted version of the paper. The main point that we intend to make here is that land surface changes on their own can have a significant influence on regional air quality via altered meteorological conditions. Therefore, we should consider the benefits and penalties of UHI mitigation strategies (i.e., since most of them modify land surface properties) from the viewpoint of both climate and air quality to achieve a comprehensive assessment. We revised the conclusion section (section 4) as follows:

'This study highlights the role that land cover properties can have on regional meteorology and air quality. We find that increases in evapotranspiration, thermal inertia, and surface roughness due to historical urbanization are the main drivers of regional meteorology and air quality changes in Southern California. …Our findings indicate that air pollutant concentrations have been impacted by land cover changes since pre-settlement times (i.e., urbanization), even assuming constant anthropogenic emissions. These air pollutant changes are driven by urbanization-induced changes in meteorology. This suggests that policies that impact land surface properties (e.g., urban heat mitigations strategies) can have impacts on air pollutant concentrations (in addition to meteorological impacts); to the extent possible, all environmental systems should be taken into account when studying the benefits or potential penalties of policies that impact the land surface in cities."

I am convinced, that the model system, combined with the emission and land surface data sets offer a promising tool for discussing air quality/meteorology interactions in large urban areas such as Los Angeles, but however think that the variety of scenarios should be increased in order to allow for a more robust results towards currently relevant issues. The authors rely mostly on previous work with equal model configuration. Therefore, the own contribution to the field and the new development does not come out clearly. The paper however is well written and easy to follow, but crucial points have to be considered in a review before being able undertake a detailed line-by-line evaluation.

The study certainly builds on our prior work, but this paper focuses on air quality impacts, whereas our previous research was on only meteorology. Thus, the most important contribution of this work is that we investigate a totally different environmental system than previous work. In order to do so, we also add a new modeling component (atmospheric chemistry) that is not presented in past work.

Other smaller additions compared to our past work is that we turn on the shadow model and incorporate GIS-based building morphologies, which make the model simulations more representative of current day weather conditions in LA. Moreover, while the influence of land use changes on regional weather has been well studied, its influence on regional air quality has been seldom studied with accurately resolved land surface data, especially in the Southern California region. Therefore, our study fills this research gap. We added several sentences in the last paragraph of introduction section to emphasize this point.

"Note that this paper builds on our prior study Vahmani et al. (2016), but focuses on air quality impacts, whereas our previous research was on meteorological impacts only. While the influence of land surface changes on regional weather has been investigated in numerous past studies, its influence on regional air quality has been seldom studied in past work."

Moreover, the focus of this study is on the impact of land surface changes on regional meteorology and air quality. Thus, the two major scenarios discussed are "Nonurban" and "Present-day" scenarios, which characterized land surface prior to human perturbation and under current conditions respectively. We also included a supplemental scenario "Present-day No Irrigation" that teases out the effects of irrigation.

1. The scope of the study should be defined more clearly in light of the above mentioned points. The experimental design should be expanded, in order to include more own ideas/developments.

As mentioned in our previous responses, we focus this study on the relative importance of land cover changes via urbanization on regional meteorology and air quality, and assume identical climate and anthropogenic emissions in both scenarios. In our simulations, we implemented real-world representation of land surface properties in the "Present-day" scenario, which made it possible to tease out the most important land surface factors. Our results indicate that land surface changes have a significant influence on regional air quality via altered meteorological conditions. This suggests that policies that impact land surface properties should take all environmental systems into account when studying the benefits or potential penalties of the policies. We feel that this is a solid focused story for the paper, and adding additional simulations would only dilute the main points we are trying to make. In other words, adding more complexity to the study would only muddle the story.

2. One interesting and highly relevant point in my opinion is the 'irrigation' module which might offer a nice tool for testing different irrigation scenarios.

We agree that the proposed research idea is an interesting topic. However, it would be more appropriate as an individual study on the influence of irrigation on regional climate and air quality. This isn't the main research question we are trying to answer. In this study, we want to keep the scope well defined in answering the posed research questions on how historical land surface changes have affected regional climate and air pollutant concentrations in Southern California. Thus, investigating the regional influence of irrigation sounds interesting but beyond our motivation and scope.

3. Why did you select a single-layer urban canopy model rather that a more complex multi-layer canopy representation (BEP/BEM)? The latter should deliver higher accuracy close to the ground I guess? What is the depth of the lowest model level?

As suggested by Kusaka et al. (2001) , the model performance of UCM and BEP/BEM with regard to studying mesoscale heat islands are comparable. Chen et al. (2011) also mentions that the UCM may be more suitable than BEP/BEM for weather and air quality prediction. In addition, coupling BEP/BEM to WRF/Chem would be an extremely complex model development exercise, and the resulting model would be prohibitively computationally expensive, but for likely little additional benefit in the quality of simulations. Therefore, we choose to couple the UCM instead of BEP/BEM to WRF/CHEM.

The averaged depth of the lowest model level is 53 m for all three domains. This information has been added to the paper at the last sentence in section 2.1.

"The average depth of the lowest model level is 53 m for all three domains."

4. Where do the input parameters for SLUCM come from?

We use NLCD impervious surface data for impervious fraction of each grid cell. For surface albedo of roof, building wall, and road, we assign the grid cell albedo value derived from MODIS. Building morphologies (including building height, standard deviation of roof height, building width and road width) are from NUDAPT where available. Where NUDAPT data are unavailable, we adopt average building and road morphology from LARIAC. This information is mentioned in section 2.2. For the other parameters in UCM (e.g., anthropogenic latent heat, surface emissivity), we use default WRF settings documented in file URBPARM.TBL. We added this information to section 2.2.

"For the other parameters in the UCM (e.g., anthropogenic latent heat, surface emissivity), we use default WRF settings documented in file URBPARM.TBL."

5. What is the additional gain of a 30 m land surface classification which has to be scaled to 2 km model resolution?

We chose to use 30 m-resolution 33-category NLCD mainly for two reasons. First, urban land use varies at spatial scales on the order of 10s of meters. So it works best to define land use at spatial scales of 10s of m, and then aggregate to the model grid resolution. It would be difficult to detect land use using data at 2km resolution. Second, the 30 m-resolution land use dataset has 33 categories of land use type, which divides urban type into three sub-types: low-intensity residential, high-intensity residential, and industrial/commercial. This allows different parameterizations for different sub- urban types, which better characterize land surface properties.

6. Is there a problem with regard to the discrepancy between emission inventory and model resolution?

No, there should not be a problem. The resulting air quality predictions are simply lower resolution than they would be if they were at 2km. We ensured that the total emissions within in the domain are kept consistent after regriding.

7. How realistic is the surrounding 'non-urban' land use classification for the 'historical' scenario?

The dominant natural land cover type surrounding Los Angeles and San Diego metropolitan areas is shrub. So it is reasonable to assume shrub as the land use type in the "Nonurban" scenario. The land surface properties of these grid cells in the "Nonurban" scenario are derived using the inverse distance weighting approach, which is mentioned in section 2.5, and consistent with our previous publication.

"For the Nonurban scenario, we assume natural land cover prior to human perturbation, and replace all urban grid cells with "shrubs" (Figure 1c). We modify MODIS-retrieved albedo, GVF and LAI in these areas based on properties for shrub lands surrounding urban regions in the Present-day scenario. A detailed explanation on this method (inverse distance weighting approach) can be found in Vahmani et al. (2016)."

8. How well does the model simulate urban AND rural parameters?

In section 3.1, we showed how well the model simulated urban variables (i.e., near surface air temperature, and pollutant concentrations). We agree with the reviewer that it is also important for the model to capture nonurban (especially shrub) air temperatures well, thus we included a discussion on this topic in section S2 in the supplemental information. Note that nonurban observational sites that measure pollutant concentrations are rare. Thus, we decided not the discuss how well the model simulated pollutant concentrations in nonurban area. The new text in the main paper is pasted below. Please see section S2 in the supplemental information for more details on the validation. We did not paste it here because it is ~3 pages.

In the main paper:
"In addition, we only include observations from monitoring sites that are located in urban grid cells in the Present-day scenario. The validation of near surface air temperatures for both urban and nonurban sites are discussed in section S2 in the supplemental information."

9. Please specify how results from this study can serve as contribution for applied urban planning.

As mentioned in our response to your second general concern, we changed the last paragraph in the conclusion section, which explains the implications of our study. Please see that response for more detail.

10. In relation to other chapters, the introduction is slightly too long. Try to focus on the relevant points here and shorten were possible.

We think that the background knowledge, brief literature review, and research gaps described in the introduction section are necessary for a clear explanation of the scope and motivation of this study.

The flow of the introduction section is as follows. First, we point out that urbanization has led to profound modification of the land surface. We then explain how changes in land surface properties can affect regional meteorological fields such as surface and air temperature, wind speed and PBL height. We go on to demonstrate how those changes in meteorology due to land surface modification can in turn affect air pollutant concentrations via different mechanisms. While there are a number of studies that have investigated the impacts of land surface changes on regional meteorology, limited studies have quantified the impact of land surface changes on regional air quality, especially for the Southern California region, which has a history of severe air pollutant problems. In addition, recent studies have made it possible to utilize satellite land surface data in model simulations, which better predict regional weather in urbanized regions, and urban versus nonurban differences. Thus, our study adopts the modified model configuration, and aims to characterize the influence of historical urbanization on urban meteorology and air quality in Southern California.

Please find below comments for specific sections, which partly have been addressed in the main points above.
Ln 11: ventilation not a good expression here

We think that "ventilation" is a proper expression here because it appropriately describes the ability of atmosphere to transport pollutants out of the studied area.

13: 'before human settlement' is a bit misleading here, as it is not entirely captured by your model setup. As mentioned before, some effort has to be put in a clear definition of the scope of your study. What problem should be addressed – also in light of recommendations for real urban planning (Lines 570-573?

The two concerns mentioned in this comment are addressed in the first two general comments respectively.

43: "Differences in surface temperature..." What was the purpose of these studies mentioned here and what do they try to answer? How does this sentence relate to your study and the intention for this work?

The UHI and UCI represents urban versus nonurban difference in surface or air temperature. They are both climate phenomena at urban scale that occur due to variability in land cover changes. Here we summarize possible ways in which land surface modifications can affect surface/air temperature difference between urban and nonurban areas, which is what causes the UHI/UCI. The temperature difference between the "Present-day" scenario and "Nonurban" scenario discussed in our study is analogous to the UHI/UCI. Thus, the background information here is necessary.

47: "UCI": How does this relate to your study?

This is explained in the response to the comment above.

67: What is the role of the atmospheric aerosol burden for UHI formation?

The role of atmospheric aerosol on UHI intensity is an active research topic, and yet no consensus has been reached. For example, Kumar et al. (2017) carried out a Global Climate Model simulation, and suggested that daytime cooling (UCI) can be partially attributed to absorbing aerosols over Indian cities. Cao et al. (2016) used satellite observations, and found positive correlation between urban–rural difference in AOD and nighttime UHI.

73: better "characteristics/shape of the PBL is dependent on..."

Changed. Thanks!

81: better "due to urbanization..."

Changed. Thanks!

86: unclear what is meant by "meteorological changes via altered emissions,..."

We changed the sentence as follows:

"Meteorology can affect emission rates, chemical reaction rates, gas-particle phase partitioning of semi-volatile species, pollutant dispersion, and deposition; thus, it plays an important role in determining air pollutant concentrations."

115: Why do higher PBLs increase $PM_{2.5}$ concentration? Please discuss the related processes here.

In our text, we mentioned that Chen et al. (2018) found that higher PBLs decrease $PM_{2.5}$ concentrations. Please find the original sentence below:

"Chen et al. (2018) studied urbanization in Beijing, and found that modification of rural to urban land surfaces has led to increases in near-surface air temperature and PBL height, which in turn led to increases (+9.5 ppb) in surface O_3 concentrations and decreases (–16.6 $\mu g/m^3$) in $PM_{2.5}$ concentrations."

119: How exactly does your experimental setup treat the "wide heterogeneity of urban land surface processes" compared to existing studies? A large number of studies already exist using model systems (e.g. WRF) which include urban canopy models with varying complexity (SLUCM, BEP), which consider a similar level of heterogeneity than your experiments? Please discuss your statement.

While previous studies have used models with different levels of complexity, most of them failed to incorporate real-world land surface property data as input. They used default WRF settings for land surface properties such as building morphology, albedo, vegetation fraction, which either is out of date, or lacks spatial heterogeneity. By contrast, in this study we use NLCD for land cover type and impervious fraction, satellite-retrieved data for albedo, vegetation fraction and leaf area index, and GIS-based data for building morphology, which resolves spatial heterogeneity of land surface properties, and better predicts regional weather and air quality. The default version of the WRF/UCM assumes that many land cover properties are spatially homogeneous, which is not realistic.

122: unclear expression "amongst"?

We changed the sentence as follows:

"In addition, only few studies investigate interactions between land surface changes and air quality for the Southern California region (Taha, 2015; Epstein et al., 2017; Zhang et al., 2018b), which is one of the most polluted areas in the United States (American Lung Association, 2012)."

134-140: It should be made clear which new aspects you aim to analyze compared to the studies mentioned above. In my opinion simply turning urban on/off does not reveal significantly new insights. Further the term "human disturbance" is unclear, as this would also involve air quality modifications.

As we mentioned in the introduction, there are limited studies on the effect of land surface change via urbanization on regional air quality, most of which do not resolve the real-world spatial heterogeneity. In addition, there are several recent studies by our group, which incorporate satellite data for land surface characterization within Southern California, and quantifies the effect of land surface changes on regional climate including temperature and wind speed. Thus, this study combines the research idea of these two types of studies together, and aims to characterize the influence of land surface changes via historical urbanization on urban meteorology and air quality in Southern California using highly resolved land surface characterization.

We focus on the land surface modifications from human disturbance in this study, and use specific phrasing about this in the paper.

Abstract
"In this study we characterize the influence of land surface changes via historical urbanization from before human settlement to present-day on meteorology and air quality in Southern California using the Weather Research and Forecasting Model coupled to chemistry and the single-layer urban canopy model (WRF/Chem-UCM)."

Last paragraph in introduction section
"Therefore, this study aims to characterize the influence of land surface changes via historical urbanization on urban meteorology and air quality in Southern California by comparing a "Present-day" scenario with current urban land surface properties and land surface processes to a "Nonurban" scenario assuming land surface distributions prior to human perturbation."

Section 2.5

"For the Nonurban scenario, we assume natural land cover prior to human perturbation, and replace all urban grid cells with "shrubs" (Figure 1c)."

First paragraph in conclusion section

"In this study, we have characterized the impact of land surface changes via urbanization on regional meteorology and air quality in Southern California using an enhanced version of WRF/Chem-UCM. ... The two main simulations of focus in this study are the real-world "Present-day" and the hypothetical "Nonurban" scenarios; the former assumes current land cover distributions and irrigation of vegetative areas, while the latter assumes land cover distributions prior to widespread urbanization and no irrigation."

174: Please specify your lowest model level.

Thanks for the suggestion. We added this information to the manuscript.

"The average depth of the lowest model level is 53 m for all three domains."

175: "process parametrization" unclear

We changed the title to:

"Land Surface Property Characterization and Irrigation Parameterization"

180: Please discuss the term "real world representation", answering the question why the WRF default land use classification in WRF is not "real" enough for your case comparing these datasets with your input. What was the idea behind using a 30m dataset? Please briefly discuss the gain of using 30 m land cover data for a maximum resolution of 2 km. How much information "is lost" by the process of "upscaling" the LU data. Would the 2011 NLCD dataset add additional benefit?

By "real world representation", we mean instead of using default land surface properties provided with WRF, we used satellite-retrieved data specifically for the Southern California region. This is beneficial for a better prediction of regional weather.

As we mentioned in response to comment 5, we chose to use 30 m-resolution 33-category NLCD mainly for two reasons. First, urban land use varies at spatial scales on the order of 10s of meters. Second, it separates urban to three sub- urban types, which allows more detailed parameterization.

Also we chose to use the 2006 NLCD dataset in order to keep consistency with previous work from our group.

205: Did you use the additional sub-tiling option in WRF?

No, we didn't. The land surface module (the unified Noah land surface model) we use doesn't have a sub-tiling option. However, the module treats impervious fraction and pervious fraction of the urban grid cell separately.

243: Do you consider daily emission profiles? Meaning, do you find two "peaks" for instance in NOx emission/concentration?

Yes we do. Figure R2.1 shows the diurnal cycle of NOx emissions. The diurnal cycle of NOx concentrations is shown in the paper in Figure 6a. We can see that the emissions of NOx shows two peaks during daytime, and stays high between the two peaks. However, for NOx concentrations, it peaks during morning, and decreases continuously until late afternoon, despite rather high emissions. This indicates that high photolysis rates and high PBL heights due to warm temperatures in the afternoon play an important role on determining NOx concentration during daytime, apart from just emissions.

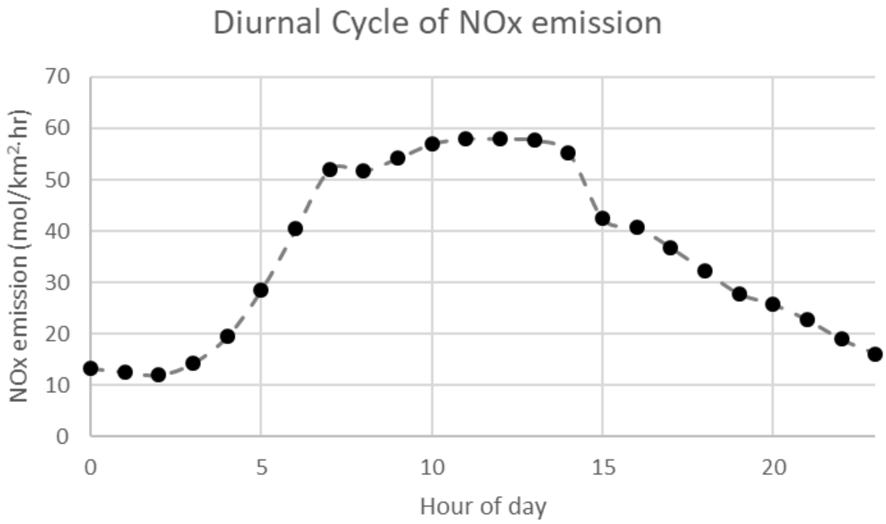

Figure R2.1 Diurnal cycle of NOx emissions. Averaged over urban grid cells within simulation period.

267: How realistic is the conversion to shrub-land for all grid cells? Would you expect different effects for a non-urban, but more heterogeneous "before human" land cover?

Please see the response to comment 7.

294: Please indicate better proof of the "good fit" mentioned here. It is not indicated by Figures S1 and 3 for $PM_{2.5}$. How does the correlation coefficient look like? What are the reasons for the poor correlation especially for the high range of the observed concentrations? How representative are the measurement stations? As the ozone concentration is highly dependent on temperature you find a good fit. Does the poor fit for $PM_{2.5}$ relate to high mixing, chemistry, both? How do correlations look like for NO2, NO, CO? Are the simulated diurnal variations realistic? Please also discuss the values from Table 1? Are they particularly good/bad?

We modified the last sentence in section 3.1 as below, and added the comparison between our evaluation results and recommended model performance benchmarks. The comparison indicates that the evaluation result are close to the ME benchmark for hourly near surface air temperature, and NMB benchmark for hourly Ozone concentrations. For daily $PM_{2.5}$ concentration, the discrepancy between the evaluation and the recommended benchmark is largely due to the underestimation of high observational values. This poor fit at high concentrations is likely occurring due to one or more of the following factors: 1) not including dust emissions in the simulation, which makes up an appreciable fraction of real-world total $PM_{2.5}$, 2) a failure of the emissions inventory to capture high emission rates on particular days, and 3) the chemistry parameterizations in WRF/Chem tending to underestimate $PM_{2.5}$ concentrations at high values, and 4) errors in simulated air pollution meteorology. We also added this information to the main paper.

'The underestimation of $PM_{2.5}$ concentrations may be occurring due to one or more of the following factors: 1) not including dust emissions in the simulation, which makes up an appreciable fraction of real-world total $PM_{2.5}$, 2) a failure of the emissions inventory to capture high emission rates on particular days, 3) the chemistry parameterizations in WRF/Chem tending to underestimate $PM_{2.5}$ concentrations at high values, and 4) errors in simulated air pollution meteorology. Table 1 shows four statistical metrics for model evaluation, including mean bias (MB) and normalized mean bias (NMB) for the quantification of bias, and mean error (ME) and root mean square error (RMSE) for the quantification of error. The statistical results indicate that model simulations underestimate near-surface air temperature, O_3 and $PM_{2.5}$ concentrations by 1.0 K, 22% and 31%, respectively. The comparison between our evaluation results and recommended model performance benchmarks is presented in the supplemental information Table S2."

The correlation coefficients for near surface air temperature, O_3 concentration and $PM_{2.5}$ concentration are 0.92, 0.82, 0.025 respectively. The observation sites for air temperature, O_3 concentration, and $PM_{2.5}$ concentration are shown is Figure S9. The sites are spread across the urban region in the model domain, and should be representative of the urban region in Southern California. On the other hand, point measurements do not capture the same spatial footprint as 2 km model grid cells. Thus, some model versus observational discrepancy is always expected, making interpretation difficult.

Figure S11 and Figure S12 shows the comparison between observed and modeled diurnal cycle for near surface air temperature and O_3 concentrations. Values for each hour are averaged over the whole simulation period for all observation sites. The results indicate that while model underestimates both observed air temperature and O_3 concentrations, it follows the diurnal pattern well. These figures are in the supplemental information, and we added a sentence in the main paper.

"…(Comparisons between observed and modeled diurnal cycles for near surface air temperatures and O3 concentrations are also presented in the supplemental information, Figure S11 and S12.)"

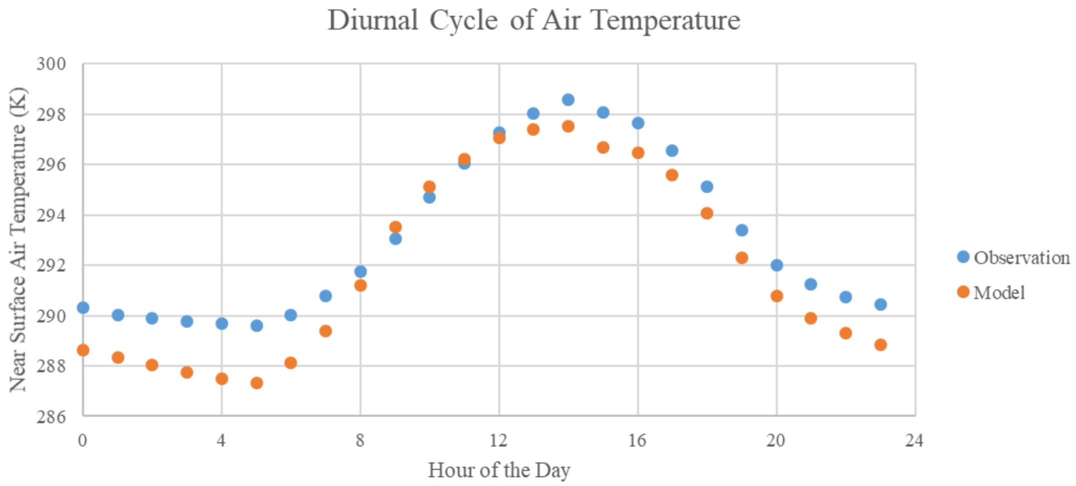

Figure S11. Diurnal cycle of observed and modeled near surface air temperature.

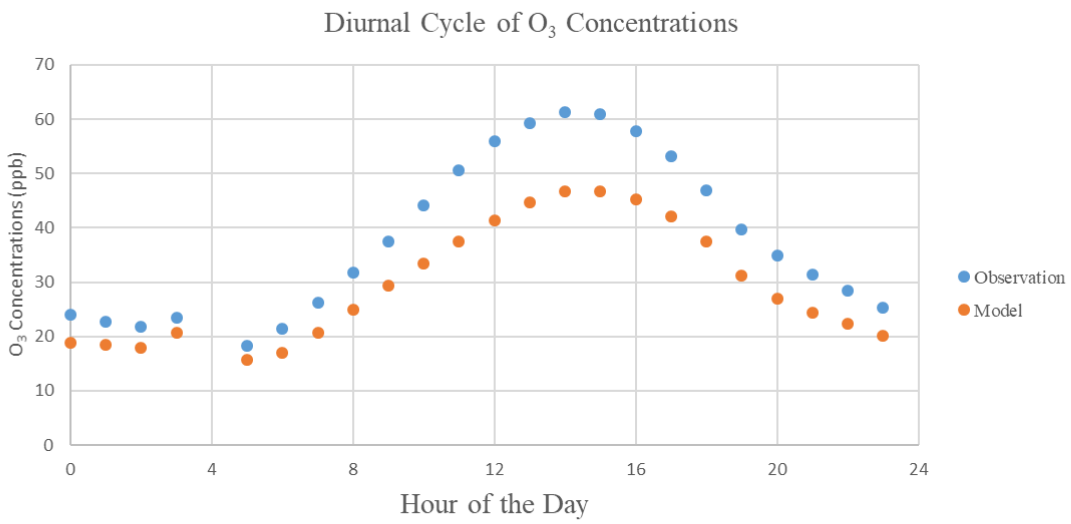

Figure S12. Diurnal cycle of observed and modeled surface O_3 concentrations (ppb).

347: Can you find impacts on the strength of the sea breeze when there is no urban area left?

The Present-day versus Nonurban difference in the strength of sea breeze is shown by Figure S14 in the supplemental information. Land surface changes via urbanization has led to decrease in wind speed throughout the day due to increase in land surface roughness. This weakening is more significant during the day, especially in the afternoon, when the baseline sea breeze is strongest.

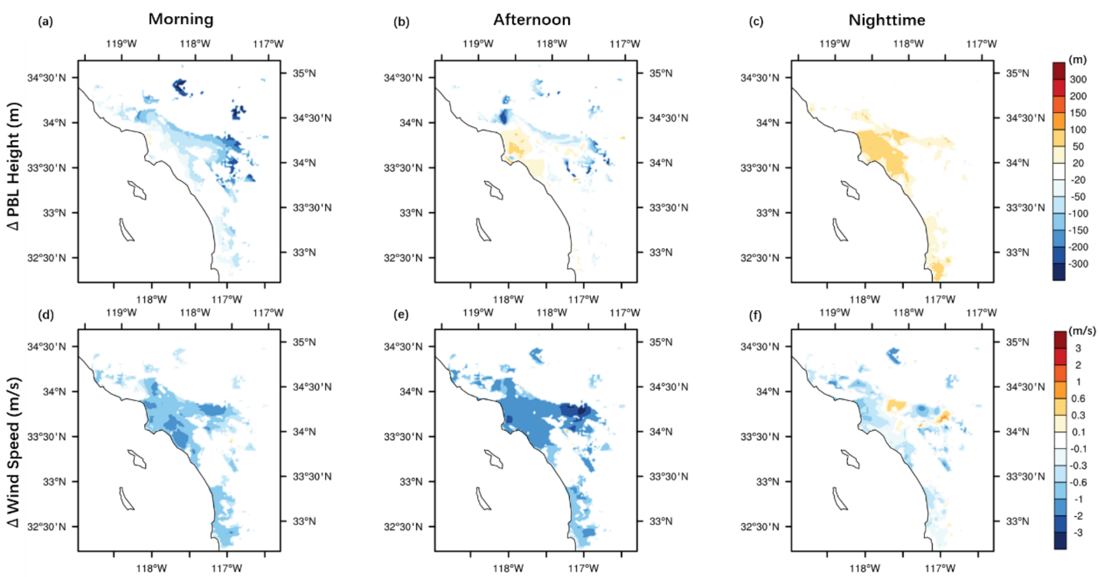

Figure S14. Spatial patterns of differences (Present-day – nonurban) in temporally averaged values during morning, afternoon and nighttime for (a,b,c) PBL heights, and (d,e,f) averaged wind speed under within PBL. Note that values are shown only for urban grid cells. Morning is defined as 7 PST to 12 PST, afternoon as 12 PST to 19 PST, and nighttime as 19 PST to 7 PST. Note that values are shown only for urban grid cells.

363: What is the order of difference between shadow model on/off?

There are two major differences in model configuration between this study and our previous publication (Vahmani and Ban-Weiss, 2016; Vahmani et al., 2016). First, in this study, we turn on the shadow model, while our previous study doesn't account for the shadow effect. Second, we use the model default calculation of surface temperature (which will affect the calculation of air temperature), while our previous study uses an alternative calculation of surface temperature. After a careful comparison among different model set-ups, we found that the parameterization of surface temperature is the more important factor that affects daytime air temperature (Figure S16 in the supplemental information). Therefore, we delete "and shading effects within urban canopies" in the sentence "In addition, increases in thermal inertia caused by use of manmade materials (e.g., pavements and buildings) and shading effects within urban canopies can contribute to simulated temperature reductions during the morning.". We also modified the last paragraph in section 3.2.3 as below.

"Note that changes in air temperature during daytime shown here disagree with Vahmani et al. (2016). While our study detects daytime temperature reductions due to urbanization, Vahmani et al. (2016) suggests daytime warming. After detailed comparison of the simulations in our study versus Vahmani et al. (2016), we find that the differences are mainly associated with UCM configuration. First, our study uses model default calculations of surface temperature for the impervious portion of urban grid cells, whereas Vahmani et al. (2016) applied an alternative calculation proposed by Li and Bou-Zeid, 2014. Li and Bou-Zeid, 2014 intended the alternate surface temperature calculations to be performed as a post-processing step rather than during runtime. After a careful

comparison among different model set-ups, we found that the parameterization of surface temperature is an important factor that affects simulated daytime air temperature (See Figure S16). Second, our study accounts for shadow effects in urban canopies, whereas Vahmani et al. (2016) assumes no shadow effects. (We note here that the default version of the UCM has the shadow model turned off. The boolean SHADOW variable in module_sf_urban.F needs to be manually switched to true to enable the shadow model calculations. With the shadow model turned off, all shortwave radiation within the urban canopy is assumed diffuse.) We suggest that it is important to include the effects of building morphology on shadows within the canopy, and to track direct and diffuse radiation separately, and therefore perform simulations in this study with the shadow model on. Note that the effect of shadows is not as significant as the parameterization of surface temperature for most of the domain in our study because the ratio between building height and road width is small."

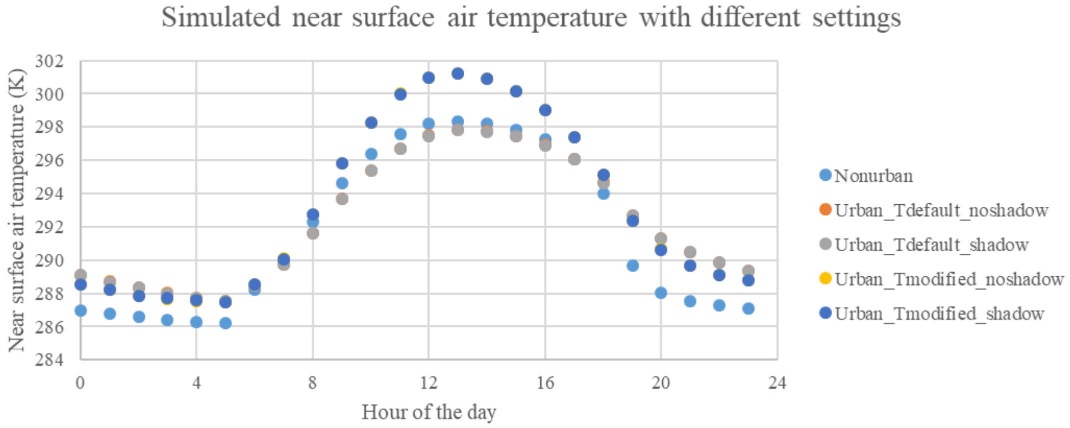

Figure S16. Diurnal cycle of near surface air temperature simulated with different model set-ups. "Tdefault" indicates that the simulation uses the default calculation of surface temperature in WRF, while "Tmodified" indicates that the simulation uses the calculation of surface temperature from Li and Bou-Zeid (2014) (which is also used in (Vahmani el al. (2016)). Dots for "Urban_Tdefault_shadow" and "Urban_Tdefault_noshadow" ("Urban_Tmodified_shadow" and "Urban_Tmodified_noshadow") are overlapping at every hour of the day because the simulation results with shadow on/off are very similar.

367: Origin of the UCM parameters?

Please refer to our response to the fourth comment.

381: Calculation of the ventilation coefficient?

We added the calculation of ventilation coefficient to the main content in section 3.2

"… The integral form of this calculation can be written as (Eq1).

$$Ventilation\ Coefficient = \int_0^{PBL\ height} U(z)\ dz \qquad (Eq1)$$

Given that the atmosphere is stratified in models, Eq1 can be discretized as Eq2:

$$Ventilation\ Coefficient = \sum_{i=1}^{m} U(z_i) \times \Delta z_i \qquad (Eq2)$$

Where $U(z_i)$ stands for horizontal wind speed within the i^{th} model layer (m/s), Δz_i is the depth of

i^{th} model layer that is within PBL (m), and m is the number of vertical layers up to PBL height."

386: Please evaluate the quantity values here? Provide relative numbers.

Thanks for the suggestion. We add relative values to the sentence.

"… the spatially averaged decreases are –726 m^2/s (–23%) and –560 m^2/s (–34%), respectively."

395: What is the relation between PBL height and surface roughness? Please provide more details. Can you find proof for this in your study?

We've discussed how surface roughness affects PBL height in the third paragraph in the introduction section. The nighttime PBL height is associated with variations in wind speed, which is related to variations in surface roughness. By changing shrubs (homogeneously throughout the urban region) to buildings (heterogeneously varies according to sub-urban types), the variation in surface roughness is increased. We modified the relevant sentence in section 3.2.4.

"Coastal (inland) regions show larger (smaller) variation in roughness length (Figure 2e), which leads to larger (smaller) increases in PBL heights (Figure S14c)."

490: Can you say something about the change of PBL dynamic comparing urban and non-urban. I suspect concentration of $PM_{2.5}$ is highly dependent on the boundary layer depth. Expecting lower PBLs in "urban-free" areas actually should decrease $PM_{2.5}$ in summer?

As we discussed in section 3.2.5, air temperature (surface roughness) changes are the major driver of PBL changes during the day (night) for urban grid cells. While land surface properties don't change among nonurban grid cells (i.e., outside the urban domain), changes among urban grid cells will affect nonurban grid cells via transport of moisture, energy and momentum. Thus, most nonurban regions show similar trends for changes in PBL height compared to urban regions (discussed in the response to the next comment).

Responding to your last sentence, lower PBLs would lead to greater $PM_{2.5}$ concentrations, not lower concentrations.

530: What happens to the PBL height in non-urban environment? Even deeper?

Figure R2.2 shows changes in PBL height for Present-day – Nonurban (showing values only for grid cells that are not deemed urban in the Present-day scenario). PBL height decreases in most regions during the day (i.e., morning and afternoon), while changes at night are negligible. The tendency of changes in these grid cells outside the urban region are similar to that in urban grid cells.

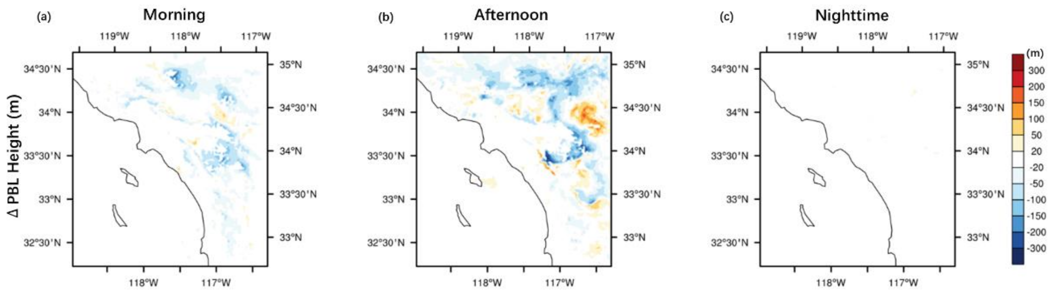

Figure R2.2 Spatial patterns of differences (Present-day – nonurban) in temporally averaged values during morning, afternoon and nighttime for (a,b,c) PBL height. Note that values are shown only for grid cells that are not deemed urban in the Present-day scenario.

535: Specify "enhanced".

We added the following sentence to the paper to specify "enhanced".

"We use satellite data for the characterization of land surface properties, and include a Southern California-specific irrigation parameterization."

541: How confident are you that the land use class in the "before-human" settlement is correct? Or is it just a guess?

Please refer to the response to comment 7.

573: As mentioned earlier I am not entirely convinced, how findings from this study could be used for applied urban planning? You mention 'mitigation and adaptation', but a complete 'removal' of the urban area should be hard to transfer into an actual applicable strategies. Maybe more 'moderate' scenarios would be better. However, avoiding a complete re-doing of model experiments, the scope of the study should be formulated differently

Please refer to our responses to the second general comment, and the first detailed comment.

Reference

Cao, C., Lee, X., Liu, S., Schultz, N., Xiao, W., Zhang, M. and Zhao, L.: Urban heat islands in China enhanced by haze pollution, Nat. Commun., 7, 12509, 2016.

Chen, F., Kusaka, H., Bornstein, R., Ching, J., Grimmond, C. S. B., Grossman-Clarke, S., Loridan, T., Manning, K.W., Martilli, A., Miao, S. and Sailor, D.: The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems, Int. J. Climatol., 31(2), 273–288, 2011.

Kumar, R., Mishra, V., Buzan, J., Kumar, R., Shindell, D. and Huber, M.: Dominant control of agriculture and irrigation on urban heat island in India, Sci. Rep., 7(1), 14054, 2017.

Kusaka, H., Kondo, H., Kikegawa, Y. and Kimura, F.: A simple single-layer Urban Canopy Model for atmospheric models: comparison with multi-layer and slab models, Bound.-Layer Meteorol., 101(3), 329–358, 2001.

Vahmani, P. and Ban-Weiss, G. A.: Impact of remotely sensed albedo and vegetation fraction on simulation of urban climate in WRF-urban canopy model: A case study of the urban heat island in Los Angeles. J. Geophys. Res. Atmos., 121(4), 1511–1531, 2016.

Vahmani, P., Sun, F., Hall, A. and Ban-Weiss, G. Investigating the climate impacts of urbanization and the potential for cool roofs to counter future climate change in Southern California. Environ. Res. Lett., 11(12), 124027, 2016.