

Reviewer2

Comment #2-0

The paper takes on the very substantial challenge of developing a global dataset of urban fractional changes at a 1km resolution from 2020 to 2100 under eight scenarios of socioeconomic pathways and climate change. The newly developed fractional urban land dataset is quite valuable and is helpful to assess the environmental impact of future urbanization. The manuscript is generally well structured, but the discussion and conclusion need to be reorganized. For example, the conclusion section is too wordy and repeats some information from the method and result sections. Though the method used in this work has been published, I still have some concerns when the model is applied to global-scale modeling. There are some missing details in the method section, and I have provided detailed comments below. Overall, I believe this paper could be publishable after major revision.

Response: thank you for your suggestions and comments. We carefully revised the discussion and conclusion part, and added details of the methodology part in our revised manuscript. The detailed point-by-point response is shown below.

Comment #2-1

P6, Line 1-2. “We characterized urban fractional change across different states in each country, using the long-term (1985-2015) urban extent data (i.e., GAIA) and the sigmoid growth model.” The sigmoid growth model is fitted in the state level, and the estimate parameters (a, b, c, d) are also shown in Fig. 6. How the sigmoid growth model performs in each state, because the urbanization stage is different among developing countries and developed countries. I suggest adding a fit performance of the sigmoid growth model to prove the model reliability and explain the model uncertainty.

Response: thank you for your suggestions. We have included the fit performance of the sigmoid growth model across states globally (Fig. S7). First, we fitted the ISA-based urban growth model with calibrated model parameters using the urban extent time series from (1985-2015). Then, we evaluated the performance of sigmoid model using the coefficient of determination (R^2) between the modeled and referred ISAs. Our results indicate the sigmoid model can well characterize the urban evolution with R^2 above 0.8.

“Additionally, we evaluated its performance of the ISA-based sigmoid growth model using the coefficient of determination (R^2) between the estimated and observed ISAs (1985-2015) at the state level.” (Page 8, Line 4-6)

“The fit performance (i.e., with R^2 above 0.8) of the sigmoid model at the state-level indicates the calibrated model can well characterize the spatially explicit urban growth patterns over past decades using satellite observations (Fig. 7). It is worth to note that there are some states with relatively low performance, probably due to limited increments with different urbanization stages, e.g., in highly urbanized regions or developing regions with low urbanization levels.” (Page 12, Line 5-9)

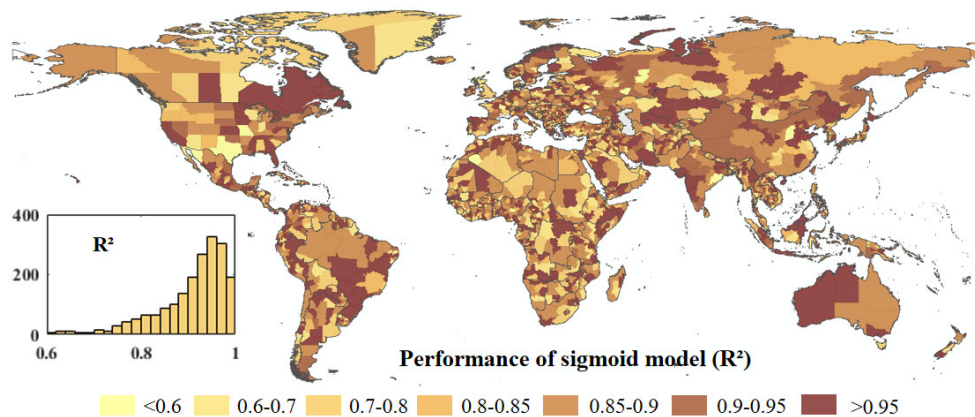


Fig 7. The fit performance (R^2) of state-level sigmoid model at the global scale. R^2 is the coefficient of determination between the simulated and referred ISAs over past decades.

Comment #2-2

P6, Line 12-15. “We incorporated the ISA-based growth mechanism with the Logistic-Trend-CA model (He et al., 2023), which incorporates temporal contexts of urban sprawl into the neighborhood configuration...” The sigmoid growth model is estimated at the state level, so did you also train the logistic regression model at the state level? The logistic regression is a binary regression model, but input data (GAIA) is fractional type. How did you implement the model training? I also see the methods in the supplementary materials, so only three spatial proxies (i.e., DEM, Distance to city centers, Distance to major road) are used to training the logistic regression model? And how many samples were selected to train the model in each state?

Response: thank you for your comments. We trained the country-specific regression model and obtained the spatially explicit suitability surface map. This is different with the sigmoid growth model because it is too homogeneous for these spatial proxies at the state level. Specifically, we identified those changed and persistent pixels and collected these two sample groups using the stratified sampling strategy. Thereafter, the suitability surface was derived using the regression model based on those spatial proxies, including DEM, slope, distance to city centers, distance to different roads, as well as those land covers. Details of these proxies can be referred in Supplementary Materials. We clarified these issues in our revised manuscript.

“Specifically, we identified those changed and persistent pixels and collected these two sample groups using the stratified sampling strategy. Thereafter, the suitability surface was derived using the regression model based on those spatial proxies, including the DEM, slope (calculated from DEM), minimum distance to city centers, minimum distance to different grades of roads, as well as various land covers (i.e., evergreen needleleaf forests, evergreen broadleaf forests, deciduous needleleaf forests, deciduous broadleaf forests, mixed forests, open shrublands, savannas, grasslands, permanent wetlands, croplands, cropland/natural vegetation mosaics, urban and built-up lands, permanent snow and ice, barren, water bodies).” (Supplementary Information)

“These spatial proxies reflect different spatial aspects related to urban sprawl, such as locations (e.g., minimum distance to major worldwide cities), traffic networks (e.g., minimum distance to major roads, minimum distance to highways, and minimum distance to local roads), terrain (e.g., digital elevation model and slope), were used to train the logistic regression model and land constraints (e.g., protected areas) (Table 1). For example, some spatial proxies (e.g., land cover and protected area) were defined as specific constraints (e.g., suitable or not), while terrain and location proxies were directly calculated from the DEM and distance to the nearest roads (or city centers), respectively.” (Page 5, Line 10-16)

Comment #2-3

P7, Line 12. “We calibrated the Logistic-Trend-ISA-CA model at the state level using historical urban extent time series data (i.e., GAIA) from satellite observations”. So how did you validate the sigmoid growth model?

Response: thank you for your comments. We evaluated the sigmoid growth model by calculating the coefficient of determination (R^2) between the simulated and referred ISA time series data. We explained this issue in our revised manuscript. Details can be referred to our response to Comment #2-1.

Comment #2-4

P7, Line 13-15. “First, we evaluated the performance of derived suitability surface using the receiver operating characteristic (ROC) method (Sunde et al., 2014). That is, the continuous values can be divided into binary maps using different thresholds to measure the agreement between threshold-derived results and the referenced urban extent.” The binary urban land map was used to evaluate the performance of the derived suitability surface. How the thresholds were determined to extract the binary urban maps for each state. Are there large differences in the thresholds among states? If you evaluate the model performance for each state, I suggest that Fig.7 (i.e., model performance at the country level) could show the AUC values at the state level.

Response: thank you for your suggestions. First, the receiver operating characteristic (ROC) approach is a threshold-based evaluation approach (Sunde et al., 2014). That is, the continuous values can be divided into binary maps using different thresholds to measure the agreement between threshold-derived results and the referenced urban extent (i.e., identified by their increased ISA during 1985-2005 with a threshold of 0.5) (Sunde et al., 2014). Second, as our response to Comment #2-2, we didn't implement the regression model at the state level, because these spatial proxies are inadequate the heterogeneity spatially at the local scale. Finally, as suggested, we also evaluated the performance of suitability surface at the state-level (Fig. S8). We clarified it in our revised manuscript.

“We evaluated the performance of derived global suitability using the receiver operating characteristic (ROC) approach, which essentially is a threshold-based evaluation approach (Sunde et al., 2014). That is, the continuous values can be divided into binary maps using different thresholds to measure the agreement between threshold-derived results and the referenced urban extent (i.e., identified by their increased ISA during 1985-2005 with a threshold of 0.5) (Sunde et al., 2014).” (page 7, line 20-23; Page 8, Line 1)

“It is worthy to note that here we used the traditional ROC approach to evaluate the suitability surface, which is only one component of the adopted urban CA model in this study, despite that our modeling target is ISA instead of binary urban extent.” (page 8, line 2-4)

“Furthermore, the suitability surface performs well in most states with relatively high AUC values (Fig S8).” (Page 12, Line 19)

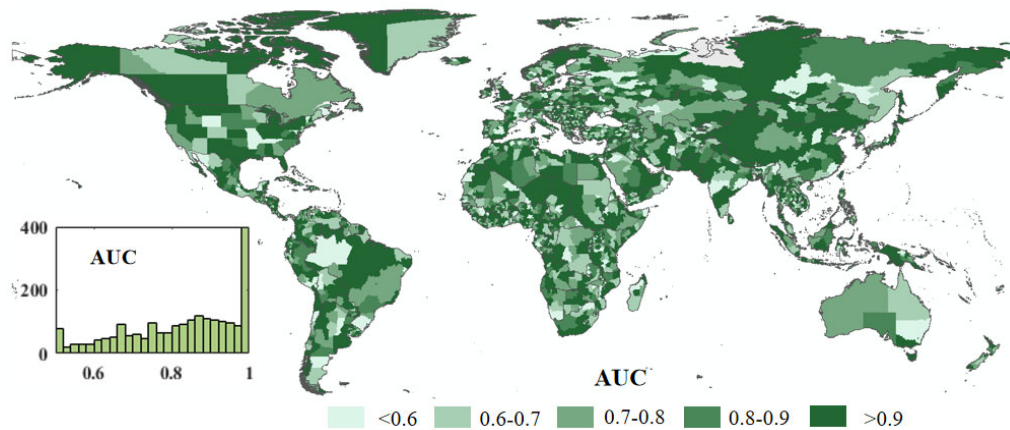


Fig S8. The model performance of derived suitability surfaces at the state-level using the indicator of the area under the curve (AUC) at the global scale.

Comment #2-5

P8, Line 5-6. Data harmonization should be cautious because the data source and definitions of GAIA and LUH2-urban are different. It is also simple to use equations (6) and (7) to harmonize the urban land from GAIA and LUH2. As I know, the overlap period for GAIA and LUH2 is 1985-2020. Did the harmonization rate change a lot during the overlap period? I suggest adding an uncertainty analysis for the data harmonization.

Response: thank you for your suggestions. First, the urban extents from both GAIA and LUH2 database were derived from remotely sensed observations and the definition of “urban” in LUH2 and GAIA are similar (i.e., pixel dominated by built-up areas). Specifically, the urban extents in LUH2 were initially estimated from spatially explicit built-up area map in 2000 from 1km DISCover dataset (Loveland *et al.*, 2000). Although the definition of “urban” in LUH2 and GAIA are similar (i.e., pixel dominated by built-up areas), the GAIA data have a finer spatial resolution (i.e., 30m) and a longer temporal span (1985-2018) at an annual step, with mean overall accuracies of above 90% across different years. Second, to keep the modeled ISA map in the future consistent with satellite derived observations, we chose the ISA map in 2015 as the start point of modeling, instead of the overlap period, for harmonization with LUH2 data. Third, as suggested, we also discussed the uncertainty of the harmonization rate (γ) in our revised manuscript. We used the quartile coefficient of dispersion (QCD) to measure the range of γ over past decades. Our results indicate the QCD is relatively small (0-0.2) in most countries, suggesting the γ is robust during the overlap period (Fig. S9). We explained these issues with added discussion in our revised manuscript.

“Due to the difference of adopted baseline urban extent in each product, there is a distinct gap regarding the urban area in these two datasets (i.e., GAIA and LUH2). Specifically, the urban extents in LUH2 were initially estimated from spatially explicit built-up data of Data and Information System Global Land Cover (DISCover) dataset at 1km resolution, which was mainly derived from the Advanced Very High Resolution Radiometer (AVHRR) satellite observations (Loveland et al., 2000; Goldewijk, 2017). While the definitions of "urban" are similar in both products, the differences in urban areas across various regions can be attributed mainly to their spatial resolutions and mapping years. In general, the urban extent in GAIA derived from Landsat has a longer temporal span and a high accuracy, with mean overall accuracies of above 90% across different years (Gong et al., 2020a).” (Page 9, Line 14-21)

“We calculated the QCD using the first (Q1) and third (Q3) quartiles for the annual harmonized rate γ in each country during the overlap period, as Eq.8. The QCD can capture the variation of harmonized rates within each country and is also comparable between countries with different deviation of future projection.

$$QCD = \frac{Q_3 - Q_1}{Q_3 + Q_1} \quad (8)$$

where Q_1 and Q_3 are the first (25th percentile) and third (75th percentile) quartiles of the annual harmonized rate in each country, respectively.” (Page 9, Line 22-23; Page 10, Line 1-5)

“Our results indicate the QCD is relatively small (0-0.2) in most countries, suggesting the γ is robust during the overlap period (Fig S9). The urban areas in most regions like America, Canada, China, India, and Australia have similar trends in GAIA and LUH2, while their gaps (i.e., QCD) are relatively large in those least developed countries (Fig S9).” (Page 14, Line 5-8)

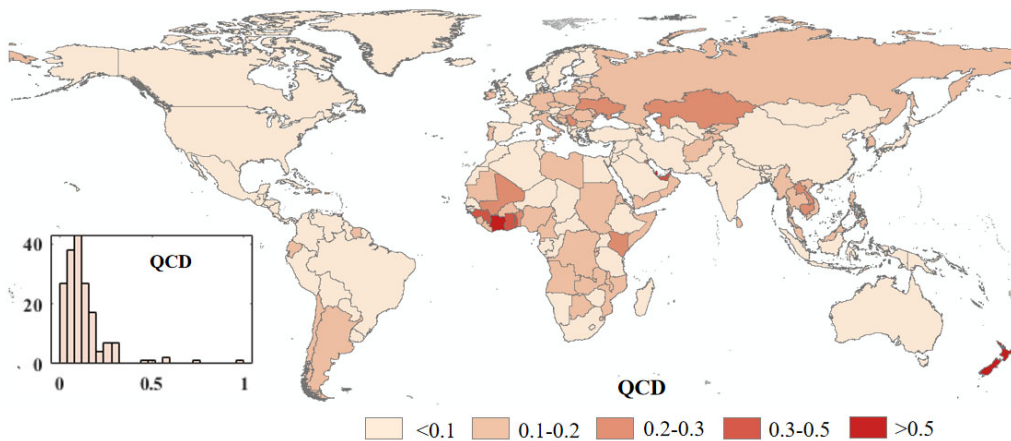


Fig. S9. The quartile coefficient of dispersion (QCD) within each country at the global scale during the overlap period (1990-2015).

Comment #2-6

P8, Line 20-24. Global South countries located in middle Asia, south America, and Africa, would likely experience more noticeable urban growth than Global North countries in the future, e.g., the growth rates of the United States of America (USA) and China are 3.95 and 1.05 times in 2100 under SSP2-RCP4.5, respectively, relative to the base year of 2015. These two sentences are contradictory, you made me confused.

Response: thank you for your comments. We improved this sentence in our revised manuscript.

“Among these scenarios, Global South countries located in middle Asia, south America, and Africa, would likely experience more noticeable urban growth than Global North countries in the future (Fig. 2), i.e., the average growth rate (i.e., 2100/2015) of Global South countries is around 5 under various future scenarios, while the average growth rate in Global North countries is mostly lower than 3.” (Page 10, Line 9-12)

Comment #2-7

P12, Section 4.3. In this section, you mainly compared the spatial pattern of the newly developed urban fractional dataset and previous datasets. I suggest adding a comparison analysis of future urban land area between the harmonized data and other available datasets.

Response: thank you for your suggestions. Although the overall trends of future global urban sprawl in our results are similar as those two global urban extent datasets (*Chen et al., 2020; Gao et al., 2020*) (Fig. S10), their magnitudes are notably different. This is mainly attributed to the variation caused by the urban area growth estimation model adopted in different products. For instance, the temporal trend of our results was mainly inherited from LUH2, which essentially was estimated from multiple integrated assessment models. However, for products in *Chen et al. (2020)* and *Gao et al. (2020)*, their urban areas were estimated using panel analysis and data-driven approaches, respectively, based on four-epoch time series data of global human settlement layer. We clarified it in our revised manuscript.

*“The overall trends of future global urban sprawl in our results are similar as those two global urban extent datasets (*Chen et al. (2020a); Gao and O’neill (2020)*), but their magnitudes are notably different (Fig. S10). These deviations are mainly attributed to the variation caused by the urban area growth estimation model adopted in different products. For instance, the temporal trend of our results was mainly inherited from LUH2, which essentially was estimated from multiple integrated assessment models. However, for products in *Chen et al. (2020)* and *Gao et al. (2020)*, their urban areas were estimated using panel analysis and data-driven approaches, respectively, based on four-epoch time series data of Global Human Settlement Layer (GHSL).” (Page 14, Line 15-23; Page 15, Line 1-4)*

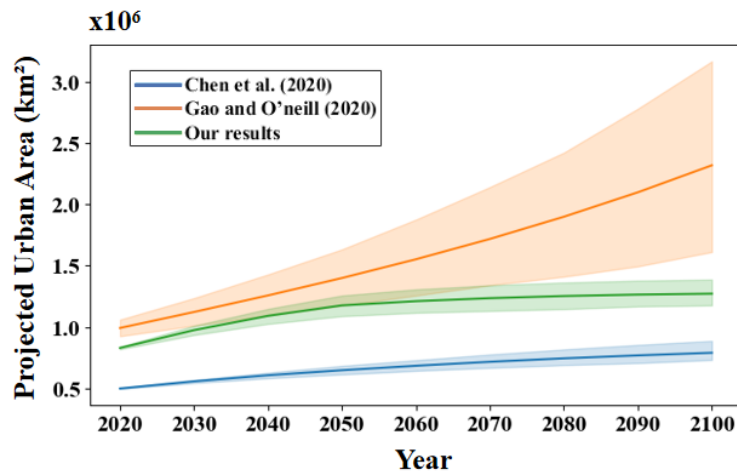


Fig. S10. The projected urban area of our results with two similar global urban datasets from Gao and O'neill (2020) and Chen et al. (2020a) under all future scenarios across years.

Comment #2-8

Fig. S6. I note that there will be no low-density ISA area in the city you show after 2060, and it seems that most of the urban area have the same ISA fraction. It also existed in other metropolitan areas (e.g., Fig. 10 and 11, New York city). so, I suggest to explain why.

Response: thank you for your comment. The projected urban area in LUH2 data will decrease in most regions due to the population decrease, especially after 2050. With the widely adopted assumption that it is irreversible for the transition from urban to non-urban, here we assumed the urban area growth would be plateaued once it reaches the peak, so do the ISA. Illustrated as the Georgia state in the US (e.g., Atlanta in Fig. S6), the total urban area reaches the peak after 2080 (Fig R1). We explained it in our revised manuscript.

“Given that the conversion from non-urban to urban is commonly assumed to be irreversible (Li et al., 2015), we assumed that the urban area in regions with projected population decline in the future would plateau after reaching its peak.” (page 9, line 7-9)

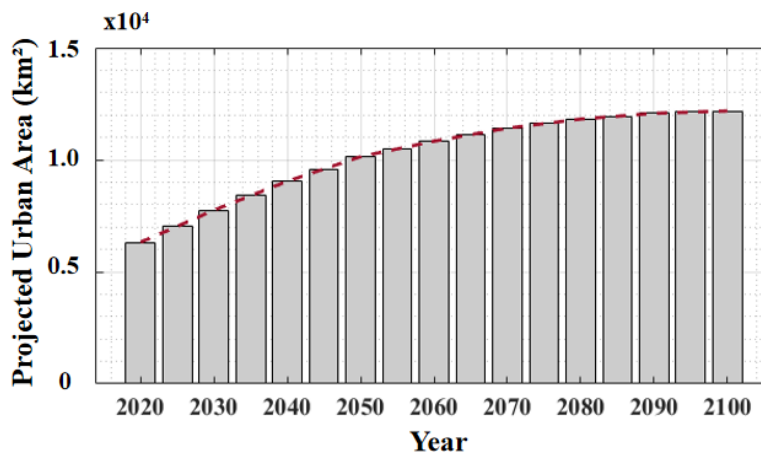


Fig. R1. The temporal trend of urban area growth in the state of Georgia.