**Data and analysis codes used in**

**Frolking S**, R Mahtta, T Milliman, T Esch, KC Seto. **2024-in press**. From spreading out to building up – the evolution of global urban structural growth over three decades, *Nature Cities.*

**PART 1. DATA FILES**

**PART 2. ANALYSIS CODES**

**PART 3. REFERENCES**

**PART 1. DATA FILES**

There are four csv data files that are read by the codes and used to generate figures and tables in the paper. There are also a few files used in QGIS, 1 shape file of MUA polygons, and some raster geotiff files; these are listed after the CSV files.

**global\_china\_total\_floor\_area\_data.csv**

• annual floor area from IEA (2022), global total, also partitioned into ‘advanced’ and ‘developing/emerging’ economies; areas in billion m2.

• Accumulating annual new floor area completed for China (billion m2) and Beijing and Shanghai (million m2). First year of data for Shanghai was 1978 so 51 million m2 had accumulated from then through 1985. Annual increments reported in online statistical yearbooks for China, Beijing, Shanghai (see reference list below).

**mua\_country\_list\_economy\_region.csv**

• list of countries with MUAs used in this study.

‘mua\_country\_code’ id number from Taubenbock et al. 2019;

‘region\_1’ is one of 12 global regions

‘economy’ is ‘advanced’ or ‘other’ (‘other’ = ‘developing/emerging’) according to IEA (2022). Sub\_regions’ divides ‘Africa’ into 5 sub-regions (north, south, east, west, central) and ‘Other Asia’ into 2 sub-regions (south, central). All other regions are listed also as sub-regions.

**mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv**

• list of individual MUAs (or cities) and some of their properties, from Taubenbock et al. (2019).

mua\_objectid: from Taubenbock et al. (2019);

mua\_country\_code: from Taubenbock et al. (2019);

mua\_country: from Taubenbock et al. (2019);

mua\_worldpop\_thousands: from Taubenbock et al. (2019) (from 2014 UN data);

mua\_area\_m2: (polygon area) from Taubenbock et al. (2019);

mua\_aggname\_short: from Taubenbock et al. (2019), shortened if compound (multi-city) name;

mua\_cell\_count: number of 0.05° grid cells touching MUA polygon, including those with water > 50%;

mua\_mean\_lon: mean longitude of all grid cells in MUA, not used in analysis

mua\_mean\_lat: mean latitude of all grid cells in MUA, not used in analysis

region\_1: region of MUA’s country

economy: economy of MUA’s country

sub\_regions: sub-region of MUA’s country

**global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv**

• multiple variables for all grid cells in all MUAs included in the study.

NOTE: the working csv file of all grid data included a number of variables that were not used in the end. These have all been replace with 'NA' and their column names with NA1 through NA57

NOTE: to upload to github this was split into two parts: PART 1 has data for first 20,000 grid cells, PART 2 has data for remaining 14,880 grid cells. Concatentate these two files before running scripts.

• Grid cell properties and names:

longitude; latitude; mua\_objectid; mua\_fid1; mua\_country\_code; mua\_cellcount; mua\_worldpop\_thousand; mua\_area\_km2; mua\_country; mua\_aggname; grid\_area\_m2; water\_pct; region\_1; economy

• summer mean backscatter PR [PR = 10^[sigma\_0/10)] with sigma\_0 in dB; sigma\_0 values from Frolking et al. (2022):

ERS (1993-2000),

QSCAT (1999-2009 in northern hemisphere, 2000-2009 in southern hemisphere)

ASCAT (2007-2021)

• backscatter decadal trends computed with ‘lm’ in R

ers\_slope; ers\_slope\_stderr; ers\_slope\_r2; ers\_slope\_pval;

qscat\_slope; qscat\_slope\_stderr; qscat\_slope\_r2; qscat\_slope\_pval;

ascat\_10\_21\_slope; ascat\_10\_21\_slope\_stderr; ascat\_10\_21\_slope\_r2; ascat\_10\_21\_slope\_pval;

• WSF3D building metrics (from World Settlement Footprint 3D) (not used in three-decade analysis); second set of values is smoothed over 3x3 grid of grid cells, using mean smoothing

WSF3D\_Vol\_sum; WSF3D\_BF\_mean; WSF3D\_BA\_sum; WSF3D\_HT\_mean; WSF3D\_Vol\_sum\_3x3smooth; WSF3D\_BF\_mean\_3x3smooth; WSF3D\_BA\_sum\_3x3smooth; WSF3D\_HT\_mean\_3x3smooth

• WSF-evolution percent built for 1985-2015 (from World Settlement Footprint Evolution)

WSFEvolution\_1985\_mean, … , WSFEvolution\_2015\_mean

• annualized rate of increase in WSFEvolution built fraction:

del\_wsf\_bf\_93\_00; del\_wsf\_bf\_99\_09; del\_wsf\_bf\_10\_15

**GIS files:**

Figure ED8 is generated in QGIS using a shape file for MUA polygons from Taubenbock et al. (2019) (**Morphological\_Urban\_Area.shp**), and a raster geotiff file identifying all 0.05° lat/lon grid cells that intersect with the MUA polygon shape file (**mua\_mask.tif**).

A second set of raster geotiff files are used to make Fig. 1. These files are generated in one of the R codes below.

**PART 2. ANALYSIS CODES**

There are six R codes that generate most of the figures in the paper

Figure 1 is generated in QGIS using a geotiff raster files for gridded bivariate values. These geotiff files are generated in **bivariate\_choropleth\_plots\_v3.R** (below).

**BF\_&\_PR\_timeseries\_and\_v\_floor\_area\_and\_accel\_decel\_plots.R**

Reads four input files:

*• global\_china\_total\_floor\_area\_data.csv*

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

*• mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv*

*• mua\_country\_list\_economy\_region.csv*

Generates seven pdf files, two 12-panel time series figures (Fig. Extended Data ED1), one 8-panel BF and PR vs. floor area figures (Fig. ED2), a single panel time series for three cities in Syria (Fig. ED7), and three multi-panel pdf files of mean PR and BF interdecadal acceleration/deceleration arrows (Fig. 3), using intercalibrated sensor trends

*• backscatter\_&\_bf\_v\_floor\_area\_plots.pdf*

*• region\_mean\_timeseries\_PR\_WSFevo.pdf*

*• mua\_mean\_timeseries\_PR\_WSFevo.pdf*

*• syria\_city\_timeseries.pdf*

• *mua\_mean\_growth\_rate\_PR\_WSFevo.pdf*

• *region\_mean\_growth\_rate\_PR\_WSFevo.pdf*

• *subregion\_\_mean\_growth\_rate\_PR\_WSFevo.pdf*

**city\_backscatter\_PR\_v\_WSF3D\_plots.R**

Reads one input file:

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

Generates one pdf file containing three 12-panel scatterplot figures of ASCAT PR vs. smoothed WSF3D building volume, building height, and building fraction area (Fig. Supplemental S2)

*• mua\_12\_ascat\_v\_wsf3d.pdf*

**backscatter\_sensor\_PR\_trend\_intercalibration.R**

Reads three input files:

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

*• mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv*

*• mua\_country\_list\_economy\_region.csv*

Generates one pdf file containing a 4-panel scatterplot of scatterometer sensor PR trends for ‘overlapping years (Fig. ED9):

*• scatterometer\_trend\_intercalibration\_plot.pdf*

**Asia\_region\_BF\_&\_PR\_trends\_by\_pop\_bin.R**

Reads three input files:

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

*• mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv*

*• mua\_country\_list\_economy\_region.csv*

Generates one pdf file containing two 4-panel box-whisker plots of BF & PR trends by decade, binned by city population, for East Asia, China, Southeast Asia, and India (Fig. ED4):

*• regional\_grid\_trends\_binned\_by\_pop.pdf*

**bivariate\_choropleth\_plots\_v3.R**

Reads three input files:

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

*• mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv*

*• mua\_country\_list\_economy\_region.csv*

Generates 9 csv files that can be used to construct Table 1:

• *region\_bivariate\_cluster\_stats\_Y1\_Y2.csv* where Y1\_Y2 Is 93\_00 r 99\_09 or 10\_21

• *sub\_region\_bivariate\_cluster\_stats\_Y1\_Y2.csv* where Y1\_Y2 Is 93\_00 r 99\_09 or 10\_21

• *economy\_bivariate\_cluster\_stats\_Y1\_Y2.csv* where Y1\_Y2 Is 93\_00 r 99\_09 or 10\_21

Generates two pdf files containing a number of three panel pages of plots of grid cell number by growth type, city population, and decade, one page of each region or sub-region (Figs. 2, ED3, S1):

*• regional\_grid\_trends\_binned\_by\_pop.pdf*

Generates three geotiff files for loading into QGIS for the city map panels in Fig. 1

• *ascat\_intercal\_wsfevo\_10\_21\_6bin\_grid.tif*

• *ers\_intercal\_wsfevo\_10\_21\_6bin\_grid.tif*

• *qscat\_wsfevo\_10\_21\_6bin\_grid.tif*

**kmeans\_cluster\_analysis\_by\_decade.R**

Reads three input files:

*• global\_mua\_grid\_variables\_table\_wat\_lt\_50\_NA.csv*

*• mua\_city\_list\_lat\_lon\_clean\_Jul2023.csv*

*• mua\_country\_list\_economy\_region.csv*

Generates 2 csv files: one giving stats on the clusters (similar to info in box-whisker plots in Fig. ED5), and includes the number of grid cells per cluster per decade (used in Table 2), and the second giving number of grid cells per region, sub-region, and country (used in Figs. 4, ED6, S3)

• *grid\_1990s\_5\_&\_2000s\_5\_&\_2010s\_5kmeans\_cluster\_stats\_array.*

• *grid\_1990s\_5\_&\_2000s\_5\_&\_2010s\_5\_region\_and\_sub\_region\_and\_country\_num\_grids.csv*

Generates one pdf file containing a series of alluvial plots of k-means cluster transitions between decades: global, regional, sub-regional (used to make Figs. 4, ED6, S3):

*• grid\_1990s\_5\_&\_2000s\_5\_&\_2010s\_5\_kMeans\_cluster\_alluvial.pdf*

Generates one pdf file containing a series box-whisker plots that characterize the k-means clusters (used to make Fig. ED5):

*• grid\_1990s\_5\_&\_2000s\_5\_&\_2010s\_5\_BFs\_&\_PRs\_state\_&\_change\_cluster\_box\_plots.pdf*

Generates three geotiff files of the clusters by decade (note that cluster 1 typology in one decade may not match cluster 1 typology in another decade, so care is needed with these)

• *kmeans\_cluster\_5\_1990s\_raster\_grid.tif*

• *kmeans\_cluster\_5\_2000s\_raster\_grid.tif*

• *kmeans\_cluster\_5\_2010s\_raster\_grid.tif*

**PART 3. REFERENCES**

H. Taubenböck, M. Weigand, T. Esch, J. Staab, M. Wurm, J. Mast, S. Dech, A new ranking of the world’s largest cities—Do administrative units obscure morphological realities? *Remote Sens. Environ.* **232**, 111353 (2019). [data in online supplement]

S. Frolking, T. Milliman, R. Mahtta, A. Paget, D. G. Long, K. C. Seto, Global Monthly and Seasonal Urban and Land Backscatter Time Series, 1993-2020, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) (2022); <https://doi.org/10.7927/GR2E-DH86>.

IEA, “World Energy Outlook 2022” (Paris, 2022); <https://www.iea.org/reports/world-energy-outlook-2022>.

*China Statistical Yearbook 2022* (China Statistics Press, 2022; <http://www.stats.gov.cn/sj/ndsj/2022/indexeh.htm>.

*Beijing Statistical Yearbook 2021* (China Statistics Press, Beijing Municipal Bureau of Statistics and NBS Survey Office in Beijing, 2021; <https://nj.tjj.beijing.gov.cn/nj/main/2021-tjnj/zk/indexeh.htm>.

*Shanghai Statistical Yearbook 2020* (2020; <https://tjj.sh.gov.cn/tjnj/zgsh/tjnj2020en.html>.

World Settlement Footprint 3D: data are available at <https://download.geoservice.dlr.de/WSF3D/files/>.

World Settlement Footprint Evolution data are available at <https://download.geoservice.dlr.de/WSF_EVO/>.