SUDS
Sea level rise and Urban infrastructure Data Set

Allan Hsiao
Stanford University

January 11, 2025

The Sea level rise and Urban infrastructure Data Set (SUDS) measures the exposure of urban infrastructure to sea level rise of up to 5 meters. The SUDS data have global coverage and include 11,422 urban clusters with a total population of 3.7 billion. This file describes the data, its construction, and the code for replication. I link the relevant files and the original paper below.

Data: `https://allanhsiao.com/files/suds/data.zip`
Code: `https://allanhsiao.com/files/suds/code.zip`

Paper: `https://allanhsiao.com/files/Hsiao_suds.pdf`

Citation

Hsiao, Allan. 2025. "Sea Level Rise and Urban Infrastructure." *AEA Papers and Proceedings*, 115.

1 Data

The `suds.csv` file contains the data. Each row records sea level rise exposure for land or infrastructure in a city or part of a city. Columns 1 through 7 are city-level information (table 1). If rows have the same value in column 1, then they have the same values in columns 2 through 7. Column 8 breaks out calculations within cities, classifying cells by quartile of night light intensity within each city (table 2). Column 9 is whether calculations are for land or a particular type of infrastructure (table 3). Column 10 is total quantities, and columns 11 through 61 are quantities impacted by sea level rise of 0 to 5 meters in 0.1-meter increments (table 4). I round land areas and road lengths to the nearest integer.

Email: `ajhsiao@stanford.edu`. Xilin Fan provided exceptional research assistance.

Table 1: City-level infomation

Column	Label	Description
1	g	City population rank
2	ghs	City ID in GHS-UCDB
3	city	City name
4	country	Country name
5	area	City land area in square kilometers
6	pop	City population in 2025
7	gdp	Country GDP quartile among all countries in 2020

Column 7 (gdp) takes values of q1, q2, q3, and q4, which record quartiles. The lowest quartile is q1, and the highest quartile is q4.

Table 2: Within-city breakdown

Column	Label	Description
8	light	Night light intensity
	all	All cells in a city
	q1	Lowest quartile among all cells in a city
	q4	Highest quartile among all cells in a city

The table omits q2 and q3, which are defined analogously.

Table 3: Land and infrastructure

Column	Label	Description
9	var	Variable
	land	Square kilometers of land area
	school	Number of schools
	hospital	Number of hospitals
	clinic	Number of clinics
	highway	Kilometers of highways
	primary	Kilometers of primary roads

The table omits college, university, kindergarten, doctors, dentist, pharmacy, secondary (road), and tertiary (road), which are defined analogously.

Table 4: Exposure to sea level rise (SLR)

Column	Label	Description
10	total	Total quantity
11	slr0	Quantity inundated with SLR of 0 meters
21	slr10	Quantity inundated with SLR of 1 meters
61	slr50	Quantity inundated with SLR of 5 meters

The table omits slr1 to slr9 and slr11 to slr49, which are defined analogously.

Table 5: Input data sources

Data	Year	Source	Citation
Cities	2025	GHS-UCDB	Uhl et al. (2024)
Elevation	2011-15	DeltaDTM	Pronk et al. (2024)
Coastlines	∼2017	GSHHG	Wessel and Smith (1996)
Roads	∼2018	GRIP	Meijer et al. (2018)
Education	2024	OSM	OpenStreetMap (2024)
Health	2024	OSM	OpenStreetMap (2024)
Night lights	2023	EOG	Elvidge et al. (2021)

2 Methodology

The SUDS data are derived from a number of frontier, high-resolution spatial data sets. Table 5 lists these input data sources. I lightly process each as needed to obtain raster data at a resolution of 1 arc second – roughly 30 meters at the equator.

Cities

I define cities with the urban center boundaries of the Global Human Settlement Urban Center Database (GHS-UCDB). I use version R2024A, as released on May 11, 2024. This data set compiles a broad range of socioeconomic and other indicators for the urban centers delineated in the GHS-SMOD R2023A settlement layer, which in turn quantifies the degree of urbanization using population and build-up surface data from GHS-POP R2023A and GHS-BUILT-S R2023A. An urban center is a contiguous cluster of cells with a population density of at least 1,500 per square kilometer and a total population of at least 50,000. Boundaries are smoothed, and holes of less than 15 square kilometers are filled. Populations are 2025 populations. The GHS-UCDB contains 11,422 urban centers that I refer to as "cities." The urban boundaries are available as polygons, which I rasterize to a resolution of 1 arc second. I draw land areas, 2025 populations, and 2020 GDPs from the same GHS-UCDB data set, and I recode the GDP values into quartiles. I note that the underlying GHS-SMOD data are of 1-kilometer resolution – roughly 30 arc seconds. In rasterizing the GHS-UCDB polygons, I downscale them from 30- to 1-arc-second resolution. I use high-resolution elevation data to refine these data by dropping pixels that are within the (coarse) GHS urban boundary but coded as water in the (fine) elevation data.

GHS-UCDB: https://human-settlement.emergency.copernicus.eu/ghs_ucdb_2024.php

Elevation

Elevation data are from the DeltaDTM global coastal digital terrain model. I use version 4, as published on December 12, 2024. These data focus on the Low Elevation Coastal Zone: low-lying coastal areas with elevation less than 10 meters above mean sea level. These data are already available as raster data at my desired resolution of 1 arc second. The data are divided into 1-degree tiles, which I merge and reproject as needed to obtain elevation over space for each city of interest. The data isolate underlying terrain by removing elevation biases from the Copernicus GLO-30 Digital Surface Model, which measures the surface of the Earth with data acquired from 2011 to 2015 on the TanDEM-X mission. The processed data are then verified with airborne lidar validation data from around the world. The underlying Copernicus data are available globally (except for Armenia and Azerbaijan) at resolutions of 1 and 3 arc seconds. They are available for Europe at even finer resolution. The Copernicus data also include water body mask tiles that distinguishes land, ocean, lakes, and rivers. The DeltaDTM data include these mask tiles alongside the processed elevation tiles.

DeltaDTM: https://data.4tu.nl/datasets/1da2e70f-6c4d-4b03-86bd-b53e789cc629
Copernicus: https://dataspace.copernicus.eu/explore-data/data-collections/copernicus-contributing-missions/collections-description/COP-DEM

Coastlines

Coastline data are from the Global Self-consistent, Hierarchical, High-resolution Geography database (GSHHG). The data were originally assembled in 1996 and have been updated over time. I use data last updated on June 15, 2017. I use these data to distinguish land from ocean for areas not covered by the elevation data, and also to calculate distance to the ocean. The crude-resolution version of the data help to speed computation, while remaining sufficiently accurate for these purposes.

GSHHG: https://www.soest.hawaii.edu/pwessel/gshhg

Infrastructure

Education and health infrastructure data are from OpenStreetMap (OSM). Education infrastructure includes amenities tagged as "school," "kindergarten," "college," or "university." Schools do not broadly distinguish between primary, middle, and high schools, although some schools record the level of education based on the International Standard Classification of Education (ISCED). Colleges are institutes of continuing or further education. Health infrastructure includes amenities tagged as "hospital," "clinic," "doctors," "dentist," or "pharmacy." OpenStreetMap details these and other amenity tags below.

OSM: https://www.openstreetmap.org
Amenities: https://wiki.openstreetmap.org/wiki/Key:amenity

I obtain the global universe of data from Geofabrik, which hosts OpenStreetMap data extracts by continent. I use the December 21, 2024 version of the data, which are updated daily. I filter the data with the Osmium Tool to extract points and polygons – OpenStreetMap "nodes" and "ways" – that correspond to my amenity tags of interest. I omit points that intersect a polygon, and I reduce polygons to their centroids. For each amenity type, I rasterize the resulting list of points to a resolution of 1 arc second by counting the number of points that fall within each cell.

Geofabrik: https://download.geofabrik.de
Osmium: https://osmcode.org/osmium-tool

Road infrastructure data are from the Global Roads Inventory Project (GRIP). I use version 4, as published on May 23, 2018. These data integrate nearly 60 geospatial datasets on road infrastructure to construct global data covering 222 countries and more than 21 million kilometers of roads. The data were published in 2018 and aim for the most recent temporal coverage possible. The data distinguish between highways, primary roads, secondary roads, tertiary roads, and local roads. The data are available as vector data, which I rasterize for each road type to a resolution of 1 arc second. I omit local roads to limit computational complexity.

GRIP: https://www.globio.info/download-grip-dataset

Night lights

Night light data are from the Earth Observation Group (EOG). These data are constructed from the Visible and Infrared Imaging Suite (VIIRS) Day Night Band (DNB) data from the Joint Polar-orbiting Satellite System (JPSS). I use the Annual VIIRS Nighttime light (VNL) V2 data, which aggregate over monthly cloud-free average radiance grids. I use version 2.2, as last modified on March 1, 2024. I focus on masked averages for 2023. The data form a raster with a resolution of 15 arc seconds, which I downscale to a resolution of 1 arc second.

EOG: https://eogdata.mines.edu/products/vnl
VIIRS: https://www.earthdata.nasa.gov/data/instruments/viirs

Exposure

I simulate inundation under sea level rise with a simple model that combines the elevation data with the principle of hydrologic connectivity. I treat cities as safe from inundation if they lie outside of the Low Elevation Coastal Zone or if they are more than 500 kilometers from the ocean. For each sea level rise scenario, I overlay predicted inundation with city boundaries to compute the amount of land area at risk of inundation. Similarly, I overlay predicted inundation with the infrastructure data to compute exposure for infrastructure. Hsiao (2025) details the model and exposure calculations.

3 Code

The main code file is `main.jl`. It calls code files in the `code` folder to process input data in the `data` folder. The `code` folder contains the `city.jl`, `coast.jl`, `elevation.jl`, `facility.jl`, `Geo.jl`, `light.jl`, `road.jl`, and `sea.jl` files. The `data` folder contains the `delta`, `ghs`, `grip`, `gshhg`, `osm`, and `viirs` subfolders. The `fig` and `tab` folders hold intermediate output: city-specific maps in `fig` and city-specific data in `tab`. The final output is `suds.csv`, which compiles SUDS data for all cities.

All code is written in Julia. I note dependency on the packages `ArchGDAL`, `CSV`, `DataFrames`, `LinearAlgebra`, `Plots`, `Rasters`, and `StatsBase`. I detail individual code files as follows.

1. `main.jl` executes the other code files and compiles the data for each city. For cities within the Low Elevation Coastal Zone and within 500 kilometers of the ocean, it simulates sea level rise with an elevation-based inundation model. It maps inundations for sea level rise of 0, 1, 2, 3, 4, and 5 meters, and it exports these maps to `fig/map*.png`. It treats other cities as safe from sea level rise. For all cities, it overlays city boundaries, infrastructure, and inundation to calculate the total amount of infrastructure and the amount impacted by each inundation scenario. It exports these calculations to `tab/data*.csv`. Finally, it collects the city-specific calculations and exports the result to `suds.csv`.

2. `code/Geo.jl` collates a number of geospatial helper functions.

3. `code/city.jl` reads city boundaries and demographics from the socioeconomic thematic domain subset of the GHS-UCDB data.

    `data/ghs/GHS_UCDB_THEME_SOCIOECONOMIC_GLOBE_R2024A.gpkg`

It sorts cities by population rank. Then for each city, it extracts the boundary polygon and other information, and it rasterizes the polygon.

4. `code/coast.jl` reads ocean coastlines from the GSHHG data.

   ```
   data/gshhg/GSHHS_c_L1.dbf
   data/gshhg/GSHHS_c_L1.prj
   data/gshhg/GSHHS_c_L1.shp
   data/gshhg/GSHHS_c_L1.shx
   ```

   For a given city, it returns the closest piece of coastline, and it identifies whether the city is inland or far inland. Inland is within the coastline, and far inland is 500 kilometers or more within the coastline.

5. `code/elevation.jl` reads elevation and mask tiles from the DeltaDTM data.

   ```
   data/delta/deltadtm_tiles.gpkg
   data/delta/Africa/DeltaDTM_v1_1_N00E006.tif [...]
   data/delta/masks/DeltaDTM_v1_1_N00E006.tif [...]
   ```

   For a given city, it identifies whether the city is covered by the DeltaDTM data, it uses `deltadtm_tiles.gpkg` to identify the relevant tiles, and it reads and merges these tiles.

6. `code/sea.jl` simulates inundation under sea level rise of 0 to 5 meters. It considers sea level rise scenarios in 0.1-meter increments, and it returns inundation in each scenario as a boolean raster.

7. `code/facility.jl` reads education and health facilities from the filtered OSM data, which are divided by continent and facility type.

   ```
   data/osm/africa-clinic.osm.pdf [...]
   ```

   For a given city, it extracts OSM points and polygons, drops points that overlap with polygons, converts polygons to centroid points, and rasterizes all points by facility type.

8. `code/road.jl` reads roads from the GRIP data, which are divided by region.

   ```
   data/grip/GRIP4_Region1_vector_shp/GRIP4_region1.shp [...]
   ```

   For a given city, it extracts roads, computes total road length, and rasterizes the roads by type.

9. `code/light.jl` reads night light intensity from the EOG VIIRS data.

   ```
   data/viirs/VNL_npp_2023_global_vcmslcfg_v2_c202402081600.aver
   age_masked.dat.tif
   ```

   For a given city, it extracts and downscales the relevant part of the global raster.

4 Replication

1. Download the input data. All data are publicly available – see table 5. The replication files do not include the input data.
2. Filter OSM planet data with the Osmium Tool, as demonstrated by the shell commands below.

```
continents=("africa" "antarctica" "asia" "australia-oceania"
 "central-america" "europe" "north-america" "south-america")

amenities=("kindergarten" "school" "college" "university"
 "pharmacy" "hospital" "doctors" "dentist" "clinic")

for c in "${continents[@]}"; do
 for a in "${amenities[@]}"; do
  echo "$c, $a"
  input="${c}-latest.osm.pbf"
  output="${c}-${a}.osm.pbf"
  osmium tags-filter "$input" "w/amenity=$a" -o "$output"
 done
done
```

3. Place the input data in the relevant subfolders of the `data` folder. The previous section specifies the data files and subfolders that correspond to each code file.
4. Run `main.jl` to execute the code. Total runtime is roughly 4 days for all cities.
5. Obtain `suds.csv`, which should match the SUDS data.

5 Changelog

1. January 11, 2025. Initial release.

References

Elvidge, Christopher, Mikhail Zhizhin, Tilottama Ghosh, Feng-Chi Hsu, and Jigar Taneja. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. *Remote Sensing*, 13(5):922, 2021.

Hsiao, Allan. Sea Level Rise and Urban Infrastructure. *AEA Papers and Proceedings*, 115, 2025.

Meijer, Johan, Mark Huijbregts, Kees Schotten, and Aafke Schipper. Global Patterns of Current and Future Road Infrastructure. *Environmental Research Letters*, 13 (6):064006, 2018.

OpenStreetMap. Planet data extracts retrieved from https://download.geofabrik.de. https://www.openstreetmap.org, 2024.

Pronk, Maarten, Aljosja Hooijer, Dirk Eilander, Arjen Haag, Tjalling de Jong, Michalis Vousdoukas, Ronald Vernimmen, Hugo Ledoux, and Marieke Eleveld. DeltaDTM: A Global Coastal Digital Terrain Model. *Scientific Data*, 11:273, 2024.

Uhl, Johannes, Martino Pesaresi, Luca Maffenini, Patrizia Sulis, Monica Crippa, Diego Guizzardi, Enrico Pisoni, Claudio Belis, Duarte Oom, Alfredo Branco, Dennis Mwaniki, Edwin Kochulem, Daniel Githira, Alessandra Carioli, Daniele Ehrlich, Pierpaolo Tommasi, Thomas Kemper, and Lewis Dijkstra. Stats in the City GHSL Urban Centre Database 2025, 2024.

Wessel, Paul and Walter Smith. A Global, Self-Consistent, Hierarchical, High-Resolution Shoreline Database. *Journal of Geophysical Research: Solid Earth*, 101 (B4):8741–8743, 1996.