README for the Reproducibility Package for

The evolution of flood exposure and protection gaps in Germany

Overview

This repository contains **three Jupyter notebooks** that reproduce all figures and maps presented in the paper:

- `Cleaning_up_Germany_Land_2020.ipynb`
- `Cleaning_up_Germany_Geimende_plots_maps.ipynb`
- `Cleaning_up_Germany_Gemeinde_Climate_Scenarios.ipynb`

Each notebook includes an **"Open in Colab"** button at the top.
Running the notebooks on Google Colab is the primary and recommended option for replication, as it provides a stable execution environment and simplifies package installation.

The notebooks are designed to be run independently, but they share common data inputs and intermediate processing steps. Several figures rely on large raw datasets and intermediate variables that are saved and reloaded during execution.

Data Availability

Data Availability

Publicly available data

Germany administrative boundary data were obtained from the Federal Agency for Cartography and Geodesy (BKG).
These data provide administrative boundary layers for Germany and are used to define spatial units in this study.
The dataset is publicly accessible at:
https://gdz.bkg.bund.de/index.php/default/open-data/verwaltungsgebiete-1-250-000-mit-einwohnerzahlen-stand-31-12-vg250-ew-31-12.html

(accessed May 2025).

To ensure reproducibility, the exact version of the dataset used in this study is included in the restricted-access data folder.

Data with restricted access

The World Settlement Footprint (WSF) Tracker dataset was developed by the German Aerospace Center (DLR).

The WSF Tracker provides bi-annual global maps of built-up areas at 10 m resolution from July 2016 to January 2025.

Access to the data is subject to DLR licensing conditions and can be requested via:
https://www.dlr.de/en/eoc/research-transfer/projects-missions/world-settlement-footprint-wsf-r

Flood data is based on the 2022 Fathom-Global 3.0 flood hazard dataset.
These data provide model-based flood inundation extents as well as depths for multiple flood return periods and are used to assess flood exposure in this study.
Access is subject to a commercial license and can be requested from:
https://www.fathom.global

Germany's dike location data were provided by the German government (Bundesanstalt fur Gewasserkunde, BfG)
These data describe the spatial distribution of structural flood defense and are used to account for flood protection in the analysis.
The data are not publicly available but may be viewed on the BfG online platform at
https://geoportal.bafg.de/karten/HWRM_Aktuell/?tabs=on

Access should be requested through the responsible government agencies.

Dataset-File mapping

Dataset name	Filename(s)	Description	Access type
Germany Administrative Boundaries (BKG VG250)	vg250_01-01.gk3.shape.ebenen.zip	Administrative boundary layers for Germany	Public
Flood Exposure and Projected Flood Exposure in Germany - Updated supplementary data for Global evidence of rapid urban growth in flood zones since 1985 (https://www.nature.com/articles/s41586-023-06468-9)	Land_all_return_periods.zip	Processed flood exposure under different return-period scenarios, aggregated at the state level	Restricted
	Gemeinde_all_return_periods.zip Gemeinde_Stats_all_return_periods.zip	Processed flood exposure under different return-period scenarios, aggregated at the Gemeinden level	
	most_optimistic.zip moderate_optimistic.zip moderate_pessimistic.zip most_pessimistic.zip	Projected flood exposure under the 100-year return period across multiple climate-change scenarios, aggregated at the Gemeinden level	
Germany dike and dune location data	dike_dune_GIS_file/	Spatial distribution of structural flood defenses	Restricted

Statement about Rights:

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

Instructions for Replicators

Option 1 (Recommended): Run on Google Colab

Google Colab is the **preferred replication environment**.

1. Click the **"Open in Colab"** button at the top of any notebook.
2. Run the first two cells to install and import all required packages.
3. Follow the **highlighted instructions inside each notebook**, which indicate:
    - when to compute large intermediate variables and store them as .pkl files,
    - when to reload previously saved .pkl files,
    - and which sections must be re-run if a session restarts.

Colab Memory Limitations

The Colab free tier has limited RAM. This primarily affects:

- `Cleaning_up_Germany_Geimende_plots_maps.ipynb`
- `Cleaning_up_Germany_Gemeinde_Climate_Scenarios.ipynb`

These notebooks process large raw files and create memory-intensive intermediate variables.

What to Do If Colab Runs Out of RAM

If the session restarts:

1. Re-run the **package installation and import cells** (first two cells).
2. Reload the **saved .pkl files** containing key intermediate variables (These variables should still be available and visible in the Colab sidebar even after the session restarts.).
3. Re-run only the sections explicitly indicated in the notebook.

Important:
Saving and reloading intermediate variables is required on Colab. Without this step, Colab will repeatedly recompute memory-intensive objects and fail before completing the workflow. Based on experimental testing, **the session typically needs to restart only once** for each of the two computationally heavy notebooks.

Option 2: Run Locally (Advanced Users)

Users may also run the notebooks locally. This avoids Colab memory limitations but requires manual environment setup.

1. Install all required Python packages listed in the **Requirements** section.
2. Ensure sufficient RAM to load large intermediate datasets.
3. Run code cells sequentially without restarting the kernel.

List of Exhibits

The provided notebooks reproduce:

- All numbers provided in text in the paper
- All tables and figures in the paper

Exhibit	**Output**	**Notebook**	**Notes**
Figure 1 & 2; Table 1	.png csv	`Cleaning_up_Germany_Land_2020.ipynb`	National- and state-level plots; Table 1 is re-summarized from `max_table_data.csv` processed in the "Combined flooding (max) – Make an order for plotting" section.
Figure 3 – 8; Figure 13 & 14	.png	`Cleaning_up_Germany_Geimende_plots_maps.ipynb`	Gemeinde-level plots and maps, including comparisons between flooding exposure and dike protection.
Figure 9 - 12	.png	Cleaning_up_Germany_Gemeinde_Climate_Scenarios.ipynb	Plots and maps of projected flooding exposure under different climate scenarios, at national and Gemeinde levels.

Requirements

Computational Requirements

- **Recommended environment:** Google Colab

- **Local execution:** Any OS capable of running Python and Jupyter Notebook

Software Requirements

If running locally, install Python and the required packages listed below.
Package versions correspond to those used on Google Colab at the time of development.

```
pandas==2.2.2
numpy==2.0.2
matplotlib==3.10.0
seaborn==0.13.2
plotly==5.24.1
kaleido==1.2.0
geopandas==1.1.1
shapely==2.1.2
fiona==1.10.1
contextily==1.7.0
mapclassify==2.10.0
folium==0.20.0
ipyleaflet==0.20.0
ipywidgets==7.7.1
scipy==1.16.3
statsmodels==0.14.6
arviz==0.22.0
pymc==5.26.1
xarray==2025.12.0
brokenaxes
matplotlib-scalebar==0.9.0
Pillow==11.3.0
```

`moviepy==1.0.3`

`geo_northarrow @ git+https://github.com/pmdscully/geo_northarrow.git`

Memory, Runtime, and Storage

Memory:

Colab (12.67GB RAM under free-tier): limited; requires intermediate variable caching

Runtime:

Individual notebooks may take 10 - 60 minutes depending on environment

Storage:

Additional space required for intermediate .pkl files and temporary outputs

Folder Structure

The repository has the following structure:

Germany_Flood_Study

- Visuals
    - visual_scripts
        - `Cleaning_up_Germany_Land_2020.ipynb`
        - `Cleaning_up_Germany_Geimende_plots_maps.ipynb`
        - `Cleaning_up_Germany_Gemeinde_Climate_Scenarios.ipynb`
- README

All three Jupyter notebooks used to generate the figures and maps are located in the Visuals/visual_scripts folder.

Data Placement and File Paths (Important)

After requesting and obtaining access to the restricted datasets, users must store the data in a location that is accessible to the notebooks and **update the file paths in the notebooks accordingly**.

This step is required for the notebooks to run correctly.

Running on Google Colab (Recommended)

When running the notebooks on Google Colab, users may choose either of the following options:

1. Upload the data files directly to the Colab session (temporary storage), or
2. Store the data in Google Drive and mount Google Drive in Colab.

In both cases, users **must modify the file paths in the notebooks** to match the location where the data are stored before running the data-loading cells.

Running Locally

Alternatively, users may download and store the data on their local machine and run the notebooks locally after installing the required Python environment.

As with the Colab workflow, users **must modify the file paths in the notebooks** to point to the local data directory.