Quantifying the current and future extent and impacts of urban gullies in Africa

A Data Management Plan created using DMPonline.be

Creator: Elise Dujardin

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 11PHF24N

ID: 204957

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

The rapid and typically uncontrolled growth of many African cities leads to a plethora of problems and challenges. One of these is the formation and expansion of large urban gullies (UGs) in many (sub)tropical cities. UGs typically lead to the destruction of houses and other infrastructures, displace large numbers of people and often claim casualties. As such, UGs form a new type of geo-hydrological hazard in the Anthropocene. Yet, this issue remains largely under-researched. So far we observed UGs in more than 100 cities in Africa, but the actual scale and impacts of UGs across the continent are currently unknown. Moreover, further urbanization and climate change will likely aggravate the problem. This project will help tackle these emerging research needs by: (i) better documenting UGs and their impacts across Africa; (ii) constructing a model that can simulate the initiation and expansion of UGs in relation to its key driving factors; (iii) applying this model at the scale of Africa to identify current hotspots of UGs and to quantify some of the key impacts (e.g. number of persons directly displaced by UGs); and (iv) conducting scenario analyses with our model to assess how the potential impacts of UGs will evolve as a result of continued urbanization and climate change.

Last modified: 09-04-2024

Quantifying the current and future extent and impacts of urban gullies in Africa
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Domain	Raster file showing the delineation of the domain	Generate new data	Digital	Compiled/aggregated data Simulation data	.geotiff	<100 MB	
UG head location	Mapped UG heads with their relevant metadata (e.g. Date)	Generate new data	Digital	Observational	.kmz/.shp	<100 MB	
UG head occurrence	Sites of 1 km² with information on the presence or absence of UG (0/1) to calibrate and validate the susceptibility and probability models	Generate new data	Digital	Observational	.shp	<100 MB	
UG susceptibility map	Raster file with susceptibility values of UG head occurrence	Generate new data	Digital	Simulation data	.geotiff	<100 GB	

UG probability map	Raster file with probability values of having an UG	Generate new data	Digital	Simulation data	.geotiff	<100 MB	
Variables	Raster files of variables describing the natural environment and the human context	Reuse existing data	Digital	Observational Simulation data	.geotiff	<1 TB	
UG stable extent	Mapped UG at a stable size	Generate new data	Digital	Observational	.kmz/.shp	<100MB	
UG extents	Mapped UG extent for specific time serie	Generate new data	Digital	Observational	.kmz/.shp	<1 GB	
UG model output	Simulated proportion of pixel affected by UG for a period from 2000 to 2023 with a timestep of 5 years	Generate new data	Digital	Simulation data	.geotiff	<100 GB	
Input layer for scenarios	Raster files of relevant layers used as an input in our UG model with scenarios for climate change and urban expansion	Reuse existing data	Digital	Simulation data	.geotiff	<1 TB	
Output of scenarios analyses	Simulated proportion of pixel affected by UG for a period from 2023 to 2100	Generate new data	Digital	Simulation data	.geotiff	<100 GB	
Code	Python scripts to run and validate our models	Generate new data	Digital	Software	.txt/.py	<100MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

First sources of raster layers used as input to our models (list to be completed during the project):

GHSL Data Package 2023, Publications Office of the European Union, Luxembourg, 2023, JRC133256, ISBN 978-92-68-02341-9 https://doi.org/10.2760/098587

Hengl, T., Miller, M.A.E., Kriˇzan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijevi´c, O., Gluˇsica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G. E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11 https:// doi.org/10.1038/s41598-021-85639-y.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific data, 5(1):1–15. https://doi.org/10.1038/sdata.2018.40

Stefan Lange, Christoph Menz, Stephanie Gleixner, Marco Cucchi, Graham P. Weedon, Alessandro Amici, Nicolas Bellouin, Hannes Müller

Schmied, Hans Hersbach, Carlo Buontempo, Chiara Cagnazzo (2021): WFDE5 over land merged with ERA5 over the ocean (W5E5 v2.0). ISIMIP Repository. https://doi.org/10.48364/ISIMIP.342217

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

Metadata documentation will be provided for each dataset, presented in the form of README.txt files and/or spreadsheets. This documentation will encompass two layers of information: 1) general details about the study as a whole, including its title, description, methodology, keywords, time period, spatial extent and resolution, projection, contact information, etc. 2) Specific descriptions of data collected (with various attributes) or variables/model output, accompanied by their respective units of measurement.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

The choice of a metadata standard or guidelines will ultimately be determined based on the data repository used. We expect, however, that the metadata will closely follow field-specific standards such as the Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata, NASA's Earth Science Data Systems' ASCII File Format Guidelines for Earth Science Data, or the International Organization for Standardization's (ISO) 19115-1:2014 Geographic Information – Metadata – Part 1: Fundamentals.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on multiple storage platforms, including OneDrive with a 2TB capacity, SharePoint with a 5TB capacity, and external hard drives situated at various locations, including the home of the PhD candidate and the office of the promoter.

How will the data be backed up?

The data will be stored on OneDrive and regularly backed up on external hard drives.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

Yes, there is currently sufficient storage and backup capacity (2TB capacity on OneDrive + 5TB capacity on SharePoint + 5TB capacity on hard drive) during the project to store the expected amount of data needed (~ 3TB).

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Access to both the original data and its backups will be tightly controlled, restricted to the principal investigator (the PhD candidate) and the promoter. In cases where selected portions of the data must be shared, such as with Bachelor's and Master's students, a separate version of the data will be created.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The expected costs for data storage and backup during the research project is ~300 euros (price of external hard drives). These costs will be covered by the PhD candidate's FWO bench fee.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

The last version of collected data, important mid-output and all final output data will be retained for at least five years.

Where will these data be archived (stored and curated for the long-term)?

The data will be preserved on the large volume storage services of KU Leuven (https://icts.kuleuven.be/sc/english/storage/large-volume-storage) and on external hard drives, kept at different physical locations.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

The expected costs for data preservation during the expected retention period is ~300 euros/year. The promotor of the research has dedicated funding to cover these costs.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository

Most of the data will be made available for reuse after publication:
- UG databases: vector file of location and extents of UG and their relevant properties and metadata.
- Model output: raster files of relevant environmental variables and final model outputs on the susceptibility to UG and proportion of pixel affected by UG for a period from 2000 to 2023 with a timestep of 5 years; raster files of relevant layers used as an input in our UG model with scenarios for climate change and urban expansion and final model ouputs of proportion of pixel affected by UG for a period from 2023 to 2100.
- Code and programs: Python scripts developed to run our model.

If access is restricted, please specify who will be able to access the data and under what conditions.

NA

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

NA

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Our key results (e.g. UG databases, model, final output) will be made freely available via Open Access platforms (e.g. KU Leuven RDR platform) and as appendices to our publications.

When will the data be made available?

The data will be made available during and after the final stages of the research project.

Which data usage licenses are you going to provide? If none, please explain why.

As we abide to an open-access philosophy, we plan to provide an Open Database License (ODbL) for all final output.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

The expected costs for data sharing is ~200 euros. These costs will be covered by the the promotor's dedicated funding and the university.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

During the research project, the PhD candidate will manage data documentation and metadata.

Who will manage data storage and backup during the research project?

During the research project, the PhD candidate will manage data storage and back up.

Who will manage data preservation and sharing?

After the research project, the promoter, in consultation with the PhD candidate, will manage data preservation and sharing.

Who will update and implement this DMP?

The PhD candidate, in consultation with the promoter, will update and implement this DMP.