
Real-time hybrid modelling of urban pluvial flood hazard maps using machine learning

A Data Management Plan created using DMPonline.be

Creator: Daan Buekenhout

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1SHFS24N

ID: 206580

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Extreme precipitation events are projected to become more frequent and intense due to the ongoing climate change. The result of these extreme events can be devastating (e.g. July 2021). The impact of pluvial floods are further exacerbated by an increasing urbanisation, also in Flanders. Due to the short response times of the hydrological system in highly urbanised catchments and the high economic and societal value of city centres, these areas are especially vulnerable to the impact of extreme precipitation. Therefore, mitigation measures such as performant flood warning systems are of high importance. Detailed real-time estimations of floods are valuable for emergency operation management, but are, however, nonexistent in most areas around the world.

The development of performant probabilistic real-time pluvial flood forecasting systems typically entails (i) probabilistic precipitation forecasting and (ii) deterministic real-time flood modelling. The focus of this research proposal is on the creation of real-time flood models. Given the long calculation times of state-of-the-art detailed flood models, real-time alternatives based on machine learning are investigated. Novel surrogate, observation-based and hybrid models will be implemented and evaluated. The city of Antwerp will be considered as case study area because the city is highly prone to flood, with large socio-economic impacts due to urban flood hazard.

Last modified: 18-04-2024

Real-time hybrid modelling of urban pluvial flood hazard maps using machine learning

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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Generate new data • Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Digital • Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA 	
Rain gauges	Rainfall observations from ground stations	Reuse existing data	Digital	Observational (timeseries)	.csv	< 100 GB	
Rainfall radar estimations	Rainfall estimates derived from radar observations	Reuse existing data	Digital	Observational (timeseries)	.hdf, .nc	< 5 TB	
Rainfall radar forecasts	Forecasts provided by RMI	Reuse existing data	Digital	Simulation data (timeseries)	.nc	< 5 TB	
Geospatial data	Digital elevation data, land cover data, ...	Reuse existing data	Digital	Observational (geospatial)	.shp, .tif	< 1 TB	
Hydrodynamic model (subsurface system)	Physically-based models for flood simulations (subsurface component)	Reuse existing data	Digital	Software (model + parameters)	NA	< 100 GB	
Hydrodynamic model (surface component)	Physically-based models for flood simulations (surface component)	Generate new data	Digital	Software (model + parameters)	NA (parts: .tif, .shp)	< 100 GB	

Synthetic storms	Synthetics rainfall generated using statistical properties	Generate new data	Digital	Simulation data	.csv, .hdf, .tif, .pkl	< 5 TB	
Flood maps	Produced by hydrodynamic model, used as training data for data driven models	Generate new data	Digital	Simulation data (geospatial)	.shp, .tif, .csv	< 50 TB	
Data driven models	Program code and selected parameters + democases	Generate new data	Digital	Software (model + parameters)	.py, .pkl, .ipynb	< 1 TB	

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:

1. Rain gauges: publicly offered by VMM and HIC (<https://www.waterinfo.be/>)
2. Rainfall radar estimations & forecasts: offered by RMI (nonpublic), (parts of) datasets will become publicly available in the near future (<https://opendata.meteo.be/>)
3. Geospatial datasets: publicly offered by Geopunt (<https://www.vlaanderen.be/geopunt/catalogus>)
4. Hydrodynamic model (subsurface component): shared by Aquafin (nonpublic)

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.

- Yes

Parts of the hydrodynamic model for the city of Antwerp was conceived by Aquafin, who share their model for research projects at KU Leuven. No formal restrictions are in place. However, disseminating their data to third parties without permission might be sensitive and might jeopardize the PhD supervisor's professional relationship with them.

The rainfall radar products offered by RMI are not public and can thus not be disseminated to third parties without permission. Formal

agreements are in place in the framework of an internship agreement between the PhD researcher and the RMI (supervised by PhD committee member prof. Ricardo Reinoso-Rondinel), meaning the data can be used in the context of this PhD project for research purposes only.

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

The rainfall radar products offered by RMI are not public and can thus not be disseminated to third parties without permission. Formal agreements are in place in the framework of an internship agreement between the PhD researcher and the RMI (supervised by PhD committee member prof. Ricardo Reinoso-Rondinel), meaning the data can be used in the context of this PhD project for research purposes only.

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).

1. Reference information, a hyperlink to the source and the path to the PDF location of the file of all background material is collected in a bibliography manager, e.g. Mendeley. These tools allow focused searching and filtering of that material as well.
2. Historical observation data is stored using the identifiers used by the issuing instance. This allows backtracing metadata/changes/...
3. Parameters of hydrodynamic models are stored automatically by the (commercial) software that will be used (InfoWorks ICM)
4. Synthetically generated input data, as well as simulation results will be stored using a descriptive folder/file structure that is detailed in a dedicated .pdf
5. Parameters of data-driven models created within this PhD project, as well as the models themselves, will be stored in appropriate formats (.csv, .pkl, ...)

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.

- Yes

For the (parts of the) data that will be published in OA repositories, a metadata standard compatible with the repository will be used (e.g. Dublin Core or Datacite).

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

Where will the data be stored?

During the research project, preliminary and intermediate results will be stored on the laptop of the PhD student as well as on external harddrives (backup). In addition, KU Leuven network drives will be used for transferring files between computer systems.

Relevant code and final results will be stored on cloud servers using services provided by KU Leuven (KU Leuven Gitlab, OneDrive, local KU Leuven network drives). These cloud services meet the FWO requirements for backups (saved at multiple locations etc.)

How will the data be backed up?

SyncBackFree will be used for scheduling and automating backups.

**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

Preliminary and intermediate results (e.g. simulation results) require storage exceeding the available storage space of a typical laptop. Hence, external hard drives are used to store these. Final results require significantly less storage, and are thus suitable for storage on laptop and cloud services.

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

Physically accessible systems (laptop, external hard drives) are password protected. Cloud services (both KU Leuven as external) require login details. Login details are managed by the PhD researcher.

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

The costs related to laptop and external hard drives are covered by the bench fee provided by FWO. Services provided by KU Leuven are covered by the university/department/supervisor.

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...).

All reused and generated data will be retained for at least 10 years after the end of the project.

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

- 1) Data used for publications will be published in Open Access with an DOI identifier.
- 2) Other data, including code/data/background material not directly used in publications, will be stored on KU Leuven systems (central servers with automatic back-up procedures).
- 3) Structured code repositories will be stored on KU Leuven's Gitlab using the git paradigm.

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

- 1) To be determined based on selected OA data repositories
- 2) KU Leuven / Departement of Civil Engineering / Supervisor
- 3) Free

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 a restricted access repository (after approval, institutional access only, ...)
- Yes, in an Open Access repository

Data used for scientific publications will be published in Open Access repositories and linked to the publications.
Other data will be restricted and only shared upon permission of the PhD supervisor.

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

Upon supervisor permission, (parts of) the data and models can be shared with interested stakeholders (e.g. other researchers and scientific collaborators, local end-users).

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.

- Yes, Intellectual Property Rights
- Yes, Other

Parts of the hydrodynamic model for the city of Antwerp was conceived by Aquafin, who share their model for research projects at KU Leuven.

The meteorological datasets used in this research are conceived and belong to Federal and Regional institutes (e.g. RMI, VMM, HIC). Not all of which are public (RMI).

Note that no formal agreements were signed with relation to the non-public datasets. However, sharing this data should not occur without approval of the data providers out of respect for them and in order to maintain good professional relationships that might result in additional datasets in the future.

The rainfall radar products offered by RMI are not public and can thus not be disseminated to third parties without permission. Formal agreements are in place in the framework of an internship agreement between the PhD researcher and the RMI (supervised by PhD committee member prof. Ricardo Reinoso-Rondinel).

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

In an Open Access repository.

When will the data be made available?

Upon publication of the research results.

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

Depending on the data that will be shared, this will likely be GNU General Public License v3.0 (or later) for computer code and Attribution-(NonCommercial-)ShareAlike 4.0 International for datasets in OA repositories. To be discussed between PhD researcher and supervisor upon publication of the code/data.

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

For Open Access datasets, yes.

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

As all data in this research project is digital, no direct costs are expected for sharing data.

6. Responsibilities

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

The PhD researcher (Daan Buekenhout) will manage data documentation and metadata during the research project.

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

The PhD researcher (Daan Buekenhout) will manage data storage and backup during the research project.

Who will manage data preservation and sharing?

The PhD supervisor (prof. Patrick Willems) will manage data preservation and sharing after the research project has concluded.

Who will update and implement this DMP?

The PhD researcher (Daan Buekenhout) bears the end responsibility of updating & implementing this DMP.