A geostatistical framework to forecast malaria incidence in the context of climatic variables

A Data Management Plan created using DMPonline.be

Creators: Jose Alejandro Rozo Posada, n.n. n.n.

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Jose Alejandro Rozo Posada

Project Administrator: n.n. n.n.

Grant number / URL: 11M6523N

ID: 198266

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Malaria is a life-threatening disease that causes a significant public health burden worldwide. Malaria incidences show considerable temporal and geographical variation, which can be linked at least partly to climatic conditions. Over the last twenty years, multiple Malaria Early Warning Systems (MEWS) have been developed. These systems primarily focus on short-term incidence predictions, but there is a growing demand to implement long-term predictions that are informed by climatic variables. This project intends to develop a geostatistical long-term forecasting framework for malaria incidence with an emphasis on the effects of climate-related conditions. Statistical methodology will be developed that makes use of spatial multivariate time series analyses, distributed lag models, and joint models for multiple data sources, in order to flexibly model spatio-temporal variation in malaria incidence and how this is affected by climatic events. These methods will be embed in a MEWS, which prioritizes applicability and interpretability for healthcare workers and policy-makers in developing countries.

Last modified: 02-05-2023

A geostatistical framework to forecast malaria incidence in the context of climatic variables
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Question not answered.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Question not answered.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

Question not answered.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Question not answered.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

Question not answered.

A geostatistical framework to forecast malaria incidence in the context of climatic variables
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

Question not answered.

A geostatistical framework to forecast malaria incidence in the context of climatic variables
GDPR

GDPR

Have you registered personal data processing activities for this project?

Question not answered.

A geostatistical framework to forecast malaria incidence in the context of climatic variables
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
Monthly malaria cases	This dataset contains monthly counts of laboratory-confirmed cases, which are reported from all health facilities in each of the 161 individual administrative districts in Mozambique. They include the dates, the district and province codes, and the total number of confirmed cases of malaria. Information is available from 2017 onward.	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
Weekly malaria cases	This dataset contains weekly counts of laboratory-confirmed cases, which are reported from all health facilities in each of the 161 individual administrative districts. They include the dates, the district and province codes, the total number of confirmed cases of malaria in the whole population, the total number of confirmed cases of malaria in children under five years old, and the total district population. The information is available from 2010 onward.	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
Weekly environmental data	This dataset contains Mean/Max/Min precipitation, Mean Temperature, mean dewpoint temp, mean surface pressure, specific and relative humidity. The information available from 2010 onwards	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
Monthly environmental information	This dataset contains information for temperature, potential evaporation, leaf area index, low vegetation, surface pressure, skin reservoir content, surface runoff, and total precipitation. The information available from 2017 onwards	Reuse existing data	Digital	Compiled/aggregated data	.csv	<100MB	
R package	New compiled R package containing the methods developed	New	Digital	software	.R and .rmd	<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:

The Ministry of Health of Mozambique provides weekly and monthly counts of malaria through the National Malaria Control Program. A data transfer agreement was signed between the two parties on November 21, 2021.

The environmental information comes from an aggregation at the weekly, monthly and district levels performed using the information from the NOAA Earth System Research Laboratories and the European Union's Earth observation programme, Copernicus.

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

There is a Data Transfer Agreement between the National Malaria Control Program and me. It was signed on November 21, 2021.

It is stated that the data is the property of the provider, and it will be used only for planning purposes at KU Leuven and UHasselt. Only my team of supervisors and I have access to it. In addition, I must use the data under contract for nonprofit proposes, implementation and capacity building to the public health benefit within the project plans.

Except for normal computations required, I should not try to modify in any way the data for different proposes other than the ones stated in the research proposal. The data cannot be transferred to anyone who is not part of the supervisory team without any prior written consent from the provider institution. Furthermore, I cannot use the data for commercial, research or consulting proposes in which I or the universities I'm affiliated with obtain rights of research results without the prior written consent of the provider institution. After using the requested data, I should share cleaned, computed or transformed data with the provider institution and agree on further management of the data.

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).

Each dataset has a data dictionary associated that contains the variable name, the readable variable name, the measurement units, the allowed values, and the definition of the variable. Additionally, the scripts used to transform and use the data are and will be commented on, explaining the purpose of the script and the expected result of each piece of code. These scripts are available in R.

Documentation of the R package will be provided following the guidelines of CRAN.

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

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

Where will the data be stored?

The data are stored in two cloud-based storage services, OneDrive from KU Leuven and Google Drive from UHasselt. The access is granted to the four supervisors and nobody else has access to it.

How will the data be backed up?

Since the data is stored in two cloud-based storage services, each one works as a backup for the other.

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

All KU Leuven personnel has access to 2 TB of data storage on OneDrive. As the estimated sizes of the datasets <100 GB, sufficient storage and backup capacity are available. Likewise, all UHasselt personnel has access to 2 TB of data storage on google drive.

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

Due to the nature of OneDrive and Google Drive, files that you are not explicitly shared are not accessible to anyone else.

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

OneDrive and Google Drive are free for staff and students of KU Leuven and UHasselt, respectively.

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...).

Upon agreement with the National Malaria Control Program and following the recommendations of KU Leuven, the weekly and monthly malaria case information, as well as the environmental information, will be retained for ten years. In addition, any manipulation of the data, e.g. aggregation, will be retained. The ten years period is for the purpose of reproducibility, verification and potential reuse.

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

The data will be archived in the file storage located in the central KU Leuven datacenters. Collaboration is possible inside and outside KU Leuven.

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

The cost is € 251.83 per year per TB. 1TB will provide enough storage capacity.

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, …)

Upon agreement with the National Malaria Control Program, the weekly and monthly malaria data will be shared with researchers interested in it. The environmental data and any modifications to it will be available for the public.

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

The epidemiological information (weekly and monthly malaria data) is available for the main researcher and the team of supervisors. Any transformation of the data will be shared with the National Malaria Control Program as agreed in the Data Transfer Agreement.

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, Other

A Data Transfer Agreement was signed, and it is upon agreement with them with who the information can be shared.

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

To be discussed with the National Malaria Control Program.

When will the data be made available?

Any transformation of the environmental data will be available upon the publication of the results. The epidemiological information is restricted and to access to it a Data Transfer Agreement should be signed with National Malaria Control Program

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

The R-package and associated functions will be uploaded on CRAN

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.

- No

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

No costs are expected

6. Responsibilities

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

Jose Alejandro Rozo Posada

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

Jose Alejandro Rozo Posada and Thomas Neyens

Who will manage data preservation and sharing?

Jose Alejandro Rozo Posada and Thomas Neyens

Who will update and implement this DMP?

Jose Alejandro Rozo Posada and Thomas Neyens