



Climate-driven vector-borne disease risk assessment

Data Management Plan

CREATED BY

Theophano Pampaka

PRINCIPAL INVESTIGATOR

Dr. Kamil Erguler

DATA MANAGER

Dr. Kamil Erguler
Dr. Yiannis Proestos
Dr. George Tsouloupas

PROJECT ADMINISTRATOR

Christiana Poyiatzea

AFFILIATION

Climate and Atmosphere Research Center (CARE-C)
The Cyprus Institute

FUNDER

Wellcome Trust (Ref. 226065/Z/22/Z)

START - END DATES

1 January 2023 - 31 December 2027

Revision history

Date	Created/modified by	Comments
19/06/2023	Theophano Pampaka	Initial content created
04/09/2023	Kamil Erguler	Style and content updated
05/09/2023	Yiannis Proestos	Comments and suggestions

Contents

Project abstract.....	3
Outputs management plan.....	3
Data outputs of the research:.....	3
Data availability:	4
Data access for researchers:	4
Data sharing:	4
Key datasets and resources:.....	5

Project abstract

The risk of vector-borne disease emergence will increase in the northern hemisphere in response to a warming climate, which drives the suitable habitats and distribution ranges of many species. Developing predictive, reliable, and widely applicable models of disease risk requires accurate representation of the physiological dependencies of vectors and pathogens. VECLim aims at vector-borne disease prediction and management by employing data-driven climate-sensitive large-scale mechanistic modelling to represent vector populations and disease transmission. An operational model service will provide access to a vector-pathogen model repository and enable customisation and evaluation of control strategies and future climate scenarios. Short, medium, and long-term predictions of local, regional, and global vector activity and outbreak risk will be delivered through a user-friendly interactive web-based geographic information platform. Model outputs and analyses will be optimised for clarity and utility, targeting wide adoption by scientists, decision makers, and the public. With the vector-pathogen model repository, comprehensive risk assessment web-platform, and an active user community, VECLim will aid the adoption of predictive tools in integrated management plans for vectors and vector-borne diseases.

Outputs management plan

VECLim adheres to the principles of free and open exchange of data, knowledge, and open-access publishing, stressing the increasing demand for open access, especially concerning scientific results obtained from publicly funded research. Intellectual property (IP) generated, as per the Intellectual Property Policy of The Cyprus Institute (Cyl), shall be the property of the Cyl. Following internal review, which will take place at least once a year, IP generated shall be shared with the community via online open access repositories.

Data outputs of the research:

VECLim generates, hosts, and serves the following data, which will be valuable assets to the wider research community:

1. Climate and environment-driven mechanistic models of vector populations and vector-borne disease transmission.
2. Short-, medium-, and long-term predictions of vector activity and outbreak risk/impact at local, regional, and global scale.
3. Computational tools for modelling and data analysis pertaining climate-sensitive vector-borne disease research.

All is shared freely for the benefit of the wider scientific community, decision makers, and the public according to the FAIR principles of open access (Findability - Accessibility - Interoperability - Reusability of data). VECLim adopts the open-access principles to share the data collected/generated, in appropriate open (platform independent) data formats, such as TXT, CSV, JSON, and NetCDF, with a permissive license while also ensuring the acknowledgement of the data producers.

In addition, VECLim hosts a local copy of the following publicly available data, and serves to the wider research community by appropriately acknowledging data sources and data collectors:

1. Meteorological variables, climate projections, and environmental covariates.
2. Vector and disease surveillance data adhering to the appropriate data standards, such as the minimum information standard for reporting arthropod abundance data (MiREAD) standards.

All is shared strictly according to the permissions granted by the licenses of the original data sources. No personal data are collected or retained pertaining the surveillance data. Data collection points are appropriately concealed by aggregating into the grid resolution of common environmental datasets (such as the ERA5 dataset available through the Climate Data Store).

Data availability:

Data generated (predictions and risk analyses) are routinely streamlined and deposited to online open access repositories, such as Figshare and Zenodo, with a unique digital object identifier (DOI), and the model code and updates are deposited to the Python Package Index in a timely manner following annual review by the Scientific Advisory Board. The development version of the code is shared with the research community via open access code sharing tools, such as GitHub, and each release is archived in open access repositories, such as Zenodo.

A subset of the environmental covariates dataset, depending on relevance to the analyses, are redistributed via the platform, provided that a permissive license, such as the Creative Commons Attribution 4.0 License - CC BY 4.0, is associated with the data. Vector abundance data and vector-borne disease incidence rates, gathered through collaborative projects or provided by international organisations, such as VectorBase (the data repository of invertebrate vectors of human pathogens) and (E)CDC, are displayed on the platform in an anonymised and gridded form only if, and as long as, the appropriate permissions are held. Data gathered through collaborative projects are displayed following data curation and deposition to online open access repositories (with a DOI). In all cases, special care is taken to properly attribute the original data source and providers.

Open-access open peer review scientific journals, such as Wellcome Open Research and F1000Research, are considered as initial targets to accelerate sharing of scientific publications which may arise as part of VECLim and for short articles describing the data.

Data access for researchers:

All data and code that can be shared with a permissive license on online data repositories are readily accessible as such following the annual review by the Scientific Advisory Board. Recent and historical predictions are available as they are generated through the platform with a permissive license. VECLim avoids using vector abundance and disease incidence data which have restrictions on access and sharing rights. In cases where such data are essential to significantly improve model predictions, data providers and their contact details are cited in relevant publications for data requests.

Data sharing:

There are no foreseeable limits to data sharing: no personal data are collected or used by VECLim in model training or testing, and data from external resources are requested only as anonymised. The vector abundance and disease incidence data obtained from external collaborators belong to data producers. The ownership is protected, and the distribution rights are given to the Cyl with data-sharing agreements, which clearly state in writing the terms of reference and the responsibilities of both parties.

Key datasets and resources:

Online trustworthy data sharing platforms, such as Zenodo and Figshare, are used to deposit key datasets, including the annual prediction and risk analyses summary, for cost-free long-term access. Additionally, the VECLim server hardware, hosted by the Cyl, is used to store all the data. The server is integrated into the High-Performance Computing (HPC) environment of the Cyl HPC Facility, for regular back-ups and maintenance for a minimum of 10 years.