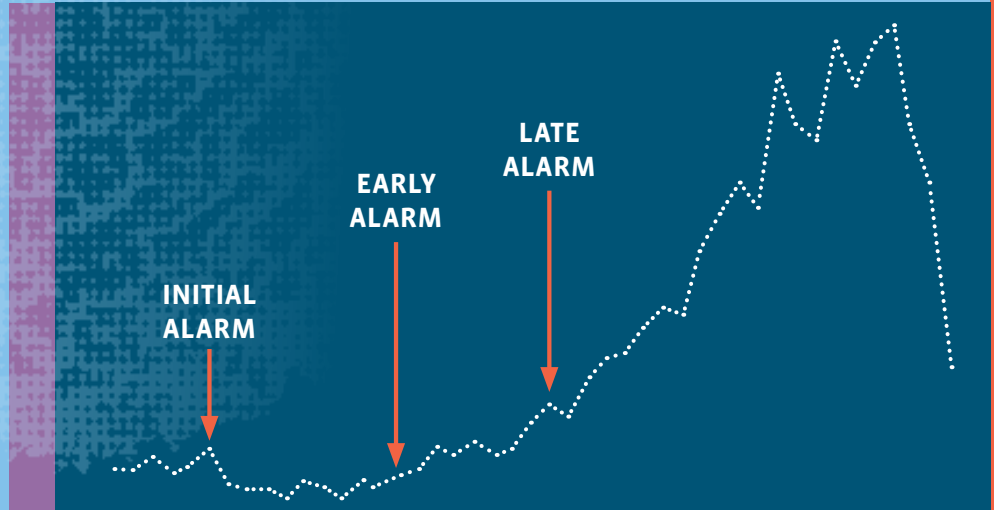


OPERATIONAL GUIDE USING THE WEB-BASED DASHBOARD



The Early Warning and Response System (EWARS) for Dengue Outbreaks



World Health
Organization



For research on
diseases of poverty

UNICEF • UNDP • World Bank • WHO

WHO Library Cataloguing-in-Publication Data

Early Warning and Response System (EWARS) for Dengue Outbreaks: operational guide using the web-based dashboard
ISBN 978-92-4-151435-4

Copyright ©World Health Organization 2018 on behalf of the Special Programme for Research and Training in Tropical Diseases

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization.

Suggested citation. Early Warning and Response System (EWARS) for dengue outbreaks: operational guide using the web-based dashboard. . Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at <http://apps.who.int/iris>.

Sales, rights and licensing. To purchase WHO publications, see <http://apps.who.int/bookorders>. To submit requests for commercial use and queries on rights and licensing, see <http://www.who.int/about/licensing>.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

Graphic Design: Lisa Schwarb, Lausanne
Printed by the WHO Document Production Services (DUP), Geneva, Switzerland

OPERATIONAL GUIDE

USING THE WEB-BASED DASHBOARD

The Early Warning and
Response System (EWARS)
for Dengue Outbreaks



**World Health
Organization**



**For research on
diseases of poverty**

UNICEF • UNDP • World Bank • WHO

Contents

- Foreword** vii
- Preface** viii
- Acknowledgements** viii
- Glossary** ix

- Chapter 1. Preparing your dataset** 1
 - 1.1 Surveillance data, contents and format 1
 - 1.1.1 General variables 1
 - 1.1.2 Outbreak indicator 2
 - 1.1.3 Alarm indicator(s) 2

- Chapter 2. Dashboard I: Data Calibration (for user at central level)** 3
 - 2.1 Accessing the EWARS dashboard 3
 - 2.2 Setting the country, password and the surveillance dataset 4
 - 2.2.1 Country code and password settings (to allow access for local users) 4
 - 2.2.2 Uploading your surveillance dataset 4
 - 2.3 Calibrating your instrument 5
 - 2.3.1 Dividing your dataset into run-in and evaluation data 5
 - 2.3.2 Defining the districts to be analysed 6
 - 2.3.3 Defining the district population 6
 - 2.3.4 Defining your outbreak indicator 7
 - 2.3.5 Defining your outbreak period 7
 - 2.3.6 Defining your alarm indicator(s) 8
 - 2.3.7 Defining your window size for the alarm indicators 8
 - 2.3.8 Defining your alarm threshold 9
 - 2.3.9 Defining the graphical outputs 9
 - 2.3.10 Defining your seasonal variation 10
 - 2.3.11 Defining your endemic channel 10
 - 2.3.12 Defining your window size for the outbreak indicator 12
 - 2.3.13 Defining your prediction distance 12
 - 2.3.14 Defining your outbreak signal 13
 - 2.3.15 Activating the process for monotonic relationship (spline) 13
 - 2.3.16 Running the program 14
 - 2.3.17 Generating the surveillance workbook 14
 - 2.4 Understanding your calibration outputs 15
 - 2.4.1 Graph: run-in period 15

| | | |
|--|---|----|
| 2.4.2 | Graph: evaluation period | 16 |
| 2.4.3 | Graph: run-in and evaluation period..... | 17 |
| 2.4.4 | Summary of calibration results, and sensitivity and PPV..... | 18 |
| Chapter 3. Dashboard II: Harvesting the results (for user at local level) | | 21 |
| 3.1 | Prospective early warning system | 21 |
| 3.1.1 | Access to and summary of parameters: where to find them in Dashboard II and their usefulness..... | 21 |
| 3.1.2 | The three sub-links in Dashboard II: “Retrospective”, “Dataset” and “Help”; their definitions and uses..... | 23 |
| 3.1.3 | The surveillance workbook tab: all you need to know for this section..... | 23 |
| 3.1.4 | The input data tab: all you need to know for this section..... | 24 |
| 3.1.5 | The outbreak tab: all you need to know for this section..... | 25 |
| 3.1.6 | The probability tab: all you need to know for this section..... | 26 |
| 3.1.7 | The alarm plus outbreak tab: all you need to know for this section..... | 27 |
| 3.1.8 | The response tab: all you need to know for this section..... | 28 |
| 3.2 | Retrieving online meteorological information on weekly alarms (datasets sub-link) | 29 |
| 3.2.1 | Setting your local district/municipality spatial information to detect nearest meteorological station..... | 29 |
| 3.2.2 | Setting your period of interest for deriving the meteorological information..... | 30 |
| 3.2.3 | Selecting a desired meteorological station from the obtained list..... | 31 |
| 3.2.4 | Using the derived information of the alarm indicators..... | 32 |
| ANNEX 1. Technical guide | | 33 |
| 1. | Introduction | 33 |
| 2. | Methodology | 33 |
| 2.1 | Retrospective phase | 33 |
| 2.1.1 | Run-in data..... | 33 |
| 2.1.2 | Evaluation data..... | 33 |
| 2.2 | Prospective early warning phase (i.e. Dashboard II) | 34 |
| 2.2.1 | Rationale for the approach..... | 34 |
| 3. | Structural design | 34 |
| 3.1 | Phase I (retrospective: Dashboard I) | 34 |
| 3.1.1 | Using the run-in data to create the prediction model (part one)..... | 34 |
| 3.2 | Phase II (prospective surveillance: Dashboard II) | 40 |
| 3.1.2 | Using the evaluation data to assess how the derived parameters from the run-in data would predict an outbreak (part two)..... | 39 |

LIST OF FIGURES

| | | |
|------------|--|----|
| Figure 1. | Summary list of variables | 1 |
| Figure 2. | Language setting and assigning password for local users to access their corresponding algorithms..... | 4 |
| Figure 3. | Uploading your dataset and entering the corresponding sheet name and number..... | 4 |
| Figure 4. | Dividing your dataset..... | 5 |
| Figure 5. | Choosing your districts | 6 |
| Figure 6. | Defining the annual number of people in your districts..... | 6 |
| Figure 7. | Labelling the number of outbreak cases..... | 7 |
| Figure 8. | Defining the outbreak period | 7 |
| Figure 9. | Defining the alarm indicator | 8 |
| Figure 10. | Defining window size for the alarm indicators | 8 |
| Figure 11. | Defining the alarm threshold | 9 |
| Figure 12. | Selecting your graphical output | 9 |
| Figure 13. | Choice of the seasonal variation..... | 10 |
| Figure 14. | Changing the endemic channel (the outbreak threshold)..... | 10 |
| Figure 15. | Modelling the endemic channels | 11 |
| Figure 16. | Defining the number of outbreak weeks (window size)..... | 12 |
| Figure 17. | Defining the prediction distance (in weeks) | 12 |
| Figure 18. | Defining the cut-off point of the outbreak signal | 13 |
| Figure 19. | Choice of the spline application..... | 13 |
| Figure 20. | Running the surveillance workbook..... | 14 |
| Figure 21. | Generating the surveillance workbook..... | 14 |
| Figure 22. | Graph of the run-in data showing the endemic channel, number of cases and outbreak period..... | 15 |
| Figure 23. | Graph of the evaluation data showing the general parameters including the probability of outbreak period..... | 16 |
| Figure 24. | Graph of the combined run-in and evaluation data showing all information..... | 17 |
| Figure 25. | Absolute values in the output table, example showing total number of outbreaks, missed outbreaks, alarms, correct alarms and false alarms..... | 18 |
| Figure 26. | Sensitivity and PPV (screenshot of the output table – success of predicting outbreaks is shown in sensitivity/PPV)..... | 19 |

| | |
|---|----|
| Figure 27. Summary of parameters from the retrospective analysis (Dashboard I) | 22 |
| Figure 28. Surveillance workbook tab: all you need to know for this section..... | 23 |
| Figure 29. The input data tab: all you need to know for this section | 24 |
| Figure 30. The outbreak tab: all you need to know for this section..... | 25 |
| Figure 31. The probability tab: all you need to know for this section..... | 26 |
| Figure 32. The alarm plus outbreak tab: all you need to know for this section | 27 |
| Figure 33. The response tab: all you need to know for this section..... | 28 |
| Figure 34. Retrieving online meteorological information: setting your district's spatial information..... | 29 |
| Figure 35. Retrieving online meteorological information: defining the period of interest..... | 30 |
| Figure 36. Retrieving online meteorological information: selecting a desired meteorological station..... | 31 |
| Figure 37. Reading the retrieved online meteorological information for alarm indicators..... | 32 |
| Figure 38. Block diagram illustrating the process and steps of the retrospective and prospective phases (I, II) | 35 |
| Figure 39. Illustrated process of measuring the alarm indicator(s) for a defined alarm window size.... | 36 |
| Figure 40. Illustration of the moving average process of measuring the outbreak indicator for a defined outbreak window size..... | 37 |

FOREWORD

Currently, dengue fever is the fastest-spreading mosquito-borne viral disease worldwide with epidemics overburdening stretched health systems and threatening the stability of societies. Although the transmission of dengue can be controlled with a combination of effective vector control interventions and an efficient vaccine, late detection and inadequate response mechanisms are compounding the effects of rapid transmission. Furthermore, despite the existence of outbreak alert indicators, the means of deploying them in early warning systems is often lacking.

With this in view, a programme led by TDR, the Special Programme for Research and Training in Tropical Diseases, conducted multi-country research into alarm signals for outbreaks and their use within early warning systems. In line with the prevailing literature,^{1,2,3} alarm variables, such as hospitalized confirmed or probable dengue cases, as well as a set of epidemiological, entomological and environmental variables evidenced predictive abilities.⁴ However, it was clear that countries are in need of a standardized and compatible approach to deploy these alarm signals in a predictive and operational way. It was on this basis that an accessible, adaptable and user-friendly early warning system was developed.⁵

This guide is an update to the previous version in 2017. This revised edition of *The Early Warning and Response System (EWARS) for Dengue Outbreaks: operational guide using the web-based dashboard* aims to provide programme managers with a user-friendly tool that can: (i) analyse and draw conclusions from historic dengue datasets; (ii) identify appropriate alarm indicators that can sensitively and specifically predict forthcoming outbreaks at smaller spatial scales; and (iii) use these results and analyses to build an early warning system to detect dengue outbreaks in real-time and respond accordingly. Together, these three components will build technical capacity and provide a standardized methodology for predicting dengue outbreaks in countries with great need. Furthermore, this web-based EWARS tool can ensure enhanced, fast and secured communication, between national and subnational levels, and standardized utilization of surveillance data.

This guide was produced by TDR together with WHO's Neglected Tropical Diseases (WHO/ NTD) and WHO regional offices in the context of a European Union-financed research programme, the International Research Consortium on Dengue Risk Assessment, Management and Surveillance (IDAMS), to develop an evidence-based, early warning system for outbreak detection and management of dengue fever outbreaks.

¹ Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis.* 6:e1908.

² Halide H, Ridd P (2008). A predictive model for dengue hemorrhagic fever epidemics. *Int J Environ Health Res.* 18:253–265.

³ Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, Ng N (2012). Optimal lead time for dengue forecast. *PLoS Negl Trop Dis.* 6:e1848.

⁴ Bowman LR, Tejada GS, Coelho GE, Sulaiman LH, Gil BS, McCall PJ et al. (2016). Alarm variables for dengue outbreaks: a multi-centre study in Asia and Latin America. *PLoS One.* 11:e0157971.

⁵ Hussain-Alkhateeb L, Kroeger A, Olliaro P, Rocklöv J, Sewe MO, Tejada G et al. (2018). Early warning and response system (EWARS) for dengue outbreaks: recent advancements towards widespread applications in critical settings. *PLoS One* 13:e0196811.

PREFACE

Welcome to *The Early Warning and Response System (EWARS) for Dengue Outbreaks: operational guide using the web-based dashboard*. This *Guide* will provide you with the information and tools necessary to use and analyse surveillance data to predict dengue outbreaks. Below are step-by-step instructions to help you organize your raw data, enter it into the web-based dashboard, run the analysis and interpret your results. At the end of this *Guide* there is an annex providing technical information on the processes and statistics you will be using. However, you do not need to use it to build your early warning system. The computerized statistical program that the EWARS will be using is called 'R', version 3.4.3. Before running analyses, it is important that the data you have collected are in the correct format; otherwise, the analytical software will not recognize your data and will not work properly.

ACKNOWLEDGEMENTS

The Early Warning and Response System (EWARS) for Dengue Outbreaks: operational guide using the web-based dashboard was written by Laith Hussain-Alkhateeb. It was coordinated and supported by Axel Kroeger and Piero Olliaro of TDR, the Special Programme for Research and Training in Tropical Diseases. Special thanks goes to Joacim Rocklöv and his team; Maquins Odhiambo Sewe for the development of the analytical programming during the transfer process from the STATA-to-R statistical package; and, to Aditya L. Ramadona for providing insight and key developmental contributions to the web-based dashboard. We also acknowledge David Benitez (Mexico) and Balvinder Gill (Malaysia) for beta testing the materials, as well as the following colleagues for contributing valuable country data to enable the development of EWARS: Roberta Carvalho (Brazil); Giovanini Coelho (Brazil); Lokman Hakim (Malaysia); Luong Quang (Viet Nam); Ronald Skewes Ramm (Dominican Republic); and Gustavo Sanchez Tejeda (Mexico). This *Guide* was developed in the context of a dengue research programme supported by a grant from the European Commission (Grant Number m281803) to the International Research Consortium on Dengue Risk Assessment, Management and Surveillance (IDAMS) within the 7th Framework Programme of the European Commission (FP7), and by TDR.

GLOSSARY

| Term | Definition |
|--|---|
| R-package | Specialized software used to process and analyse your data and to generate meaningful results. |
| Endemic channel | This represents the number of cases within the expected normal seasonal range of specific area; anything above this moving threshold would be considered representative of an unprecedented number of cases, i.e. an outbreak. |
| Outbreak indicator | The dependent variable(s) used to define outbreaks. Usually probable or hospitalized dengue cases. |
| Alarm indicator | The independent variable(s) used to predict outbreaks. This could be one of a number of meteorological variables, e.g. rainfall, temperature, or other entomological/epidemiological variables. |
| Spline | A function to capture both positive and non-linear associations between the same alarm indicator(s) and outbreak indicator(s). |
| Sensitivity | The proportion of outbreaks correctly predicted by alarms. A higher number indicates higher true positive alarms. |
| Positive predictive value (PPV) | The proportion of alarms that successfully predicted outbreaks. A higher number indicates lower false positive alarms. |
| Standard deviation (SD) | This is the standard deviation, which is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points scatter close to the mean (average). |

Chapter 1

Preparing your dataset

1.1 Surveillance data, contents and format

Before running the analysis, your dataset should contain the list of variables described below. Variables “year”, “week” and “district” need to be written in lower-case letters, as presented in figure 1. At least one alarm indicator is required (e.g. mean temperature) but additional alarm indicators can also be included. A minimum of three years’ data records is needed to run this program.

Figure 1. Summary list of variables

| A | B | C | D | E | F | G | H | I | J | K | L | M |
|------|----------|------|--------------------------|---------------------------|----------------------|---------------------|-------------------------|---|---|---|---|---|
| year | district | week | population_annual_number | weekly_hospitalized_cases | weekly_humidity_mean | weekly_rainfall_sum | weekly_temperature_mean | | | | | |
| 2015 | 1 | 39 | 200000 | 3 | 62.0 | 0.0 | 0.0 | | | | | |
| 2015 | 1 | 40 | 200000 | 0 | 65.1 | 11.7 | 23.9 | | | | | |
| 2015 | 1 | 41 | 200000 | 1 | 72.3 | 0.0 | 26.9 | | | | | |
| 2015 | 1 | 42 | 200000 | 1 | 66.0 | 2.8 | 23.8 | | | | | |
| 2015 | 1 | 43 | 200000 | 2 | 71.1 | 1.3 | 26.8 | | | | | |
| 2015 | 1 | 44 | 200000 | 3 | 71.0 | 0.0 | 27.0 | | | | | |
| 2015 | 1 | 45 | 200000 | 1 | 72.7 | 23.5 | 28.1 | | | | | |
| 2015 | 1 | 46 | 200000 | 4 | 70.0 | 30.4 | 29.4 | | | | | |
| 2015 | 1 | 47 | 200000 | 5 | 68.7 | 25.0 | 29.9 | | | | | |
| 2015 | 1 | 48 | 200000 | 3 | 66.7 | 15.0 | 30.1 | | | | | |
| 2015 | 1 | 49 | 200000 | 2 | 70.0 | 50.5 | 30.9 | | | | | |
| 2015 | 1 | 50 | 200000 | 1 | 71.4 | 90.2 | 31.0 | | | | | |
| 2015 | 1 | 51 | 200000 | 0 | 74.1 | 139.4 | 31.0 | | | | | |
| 2015 | 1 | 52 | 200000 | 0 | 84.0 | 369.7 | 30.9 | | | | | |
| 2015 | 2 | 1 | 3450000 | 2 | 62.0 | 0.0 | 0.0 | | | | | |
| 2015 | 2 | 2 | 3450000 | 2 | 65.1 | 11.7 | 23.9 | | | | | |
| 2015 | 2 | 3 | 3450000 | 4 | 72.3 | 0.0 | 26.9 | | | | | |
| 2015 | 2 | 4 | 3450000 | 3 | 66.0 | 2.8 | 23.8 | | | | | |
| 2015 | 2 | 5 | 3450000 | 5 | 71.1 | 1.3 | 24.8 | | | | | |
| 2015 | 2 | 6 | 3450000 | 4 | 71.0 | 0.0 | 27.0 | | | | | |
| 2015 | 2 | 7 | 3450000 | 3 | 72.7 | 0.0 | 28.1 | | | | | |
| 2015 | 2 | 8 | 3450000 | 2 | 70.0 | 0.5 | 29.4 | | | | | |
| 2015 | 2 | 9 | 3450000 | 1 | 68.7 | 1.3 | 29.9 | | | | | |
| 2015 | 2 | 10 | 3450000 | 1 | 66.7 | 0.0 | 30.1 | | | | | |
| 2015 | 2 | 11 | 3450000 | 3 | 70.0 | 1.8 | 30.9 | | | | | |
| 2015 | 2 | 12 | 3450000 | 2 | 71.4 | 0.3 | 31.0 | | | | | |
| 2015 | 2 | 13 | 3450000 | 0 | 74.1 | 139.4 | 31.0 | | | | | |
| 2015 | 2 | 14 | 3450000 | 1 | 84.0 | 369.7 | 30.9 | | | | | |

The user can add more "alarm" indicators here!

1.1.1 General variables

- “year”: the year when the data were collected. The year must be entered in full using “four” numbers (e.g. 2015, 2016, etc.);
- “week”: the number of the epidemiological week (Sunday to Saturday) when data were collected/obtained. The week number must be entered in full (e.g. 1, 2, 3, etc.);
- “district”: a number (code) that represents the district, locality or municipality where data were captured (e.g. 1, 2, 15, 22, etc.);
- “population_annual_number”: the annual population size of a district is reported in absolute numbers in the surveillance data. See column G in figure 1, above.

CHAPTER 1

Preparing your dataset

1.1.2 Outbreak indicator

- “weekly_hospitalized_cases”: this is the number of hospitalized cases in a given district per epidemiological week, based on the date of hospital admission. Unless rigorous analysis using statistical imputation and validation methodologies have been performed on other outbreak indicators beforehand, we recommend using hospitalized cases as the most appropriate records. You could substitute probable case data where hospitalized data are missing, but this must be distinguished and consistent across all datasets. You cannot mix hospitalized and probable case data.

1.1.3 Alarm indicator(s)

You need a minimum of one alarm indicator, but there is no maximum limit. Please ensure that you enter the exact variable name of the alarm indicator in the Dashboard I interface but avoid alarm indicators with substantial missing records. Some examples of alarm indicators are listed below:

- “weekly_humidity_mean”: the weekly mean humidity (as a percentage), for a corresponding district and year;
- “weekly_rainfall_sum”: the total weekly rainfall (in mm) for a corresponding district and year;
- “weekly_temperature_mean”: the weekly mean temperature in either Celsius or Fahrenheit (do not use Celsius and Fahrenheit data in the same spreadsheet: choose one or the other), for a corresponding district and year;
- “ovitrap index”, measured as the proportion of positive ovitraps per block or other indicators (i.e. infested with *Aedes* eggs).

Please note, it is important that the corresponding variable names are entered correctly where specified in the Dashboard I interface, i.e. they should have the exact spelling and format as in the original surveillance dataset (column).

Chapter 2

Dashboard I: Data calibration (for user at central level)

In this web-based interface (Dashboard I), you are going to make necessary settings and calibrations and, eventually generate the algorithm and parameter coefficients that users at district levels will be using for their prospective analyses (in Dashboard II). This process can take place once a year. In addition, and prior to doing any calibration, you will need to assign local users at districts level an 8-digit password to access their corresponding district account and run their prospective analysis.

The instructions below relate to these steps before linking to the prospective analysis (Dashboard II).

2.1 Accessing the EWARS dashboard

Countries can contact ewars@post.com to request their own EWARS accounts for free access and applications. The use of a country's own dashboard account increases the usage capacity and ensures privacy and data protection.

Once requested, each country will be provided with usernames and passwords to access their own dashboard accounts as well as a Google drive account to link both dashboards. Users at central (national) level will need to use Dashboard I for the setting and calibration process and, users at local (district/provincial) level will need Dashboard II for declaring alarm signals (if any) and responding to alarms accordingly (prospective phase). Users at the central level will need to log into their corresponding Google drive accounts and install into their local PC, where they can view all files generated from dashboard I. You can access your account by simply clicking on the link provided in the EWARS COUNTRY PACKAGE then use the given user name and password to run.

Along with the EWARS package, the R-program script and additional instructions are also provided (under the “HELP” tab in Dashboard I) to assist countries with interest in building their own national hub or integrating this EWARS tool into their existing national surveillance program. This is recommended as it can facilitate a more feasible and sustainable application.

For the purpose of demonstration and training, demo Dashboard accounts and a demo dataset are provided in the box below:

Dashboard I: <http://umuccgh.cloudapp.net:8787/p/5544/>
Username: **demo**
Password: **uMu2018**

Dashboard II: <https://alamadona.shinyapps.io/EWARSv2/>
Google drive account:
Username: **ewars.dashboard1@gmail.com**
Password: **ewars2018**

Once you access Dashboard I, you can download the “Demo Dataset” from the “HELP” tab.

CHAPTER 2**DASHBOARD I: Data calibration (for user at central level)**

2.2 Setting the country, password and the surveillance dataset

2.2.1 Country code and password settings (to allow access for local users)

Before running Dashboard I, you can choose your country code from the list (for the demo account only!).

Users at central level need to assign secured 8-digit passwords for local users at district-level (Dashboard II) to use when accessing their corresponding district information (algorithm and coefficients) they require on a weekly basis for the prospective analysis.

Figure 2 shows where you can find and change these settings. More settings follow after the figure and are illustrated in the section below.

Figure 2. Language setting and assigning password for local users to access their corresponding algorithms



The screenshot shows two input fields with red arrows pointing to them from yellow callout boxes. The first field is a dropdown menu labeled "country code" with "XX" selected. The second field is a text input labeled "8-digit password" with "password" entered.

country code
XX

Select your country code from the drop-down list.

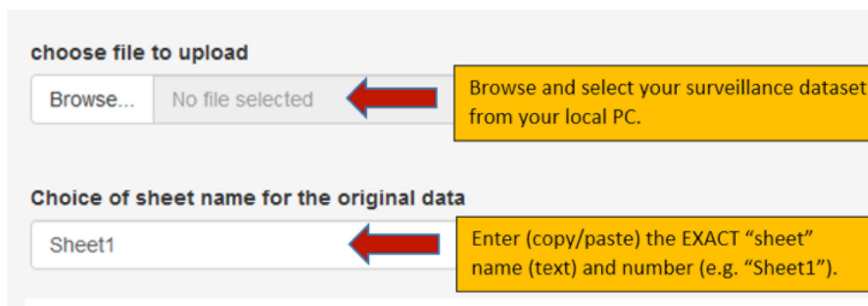
8-digit password
password

Enter 8-digit password, which will be used by local users to access their algorithm.

2.2.2 Uploading your surveillance dataset

Browse and select your dataset in order to upload it and run the calibration process. In Excel, a spreadsheet is divided into “sheets”. These sheets can be given a name, but unless you do, the usual name is “Sheet1”. Enter (copy/paste) the EXACT sheet name (text) and number (e.g. “Sheet1”) into the corresponding box in your dashboard interface so the analytical tool can allocate your dataset.

Figure 3. Uploading your dataset and entering the corresponding sheet name and number



The screenshot shows two input fields with red arrows pointing to them from yellow callout boxes. The first field is a file upload button labeled "choose file to upload" with "Browse..." and "No file selected" text. The second field is a text input labeled "Choice of sheet name for the original data" with "Sheet1" entered.

choose file to upload
Browse... No file selected

Browse and select your surveillance dataset from your local PC.

Choice of sheet name for the original data
Sheet1

Enter (copy/paste) the EXACT “sheet” name (text) and number (e.g. “Sheet1”).

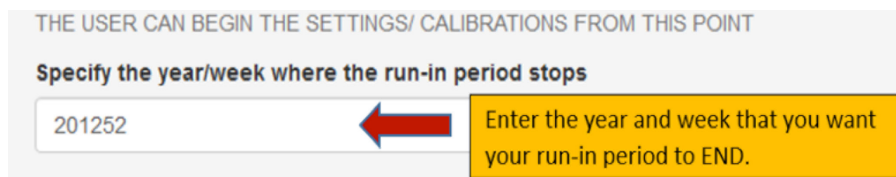
2.3 Calibrating your instrument

2.3.1 Dividing your dataset into run-in and evaluation data

Before running Dashboard I, you must change the model settings in accordance with your data and the analysis you want to run (changing values, alarm/outbreak thresholds, etc.). This will define and improve the prediction of outbreaks.

The figures in this section show you where to find these settings so that you can change them.

Figure 4. Dividing your dataset



- To detect alarm indicators that can help you predict outbreaks, you need to first run a retrospective analysis of your data. For this step, the model requires that you divide your dataset into two time periods. Now you must choose the dates of your historic/run-in period and the evaluation period.
- See sections 2.1.1 and 2.1.2 in Annex 1 for more details on the “run-in period” and “evaluation period”!
- To do so, enter the year and week that you want your “run-in period” to END. The analytical tool will automatically use all data after this date as the “evaluation period”. For example, enter 201326 with no space between the year and week digits for year “2013” and week “26”.
- *This corresponds to **steps 1 and 10** in the annex.*

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

2.3.2 Defining the districts to be analysed

Figure 5. Choosing your districts

Enter the corresponding code(s) for the District/ Municipality you wish to analyse. If you choose more than one, use comma between each District/ Municipality

15, 10, 8

Enter the desired district code(s) of interest. Use comma to separate multiple districts.

- Here you can choose to analyse specific districts or all districts.
- Enter the desired district code(s) of interest.
- The user can enter more than one district code by using ‘comma’ between each district code.
- *This precedes **step 1** in the annex.*

2.3.3 Defining the district population

Figure 6. Defining the annual number of people in your districts

Enter the variable name which represents the annual total Population of the corresponding district/ municipality

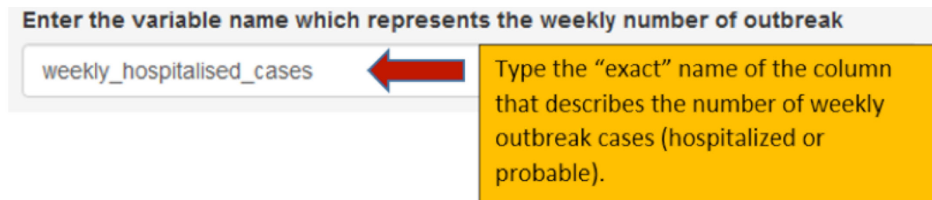
population

Type the “exact” name of the column that indicates for the size of each district population.

- In each district you will have a different human population.
- Here, you must type the name of the variable that tells the analytical program the size of each district population, e.g. if your spreadsheet column is labelled population, please type “population” into the box.
- It is important to type the EXACT text of the variable names (i.e. the title of the column) as it appears in the surveillance data.
- *This corresponds to **step 3a** in the annex.*

2.3.4 Defining your outbreak indicator

Figure 7. Labelling the number of outbreak cases



Enter the variable name which represents the weekly number of outbreak

weekly_hospitalised_cases

Type the “exact” name of the column that describes the number of weekly outbreak cases (hospitalized or probable).

- Here you must consider what incident case data have been captured. For example, it might be the number of “weekly_hospitalised_cases” (recommended) or weekly probable clinical cases or other possible case indicators.
- Please type in the indicated box, the column name that describes these data – it must be the same as in your surveillance dataset.
- *This corresponds to **step 3a** in the annex.*

2.3.5 Defining your outbreak period

Figure 8. Defining the outbreak period



Choice of the number of outbreak weeks to declare outbreak period/stop outbreak

3

Enter the desired choice of the minimum number of outbreak weeks e.g. 3.

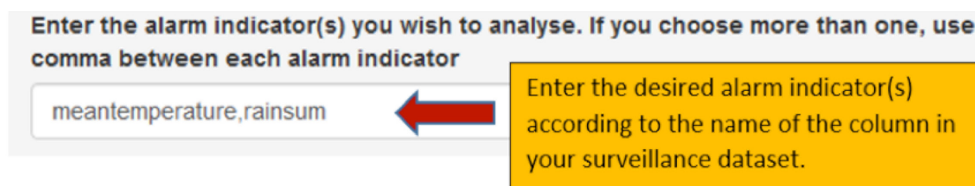
- A collection of consecutive “outbreak weeks” defines an outbreak. It is up to you how to define an outbreak, but we recommend 2 or 3 weeks.
- Enter the desired choice of the minimum number of outbreak weeks needed to define the outbreak period. For example, if you enter 3 in this box then a minimum of “3” consecutive outbreak weeks is required to define an outbreak (outbreak period), and a minimum of “3” consecutive NON-outbreak weeks is required to declare a no-outbreak period.
- *This corresponds to **step 3b** in the annex.*

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

2.3.6 Defining your alarm indicator(s)

Figure 9. Defining the alarm indicator



Enter the alarm indicator(s) you wish to analyse. If you choose more than one, use comma between each alarm indicator

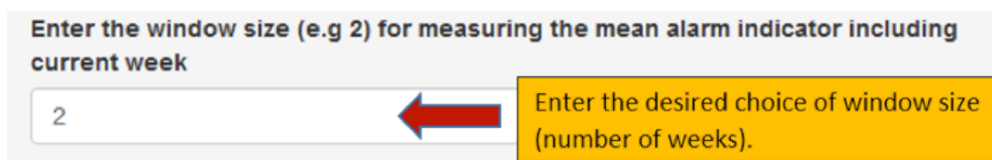
meantemperature,rainsum

Enter the desired alarm indicator(s) according to the name of the column in your surveillance dataset.

- Alarm indicators are defined as an alarm that can predict a forthcoming outbreak.
- Here you can choose which “alarm indicator”(s) you want to test for predictive capacity.
- Enter the desired alarm indicator(s) according to the name of the column in your surveillance dataset; you may include an unlimited number of alarm indicators in this command.
- Missing data will negatively affect the results so be sure that you have a complete dataset before running any or multiple alarm indicators.
- Do not alter the alarm indicator(s) text.
- It is important that the name of the alarm indicator you enter in this option is EXACTLY the same as the variable name in the surveillance data (i.e. the same text you find in the column title in your surveillance data).
- *This corresponds to **step 4** in the annex.*

2.3.7 Defining your window size for the alarm indicators

Figure 10. Defining window size for the alarm indicators



Enter the window size (e.g 2) for measuring the mean alarm indicator including current week

2

Enter the desired choice of window size (number of weeks).

- The values of each “alarm indicator” will be recalculated by the program to produce an average over a given time period. Here you can alter this time period.
- Enter the desired choice of window size (number of weeks) from which the mean value of the alarm indicator can be calculated.
- For example, entering a value of “3” for this option suggests that you are calculating the mean value (average) of the alarm indicator (e.g. temperature) of the current week and two previous weeks, consecutively. A minimum value of “1” means you are measuring the alarm indicator unit of the current week!
- *This corresponds to **step 4** in the annex.*

2.3.8 Defining your alarm threshold

Figure 11. Defining the alarm threshold

Enter the choice of alarm threshold, which is the value from which the alarm signal can be declare (e.g 0.12)

0,12

Enter the desired choice of alarm threshold. A range of 0.05-0.2 is recommended.

- Here you decide what “alarm threshold” is used to signal an alarm.
- At this stage, a probability of an outbreak is calculated based on the user’s input. The alarm threshold is, therefore, a value you need to enter to define a possible alarm signal (an alarm signal is declared when the calculated outbreak probability > the alarm threshold you entered).
- If the threshold you entered is too high, you may not have any alarms and, therefore, no alarm outbreaks will be predicted. If the threshold is too low, you may have many alarms that detect all outbreaks but you will also have many false alarms. You need to experiment with the appropriate value by changing the threshold to get a good balance between a low number of false alarms (positive predictive value or PPV) and high outbreak prediction rate (sensitivity).
- The optimal threshold value has often been shown to be in the range of 0.05–0.2.
- *This corresponds to **step 13** in the annex.*

2.3.9 Defining the graphical outputs

Figure 12. Selecting your graphical output

Specify graph per district/municipality option

Tick this box to generate graphs.

- Here you can generate a graph for a specific district, many districts or all districts.
- Tick the indicated box for the graph presentation of the outcome.
- Leaving this option blank will not generate the basic “run-in” analysis graph.
- *This corresponds to **step 14** in the annex.*

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

2.3.10 Defining your seasonal variation

Figure 13. Choice of the seasonal variation

The screenshot shows a dashboard input field labeled "Season length" with the value "52" entered. A red arrow points from a yellow callout box on the right to the input field. The callout box contains the text: "Enter season length in weeks."

- If you believe that some alarm indicators are better predictors at different times of the year, you can allow for this by dividing the year into different periods.
- For example, “season length”=4 means that the first alarm indicator analysis is based on the first 4 weeks of the year, the second is based on week 5–8, the third is week 9–12, etc.
- A maximum season length value of 52 is acceptable. Your minimum season length value should be > the value you entered for the window size (e.g. 2) for measuring the mean alarm indicator including the current week.
- *This corresponds to **step 8** in the annex.*

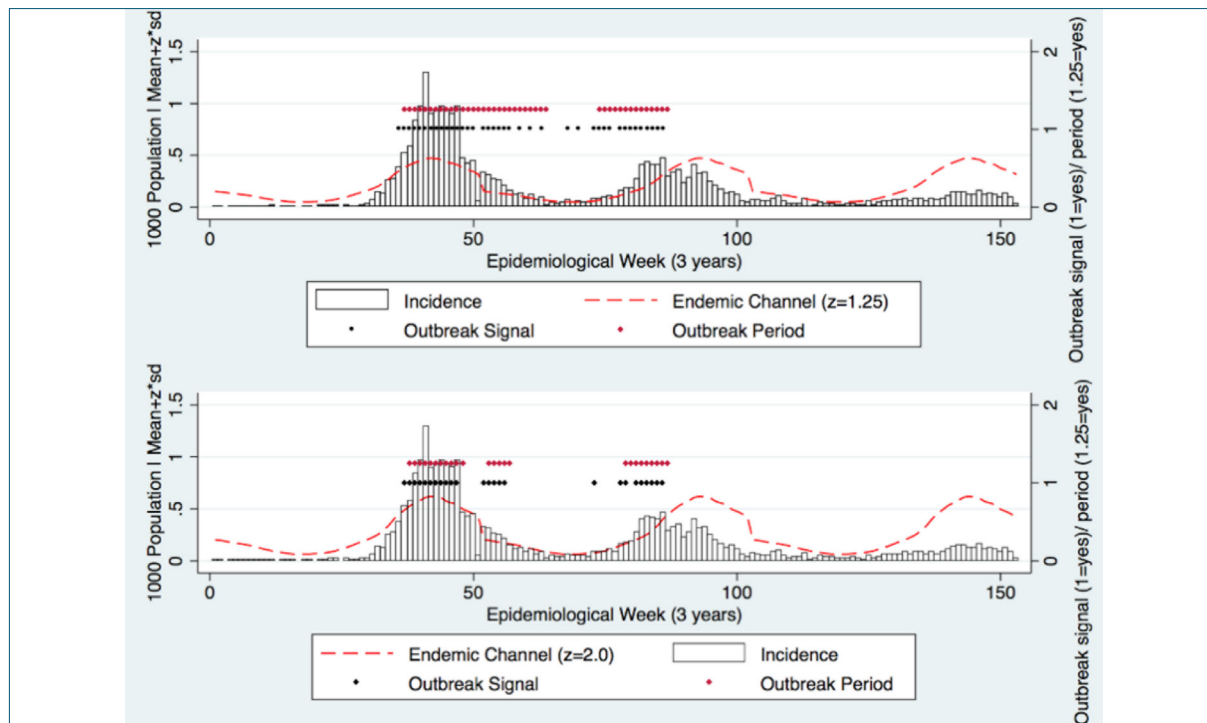
2.3.11 Defining your endemic channel

Figure 14. Changing the endemic channel (the outbreak threshold)

The screenshot shows a dashboard input field with the text "Enter the multiplier of the standard deviation to vary the endemic channel within the evaluation period (e.g. 1.25)" above it. The value "1,25" is entered in the field. A red arrow points from a yellow callout box on the right to the input field. The callout box contains the text: "Enter the desired value to define the multiplying value by the standard deviation (SD)."

- Here you can change the number of cases required to form an outbreak week and therefore an outbreak. By increasing this threshold you will define fewer outbreaks, and by decreasing this you will increase the number of recorded outbreaks (figure 15).
- Enter the desired value to define the “multiplying value” by the “standard deviation (SD)”, for example, $z=1$ is the same SD, $z=1.5$ is one and half times the SD, $z=2$ is two times the SD, etc. See figure 15 below.
- *This corresponds to **step 2** in the annex.*

Figure 15. Modelling the endemic channel^a




^a The modelling illustrates two z-values (top: z=1.25 and bottom: z=2.0) to form the endemic channel – outbreak signals are fewer when z-value is increased.

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

2.3.12 Defining your window size for the outbreak indicator

Figure 16. Defining the number of outbreak weeks (window size)



Enter the choice of outbreak window (e.g 3)

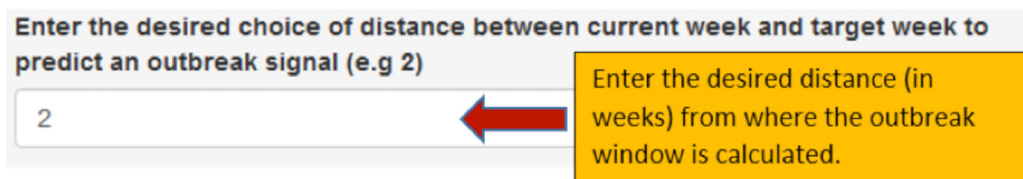
4

Enter the desired value required to define the number of outbreak weeks.

- In section 2.3.5, you defined the outbreak period. In this section, you can choose a suitable window size (or denominator) for calculating the proportion from these outbreak periods. For example, if you choose a window size of “4”, the program will take the “sum of values” of 4 consecutive outbreak periods and divide it by 4 to generate a proportion – this proportion is then needed in the next steps to formulate the outbreak signal.
- Decreasing this window size can increase the sensitivity of predicting an outbreak. At the same time, you may have to choose to increase this window size when you have some missing records in your data.
- The choice of your window size depends on the evaluation criteria that give the best outbreak predictions, see step 14 in the annex.
- *This corresponds to **step 5** in the annex.*

2.3.13 Defining your prediction distance

Figure 17. Defining the prediction distance (in weeks)



Enter the desired choice of distance between current week and target week to predict an outbreak signal (e.g 2)

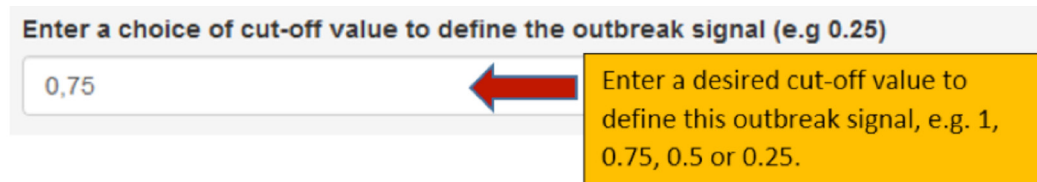
2

Enter the desired distance (in weeks) from where the outbreak window is calculated.

- Enter the desired “choice of distance between current week and target week to predict an outbreak signal”.
- For example, at week 10, if prediction distance=2 and outbreak window=4 (previous option), this will mean that you start counting from week 12 up to week 15 (4 weeks) to predict an outbreak signal.
- *This corresponds to **step 5** in the annex.*

2.3.14 Defining your outbreak signal

Figure 18. Defining the cut-off point of the outbreak signal



Enter a choice of cut-off value to define the outbreak signal (e.g 0.25)

0,75

Enter a desired cut-off value to define this outbreak signal, e.g. 1, 0.75, 0.5 or 0.25.

- In a previous step (2.3.12), you computed the “proportions” from the outbreak periods. The proportion is a value between zero and one and in order to process this value to generate an outbreak signal, it needs to be converted into either zero (which means no outbreak) or one (which means an outbreak).
- Here, in this step, you can choose a “cut-off value” to define this outbreak signal. For example, if you choose a cut-off value of 0.5, then every proportion value above this cut-off value (e.g. $0.7 > 0.5$) will be given the code one (indicating that there is an outbreak), and any proportion value less than this cut-off (e.g. $0.3 < 0.5$) will be given the code zero (indicating no outbreak). This way you have defined your outbreak outcome (a binary variable 0,1), which is the dependent variable needed for the logistic regression.
- The choice of a relevant value depends on the evaluation criteria that give the best prediction of outcome, see step 14 in the annex.
- *This corresponds to **step 7** in the annex.*

2.3.15 Activating the process for monotonic relationship (spline)

Figure 19. Choice of the spline application



Spline option

Tick the box to trigger this option.

- “Spline” is a function that allows the program to treat a certain type of relationship between alarm indicators and dengue cases.
- Without spline, the program will assume linear (non-monotonic) relationships, i.e. as the temperature increases, dengue cases increase.
- With spline, the program will assume monotonic relationships, e.g. as the temperature increases, dengue cases go down. (This is due to decreased vector activity because of excessive temperatures. Another example is increased rainfall where the risk of outbreak increases with increased rainfall, which provides more water for breeding. However, too much rain, e.g. flooding, can flush out larvae and pupae, hence, temporarily reducing the outbreak risk.) Here you can choose between two options.

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

- Enter the desired spline option, i.e. code 0=No (non-monotonic relation) or 1=Yes (monotonic relation).
- You can see how spline affects your (sensitivity and PPV) results by first running without spline and then with spline.
- *This corresponds to **step 6** in the annex.*

2.3.16 Running the program

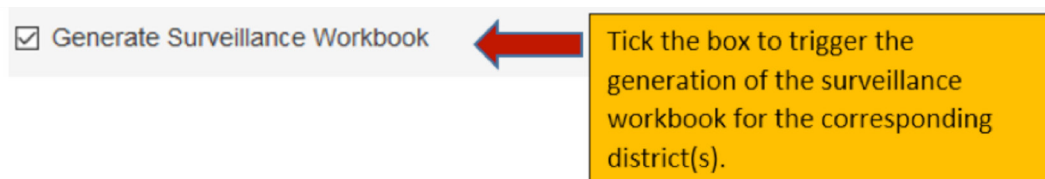
Figure 20. Running the surveillance workbook



- Once you have completed the changes, please run by clicking on the “Run!” button.

2.3.17 Generating the surveillance workbook

Figure 21. Generating the surveillance workbook



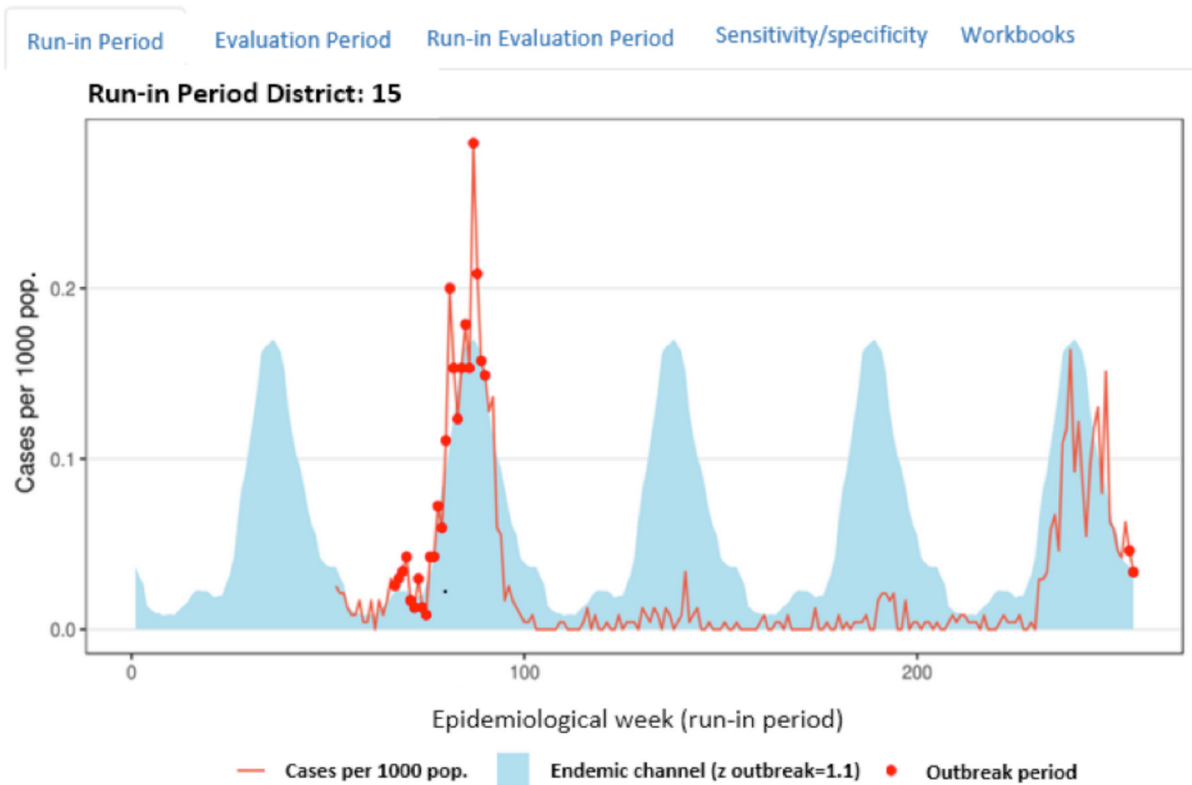
- Leaving the “Generate Surveillance Workbook” box empty (e.g. un-ticked) runs a retrospective analysis of your dataset that allows you to find your alarm indicators. You should do this first (leaving an empty box) every time you make a change to the settings (calibration stage).
- Ticking this box generates the prospective early warning system in Excel and automatically links it to the prospective analysis of the local district (Dashboard II). Only choose this if you have already run the retrospective analysis, made all necessary calibrations, already identified your alarm indicators, and are satisfied with the resulting sensitivity and PPV.
- *This corresponds to **step 16** in the annex.*

2.4 Understanding your calibration outputs

Each time you run the program after making or changing your settings, one table of results will be displayed on the dashboard screen, from which you will need to verify your calibration. Graphs will also be generated to help you understand the current process and whether or not there are any errors or gaps in your data.

2.4.1 Graph: run-in period

Figure 22. Graph of the run-in data showing the endemic channel, number of cases and outbreak period



This graph summarizes the first half of your data, which is the “Run-in Period” data.

- Here you can learn about the duration of the data (from the X-axis; every 51 epidemiological week is one year), the average number of cases/1000 throughout the “Run-in Period” and the size of your “Endemic channel” defined by the given “z” value.

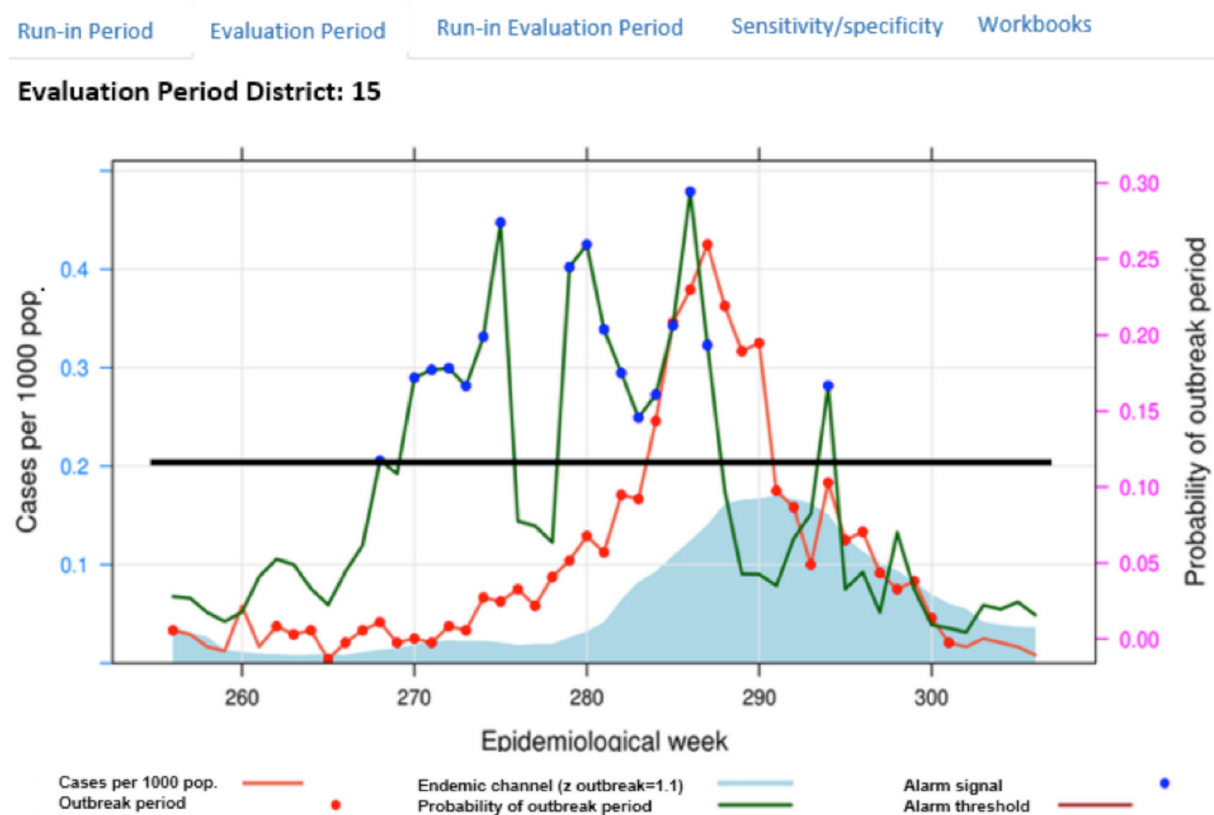
CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

- Cases that exceed the endemic channel and trigger an “Outbreak period” are also presented.
- This graph can also indicate the quality of your data, for example a gap in the endemic channel indicates that there are missing data. At this point, you may need to change your configuration to take into account these missing records by increasing the outbreak and alarm window size!

2.4.2 Graph: evaluation period

Figure 23. Graph of the evaluation data showing the general parameters including the probability of outbreak period

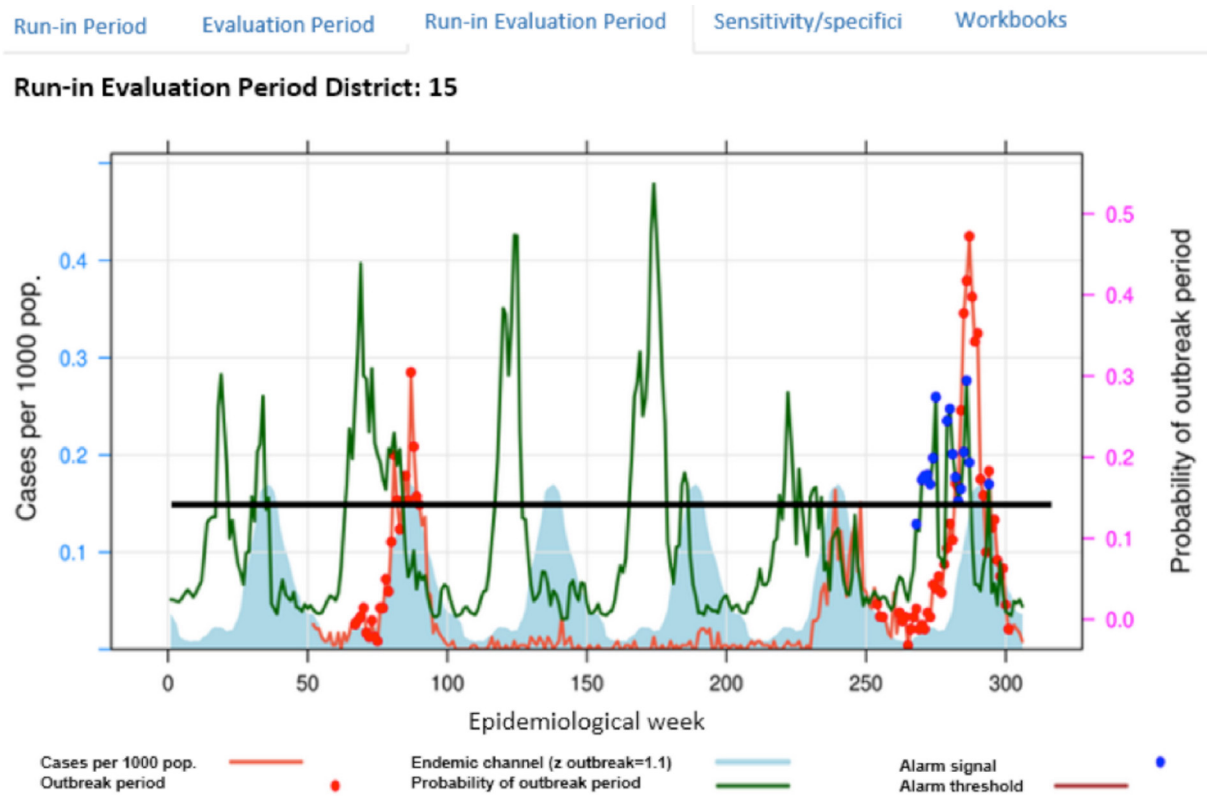


- This graph summarizes the second half of your data, which is the “Evaluation Period” data. It simply tells how your model performs according to the given settings/calibrations in the run-in data.
- Here you can view the continuation of analysis duration (from the X-axis; every 51 “Epidemiological week” is one year, now continuing from the first half of the run-in data).

- More information is provided here. In addition to the information presented in the run-in data, you also find details on “Alarm signal”, “Alarm threshold” and the “Probability of outbreak period”.
- When the probability of outbreak exceeds the given alarm threshold during a particular epidemiological week, the alarm signal (blue dots) is triggered, indicating an upcoming outbreak.

2.4.3 Graph: run-in and evaluation period

Figure 24. Graph of the combined run-in and evaluation data showing all information



- This graph summarizes both the run-in and evaluation data and their corresponding information.
- It provides you with a comprehensive picture of the duration of the analysis and an overview of how well the model is performing.

CHAPTER 2

DASHBOARD I: Data calibration (for user at central level)

2.4.4 Summary of calibration results, and sensitivity and PPV

For the purpose of providing detailed illustration of the derived terms and values to guide you through this process, we split this table into two parts; the first table (figure 25), demonstrates the absolute numbers of the results from the retrospective phase and the success/failure of the parameters (alarm threshold, z value, etc.) that were used to analyse the dataset (this table also summarizes “all cases” and “cases below threshold”, which are not displayed in this figure). The second table (figure 26) describes the proportion of successfully detected outbreaks (sensitivity) and the proportion of false alarms (PPV). The sensitivity and PPV can directly guide you to make appropriate choices for the calibration inputs.

Figure 25. Absolute values in the output table, example showing total number of outbreaks, missed outbreaks, alarms, correct alarms and false alarms

| District | Weeks | Outbreak Weeks | Outbreak Periods | Alarms | Correct Alarms | False Alarms | Missed Outbreaks | No Alarms, No Outbreaks |
|--------------|------------|----------------|------------------|-----------|----------------|--------------|------------------|-------------------------|
| 12 | 51 | 32 | 34 | 12 | 10 | 2 | 25 | 14 |
| 15 | 51 | 34 | 35 | 17 | 16 | 1 | 18 | 16 |
| Total | 102 | 66 | 69 | 29 | 26 | 3 | 43 | 30 |

The definitions of each of the column headings are presented below.

1. “Weeks”: indicates the size of your evaluation data presented by the total number of weeks, e.g. 51=almost one year, 102=two years of data records were used in the evaluation data after you have initially split your surveillance data into run-in and evaluation datasets.
2. “Outbreak Weeks”: a week where the number of cases is above the endemic channel ($z \cdot SD + \text{moving average}$).
3. “Outbreak Periods”: consecutive outbreak weeks (for our definition, this is an outbreak).
4. “Alarms”: total number of alarm signals.
5. “Correct Alarms”: alarm that correctly predicted the outbreak.
6. “False Alarms”: alarms that falsely predicted the outbreak (false positive).
7. “Missed Outbreaks”: alarms that could predict the outbreak if the alarm threshold was lower.
8. “No Alarms, No Outbreaks” (negative correct alarms): no alarms when there is no outbreak.

Figure 26. Sensitivity and PPV (screenshot of the output table – success of predicting outbreaks is shown in sensitivity/PPV)

The screenshot shows a software interface with a tab labeled 'Sensitivity/specificity' and a sub-tab labeled 'Workbooks'. Below the sub-tab is a table with two columns: 'Sensitivity' and 'PPV'. The table contains two rows of data.

| | Sensitivity | PPV |
|--|-------------|------|
| | 0.77 | 0.97 |
| | 0.80 | 0.47 |

Definitions (tests of validity)

- “Sensitivity”: the percentage of outbreaks correctly predicted by alarms. This should be as close to 100% as possible. Minimum value of 50% is acceptable.
- “PPV”: the percentage of alarms that correctly predicted outbreaks. This should be as close to 100% as possible. Minimum of 50% is acceptable.
- To increase the number of correct alarms and decrease the number of false alarms, we suggest that you alter the z value and alarm threshold, then, if necessary, alter the remaining parameters as desired. Then re-run the program.

Operational perspective

- If sensitivity is, for example 60%, then for every 10 outbreaks, the early warning system will detect 6 outbreaks correctly. It will miss 4 outbreaks.
- If PPV is, for example 70%, then 7 out of 10 alarms will be correct, 3 out of 10 alarms will be false. That means that 3 times out of 10, resources will be incorrectly mobilized, i.e. wasted.
- Consider this: in Dashboard II below, you will build the early warning system using your retrospective data. The results above are based on the presence of 1 alarm signal before an outbreak. The early warning system below uses a total of 3 alarm signals to better warn of outbreaks. Therefore, in the presence of 2 or 3 (or more) alarms, the chances of detecting an outbreak increases and the chances of wasting resources decreases with each alarm.

Chapter 3

Dashboard II: Harvesting the results (for users at local level)

3.1 Prospective early warning system

This section summarizes the parameters and settings that were found to be most relevant and provided the best prediction model (highest sensitivity and PPV).

We discuss how to build your prospective early warning system using results from your retrospective analyses in Dashboard I (users at central level).

At this stage, you know which alarm indicators are the best predictors of outbreaks in your district/country. Now we need to apply the derived results within an early warning system that will allow you to detect outbreaks in real-time.

By the end of this chapter, you will be familiar with the analysis outputs, data entry and interpretation of alarms.

PLEASE NOTE, THIS PROSPECTIVE PROCESS NEEDS TO BE PERFORMED ON A WEEKLY BASIS!

3.1.1 Access to and summary of parameters: where to find them in Dashboard II and their usefulness

Dashboard II (provided for users in the EWARS PACKAGE, see section 2.1 above).

- This section under “## parameters”, provides a useful summary of all parameters and settings used in the retrospective analysis (Dashboard I) that were found to be the most relevant and generated the best prediction model (highest sensitivity and PPV) (figure 27).
- Under “## parameters”, the “alarm indicator” headings display the actual alarm indicator(s) that were used in the retrospective analysis and need to be used again in the prospective process for the outbreak prediction. For instance, here the user from Dashboard I used “meantemperature” and “rainfallsum”. Therefore, for the outbreak prediction in the prospective analysis, weekly records from both of these parameters need to be used in Dashboard II.
- Other parameters are also presented. The “prediction distance” is another useful factor in informing the time lag for when to expect an outbreak, in case the alarm signal is triggered (i.e. we have a forthcoming outbreak). For instance, if an alarm signal were triggered at the current week (say week 7 when you enter the prospective information) and the chosen “prediction distance”=2, then you would expect an outbreak to occur during week 9.

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

Figure 27. Summary of parameters from the retrospective analysis (Dashboard I)

The screenshot displays the EWARS v.0.2a interface. On the left is a dark sidebar with navigation links: 'Retropective' (with a red arrow pointing to it), 'Datasets' (with a red arrow pointing to it), and 'Help' (with a red arrow pointing to it). Below these links is the text 'TDR | World Health Organization'. The main content area has a blue header with the title 'EWARS v.0.2a' and a hamburger menu icon. Below the header, the text '##Parameters' is displayed with a red arrow pointing to it. A list of parameters follows:

- Alarm indicator 1:** Meantemperature
- Alarm indicator 2:** Rainfallsum
- Z-outbreak:** 1.25
- Prediction distance:** 2
- Outbreak window size:** 4
- Alarm window size:** 3
- Outbreak threshold:** 0.75
- Alarm threshold:** 0.13
- Outbreak week length:** 3

3.1.2 The three sub-links in Dashboard II: “Retrospective”, “Dataset” and “Help”; their definitions and uses

As outlined at the top left-hand corner in figure 27, there are three main sub-links to follow.

1. The “Retrospective” tab is the main location for running the prospective analysis and interpreting the findings.
2. The “Datasets” tab: this is where you can locate and access the nearest meteorological stations to retrieve the needed prospective alarm indicator information (e.g. weekly mean temperature, rainfall and humidity).
3. The “Help” tab enables you to access the user’s operation workbook and the short training video.

3.1.3 The surveillance workbook tab: all you need to know for this section

Figure 28. Surveillance workbook tab: all you need to know for this section

| Epi_week | End_channel | Out_prob. | Cases | Alarm1 | Alarm2 |
|----------|-------------|-----------|-------|--------|--------|
| 2 | 0.0292372 | 0.1562352 | | | |
| 3 | 0.0279832 | 0.1562352 | | | |
| 4 | 0.0246538 | 0.1562352 | | | |
| 5 | 0.0251035 | 0.1562352 | | | |

- Under the “surveillance workbook” tab, you can set the corresponding “Country code” and “District/municipality code”, enter the assigned password to access your district information and perform the prospective analysis (e.g. CO=Colombia, MX=Mexico, DO=Dominican Republic).
- To demonstrate how to use, view and apply the applications in Dashboard II, we include demonstration data that you can select from the drop-down “country code” menu with the code “XX”.
- The “District/municipality code”: this follows the same district code given in the original surveillance dataset. You need to enter your corresponding district/municipality code to allow access to your account for doing the prospective analysis.

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

- Under “Password” enter the assigned password that was generated and sent to you during the retrospective stage. This is an 8-digit password usually unique to each district to allow access to this account.
- You need to tick the “RE-load data” button to upload the new prospective data when you enter new weekly information under the “Input data” tab. Every time you add more weekly information under this tab, you need to come here and un-tick then tick the “RE-load data” box to upload fresh information. You will need to repeat the process if you have entered incorrect weekly information. After you have entered the correct information, you need to come to this section to un-tick and tick the “RE-load data” box, to see the refreshed information appearing in the table.
- The table displays information on the epidemiological week (“epi_week”), endemic channel (“end_channel”), outbreak probability (“out_prob”), number of “cases”, and average weekly alarm indicator (“alarm 1”, “alarm 2”, etc.). The numbering of the alarm follows that displayed under “##parameters”.

3.1.4 The input data tab: all you need to know for this section

Figure 29. The input data tab: all you need to know for this section

The screenshot shows the 'Input data' tab selected in a dashboard. The tab is highlighted with a red box. Below the tab are several input fields with red arrows pointing to the input values:

- Year:** 2017
- Epidemiological week:** 4
- Population:** 275000
- Hospitalized cases:** 6
- Alarm indicator 1:** 29
- Alarm indicator 2:** 31

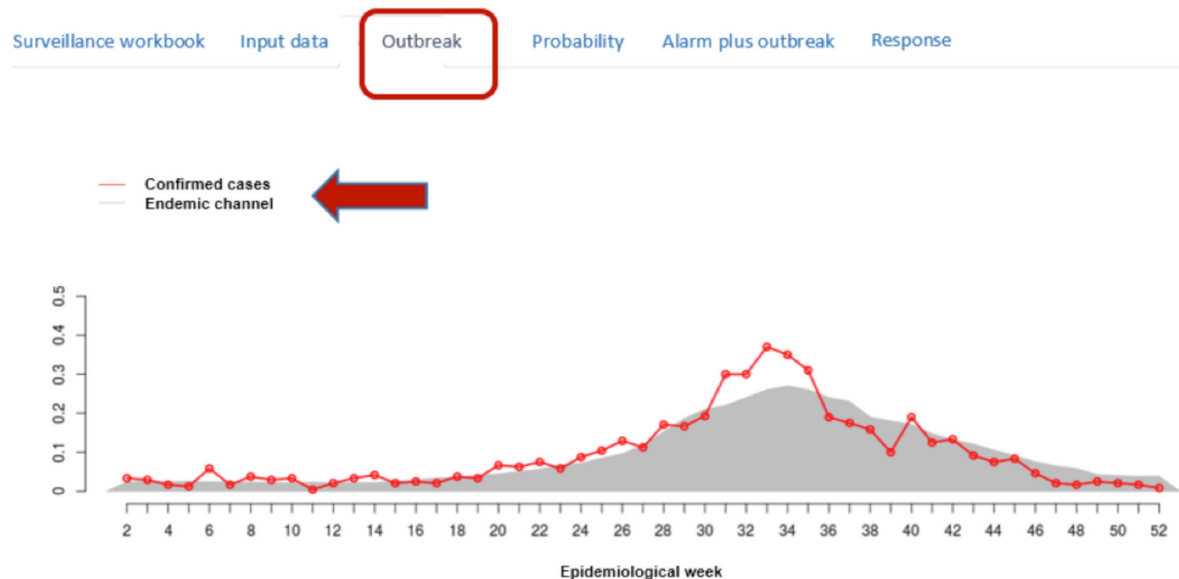
A 'Send' button is also highlighted with a red box at the bottom left of the input area.

- Under the “Input data” tab, you can enter the corresponding information of “Year”, “Population”, “Epidemiological week”, number of “Hospitalized cases” and the average number of alarm indicators to run your prospective analysis of the current week.
- The “Send” button is pressed once you have entered all corresponding information under this tab. After you send this information to the prospective analysis, you need to go back to the previous “Surveillance workbook” tab and click on the “Re-load data” button so that the information is uploaded into the EWARS programming process. Once the “Re-load data” button is clicked, you can see the updated information displayed in the table under the “Surveillance workbook” tab.

- In case you discovered that you have entered incorrect weekly information, you can simply re-enter the corresponding weekly information, send it and then tick “Re-load” data, and you will see the new records in the table.

3.1.5 The outbreak tab: all you need to know for this section

Figure 30. The outbreak tab: all you need to know for this section



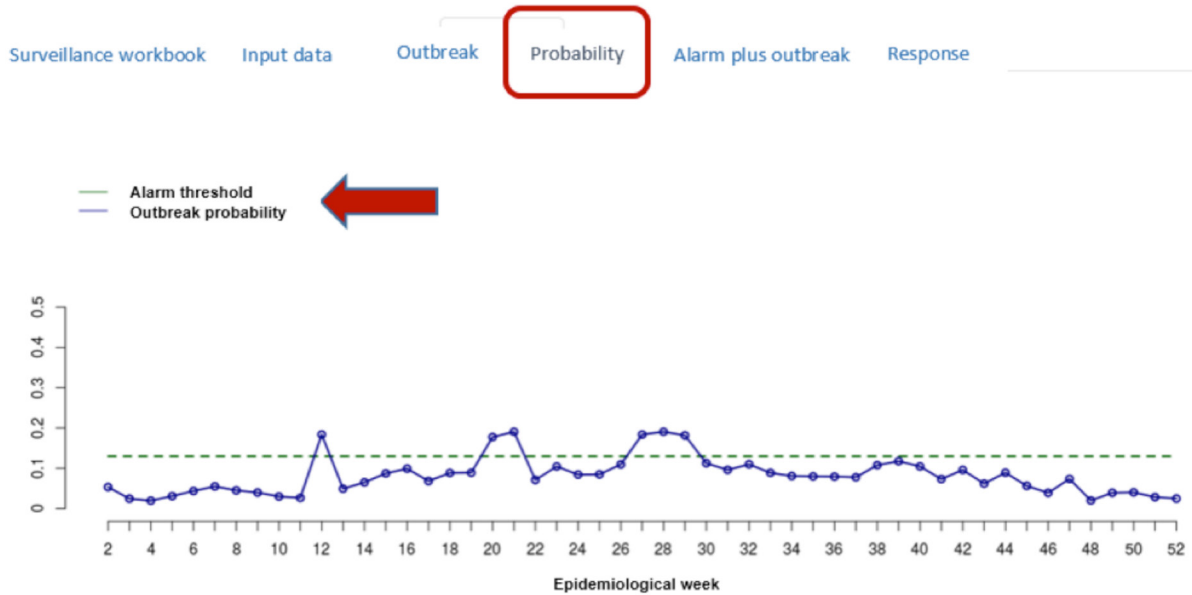
- Under the “Outbreak” tab, you can see a summary of previous history and the current week with respect to the “Endemic channel” and the rate of confirmed cases (or alternative choices of outbreak indicators such as probable cases).
- When “Confirmed cases” (RED LINE) exceed the “Endemic channel” (GREY SHADED AREA), we can say we have an outbreak week!

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

3.1.6 The probability tab: all you need to know for this section

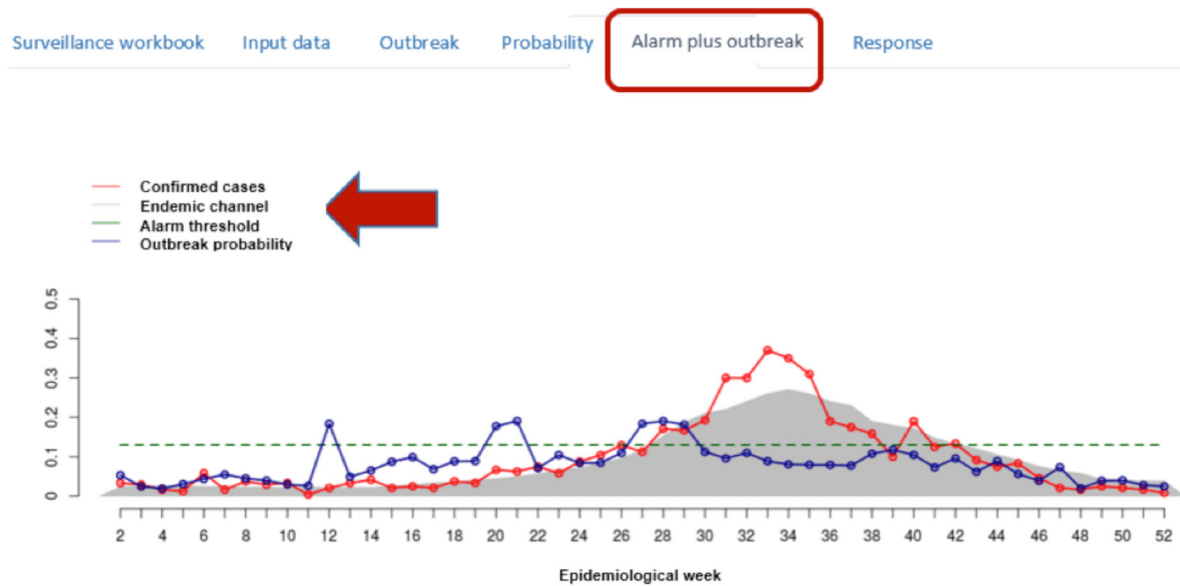
Figure 31. The probability tab: all you need to know for this section



- Under the “Probability” tab, you can see a summary of previous history and the current week with respect to the “Alarm threshold” and the “Outbreak probability”.
- When the “Outbreak probability” (BLUE LINE) crosses the “Alarm threshold” (GREEN LINE), then the outbreak prediction model is alerting you to an upcoming outbreak (alarm signal). By looking back at the “##parameters” summary to see what prediction distance was chosen, for example if prediction distance=2, we say there will be an outbreak happening in next two weeks!

3.1.7 The alarm plus outbreak tab: all you need to know for this section

Figure 32. The alarm plus outbreak tab: all you need to know for this section



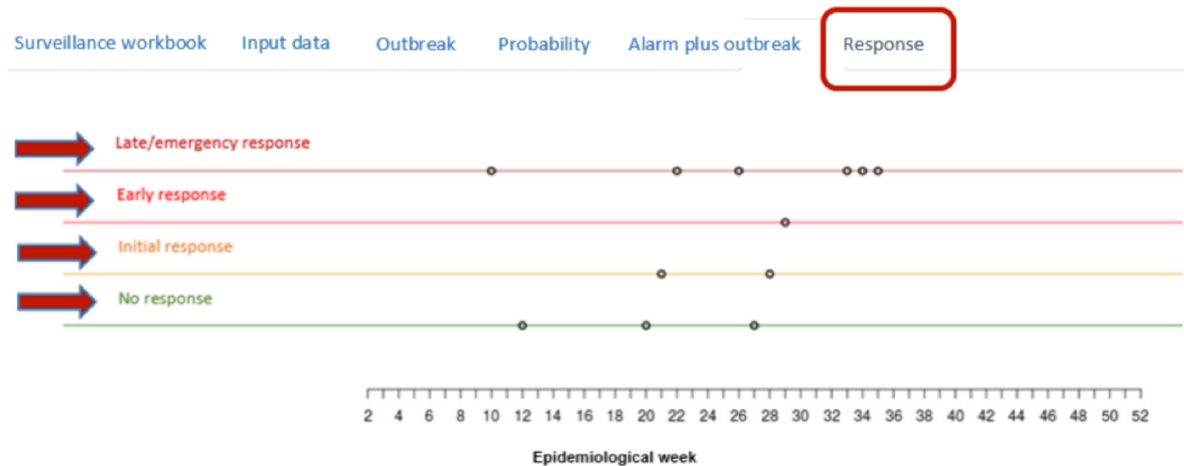
- Under the “Alarm plus outbreak” tab, you can see a summary of previous history and the current week with respect to all the “Confirmed cases”, “Endemic channel”, “Alarm threshold” and the “Outbreak probability”.
- This tab is useful in putting together the relevant information on past outbreak trends and carrying out predictive modelling of the current week, and the probability of an outbreak (alarm signal).

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

3.1.8 The response tab: all you need to know for this section

Figure 33. The response tab: all you need to know for this section



- Under the “Response” tab, you can see a summary of previous history and the current week with respect to the type of response for an upcoming outbreak.
- This section follows the outbreak prediction as outlined in the previous tab (“Alarm plus outbreak”), mainly for the relationship between the outbreak probability and the alarm threshold and, following the national recommendation for outbreak responses at district/municipality level.
- According to national guidelines:
 1. “Late/emergency response” is technically declared when more than three consecutive *outbreak weeks* take place! Furthermore, the “Late/emergency response” is declared when four or more consecutive *alarm signals* occur. Figure 33 shows the “Late/emergency response” in week 10 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) happening at weeks 7, 8 and 9. Then, due to this consecutive occurrence of alarm signals, the program notified a “Late/emergency response” to be considered at week 10. In this scenario, the dashboard must have already declared “Initial response” at week 8 (since there had already been two consecutive alarm signals) and an “Early response” in week 9 (since there had already been three consecutive alarm signals)!
 2. “Early response” is declared when three consecutive alarm signals occur, to avoid “Late/emergency response”. Another example from figure 33 shows that early response in week 29 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) at weeks 27, 28 and again in week 29. Then, due to this consecutive occurrence of alarm signals, the program notified that an early response should be considered at week 29. In this scenario, the dashboard must have already declared “Initial response” at week 28 since there had already been two consecutive alarm signals!

3. “Initial response” is declared when two consecutive alarm signals occur. Let’s take an example from figure 33, the initial response in weeks 21 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) happening at weeks 20 and 21. Then, due to this consecutive occurrence of alarm signal, the program notified that an initial response should be considered at week 21.
4. “No response” is declared when there are no alarm signals or only one alarm signal in the current week.

3.2 Retrieving online meteorological information on weekly alarms (datasets sub-link)

PLEASE NOTE, ONLY THE WEATHER APPLICATION IS CURRENTLY ACTIVE!

3.2.1 Setting your local district/municipality spatial information to detect nearest meteorological station

Figure 34. Retrieving online meteorological information: setting your district’s spatial information

The screenshot shows the EWARS v.0.2a interface. On the left is a dark sidebar with a menu containing 'Retropective', 'Datasets', and 'Help'. The main content area has a light blue header with tabs for 'Weather', 'Twitter', 'Wikipedia', and 'Google search-term'. Below the tabs, the 'Study area' section is visible. It contains three input fields: 'Latitude' with the value '3,139', 'Longitude' with the value '101,6869', and 'Radius (in km)' with the value '100'. Below these fields is a button labeled 'Click to search weather stations...'. Red arrows point to each of the input fields and the search button.

- Under the “Weather” tab, you can set your spatial information of your district/municipality to detect your nearest meteorological station.
- In the “Latitude” box, enter the corresponding latitude value of your district/municipality.
- In the “Longitude” box, enter the corresponding longitude value of your district/municipality.
- Hint! You may use the Google search engine to obtain this information and enter it into this box.
- In the “Radius” box, enter the size of the area for your search. For example, entering a value of 100, the search for nearest meteorological station will cover an area of 100 sq km within your district/municipality.
- Once all the information is entered, “Click to search weather stations” button and a map will appear to show the locations of nearest meteorological stations.

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

3.2.2 Setting your period of interest for deriving the meteorological information

Figure 35. Retrieving online meteorological information: defining the period of interest

Observation period

The earliest date (yyyy-mm-dd)

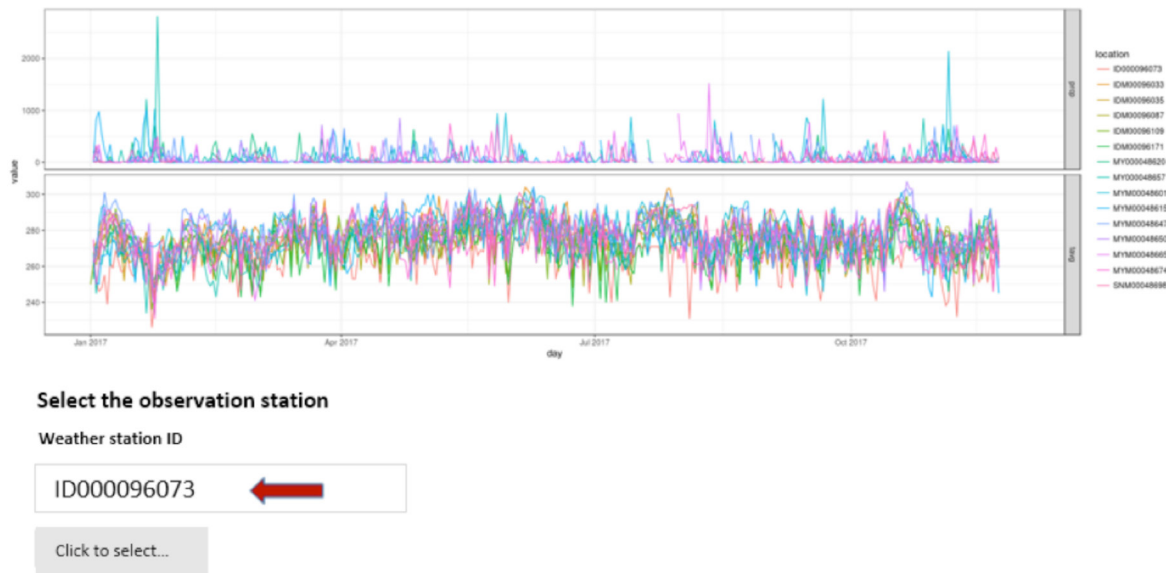
The latest date (yyyy-mm-dd)

The map displays the Kuala Lumpur region with various districts and municipalities. A red circle highlights the study site in Petaling Jaya. Several blue circles represent meteorological stations located in areas like Sungai Buloh, Hulu Langat, and Sepang. The map includes labels for major locations such as Kuala Lumpur, Petaling Jaya, Shah Alam, Kajang, Putrajaya, and Seremban.

- This section defines the period for the meteorological information to be obtained.
- In this map, the blue circles are the meteorological stations and the red circle is your study site (district/municipality).
- Enter the date for the start of the meteorological information under “The earliest date” and the end date under “The latest date”. For example, if you are interested in obtaining all records of meteorological information for last two weeks from your current week, say, 26 November 2017, then the start date will be 12 November 2017 (“2017-11-12”) and the end date will be 26 November 2017 (“2017-11-26”).
- Then, click on the “Click to proceed...” button.

3.2.3 Selecting a desired meteorological station from the obtained list

Figure 36. Retrieving online meteorological information: selecting a desired meteorological station



This section displays the list of all meteorological stations available to retrieve the alarm information. The list on the right-hand side refers to each one of these stations.

- The graph shows an overview of how each station presents its meteorological information and whether there is an agreement between them. As a user, you may select anyone that you find representative and relevant in terms of being in the nearest location to your district.
- Enter the corresponding station code (from the list presented at the right-hand side) in the “Weather station ID” box, then “Click to select...” button.
- Once this is done, you will see a list of the alarm indicators presented for the period you initially selected.

CHAPTER 3

DASHBOARD II: Harvesting the results (for users at local level)

3.2.4 Using the derived information of the alarm indicators

Figure 37. Reading the retrieved online meteorological information for alarm indicators

Show entries Search:

| | YM | tavg | prcp |
|----|----------|-------|------|
| 1 | Jan 2017 | 26.97 | 2448 |
| 2 | Feb 2017 | 27.21 | 1688 |
| 3 | Mar 2017 | 27.65 | 2423 |
| 4 | Apr 2017 | 27.73 | 1171 |
| 5 | May 2017 | 28.29 | 655 |
| 6 | Jun 2017 | 28.19 | 389 |
| 7 | Jul 2017 | 28.03 | 138 |
| 8 | Aug 2017 | 27.51 | 358 |
| 9 | Sep 2017 | 27.27 | 1168 |
| 10 | Oct 2017 | 27.85 | 507 |
| 11 | Nov 2017 | 27.43 | 1306 |

Showing 1 to 11 of 11 entries Previous Next

- This section displays the derived information of alarm indicators from the selected meteorological station. You can use the available information for your data input to run the prospective analysis.
- The column “YM” represents the period corresponding to when the alarm indicator information was collected.
- The column “tavg” represents the temperature average.
- The column “prcp” represents the precipitation or the rainfall in millimetres.
- Based on this information, you can calculate weekly alarm indicator averages.

Annex 1.

Technical guide

1. Introduction

This technical guide explains the scientific rationale behind *The Early Warning and Response System (EWARS) for Dengue Outbreaks: operational guide using the web-based dashboard*.

This project has focused on developing a validated R-based model that can enable the prediction of “out of control” dengue cases (outbreaks) as defined by dengue incidence (probable/hospitalized cases). The early warning system detects changes in the alarm indicators (entomological, meteorological and epidemiological) to predict dengue outbreaks. The purpose of this *Guide* is to provide an overview of the applied method and its rationale. Additionally, it presents block diagrams to describe the applied method with further details on each step. This will assist the user to follow the *Guide* and link to the corresponding step number. It also discusses general requirements for the R program and potential errors while using the available do-files.

2. Methodology

The general methodological concept of this analysis follows two major phases that are summarized below.

2.1 Retrospective phase

This phase uses retrospective surveillance data to create two datasets: (i) run-in data, used to develop the prediction model; and (ii) evaluation data, used to evaluate the derived parameters from the prediction model.

2.1.1 Run-in data

This dataset uses past records to estimate/calibrate the model parameters of the relationship between the outbreak indicator and alarm indicator(s). This parameterization is then tested during the evaluation process (step 2, below) and applied by the user to predict an outbreak. These data include information on the year, week and district, outbreak indicator (probable, confirmed cases, or other forms of outbreak indicators), and alarm indicator(s), such as the weekly mean temperature, sum of rainfall, mean humidity and probable cases.

2.1.2 Evaluation data

This dataset is used to: (a) evaluate the prediction model; and (b) provide summary statistics that are used to build the prospective early warning system.

ANNEXES

2.2 Prospective early warning phase (i.e. Dashboard II)

The populated file of results (final parameters) in Dashboard II, allows the user to enter prospective information to estimate the probability of an outbreak in a foreseeable period. This is simply derived by inserting weekly data of outbreak and alarm indicator(s) for the district(s) of interest.

2.2.1 Rationale for the approach

The final model derived for estimating the probability of an outbreak is generated via systematic steps (creating/evaluating parameters) by assessing the association between the level in the alarm indicators (i.e. mean rainfall, mean temperature, mean of humidity, etc.) and dengue outbreaks. Logistic regression is used to assess this association. This regression model processes the computed proportions of the outcome (computed via a cut-off value given by the user), to predict the probability of dengue outbreaks for a forthcoming period. Throughout these steps, the user is able to define relevant measures such as the size of the endemic channel, the outbreak duration, the alarm threshold and the prediction distance. Descriptions of these measures can be found in the following chapter.

3. Structural design

The block diagram in figure 38 outlines 16 steps in the overall process distributed across two different phases. Phase I (retrospective) is divided into: (i) the “Run-in data”; and (ii) the “Evaluation data”. Phase II (prospective) is the final analysis using summary statistics to populate an Excel-based early warning system that can be used in real-time to detect future dengue outbreaks. A summary of each step is presented below.

3.1 Phase I (retrospective: Dashboard I)

3.1.1 Using the run-in data to create the prediction model (part one)

Step 1

In this step, the original data are divided into “Run-in data” and “Evaluation data”. The user determines this by entering the period (cut-off), in year-week, when the run-in data end and the evaluation data begin. A minimum of two years’ data are required for the run-in data though more than two years’ data are recommended.

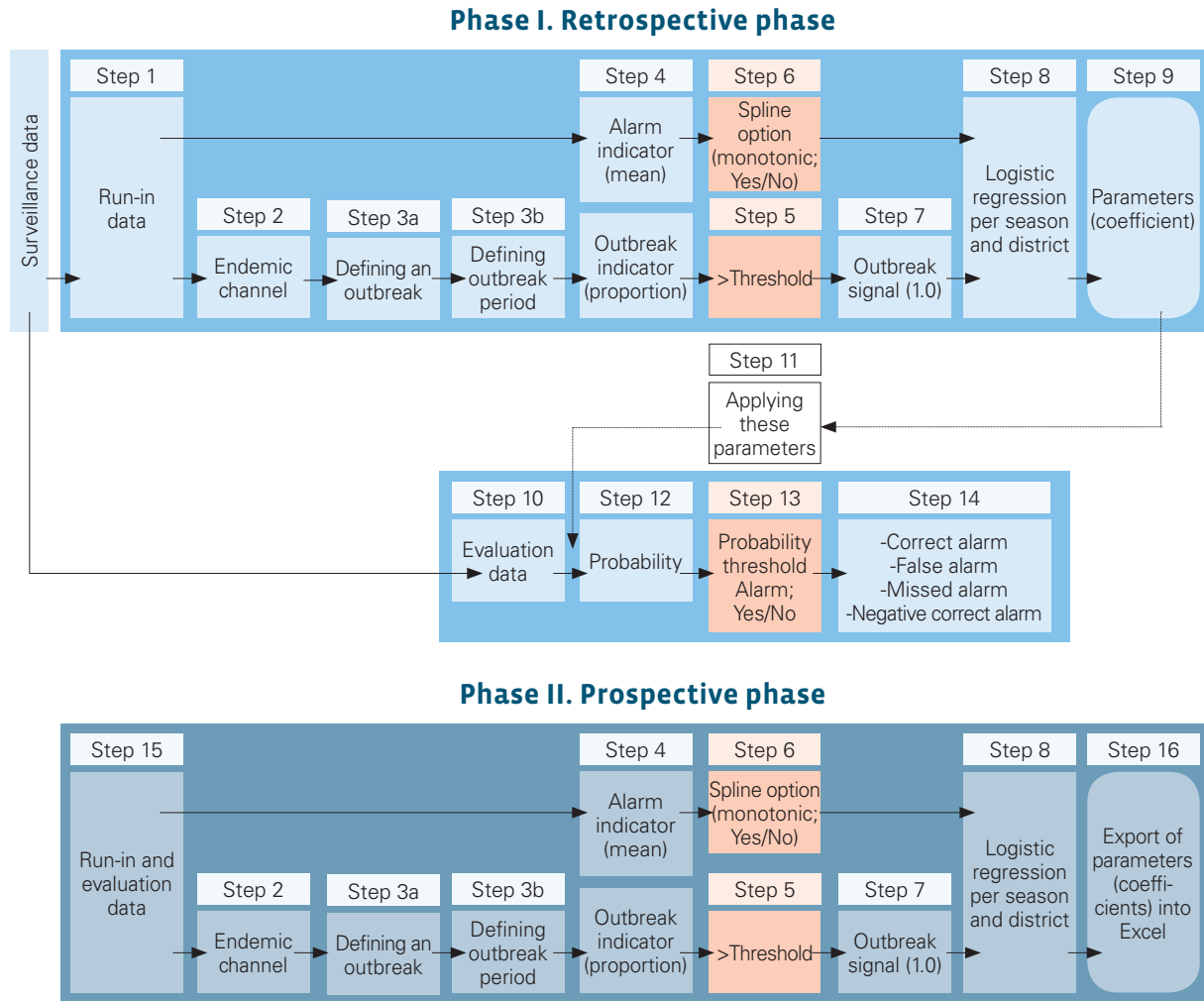
Step 2

This step refers to the “Endemic channel” and is represented by the following equation:

$$\text{Endemic channel} = (Z * SD) + \text{moving average}$$

“Z” is a multiplier value of the SD provided by the user to vary the breadth of the endemic channel. This part will assist in declaring the “out of control” status. For instance, a value of $Z=1.5$ would increase the breadth one and a half times the expected normal range of the number of dengue cases. The moving average is the mean number of dengue cases within the expected normal or seasonal range, calculated for a window size of three preceding and three succeeding weeks from the point (week) of measure.

Figure 38. Block diagram illustrating the process and steps of the retrospective and prospective phases (I, II)



ANNEXES

Step 3a

In this step, an “outbreak” is defined by the proportion of number of incident cases in relation to the annual population for a corresponding district. The user is given the option to choose which outbreak indicator to use, for example, probable or hospitalized cases.

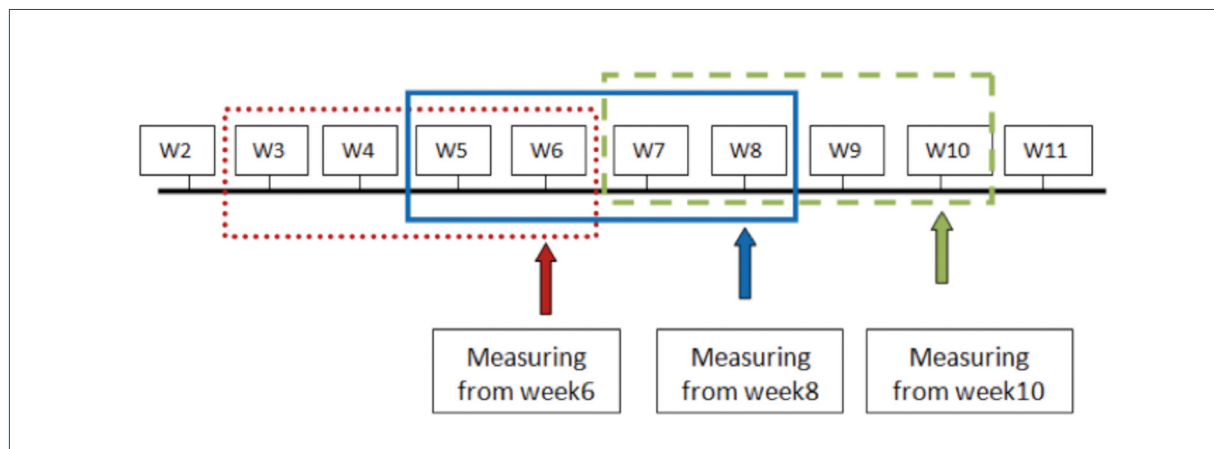
Step 3b

In this step, the user may choose the desired length of an “outbreak period”. The user can define the length of the outbreak period by determining the number of consecutive weeks (1, 2, 3, etc.) that dengue cases must exceed the endemic channel. For example, for “outbreak_week_length=3”, we can say that our outbreak period is defined when the dengue cases continue to exceed the endemic channel value for 3 consecutive weeks.

Step 4

This step illustrates how data from the “Alarm indicator” is used in relation to the alarm window (figure 39). We measure the mean of each alarm indicator (mean temperature, weekly rainfall, etc.) during the preceding pre-defined alarm window size (e.g. choosing alarm window size of 4, this step will measure the mean of each of these alarm indicators during the last four consecutive weeks including the week we are measuring from).

Figure 39. Illustrated process of measuring the alarm indicator(s) for a defined alarm window size



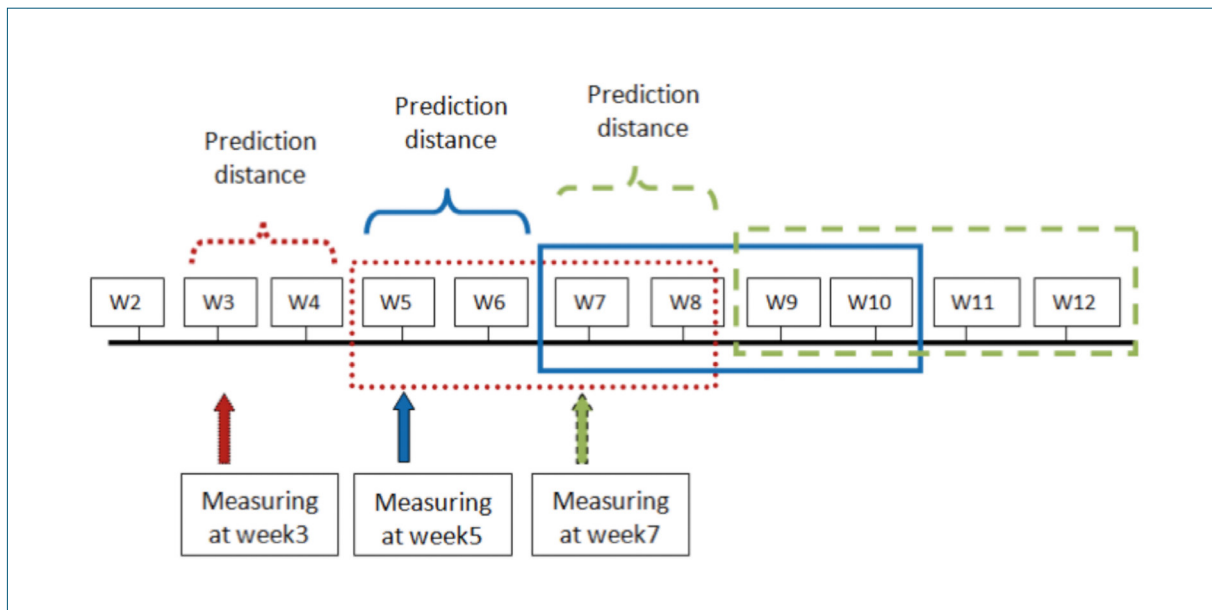
Step 5

This step illustrates how data from the outbreak period are used with the window size to define an outbreak (figure 40).

Here, the formulated outbreak period (from step 3b) is converted into proportions by dividing the “sum of values” from the outbreak periods by a given window size (or denominator) which generates values that lie between “0” and “1” (proportion). The user usually defines the window size during the calibration stage, which defines these proportions and which will, in later steps, be used to form the outbreak outcome (using the outbreak threshold).

For example, if the user chooses a window size of 4, then the analytical program will take the “sum of values” of 4 consecutive outbreak periods and divide it by 4 to generate a proportion. Figure 40 illustrates this process by also describing how other information (also defined by the user) can be used in this step. The prediction distance is another factor used in this process that defines the point where computing the proportion should begin in the data.

Figure 40. Illustration of the moving average process of measuring the outbreak indicator for a defined outbreak window size



ANNEXES

Step 6

The “Spline option” handles a non-monotonic relationship between the alarm indicator(s) and the outbreak indicator(s). A non-monotonic relationship occurs when there is an increase in the mean alarm indicator during a particular period with a decreased number of dengue cases (outbreak) for the same period. Applying this option (yes=1 (non-monotonic)/no=0 (monotonic)) is decided during the evaluation stage (section 2.3.15 Choice of the spline application; line 35 in the do-file).

The user can observe the performance (evaluation table) of the prediction model when the spline is selected or removed. The method of checking the performance of the prediction model is illustrated in step 14 below.

Step 7

In step 5, proportions from the outbreak periods were computed using a pre-defined window size and prediction distance. In this step, the outbreak signal is formulated by converting the derived proportions into a binary variable (0=no outbreak or 1=outbreak), which is necessary for running the logistic regression in the next step.

In this step (7), the user chooses a suitable cut-off point (threshold) which defines an “Outbreak signal” (i.e. code “1”) or not (i.e. code “0”). For example, if you choose a threshold value of 0.5, then any proportion above this threshold (e.g. $0.7 > 0.5$) will be given the code one (an outbreak signal) and any proportion value less than this threshold (e.g. $0.3 < 0.5$) will be given the code zero (not an outbreak signal). This will result in a binary outbreak outcome variable that is ready to enter the logistic regression process.

Step 8

In this step, we assess the association between the binary outcome of the outbreak signal (0, 1) and the alarm indicators using logistic regression. This relationship will be influenced by the “season_length” input. For example, if season_length=4 weeks, the logistic regression model will generate a coefficient for this period length alone. Then the next coefficient will be specific to the following 4 consecutive weeks and so forth.

Step 9

The coefficients generated from this regression model will be stored/applied during the evaluation stage.

3.1.2 Using the evaluation data to assess how the derived parameters from the run-in data would predict an outbreak (part two)

Step 10

Refers to the second part after splitting the surveillance data, which we call the “Evaluation data” (the first part of the original data was used in step 1, i.e. the “Run-in data”). The user determines this by entering the period, in year-week (e.g. 201452) for when the run-in data ends and the evaluation data begins. A minimum of two years’ data records is required for each dataset.

Step 11

In this step, we are evaluating the coefficients that were initially generated by the “Run-in data”. This evaluation is performed by applying these coefficients to the “Evaluation data” to observe the performance of the prediction model in detecting outbreaks.

Step 12

The derived function from the logistic regression step 8 and its parameters are used to estimate the “Probability” of an outbreak occurrence.

Step 13

The generated probability from step 12 will require a threshold to determine an alarm signal. For instance, for a probability of 0.4, with a threshold of 0.3 (i.e. the probability \geq threshold), this record is said to be an alarm signal. If, however, the probability were less than the threshold, then it is not an alarm signal.

Step 14

In order to ensure that the given threshold (in step 13) is reliable to predict an outbreak, the program will further present four evaluation criteria to assess the choice given in step 13. These criteria are:

1. correct alarm: probability \geq threshold with a true outbreak (for a target period);
2. false alarm: probability \geq threshold with no true outbreak (for a target period);
3. missed outbreak: probability \leq threshold with a true outbreak (for a target period);
4. no alarm, no outbreak: probability \leq threshold with no true outbreak (for a target period).

An optimal threshold (option in step 13) would lead to an increased number of correct alarms and negative correct alarms but a decreased number of false alarms and missed alarms. By observing these results (which will be displayed on the screen of the program – output), the user can alter the threshold in step 13, accordingly.

ANNEXES

3.2 Phase II (prospective surveillance: Dashboard II)

Step 15

In this step, the program will run through the entire dataset (i.e. both run-in + evaluation) to populate the output file (Dashboard II) that includes information on the probability formula, model parameters and details on the applied settings from the evaluation period. Once the user has entered real-time prospective data into the empty columns, the file will automatically estimate and graph the probability of an outbreak to predict an outbreak.

Step 16

In this final step, the program will populate the surveillance workbook, which is directly linked and displayed in Dashboard II. The user will input data on year, week, outbreak indicator, and alarm indicator(s) for a corresponding district directly into Dashboard II, which can automatically calculate the probability of an outbreak and produce an instant graphical presentation and alarm signals/responses.

NB: Settings from steps 2–8 in phase II (the prospective surveillance) will be automatically fixed according to the performed settings during the evaluation stage, i.e. the tool applies the same settings that the user approved during the evaluation stage to provide best prediction.

The Special Programme for Research and Training in Tropical Diseases (TDR) is an independent global programme of scientific collaboration established in 1975. It has a twin mission to improve existing and develop new approaches for preventing, diagnosing, treating and controlling neglected infectious diseases, and to strengthen the capacity of developing endemic countries to undertake this research and implement the new and improved approaches. TDR is sponsored by the following organizations:



TDR/World Health Organization
20, Avenue Appia
1211 Geneva 27
Switzerland

Fax: (+41) 22 791-4854
tdr@who.int
www.who.int/tdr

ISBN 978 92 4 151435 4

