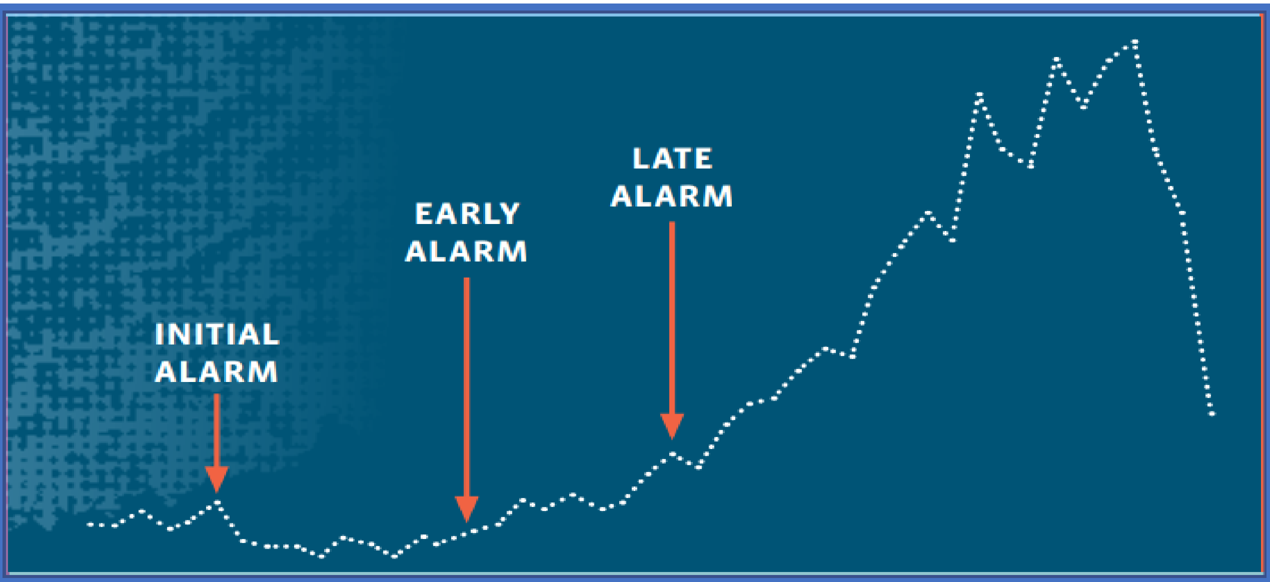
Operational guide
Using the Web-Based Dashboard



The Early Warning and Response System (EWARS+) for Selected Climate Sensitive Disease Outbreaks

Operational guide

Using the Web-Based Dashboard
(Ethiopian context)

The Early Warning, Alert and Response System (EWARS+) for Selected Climate Sensitive Disease Outbreaks

Contents

LIST OF FIGURES

PREFACE

Welcome to the Early Warning, Alert and Response System (EWARS+) for selected climate sensitive diseases outbreak: operational guide using the web-based dashboard. This guide will provide you with the information and tools necessary to use and analyze surveillance data to predict diseases 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. The computerized statistical program that the EWARS will be using is called 'R', version 3.4.3 and later. 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

Using the web-based dashboard was initially 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. Maquins Odhiambo Sewe developed the analytical programming during the transfer process from the STATA-to-R statistical package; and, Aditya L. Ramadona provided insight and key developmental contributions to the web-based dashboard. Further, David Benitez (Mexico) and Balvinder Gill (Malaysia) helped for beta testing the materials, as well as the following colleagues contribute valuable country data to enable the development of EWARS: Roberta Carvalho (Brazil); Giovanini Coelho (Brazil); Lokman Hakim (Malaysia).

For local utilization of the modified EWARS+ tool, the original operational guideline is customized to the current version by Sisay Wondaya (Ethiopia) and other members of the early warning case team of Ethiopia Public Health Institute (EPHI).

Acronyms

EPHI	Ethiopian Public Health Institute
WHO	World Health Organization
PHEM	Public Health Emergency Management
CSDS	Climate Sensitivity Disease Surveillance
EWARS+	Early Warning, Alert and Response System $_{plus}$
NTD	Neglected Tropical Diseases
HMIS	Health Management information System
MOH	Ministry of Health
RHB	Regional health Bureau
ZHD	Zonal Health Department
WASH	Water Sanitation and Hygiene
CSA	Central Statistics Agency
SD	Standard Deviation
DIR	Disease Incidence Rate
RR	Relative Risk
ROC	Receiver Operating Characteristics

Glossary

Terms	Definition
R-package	Specialized software used to process and analyze 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).
Late/emergency response	A response type that declared when more than three consecutive *outbreak weeks* take place
Early response	A response type that declared when three consecutive alarm signals occur
Initial response	A response type that declared when two consecutive alarm signals occur
No response	A response type that declared when there are no alarm signals or only one alarm signal in the current week

Chapter one

1 Introduction

1.1 Background

Ethiopia is one of the most vulnerable countries to climate variability and change which in turn putting disastrous impact on the livelihood of the people. Frequently faced with climate-related hazards, drought, floods and disease outbreaks. Climate variability and change are one of the main drivers for diseases like malaria, yellow fever, dengue fever, meningitis, leishmaniasis, chikungunya, cholera and diarrhea. Hence, there was a need to develop an evidence-based early warning system for detection and management of different outbreaks.

Effective disease surveillance and early warning system is essential to detecting disease outbreaks quickly before they spread, cost lives and become difficult to control. Early Warning, Alert and Response (EWAR) is one of the cornerstones of public health information system (PHIS) and detection of acute stage public health emergency. Early Warning, Alert and Response System (EWARS) is designed to improve disease outbreak detection on climate sensitive infectious diseases. If effectively implemented, EWARS can contribute to the reduction of risks to climate sensitive diseases and also contribute in the establishment of climate resilient health system in the country. It also contributed to the achievement of the Sustainable Development Goals (SDGs) in reducing the loss of life.

The Ethiopian Public Health Institute (EPHI) in collaboration with different stake holders has been working to advance the health early warning system (EWARS). Ethiopian Public Health Institute (EPHI) and WHO have been implementing a joint initiative to advance and validate decision supporting tools (EWARS) to enhance detection, preparedness and timely response to climate change sensitive disease outbreaks.
It is clear that there is a need of a standardized and compatible approach to deploy alarm signals in a predictive and operational way. This is the base that an accessible, adaptable and user-friendly web-based Early Warning, Alert and Response (EWARS) tool was developed.

The tool can ensure enhanced, fast and secured communication between national and regional (subnational) levels, and standardized utilization of surveillance data.

This tool is developed by WHO's Neglected Tropical Disease (NTD) team and locally customized and implemented by EPHI early warning case team. This is also in collaboration with Resolve to Save Lives (RTSL) for proper utilization and scale-up the pilot project to the wider health system. The tool is developed for outbreak prediction and risk mapping of different climate sensitive diseases such as chikungunya, dengue, malaria, yellow fever, cholera and meningococcal meningitis.

1.2 Introduction to the model

The tool is upgraded to EWARS+ with improvement and additional features. This modeling tool is developed taking advantage of the hierarchical nature of the country's surveillance data: space (districts/Woredas) and time (weekly report of incidence and alarm indicators). The model prediction is based on exposure–lag–response functions of key climate variables (e.g., rainsum, mean temperature, mean of humidity, …etc.).

Negative binomial regression (alternatively zero inflated) is used to assess the association between alarm indicators and outcome variable (outbreak). A novel Bayesian computational method called the Integrated Nested Laplace Approximation (INLA) is used for parameter estimation. In order to describe the model, consider the response observation $y_i, i = 1,2, \dots n$ (number of outbreaks in ith district). The mean μ_i of observation y_i can be linked to the linear predictor η_i via a convenient link function $g(.)$, so that $g(\mu_i) = \eta_i$. Then, the structured additive predictor η_i accounts for effects of various covariates in an additive way:

$$\eta_i = \alpha + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f^{(k)}(u_{ki}) + \varepsilon_i , \qquad i = 1,2, \dots n.$$

Here, α is the intercept, β_j, coefficients of some covariates $z_j, j = 1 \dots n_\beta$, functions $f^{(k)}$ defines n_f random effects on some vector of covariates $u_k, k = 1 \dots n_f$ and the ε_i 's are unstructured terms. For estimation of parameters, spatial-temporal covariance included to provide robust estimates using distributed lag non-linear Bayesian framework (Bayesian inference with INLA).

EWARS+ is designed to complement the Health Management information System (HMIS), which is one of the building blocks that are essential for strengthening health system, by providing timely information in a systematic way prior to an outbreak in order to make informed decisions and take action. The Ministry of Health (MOH) Government of Ethiopia gives due recognition to HMIS as a management support system for improving the health system in Ethiopia by providing continuous information support to decision making process at each decision-making level– federal MOH/EPHI, Regional Health Bureau (RHB), Zonal Health Department (ZHD), Woreda Health Office (WorHO), and health facility. EWARS objectives includes:

- Detect the likelihood of an outbreak as early as possible
- Provide advice and warning about impending outbreak in a given area. So that the necessary preparedness measures can be taken
- Increase risk knowledge about possible hazards/outbreak
- Suggest types of response assistances required
- Continuously monitor hazard parameters and precursors
- Regularly assess implementation of safety measures specially in most seriously affected areas
- Conduct disaster area assessments whenever disasters occurred

Developing and implementing an effective early warning system requires the contribution and coordination of a diverse range of individuals and groups. These includes communities, local and national governments, regional institutions and organizations, international bodies, non-governmental organizations, the private sector, and the science and academic community. Hence, for effective EWARS, the involvement and active participations of these key actors are needed. Only this multi-sectoral approach can reduce public health treats in a country.

Chapter two

2 Preparing dataset

2.1 Surveillance data, contents and format

Before running the analysis, your dataset should contain the list of necessary variables. All necessary data can be collected from different sectors like Meteorology, WASH, CSA, etc. These data then be organized, cleared, validated and integrated. All variables should be aggregated in to weekly data. For example, weekly temperature may be obtained by averaging the daily temperatures of the week. Variables need to be written (in lower cases) as presented in Figure 1. Any change of the variable name needs programing update on the source file. At least one alarm indicator is required (e.g. mean temperature) but additional alarm indicators can also be included. A minimum of three or more years' spatiotemporal data records is needed to run (calibrate) this tool.

Figure 1. Example of surveillance data to use in EWARS+

A	B	C	D	E	F	G	H	I	J	K	L	M
year	district	pulation	week	weekly_hospitalised_cases	rhdailymean	rainsum	meantemperature	WindSpeed				
2017	30115	225802	1	93	28.82285714	0	20.15142857	1.984285714				
2017	30115	225802	2	78	27.24142857	0	20.37857143	2.241428571				
2017	30115	225802	3	68	28.85857143	0	21.84357143	1.755714286				
2017	30115	225802	4	127	32.54571429	0	22.88642857	2.02				
2017	30115	225802	5	92	42.82142857	1.32	23.09928571	2.347142857				
2017	30115	225802	6	88	38.80285714	3.46	24.28642857	1.964285714				
2017	30115	225802	7	83	54.06285714	18.85	21.57357143	2.212857143				
2017	30115	225802	8	82	52.94571429	5.91	21.67714286	2.234285714				
2017	30115	225802	9	65	32.98285714	0.86	23.23071429	2.1				
2017	30115	225802	10	90	41.48285714	7.53	24.22285714	2.164285714				
2017	30115	225802	11	57	39.75142857	6.55	24.78071429	2.225714286				
2017	30115	225802	12	64	31.33857143	5.73	24.19	2.541428571				
2017	30115	225802	13	73	33.59714286	5.6	24.70785714	2.325714286				
2017	30115	225802	14	49	35.85714286	1.63	26.565	2.702857143				
2017	30115	225802	15	36	45.15857143	20.74	25.91785714	1.895714286				
2017	30115	225802	16	47	44.73142857	1.25	26.25214286	2.17				
2017	30115	225802	17	52	61.66	63.95	22.79285714	2.385714286				
2017	30115	225802	18	55	55.89285714	13.75	24.01857143	1.584285714				
2017	30115	225802	19	51	63.37571429	36.55	23.92	1.562857143				
2017	30115	225802	20	63	81.59714286	79.87	21.23357143	1.732857143				
2017	30115	225802	21	68	74.08285714	38.94	22.56214286	1.808571429				
2017	30115	225802	22	71	73.82285714	19.3	22.26142857	1.904285714				
2017	30115	225802	23	94	73.96285714	33.75	22.84928571	1.89				
2017	30115	225802	24	93	76.40857143	34.58	22.58071429	1.858571429				

The user can add more "alarm" indicators here

This historical data is required to do the risk mapping analysis and provide users with an overview of 'hotspot' patterns over time. The district code in column two should match with the code in area shapefile.

Following the model calibration, a prospective data is needed to forecast the probability of an outbreak with suggested response type in a foreseeable period. The data structure is similar with Figure 1; however, it shall be the most recent data. The longer the historical period the better but it is recommended to have at least 4 weeks of most recent disease and alarm indicator records (i.e., data from prediction time week plus previous weeks).

2.1.1 General variables

1. **year**: the year when the data were collected. The year must be entered in full using four numbers (e.g. 2017, 2018, etc.);
2. **week**: the ISO 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.);
3. **district**: a number (code) that represents the district, locality or municipality where data were captured (e.g., 30115, 30215, 30605, 30712, etc.);
4. **population**: the mid-year population size of a district reported in absolute numbers in the surveillance data. See column C in figure 1, above.

2.1.2 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.

2.1.3 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:

1. **rhdailymean**: the weekly mean relative humidity, for a corresponding district and year;
2. **rainsum**: the total weekly rainfall (in mm) for a corresponding district and year;
3. **meantemperature**: 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;
4. **WindSpeed**: the weekly mean wind speed (in m/s at 2 m height) for a corresponding district and year.

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 three

3 Dashboard I: Model Calibration Process (for user at central level)

In this web-based interface (Dashboard I), the person(s) in charge 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 on some time based (typically once per year). In addition, and prior to doing any calibration, the user at central level need to assign local users at districts level with 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).

3.1 Accessing the EWARS dashboard

The central (national) level users will create a usernames and passwords to access their own dashboard (shinyapps.io) account as well as a Google drive account to link both dashboards. They need to use Dashboard I for the setting and calibration process and, users at local (district/woreda) level will need Dashboard II for declaring alarm signals (if any) and responding to alarms accordingly (prospective phase).

In the dashboard, there are additional tabs called 'Help' and 'R scripts and files'. The Help tab contains user guide for dengue outbreak, demo data and demo files for practice. Under R scripts and files tab, all necessary R-program scripts and files are provided 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.

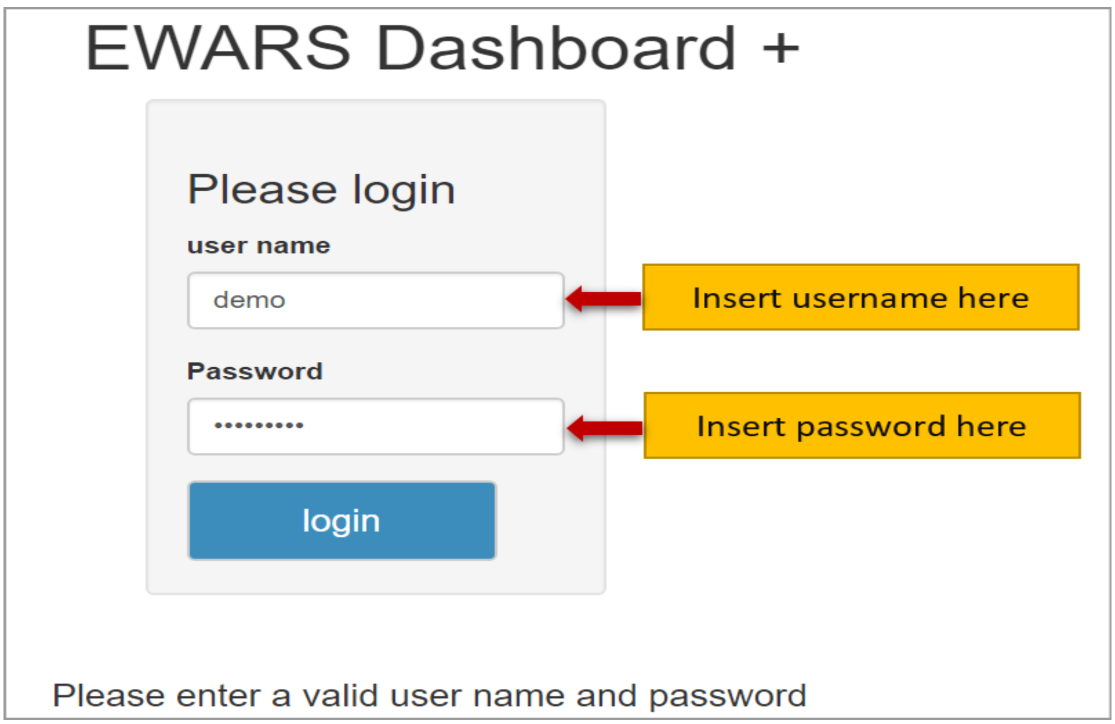
3.2 Setting password, shapefile and the surveillance dataset

3.2.1 Password settings (to allow access for local users)

Before running Dashboard I, central level users can choose their username and password to login to EWARS+ dashboard (See Figure 2 below).

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. This will be done on the dashboard database system.

Figure 2. EWARS+ login feature

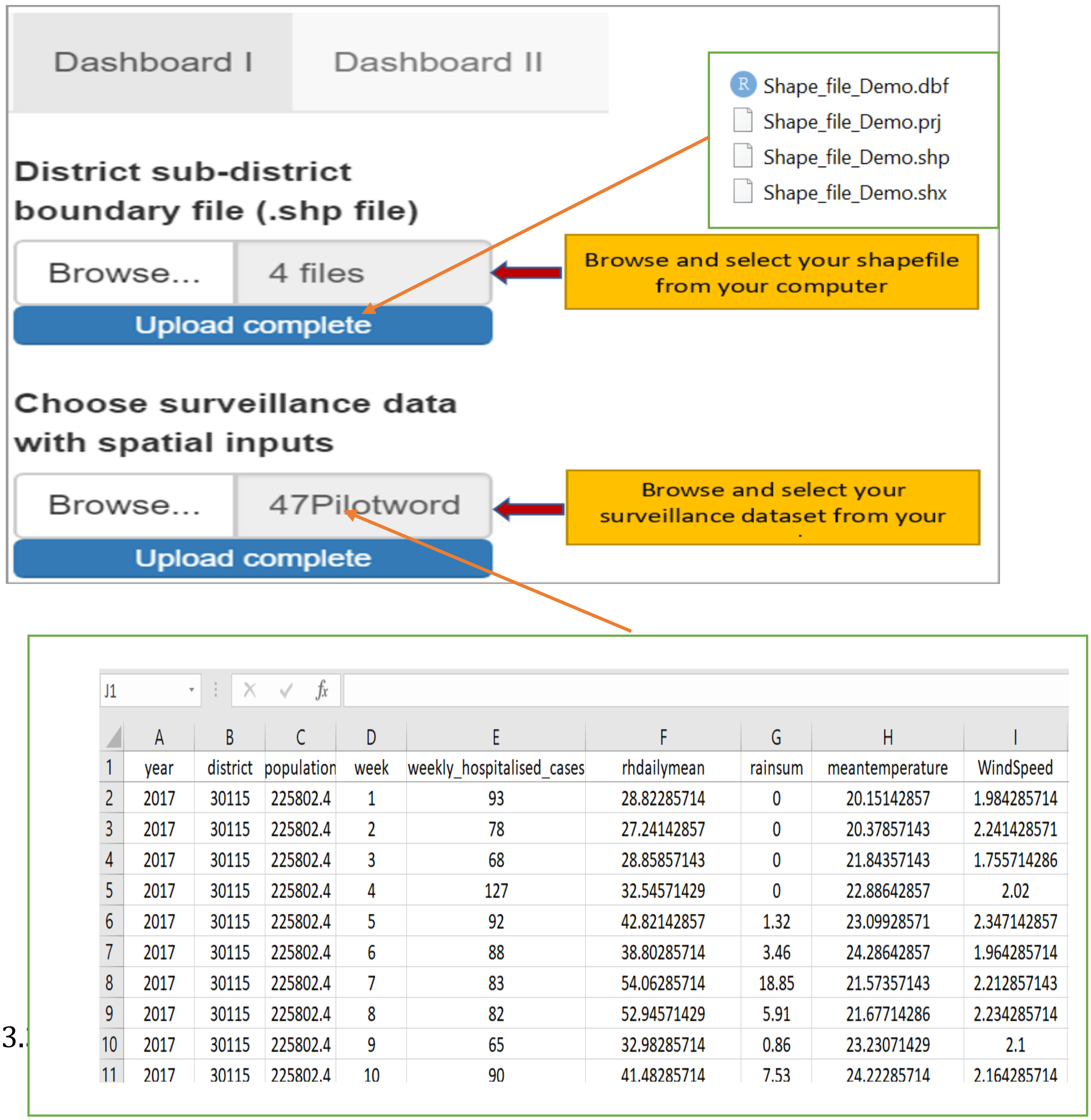


3.2.2 Uploading shapefile and surveillance dataset

Under Dashboard I, there are tabs for uploading woreda/district boundary shapefile and surveillance dataset with spatial inputs. Browse and select the shapefiles as well as the surveillance 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, however, the default data sheet in this case is 'sheet1'. As an alternative, the dataset can be saved and uploaded as CSV format.

When the data is uploaded, you will see ‘upload complete’ notification. Figure 3 below shows the data uploading tabs in EWARS+ dashboard. Model running immediately started as soon as the shapefile and surveillance dataset are uploaded. In addition, there are additional tabs for variables definition. We will discuss them in next sections.

Figure 3. Uploading your shapefile and surveillance dataset

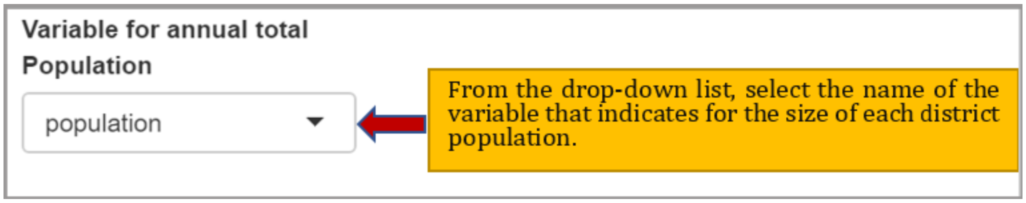


	A	B	C	D	E	F	G	H	I
1	year	district	population	week	weekly_hospitalised_cases	rhdailymean	rainsum	meantemperature	WindSpeed
2	2017	30115	225802.4	1	93	28.82285714	0	20.15142857	1.984285714
3	2017	30115	225802.4	2	78	27.24142857	0	20.37857143	2.241428571
4	2017	30115	225802.4	3	68	28.85857143	0	21.84357143	1.755714286
5	2017	30115	225802.4	4	127	32.54571429	0	22.88642857	2.02
6	2017	30115	225802.4	5	92	42.82142857	1.32	23.09928571	2.347142857
7	2017	30115	225802.4	6	88	38.80285714	3.46	24.28642857	1.964285714
8	2017	30115	225802.4	7	83	54.06285714	18.85	21.57357143	2.212857143
9	2017	30115	225802.4	8	82	52.94571429	5.91	21.67714286	2.234285714
10	2017	30115	225802.4	9	65	32.98285714	0.86	23.23071429	2.1
11	2017	30115	225802.4	10	90	41.48285714	7.53	24.22285714	2.164285714

3.

Under this tab, you need to enter the variable name which represents the annual (mid-year) total population of the corresponding woreda/district.

Figure 4. Defining the annual number of people in your districts

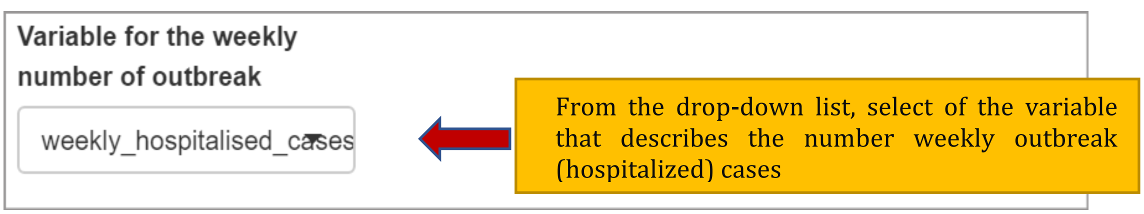


- In each district you will have a different human population size.
- Here, you must select the name of the variable that tells the analytical program the size of each district population, e.g. if your spreadsheet column is labelled population, please select 'population' from the drop-down list.
- *This corresponds to* **step 1** *in the annex*.

3.3.2 **Defining your outbreak indicator**

Under this tab, you need to enter the variable name which represents the number of weekly outbreak cases (hospitalize or probable case may be considered depending on the disease character).

Figure 5. Labelling the number of outbreak cases

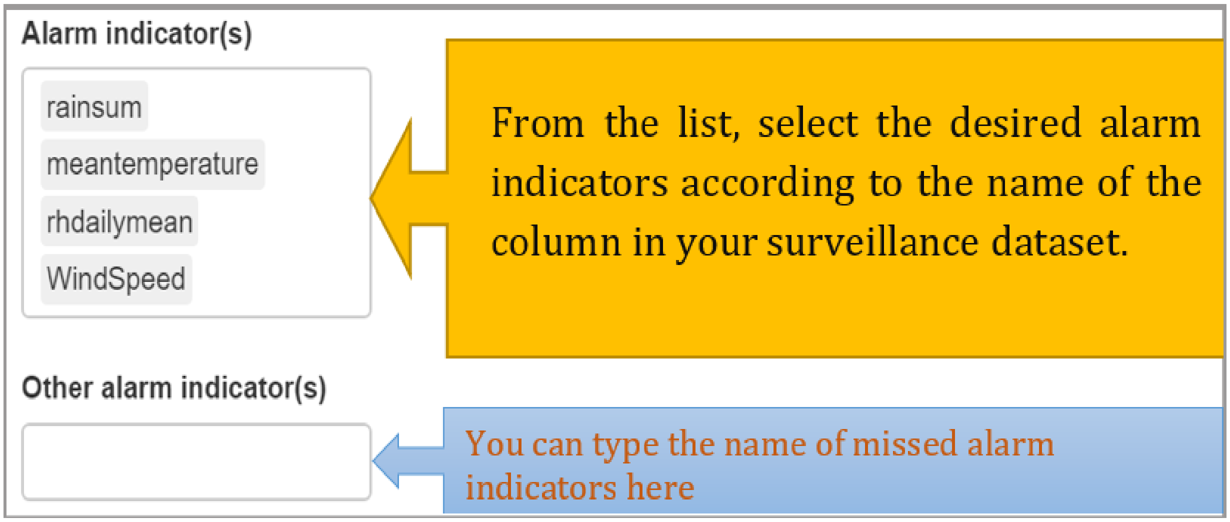


- Here you must consider what incident case data have been captured. For example, it might be the number of 'weekly_hospitalised_cases' (recommended) or weekly probable clinical cases or other possible case indicators.
- Please select in the indicated box, the column name that describes these data.
- *This corresponds to* **step 1** *in the annex*.

3.3.3 Defining your alarm indicator(s)

An **alarm indicator** are those metrological, entomological, epidemiological, and social media variables that are potentially useful for outbreak warning.

Figure 6. Defining the alarm indicator



- 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.
- Type/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.
- Missed alarm variables can be typed in the 'Other alarm indicator(s)' tab.
- 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 1** *in the annex*.

3.4 Calibrating, Running and Generating tool results in Dashboard I

Once you have completed data and file uploading, and defining the exposure and response variables, the program automatically starts the analytical process and generate results. During a running time, command status notified to show the progress. When running is finished, different tables, plots, graphs and summary will be generated. Some results are district specific, while others are for overall. We discuss those displayed results in the next sections.

Each time you run the program after making or changing your settings, updated results will be displayed, from which you will need to verify your calibration. The different graphs, plots and summary results will help you understand the current process and whether or not there are any errors or gaps in your data.

3.4.1 Calibrating your instrument

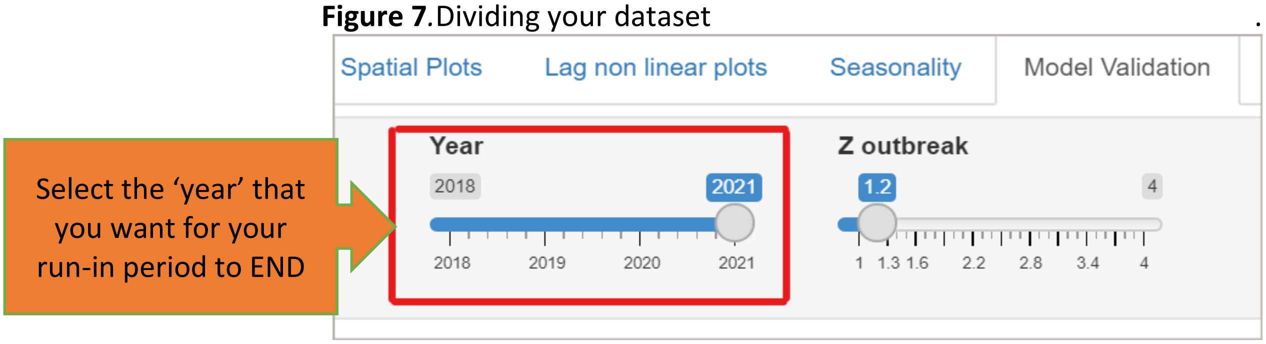
After running Dashboard I, you must adjust (calibrate) the model settings in accordance with your data and the analysis you want to augment (endemic channel multiplying value, reference values, etc.). This will define and improve the prediction of outbreaks. The illustrations in this section show you where to find these settings so that you can change them.

3.4.1.1 Dividing your dataset into run-in and validation data

The entire dataset is divided in to two parts: run-in and validation data. The run-in data is used to fit (calibrate) the model while the validation data is used to evaluate the model prediction. In this dashboard, there is a slider to choose the run-in and validation data (see *Figure **7*** **Figure *7***.Dividing your datasetbelow).

This option is found under 'Model Validation' tab in Dashboard I.

Figure 7.Dividing your dataset .



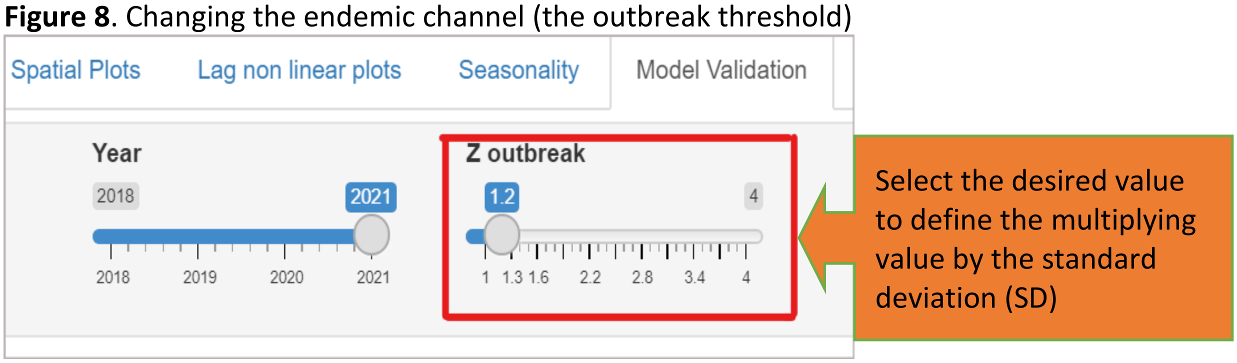
- To detect alarm indicators that can help you predict outbreaks, you need to first run a retrospective analysis of your data. For this step, the model requires that you divide your dataset into two time periods. Now you must choose the cut-off of your historic/run-in period and the evaluation period.
- 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, select the year that you want your 'run-in period' to END. That is, until selected period, it is the run-in period. The analytical tool will automatically use all data after this year as the 'validation period'. For example, select 2021 to consider the run-in period to be till end of 2020, and validation period from beginning of 2021.
- *This corresponds to* **steps 3** *and* **9** *in the annex*.

3.4.1.2 Calibrate your endemic channel (the outbreak threshold)

The slider labeled 'Z outbreak' defines the multiplying value of the standard deviation in computing the endemic channel (the outbreak threshold). Here you decide the appropriate value.

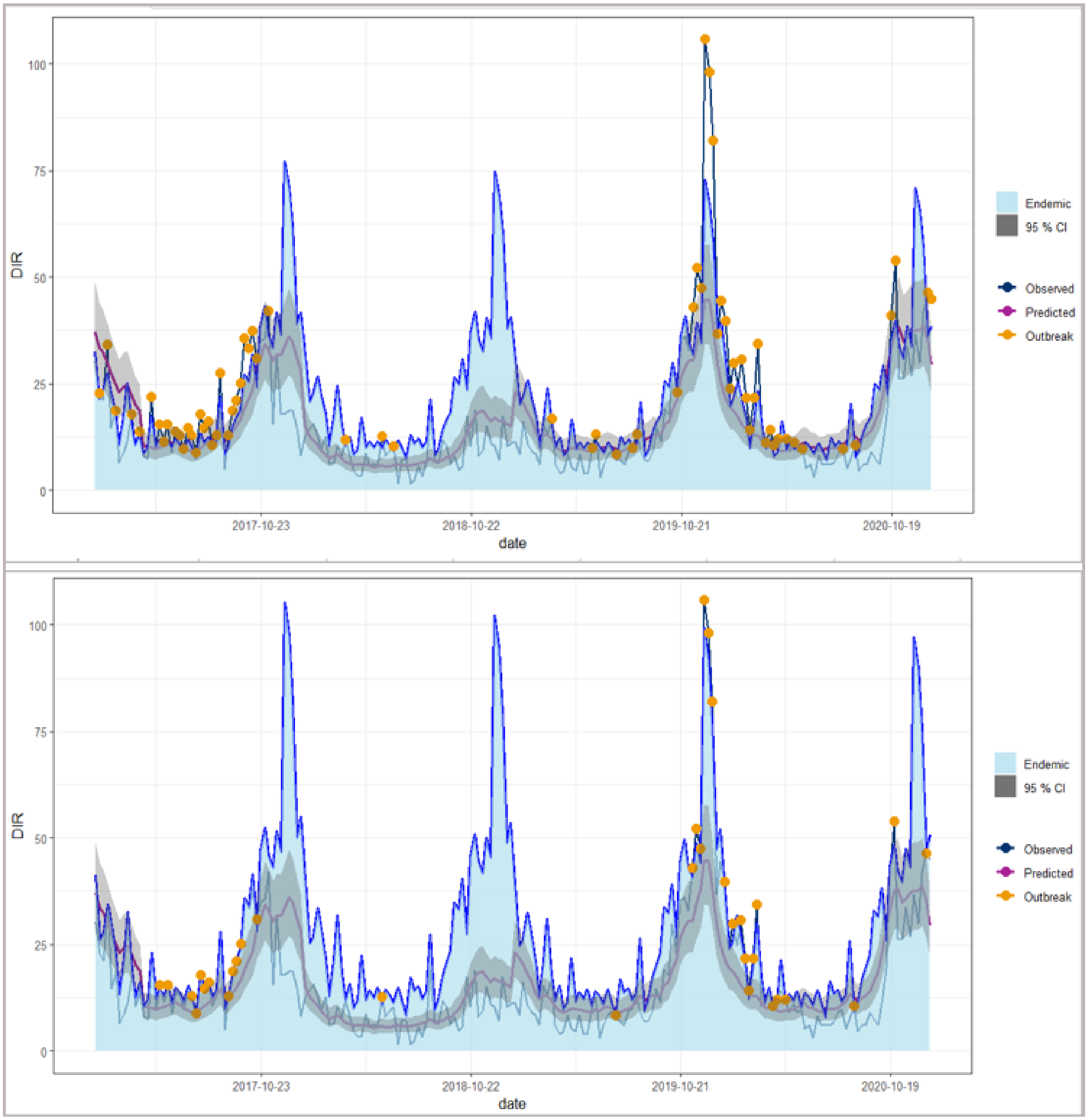
This option is found under 'Model Validation' tab in Dashboard I.

Figure 8. Changing the endemic channel (the outbreak threshold)



- 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 9***).
- Along the slider, Select 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 4** *in the annex*.

Figure 9. 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.

3.4.1.3 Lag Weeks

Biologically its takes sometime after experiencing some weather conditions (e.g. increased temperature, humidity, rainfall…etc.) to observe for example increase in vector population. This period is called the lag time and can be in *day*, *weeks*, *months* etc.

The various components of the time lag include:

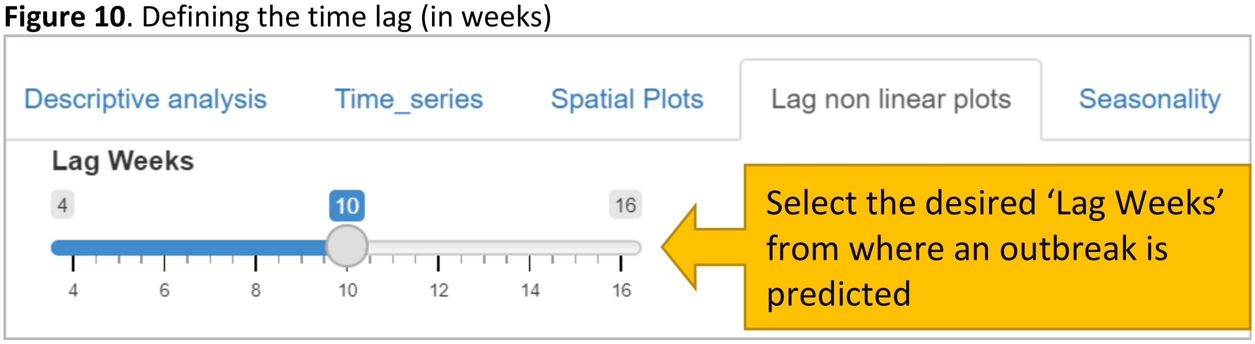
- The period for mosquito larval production to increase
- The development of the mosquito and the virus within the mosquito (extrinsic incubation period)
- The time before the first blood meal in which the mosquito transmits the virus to a human host
- The time before the appearance of clinical manifestations of dengue

The lag specifies the temporal dependency between exposure and outcome. The climate-disease risk relationship including the lag period is then modelled.

Here you select an appropriate lag week, which helps inform 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 'lag week'=10, then you would expect an outbreak to occur during week 17.

This option is found under 'Lag nonlinear plots' tab in Dashboard I.

Figure 10. Defining the time lag (in weeks)



- Select the desired 'choice of distance between current week and target week to predict an outbreak signal'.

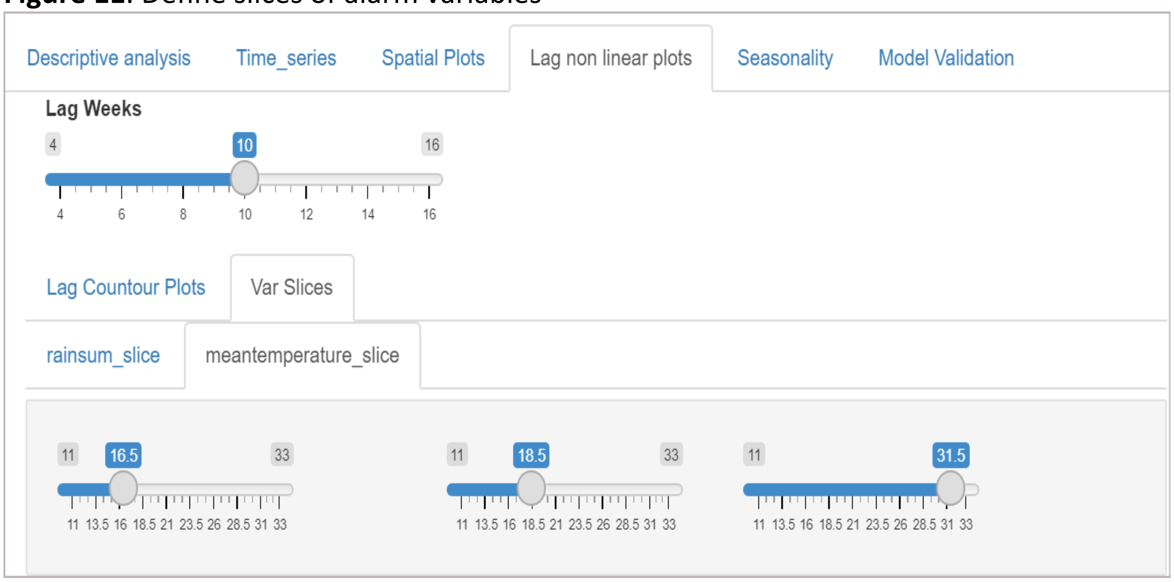
- For example, at week 7, if 'lag week'=10, this will mean that you start counting from week 17 to predict an outbreak signal. Similarly, for each of consecutive weeks, the predicted outbreak is at week of each week plus 10.
- *This corresponds to* **step 2** *in the annex.*

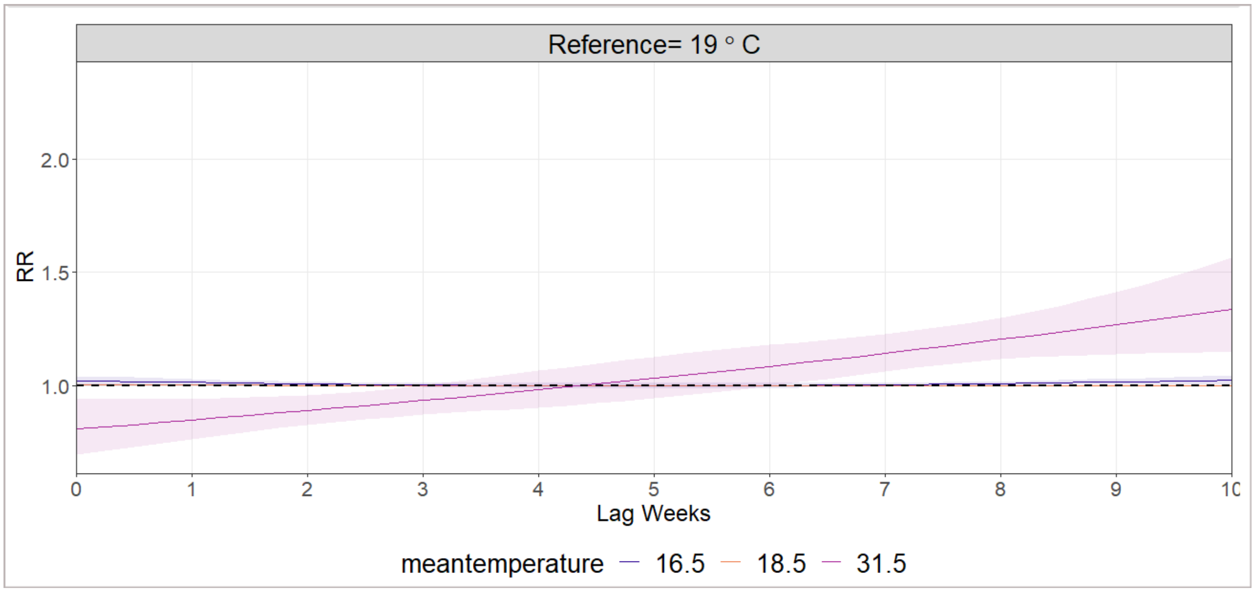
3.4.1.4 Variable slices (exposure-lag-response association)

The relative risk of disease incidence rate in different circumstances is compared to the standard (reference) to see the exposure-lag-response association. You choose different (three in this case) slices of exposure levels and you can observe the effect of the exposure level.

This option is found under 'Lag nonlinear plots ⇨ Var Slices' tab in Dashboard I.

Figure 11. Define slices of alarm variables



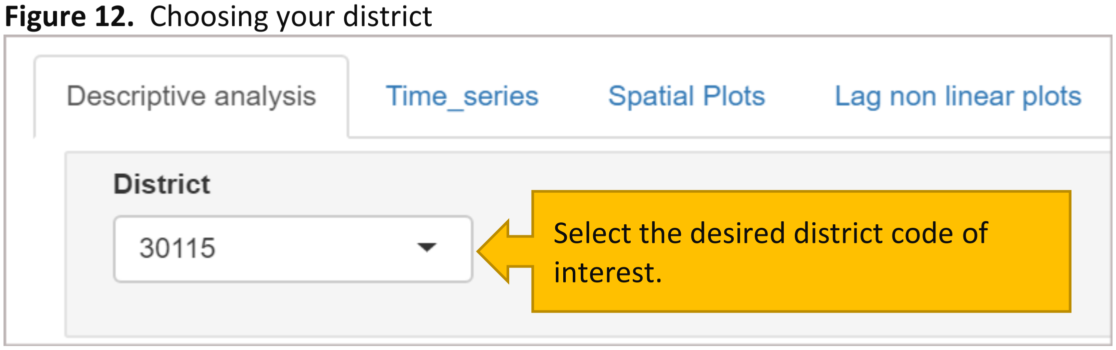


- Select the three desired exposure (alarm variables) slices. The three groups are to represent low, moderate and high levels of the exposure.
- Here reference category is the mean of the variable.
- The slicing should be done for all alarm variables.

3.4.1.5 Selecting a district to see results

This option is found under 'Descriptive analysis', 'Seasonality' and 'Model validation' tab in Dashboard I.

Figure 12. Choosing your district



- Here you can choose to display results of a specific district
- From the drop-down list, select the desired district code of interest
- The user can enter only one district code at a time

3.4.2 Understanding your calibration results

3.4.2.1 Exploratory analysis

Descriptive summaries of uploaded data are described here. Tabular and graphical presentations methods are used. For the tabular presentation of the data, the following summaries are included;

- minimum, maximum, mean and median
- 25th and 75th percentiles
- Percent of missing data by year. It indicates and helps user to spot missing or inconsistencies records
- For example, ***Figure 13*** show summaries for hospitalize case and rain sum.

Note that, the explanatory analysis includes for;

- all exposure (alarm variables)
- district level population
- response variable

Figure 13. Descriptive summaries of variables

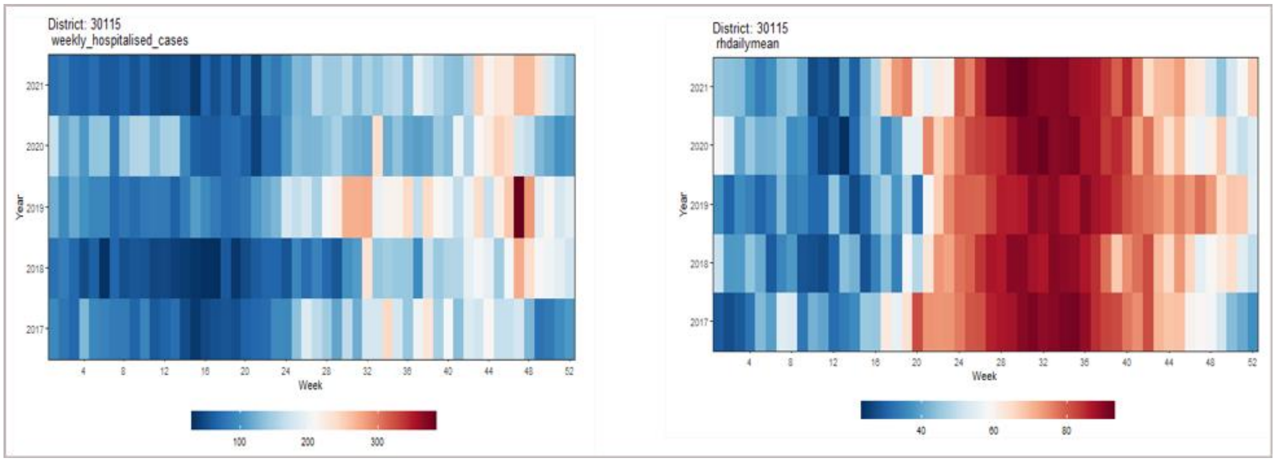
District 30115

Year	Min	Max	Mean	Median	25th Percentile	75th Percentile	% Missing
weekly_hospitalised_cases							
2017	36.0	243.0	117.9	100.5	72.5	165.2	0%
2018	30.0	273.0	106.7	79.5	54.8	153.2	0%
2019	63.0	384.0	160.1	163.5	90.8	211.5	0%
2020	42.0	248.0	128.5	126.5	105.8	149.2	0%
2021	36.0	263.0	121.6	128.5	65.0	158.2	0%
rainsum							
2017	0.0	153.2	26.6	11.2	1.4	39.7	0%
2018	0.0	105.3	20.6	7.2	0.5	37.0	0%
2019	0.0	89.3	22.5	8.1	0.7	40.8	0%
2020	0.0	172.0	32.5	6.9	0.4	52.7	0%
2021	0.0	274.4	37.8	5.8	0.8	46.8	0%

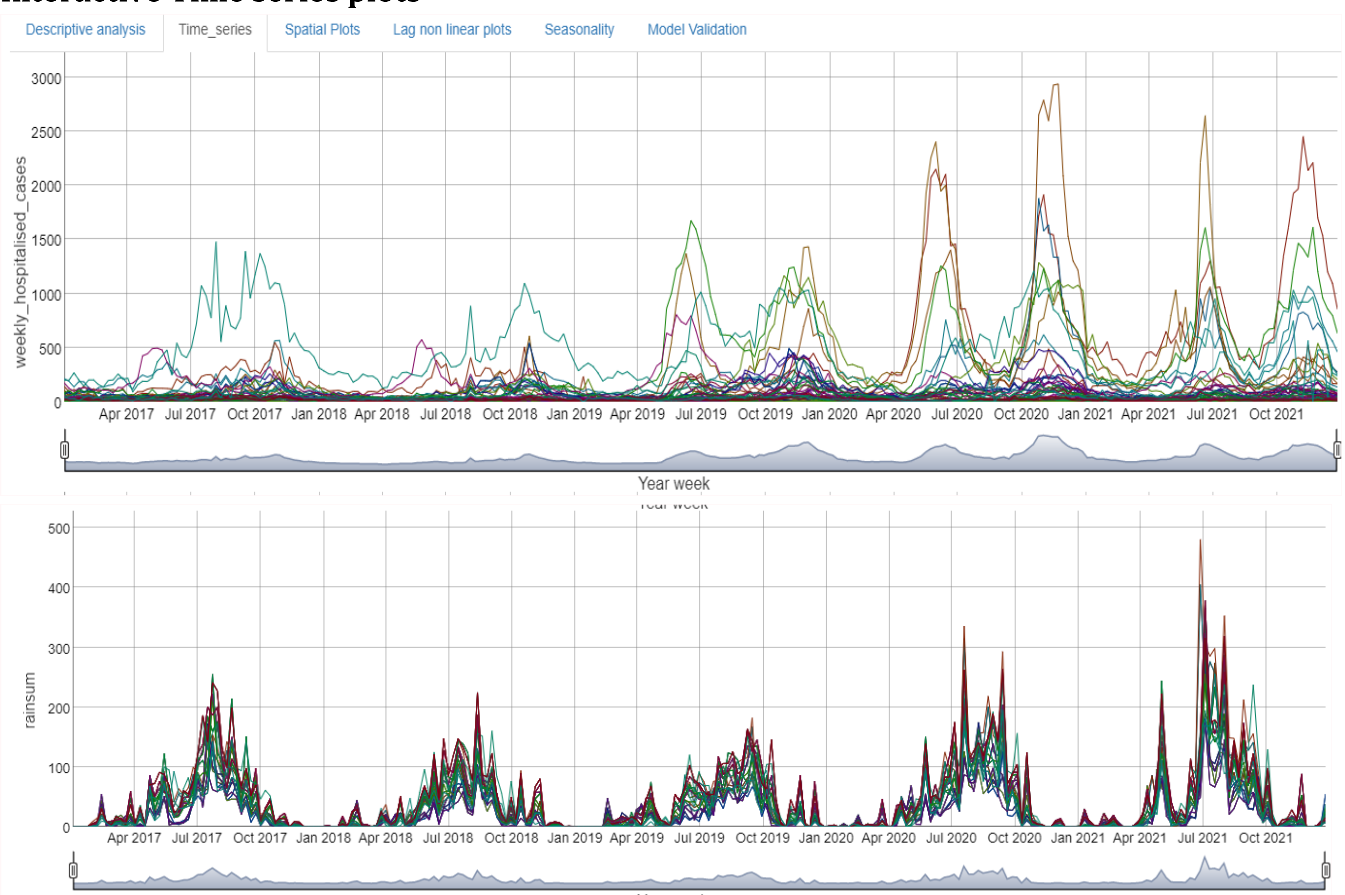
Heatmap visualization:

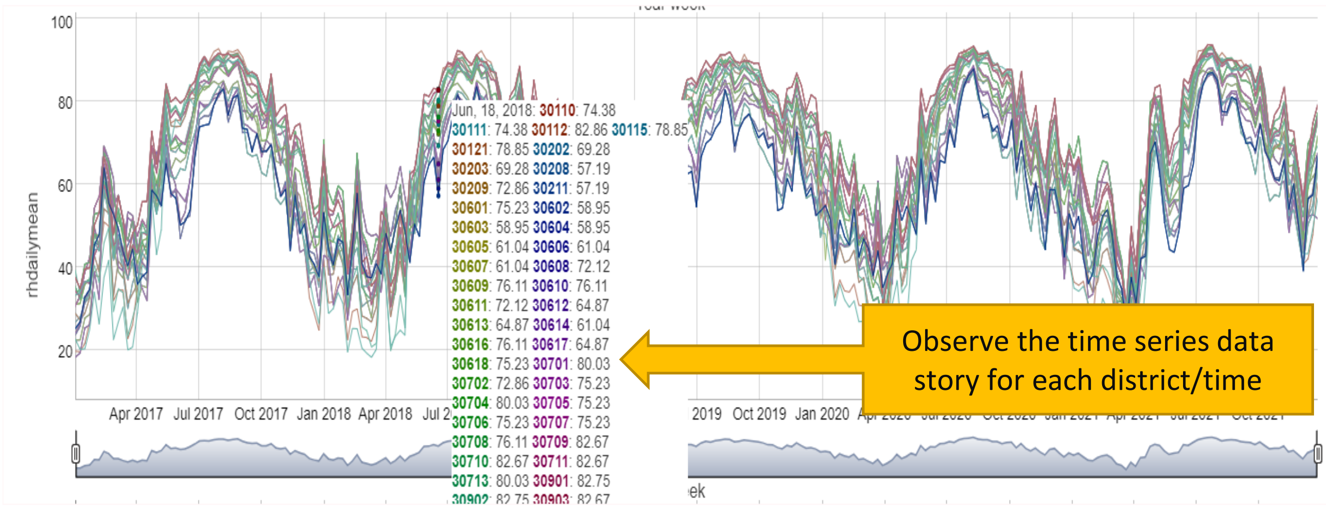
- This visualizes magnitude over the epidemiological weeks
- For one district, ***Figure 14*** shows magnitude of case (hospitalized case) and alarm indicators, e.g mean humidity

Figure 14. Heatmap visualization



Interactive Time series plots





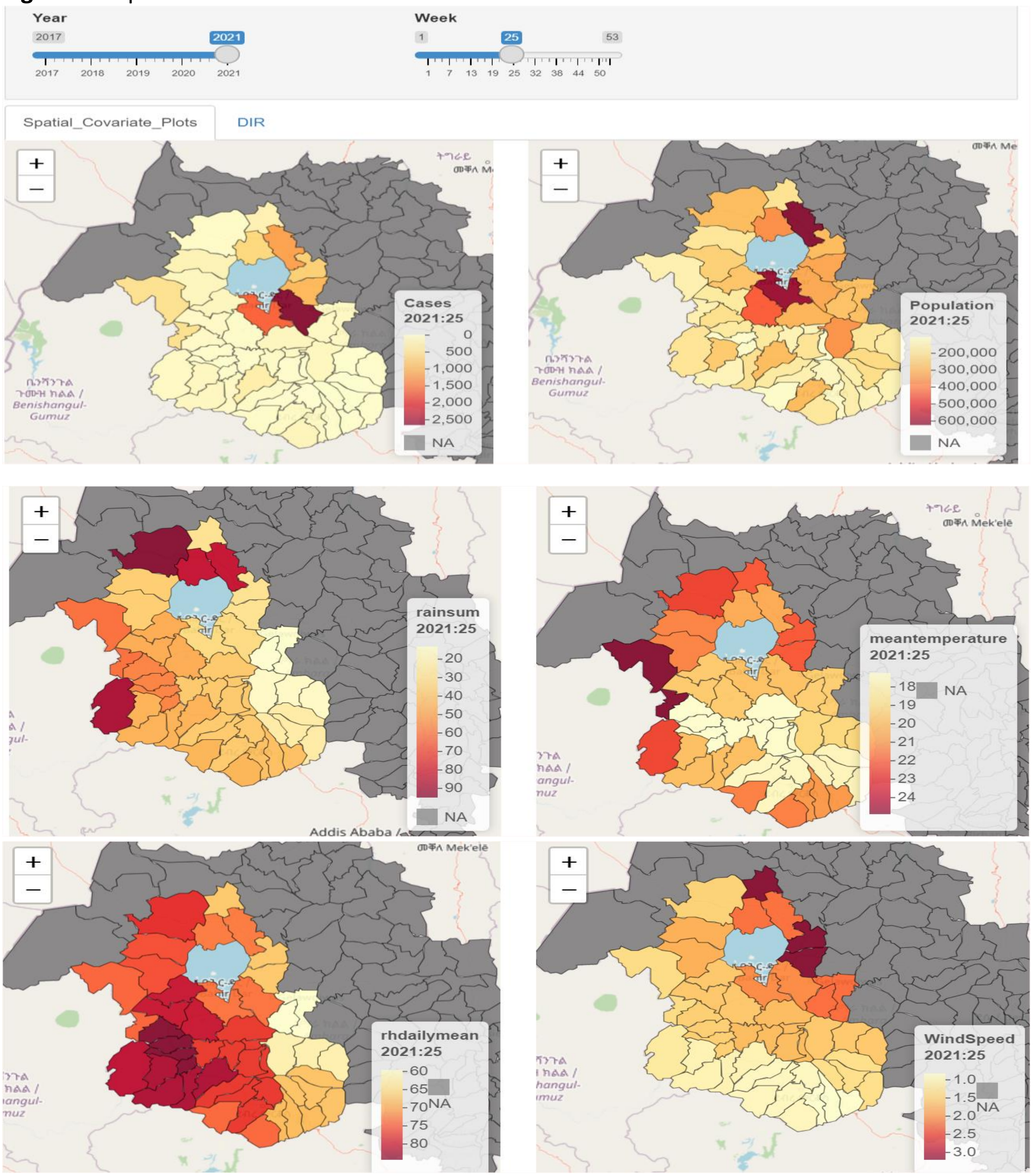
- Time series plot provides trend of data for one district (woreda) which informs profile of the data summarized i.e., the magnitude of measures being summarized and its distribution
- The interactive time series plot provides comparison of data (of each variable) across all districts. Indicate how different disease trends compares in difference districts (woredas) within the study area.

3.4.2.2 Graph: Risk Mapping Plot

This shows spatial description based on variations in parameters and alarm indicators. Note that, this is yet not an analytical step. **Figure 15** shows the different spatial plots.

This option is found under 'Spatial Plots' tab in Dashboard I.

Figure 15. Spatial Plots



- The 'spatial plots' show details of the pattern of disease cases, the population in each district and, the intensity of the alarm indicators in each district.
- This step illustrates 'hot spots' as in
  i) higher number of cases are observed in specific district and,
  ii) when alarm indicators are showing variations (for a range) making the risk of vector breeding increased/decreased!

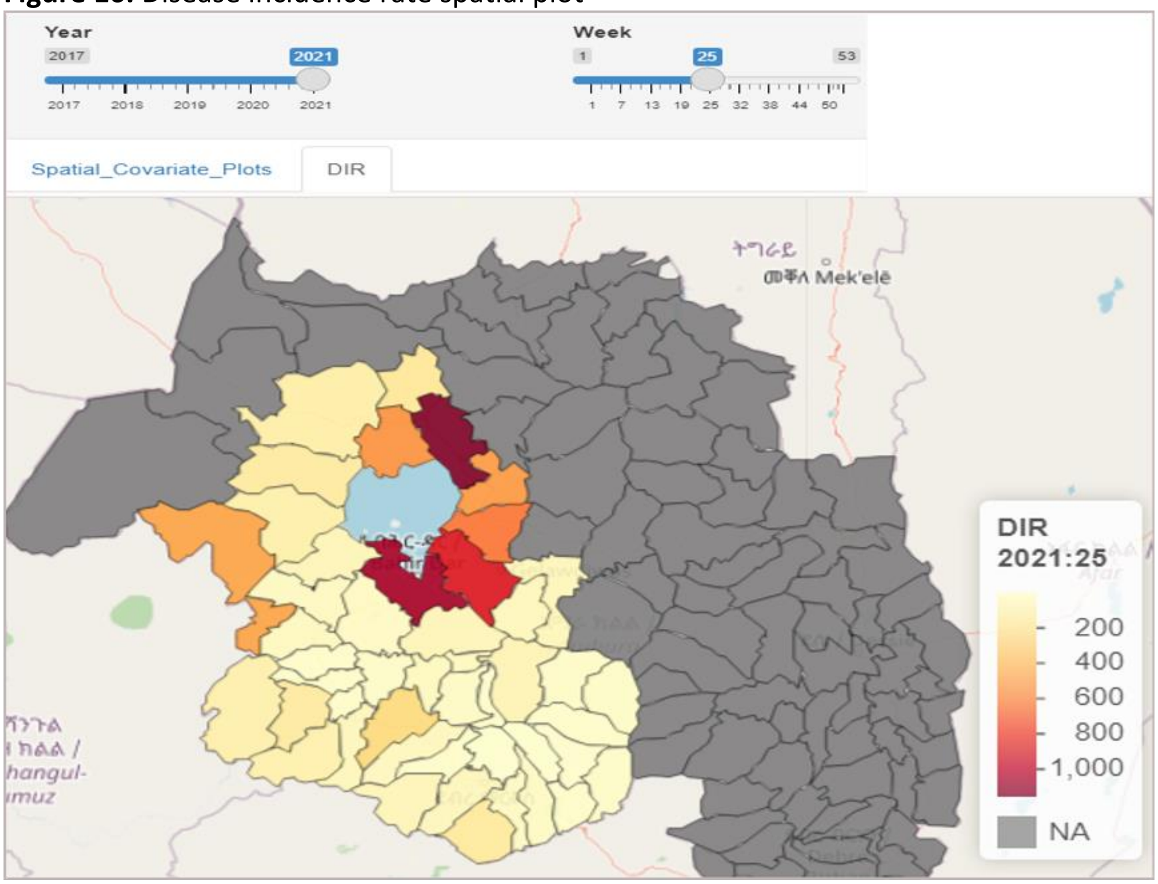
Disease Incidence Rate (DIR)

This step provides analytical spatial statistics result.

This option is found under 'Spatial Plots ➡ DIR ' tab in Dashboard I.

- At this step, the model processes all covariates input in the model i.e. alarm indicators, population size of each district, number of cases, the magnitude of cases in neighboring districts… etc
- The generated 'risk map' indicates for 'hot spots' based on the model analysis
- DIR=malaria incidence rate per 10000
- Dark red color refers to higher incidence rate of the disease (in this case malaria)

Figure 16. Disease incidence rate spatial plot



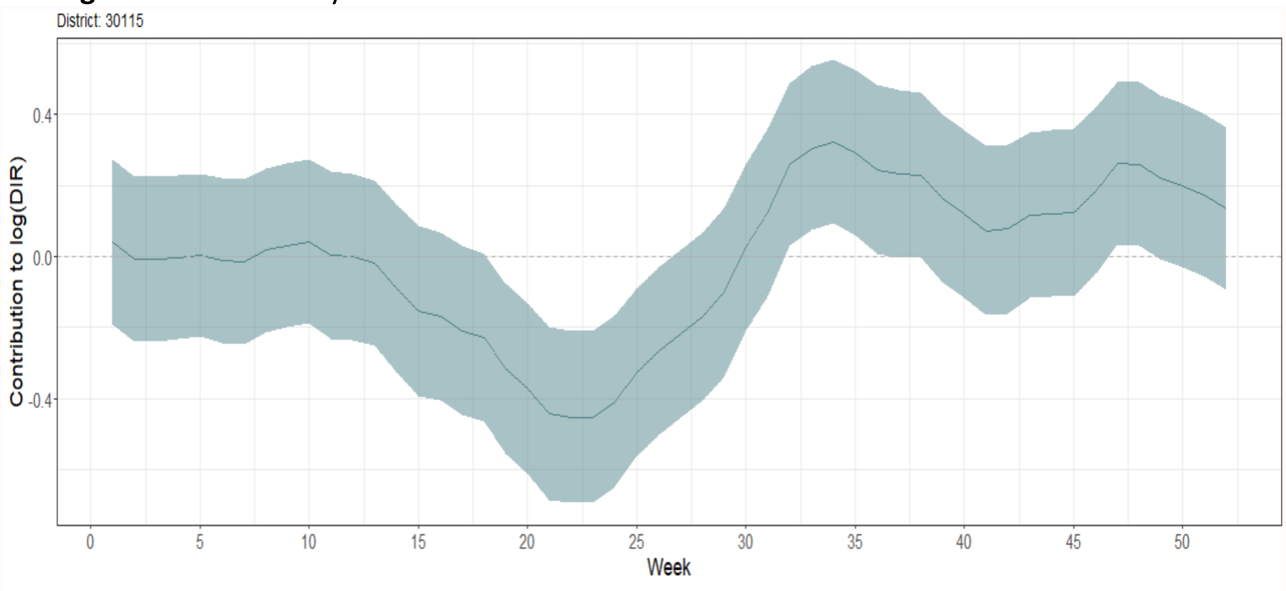
- Note: Gray shaded areas indicate no data (no risk mapping).

3.4.2.3 Graph: Seasonality plot

This graph shows the seasonality or the unmeasured/ unknown annual variability is estimated with 95% CI. E.g for district 30115, week 30 and above has positive contribution for log DIR.

This option is found under 'Seasonality plot' tab in Dashboard I.

Figure 17. Seasonality Plot



3.4.2.4 Graph: Model Validation (Run-in period)

Graph of the run-in data shows the endemic channel, observed and predicted cases of run-in period. Interactive plot is also provided to see details (***Figure 18***). The model produces out-of-sample[1] predicted probabilities of exceeding the outbreak threshold:

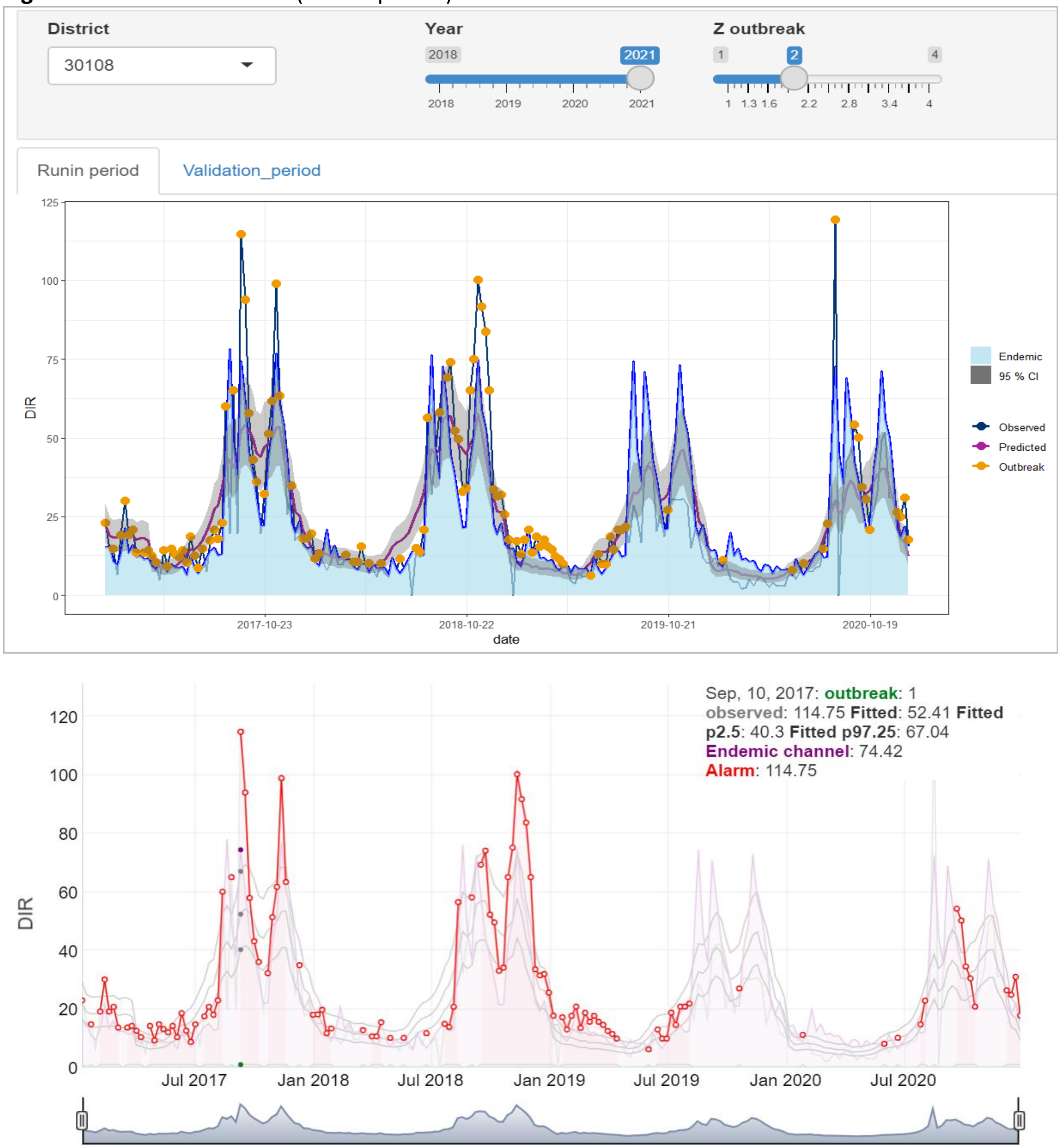
1. Weekly predictive probabilities of cases (incidence rate) for each district are first computed from alarm indicators parameters;
2. This predicted disease probability is compared against thresholds (i.e. case number based on endemic channel);
3. The probability of exceeding the endemic channel (threshold) is calculated. Then follows with triggering or averting of Alarm signals.

---

[1] *Out-of-sample= forecasting for weeks that was not part of the data sample*

4. Each week in the validation period was predicted. This step uses for ‘model validation’ in the next sub-section.

Figure 18. Model validation (Run-in period)



This option is found under ‘Model validation ➡ Runin Period’ tab in Dashboard I.

The model calibration provides (for each district) different information based on the training ('Run-in Period') data:

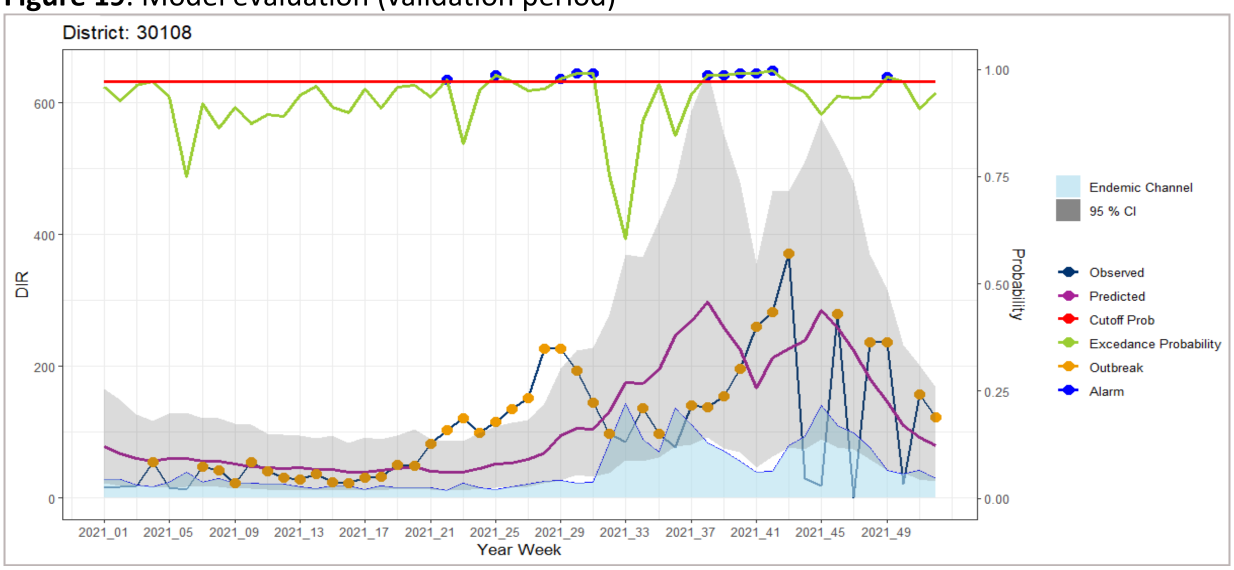
- Here you can learn about the data used for calibration (from the X-axis). In this case the data used is from 2017 to 2020.
- The DIR throughout the 'Run-in Period' (y-axis), and
- The size of your 'Endemic channel' defined by the given 'Z outbreak' value. Note that, Z outbreak is the multiplier value of the SD provided by the user to vary the breadth of the endemic channel, i.e Endemic channel= (Z*SD) + moving average. E.g a value of Z=2 would increase the breadth two times the expected normal range of the number of disease cases. The moving average is the mean number of disease cases within the expected normal or seasonal range.
- Confirmed cases that exceed the endemic channel and trigger an 'Outbreak period' are also presented.

3.4.2.5 Graph: Model Validation (Validation period) Plot

This is the stage where model performs according to the given settings/calibrations in the run-in data is evaluated. Evaluation of the model performance is based on how the predicted incidence rate (Step 4, above) exceeds the threshold (endemic channel) compared with the observed number of cases.

This option is found under 'Model validation ➡ Validation Period' tab in Dashboard I.

Figure 19. Model evaluation (validation period)



- This graph summarizes the second half of your data, which is the 'Validation 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; 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 (cutoff probability)' and the 'Exceedance Probability (or the probability that predicted case exceeds the endemic channel)'.
- When the exceedance probability exceeds the cutoff probability during a particular epidemiological week, the alarm signal (blue dots) is triggered, indicating an upcoming outbreak.

3.4.2.6 Summary of model evaluation (Sensitivity/Specificity)

Different metrics (test of validity) are used to evaluate the model performance.

- ROC[2] is calculated to determine optimal thresholds to issue alarm signals for the binary events of exceeding the moving outbreak threshold (exceeds/ not exceeds).
- The model is then said to be able to (or not) correctly distinguish between outbreak and non-outbreak weeks.

This option is found under 'Model validation ➡ Validation Period ➡ Sensitivity/Specificity' tab in Dashboard I.

Figure 20. Sensitivity/specificity

var	val	CI_Lower	CI_Upper
Cutoff probability	0.594	NA	NA
Area under the Curve (AUC)	0.820	0.706	0.934
Accuracy	0.788	0.782	0.795
Sensitivity	0.750	0.590	0.910
Specificity	0.833	0.684	0.982
Positive Predictive Value (PPV)	0.840	0.696	0.984
Negative Predictive Value (NPV)	0.741	0.575	0.906

[2] *ROC: Receiver Operating Characteristics*

Definition

1. **Sensitivity**: the percentage of outbreaks correctly predicted by alarms. This should be as close to 100% as possible. Minimum value of 50% is acceptable.
2. **Specificity:** the proportion of events that were predicted to occur but did not occur.
3. **Positive predictive value (PPV)**: the percentage of alarms that correctly predicted outbreaks. This should be as close to 100% as possible. Minimum of 50% is acceptable.
4. **Negative predictive value (NPV):** proportion of correct false predictions.
5. To increase the number of correct alarms and decrease the number of false alarms, we suggest that you alter the z-outbreak value as desired, then, if necessary, also alter other parameters (if any) as desired.

**** High sensitivity and PPV indicates the most relevant and best prediction model.**

Operational perspective

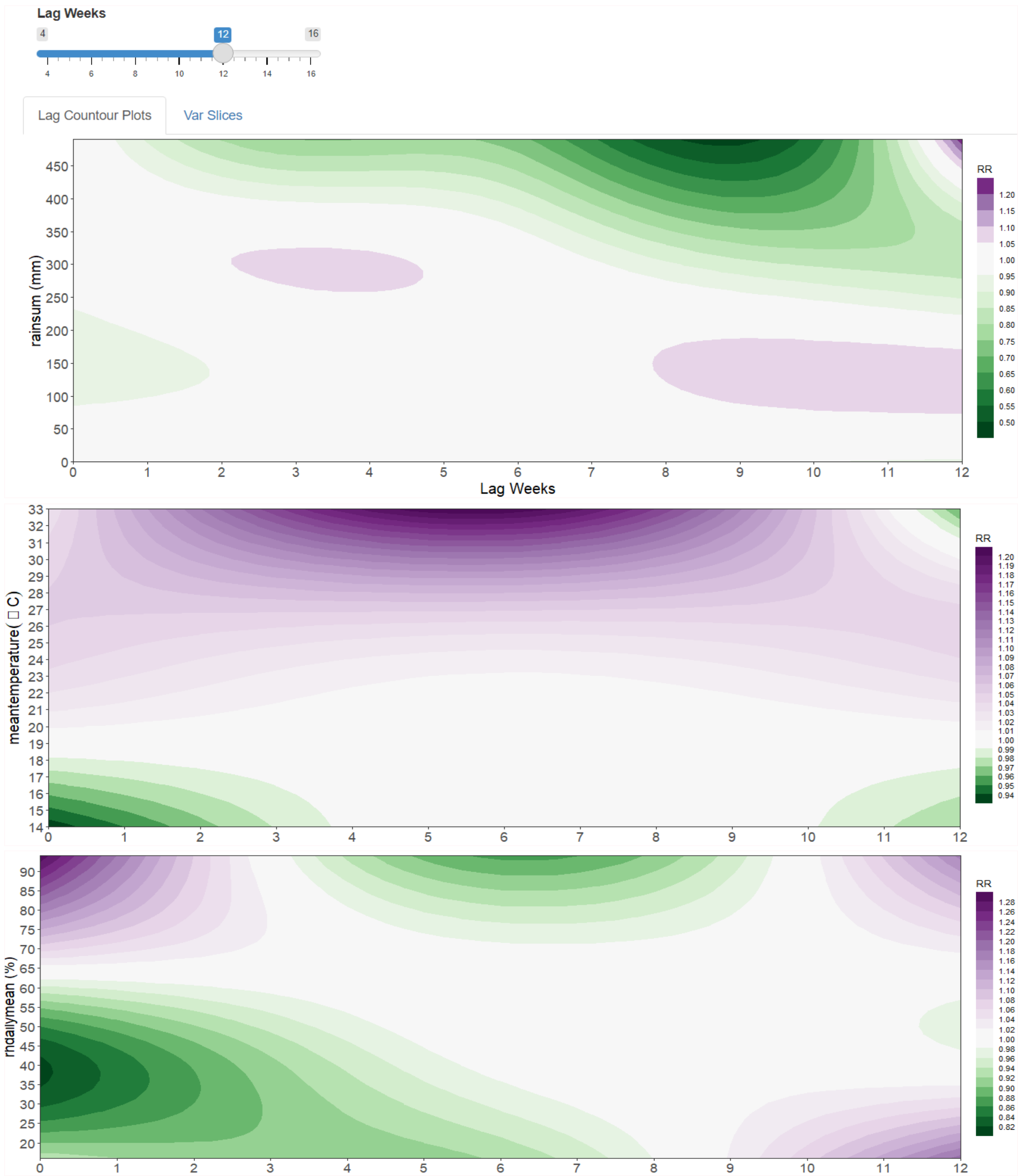
- If sensitivity is, for example 60%, then for every 10 outbreaks, the early warning system will detect 6 outbreaks correctly. It will miss 4 outbreaks.
- If PPV is, for example 70%, then 7 out of 10 alarms will be correct, 3 out of 10 alarms will be false. That means that 3 times out of 10, resources will be incorrectly mobilized, i.e, wasted.

3.4.2.7 Exposure-Lag-Response association

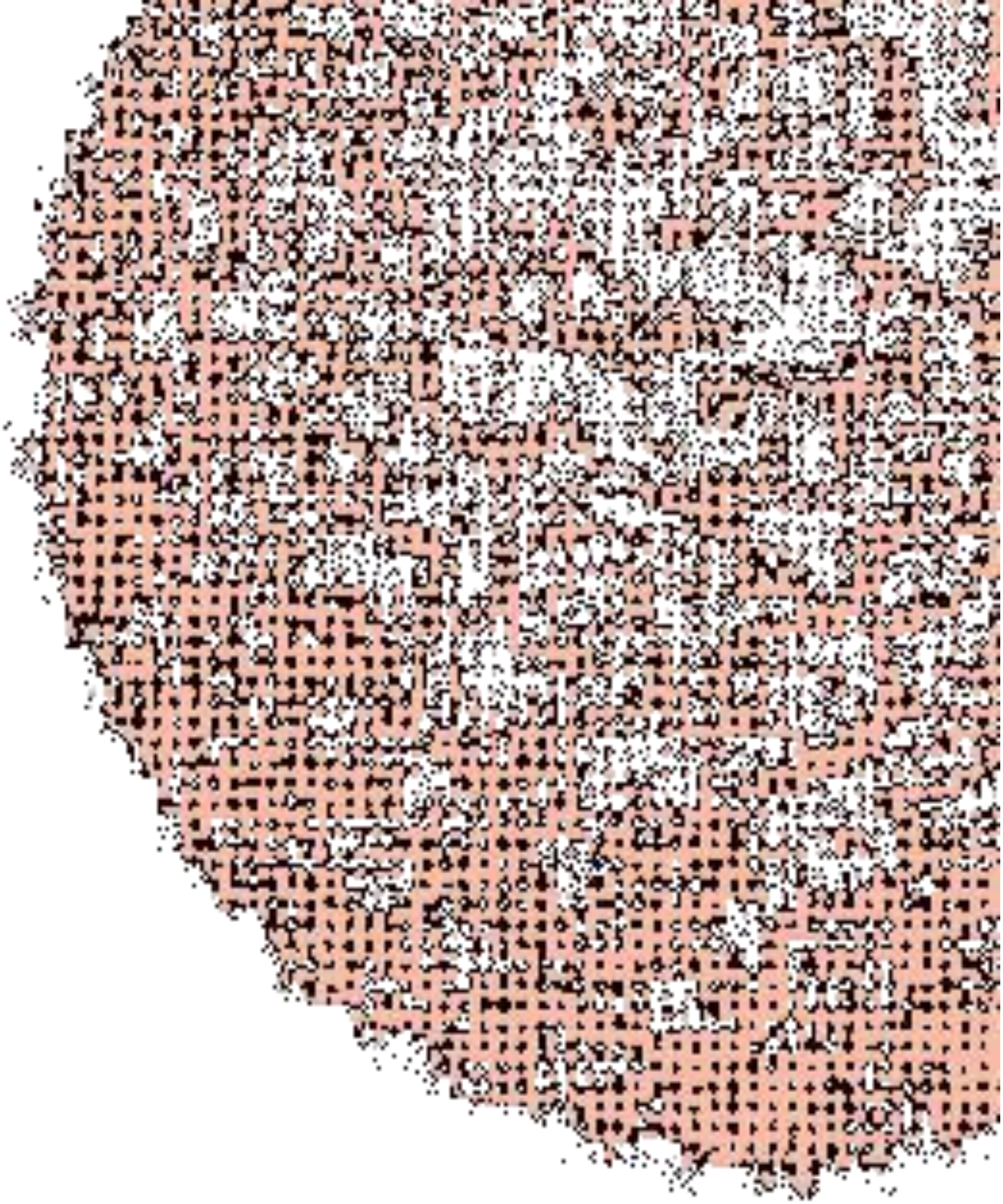
The (marginal) effect of alarm indicators on relative risk (RR) of disease are presented as contour plot.

This option is found under 'Model validation ➡ Lag nonlinear plots' tab in Dashboard I.

Figure 21. Exposure- Lag - Response Plot



- The dark purple associated with high RR of disease (in this case malaria case) while dark green associated with low.
- E.g Malaria RR is greatest after 3 to 9 week of high mean temperature (> 30 °C). On the other hand, its lowest for 0 to 2 weeks lag.



Chapter Four

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

In this chapter, we discuss the elements of dashboard II (prospective analyses phase). In the following sections, we discuss how to build the prospective early warning, alert and response system using results from the 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/woreda. Now we need to apply the derived results (calibrated model) 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 results.

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

4.1 Prospective data entry of weekly alarm indicators

4.1.1 Uploading prospective data

Seasonal climate forecasts offer the possibility of forecasting disease epidemics many months in advance (a clear advantage for disease control planners) - despite issues related to accuracy and accessibility of forecast data.

EWARS+ can process prospective climate information (on weekly basis) as well as climate forecast data (uploading forecast data of weeks or months ahead). That is, by entering the forecasted 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 (at certain prediction distance). EWARS+ keep a fixed 12 weeks’ time-lag (prediction distance), but can be adjusted as indicated above in section 3.4.2.7.

The prospective data can be organized in excel and uploaded to the tool in dashboard II. The format is similar with the retrospective dataset.

This option is found under ‘Upload prospective data’ tab in Dashboard II.

Figure 22. Prospective data uploading tab

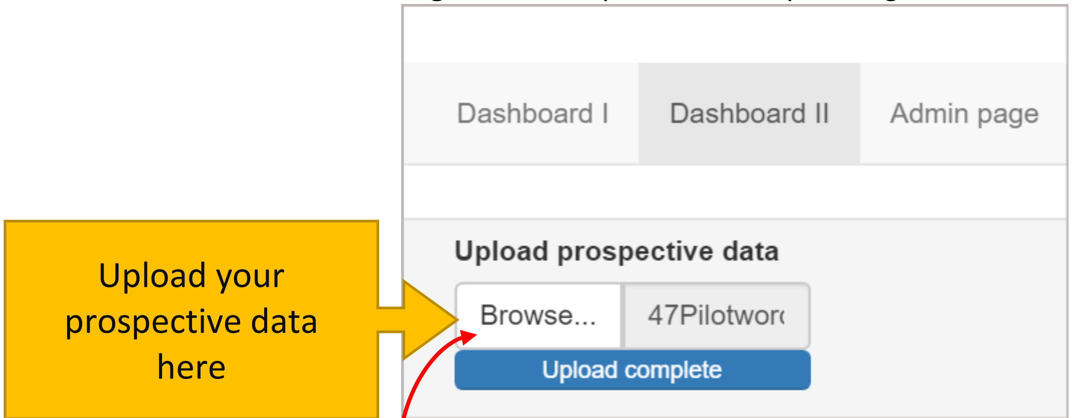


Figure 23. Prospective data

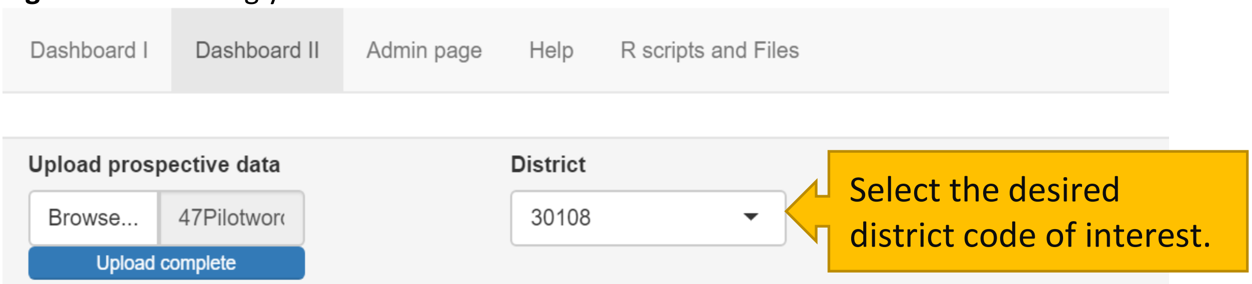
	A	B	C	D	E	F	G	H	I
1	year	district	population	week	ly_hospitalised_cases	rhdailymean	rainsum	meantemperature	WindSpeed
2	2022	30115	257472.5126	1	98	61.5	10.44	19.66071429	1.798571429
3	2022	30115	257472.5126	2	100	51.57857143	0.37	21.79	1.84
4	2022	30115	257472.5126	3	105	57.25857143	0.99	20.27857143	2.007142857
5	2022	30115	257472.5126	4	105	47.52571429	0	21.38071429	1.901428571
6	2022	30115	257472.5126	5	93	45.64142857	0.16	21.90142857	2.532857143
7	2022	30115	257472.5126	6	70	45.23285714	0.56	22.16	2.36
8	2022	30115	257472.5126	7	65	34.11571429	0.09	23.25714286	2.355714286
9	2022	30115	257472.5126	8	47	40.01714286	2.02	23.235	2.164285714
10	2022	30115	257472.5126	9	55	36.39428571	0.31	25.18357143	2.168571429
11	2022	30115	257472.5126	10	70	31.08714286	0.44	26.68571429	2.377142857
12	2022	30115	257472.5126	11	54	40.87571429	5.34	24.755	2.382857143
13	2022	30115	257472.5126	12	49	29.83857143	0.07	25.12285714	2.785714286
14	2022	30115	257472.5126	13	43	31.01	0	26.43071429	2.624285714
15	2022	30115	257472.5126	14	41	33.46428571	3.52	26.71928571	2.587142857
16	2022	30115	257472.5126	15	45	46.43571429	74.2	25.385	2.165714286

- When the data is uploaded, 'upload complete' notification will pop-up
- Once the upload completed, the program automatically starts the analytical process and generate results. However, the result generation takes couple of minutes depending on the computer/server capacity.
- Consideration: make sure that the column (variable) name here in the prospective data is same with that of the retrospective counterpart.

4.1.2 Selecting a district to see prospective analysis results

This option is found under 'District' tab in Dashboard II.

Figure 24. Choosing your district

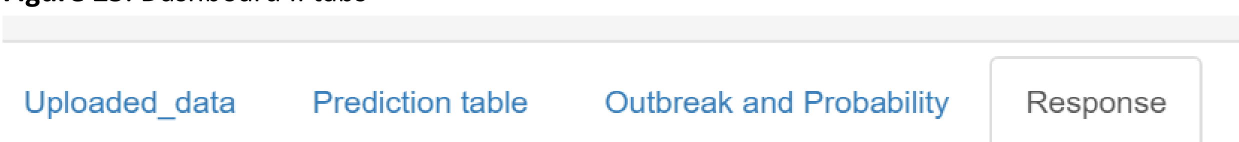


- Here you can choose to display results of a specific district
- From the drop-down list, select the desired district code of interest
- The user can enter only one district code at a time

4.2 Understanding your prospective analyses in dashboard II

In the EWARS+, there are four tabs with different analyses result. Next, we discuss each tab result.

Figure 25. Dashboard II tabs

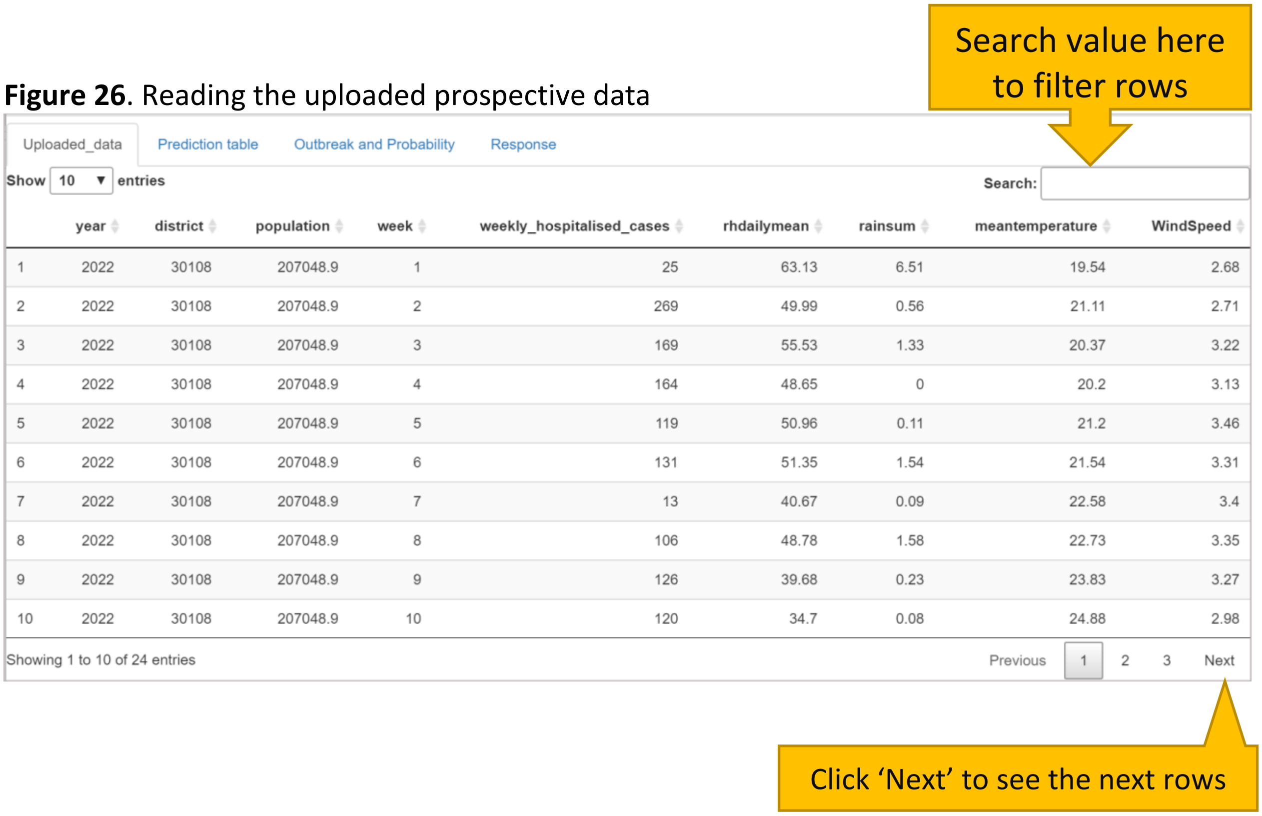


4.2.1 Browse through the prospective data

For the purpose of providing detailed look up (or checking for error) of entered prospective data, you can browse through the prospective data for your district. Here you can view part of the data in a single page. Click next to see the remaining rows of the data. To filter rows with a particular value of response or average alarm indicators, you can search the value on the search bar at the top-right corner of the table.

This option is found under 'Uploaded data' tab in Dashboard II.

Figure 26. Reading the uploaded prospective data



	year	district	population	week	weekly_hospitalised_cases	rhdailymean	rainsum	meantemperature	WindSpeed
1	2022	30108	207048.9	1	25	63.13	6.51	19.54	2.68
2	2022	30108	207048.9	2	269	49.99	0.56	21.11	2.71
3	2022	30108	207048.9	3	169	55.53	1.33	20.37	3.22
4	2022	30108	207048.9	4	164	48.65	0	20.2	3.13
5	2022	30108	207048.9	5	119	50.96	0.11	21.2	3.46
6	2022	30108	207048.9	6	131	51.35	1.54	21.54	3.31
7	2022	30108	207048.9	7	13	40.67	0.09	22.58	3.4
8	2022	30108	207048.9	8	106	48.78	1.58	22.73	3.35
9	2022	30108	207048.9	9	126	39.68	0.23	23.83	3.27
10	2022	30108	207048.9	10	120	34.7	0.08	24.88	2.98

4.2.2 **Prediction table**

Under the 'Prediction table' tab, you can see a summary of prospective weeks with respect to all the 'Endemic channel', 'Predicted case (with 95% CI)', 'Outbreak case', 'Alarm threshold', 'Outbreak probability, and 'Alarm signal'.

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 forthcoming outbreak (alarm signal).

This option is found under 'Prediction table' tab in Dashboard II.

Figure 27. Prediction table

Uploaded_data | Prediction table | Outbreak and Probability | Response

Show 10 entries

	district	year	week	pop	endemic_chanel	predicted	p25	p975	outbreak
1	30202	2022	1	297646.537397737	40.2774	53.796	13.4304	138.4444	18.0738
2	30202	2022	2	297646.537397737	17.9097	55.5908	12.7668	154.9069	18.6768
3	30202	2022	3	297646.537397737	16.6665	26.771	6.0474	72.5861	8.9942
4	30202	2022	4	297646.537397737	19.2921	49.1563	12.0865	128.0126	16.515
5	30202	2022	5	297646.537397737	18.6342	50.0668	12.4309	125.35	16.8209
6	30202	2022	6	297646.537397737	18.5273	37.8993	8.7352	97.7754	12.733
7	30202	2022	7	297646.537397737	13.3659	39.0735	9.0628	97.4478	13.1275
8	30202	2022	8	297646.537397737	11.5908	26.8073	6.3834	70.8895	9.0064
9	30202	2022	9	297646.537397737	10.3917	29.1077	6.7194	72.2501	9.7793
10	30202	2022	10	297646.537397737	13.3086	30.4828	7.0553	78.9611	10.2413

Showing 1 to 10 of 52 entries

Previous

Search:

alarm_threshold	outbreak_probability	alarm	alarm_signal	response_cat
0.771	0.612	0	0	0.5
0.771	0.9603	1	1	0.5
0.771	0.7309	0	0	0.5
0.771	0.9223	1	1	0.5
0.771	0.9326	1	1	1
0.771	0.8609	1	1	1.5
0.771	0.949	1	1	2
0.771	0.8832	1	1	2
0.771	0.933	1	1	2
0.771	0.889	1	1	2

1 2 3 4 5 6 Next

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

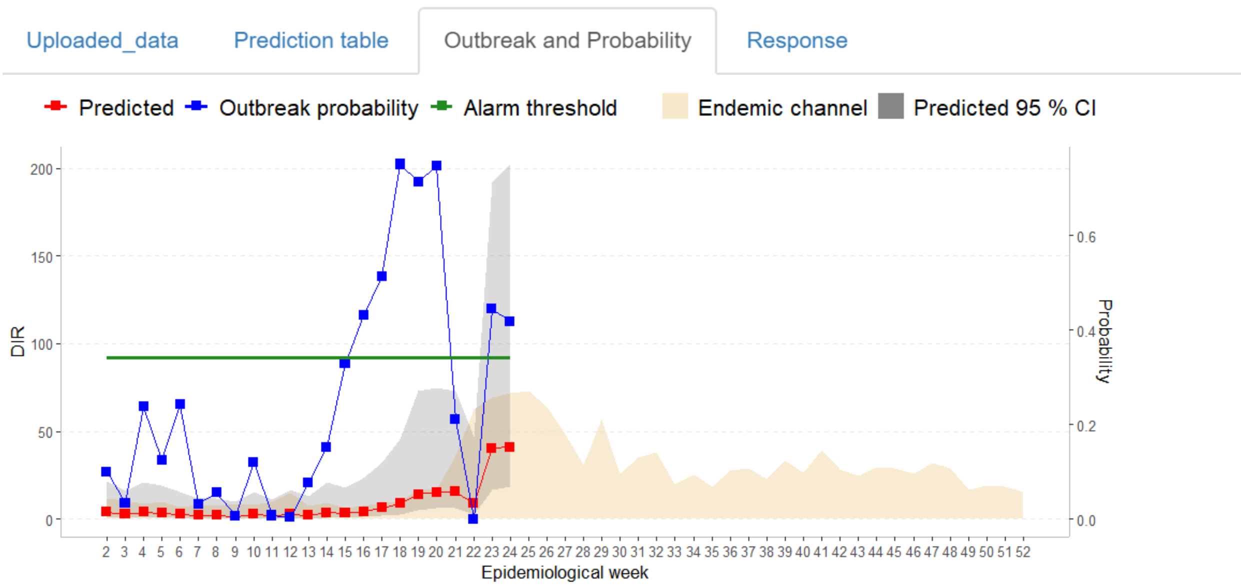
1. **District**: a number (code) that represents the district, woreda or municipality where data were forecasted /captured (e.g. 30115, 30201,30202, etc.);
2. **Year**: the year when the prospective data were forecasted/collected. The year must be entered in full using four numbers (e.g. 2022, etc.);
3. **Weeks**: the epidemiological week (Sunday to Saturday) number at which the prospective data were forecasted/ obtained. The week number must be entered in full, i.e., 1, 2, 3, ….52);
4. **Population**: the annual (mid-year) population size of the corresponding district.
5. **Endemic channel**: the number of cases within the expected normal seasonal range (z*SD + moving average) of specific district; anything above this moving threshold would be considered representative of an outbreak. Where 'Z' (previously Z-outbreak) is a multiplier value of the SD provided by the user to vary the breadth of the endemic channel.
6. **Predicted**: predicted number of cases by the prediction model.
7. 'p25' , 'p75': the lower and upper, respectively, 95% confidence limit of the predicted cases.
8. **Outbreak**: expected outbreak cases among predicted cases. You can note that, the number of outbreaks is less than or equal to the number of predicted.
9. **Alarm threshold**: a threshold in which the probability of outbreak above it indicates possibility of upcoming outbreak.
10. **Outbreak probability**: the fraction that indicate the likelihood of the outbreak to happen.
11. **Alarm**: Indicator of existence of alarm (yes=1, no=0) at the current epidemiological week.
12. **Alarm signal**: Triggering of the alarm signal (yes=1, no=0).
13. **Response category**: staged-response category (No response =0.5, Initial response =1, Early response=1.5, and Late/emergency response =2).

4.2.3 Outbreak and Probability

In this tab, a graphical representation of the prediction table is presented. You can choose one district at a time to see the result. Different summaries that are helpful for response are presented.

This option is found under 'Outbreak and Probability' tab in Dashboard II.

Figure 28. Outbreak and probability plot



- Under the 'outbreak and probability' tab, you can see a summary of prospective week with respect to all the 'predicted case (including 95% CI)', 'outbreak probability' 'alarm threshold', and 'endemic channel'.
- Here 'blue dots' represent the probability that the **predicted case** (red dots, 95% CI in gray) be an outbreak at each prospective week (current week plus prediction distance);
- Once the **outbreak probability** (blue dots) exceeds the '**Alarm threshold**' (green line) the 'alarm signal' is generated! i.e., there is a significant probability that outbreak will take place in the weeks ahead, as indicated by the prediction distance.

  For example, at week 16 blue dots are above green line, which indicates there will be outbreak probability on ISO week 28 (i.e, 16+12, a forthcoming week). Suggested response is presented in the staged-response plot below (No response needed on week 28 (i.e 16+12), however, action is needed for the next weeks, 29-32).

4.2.4 Response

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 from the outbreak prediction as outlined in the 'outbreak and probability' tab, mainly for the relationship between the outbreak probability and the alarm threshold and, following the national recommendation for outbreak responses at district/woreda 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. The response tab as we see in the next Figure (for particular district) shows the 'Late/emergency response' on week 19 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) happening at weeks 16, 17 and 18.

   Then, due to this consecutive occurrence of alarm signals, the program notified a 'Late/emergency response' to be considered at week 31 (i.e 19+12). In this scenario, the dashboard already declared 'Initial response' at week 17 (since there had already been two consecutive alarm signals) and an 'Early response' in week 18 (since there had already been three consecutive alarm signals)!

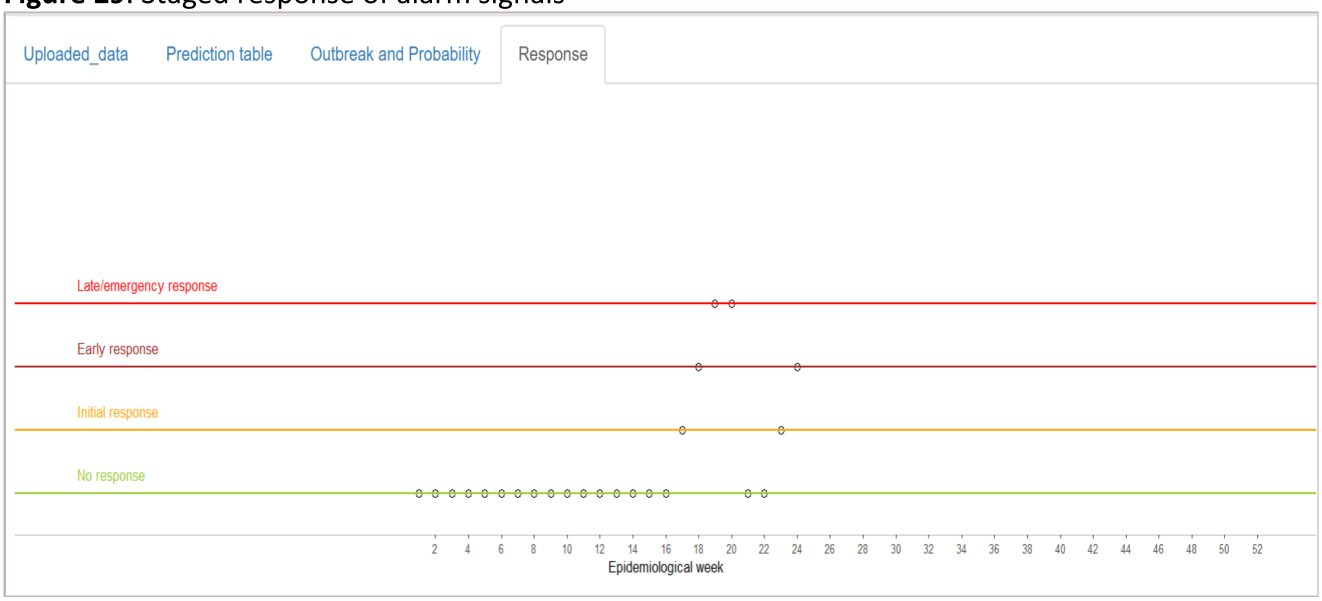
2. **Early response**: is declared when three consecutive alarm signals occur, to avoid 'Late/emergency response'.

   Another example from next figure shows that early response in week 18 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) at weeks 16, 17 and again in week 18. Then, due to this consecutive occurrence of alarm signals, the program notified that an early response should be considered at week 18. In this scenario, the dashboard must have already declared 'Initial response' at week 17 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 below, the initial response in weeks 17 occurred due to an alarm signal (outbreak probability exceeding the alarm threshold) happening at weeks 16 and 17. Then, due to this consecutive occurrence of alarm signal, the program notified that an initial response should be considered at week 17.
4. **No response**: is declared when there are no alarm signals or only one alarm signal in the current week.

This option is found under 'Response' tab in Dashboard II.

Figure 29. Staged response of alarm signals



Chapter five

5 Accessing and additional tabs in EWARS+ dashboard

5.1 Background

Climate influences the transmission of many infectious diseases, some of which being among the most important causes of death and morbidity in developing countries including Ethiopia. Commonly, these diseases occur as epidemics which may be triggered by variations in climatic conditions that imply higher transmission rates. Strengthening early surveillance, warning, alert and response systems (EWARS) for climate sensitive health hazards becomes fundamental under conditions of rapid global environmental change, population movements, disease vectors and infections.

Recent developments in machine learning and the use of big data applications enable complex analysis and enhanced predictions of early warning systems, which can fundamentally innovate response effectiveness, if structured and deployed strategically. Innovating the EWARS into a fast, reliable and harmonized tool, will aid users in endemic countries to apply more efficiently at wider-spread approach.

EWARS+ dashboard is an easy-to-use tool that provide information, using visuals to communicate the stories behind the data. It guides decision-makers through the relationships of complex, big informative data. It presents visuals in a practical order enabling quicker understanding and appreciation of data to the health system. It is developed based on Shiny web application. Specifically, the Shinny dashboards let you access a complete web application framework within the R environment. That is, it easily turns your work in R, analyses and visualizations, machine learning models, and more into web applications. In EWARS+, end-users don't need an understanding of R to use it.

5.2 Accessing the EWARS dashboard

The shiny server used for this new version is an open source from the R package that provides an elegant and powerful web framework for building web applications using R. Shiny is more efficient in running and maintaining which has a professional annual cost of 3300 USD. It is also easy to install therefore, it facilitates a smoother process of transferring the EWARS into a national/subnational program.

Users at central level can assign unique passwords for users to access district-level information (for running prospective data). Through the link, district-level users can access the dashboard by entering given passwords. With the new advancement, the platform allows to integrate with the national surveillance program without significant IT running and maintenance cost. A package including information on the programming script, operational guide, demo data/file will be provided with the dashboard as discussed next.

5.3 **Admin tab**

Here the district level admin/manager/officer or any privileged personnel can enter password to get access of the dashboard.

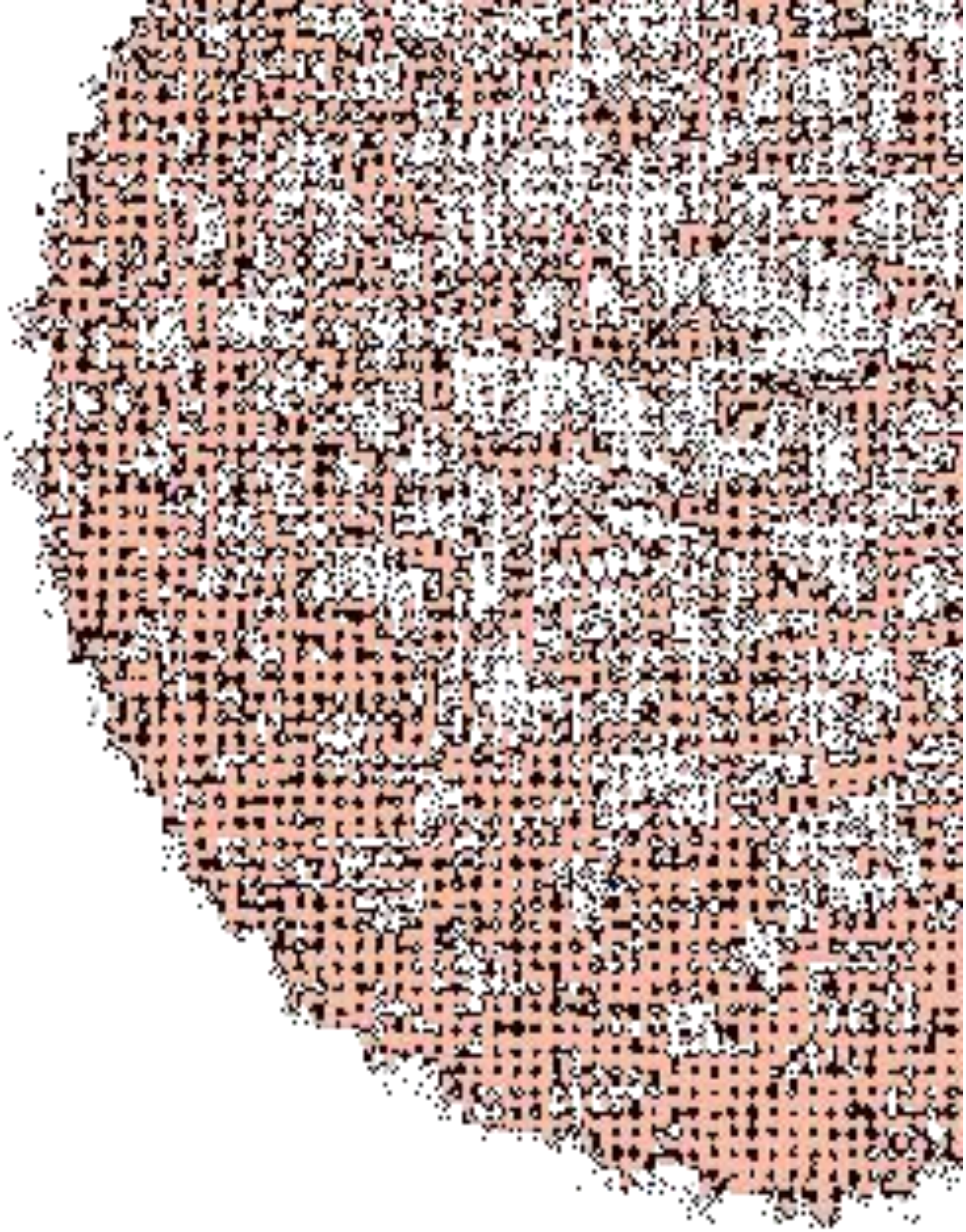
5.4 **Help tab**

Different files including the operational guide for Dengue outbreak, demo data and files are provided.

5.5 **R scripts and Files tab**

Here R scripts and files for running the demo data are provided that can be accessed via GitHUB.

To read and understand more about the predictive abilities of the Early Warning and Response System please refer to: 1) The Early Warning and Response System (EWARS-TDR) for dengue outbreaks: can it also be applied to chikungunya and Zika outbreak warning? | BMC Infectious Diseases | Full Text (biomedcentral.com). 2) Early warning systems (EWSs) for chikungunya, dengue, malaria, yellow fever, and Zika outbreaks: What is the evidence? A scoping review | PLOS Neglected Tropical Diseases



Annex 1.
Technical guide

1 Introduction

This technical guide explains the scientific rationale behind *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' outbreaks as defined by disease incidence (probable/hospitalized cases). The early warning system detects changes in the alarm indicators (meteorological, entomological and epidemiological) to predict disease 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.

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) validation 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 validation process 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 Validation 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.

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 uploading 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/validating parameters) by assessing the association between the level in the alarm indicators (i.e. mean rainfall, mean temperature, mean of humidity, etc.) and disease outbreaks. Negative binomial regression is used to assess this association. This regression model processes the computed proportions of the outcome (computed via a cut-off value from ROC curve), to predict the probability of disease outbreaks for a forthcoming period. Throughout these steps, the user is able to define relevant measures such as the size of the endemic channel, different thresholds/references, the lag week (prediction distance) and so on. Descriptions of these measures can be found in the following chapter.

3 Structural design

The overall process of the analysis distributed across two different phases. Phase I (retrospective) is divided into: (i) the 'Run-in data'; and (ii) the 'Validation data'. Phase II (prospective) is the final analysis using summary statistics to populate a web-based early warning system that can be used in real-time to detect future disease 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, variables in the original data are assigned as alarm indicator, number of hospitalize cases and district population. Probable cases might be substituted when hospitalizes cases are unavailable, however, it should be consistent across all dataset. A minimum of one alarm indicator is required to run the model, but there is no maximum limit.

Step 2

In this step, the user needs to choose desired lag week (prediction distance). That is, choose the distance between current week and target week to predict an outbreak signal. EWARS+ keeps a fixed 12 weeks' time-lag (prediction distance), however, users can choose a desired value.

Step 3

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

Step 4

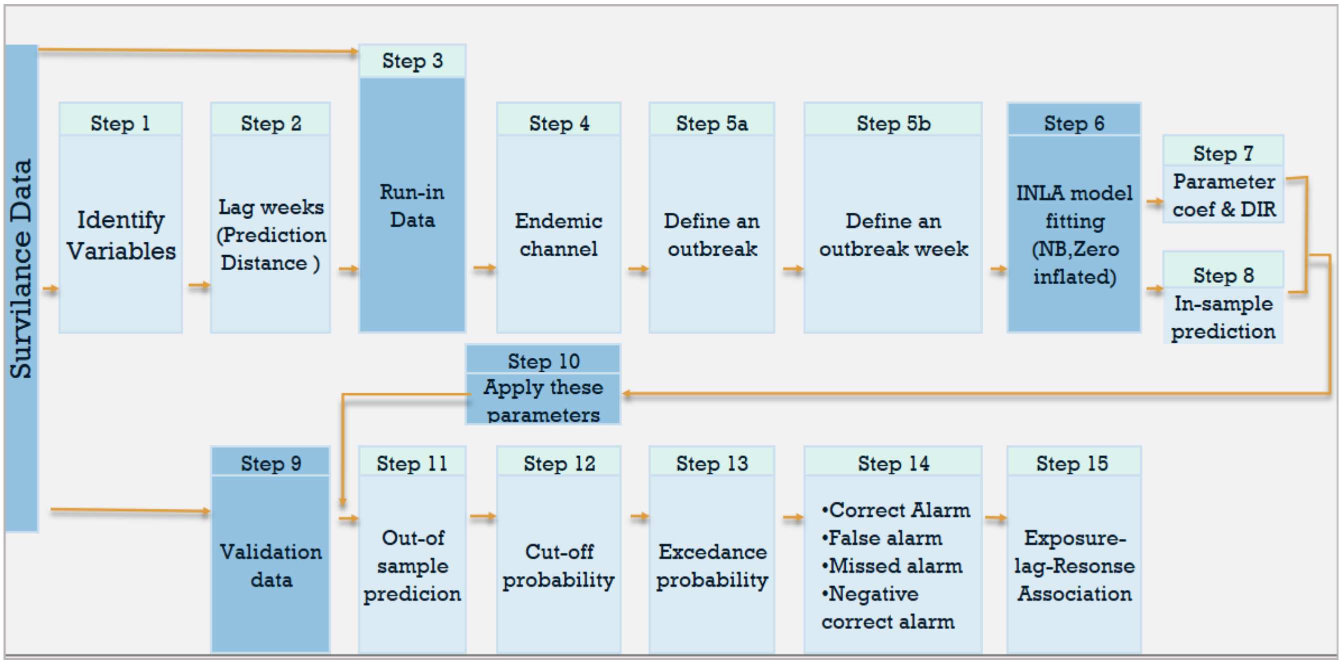
This step refers to the 'Endemic channel' and is represented by the following equation:

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

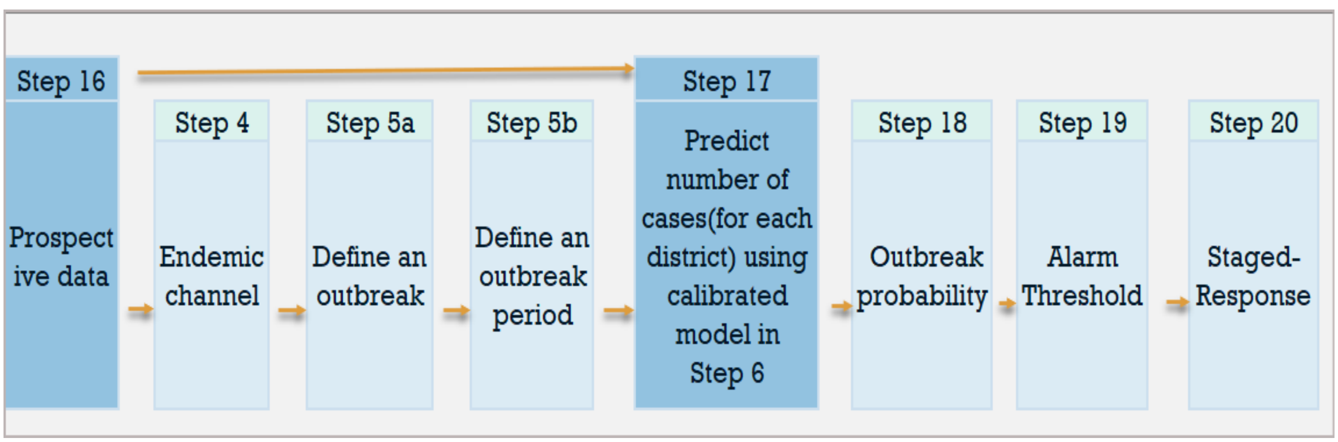
'Z' is a multiplier value of the SD provided by the user to vary the breadth of the endemic channel. This part will assist in declaring the 'out of control' status. For instance, a value of Z=1.5 would increase the breadth one and a half times the expected normal range of the number of disease cases. The moving average is the mean number of disease 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 30. Block diagram illustrating the process and steps of the retrospective and prospective phases (I, II)

Phase I: Retrospective phase



Phase II: Prospective phase



Step 5a

In this step, an 'outbreak' is defined by the proportion of disease incidence rate (DIR) per 1000 population for a corresponding district. The disease probability is compared against thresholds (i.e. case number based on endemic channel) for an outbreak signal.

Step 5b

In this step, an 'outbreak week' is the epidemiological week at which the predicted disease probability is higher than threshold value. Optimal threshold is determined by calculating ROC curve so as to issue alarm signals for the binary events of exceeding the moving outbreak threshold (exceeds/ not exceeds).

Step 6

In this step, we fit a generalized linear regression model to assess the association between the count outcome of the number of hospitalized (or probable) cases and the alarm indicators. Among the generalized linear models, we can fit Negative binomial regression or Zero-inflated negative binomial as desired. These models are used to generate the incidence rate per population.

The Bayesian hierarchal model will be used to conduct leave-one-out cross validation process to generate the model parameters. The Integrated Nested Laplace Approximation (INLA) will be used in the Bayes estimation of the parameters.

Step 7

The coefficients and DIR generated from those regression models will be stored/applied during the validation stage.

Step 8

In this step, the fitted model is used to predict the number of disease cases for run-in period. This is helpful to evaluate the predictive capabilities of the model developed using observed data to see how effective the model is in reproducing data.

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 9

Refers to the second part after splitting the surveillance data, which we call the 'Validation data' (the first part of the original data was used in step 3, i.e. the 'Run-in data'). The user determines this by entering the period, in year (e.g. 2021) for when the Evaluation data begins (on first week of 2021). A one year or above data records is preferable for the Evaluation dataset.

Step 10

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 'Validation data' to observe the performance of the prediction model in detecting outbreaks.

Step 11

In this step, the calibrated model is used to produce out-of sample (i.e. for validation period) prediction of disease cases and probability of exceeding the endemic channel (exceedance probability). This step is another means to exam the model.

Step 12

The 'cut-off probability' is an optimal probability that the predicted case is above the endemic channel. The probability that the predicted case (incidence rate) is above the endemic channel is higher than the 'cut-off probability' indicates signal of an outbreak at the corresponding week. Both the threshold and the probability for the predicted incidence rate to exceed the endemic channel will be assessed using the ROC curve.

Step 13

In this step, the model produces the exceedance probability (the probability for the predicted incidence rate to exceed the endemic channel). This helps to predict outbreak and non-outbreak weeks. For instance, for an exceedance probability of 0.4, with a threshold of 0.3 (i.e. exceedance probability > threshold), this record is said to be an alarm signal. If, however, the exceedance 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 12) is reliable to predict an outbreak, the program will further present four evaluation criteria to assess the choice given in step 12. These criteria are:

1. Correct alarm: Exceedance probability ≥ threshold with a true outbreak (for a target week);
2. False alarm: Exceedance probability≥ threshold with no true outbreak (for a target week);
3. Missed outbreak: Exceedance probability ≤ threshold with a true outbreak (for a target week);
4. No alarm, no outbreak: Exceedance probability ≤ threshold with no true outbreak (for a target week).

An optimal threshold (option in step 12) 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 dashboard output), the user can alter calibrations (Z-outbreak and other), accordingly.

Step 15

In this step, the model produces the temporal dependency between exposure and outcome. Using heat-map plot, the dashboard displays the association between relative risk (RR) of the disease and the alarm variable over the lag time. The user can alter the lag time (as desired), to see the exposure-lag-response association.

3.2 **Phase II (prospective surveillance: Dashboard II)**

Step 16

In this step, the program will run through the prospective dataset to populate the output file (Dashboard II) that includes information on the predicted probability of outbreak and suggested staged-response on the forthcoming weeks. Once the user has uploaded prospective data, the file will automatically estimate and graph both the probability of an outbreak to predict an outbreak and the staged-response.

Step 17-19

In these steps, the tool calculates information including the predicted case, outbreak probability (i.e. similar to exceedance probability in dashboard I) and alarm threshold (i.e. similar to Cut-off probability in dashboard I).

Step 20

In this final step, the program will automatically produce an instant graphical presentation and alarm signals/responses.

References

1. Programme for Research and Training in Tropical Diseases. (2020). Operational guide using the web-based dashboard: Early Warning and Response System (EWARS) for dengue outbreaks, 2nd ed. World Health Organization. https://apps.who.int/iris/handle/10665/332323.
2. Cardenas R, Hussain-Alkhateeb L, Benitez-Valladares D, Sánchez-Tejeda G, Kroeger A. The Early Warning and Response System (EWARS-TDR) for dengue outbreaks: can it also be applied to chikungunya and Zika outbreak warning? BMC Infect Dis. 2022 Mar 7;22(1):235.
3. Hussain-Alkhateeb L, Rivera Ramírez T, Kroeger A, Gozzer E, Runge-Ranzinger S. Early warning systems (EWSs) for chikungunya, dengue, malaria, yellow fever, and Zika outbreaks: What is the evidence? A scoping review. PLoS Negl Trop Dis. 2021 Sep 16;15(9).