THE MOVING EPIDEMIC METHOD

The MEM Web Application: technical manual

José E. Lozano[1]
Tomás Vega[2]

[1] Foundation Institute of Health Sciences Studies of Castilla y León (Spain)
[2] Public Health Directorate. Junta de Castilla y León (Spain)

Disclaimer

This document has been created using the version 2.16 of the MEM Web Application. Some features might have been added, deleted, moved or changed since the last revision of this document.

Last updated on 24 May 2024

Contents

List of tables

List of figures

1   Background & objectives

The Moving Epidemics Method (MEM) is a tool developed in the Health Sentinel Network of Castilla y León (Spain) to help in the routine influenza surveillance in health systems. It gives a better understanding of the annual influenza epidemics and allows the weekly assessment of the epidemic status and intensity.

Although in its conception it was originally created to be used with influenza data and health sentinel networks, MEM has been tested with different respiratory infectious diseases and surveillance systems so nowadays it could be used with any parameter which present a seasonal accumulation of cases that can be considered an epidemic.

MEM development started in 2001 and the first record appeared in 2003 in the Options for the Control of Influenza V[1].

It was presented to the baselines working group of the European Influenza Surveillance Scheme (EISS) in the 12th EISS Annual Meeting (Malaga, Spain, 2007), with whom started a collaboration that continued when, in 2008, was established the European Influenza Surveillance Network.

In 2009 MEM is referenced in an official European document: the Who European guidance for influenza surveillance in humans[2]. A year later MEM was implemented in The European Surveillance System (TESSy), of the European Centre for Disease Prevention and Control (ECDC), and in 2012, after a year piloting, in the EuroFlu regional influenza surveillance platform, of the World Health Organization Regional Office for Europe (WHO-E).

As a result of the collaboration with ECDC and WHO-E, two papers have been published, one related to the establishment of epidemic thresholds[3] and other to the comparison of intensity levels in Europe[4].

In 2010 a tool was created to help users around the world to apply mem on their data. It was released in November 2010 as a package for R[5], a free software environment for statistical computing and graphics. This is the first version[6] of the MEM R package which was published at CRAN (The Comprehensive R Archive Network), the official R repositories.

The source code was rewritten completely for the second version of the package, including a lot of new features and graphics. It was published in November 2015 as an open source project at GitHub, a web-based Git or version control repository and Internet hosting service. This version is the development version[7] and it is constantly being updated with latest bug fixes and improvements.

This second version of the mem R package was incorporated to CRAN two years later, in June 2017, and since then, this is the official stable version of the mem R package[6].

The first web application for mem was created by Jakob Bergström for the The Public Health Agency of Sweden. Using his code as the starting point, the memapp R package was created to serve as a graphical user interface for the R mem library. It was coded using Shiny, a web application framework for R. The development version was released at github in March 2017[8]. Three months later, in June 2017, the official stable version[9] of the web application was uploaded to the CRAN repositories as the stable official version of the memapp package.

The objective of this manual is to give to the memapp users instructions of how to download, update, install and run the application, as well as how the data must formatted to be used with the application. The manual describe different procedures and options and several features

required in some particular cases. Finally, it shows how use the memapp outputs for monitoring and surveillance.

This manual does not include specific explanations about the algorithms and statistical methods behind the estimators or indicators calculated in memapp. The users could find most of the methods implemented in the referred articles.

This manual is the complement to the 'Guidelines monitor influenza and other respiratory virus infections epidemics and pandemic' where the authors give a general overview of the concepts and terminology in influenza and other respiratory virus modelling for surveillance purposes and propose a step by step guide to best model a particular country and surveillance parameter using the MEM methodology.

2   Installation

2.1   The R language

The mem Shiny web application (memapp) is based on the mem R package and requires R to work. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS. There are binaries for most operating systems at its official web page[5]. To install it, download the binaries appropriate for your system and proceed to install it.

R is a command line program but there are a lot of graphical user interfaces (gui) available to users that wants a friendlier environment. The most popular is RStudio an open source powerful and productive user interface for R. Binaries can be downloaded from its official web[10] and installed on Windows, Mac, and Linux.

Shiny is a web application framework for R created by the RStudio team, there is no need to install separately because it acts as a package of the R language and will be installed with the rest of dependencies.

A Shiny app is a set of code files that the Shiny framework is able to interpret in order to start the web application. They can be run directly from a remote web server or in a local directory of our hard disc. If run in local, the Shiny framework start a local server and launch the application. If run remotely, then the server is the one that might be running R and start the application.

2.2   Packages and repositories

The core of R is the packages, collections of R functions, data, and compiled code in a well-defined format that perform specific tasks. The directory where packages are stored is called the library. R is based on the contribution of the community, which creates their own code as packages and share them, extending the features of the base package.

Most packages need other packages to run not to replicate code others have done before. These are the dependencies of a package. When a package is installed, all the dependencies are installed as well.

The R packages are listed and archived in The Comprehensive R Archive Network (CRAN), a network of mirrors of the official repositories where all official packages are distributed.

There is another option to store a package for R: to use sharing code platform. The most popular is GitHub, a web-based Git or version control repository and Internet hosting service. Users can install packages from these hosters compiling them from its source code.

Most packages have the official version at CRAN and a development version at GitHub and users can install one or the other. Generally, packages at CRAN are stable versions and pass more tests to be published.

2.3   The MEM app

The official web of memapp is:

https://github.com/lozalojo/memapp

[9]

There are two options to run the web application:

- Locally.
- Remotely.

To run it locally you need to install R in your computer, set up the app dependencies and start it.

To run it remotely, you only need an internet connection and the url where the application is hosted, no need to install R or worry about dependencies since all is already set up in the server side.

2.3.1   Run memapp locally

2.3.1.1   Installation

MEM Shiny app is shipped as an R package and requires a set of packages (dependencies) to run, all of them are installed automatically. It has two version: the stable hosted at the official repositories (CRAN) and the development version hosted at github.

To install the *stable version* copy and paste the following code in the command line prompt. It is recommended to install memapp from the RGui (the default front end shipped with R).

```
# install the memapp package
install.packages("memapp")
```

The development version includes the latest updates. Before installing it, another package called devtools is required. To install memapp, copy and paste the following code to the R console:

```
# check if devtools is installed, and install it otherwise
if(!require("devtools")) install.packages("devtools")
library("devtools")
# install the memapp development version from GitHub
install_github("lozalojo/memapp", ref="development")
```

When installing the development version of memapp, you are also installing development versions of the most important packages that memapp uses to have them updated.

Along with memapp, there are other libraries that might be worth installing. Some memapp features will not be enabled if those libraries are not installed. Here is a list of suggested packages to install:

- openxlsx: to read MSExcel files (.xlsx) and to export data to excel.
- readxl: to read MSExcel 97-2003 files (.xls)
- foreign: to read dBase (dbf), SPSS (sav) and Stata (dta) files
- haven: to read SAS (sas7bdat) files
- readODS: to read Opendocument spreadsheets (.ods)
- RODBC: to read MSAccess files (mdb, accdb) on Windows platforms (wont work for Linux, though there is a workaround, see 5.3 for more information)
- animation: to create the surveillance graph. It will require to have graphicsmagick or imagemagick installed (see 5.5)

To install any package, simply write at the console:

```
install.packages("packagename")
```

As an example:

```
# install openxlsx the package
install.packages("openxlsx")
```

If you find any troubles installing dependencies, please refer to the installation chapter of the manual of each specific library that failed to install to find a solution for your specific OS.

You can find installation guides for Windows, Linux and MacOSX at the following youtube channel:

https://www.youtube.com/playlist?list=PLhUpdbxODE7DizLt2TA-Hgw0cQ5ycM8BB

2.3.1.2   Running

To run the application, start it by typing:

```
memapp::runmemapp()
```

The application will start the application in your default browser.

2.3.2   Run memapp remotely

To run the app remotely there is no need to install R, memapp or its dependencies, it's the server the one that runs the code, not the local computer.

Only a web browser with an active internet connection is required. The app is hosted in a web server at the Foundation Institute of Health Sciences Studies of Castilla y León (Spain), the address is:

http://memapp.iecscyl.com:8080/

3 The MEM Web Application

The application has three different panels (Figure 1 – memapp main page) to navigate:

- Top panel: Main menu of the MEM different procedures.
- Left panel: Allows to load data and has options of the procedures that affects directly to the outputs of the application, like model or thresholds selection.
- Right panel: By default, it is hidden, but can be activated by clicking the gear button at the top right corner. Has options to modify graphs, like the palette to use in plots or the axis names, and to change mem parameters.
- Main panel: Where all outputs are displayed.
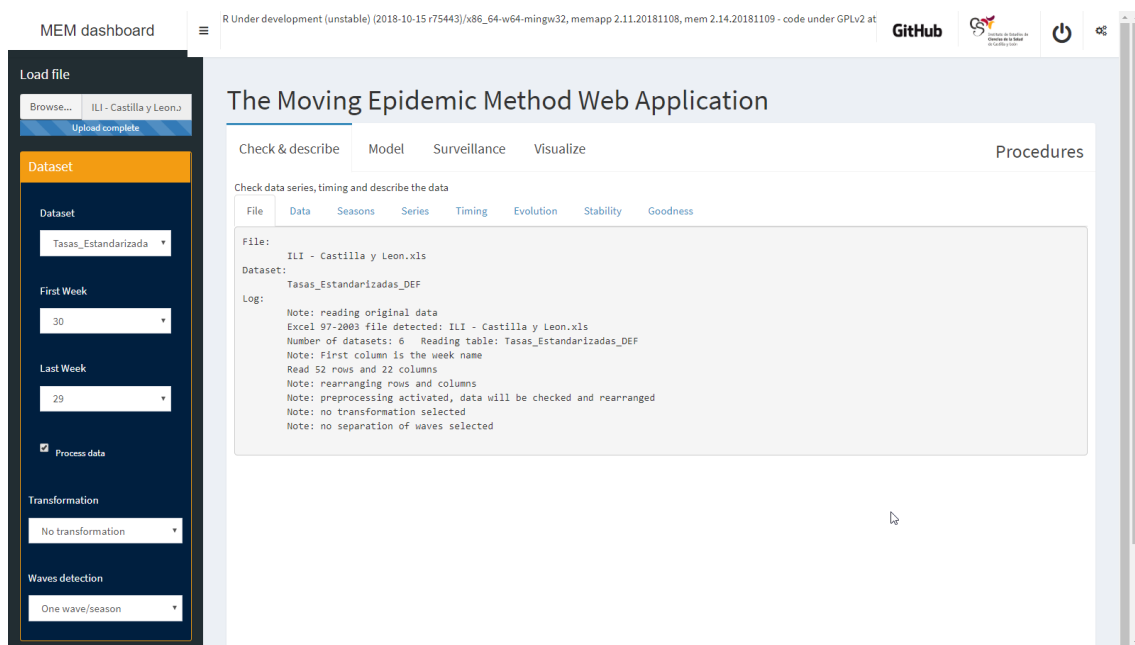
Figure 1 – memapp main page

3.1 Data

3.1.1 Loading data

The upper left button Load file allows to upload a file that contains data. Currently memapp supports the following formats:

- Text files (csv, dat, prn, txt): text files, with any encoding, any separator and any decimal point.
- R files (rda, rdata, rds): formats used by R.
- MS Excel files (xls, xlsx): excel files, both Excel 97-2003 and Excel 2007+.
- MS Access files (mdb, accdb): access files, both Access 97-2003 and Access 2007+.
- dBase files (dbf): dBase files.
- SPSS files (sav): SPSS files.
- Stata files (dta): Stata files.
- SAS files (sas7bdat): SAS files.

When loading a file, in the left panel will show up a box called **Dataset** (Figure 2 – dataset box) to set up different options.

[12]

All available formats, except text files, are able to store more than one dataset. The combo box **Dataset** allows selecting the dataset you want to use.
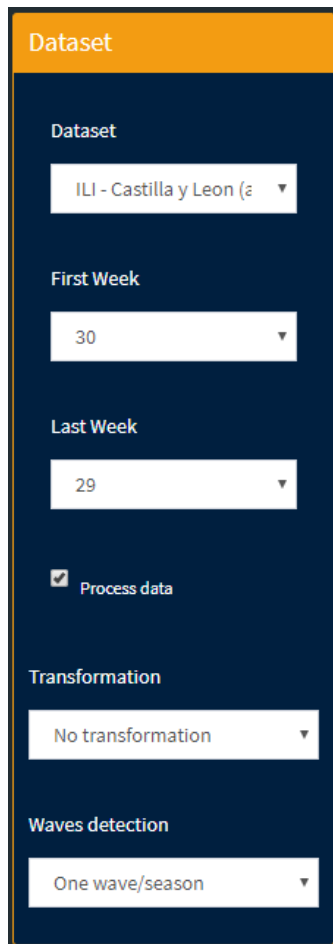
Figure 2 – dataset box

In addition, a specific surveillance period can be selected using **First week** and **Last week**.

Note that using these options the data will be automatically transformed to fit the new structure of your surveillance period.

If the surveillance period of the original dataset is different from the one selected in the box, then the data will be transformed. This option can be used to transform data from northern hemisphere-like surveillance (Table 2 – northern hemisphere surveillance) to southern hemisphere-like surveillance (Table 3 – southern hemisphere surveillance), or the contrary.

Example: you use as input a set of 5 seasons with southern-like surveillance period involving one single year: 2000, 2001, 2002, 2003 and 2005 going from week 1 to week 52. When seeing data you decide that the surveillance period should start at week 40 because the epidemic occurs between weeks 45 and 10 next year, so you set the first week to 40 and the last week to 39. The program automatically change the notation of the seasons to a northern-like structure, ending up with SEVEN (not five!) seasons: 1999/2000 (which contains data from 2000/01 to 2000/39), 2000/2001, 2001/2002, 2002/2003, 2003/2004, 2004/2005 and 2005/2006 (data from 2005/40 to 2005/52).

The **Process Data** checkbox is used to apply pre-processing to the original data. Pre-processing the data implies:

- Re-estructure the data using the First and last week of surveillance.
- Check row names and rearrange them if there are missing rows (f.i. if dataset jumps from week 13 to 15, the program adds week 14 and fill in with missing values), reorder them and check if there is week 53 to accommodate according with 53rd week rules (see Week 53), and summarise rows (mean) in case of duplicates.
- Check column names to see if there are duplicated data (for example, having two columns of the same season, numbered (1) and (2) and having data for the same week).
- Removing columns with zero ony values.

If the checkbox is not tickmarked, then there is no pre-processing, but also will ignore the transformation and waves options that follow.

The **transformation** option serves of modify the original data. Transform applies a transformation to each column of the data and uses it in the application.

- No transformation: applies no transformation.
- Odd: Calculate the odds for percentage-like values. If the values are above 1, it converts first to 0-1 scale and then calculates the odd.

$$odd = {p}/{1-p}$$

- Fill missing: Uses a loess regression to impute missing values in the original dataset. This transformation only applies to missing in the middle of the surveillance period (Table 1 – fill missings scheme). It does not impute starting or ending values that are missing.

-	-	43	56	-	71	81	-	-	102	82	-	43	-	-
Imputation: Yes (green), No (red)														

Table 1 – fill missings scheme

- Smoothing regression: Replaces all data with predicted values of a smoothing regression using an optimal window parameter.
- Loess: Replaces all data in the original datasets with those coming from a loess regression. **Loess span** option (advanced feature) is the span parameter of the loess.
- Splines: Replaces all data with the predicted values of a spline regression.
- Moving average: Uses moving average. **MA weeks** parameter (advanced feature) is the number of weeks previous to the current one to average.

MEM was designed for finding only one epidemic wave in each season and column. There have been implemented several procedures when there are more than one epidemic per season. These algorithms are under **waves detection** option:

- One wave/season: applies no transformation, only one wave per season.
- Two waves/season (observed): when there are two waves per season or there are seasons with one and others with two waves. This algorithm fits a mixture of two normal distributions to each season to separate two waves in each original season, then separate the two waves in different columns, adding (1) and (2) to the original names. The **minimum proportion** option (advanced feature) tells memapp to remove any detected wave that represents less than this proportion and to fit only one normal distribution to that season.

- Two waves/season (expected): is the same algorithm as two waves/season (observed), but in this case, weekly rates are replaced with predicted data from the normal mixture.
- Multiple waves/series (experimental feature): can be used where there is no a regular pattern in the waves per season or when there is no fixed cut point between seasons. This algorithm detects and separates waves in a whole series. The No. waves (advanced feature) option tells the algorithm to fit a fixed number of waves in the series. Separation (advanced feature) option forces to join to waves separated with no more than a given number of weeks, considering they are the same wave because they are close enough. Param 1 (advanced feature) rules how to detect an epidemic in the whole series, the higher the shorter are the epidemics detected. Param 2 (advanced feature) rules when to stop detecting waves, the higher the less waves are detected. See Multiple waves algorithm for more information.

3.1.2   Data structure

It is important to know how to arrange information in order to use with memapp. The key points are:

- One single epidemic wave each season (see 4.2.3).
- Never delete a rate inside an epidemic (see 4.2.1).
- Accommodate week 53 (see 4.2.2).
- Do not inflate missing values with zeroes (see 4.2.1).

Data must contain information from the historical series. Surveillance period can start and end at any given week (typically start at week 40th and ends at week 20th), and data can have any units and can be expressed in any scale (typically rates per 100,000 inhabitants or consultations).

The table must have one row per epidemiological week and one column per surveillance season. A season is a full surveillance period from the beginning to the end, where occurs at some point one single epidemic wave on it. No epidemic wave can be spared in two consecutive seasons. If so, you have to redefine the start and end of the season defined in your dataset. If a season have two waves, it must be split in two periods and must be named accordingly with the seasons name conventions described below. Each cell contains the value for a given week in a given season.

The first column should contain the names of the weeks. When the season contains two different calendar years, the week will go from 40th of the first year to 52nd, and then from 1st to 20th. When the season contains one year, the weeks will go from 1st to 52nd.

Note: If there is no column with week names, the application will name the weeks numbering from 1 to the number of rows.

- In the northern hemisphere countries (Table 2 – northern hemisphere surveillance), the surveillance period usually goes from week 40 to 20 of the following year (notation: season 2016/2017).

Week No	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
40					
41					
…					
17					

18					

Table 2 – northern hemisphere surveillance

- In the southern hemisphere countries (Table 3 – southern hemisphere surveillance), the surveillance period usually goes from week 18 to 39 same year (notation: season 2017).

Week No	2013	2014	2015	2016	2017
18					
19					
…					
37					
38					
39					

Table 3 – southern hemisphere surveillance

The first row must contain the names of the seasons. This application understand the naming of a season when it contains one or two four digits year separated by '/' and one one-digit number between parenthesis to identify the wave number. The wave number part in a name of a season is used when a single surveillance period has two epidemic waves that have to be separated in order to have reliable results. In this case, each wave is placed in different columns and named ending with (1) for the first period, (2) for the second, and so on.

Examples:

2016/2017: Season starting at any point of 2016 (normally at 40th week) and ending in 2017 (normally at 20th week). This is generally used in northern hemisphere countries.

2016 or 2016/2016: Season starting and ending at 2016. Used in southern hemisphere countries.

2016/2017(1): First wave of season 2016/2017.

2016(2) or 2016/2016(2): Second wave of season 2016.

Starting with version 2.10 of the application, memapp now supports using input files in the following format:

year	week	rate
2010	1	15.3
2010	2	16.4
2010	3	10.4
2010	[…]	[…]
2010	52	5.9

For memapp to detect this structure there might be only three columns and two of them must be named "year" and "week". The program will use the remaining column as values of the corresponding year and week. Once it reads the file, it will transform to the suitable format of week, season1, season2, …, seasonN for the use with memapp.

See more on the data structure in 4.2.

3.2 Procedures

The MEM Web App has four procedures in different tabs of the main panel:

- Check & describe
- Model
- Surveillance
- Visualize

While describing different outputs of the application we will explore more of the options in the left and right panels.

3.2.1 Check & describe

This tab allows to see the full dataset and check if data is correct, see the timing and inspect the evolution of the data across time.

File

It gives information of the file and dataset loaded (Figure 3 – check & describe file) and also gives a log of the reading process.
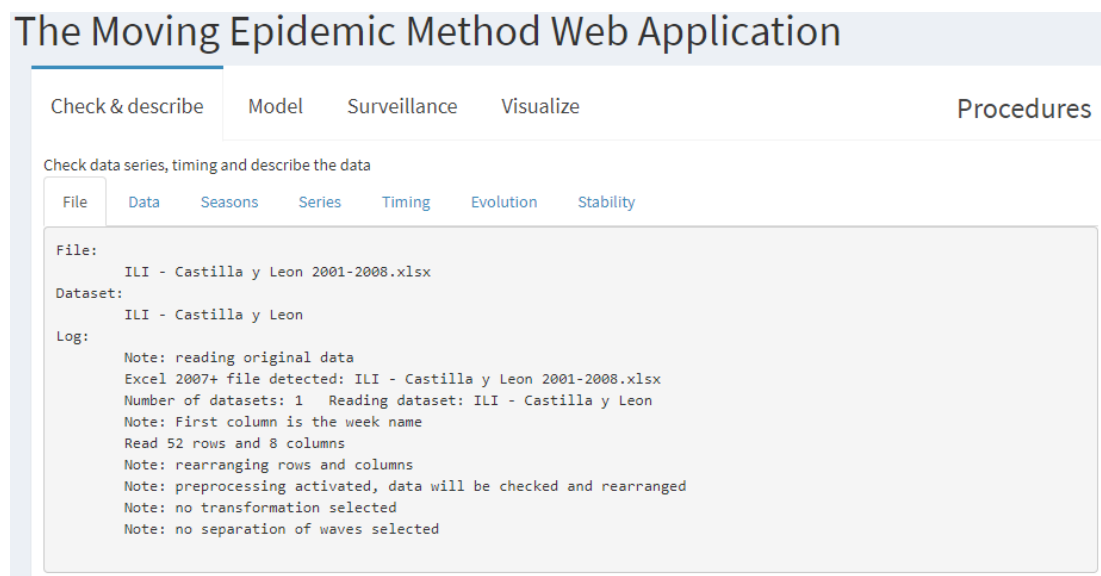

Figure 3 – check & describe file

When waves detection is set to Multiple waves and advanced options are activated also shows a graph (Figure 4 – Multiple waves separation graph) with the separation that MEM has created.
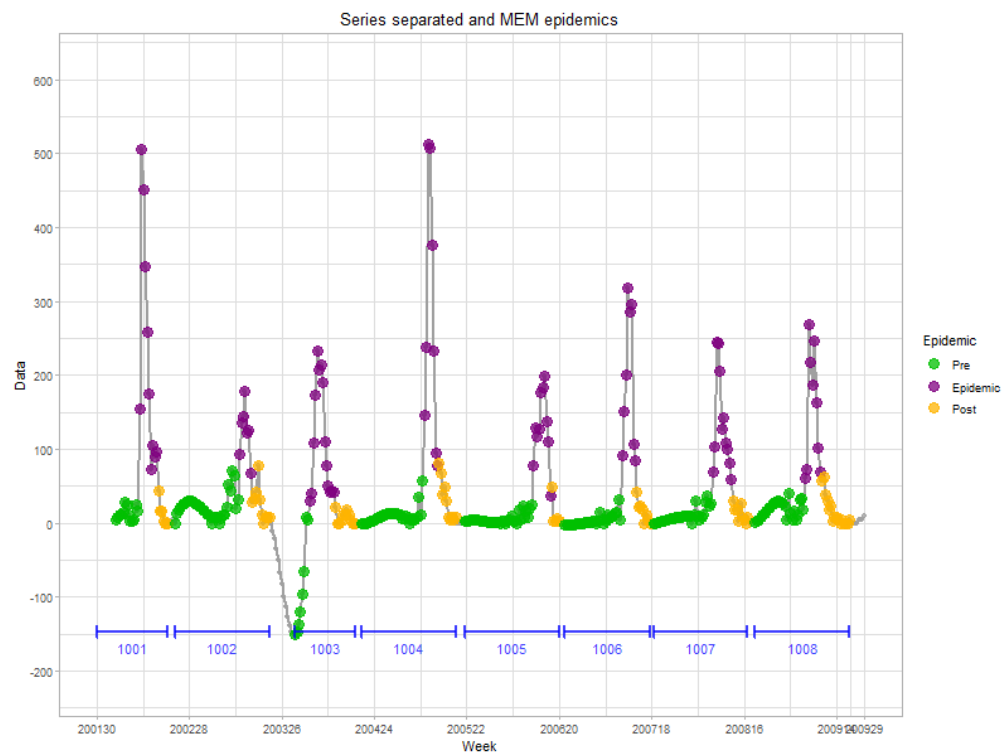
[17]

Figure 4 – Multiple waves separation graph

Data

Shows the data loaded (Figure 5 – check & describe data).
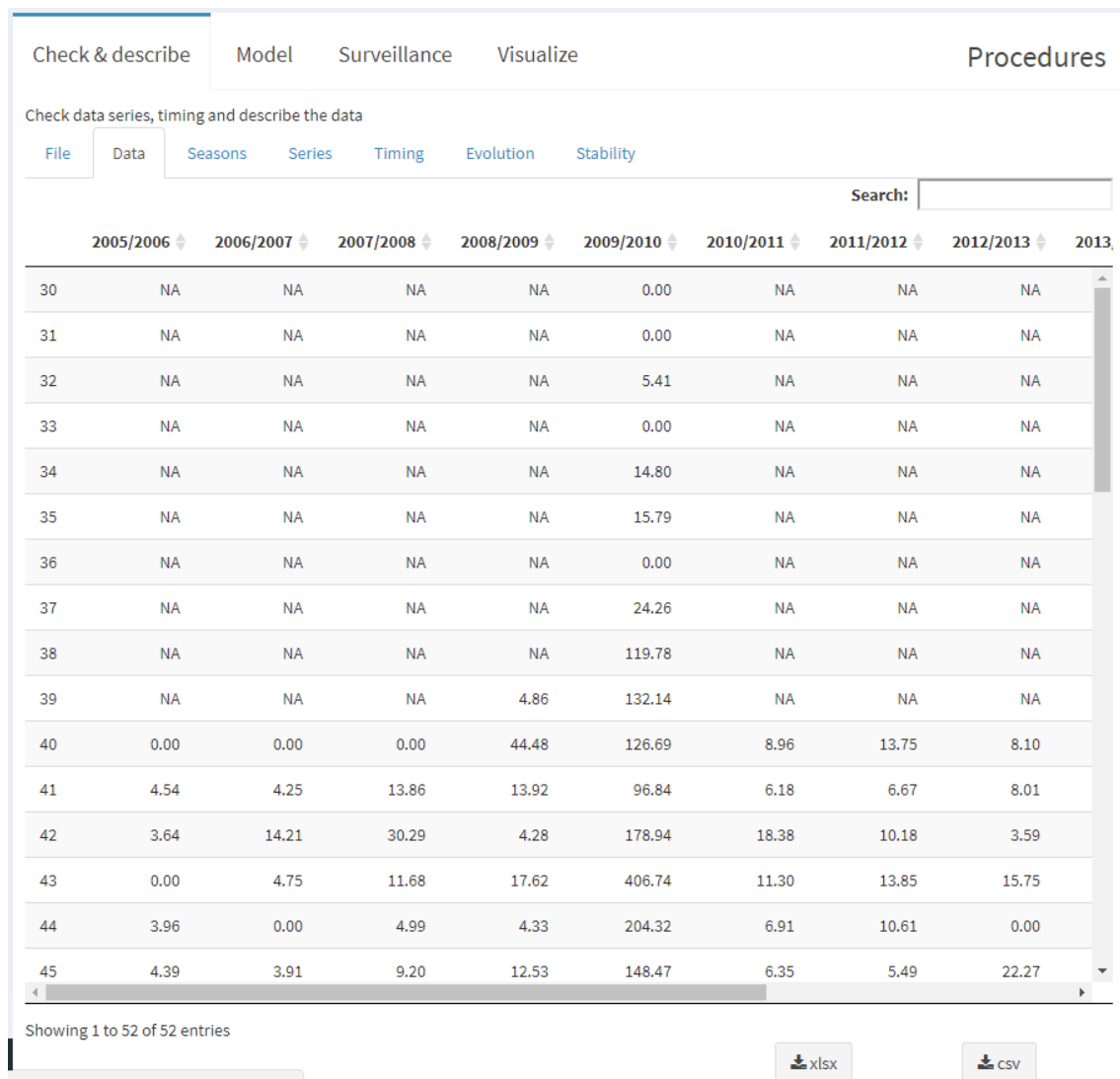
[18]

Figure 5 – check & describe data

This table can be exported to excel or csv files.

Seasons

Makes an overlapped graph (Figure 6 – check & describe seasons) of all seasons to compare magnitudes and timings of all epidemics in the dataset.

[19]

Figure 6 – check & describe seasons

Pre-epidemic, post-epidemic and intensity thresholds can be activated or deactivated in the thresholds box (Figure 42 – thresholds box).

Series

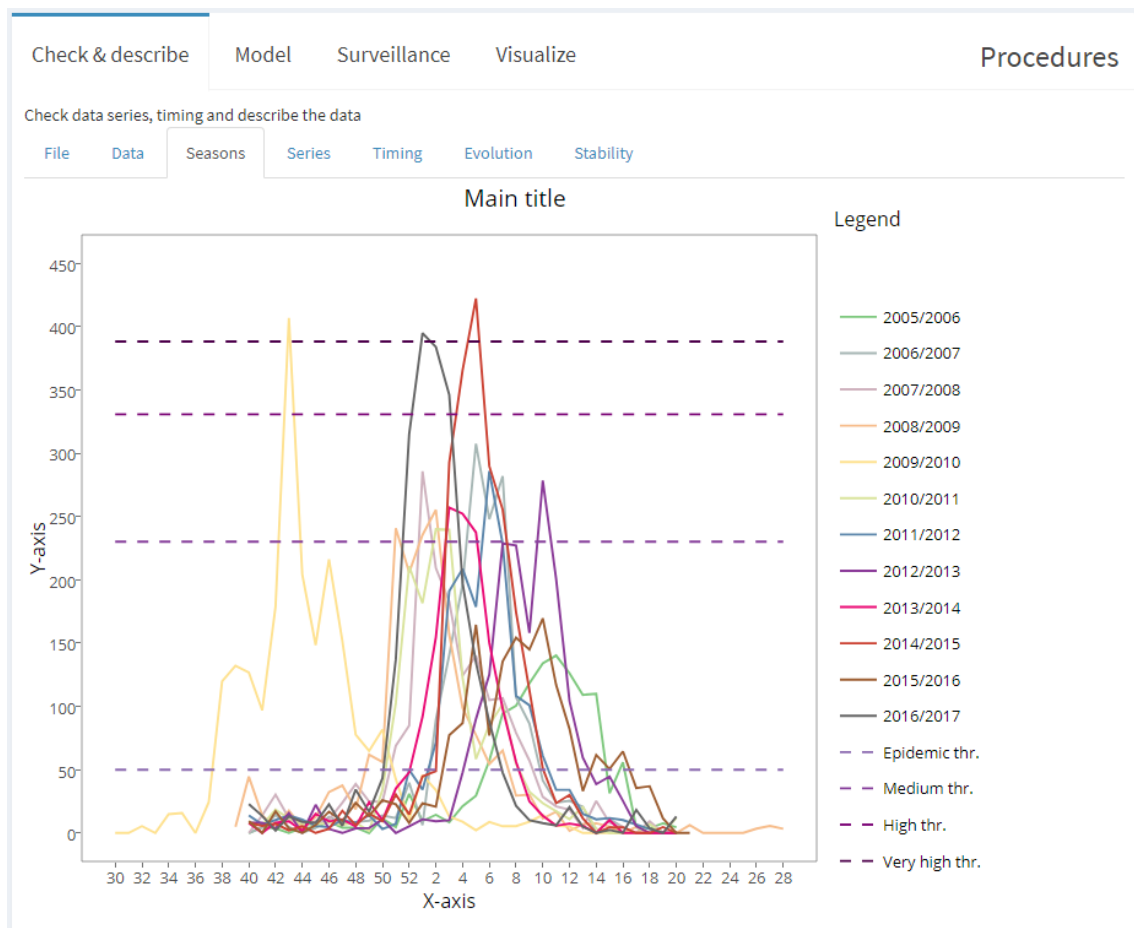Makes a graph of all series across time (Figure 7 – check & describe series).

Pre-epidemic, post-epidemic and intensity thresholds can be activated or deactivated in the thresholds box (Figure 42 – thresholds box).

Figure 7 – check & describe series

Timing

Shows the timing of each season (Figure 8 – check & describe timing) showing pre-epidemic, epidemic and post-epidemic period in different colours.
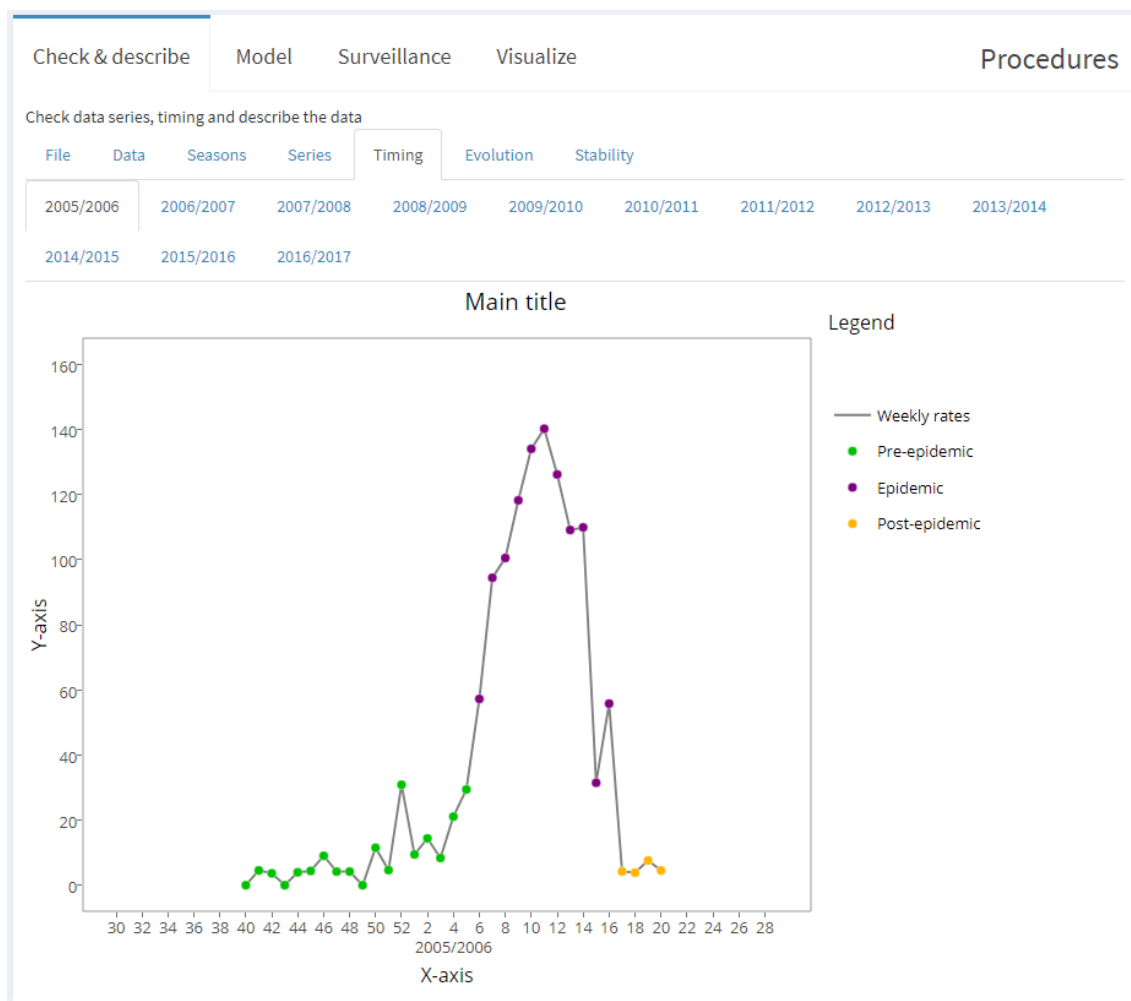
Figure 8 – check & describe timing

When advanced options are activated, two additional graphs showing how MEM detects the epidemic according to the method selected.

The MAP curve shows the maximum percentage reached by any number of consecutive weeks respective to the total sum of rates.
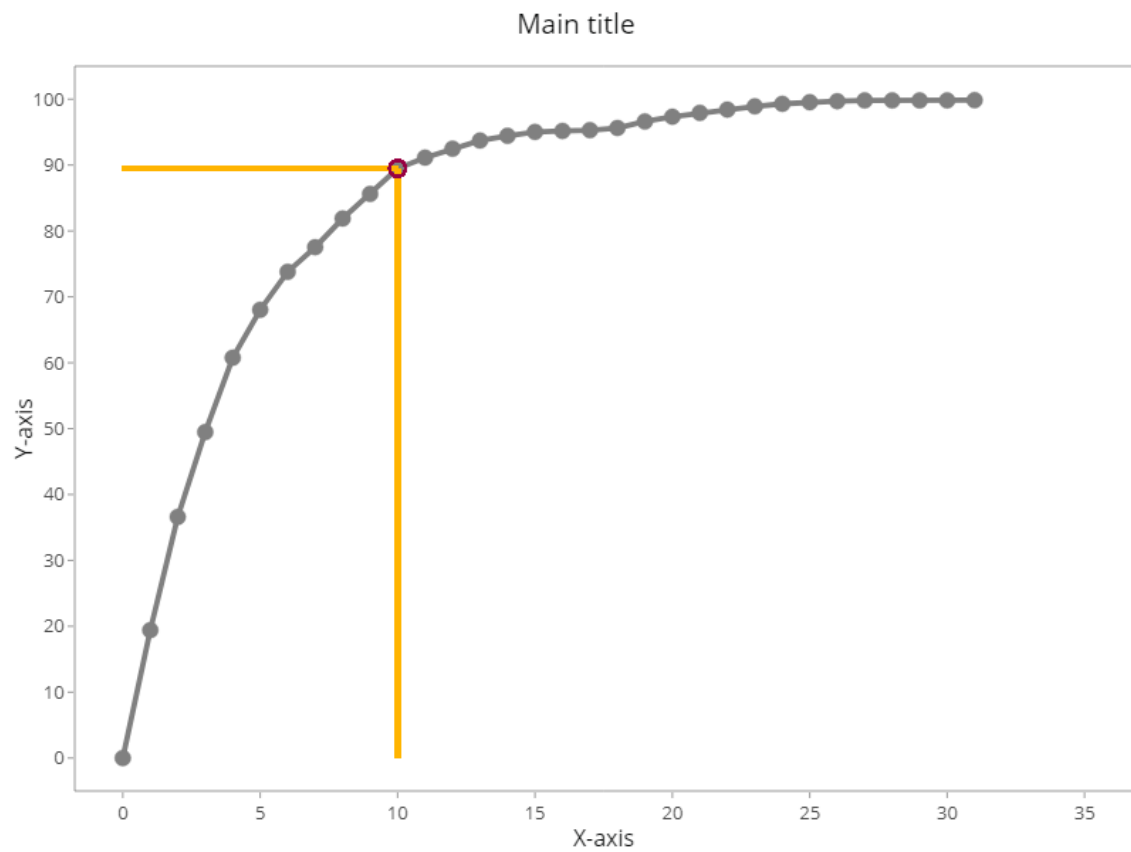
Figure 9 – MAP curve

The Slope curve is the slope of the MAP curve. Is used to determine the optimum depending on the method chosen in MEM options. For second derivative method, this graph shows the slope of slope curve.

Main title

Figure 10 – Slope curve

Evolution

Shows the evolution of some indicators across time. The indicators are: duration (Figure 11 – check & describe evolution) of the epidemic, start of the epidemic, epidemic percentage, pre-epidemic, post-epidemic and intensity thresholds.
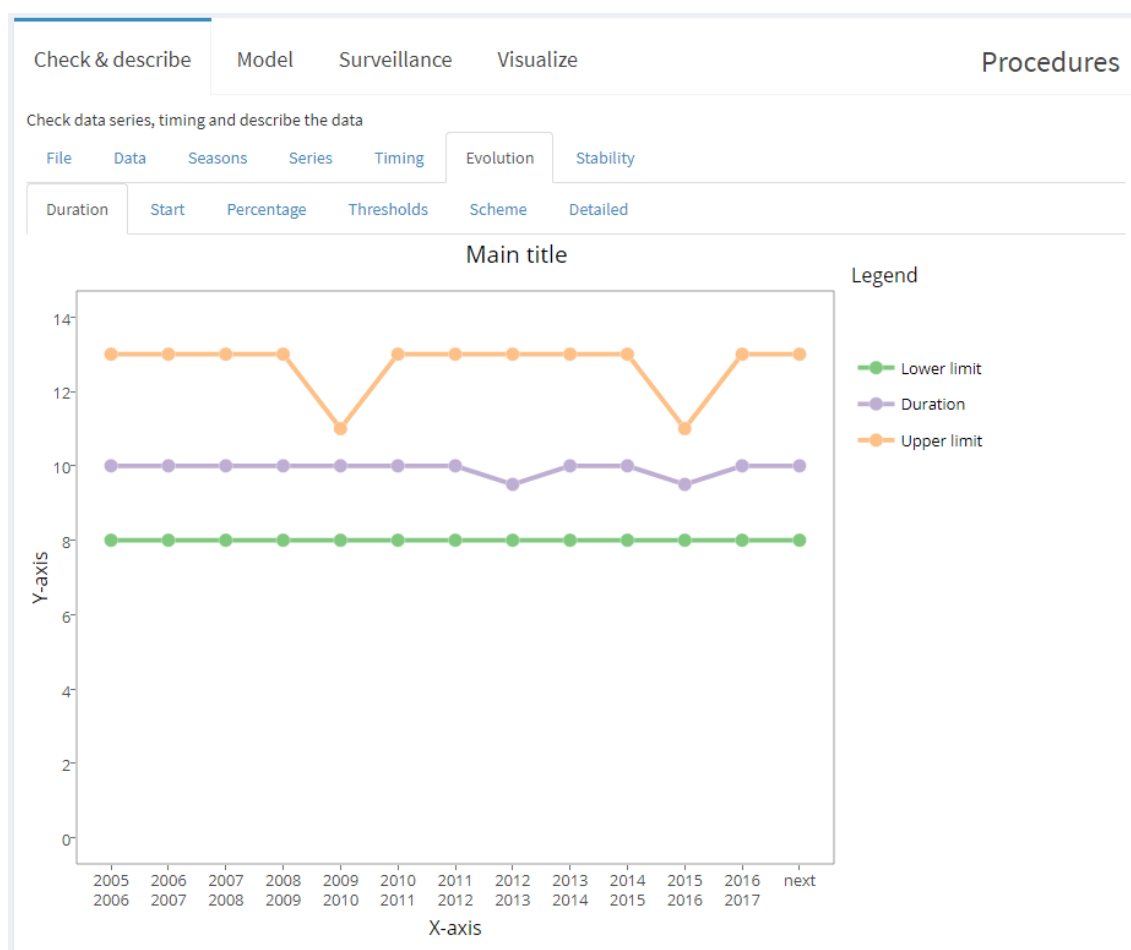
Figure 11 – check & describe evolution duration

Values per season are the estimates of the indicator for this specific season calculated according the MEM options (Figure 40 – MEM options box): cross or sequential (validation option), maximum number of seasons to use in calcula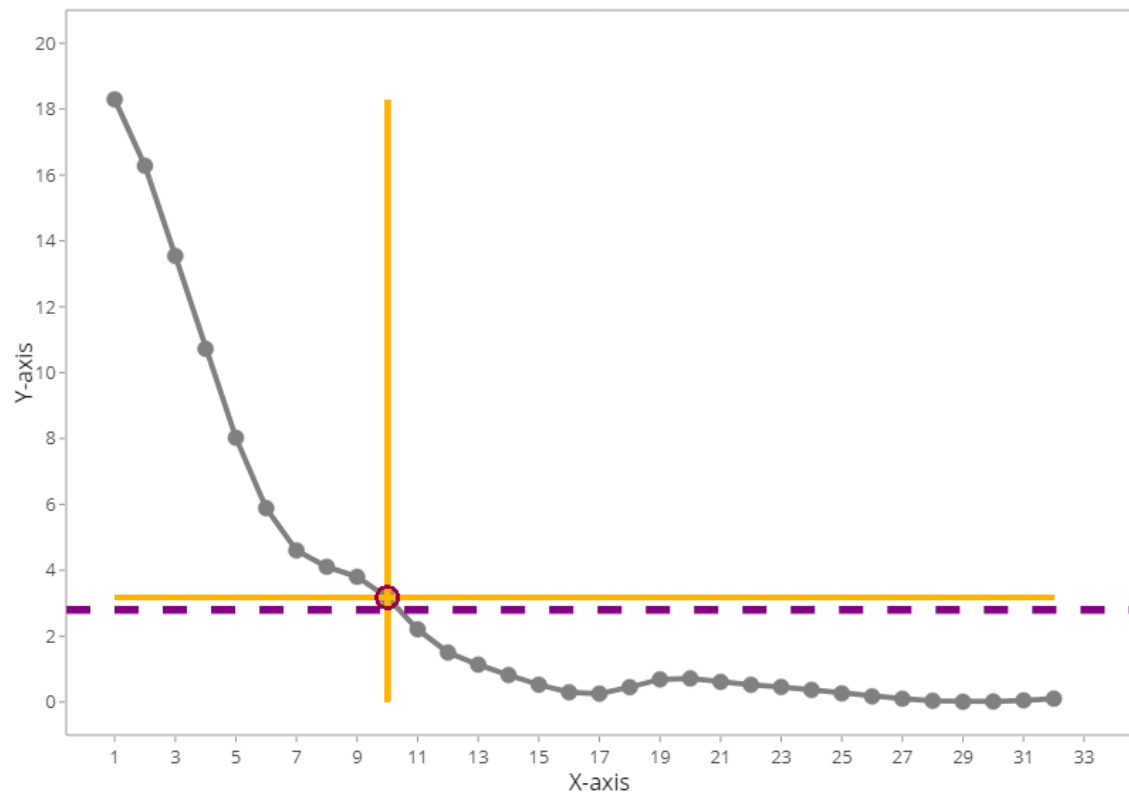tions, and the selected seasons in the Model options in the left panel (Figure 19 – model box). In MEM options is it also possible to select the method used to calculate indicators and the level of confidence of the confidence intervals.

To clarify how it is calculated, the tab scheme (Figure 12 – check & describe evolution scheme (cross)) shows for each x-axis point the seasons selected to calculate thresholds. Each row is a different season and in each column it is shown if that specific season has been used to calculate the indicators. If Yes, it has been used, if No, it hasn't.

In the last row, called Next, the indicators are calculated for the upcoming next season.

Check & describe   Model   Surveillance   Visualize

Check data series, timing and describe the data

File   Data   Seasons   Series   Timing   **Evolution**   Stability

Duration   Start   Percentage   Thresholds   **Scheme**   Detailed

	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
2005/2006	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2006/2007	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2007/2008	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2008/2009	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2009/2010	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2010/2011	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
2011/2012	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
2012/2013	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes
2013/2014	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes	✔ Yes
2014/2015	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes	✔ Yes
2015/2016	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✔ Yes
2016/2017	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
next	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes

Figure 12 – check & describe evolution scheme (cross)

Note that if you select sequential validation (Figure 13 – check & describe evolution scheme (sequential)), first points in the graphs calculate thresholds using less data (there are less seasons before the current data than the Maximum seasons value), and this will affect the range of confidence intervals (see 4.4).

Check & describe   Model   Surveillance   Visualize

Check data series, timing and describe the data

File   Data   Seasons   Series   Timing   **Evolution**   Stability

Duration   Start   Percentage   Thresholds   **Scheme**   Detailed

	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
2007/2008	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No
2008/2009	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No
2009/2010	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No
2010/2011	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No
2011/2012	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No
2012/2013	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No	✖ No
2013/2014	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No	✖ No
2014/2015	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No	✖ No
2015/2016	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No	✖ No
2016/2017	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✖ No
next	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes

Figure 13 – check & describe evolution scheme (sequential)

Example: For season 2010/2011 in the cross validation all seasons around 2010/2011 is used, from 2005/2006 to 2009/2010 and from 2011/2012 to 2015/2016). Since the maximum seasons is set to 10, it won't get more seasons to calculate the indicators. In the sequential validation, only preceding seasons are used: from 2005/2006 to 2009/2010, so only 5 seasons are used instead of the maximum number of 10 seasons allowed. Note that in cross validation always search for the closest seasons around the one selected. For the 2010/2011, there are 5 seasons

[26]

to the right and 5 to the left, but for season 2005/2006 uses only seasons to the right of it, because there are no seasons to the left. Also note that for sequential validation, first two seasons in the dataset are discarded (there are no rows for 2005/2006 and 2006/2007) since computation requires at least two seasons, and there is no enough data to the left of those seasons to calculate the indicators.

There are four plots for different indicators:

- Duration: Average duration of the epidemic and its confidence interval.
- Start: Average start of the epidemic and its confidence interval.
- Percentage: the epidemic percentage. The sum of the values (cases/rates) in the epidemic period divided by the total sum of values of the whole surveillance period. It's a coverage percentage of the epidemic period. And its confidence interval.
- Thresholds: pre-epidemic, post-epidemic and intensity thresholds.

The tab Details (Figure 14 – check & describe evolution detailed) shows the detailed information of each season calculations: number of seasons used, indicators and their confidence intervals and thresholds.

	Seasons	Duration (lower limit)	Duration	Duration (upper limit)	Start (lower limit)	Start	Start (upper limit)	Epidemic percentage (lower limit)	Epidemic percentage
2005/2006	10.00	8.00	10.00	13.00	9.00	24.00	26.00	78.42	86.00
2006/2007	10.00	8.00	10.00	13.00	9.00	24.00	27.00	78.42	85.30
2007/2008	10.00	8.00	10.00	13.00	9.00	25.00	27.00	78.42	86.12
2008/2009	10.00	8.00	10.00	13.00	9.00	25.00	27.00	81.89	86.12
2009/2010	10.00	8.00	10.00	11.00	21.00	25.00	27.00	78.42	85.30
2010/2011	10.00	8.00	10.00	13.00	9.00	25.00	27.00	78.42	86.12
2011/2012	10.00	8.00	10.00	13.00	9.00	22.00	26.00	78.42	86.04
2012/2013	10.00	8.00	9.50	13.00	9.00	22.00	26.00	78.42	85.21
2013/2014	10.00	8.00	10.00	13.00	9.00	24.00	26.00	78.42	85.21
2014/2015	10.00	8.00	10.00	13.00	9.00	22.00	26.00	78.42	85.21
2015/2016	10.00	8.00	9.50	11.00	9.00	22.00	26.00	78.42	86.04
2016/2017	10.00	8.00	10.00	13.00	9.00	24.00	26.00	78.42	86.00
next	10.00	8.00	10.00	13.00	9.00	22.00	26.00	78.42	85.21

Figure 14 – check & describe evolution detailed

The confidence intervals for these indicators are calculated using the Other CI. and Average curve/Other CI. level parameters (Figure 40 – MEM options box).

Shows the stability of the main indicators (Figure 15 – check & describe stability thresholds) according the number of seasons used to calculate them.

That allows to decide the minimum and maximum number of seasons to be included in the model for this specific country and parameter.
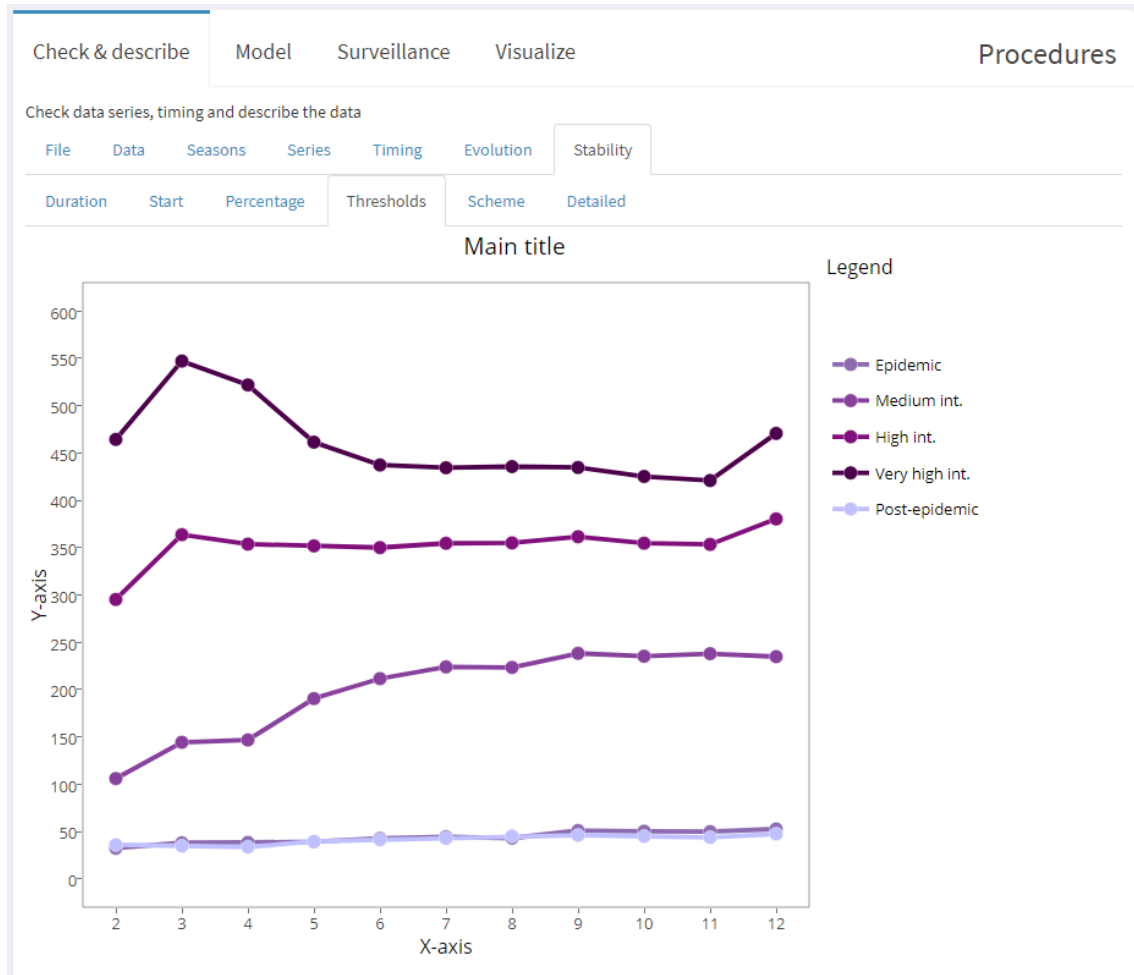

Figure 15 – check & describe stability thresholds

These indicators are calculated for the next upcoming season, so for all the iterations are used the last number of seasons determined by the x-axis, so that 2 means "last two seasons", 10 means "last ten seasons" and so on.

For stability there is no need to specify a validation option (it is always sequential) or a maximum number of seasons since our goal is to see what's the progress of the estimators when adding more seasons to the calculations up to the maximum of the dataset.

The tab Scheme (Figure 16 – check & describe stability scheme) shows data used to calculate each particular point in the graph.

	2005/2006	2006/2007	2007/2008	2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017
2	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes
3	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes
4	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes
5	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
6	✖ No	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
7	✖ No	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
8	✖ No	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
9	✖ No	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
10	✖ No	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
11	✖ No	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes
12	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes	✔ Yes

Figure 16 – check & describe stability scheme

Example: When 2 seasons are used (row labelled as 2), the program uses the last ones: 2015/2016, when 5 seasons are used (row labelled as 5), the program uses 2012/2013 to 2015/2016, when 12 seasons are used (row labelled as 12), all dataset is used.

The tabs are the same as the Evolution tab, including the Details tab (Figure 17 – check & describe stability detailed):

Figure 17 – check & describe stability detailed

Goodness (advanced feature)

This tab is exactly the same as the one in the Model tab. The difference is that the goodness of Check & Describe uses all data from the dataset and the one at the Model uses only the columns selected for the model thus specified in the Model box (From/To/Exclude/Maximum seasons).

Figure 18 – check & describe goodness

Refer to section 3.2.2 to learn about goodness tab.

3.2.2 Model

Shows a summary of the model, graphs, goodness and optimization of the MEM model. The model box (Figure 19 – model box) of the left panel selects the columns that will be used in calculations.
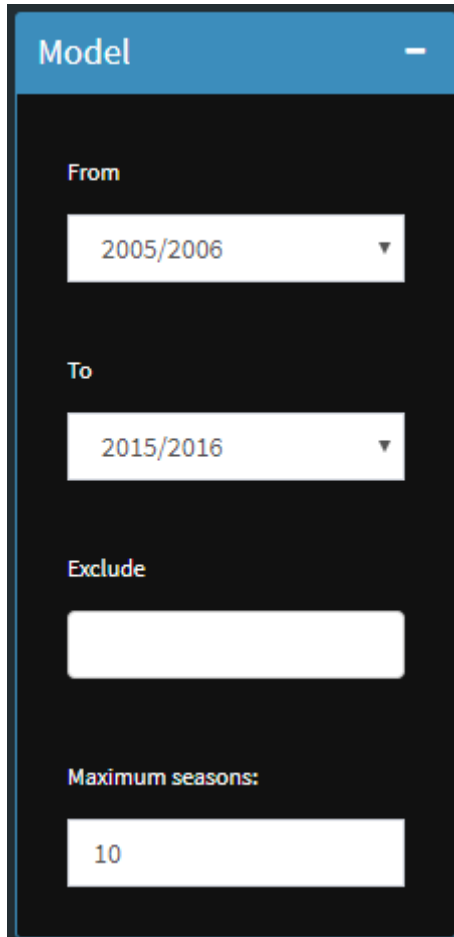
Figure 19 – model box

To select the columns to model, select the first season in **From** and the last season in **To**. If you want to exclude any season from the model, select it in **Exclude**. Finally, use the option **Maximum seasons** to limit the seasons selected to a number of seasons. If this number is lower than the column selection with From/To/Exclude, the application will take the last Maximum seasons of that selection.

If the current selection with From/To/Exclude has less seasons than the Maximum seasons parameter, then it won't affect the selection.

Data, seasons, series and timing

These tabs are equal of those described in the Check & describe procedure, except that in Model only the selected seasons of the model box are shown.

Mem

Calculate the estimators using MEM with data provided by the Model box.

See more on MEM options in section 4.1.

Estimators

Shows the number of seasons used in the model, the average epidemic start week, the average epidemic length, the epidemic percentage (percentage of the total cases in the epidemic period) and the epidemic and intensity thresholds (Figure 20 – model MEM).
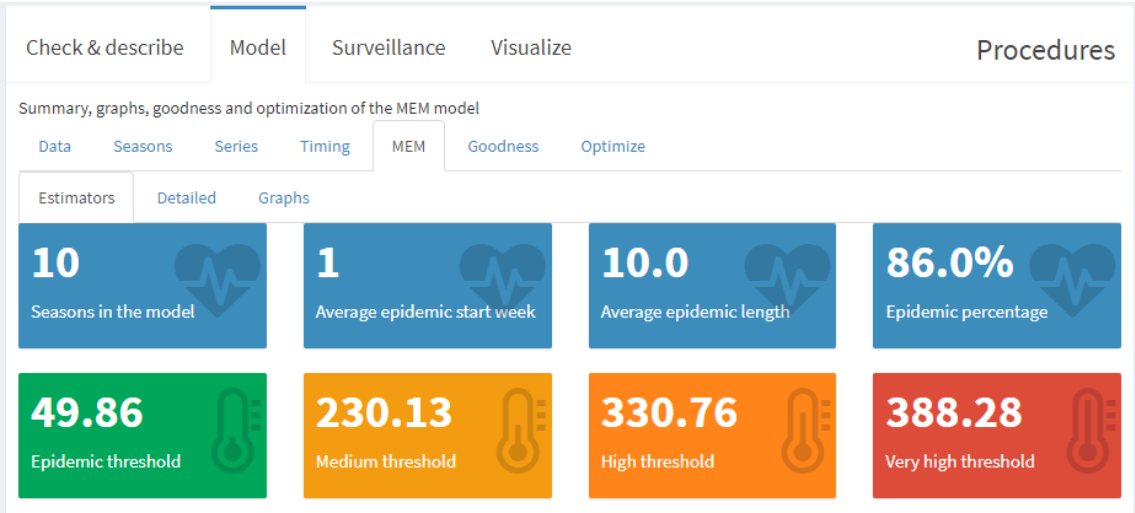
Figure 20 – model MEM estimators

Detailed

Detailed information of the mem output, including the parameters used and the outputs.
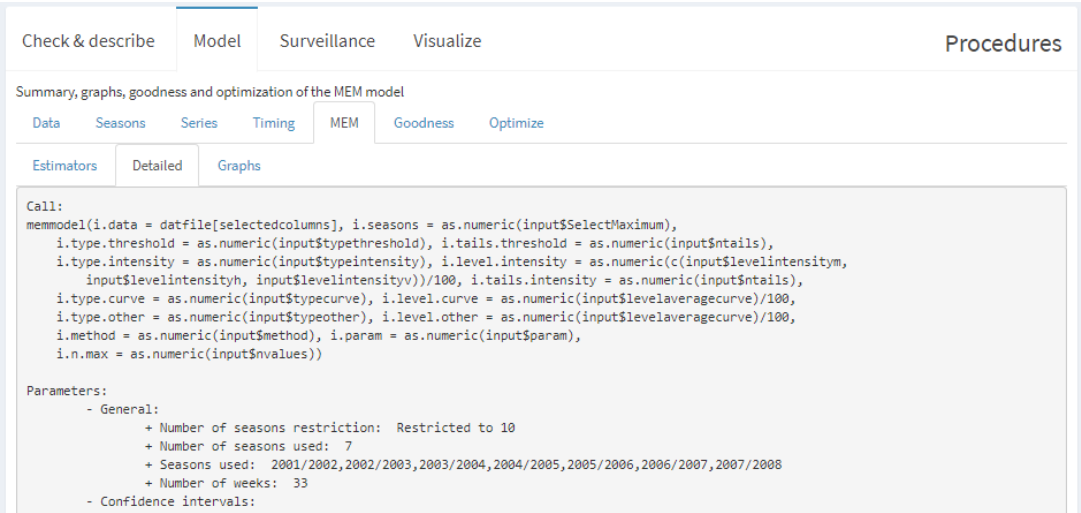
Figure 21 – model MEM detailed information

Graphs

Shows two graphs, the Moving epidemics graph (Figure 22 – model MEM moving epidemics), which shows all the seasons in the model centred on their epidemic periods, which the average start and duration.
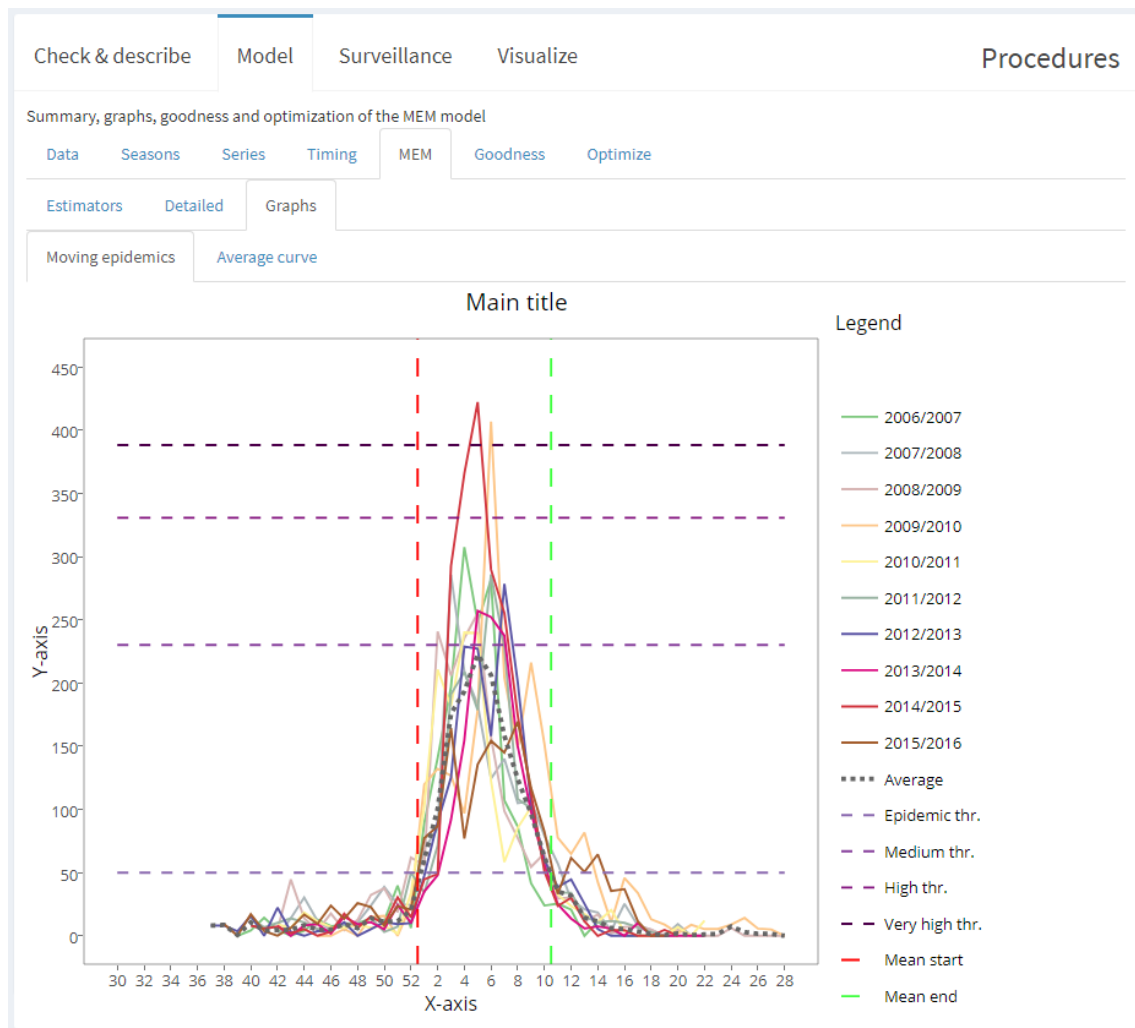
Figure 22 – model MEM moving epidemics

And the Average curve graph (Figure 23 – model MEM average curve), which shows the average epidemic along with the thresholds and the average start and duration.
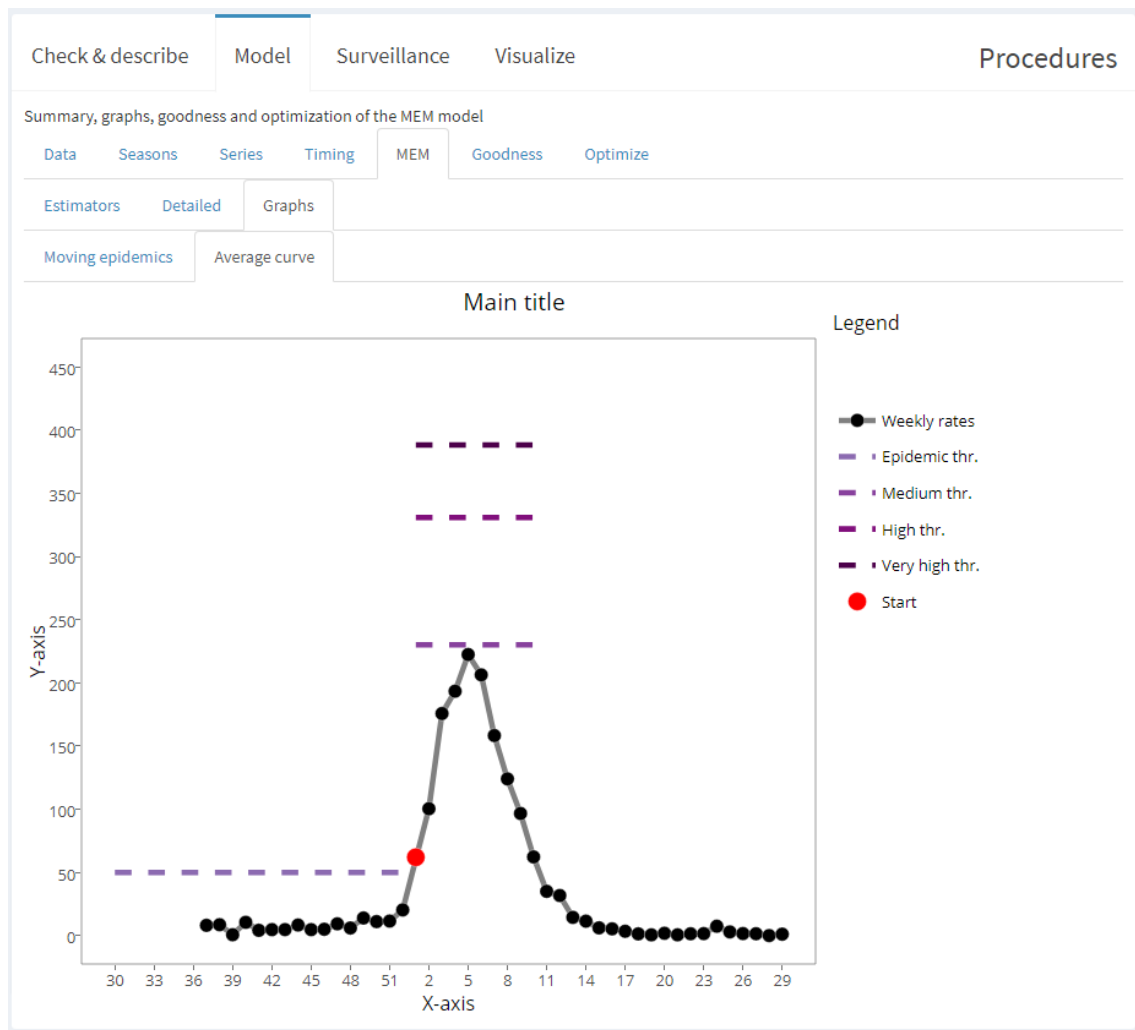
Figure 23 – model MEM average curve

Goodness

Calculate different indicators related to the goodness of the MEM for detecting the epidemics, using data from the model selected.

The Indicators tab (Figure 24 – model goodness indicators) shows sensitivity, specificity, positive predictive value, negative predictive value, percent agreement and the Matthews correlation coefficient.
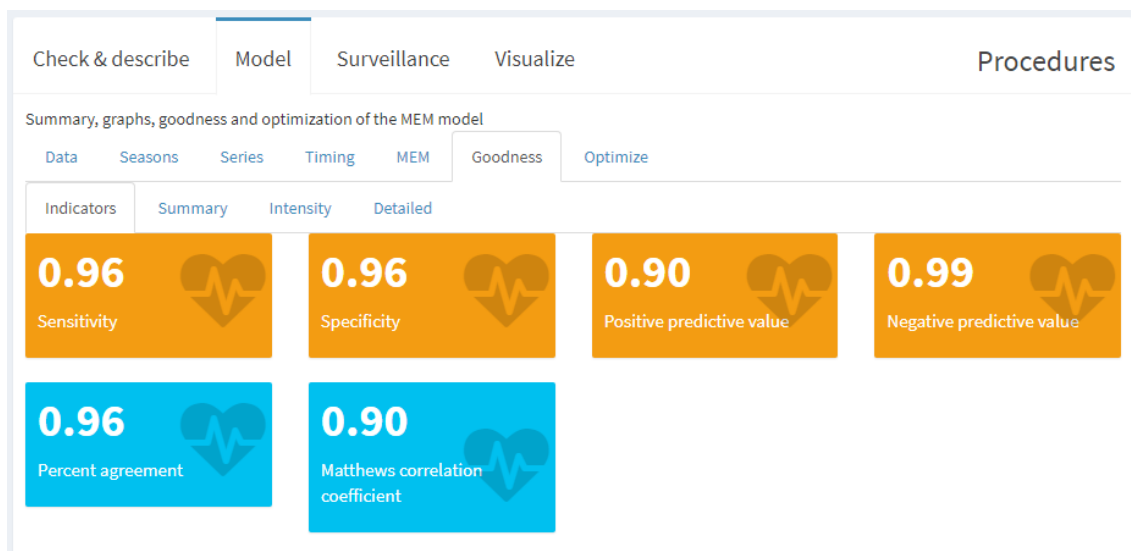
Figure 24 – model goodness indicators

When advanced features is activated, a new value box in the intensity tab is shown.
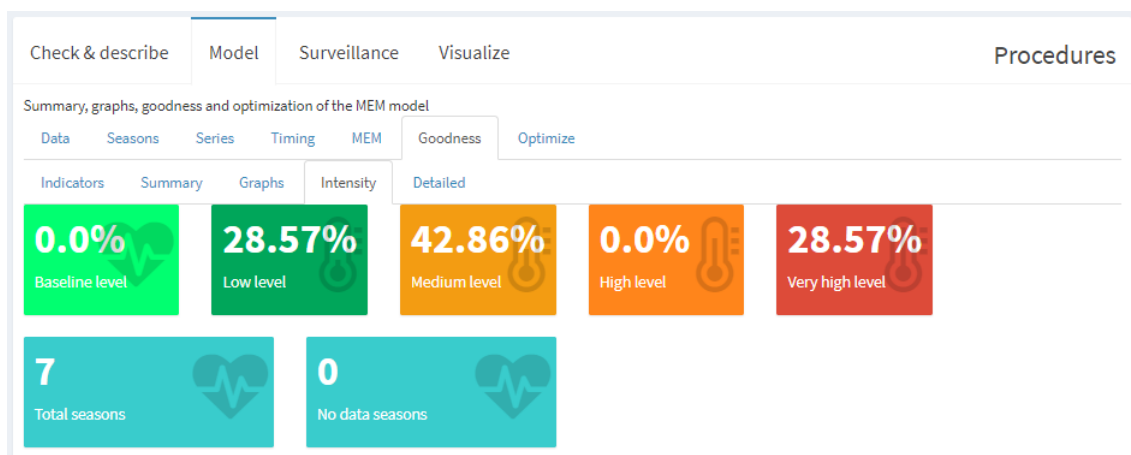
Figure 25 – no data seasons valuebox

The value **No data seasons** shows the number of seasons for which intensity was impossible to be calculated.

The summary tab (Figure 26 – model goodness summary) shows particular information of the above indicator for each iteration of the goodness process.

|	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Percent agreement	Matthews correlation coefficient
2006/2007	0.94	1.00	1.00	0.98	0.99	0.96
2007/2008	0.99	0.97	0.94	1.00	0.98	0.95
2008/2009	0.99	0.95	0.85	1.00	0.96	0.89
2009/2010	1.00	0.90	0.80	1.00	0.93	0.85
2010/2011	0.97	0.99	0.99	0.99	0.99	0.97
2011/2012	0.95	0.95	0.89	0.98	0.95	0.89
2012/2013	0.93	0.95	0.89	0.97	0.95	0.87
2013/2014	0.92	1.00	1.00	0.97	0.98	0.94
2014/2015	1.00	0.95	0.86	1.00	0.96	0.90
2015/2016	0.93	0.94	0.89	0.96	0.93	0.86
Total	0.96	0.96	0.90	0.99	0.96	0.90

Figure 26 – model goodness summary

When advanced features is activated another tab is shown under Goodness: graphs.

This tab allows to explore how each season performs when calculating the goodness indicators.

For each season, the real epidemic is determined using the mem algorithm. The epidemic is marked in purple in the plot and in the "real" line as rectangles of different colours, green and yellow for pre and post-epidemic. The observed epidemic is also coloured in purple/green/yellow in the "default" line. See section 4.4 for more information.
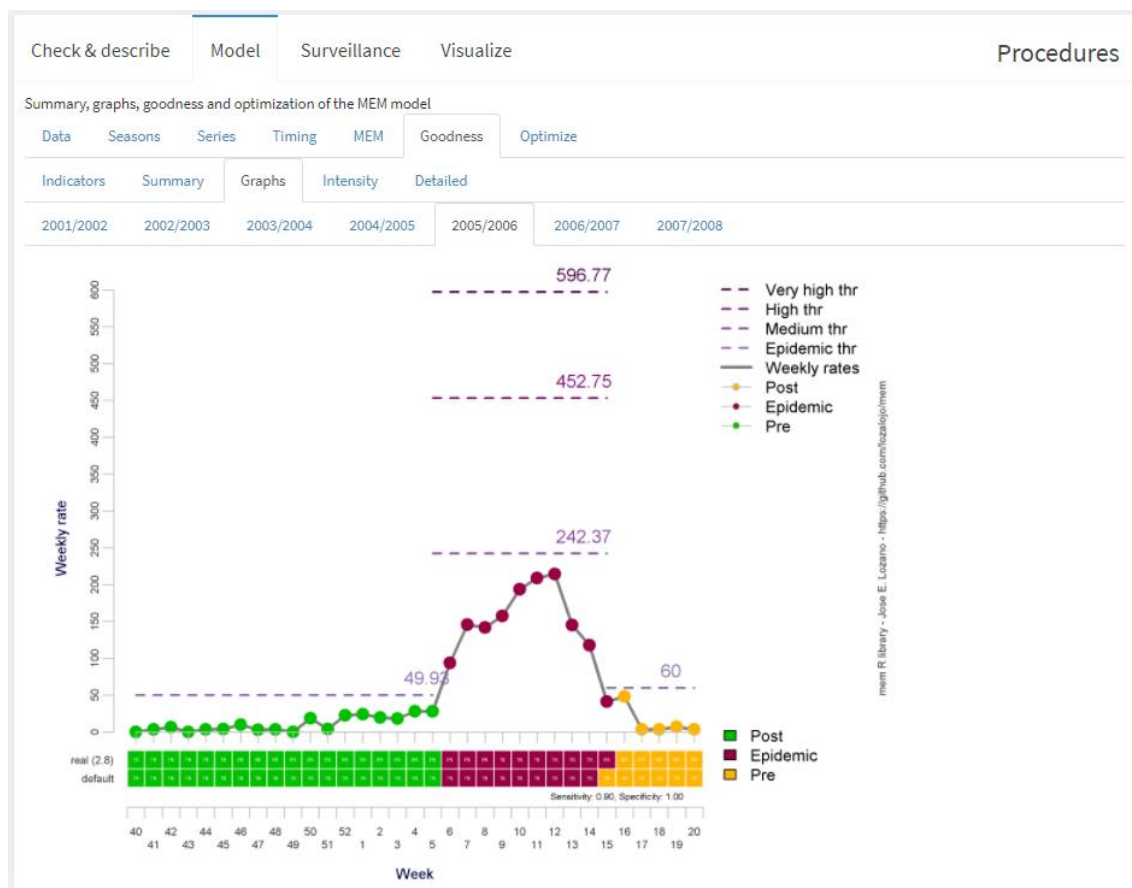
Figure 27 – goodness graphs

The Intensity tab (Figure 28 – model goodness intensity) shows the distribution of peaks intensity levels, the number of seasons used and the number of seasons for which has been impossible to calculate the intensity level.
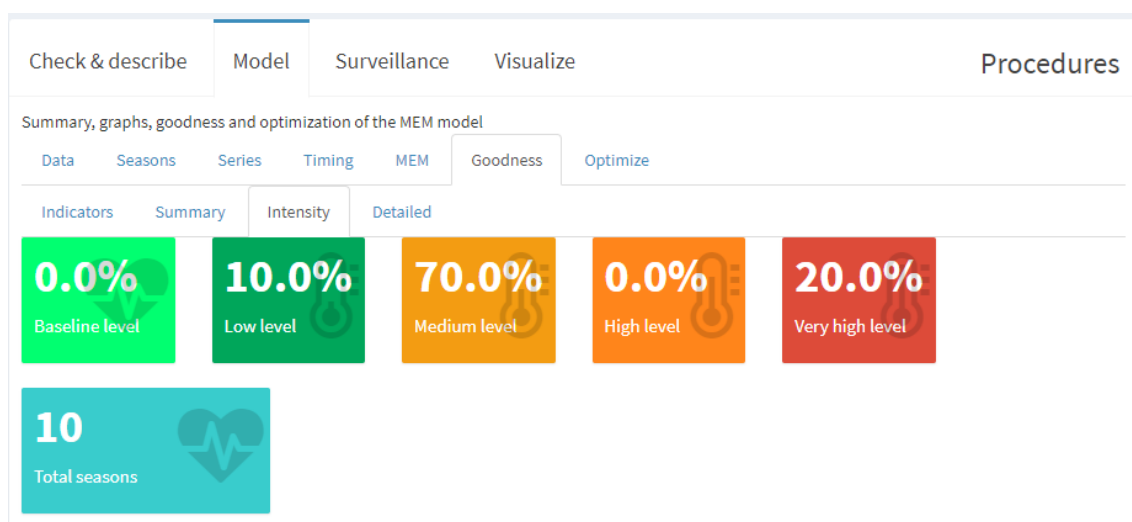
Figure 28 – model goodness intensity

The detailed tab (Figure 29 – model goodness detailed) shows information of the indicators shown above for each season in the model, the peak, the thresholds and the level reached.

Figure 29 – model goodness detailed

See section 4.4 to know how goodness is calculated.

Optimize

Optimize is an iterative process that calculates goodness indicators using different slope parameters for the fixed criterium and compares all estimators in order to find the optimum slope parameter.

The indicators tab (Figure 30 – model optimize indicators) shows the goodness indicators of the best slope parameter found.

Figure 30 – model optimize indicators

The Detailed tab (Figure 31 – model optimize detailed) shows the indicators of each slope parameter tested.

[38]

Check & describe   Model   Surveillance   Visualize

Summary, graphs, goodness and optimization of the MEM model

Data   Seasons   Series   Timing   MEM   Goodness   Optimize

Indicators   Detailed   Graphs

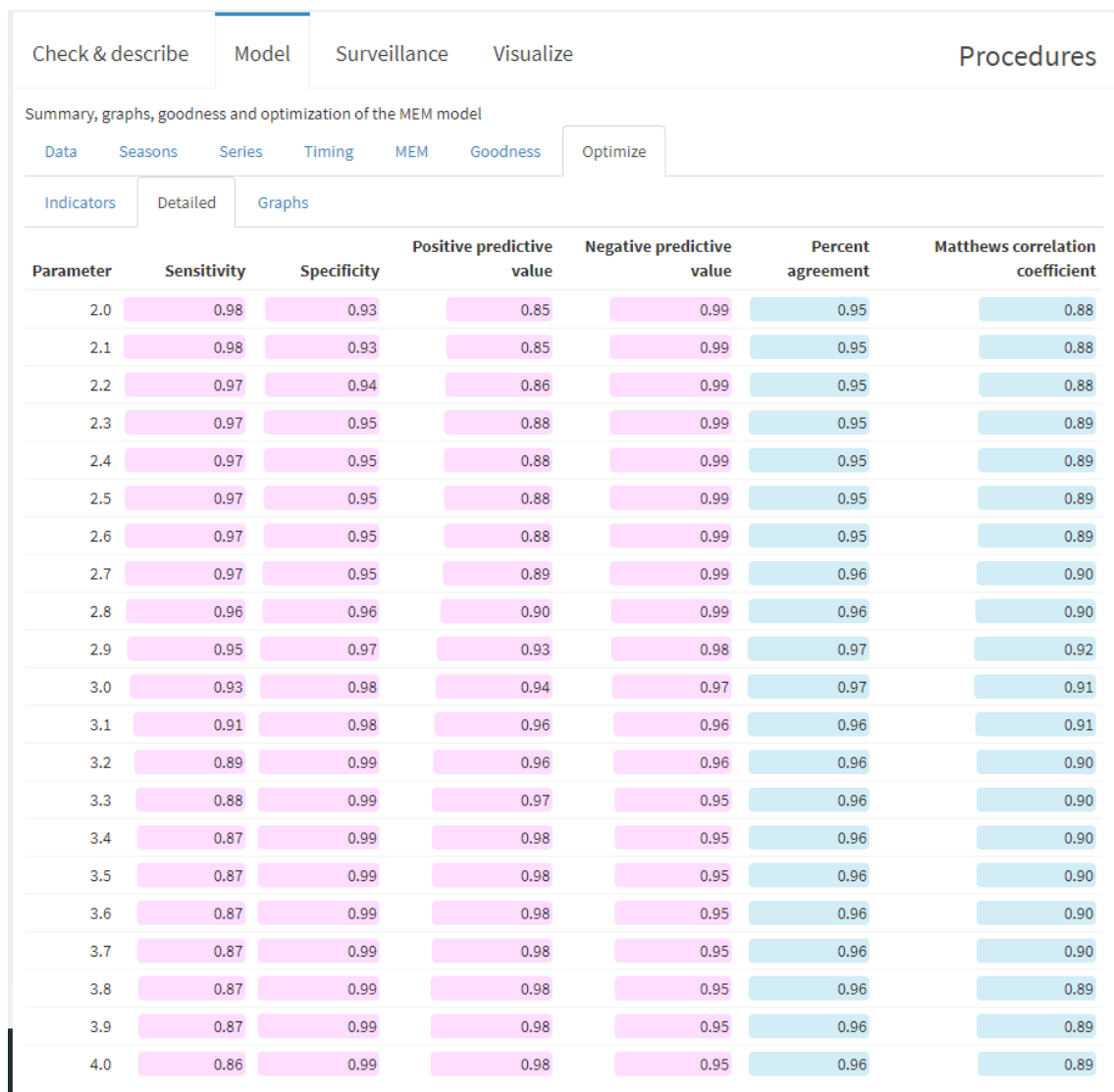| Parameter	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Percent agreement	Matthews correlation coefficient
2.0	0.98	0.93	0.85	0.99	0.95	0.88
2.1	0.98	0.93	0.85	0.99	0.95	0.88
2.2	0.97	0.94	0.86	0.99	0.95	0.88
2.3	0.97	0.95	0.88	0.99	0.95	0.89
2.4	0.97	0.95	0.88	0.99	0.95	0.89
2.5	0.97	0.95	0.88	0.99	0.95	0.89
2.6	0.97	0.95	0.88	0.99	0.95	0.89
2.7	0.97	0.95	0.89	0.99	0.96	0.90
2.8	0.96	0.96	0.90	0.99	0.96	0.90
2.9	0.95	0.97	0.93	0.98	0.97	0.92
3.0	0.93	0.98	0.94	0.97	0.97	0.91
3.1	0.91	0.98	0.96	0.96	0.96	0.91
3.2	0.89	0.99	0.96	0.96	0.96	0.90
3.3	0.88	0.99	0.97	0.95	0.96	0.90
3.4	0.87	0.99	0.98	0.95	0.96	0.90
3.5	0.87	0.99	0.98	0.95	0.96	0.90
3.6	0.87	0.99	0.98	0.95	0.96	0.90
3.7	0.87	0.99	0.98	0.95	0.96	0.90
3.8	0.87	0.99	0.98	0.95	0.96	0.89
3.9	0.87	0.99	0.98	0.95	0.96	0.89
4.0	0.86	0.99	0.98	0.95	0.96	0.89

Figure 31 – model optimize detailed

The Graph tab (Figure 32 – model optimize graphs) shows a plot of the indicators for each slope parameter value (in the x-axis).
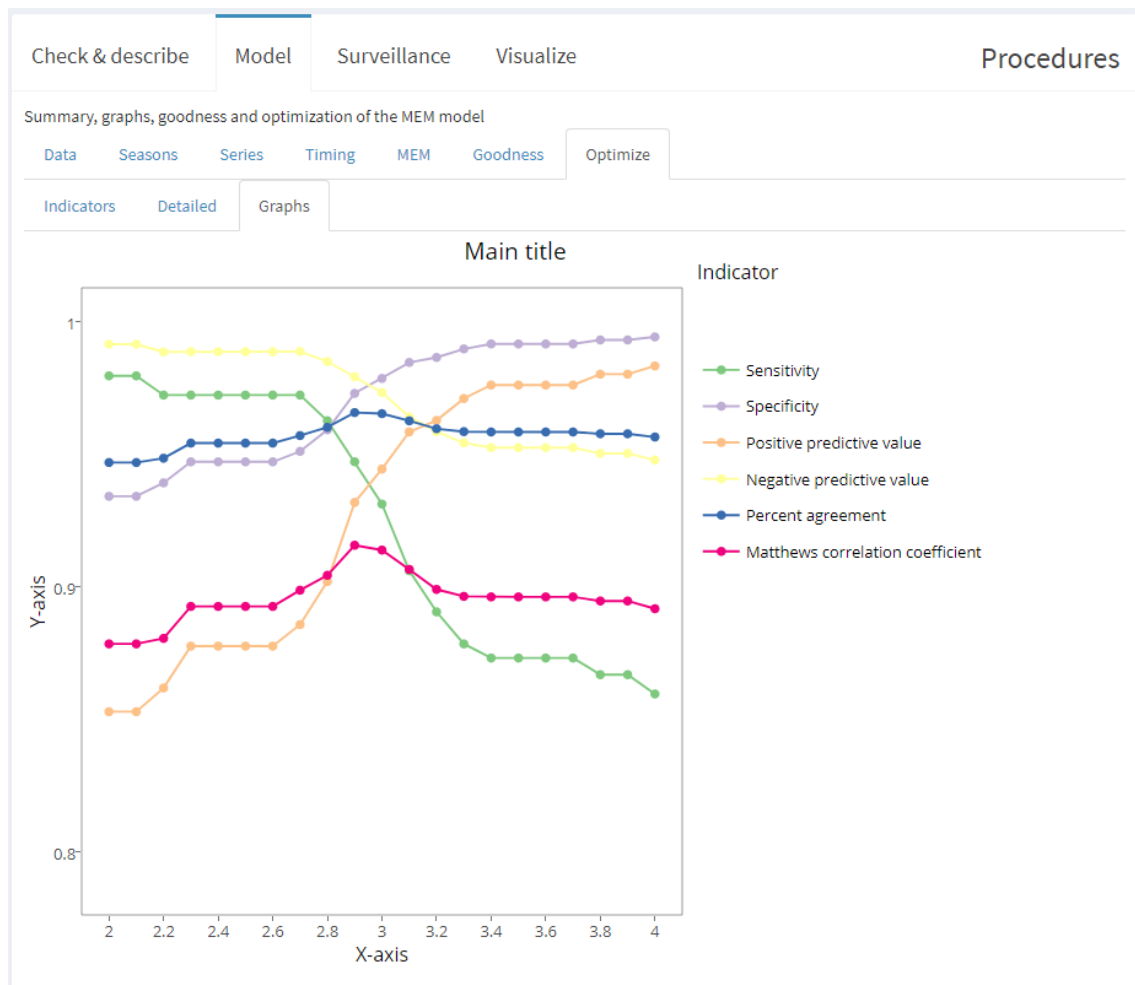
Figure 32 – model optimize graphs

When advanced features is activated, another optimization method is allowed: Manual optimization.

This procedure searches the slope parameter in the range defined by **Parameter range** (under Goodness & optimize) which match better with an observer prediction by inspection.

The procedure is as follows:

1. The observer determines the position of the epidemic by inspection, clicking on each plot two times, one for the first point of the epidemic and another one for the last point of the epidemic. The point just clicked will be coloured in red for start and green for end (Figure 33 – manual optimize, epidemic location by inspection). When there are more than two clicks, the last two ones are chosen and coloured according to its position (the first one is the starting point of the epidemic, the second one the ending point).
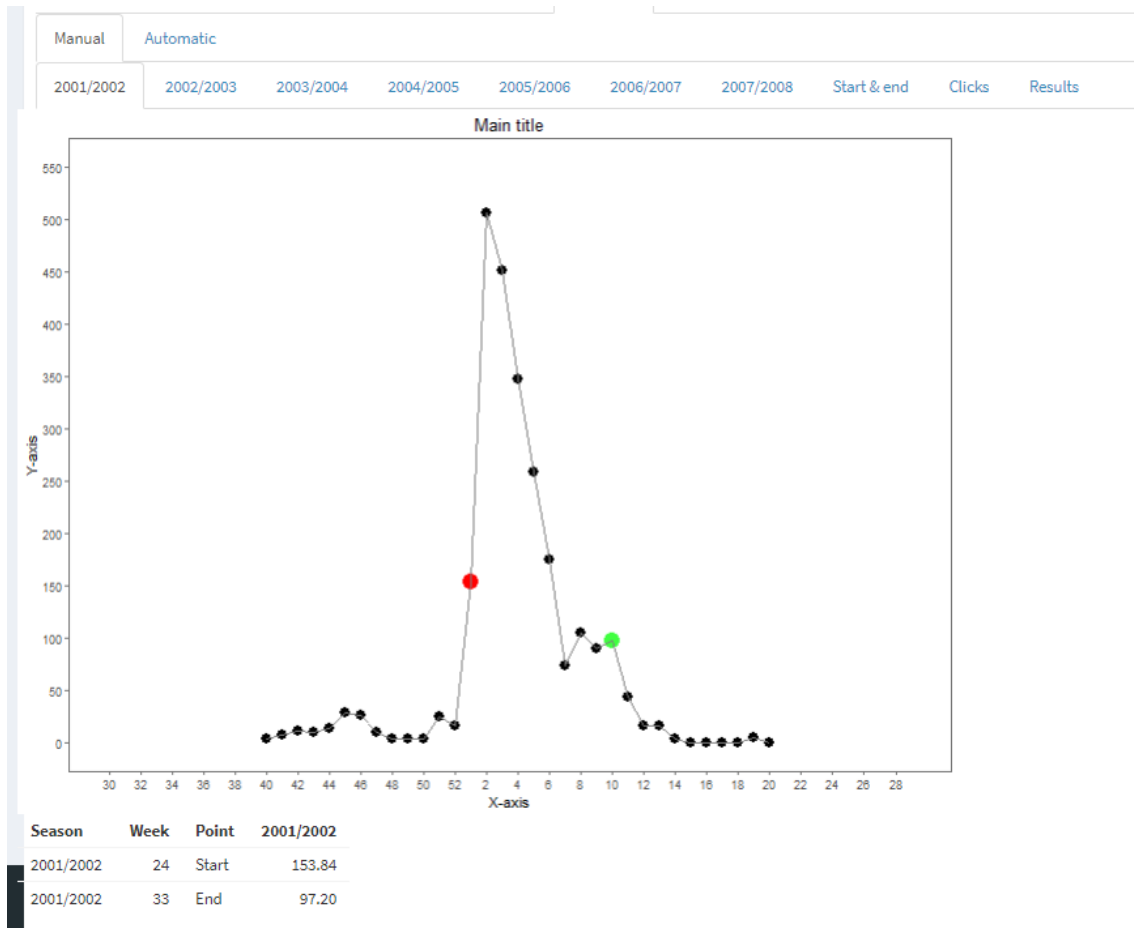
Figure 33 – manual optimize, epidemic location by inspection

2. The process is repeated for every season (in a different tab) of the whole dataset.
3. The **Start & End** tab shows the current selection of starting and ending points of all the epidemics in the dataset.
4. The **Clicks** tab shows all points that have been selected in the process.
5. When all epidemic location have been selected:
   a. The **Results** tab will match all possible slope parameters with the epidemics determined by inspection and will show different statistics for each parameter and will choose between all the parameters tested using the criterium selected at **Optimization method**. The value at **Optimum parameter** (Figure 34 – manual optimize, results with optimumm) will show the slope parameter closest to the selection made by the observer.
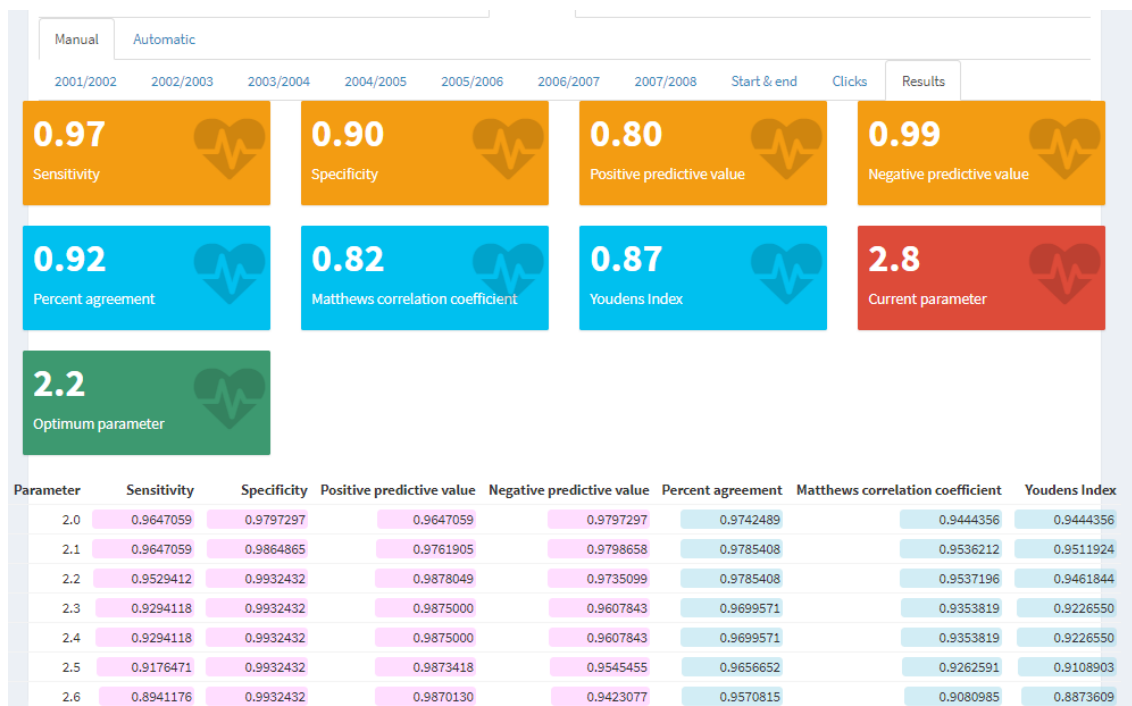
Figure 34 – manual optimize, results with optimumm

b. Tabs with plots for each season will be updated with a graph at the bottom (Figure 35 – manual optimize, comparison of optimum and inspection) showing the correspondence between the inspection and the MEM with the optimum slope parameter just calculated.

Season	Week	Point	2001/2002
2001/2002	24	Start	153.84
2001/2002	33	End	97.20
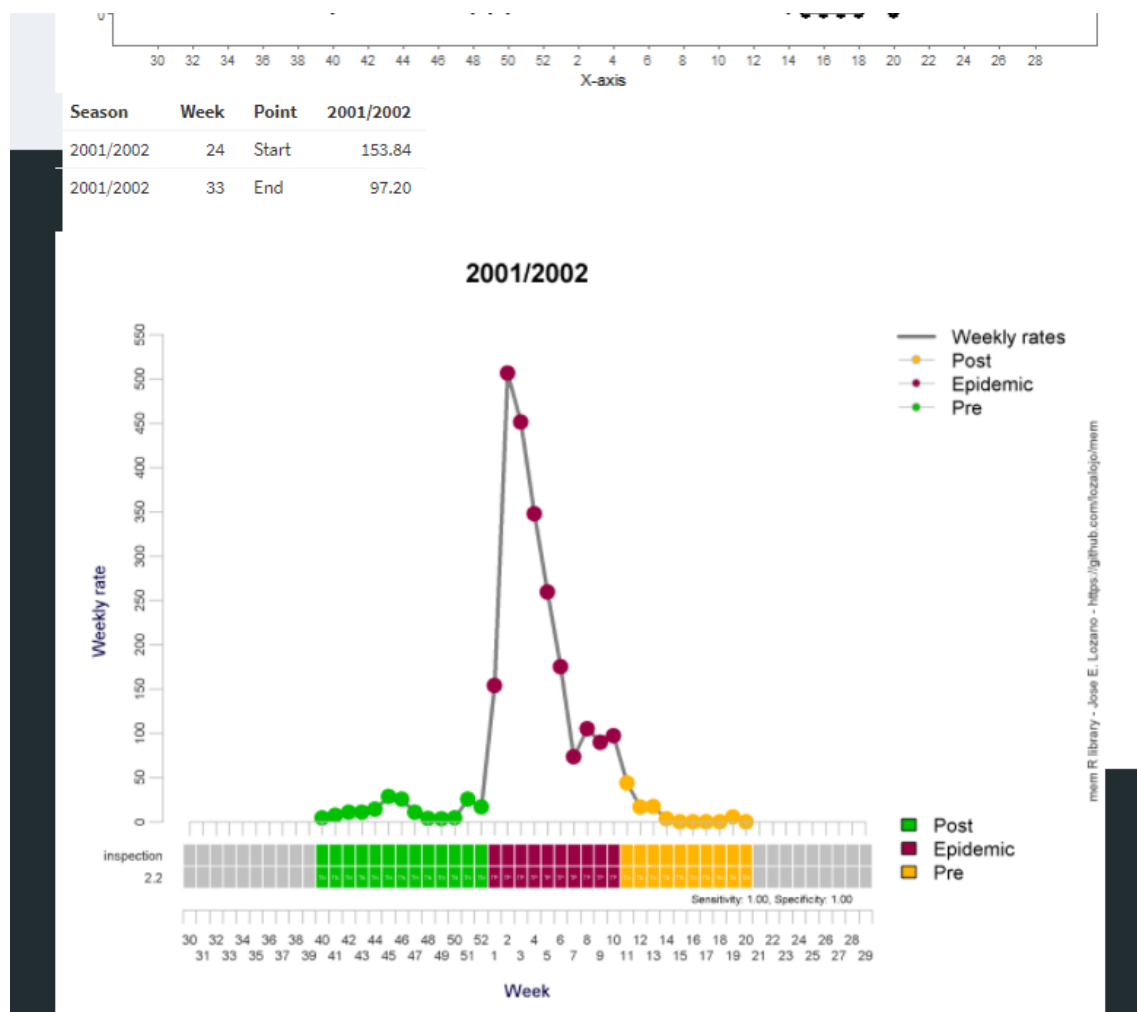
Figure 35 – manual optimize, comparison of optimum and inspection

3.2.3   Surveillance

Using the seasons selected in the Model box and the options selected in the MEM options box, the thresholds are applied to the season selected in the Surveillance box of the left panel and simulates how surveillance will look like at a given week.

[43]

Figure 36 – surveillance box

Select the season to use in the **Surveillance** combo box (Figure 36 – surveillance box). Only one season can be selected.

Surveillance consist on:

1. Draw weekly values along with the pre-epidemic threshold.
2. When the weekly value rises above the threshold a marker of the start of the epidemic is placed and the intensity thresholds are drawn.
3. When the weekly value goes down the post-epidemic threshold, the marker of the end of the epidemic is placed and the post-epidemic threshold is added to the graph.

The **Surveillance Week** allows select the week to use in the surveillance, the values of the surveillance season are shown only up to this week, and the program will ignore values past this week.

The **Force epidemic start** allows to force the placement of the epidemic start marker at a given week instead of using the first week above the epidemic threshold.

This graph is affected by the Thresholds box (Figure 42 – thresholds box), to plot the start and end markers and to plot the intensity thresholds in the graph.

The Week tab (Figure 37 – surveillance graph) shows the surveillance graph of the current selection of surveillance season, surveillance week, force start and model selection.
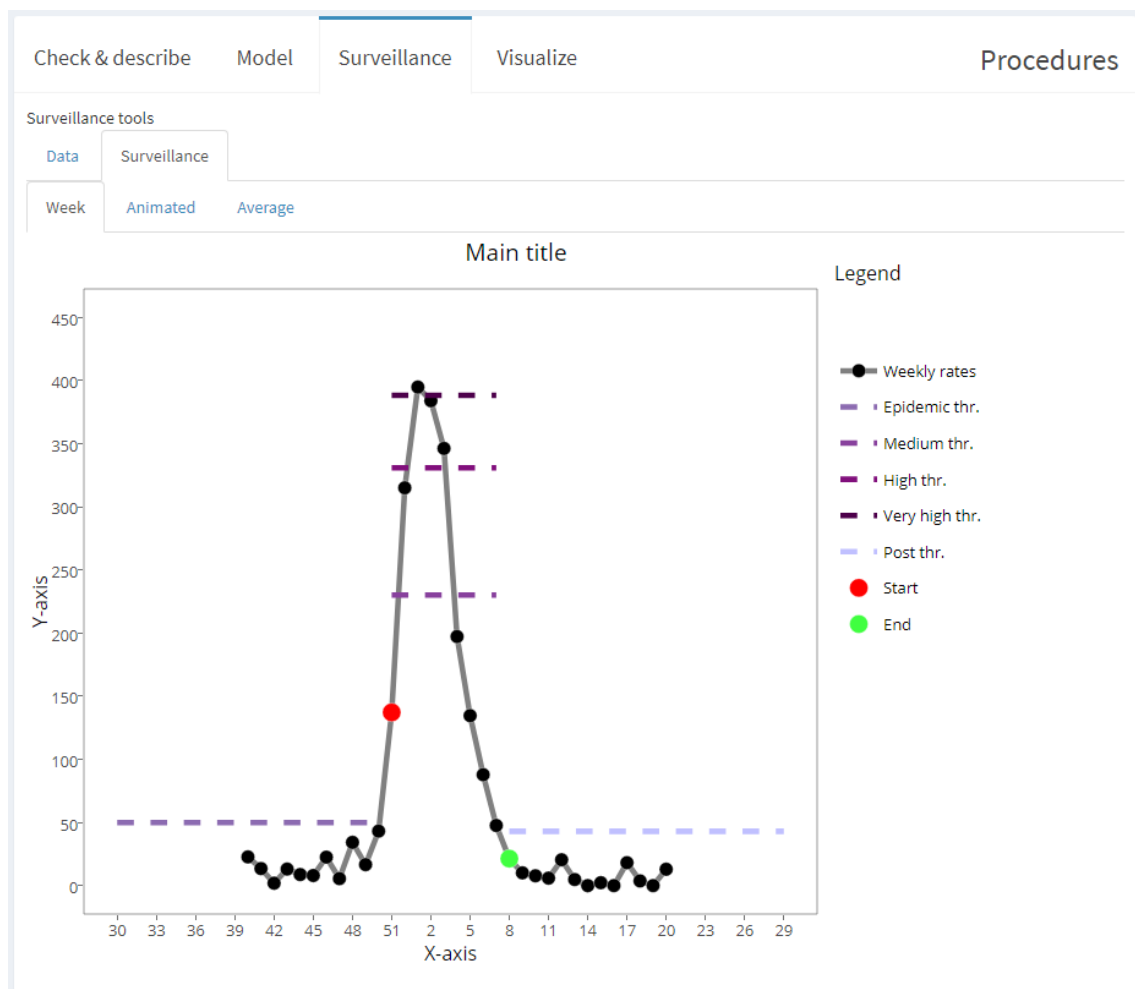
Figure 37 – surveillance graph

The Animated tab shows an animated graph of the surveillance to see the evolution of the epidemic in the current season using a model.

The Average tab (Figure 38 – surveillance average) shows the surveillance season with the average curve and the confidence limits.
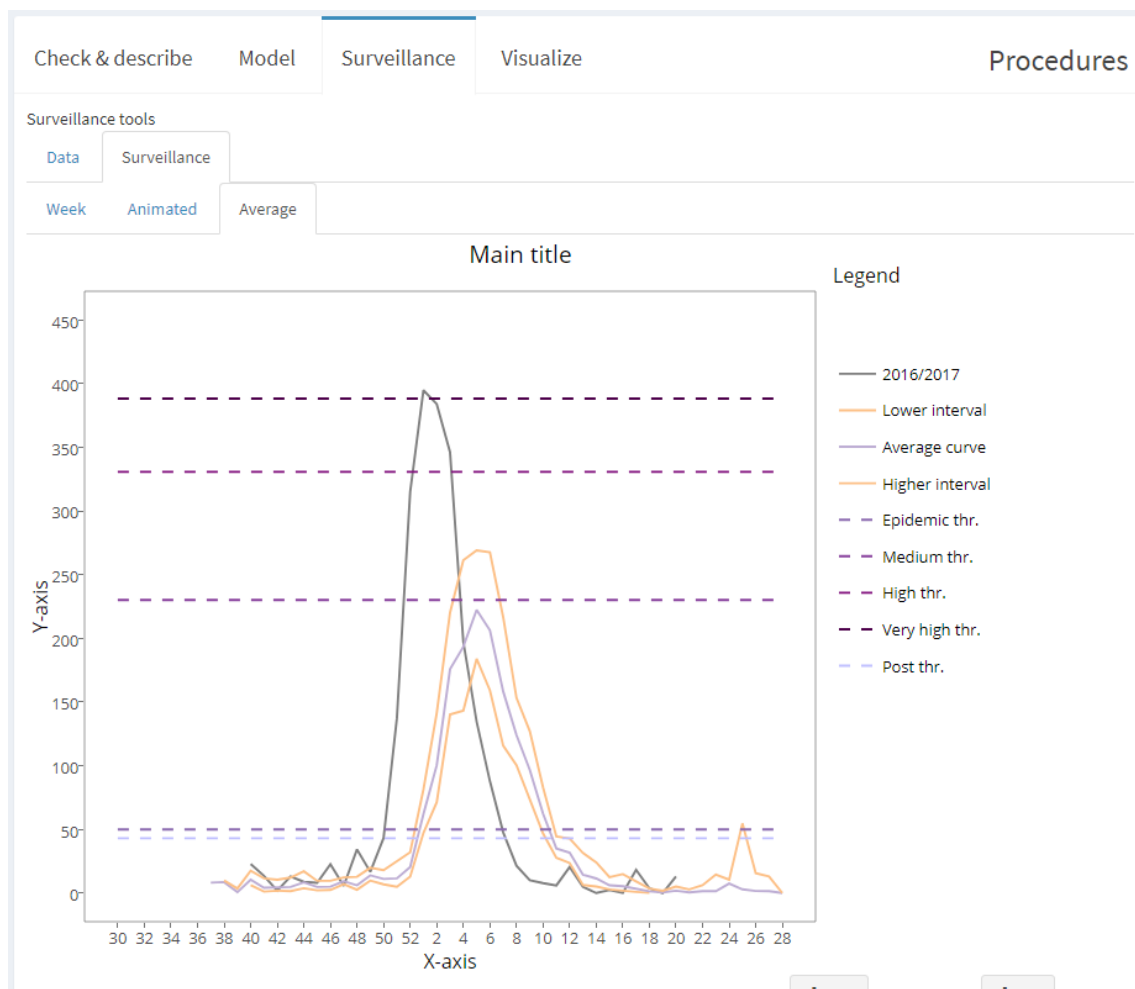
Figure 38 – surveillance average

3.2.4   Visualize

Visualize shows only a set of seasons, selected in the Visualize box (Figure 39 – visualize box) of the left panel.

Options are the same as previous tabs: Data to show the data of the selected seasons, Seasons to see the overlapped graph, Series to see the time series graph and Timing to see the timing of each season.
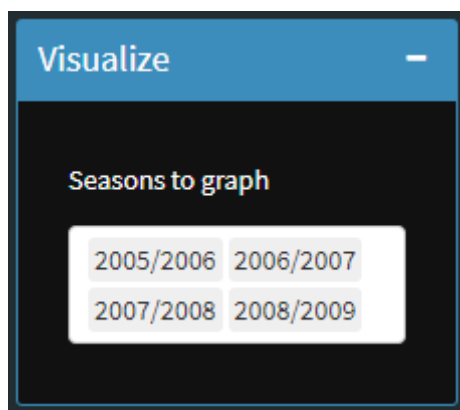
Figure 39 – visualize box

All the thresholds shown in graphs are calculated using the model selected in the Model box (Figure 19 – model box).

4 Other memapp options and details

4.1 MEM procedures options

The MEM options box (Figure 40 – MEM options box) in the right panel allows to configure different parameters used by MEM to calculate thresholds, goodness and optimize.

Figure 40 – MEM options box

4.1.1 Timing

- Method for timing: there are four methods in MEM to calculate the timing of the epidemic. The default method is the fixed criterium.
  - Original: uses the process shown in the original paper.
  - Fixed criterium: uses the slope of the MAP curve fo find the optimum, which is the point where the slope is lower than a predefined value.
 - Slope: calculates the slope of the MAP curve, but the optimum is the one that matches the global mean slope.
  - Second derivative: calculates the second derivative and equals to zero to search an inflexion point in the original curve.
- Slope parameter: the slope parameter used by the fixed criterium method. It will not affect the rest of methods.

4.1.2 Thresholds

- Values per season: Number of values per season to calculate epidemic and intensity thresholds.
- Tails: Use one or two tailed confidence intervals to calculate the epidemic and intensity thresholds.
- Epidemic thresholds: Method to calculate the epidemic (pre and post) thresholds.
  - Arithmetic mean and mean confidence interval: uses mean of the sample and the standard error of the mean to calculate CI.
  - Geometric mean and mean confidence interval: uses the geometric mean, and the standard error of the mean of the log-transformed data to calculate CI.
  - Median and Nyblom confidence interval: uses the median and the CI as described by Hettmansperger and Nyblom.
  - Median and bootstrap confidence interval: uses the median and the CI using resampling bootstrap.
  - Arithmetic mean and point confidence interval: uses mean of the sample and the standard deviation to calculate CI.
  - Geometric mean and point confidence interval: uses the geometric mean, and the standard deviation of the log-transformed data to calculate CI.
- Intensity thresholds: Method to calculate the intensity thresholds.
- Use t distribution (advanced feature): use t-value, coming from a t distribution instead of z-values from a normal distribution in all confidence intervals based on the mean (arithmetic and geometric based thresholds). It also uses a correction for low number of observations:

$$\hat{y} \pm t_{\alpha/2} s.\,e.\,p.\,(Y),$$

$$s.\,e.\,p\,(Y) = \sqrt{s^2 + \frac{s^2}{n}} = s\sqrt{1 + \frac{1}{n}}$$

- Levels: Levels of the confidence intervals used to calculate the intensity thresholds.

4.1.3 Goodness & optimize

- Validation: Used by Goodness and Evolution to determine which data to include in the model. Cross allows to use data after or before the season we are testing. Sequential

allows to use only data before the season that it is being tested. See section 4.4 to see how goodness is calculated using these two methods.

- Parameter range: Used by goodness to test the data with different timings using different parameters and averaging the results. Also this range is used by optimize to find the optimum slope parameter. It will only test values within this range.

4.1.4 Other

- Average curve CI: Method to calculate the average curve.
- Other CI: Method to calculate estimates for epidemic start, epidemic duration, epidemic percentage.
- Average curve/Other CI. level: level of the confidence intervals used for the average curve and other estimators.

4.2 Data details

4.2.1 Missing values and zero data

Be sure that a cell is blank when there is no data (missing value) and 0 when the data is 0 cases/rate. To include artificial 0 values is not recommended, it will affect thresholds. MEM automatically applies imputation techniques to the original data to fill the missing values.

4.2.2 Week 53

Since not all the seasons contain week 53rd, data must be adjusted in order to avoid gaps. When a season has a rate for 53rd week it must be pulled forward one week (Table 4 – week 53). So week 53 becomes week 1 of the following year, week 1 then turns to be week 2 and so on:

Original season	
Week No	2000/2001
50	112.1
51	150.4
52	170.7
53	210.2
1	270.0
2	280.0
3	160.2
4	120.4

Modified season	
Week No	2000/2001
50	112.1
51	150.4
52	170.7
1	210.2
2	270.0
3	280.0
4	160.2
5	120.4

Table 4 – week 53

Rates for week 53 must not be deleted in northern-like surveillance periods (two years each season), since epidemics occur in winter time, and it is probable that week 53 is part of the epidemic. Deleting one rate will shorten the epidemic period by one week and also would affect the estimators, which probably should use that weekly rate to calculate the thresholds.

In southern-like regions (Table 3 – southern hemisphere surveillance), week 53 is at the end of the surveillance period. There shouldn't be any epidemic activity there, so there is no problem on removing it. In case there is actually epidemic activity in week 53, please revise the declaration of the surveillance period for your region, since the end of the period (week 53) should not be set in the middle of an epidemic, and probably it is better to set seasons with two years, as in the northern hemisphere countries (Table 2 – northern hemisphere surveillance).

4.2.3   Two waves in the same surveillance season

Mem is designed to identify one single epidemic wave. When a season has two epidemics waves, data must be split in two (Table 5 – two waves in a season), so that each column have information from one single wave. MEM can do this automatically in the data loading transformation.

Original season	
Week No	**2000/2001**
30	30.1
31	120.3
32	240.3
33	231.6
34	160.3
35	100.1
36	40.9
…	
52	60.4
1	140.7
2	210.2
3	270.0
4	120.0
5	40.2
6	30.4

first epidemic

second epidemic

Modified season		
Week No	**2000/2001(1)**	**2000/2001(2)**
30	30.1	
31	120.3	
32	240.3	missing
33	231.6	values
34	160.3	
35	100.1	
36	40.9	
…		
52		60.4
1		140.7
2	missing	210.2
3	values	270.0
4		120.0
5		40.2
6		30.4

Table 5 – two waves in a season

Seasons thus split must be named using the same notation and adding 1 or 2 at the end between parentheses.

4.3   Graph details

4.3.1   Interactive plots

Most graphs created by memapp are interactive. Moving the cursor over one of the series give information of the element currently selected.

It is possible to remove any element in the graph clicking its name in the legend.

Plots can be zoom in and out and exported as png files (Figure 41 – interactive plots).

Figure 41 – interactive plots

4.3.2 Thresholds

You can add to most plots different thresholds, the amount of information in the plot is determined by the left panel box Thresholds (Figure 42 – thresholds box), which allows to include or exclude pre-epidemic, post-epidemic and intensity thresholds to the original graph.
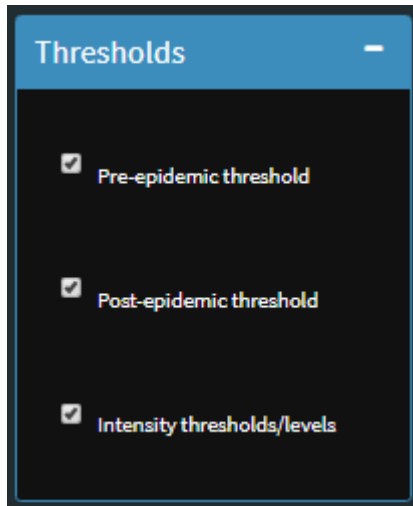
Figure 42 – thresholds box

These thresholds added to the plot are calculated using the model currently selected in the model box.

It is possible, then, to plot a graph of a set of seasons along with the thresholds of a model calculated using a totally different set of seasons.

4.3.3 Text options

The right panel box called Text options (Figure 43 – text options box) can be used to change the main title, the x-axis label and the y-axis label of the current graph.

Figure 43 – text options box

4.3.4 Graph options

The right panel box called Graph options (Figure 44 – graph options box) change the colours used in the plots, allowing to use different palettes for different graphs.

- Observed (line): This is the colour of the line of the surveillance season and affects the plots: Series, Timing, Average curve and Surveillance.
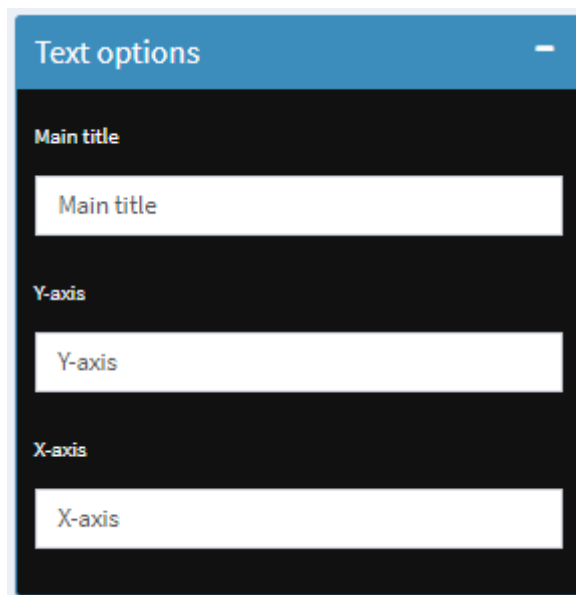- Observed (points): The colour of the points of the surveillance season and affects the plots: Average curve and Surveillance.
- Epidemic start: The colour of the marker that shows the start of the epidemic period (the first value above the pre-epidemic threshold) and affects the plots: Moving epidemics, Average curve and Surveillance.
- Epidemic end: The colour of the marker that shows the end of the epidemic period (the first value below the post-epidemic threshold after the peak has been reached) and affects the plots: Moving epidemics, Average curve and Surveillance.
- Thresholds palette: The colour of the pre-epidemic, post-epidemic and intensity thresholds and affects the plots: Seasons, Series, Moving epidemics, Average curve and Surveillance.
- Seasons palette: The colour of the different series when there are different plots of the same kind in the graph and affects the plots: Seasons, Moving epidemics and Optimize.
- Timing palette: The colour of the points included in the pre-epidemic, epidemic and post-epidemic period and affects the plots: Series and Timing.

Figure 44 – graph options box

4.4 How goodness is calculated

MEM calculates goodness indicators in an iterative process. In each iteration:

1. For one particular season the timing is calculated to determine which weeks are inside the pre, post and epidemic periods. This is used as the real data: a real positive outcome (epidemic weeks) and a real negative outcome (pre and post-epidemic weeks).

2. With a set of seasons, pre-epidemic threshold is calculated. This threshold is compared with values from the season selected in the first step and see if values are above or below the threshold. This is used as the observed data: an observed positive outcome (week value above the threshold), and observed negative outcome (week value below the threshold).

3. Each week has a real and an observed outcome, so it can be classified in:
   - True positives (TP): real positive, observed positive: values of the epidemic period above the threshold.

[55]

- True negatives (TN): real negative, observed negative: values of the non-epidemic period below the threshold.
- False positives (FP): real negative, observed positive: values of the non-epidemic period above the threshold.
- False negatives (FN): real positive, observed negative: values of the epidemic period below the threshold.

4. The process is repeated for each season in the dataset (each iteration a different value until all seasons have been processed).
5. All TP, TN, FP and FN are pooled together and sensitivity, specificity, positive predictive value, negative predictive value, percent agreement and the Matthews correlation coefficient are calculated.

There are two ways of deciding the set of seasons used to calculate the pre-epidemic threshold in each iteration and it is determined by the Validation parameter in the right panel MEM options box.

- Cross: For each value, the surrounding seasons (after or before the current value) are selected up to the number of Maximum seasons (parameter of the Model box). To calculate the thresholds for season 2010/2011, data from 2005/2006 to 2009/2010 and from 2011/20012 to 2015/2016 will be taken.
- Sequential: Only preceding seasons are used (before the current value) up to the number of Maximum seasons. To calculate the thresholds for season 2010/2011, data from 2000/2001 to 2009/2010 are taken.

4.5 Support

The box support (Figure 45 – support box) has a link to the technical manual, another one to submit issues with the application and a checkbox to activate the advanced features of memapp. Advanced features have been described along this manual.
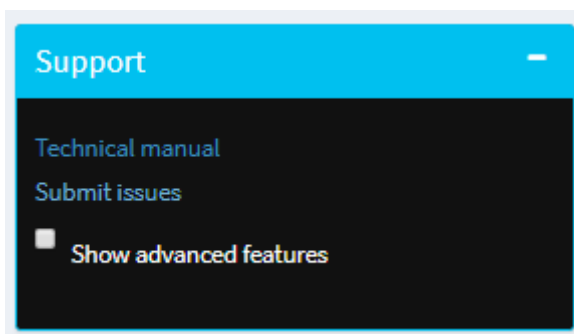

Figure 45 – support box

4.6 Language

This combo box allows changing the language of the web application; currently English, Spanish and French translation are available.

Starting with version 2.6, memapp enabled an option to localize the web app. Any user interested on translating memapp to another language can edit the semicolon separated values file at github:

https://github.com/lozalojo/memapp/blob/master/inst/shinyapp/lang/en-GB.txt

With a text editor, translating the second column and sending it back to the maintainer to be added to the app.

4.7   Multiple waves algorithm

This algorithm separates different waves of a whole dataset.

Prepare data as a single series of data: $y_i, i = 1, …, N$ and calculate another series of the loess predicted data $\tilde{y}_i, i = 1, …, N$. Replace missing values by predicted data:

$$\breve{y}_i = \begin{cases} \tilde{y}_i & if \ y_i \ is \ missing \\ y_i & otherwise \end{cases}$$

Input parameters: $\delta, \gamma, \tau, \varphi$

Main algorithm

1. The sum of all weekly values is:

$$S = \sum_{i=1}^{N} \breve{y}_i$$

2. Detect one epidemic using the *single epidemic detection algorithm* (see below).
3. The epidemic is detected at position r and has a length of n weeks, the percentage covered by that epidemic is:

$$P = {\sum_{i=r}^{r+n-1} \breve{y}_i} / S$$

4. If $P < \boldsymbol{\gamma}$ stop, the new epidemic represent less that a predefined percentage, thus is not included in the final results, the epidemics are those defined in the previous step.
5. Else, assign missing values to the current epidemic:

$$\breve{y}_i = missing, i = r, …, r + n - 1$$

And return to step 2.

Post-processing

1. Join epidemics whose distance is less than τ to create new bigger epidemics.
2. To define seasons (which will include the epidemics), calculate the cut points between seasons as the lowest **loess** value of inter-epidemics periods.

3. Once the cut points are defined, the seasons are defined as the period between one cut point and the following one.
4. The length of each season is shortened to length φ not to have seasons longer than 52.

Single epidemic detection algorithm

1. Define the percentage given of a set starting at point r of length n.

$$p_r^n = {\sum_{i=r}^{r+n-1} \breve{y}_i} \Big/ {\sum_{i=1}^{N} \breve{y}_i}$$

2. Calculate the maximum percentage for a given length n.

$$p_{max}^n = max\{p_r^n : r = 1, \dots, N - r + 1\}$$

3. Calculate the addition of the lowest rate to that maximum percentage of length n.

$$r^{max} = \{r / p_r^n = p_{max}^n\}$$

$$t_{min}^n = {min\{\breve{y}_i, i = r^{max}, \dots, r^{max} + n - 1\}} \Big/ {p_{max}^n}$$

4. Determine the length of the first epidemic (maximum length of an epidemic on which the minimal rate represent a percentage of at least $\boldsymbol{\delta}$).

$$n^* = max\{n / t_{min}^n \geq \boldsymbol{\delta}\}$$

5. The epidemic start at r^* and has a length of n^*.

$$r^* = \{r / p_r^{n^*} = p_{max}^{n^*}\}$$

4.8 Experimental features

Some features in the present manual have been tagged as "Experimental". Those features are only activated by default in the development version (github) of the application. There are no experimental features in the stable version (CRAN) of the package.

5   Troubleshooting

5.1   Linux systems

In Linux systems, there are R libraries that need some specific packages installed on your system. When installing each R library, if it fails R will tell you which system package is missing and will point you to a possible solution.

For example, for debian-based distros this is the list of required system packages required to build memapp:

```
sudo apt-get install libssl-dev libcurl4-openssl-dev
unixodbc-dev zip
```

5.2   Macosx systems

In MacOSX, you'll need to install an X Windows System: quartz.

First install homebrew. From the command line:

```
ruby -e "$(curl -fsSL
https://raw.githubusercontent.com/Homebrew/install/master/in
stall)"
```

And then install XQuartz using brew.

```
brew cask install xquartz
```

Install GCC compilers:

```
brew install gcc
```

Brew also lets you to install R:

```
brew cask install r-app
```

And RStudio:

```
brew cask install --appdir=/Applications rstudio
```

Some packages will require to install additional programs. For RODBC package:

```
brew install unixodbc
```

For openssl package:

```
brew install openssl
```

5.3   Ms Access files

The memapp is able to read MS Access files.

In windows systems, it requires to have the ODBC driver installed in your computer (normally if you have MS Office installed, there will be no further problems).

In Linux systems it is required to install a package (for your OS, not for R) called mdbtools:

http://mdbtools.sourceforge.net/

Which probably will be in your distro repositories, for debian-based distros:

```
sudo apt-get install mdbtools
```

[59]

In MacOSX, it is possible to install mdbtools using brew command:

```
brew install mdbtools
```

5.4   Exporting to MS Excel

The package in charge of exporting files to MS Excel is openxlsx. To install it:

```
# check if openxlsx is installed, and install it otherwise
if(!require("openxlsx")) install.packages('openxlsx')
```

The function that exports data requires a zip executable in your system.

In most Linux distros is enough to have the package zip (or similar) installed and the executable in the /usr/bin/ directory.

In Windows, memapp will look for Rtools, a set of tools for R, can be downloaded from its webpage and installed:

https://cran.rstudio.com/bin/windows/Rtools/

Or, alternatively, you can do it inside R:

```
# check if installr is installed, and install it otherwise
if(!require("installr")) install.packages('installr')
library("installr")
# install Rtools
install.Rtools()
```

Whatever option you chose, when the installer ask you if you want to automatically modify the PATH environment variable, check the tickmark and restart after all the process is done.

In MacOSX zip is probably installed in your system by default and will work with no issues.

If memapp is not able to find the zip file, it will automatically hide the export to xlsx button.

5.5   Creating animated graphs

The graph under Surveillance/Animated is an animated gif that requires some configuration.

The easiest option is to install the *animation* package, is listed as suggested for the memapp package, but not mandatory, so it is not installed by default.

In windows environments will work out of the box, but in linux systems, it requires another package to be installed:

```
sudo apt-get install libmagick++-dev
```

Under MacOSX also requires:

```
brew install imagemagick
```

The second and recommended option (Windows only) is to use the *animation* package. It works in combination with any of the two following programs: GraphicsMagick (recommended) or ImageMagick.

They can be downloaded and installed from their respective web pages:

www.graphicsmagick.org

www.imagemagick.org

Alternatively, the package installr allows to install those programs directly from the R console:

```
# check if installr is installed, and install it otherwise
if(!require("installr")) install.packages('installr')
library("installr")
# install GraphicsMagic
install.GraphicsMagick()
# install ImageMagick
install.ImageMagick()
```

6 References

1. Vega Alonso, T., Lozano Alonso, J. E., Ortiz de Lejarazu, R. & Gutiérrez Pérez, M. Modelling influenza epidemic—can we detect the beginning and predict the intensity and duration? *Int. Congr. Ser.* **1263,** 281–283 (2004).

2. Caroline Sarah Brown *et al.* WHO Regional Office for Europe guidance for sentinel influenza surveillance in humans (2009). (2009).

3. Vega, T. *et al.* Influenza surveillance in Europe: establishing epidemic thresholds by the moving epidemic method. *Influenza Other Respir. Viruses* **7,** 546–558 (2013).

4. Vega, T. *et al.* Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method. *Influenza Other Respir. Viruses* **9,** 234–246 (2015).

5. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2017).

6. Lozano Alonso, J. E. *The Moving Epidemic Method R Package*. (CRAN, 2010).

7. Lozano, J. E. lozalojo/mem: Second release of the MEM R library. *Zenodo* doi:10.5281/zenodo.165983

8. Lozano, J. E. & Bergström, J. *lozalojo/memapp: Second release of the MEM Shiny Web Application R package*. (Zenodo, 2018). doi:10.5281/zenodo.1173518

9. Lozano Alonso, J. E. *The Moving Epidemic Method Web Application*. (CRAN, 2017).

10. RStudio Team. *RStudio: Integrated Development for R*. (RStudio, Inc., 2016).