

Software

Open Access

Online detection and quantification of epidemics

Camille Pelat^{1,2}, Pierre-Yves Boëlle^{1,2,3}, Benjamin J Cowling⁴,
Fabrice Carrat^{1,2,3}, Antoine Flahault^{1,2,5}, Séverine Ansart^{2,6} and Alain-
Jacques Valleron*^{1,2,3}

Address: ¹Université Pierre et Marie Curie-Paris 6, UMR-S 707, Paris, 75012 France, ²INSERM, U707, Paris, 75012 France, ³AP-HP, Hôpital Saint Antoine, Unité de Santé Publique, Paris, 75012 France, ⁴Department of Community Medicine and School of Public Health, The University of Hong Kong, Hong Kong SAR, China, ⁵AP-HP, Hôpital Tenon, Unité de Santé Publique, Paris, 75012 France and ⁶Hôpital de la Cavale Blanche, Maladies infectieuses, Brest, 29609 cedex, France

Email: Camille Pelat - pelat@u707.jussieu.fr; Pierre-Yves Boëlle - boelle@u707.jussieu.fr; Benjamin J Cowling - bcowling@hku.hk;
Fabrice Carrat - carrat@u707.jussieu.fr; Antoine Flahault - flahault@u707.jussieu.fr; Séverine Ansart - severine.ansart@chu-brest.fr; Alain-
Jacques Valleron* - valleron@u707.jussieu.fr

* Corresponding author

Published: 15 October 2007

Received: 29 June 2007

BMC Medical Informatics and Decision Making 2007, **7**:29 doi:10.1186/1472-6947-7-29

Accepted: 15 October 2007

This article is available from: <http://www.biomedcentral.com/1472-6947/7/29>

© 2007 Pelat et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Time series data are increasingly available in health care, especially for the purpose of disease surveillance. The analysis of such data has long used periodic regression models to detect outbreaks and estimate epidemic burdens. However, implementation of the method may be difficult due to lack of statistical expertise. No dedicated tool is available to perform and guide analyses.

Results: We developed an online computer application allowing analysis of epidemiologic time series. The system is available online at http://www.u707.jussieu.fr/periodic_regression/. The data is assumed to consist of a periodic baseline level and irregularly occurring epidemics. The program allows estimating the periodic baseline level and associated upper forecast limit. The latter defines a threshold for epidemic detection. The burden of an epidemic is defined as the cumulated signal in excess of the baseline estimate. The user is guided through the necessary choices for analysis. We illustrate the usage of the online epidemic analysis tool with two examples: the retrospective detection and quantification of excess pneumonia and influenza (P&I) mortality, and the prospective surveillance of gastrointestinal disease (diarrhoea).

Conclusion: The online application allows easy detection of special events in an epidemiologic time series and quantification of excess mortality/morbidity as a change from baseline. It should be a valuable tool for field and public health practitioners.

Background

The generalization of electronic data capture in health care has made time series data increasingly available for public health surveillance [1]. How to best analyse these data will likely be case dependent and require expert sta-

tistical advice. There is however a well agreed "good analysis practice" in particular classes of surveillance problems, so that less expert users may consider undertaking the analysis themselves. This requires making software available online and providing guidance on its use: this is

exactly what was done with online tools for DNA sequences alignment (BLAST, FASTA), allowing biologists to successfully use these methods on their own data.

Here, we focus on epidemic detection and quantification from time series data. There is a widely used approach for this purpose originating from Serfling's work on influenza [2]. He proposed calculating excess P&I mortality due to seasonal influenza using deviations from a periodic regression model that captured the annual seasonality of the data. It was first necessary to (subjectively) select years without excess death to train the baseline regression model. The approach has then been extended to address several issues: refining regression equations and extracting baseline model information without subjective filtering of the data [3-5]. Algorithms for prospective outbreak detection were also proposed in this framework [6-8].

In this paper, we describe an online tool allowing users to detect unexpected events, eg outbreaks, in a seasonal epidemiologic time series. Two applications are detailed to illustrate how results are obtained.

Implementation

Two types of analysis exist for surveillance time series: retrospective analysis, to locate and quantify the impact of past epidemics, and prospective analysis, for real time detection of epidemics. In all cases, four steps are necessary. First, a subset of data ("training data") is selected from the whole time series to estimate the baseline level. Second, an algorithm or a rule is used to selectively discard epidemic events from the training data, so that the baseline level is estimated from truly non epidemic data. Third, a periodic regression model is fitted to the training data. Finally, the model is used to define an epidemic threshold and/or estimate excess morbidity/mortality. We review how these issues have been addressed in the literature, using the detection of influenza epidemics in time series as an illustration. Table 1 summarizes all inputs required from the user, and describes the default options retained in our system.

Training period

Even if long time series are available, it is not generally the case that all data should be included in the training period [9]. Indeed, changes in case reporting and demographics will likely be present over long time periods, and this may affect how well the baseline model fits the data. Modelling of influenza mortality typically uses the five preceding years in baseline determination [2,10,11]. Including more past seasons improves the seasonal components estimates, while limiting the quantity of data allows capturing recent trends. In our system, we propose using the whole dataset in the model fitting for retrospective analysis (as done, for example, in [12,13]), and to limit to a past few years in the case of prospective detection of epidemics (as, for example, in [7,8]). In the latter case, the user is invited to specify the length of the training data in an input field. He can define it in number of years or in number of observations. In either case, the minimal time span accepted is one year.

Purge of the training period

In order to model the non-epidemic baseline level, the model must be fitted on non-epidemic data. For seasonal diseases such as influenza in the Northern hemisphere, it is difficult to find long epidemic-free periods since epidemics typically occur every year. There are two choices to deal with the presence of epidemics in the training data: excluding the corresponding data from the series, or explicitly modelling the epidemics.

In the first choice, epidemics must first be identified. Several rules have been suggested in this respect. Viboud et al. excluded the 25% highest values from the training period [13]. Costagliola et al. removed all data above a given threshold (more than three influenza-like illness cases per sentinel general practitioner) [14]. Olson et al. excluded the months with "reported increased respiratory disease activity or a major mortality event" [4]. Others deleted entire periods: e.g. December to April [12], or September to mid-April [15].

Table 1: Required inputs from the user for baseline model fitting

Parameter	Possible values	Default value
Length of the training period	Number of years, number of observations	Retrospective : All data Prospective : Half dataset
Purge of the training period	Data above selected percentile, above cut-off value, or in user-defined periods	Above the 15% highest percentile
Regression equation	Linear, quadratic or cubic terms. Annual, semi-annual or quarterly periodicity	Automated model selection
Upper forecast limit (UFL)	Percentile between 50% and 100%	95%
Minimum duration above UFL defining an unexpected change	Number of observations	14 days/2 weeks/1 month

The second choice, less common, requires explicit modelling of the epidemic periods during the training data. In this case, an epidemic indicator must be included as a covariable in the model. For influenza epidemic, one may choose the number of laboratory influenza A and B isolates [5,16]. However, the availability of an independent epidemic indicator is uncommon in practice.

In summary, data points may be excluded either because they exceed a (possibly data determined) threshold, because they were collected during a period known to be epidemic prone (for example winters), or because the user wishes to exclude the points. These three options are available in our system.

Regression equation

A variety of formulations may be used for the regression equation, including linear regression [14], linear regression on the log-transformed series [6], Poisson regression [17], and Poisson regression allowing for over-dispersion [18]. Linear regression is suitable when working with large frequencies or incidences, while working with the log transformed series or applying Poisson regression is advised when observations are small in magnitude.

In the regression equation, the trend is generally modelled using a linear term [2,4,11], or a second degree polynomial [3,7,19]. In our application we propose these two trends plus the third degree polynomial, to offer more flexibility. When the model is used for prospective detection of epidemics, it is often safer to use only a linear trend to avoid inconsistencies when the model will be extrapolated into the future. Thus, the application restrains the user's choice to the models that have linear trend. For retrospective analysis, where extrapolation is not an issue, more complex trends may improve the fit of the baseline model. So, the application allows the user to choose among all the proposed models with linear, quadratic and cubic trends. For the seasonal component, a simple yet effective description may be obtained using sine and cosine terms with period one year [2]. Refined models are found in the literature, often with terms of period 6 months [14], sometimes 3 months [3], and, rarely, smaller [11]. In our application, we chose to propose the most widely used periodicities, ie 12, 6 and 3 months. As a result, all regression equations for the observed value $Y(t)$ are special cases of the following model: $Y(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \gamma_1 \cos(2\pi t/n) + \delta_1 \sin(2\pi t/n) + \gamma_2 \cos(4\pi t/n) + \delta_2 \sin(4\pi t/n) + \gamma_3 \cos(8\pi t/n) + \delta_3 \sin(8\pi t/n) + \epsilon(t)$. For prospective modelling, α_2 and α_3 are always 0. Model coefficients are estimated by least squares regression.

In our system, automatic selection of the best fitting model is made possible by a selection algorithm (see Fig-

ure 1, which illustrates the process on an example detailed in the result section). It relies on ANOVA comparison (significance level : 0.05) to select between nested models, and on Akaike's Criterion to select between non-nested models [20]. The algorithm starts comparing, by ANOVA, the simplest model M11 ($Y(t) = \alpha_0 + \alpha_1 t + \gamma_1 \cos(2\pi t/n) + \delta_1 \sin(2\pi t/n) + \epsilon(t)$) with the two models in which it is nested: M12 ($Y(t) = \alpha_0 + \alpha_1 t + \gamma_1 \cos(2\pi t/n) + \delta_1 \sin(2\pi t/n) + \gamma_2 \cos(4\pi t/n) + \delta_2 \sin(4\pi t/n) + \epsilon(t)$) and M21 ($Y(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \gamma_1 \cos(2\pi t/n) + \delta_1 \sin(2\pi t/n) + \epsilon(t)$). If none of the alternative models (M12 and M21) is significantly better than the initial one (M11), the algorithm keeps M11 and stops. If one of the two alternative models is better than the initial one, the algorithm keeps it and goes on. If the two alternative models are better than the initial one, the algorithm keeps the one with the lowest AIC and goes on. The process is repeated until finding the "best overall" model over the nine proposed models.

Alert notification

As the baseline model is fitted to the observations, the variation around the model fit may be estimated by the standard deviation of the residuals (difference between observed and model value). It is therefore possible to calculate forecast intervals for future observations, assuming that the baseline model holds in the future. Thresholds signalling an unexpected change are typically obtained by taking an upper percentile for the prediction distribution (assumed to be normal), typically the upper 95th percen-

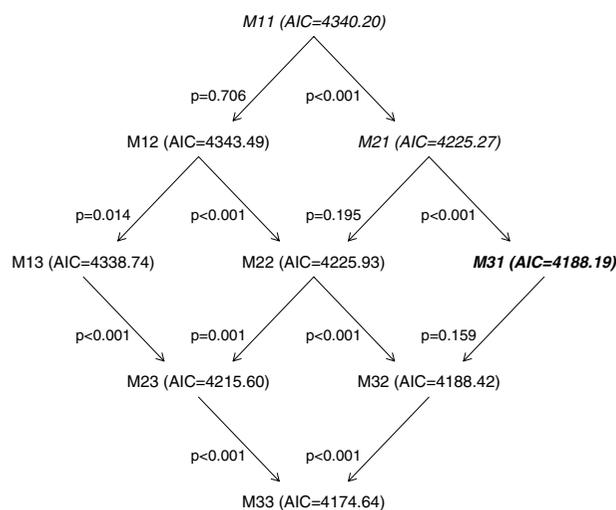


Figure 1
Model selection algorithm. Graphical output of the model selection algorithm. Data and models are described in Table 2. Models selected through the algorithm pathway are in italics. The model finally kept is in bold italics.

Table 2: Retrospective evaluation of the excess P&I mortality in France for 1968–1999, using nine periodic regression models. The components included in each model are indicated by a*. #Model options: exclusion of the top 15% percentile from the training period; forecast interval: 95%

Model#	Trend			Periodicity			AIC	Cumulated excess mortality over the whole period
	t	t ²	T ³	1 year	6 months	3 months		
M11	*			*			4 340	88 442
M12	*			*	*		4 343	87 260
M13	*			*	*	*	4 339	88 266
M21	*	*		*			4 225	85 083
M22	*	*		*	*		4 226	83 245
M23	*	*		*	*	*	4 216	83 505
M31	*	*	*	*			4 188	85 175
M32	*	*	*	*	*		4 188	83 337
M33	*	*	*	*	*	*	4 175	82 465

tile [14], or upper 90th percentile [11]. A rule is then used to define when epidemic alerts are produced: for example as soon as an observation exceeds the threshold [7], or if a series of observations fall above the threshold, for example during 2 weeks [13], or 1 month [21].

Results

We developed a web-based application allowing users to construct periodic regression models for analysis of epidemiologic time series. It is written in HTML, PHP, and JavaScript for the user interface, and interfaced with the R system (2.5.0) for statistical computations [22]. The application is available online [23]. The R codes are freely available in Additional file 1 and on the application web site.

Users may input their own dataset (eg incidences, mortalities, medication sales) as a plain text file (ie ASCII file) containing the time series as a single column, i.e. the values are separated by a carriage return. Observations must be aggregated by day, week or month. The user will be invited to specify this time step in a scrolling list. Missing values are allowed, provided they are coded by "NA". It is assumed that the dataset will contain at least one year of data. Several example datasets from France are included in the system: incidence rates per 100,000 population for influenza-like illness and diarrhoea for 1991–2001, and P&I mortality series for 1968–1999 [24]. They are available as daily, weekly or monthly time series.

Retrospective analysis of influenza epidemics

The first example uses monthly P&I mortality in France over the period 1968–1999. The user wishes to retrospectively identify the epidemic periods and quantify the cumulated mortality in these epidemics. Use of the system begins with selecting the corresponding dataset on the main page.

After data input, the user is taken through three successive webpages to specify the baseline model parameters (Table 1). The first page allows choosing the type of analysis. Here, the user selected to conduct retrospective analysis, therefore the whole time series is included in the training period.

The second page allows excluding observations from the training period (Figure 2). Three options for excluding data are proposed. The user may select the upper percentile between 0% and 60% above which all data are excluded. Excluding all observations greater than a specified cut-off value is the second option. In the third option, the user provides a file of the same length as the training period flagging the observations as "keep" (value 0) or "exclude" (value 1). To guide the percentile or cut-off selection, histograms and cumulated density plot are provided. In Figure 2, the user selected to exclude all observations greater than the 15% upper percentile.

The third page allows the user to select the mathematical form for the baseline model. This page is dependent on the type of analysis, prospective or retrospective. For the retrospective analysis, nine models are available, combining the three choices for the trend and periodicity (see Table 2). Using the automated selection feature, the model with cubic trend and annual periodicity is chosen for baseline P&I mortality. Figure 1 presents the detail of the selection algorithm.

In the third page, the user defines the epidemic threshold by selecting a percentile of the prediction distribution, between 50% and 100%. Here, default value (95%) was selected. Increasing this value will lead to less observations outside the thresholds and more specific detection.

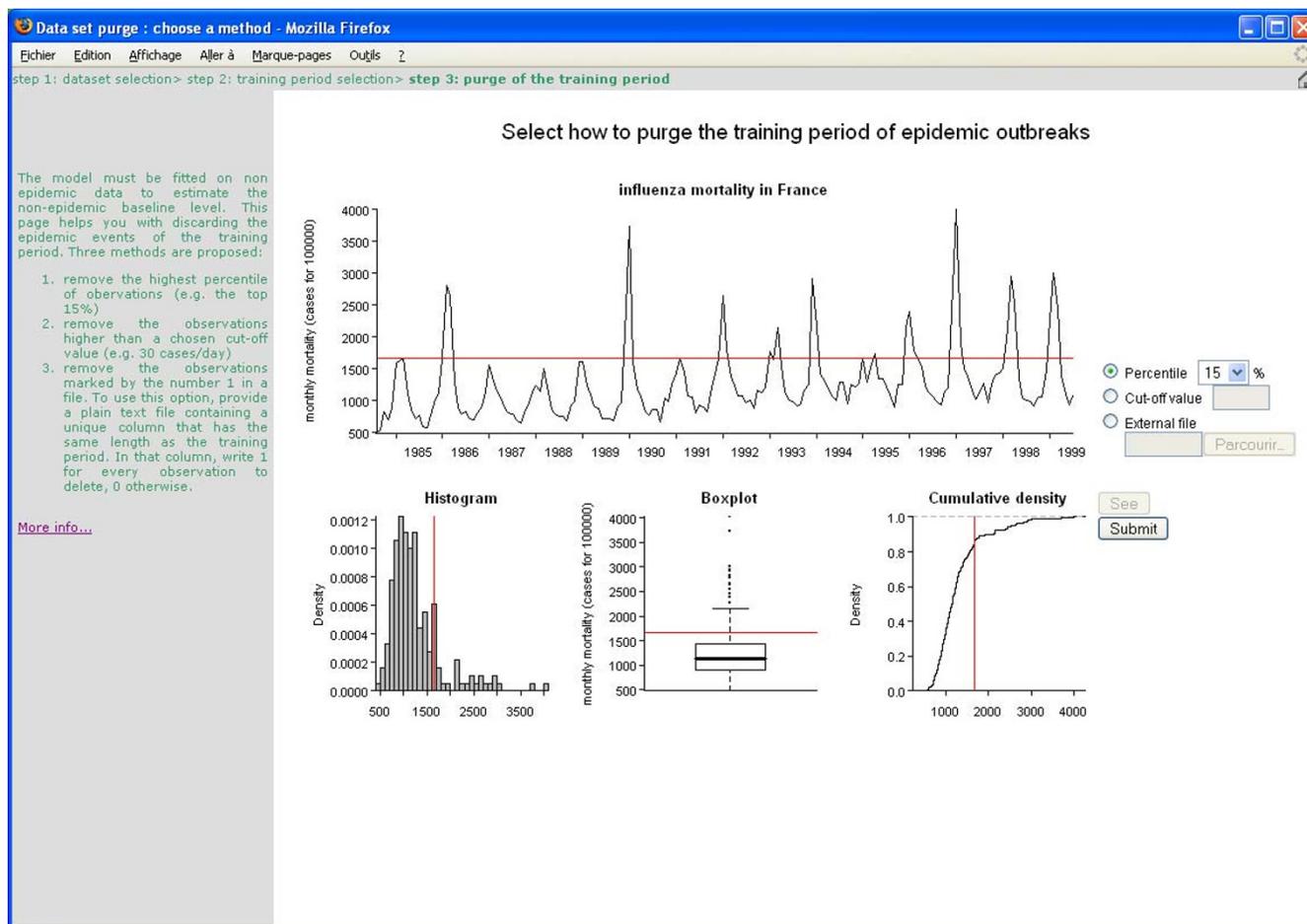


Figure 2
Purge of the training period. Interactive selection of the method used to purge the training period of past epidemic outbreaks. Option 1 (delete the highest percentile of observations) was chosen. The percentile was set to 15% in a scrolling list ranging 0% to 60%.

On the contrary, decreasing the threshold will increase sensitivity and timeliness of the alerts.

To avoid making alerts for isolated data points, a minimum duration above the threshold may be required. Default values are 14 days (2 weeks) for daily and weekly data, and 1 month if the data are monthly-aggregated. The beginning of the epidemic is the first date the observations exceed the threshold, and the end the first time observations return below the threshold. Here, the default value (1 month) was selected.

The application provides plots of the time series, the baseline level, the threshold and the detected epidemics (Figure 3a). A first table is output with the expected baseline and threshold values at each date in the dataset. A second table shows the dates and excess mortality for all detected epidemics (summarized in Table 2). The excess mortality

is defined as the cumulated difference between observations and baseline over the entire epidemic period. Excess percentages are also provided, calculated as the observed size divided by the sum of expected values throughout each epidemic.

Prospective surveillance of gastrointestinal diseases

In this analysis, the user wishes to define epidemic thresholds for prospective monitoring of diarrhoea. We briefly summarize the differences between this analysis and the retrospective case. As above, a time series must be first provided (here we selected diarrhoea with weekly observations). After choosing "prospective analysis", the user must select the duration of the training period, typically a few years. We select five years for the training data. Data exclusion before fitting the baseline model is carried out as in the first example. The regression equation is limited to a linear trend, but all three periodicities are available.

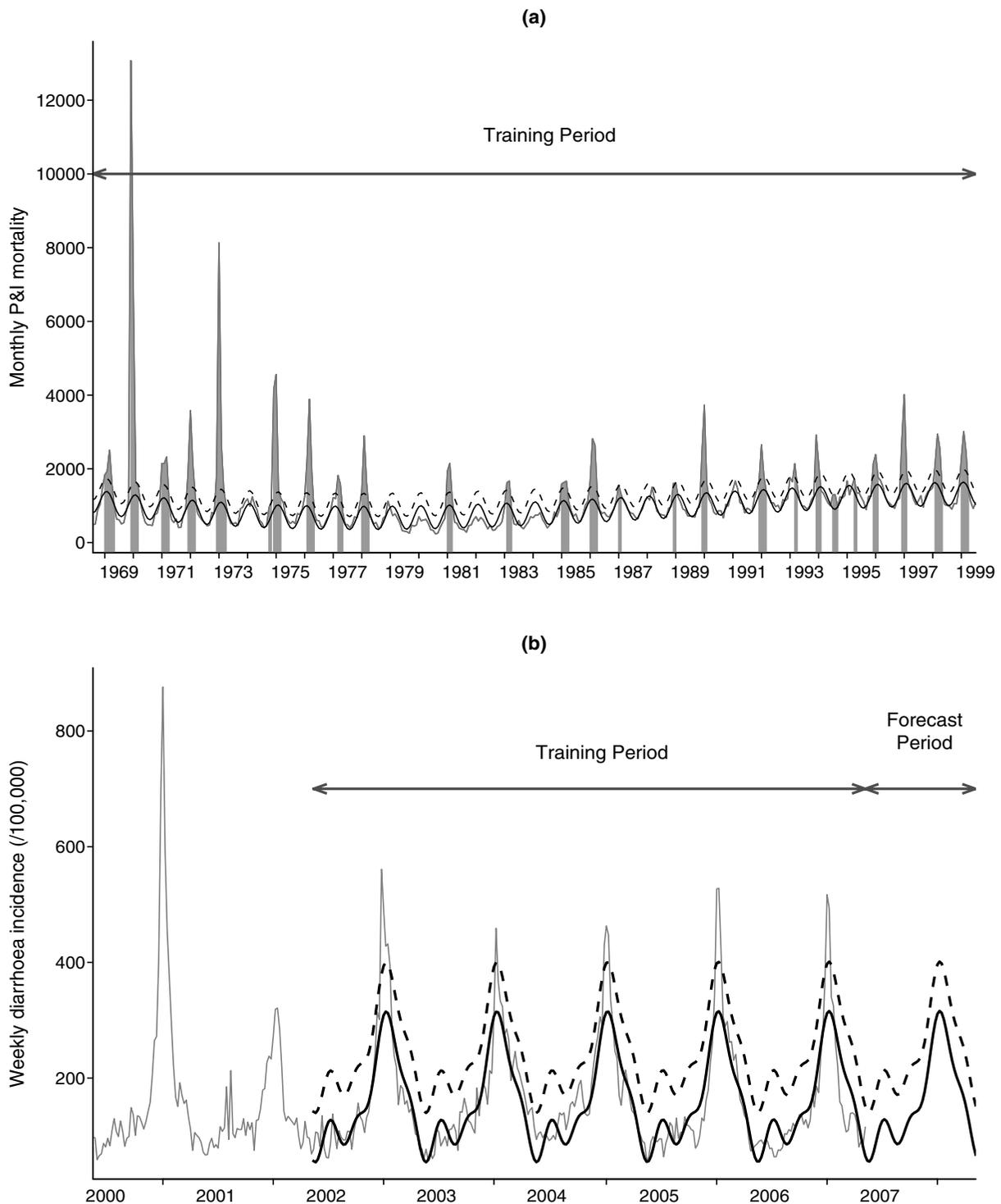


Figure 3
Graphical output of the software. (a) Retrospective detection of influenza epidemics from monthly P&I mortality in France, 1968–1999. (b) Prospective analysis of gastrointestinal disease (2002–2007) and model-based extrapolation for 2008 with epidemic threshold. In all graphs: observed (grey), model (black), upper forecast limit (dashed).

Here, the automated selection leads to a model with a linear trend, and annual, semi-annual and quarterly periodic terms.

Alert thresholds are defined by selecting a percentile of the prediction distribution, between 50% and 100%. A typical choice is 95%. The application then generates a plot showing the whole data, the baseline and threshold values over the training period and model extrapolation for the following year (Figure 3b). An output table contains the expected baseline and threshold values for each date in the dataset and the following year.

Discussion

We have presented an online application for analysing epidemiologic surveillance time series. The program can be used to extract the dates and size of past epidemics, or to establish epidemic thresholds for prospective surveillance. We intend this application to be a practical tool for field and public-health practitioners. We designed a user-friendly interface that provides default-values options and interactive graphical feedback. Since all the parameters can be changed by the user, the program provides an easy way to check how the analysis changes with different choices.

The epidemiologic time series most suitable for analysis are those where the monitored signal consists of a seasonal background with outbreaks. This is clearly the case for influenza surveillance data. Influenza-like syndromes occur at all times of the year, although typically more in the winter than the summer, even when no influenza viral strain is circulating. Viral testing is considered the gold standard method to provide the real number of influenza-affected patients but since this test is not part of routine diagnoses, morbidity and mortality in a population can not be specifically attributed to influenza. One way to estimate the impact of influenza in a population from surveillance data including surveillance of influenza-like syndromes, pneumonia or influenza associated admissions, or cause-specific mortality, is to use statistical methods such as periodic regression. This hypothesis also holds for other infectious diseases, for example gastroenteritis where syndromes under surveillance (diarrhoea, fever) can be due to various pathogens which are more active in some seasons than others. Alternative detection methods exist that do not rely on the hypothesis of a seasonal baseline. For instance, Hidden Markov Models assume that the observations are generated from a finite mixture of distributions governed by an underlying Markov chain [25,26]. These methods have shown good aptitude in distinguishing epidemic and non epidemic phases in seasonal and non-seasonal time series. Another alternative is control-chart methods, which may be cali-

brated on data from recent months rather than from previous years [27].

A minimum of one year historical data is required to fit the models discussed here, but we note that more reliable predictions require at least two or three year historical data to calculate the baseline level. Other methods have been developed for disease surveillance with limited historical data sets [27,28]. We also recommend, for the prospective setting, to make sure that the one year long predictions begin outside the epidemic season, in order to highlight the incoming epidemic in its entirety. While first and second degree polynomial trends are frequently used in periodic regression models in the literature [2,3], we have added the option of a third degree polynomial to offer more flexibility, only for the retrospective analysis. For the seasonal components, we included the most widely used periodicities, ie 12, 6 and 3 months. We did not propose higher degree polynomials or seasonal terms because higher order terms may be more prone to result in unidentifiable models or other problems with model fit.

The application is based on a general periodic regression model that contains most previous published models as special cases. Yet, we did not implement some specialised models encountered in the literature. For example, some authors modelled the secular trend with a smoothing spline fitted on summer months [12,29]. Others included autoregressive terms in their models [5,30,31]. Additional variables may also be incorporated into the regression model, for example day of the week, holiday, and post-holiday effects [7], sex and age [32], or temperature and humidity [5]. A few authors replaced the epidemic values in the training period by expected non-epidemic values, rather than deleting them [10,33]. We have not included these options in the application for reasons of parsimony. One of the most important features of an online tool such as the one presented here is that it should allow inferences to be made by front-line practitioners who often do not have detailed knowledge of statistical software. We have attempted to balance the desire to provide a user-friendly interface while at the same time offering sufficient options to cover the needs of most surveillance datasets.

Conclusion

The online application presented here should be a valuable tool for public health surveillance. Its user-friendly interface facilitates fairly complex modelling, offering public health practitioners the possibility to rapidly investigate the burden of epidemics, or to utilise the same statistical approaches to set epidemic thresholds for prospective surveillance.

Availability and requirements

- **Project name:** Periodic regression models
- **Project home page:** http://www.u707.jussieu.fr/periodic_regression/
- **Operating systems:** Web based application
- **Programming language:** R, PHP, Javascript
- **Other requirements:** Javascript supported and activated on the web browser (tested with Mozilla 5.0 and Internet Explorer 7.0).

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

CP designed and programmed the application and drafted the manuscript. PYB helped to conceive the application and to draft the manuscript. BJC helped with the program and with drafting the manuscript. FC participated in the program design. AF helped to draft the manuscript. SA helped designing the application. AJV participated in designing the application and drafting the manuscript. All authors have read and approved the final manuscript.

Additional material

Additional file 1

R codes. The zip file contains all the R codes used in the web site and an instruction file. You can directly run the scripts with the R software, following the directions in the instruction file, to obtain the graphical outputs and the tables.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1472-6947-7-29-S1.zip>]

Acknowledgements

This work was supported by the EU Sixth Framework Programme for research for policy support (contract SP22-CT-2004-511066). The funding source had no involvement in the work process.

References

- O'Carroll PW: **Introduction to Public Health Informatics.** In *Public Health Informatics and Information Systems* Edited by: O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH, Martin EL. New York, Springer; 2003:3-15.
- Serfling RE: **Methods for current statistical analysis of excess pneumonia-influenza deaths.** *Public Health Reports* 1963, **78**:494-506.
- Housworth J, Langmuir AD: **Excess mortality from epidemic influenza, 1957-1966.** *Am J Epidemiol* 1974, **100**(1):40-48.
- Olson DR, Simonsen L, Edelson PJ, Morse SS: **Epidemiological evidence of an early wave of the 1918 influenza pandemic in New York City.** *Proc Natl Acad Sci U S A* 2005, **102**(31):11059-11063.
- Wong CM, Yang L, Chan KP, Leung GM, Chan KH, Guan Y, Lam TH, Hedley AJ, Peiris JS: **Influenza-associated hospitalization in a subtropical city.** *PLoS Med* 2006, **3**(4):e121.
- Brillman JC, Burr T, Forslund D, Joyce E, Picard R, Umland E: **Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance.** *BMC Med Inform Decis Mak* 2005, **5**(1):4.
- Mostashari F, Fine A, Das D, Adams J, Layton M: **Use of ambulance dispatch data as an early warning system for community-wide influenzalike illness, New York City.** *J Urban Health* 2003, **80**(2 Suppl 1):i43-9.
- Tsui FC, Wagner MM, Dato V, Chang CC: **Value of ICD-9 coded chief complaints for detection of epidemics.** *Proc AMIA Symp* 2001:711-715.
- Sebastiani P, Mandl K: **Biosurveillance and Outbreak Detection.** *Data Mining: Next Generation Challenges and Future Directions* 2004:185-198.
- Choi K, Thacker SB: **An evaluation of influenza mortality surveillance, 1962-1979. I. Time series forecasts of expected pneumonia and influenza deaths.** *Am J Epidemiol* 1981, **113**(3):215-226.
- Lui KJ, Kendal AP: **Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985.** *Am J Public Health* 1987, **77**(6):712-716.
- Simonsen L, Reichert TA, Viboud C, Blackwelder WC, Taylor RJ, Miller MA: **Impact of influenza vaccination on seasonal mortality in the US elderly population.** *Arch Intern Med* 2005, **165**(3):265-272.
- Viboud C, Boelle PY, Pakdaman K, Carrat F, Valleron AJ, Flahault A: **Influenza epidemics in the United States, France, and Australia, 1972-1997.** *Emerg Infect Dis* 2004, **10**(1):32-39.
- Costagliola D, Flahault A, Galinec D, Garnerin P, Menares J, Valleron AJ: **A routine tool for detection and assessment of epidemics of influenza-like syndromes in France.** *Am J Public Health* 1991, **81**(1):97-99.
- Vellinga A, Van Loock F: **The dioxin crisis as experiment to determine poultry-related campylobacter enteritis.** *Emerg Infect Dis* 2002, **8**(1):19-22.
- Wong CM, Chan KP, Hedley AJ, Peiris JS: **Influenza-associated mortality in Hong Kong.** *Clin Infect Dis* 2004, **39**(11):1611-1617.
- Thompson WW, Shay DK, Weintraub E, Brammer L, Cox N, Anderson LJ, Fukuda K: **Mortality associated with influenza and respiratory syncytial virus in the United States.** *Jama* 2003, **289**(2):179-186.
- Vergu E, Grais RF, Sarter H, Fagot JP, Lambert B, Valleron AJ, Flahault A: **Medication sales and syndromic surveillance, France.** *Emerg Infect Dis* 2006, **12**(3):416-421.
- Housworth WJ, Spoon MM: **The age distribution of excess mortality during A2 Hong Kong influenza epidemics compared with earlier A2 outbreaks.** *Am J Epidemiol* 1971, **94**(4):348-350.
- Burnham KP, Anderson D: **Model Selection and Multi-Model Inference.** 3rd edition. Springer; 2003.
- Zucs P, Buchholz U, Haas W, Uphoff H: **Influenza associated excess mortality in Germany, 1985-2001.** *Emerg Themes Epidemiol* 2005, **2**:6.
- R: A language and environment for statistical computing** [<http://www.R-project.org/>]
- An online tool for detecting and measuring epidemics in time series data** [http://www.u707.jussieu.fr/periodic_regression/]
- Garnerin P, Saidi Y, Valleron AJ: **The French Communicable Diseases Computer Network. A seven-year experiment.** *Ann N Y Acad Sci* 1992, **670**:29-42.
- Le Strat Y, Carrat F: **Monitoring epidemiologic surveillance data using hidden Markov models.** *Stat Med* 1999, **18**(24):3463-3478.
- Rath TM, Carreras M, Sebastiani P: **Automated detection of influenza epidemics with Hidden Markov Models.** *LECT NOTES COMPUT SC* 2003, **2810**:521-532.
- Cowling BJ, Wong IO, Ho LM, Riley S, Leung GM: **Methods for monitoring influenza surveillance data.** *Int J Epidemiol* 2006, **35**(5):1314-1321.
- Hutwagner LC, Maloney EK, Bean NH, Slutsker L, Martin SM: **Using laboratory-based surveillance data for prevention: An algorithm for detecting Salmonella outbreaks.** *Emerging Infectious Diseases* 1997, **3**(3):395-400.

29. Viboud C, Bjornstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT: **Synchrony, waves, and spatial hierarchies in the spread of influenza.** *Science* 2006, **312(5772)**:447-451.
30. Ozonoff A, Forsberg L, Bonetti M, Pagano M: **Bivariate method for spatio-temporal syndromic surveillance.** *MMWR Morb Mortal Wkly Rep* 2004, **53 Suppl**:59-66.
31. Wang L, Ramoni MF, Mandl KD, Sebastiani P: **Factors affecting automated syndromic surveillance.** *Artif Intell Med* 2005, **34(3)**:269-278.
32. Brinkhof MW, Spoerri A, Birrer A, Hagman R, Koch D, Zwahlen M: **Influenza-attributable mortality among the elderly in Switzerland.** *Swiss Med Wkly* 2006, **136(19-20)**:302-309.
33. Grigoryan VV, Wagner MM, Waller K, Wallstrom GL, Hogan WR: **The Effect of Spatial Granularity of Data on Reference Dates for Influenza Outbreaks.** *RODS Laboratory Technical Report* 2005.

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6947/7/29/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

