Méthodes de détection d'événements inhabituels

Yann Le Strat

Département des Maladies Infectieuses
Institut de Veille Sanitaire (InVS)
Saint-Maurice, France

Objectifs de la surveillance

Les trois objectifs classiquement attribués à la surveillance sont :

- Une meilleure connaissance de l'évolution de la maladie dans un but d'action
- Apporter une aide à l'évaluation des politiques et des stratégies de contrôle et de prévention
- Détecter des phénomènes inhabituels pour alerter

Contexte

On se place dans le cadre d'une surveillance (spécifique ou non)

- qui collecte de manière continue les cas
- ayant une certaine ancienneté (historique)
- ayant une certaine stabilité dans le temps (exhaustivité, nombre/type de participants)

Définition

On dit qu'un **événement** est **inhabituel** lorsque le nombre observé de cas est significativement supérieur au nombre attendu de cas

- soit dans un intervalle de temps (jour, semaine)
- soit dans une zone géographique (pays, région, département)
- soit dans un intervalle de temps et dans une zone géographique

Les outils

Des modèles statistiques

On dispose aujourd'hui :

- d'une vingtaine de méthodes statistiques
- d'un logiciel permettant de les exécuter facilement (librairie R `surveillance`)

Une excellente revue des méthodes de détection

Journal of the
Royal Statistical Society
SERIES A
Statistics
in Society

J. R. Statist. Soc. A (2012)
175, *Part* 1, *pp.* 49–82

Statistical methods for the prospective detection of infectious disease outbreaks: a review

Steffen Unkel, C. Paddy Farrington and Paul H. Garthwaite

The Open University, Milton Keynes, UK

Chris Robertson

University of Strathclyde, Glasgow, UK

and Nick Andrews

Health Protection Agency, London, UK

[Received August 2010. Revised March 2011]

Summary. Unusual clusters of disease must be detected rapidly for effective public health interventions to be introduced. Over the past decade there has been a surge in interest in statistical methods for the early detection of infectious disease outbreaks. This growth in interest has given rise to much new methodological work, ranging across the spectrum of statistical methods. The paper presents a comprehensive review of the statistical approaches that have been proposed. Applications to both laboratory and syndromic surveillance data are provided to illustrate the various methods.

Keywords: Biosurveillance; Clusters; Control chart; Epidemics; Infectious diseases; Outbreak; Prospective detection; Surveillance

Librairie R surveillance

Package ‘surveillance’

February 20, 2015

Title Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena

Version 1.8-3

Date 2015-01-04

Author Michael Höhle [aut, cre, ths], Sebastian Meyer [aut], Michaela Paul [aut], Leonhard Held [ctb, ths], Thais Correa [ctb], Mathias Hofmann [ctb], Christian Lang [ctb], Juliane Manitz [ctb], Andrea Riebler [ctb], Daniel Sabanés Bové [ctb], Maëlle Salmon [ctb], Dirk Schumacher [ctb], Stefan Steiner [ctb], Mikko Virtanen [ctb], Valentin Wimmer [ctb], R Core Team [ctb] (A few code segments are modified versions of code from base R)

Maintainer Michael Höhle <hoehle@math.su.se>

Depends R (>= 3.0.2), methods, grDevices, graphics, stats, utils, sp (>= 1.0-15), xtable, polyCub (>= 0.4-3)

Imports Rcpp (>= 0.11.0), MASS, Matrix, spatstat (>= 1.36-0)

LinkingTo Rcpp

Suggests parallel, grid, gridExtra, lattice, colorspace, scales, animation, msm, spc, quadprog, memoise, polyclip, rgeos, gpclib, maptools, intervals, spdep, numDeriv, maxLik, testthat, coda, splancs, gamlss, INLA, runjags

Description A package implementing statistical methods for the modeling and

Trois points importants

1. Quelle(s) méthode(s) choisir ?
2. Doit-on prendre en compte les délais de déclaration ?
3. Comment les rendre facilement utilisables ?

1. Quelle(s) méthode(s) choisir ?

Performances des méthodes

1. Intérêt d'utiliser plusieurs méthodes
2. Les méthodes n'ont pas les mêmes performances en termes de sensibilité, spécificité, VPP, VPN, etc.
3. Veut-on être sensible, spécifique, les deux ?
   $\rightarrow$ le choix de l'épidémiologiste
4. Absence d'un travail mesurant les performances des méthodes
   $\rightarrow$ le travail du statisticien

Performances des méthodes

Évaluation et comparaison de 21 méthodes (thèse en cours) basées sur une étude de simulation anglaise (Noufaily *et al*, 2012).

Les méthodes les plus sensibles		Les méthodes les plus spécifiques		Un pack intéressant
Bayes 3	(88%)	Improved Farrington	(99%)	Bayes 3
Bayes 2	(86%)	Original Farrington	(98%)	RKI 3
Cusum GLM	(84%)	GLR neg binomial	(98%)	CDC
Bayes 1	(82%)	CDC	(96%)	Improved Farrington
Cusum GLM Rossi	(81%)	RKI 2	(95%)	Cusum GLM Rossi

Noufaily A et al (2012) An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*.

2. Doit-on prendre en compte les délais de déclaration ?

Impact des délais de déclaration

- Les délais dégradent la rapidité (*timeliness*) de la détection
- Cela peut retarder les investigations, les prises de décisions
  $\rightarrow$ développements statistiques récents : nowcasting

Modélisation des délais de maladies infectieuses au RU

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2015)
178, *Part* 1, *pp.* 205–222

Modelling reporting delays for outbreak detection in infectious disease data

Angela Noufaily, Yonas Ghebremichael-Weldeselassie, Doyo Gragn Enki and Paul Garthwaite,

The Open University, Milton Keynes, UK

Nick Andrews and André Charlett

Public Health England, London, UK

and Paddy Farrington

The Open University, Milton Keynes, UK

[Received June 2013. Final revision November 2013]

Summary. The delay that necessarily occurs between the emergence of symptoms and the identification of the cause of those symptoms affects the timeliness of detection of emerging outbreaks of infectious diseases, and hence the ability to take preventive action. We study the delays that are associated with the collection of laboratory surveillance data in England, Wales and Northern Ireland, using 12 infections of contrasting characteristics. We use a continuous time spline-based model for the hazard of the delay distribution, along with an associated proportional hazards model. The delay distributions are found to have extremely long tails, the hazard at longer delays being roughly constant, suggestive of a memoryless process, though some laboratories appear to stop reporting after a certain delay. The hazards are found typically to vary strongly with calendar time, and to a lesser extent with season and recent organism frequency. In consequence, the delay distributions cannot be assumed to be stationary. These findings will inform the development of outbreak detection algorithms that take account of reporting delays.

Keywords: Delay; Hazard; Infectious disease; Penalized likelihood; Spline; Surveillance

Nowcasting - Méthode

BIOMETRICS 70, 993–1002
December 2014

DOI: 10.1111/biom.121

Bayesian Nowcasting during the STEC O104:H4 Outbreak in Germany, 2011

Michael Höhle[1,2,*] and Matthias an der Heiden[2]

[1]Department of Mathematics, Stockholm University, Kräftriket, 106 91 Stockholm, Sweden
[2]Department for Infectious Disease Epidemiology, Robert Koch Institute, Seestraße 10, 13353 Berlin, Germany
[*]*email:* hoehle@math.su.se

SUMMARY. A Bayesian approach to the prediction of occurred-but-not-yet-reported events is developed for application in re
time public health surveillance. The motivation was the prediction of the daily number of hospitalizations for the hemolyt
uremic syndrome during the large May–July 2011 outbreak of Shiga toxin-producing *Escherichia coli* (STEC) O104:H4
Germany. Our novel Bayesian approach addresses the count data nature of the problem using negative binomial sampl
and shows that right-truncation of the reporting delay distribution under an assumption of time-homogeneity can be hand
in a conjugate prior-posterior framework using the generalized Dirichlet distribution. Since, in retrospect, the true numl
of hospitalizations is available, proper scoring rules for count data are used to evaluate and compare the predictive qual
of the procedures during the outbreak. The results show that it is important to take the count nature of the time series i
account and that changes in the delay distribution occurred due to intervention measures. As a consequence, we extend t
Bayesian analysis to a hierarchical model, which combines a discrete time survival regression model for the delay distribut
with a penalized spline for the dynamics of the epidemic curve. Altogether, we conclude that in emerging and time-criti
outbreaks, nowcasting approaches are a valuable tool to gain information about current trends.

KEY WORDS: Infectious disease epidemiology; Real-time surveillance; Reporting delay; Truncation.

Nowcasting (package surveillance)

nowcast	*Adjust a univariate time series of counts for observed but-not-yet-reported events*

Description

Nowcasting can help to obtain up-to-date information on trends during a situation where reports about events arrive with delay. For example in public health reporting, reports about important indicators (such as occurrence of cases) are prone to be delayed due to for example manual quality checking and reporting system hierarchies. Altogether, the delays are subject to a delay distribution, which may or may not vary over time.

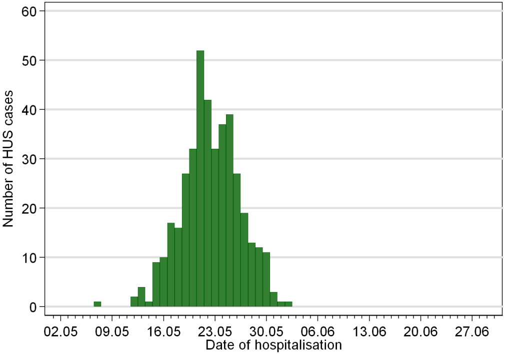
Usage

```
nowcast(now,when,data,dEventCol="dHospital",dReportCol="dReport",
        method=c("bayes.notrunc","bayes.notrunc.bnb","lawless","bayes.trunc",
                 "unif","bayes.trunc.ddcp"),
        aggregate.by="1 day",
        D=15, m=NULL,
        control=list(
            dRange=NULL,alpha=0.05,nSamples=1e3,
            N.tInf.prior=c("poisgamma","pois","unif"),
            N.tInf.max=300, gd.prior.kappa=0.1,
            ddcp=list(ddChangepoint=NULL,
                logLambda=c("iidLogGa","tps","rw1","rw2"),
```

Nombre quotidien d'hospitalisations pour SHU en Allemagne au 3 juin 2011

Situation on June 3

HUS cases by date of hospitalisation



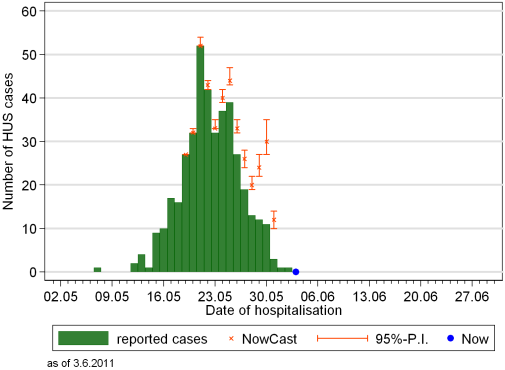
as of 3.6.2011

On aurait envie d'estimer les cas qui n'ont pas encore été rapportés

Evaluation

Now-Cast on June 3 (with pointwise confidence intervals)

HUS cases by date of hospitalisation



an der Heiden, Wadl and Höhle Now-Casting 9/ 15

Nombre quotidien d'hospitalisations pour SHU en Allemagne (01/05/2011-06/07/2011)

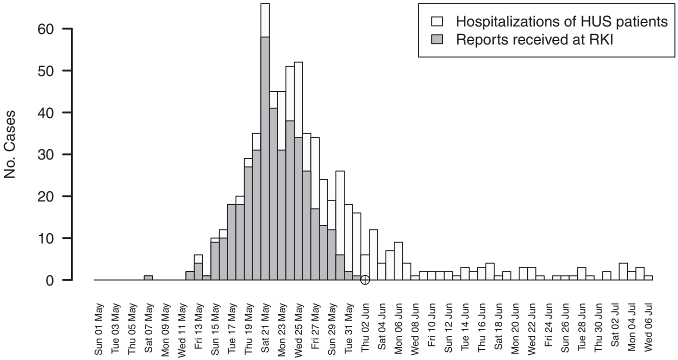


Figure 1. Daily number of hospitalizations due to HUS during the outbreak as available in retrospect. Also shown in darkgray are the number of available hospitalization reports at the RKI as of 2011-06-02 (indicated by the chrosshairs symbol).

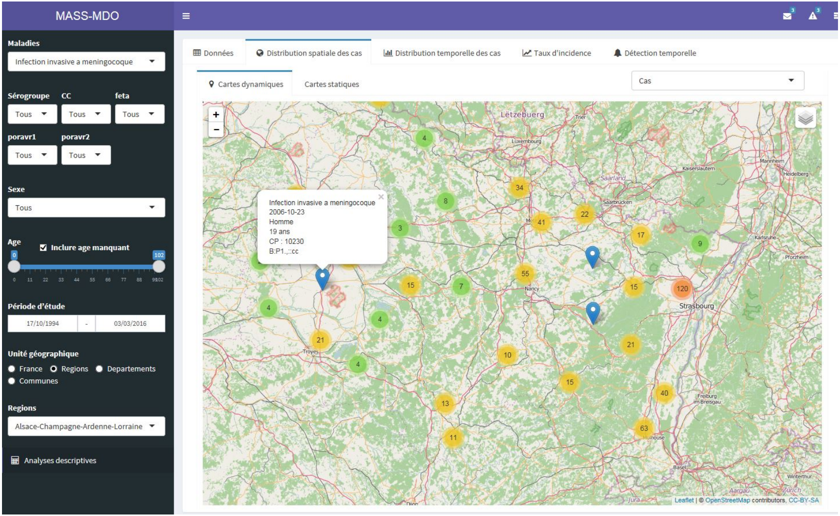
3. Comment rendre ces méthodes facilement utilisables ?

Approche efficiente

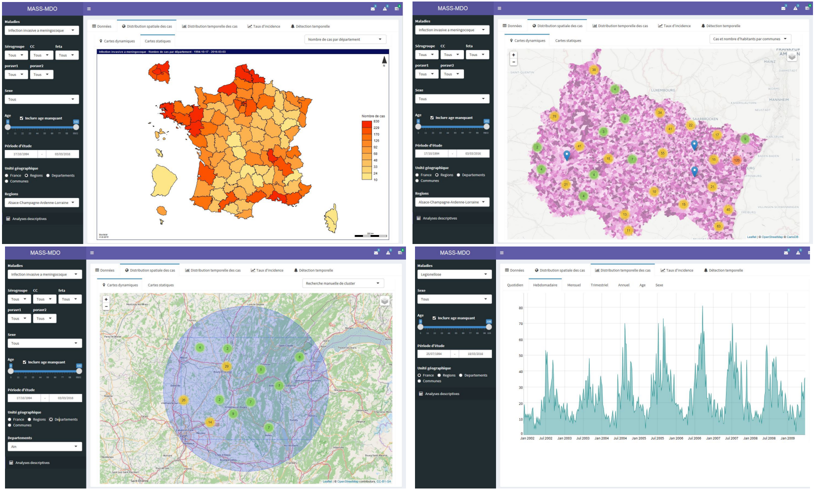
L'approche la plus efficiente est d'intégrer ces méthodes dans une application web :

- On peut aujourd'hui créer en quelques mois une application web (avec R `shiny` par exemple)
- Idée : créer des modules d'analyses, greffés aux systèmes de surveillance
- Applications qui répondent aux demandes des utilisateurs
- Aucune connaissance statistique et/ou informatique requise pour l'utilisateur
- Expériences récentes à l'Anses (Acteolab) et à l'InVS (SursauD, MDO)

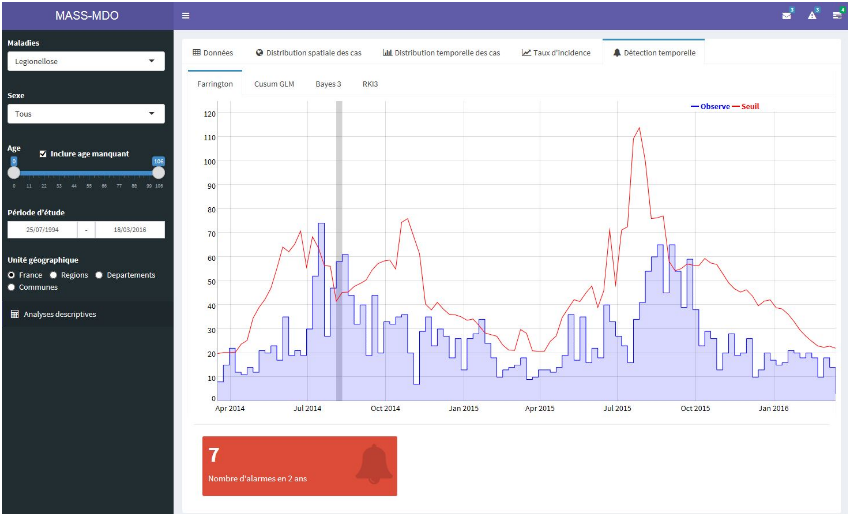
Une application web pour la surveillance des MDO



Des sorties diverses



Des méthodes de détection temporelle



Discussion

- En méthodologie, l'intégration des délais de déclaration dans les méthodes de détection semble être une perspective intéressante.
- Il faut pour cela faire un travail spécifique sur les délais en amont.
- Concernant les outils, R est devenu un logiciel incontournable par sa richesse de fonctions et sa puissance de calcul.
- Ces outils sont gratuits et leur utilisation devraient faire partie intégrante de l'environnement informatique pour la surveillance.