Automated and Early Detection of Disease Outbreaks

Kasper Schou Telkamp

Supervisors: Jan Kloppenborg Møller, Lasse Engbo Christiansen

Master Thesis Defence

14th of August 2023

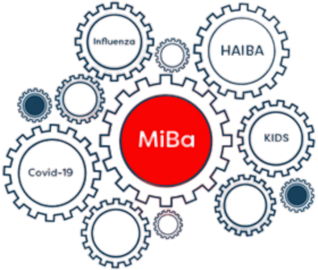
Technical University of Denmark

DTU Compute

Department of Applied Mathematics and Computer Science

Motivation

- Establishment of the Danish Microbiology Database (MiBa) by Statens Serum Institut (SSI) in 2010
- Great opportunity for data analysis
- No fully automated procedures in place at SSI



Research goals

- Review of existing literature on statistical methods for detecting disease outbreaks
- Identification and implementation of state-of-the-art methods for detection of disease outbreaks
- Formulation of hierarchical models for the individually notifiable diseases
- Development of an automated method, based on the hierarchical models, for automated and early detection of disease outbreaks
- Comparison of the developed method and state-of-the-art methods in one or more case study
- Comparison of the developed method and state-of-the-art methods in a simulation study

Research goals

- Review of existing literature on statistical methods for detecting disease outbreaks
- Identification and implementation of state-of-the-art methods for detection of disease outbreaks
- Formulation of **hierarchical** models for the individually notifiable diseases
- Development of an **automated method**, based on the hierarchical models, for automated and early detection of disease outbreaks
- Comparison of the developed method and state-of-the-art methods in **one** or more case study
- Comparison of the developed method and state-of-the-art methods in a **simulation study**

State-of-the-art algorithms

State-of-the-art algorithms for aberration detection is presented in Salmon, Schumacher, and Höhle 2016 and implemented in the R package **surveillance**. The R package includes the method introduced by Farrington et al. 1996 together with the subsequently improved method proposed by Noufaily et al. 2013.

Novel algorithm

The novel algorithm utilizes a generalized mixed effects model or a hierarchical mixed effects model as a modeling framework to model the count case observations $\boldsymbol{y}$ and assess the unobserved random effects $\boldsymbol{u}$. These random effects are used directly to characterize an outbreak.

Step 1: Modeling framework

- Assume a hierarchical Poisson Normal or Poisson Gamma model to reference data using a log link
- Incorporate covariates by supplying a model formula on the form

$$\log(\lambda_{it}) = \boldsymbol{x}_{it}\boldsymbol{\beta} + \log(n_{it}), \quad i = 1, \ldots, m, \quad t = 1, \ldots, T \tag{1}$$

- Account for structural changes in the time series using a rolling window of width k

Step 2: Inference of random effects

- Infer one-step ahead random effects u_{it_0} for each group using the fitted model
- Define outbreak detection threshold U_{t_0} as a quantile of the second stage model's random effects distribution
- Use either a Gaussian or Gamma distribution with respective plug-in estimates

Step 3: Parameter estimations and outbreak detection

- Compare inferred random effects u_{it_0} to an threshold U_{t_0}
- Raise and alarm if the inferred random effect exceeds the threshold, i.e. $u_{it_0} > U_{t_0}$
- Omit outbreak related observations from future parameter estimation

Formulation of hierarchical models

Poisson Normal

$$\boldsymbol{Y}|\boldsymbol{u} \sim \mathrm{Pois}\left(\boldsymbol{\lambda}\exp(\boldsymbol{u})\right)$$
$$\boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, I\sigma^2)$$

Poisson Gamma

$$\boldsymbol{Y}|\boldsymbol{u} \sim \mathrm{Pois}(\boldsymbol{\lambda}\boldsymbol{u})$$
$$\boldsymbol{u} \sim \mathrm{G}(\boldsymbol{1}/\phi, \phi)$$

Case study

Shiga toxin (verotoxin)-producing *Escherichia coli* (STEC)

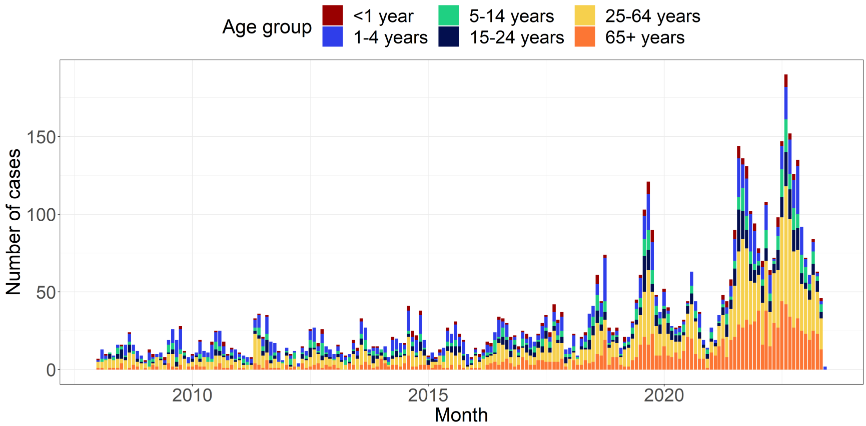


Figure: A stacked bar graph illustrating the number of monthly STEC cases observed in the period from 2008 to 2022 for the six age groups.

Constant model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \log(n_{it}) \tag{2}$$

- λ_{it} is the outbreak intensity at time t for age group i.
- $\beta(ageGroup_i)$ is the fixed effect specific to age group i.
- $\log(n_{it})$ acts as an offset, accounting for the population size at time t for age group i.

Trend model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \beta_{trend}t + \log(n_{it}) \quad (3)$$

- In addition to Model 1, includes a trend component.
- β_{trend} quantifies the rate of change in the outbreak intensity over time.

Seasonality model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \sin\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\sin} + \cos\left(2\frac{\pi \cdot \tau_t}{12}\right)\beta_{\cos} + \log(n_{it}) \tag{4}$$

- In addition to Model 1, incorporates an annual seasonality pattern.
- τ_t represents the time period t within a year (1-12).
- $\beta_{\sin}$ and $\beta_{\cos}$ capture the effect of the seasonal pattern.

Combined trend and seasonality model

$$\log(\lambda_{it}) = \beta(ageGroup_i) + \beta_{trend}t + \sin\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\sin} + \cos\left(\frac{2\pi \cdot \tau_t}{12}\right)\beta_{\cos} + \log(n_{it}) \quad (5)$$

- Builds upon previous models, combining trend and seasonality components.
- Includes both β_{trend}, $\beta_{\sin}$, and $\beta_{\cos}$ parameters.

Estimated one-step ahead random effects

- The upper bound U_{t_0} is calculated based on the 90% quantile of the distribution of the random effects
- If the one-step ahead random effects u_{it_1} exceeds the upper bound U_{t_0} an alarm is raised
- In the hierarchical Poisson Normal model (left), the random effects are exponentiated to transform them into the same domain as the hierarchical Poisson Gamma model (right)

- 30 alarms are generated using the hierarchical Poisson Normal framework, while 31 alarms are generated using the hierarchical Poisson Gamma framework.
- A great number of alarms are generated in the period from March 2021 to March 2022

Case study

Estimated one-step ahead random effects

- The upper bound U_{t_0} is calculated based on the 90% quantile of the distribution of the random effects
- If the one-step ahead random effects u_{it_1} exceeds the upper bound U_{t_0} an alarm is raised
- In the hierarchical Poisson Normal model (left), the random effects are exponentiated to transform them into the same domain as the hierarchical Poisson Gamma model (right)

- 30 alarms are generated using the hierarchical Poisson Normal framework, while 31 alarms are generated using the hierarchical Poisson Gamma framework.
- A great number of alarms are generated in the period from March 2021 to March 2022

References

Farrington, C. P. et al. (1996). “A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159.3, pp. 547–563. ISSN: 09641998, 1467985X. URL: http://www.jstor.org/stable/2983331 (visited on 01/27/2023).

Noufaily, Angela et al. (2013). “An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems”. en. In: *Online Journal of Public Health Informatics* 32.7, pp. 1206–1222.

Salmon, Maëlle, Dirk Schumacher, and Michael Höhle (2016). “Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance”. In: *Journal of Statistical Software* 70.10, pp. 1–35. DOI: 10.18637/jss.v070.i10. URL: https://www.jstatsoft.org/index.php/jss/article/view/v070i10.