L01 Introduction

Michael Höhle[1]

[1]Department of Mathematics, Stockholm University, Sweden

🐦 m_hoehle

STA427 FS2021
Statistical Methods in Infectious Disease Epidemiology
Epidemiology, Biostatistics and Prevention Institute
University of Zurich, Switzerland

University of
Zurich^UZH

Outline

Outline

Definitions and aim of this lecture

Infectious disease epidemiology

Characterizes the epidemiological analysis of *infectious diseases*. Interest lies in the detection and understanding of epidemics. One possible aim would be the ability to better control outbreaks.

Aims of this lecture:

- Give a taste of how statistical modelling can be of use in infectious disease epidemiology
- Illustrate this by plenty of examples from theory and practice

Statistical modelling of infectious diseases

Three reasons that classical statistical approaches are not immediately applicable for infectious disease data:

1. Data are rarely a result of planned experiments
2. Individuals are not independent (a case may also be a risk factor)
3. The infection process is only partially observable

Mathematical models for communicable diseases (1)

- Mathematical modeling of infectious diseases has become a key tool in order to understand, predict and control the spread of infections.
- The intention of epidemic modeling is to model the spread of a disease in a population made up of a (possible large) integer number of individuals.
- To simplify the description of the population, it is common to use a compartmental approach to modeling. Here, the population is divided into classes of *susceptible*, *infective* and *recovered* individuals.

Mathematical models for communicable diseases (2)

- Disease dynamics can then be characterized by a mathematical description of each individual's transitions between classes, subject to the state of the other individuals in the population.
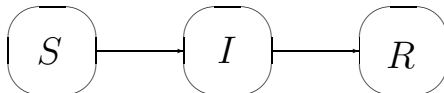
- Mathematical models differ in whether they consider the infection process as *deterministic* or *stochastic*.

- Another distinction between models is whether they operate in continuous-time or discrete-time.

Biostatistical thoughts on the mathematical modelling

- The underlying phenomena are of stochastic nature – thus modelling should, as far as possible, rely on *stochastic* approaches
- Inference on model structure and model parameters needs to be guided by adequate data
- Handling of *uncertainty* is a key virtue of a *statistical* approach to this task

Aim of this course:

Give an impression of what this statistical and data-driven approach towards epidemiological modelling of infectious diseases looks like

Outline

SIR transmission model (1)

- Continuous-time **S**usceptible-**I**nfected-**R**ecovered (SIR) type model.
- Example: Homogeneous SIR model in a closed population of size N, $S(t) + I(t) + R(t) = N$ at all times.
- Continuous time Markov chain (CTMC) model:

$$
\begin{array}{ccc}
 & \text{Transition} & \text{Rate} \\
(S(t), I(t)) & \begin{array}{l} \to (S(t) - 1, I(t) + 1) \\ \to (S(t), I(t) - 1) \end{array} & \begin{array}{l} \frac{\beta}{N} \cdot S(t) \cdot I(t) \\ \gamma \cdot I(t) \end{array}
\end{array}
$$

- Deterministic ordinary differential equation (ODE) system:

$$
\frac{dS(t)}{dt} = -\frac{\beta}{N} S(t) I(t), \quad \frac{dI(t)}{dt} = \frac{\beta}{N} S(t) I(t) - \gamma I(t), \quad \frac{dR(t)}{dt} = \gamma I(t).
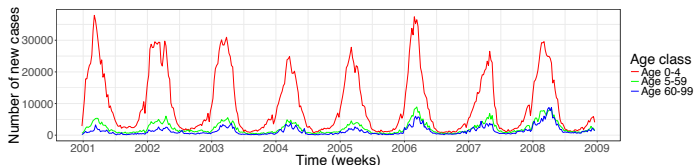$$

SIR transmission model (2)

- Example with $\beta = 1$, $\gamma = 0.5$ and $I(0) = \lceil 0.001 \cdot N \rceil$ and $S(0) = N - I(0)$.
- Illustration of the no. of new infections per week, i.e. $N_{S \to I}(t) - N_{S \to I}(t-1)$, for two population sizes (from 1,000 sims):
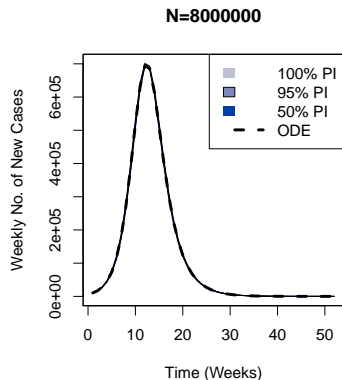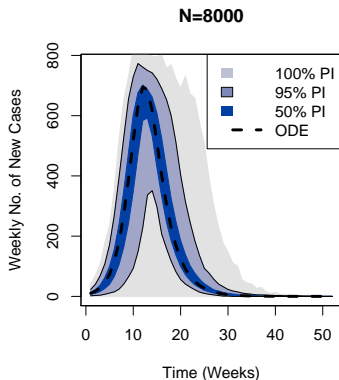
Rotavirus Dynamics in Germany (1)[1]

- Rotavirus is a gastrointestinal disease occurring mainly in children.

- Epidemiological interest is in the disease dynamics at population level in order to prepare a model to study the effect of vaccination
- Perform parameter inference from data, use model selection

[1]Stocks, T., T. Britton, and M. Höhle, 2020, Biostatistics

Rotavirus Dynamics in Germany (2)

- Age-structured dynamic model based on an CTMC with both structural noise and observational noise

- Weekly number of incidence cases in age group k is assumed to be

 $$X_k^*(t_n) \sim \mathrm{NegBin}(\Delta N_{S_k, I_k}(t_n), d),$$

  where $\Delta N_{S_k, I_k}(t_n) = N_{S_k, I_k}(t_n) - N_{S_k, I_k}(t_{n-1})$ is the number of newly infected during the time interval $[t_{n-1}, t_n)$ in age-class k

- Inference using the iterated filtering method (Ionides, Bretó, and King 2006) and model selection using AIC
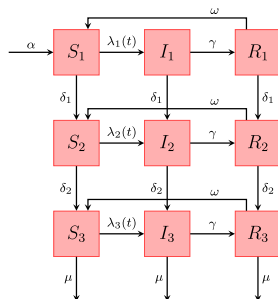
Example: Invasive Meningococcal Disease (IMD)

- IMD is a life-threatening infectious disease triggered by the bacterium *neisseria meningitidis (aka meningococcus)*
- *Two most common finetypes in Germany in 2002–2008: 336 cases of B:P1.7-2,4:F1-5, 300 cases of C:P1.5,2:F3-3*
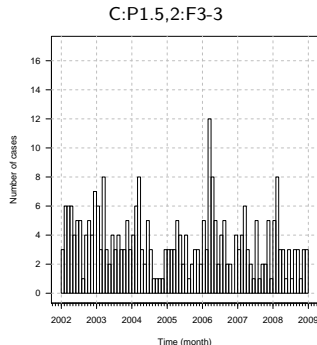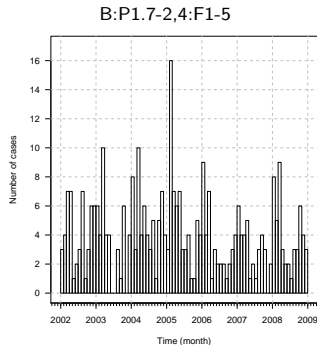
Count Data Time Series Modelling and Monitoring (1)

- Development of automated algorithms for the prospective detection of abnormalities in such univariate count data times series
- Initial step: Count data time series regression modelling taking seasonality and other factors into account, e.g.

$$Y_t | \boldsymbol{x}_t \sim \text{NegBin}(\mu_{0,t}, d), \quad t = 1, 2, \ldots$$

with expectation $\log(\mu_{0,t}) = \beta_0 + \boldsymbol{x}_t' \boldsymbol{\beta}$ and variance $\mu_{0,t} + d\mu_{0,t}^2$.

- Next step: Prospective change-point detection from β_0 to $(\beta_0 + \kappa)$ within a SPC framework, i.e. Shewhart-like, likelihood ratio or generalized likelihood ratio CUSUM[2]

---

[2] Höhle, M., and M. Paul, 2008, Computational Statistics & Data Analysis

Count Data Time Series Modelling and Monitoring (2)

- Example: Trend + penalized periodic spline model with CUSUM detection of 50% increase in $\mu_{0,t}$ for type C
- Fix alarm threshold s.t. $P(\text{first alarm time} \leq 12|\text{no change}) = 0.02$
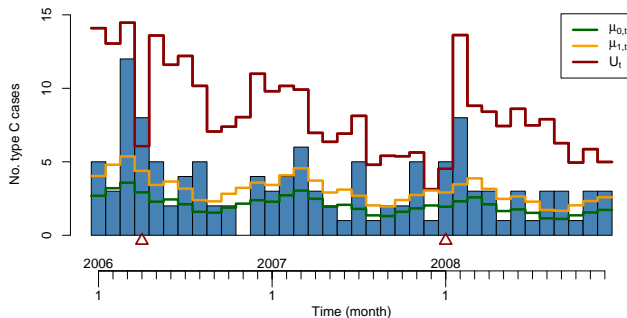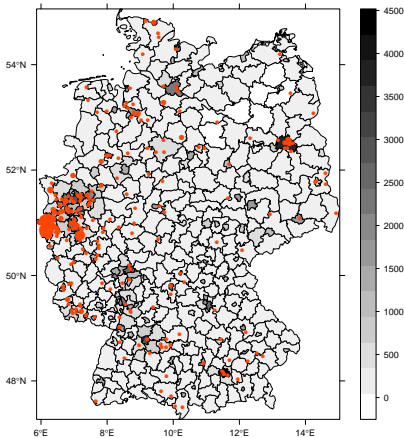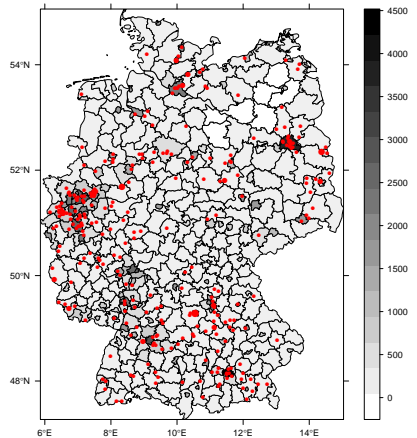
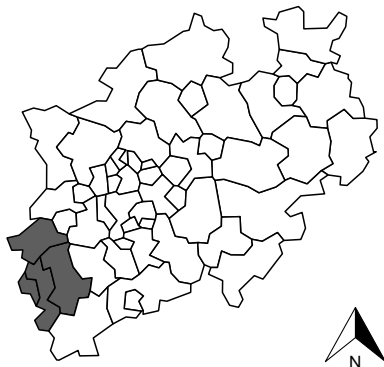# Spatial Distribution

B:P1.7-2,4:F1-5

C:P1.5,2:F3-3

Spatio-temporal cluster detection

- Monitoring the multivariate time series at district level using spatio-temporal cluster detection techniques detects a signal Feb-Aug 2005

- The core cluster consists of four districts in North Rhine-Westphalia, one of them the city Aachen

Nowcasting (1)

- **Structural delay** between exposure, day of symptom onset (DSO), testing and the recording of the case in the database
- Relevant for the understanding of disease dynamics is the exposure event of cases (alternatively: DSO) not the reporting date
- However, due to the above delays cases might not be in the central surveillance database yet (occurred-but-not-yet-reported events, OBNYR)

Nowcasting (2)

Current nowcast for Bavaria[3]:
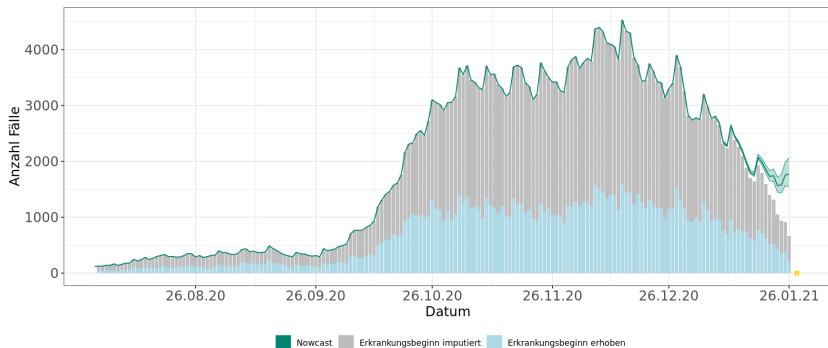
Statistisches Beratungslabor StaBLab, LMU München; Department of Mathematics, Stockholm University
Daten: Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit LGL

---

[3] https://corona.stat.uni-muenchen.de/nowcast/

Nowcasting (3)

- We[4] perform a prediction of OBNYR events through a two-component Bayesian hierarchical model extending previous work from the large 2011 STEC O104:H4 outbreak in Germany[5]:
  - Model for the dynamics of the reporting delay distribution
  - Model for the dynamics of the epidemic curve
- Inference using MCMC in STAN
- Evaluation of the models on Mar-Jun data by proper scoring rules
- Use nowcast → back-projection → segmented regression pipeline to get real-time assessment of interventions

---

[4]Günther, F., A. Bender, K. Katz, H. Küchenhoff, and M. Höhle, , Biometrical Journal

[5]Höhle, M., and M. an der Heiden, 2014, Biometrics

Outline

1. Introduction to infectious disease epidemiology

2. Examples

3. Software Implementations

4. Discussion

The R package `surveillance`

- Philosophy: Methodology has to be supported by an implementation in order to ensure transparency, reproducibility and to facilitate its use

- The R package `surveillance` provides a toolbox for statistical inference in infectious disease epidemiology[6,7]

- The pkg is in use in several European public health institutes and forms backbone of the AVOID System at the Robert Koch Institute monitoring more more than 100,000 time series on a weekly basis[8]

---

[6] Salmon, M., D. Schumacher, and M. Höhle, 2016, Journal of Statistical Software

[7] Meyer, S., L. Held, and M. Höhle, 2017, Journal of Statistical Software

[8] Salmon, M., D. Schumacher, H. Burman, C. Frank, H. Claus, and M. Höhle, 2016, Eurosurveillance

R Epidemics Consortium (RECON)

- RECON is a NGO dedicated to making analytical tools for informing the response to disease outbreaks, health emergencies and humanitarian crises in R
- RECON curates a number of R packages for or handling, visualising, and analysing outbreak data
- https://www.reconlearn.org/ provides free and open training resources to respond to outbreaks, health emergencies and humanitarian crises

Detecting public health threats in Twitter data

- Instead of structured surveillance data, time series could also be generated from Twitter search query data
- ECDC provides an open-source R package `epitweetr` (Orchard and Espinosa 2020) for **epidemic intelligence**
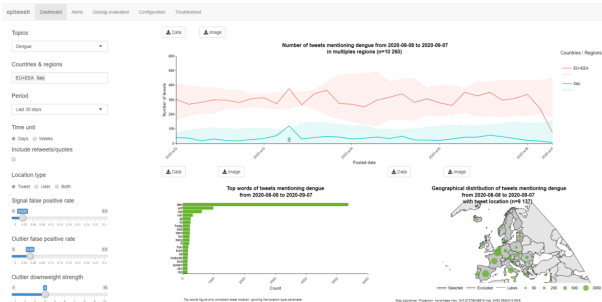
- Challenges: unstructured data, machine learning for geo-tagging tweets, representativeness, data volume

Outline

The Course

Course contents

1. Introduction to epidemic modelling
   - Compartmental models Deterministic and stochastic SIR-type models
   - Simulation and inference for SIR-type models
2. Reproduction numbers and their estimation
3. Delays
   - Back-calculation method
   - Nowcasting
4. Vaccination: effectiveness, efficacy and safety
5. Temporal and spatio-temporal detection of clusters
6. COVID-19 modelling

Literature I

📄 Günther, F., et al. "Nowcasting the COVID-19 pandemic in Bavaria".
    Biometrical Journal n/a (n/a).
    doi:https://doi.org/10.1002/bimj.202000112. eprint: https:
    //onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.202000112.
    https:
    //onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202000112.

📄 Höhle, M., and M. an der Heiden. 2014. "Bayesian Nowcasting during
    the STEC O104:H4 Outbreak in Germany, 2011". Animations available
    from http://dx.doi.org/10.1111/biom.12194, Biometrics 70 (4):
    993–1002. doi:10.1111/biom.12194.

📄 Höhle, M., and M. Paul. 2008. "Count data regression charts for the
    monitoring of surveillance time series".
    Computational Statistics & Data Analysis 52 (9): 4357–4368.

Literature II

📄 Ionides, E. L., C. Bretó, and A.A. King. 2006. "Inference for nonlinear dynamical systems".
Proceedings of the National Academy of Sciences of the USA 103, no. 49 (): 18438–18443.

📄 Meyer, S., L. Held, and M. Höhle. 2017. "Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance". Also available as vignettes of the R package surveillance.
Journal of Statistical Software 77 (11). doi:10.18637/jss.v077.i11.

📄 Orchard, Francisco, and Laura Espinosa. 2020.
epitweetr: Early Detection of Public Health Threats from Twitter Data.
R package version 0.1.20.
https://CRAN.R-project.org/package=epitweetr.

Literature III

📄 Salmon, M., D. Schumacher, and M. Höhle. 2016. "Monitoring Count
   Time Series in R: Aberration Detection in Public Health Surveillance".
   Also available as vignette of the R package surveillance.
   Journal of Statistical Software 70 (10). doi:10.18637/jss.v070.i10.

📄 Salmon, M., et al. 2016. "A system for automated outbreak detection of
   communicable diseases in Germany". Eurosurveillance 21 (13):
   pii=3018. doi:10.2807/1560-7917.ES.2016.21.13.30180.

📄 Stocks, T., T. Britton, and M. Höhle. 2020. "Model selection and
   parameter estimation for dynamic epidemic models via iterated
   filtering: application to rotavirus in Germany". Biostatistics 21, no. 3
   (): 400–416. ISSN: 1465-4644. doi:10.1093/biostatistics/kxy057.
   eprint: https://academic.oup.com/biostatistics/article-
   pdf/21/3/400/33416299/kxy057.pdf.
   https://doi.org/10.1093/biostatistics/kxy057.