

Social contact data in endemic-epidemic models and probabilistic forecasting with surveillance

Sebastian Meyer

Institute of Medical Informatics, Biometry, and Epidemiology

Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

5 July 2017

BRUSSELS
user!
2017!

Joint work with Johannes Bracher and Leonhard Held (University of Zurich)

World Health Organization 2014

Forecasting disease outbreaks is still in its infancy, however, unlike weather forecasting, where substantial progress has been made in recent years.

World Health Organization 2014

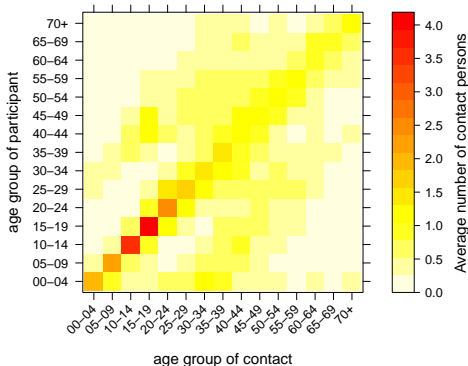
Forecasting disease outbreaks is still in its infancy, however, unlike weather forecasting, where substantial progress has been made in recent years.

Key requirements to forecast infectious disease incidence

- **Multivariate view** to predict incidence in different regions and subgroups
- Stratified **count time series** from routine public health surveillance
- Useful statistical models to reflect **forecast uncertainty**
- **Predictive model assessment**

Infectious disease spread ~ social contacts

```
library("hhh4contacts")
plotC(contactmatrix(grouping = NULL))
```

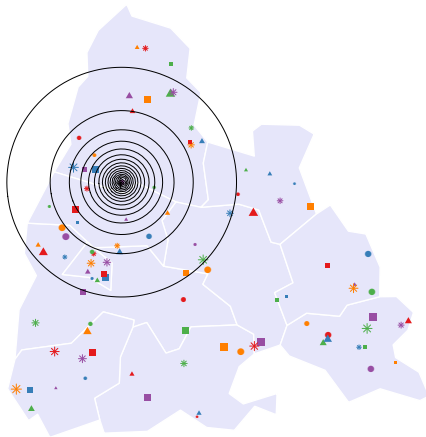


EU-funded POLYMOD study

[Mossong et al. 2008]:

- 7 290 participants from eight European countries recorded contacts during one day
- Contact characteristics were similar across countries
- Remarkable mixing patterns with respect to age

Infectious disease spread ~ location and distance



Tobler's First Law of Geography:

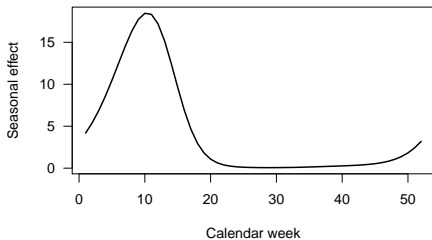
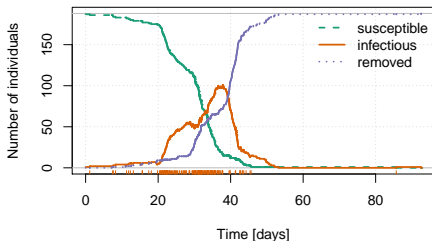
Everything is related to everything else, but near things are more related than distant things.

Specifically [e.g., Meyer and Held 2017]:

Spatial interaction decays as a power law.

Regional characteristics may also affect disease spread, e.g., rural vs. urban municipalities

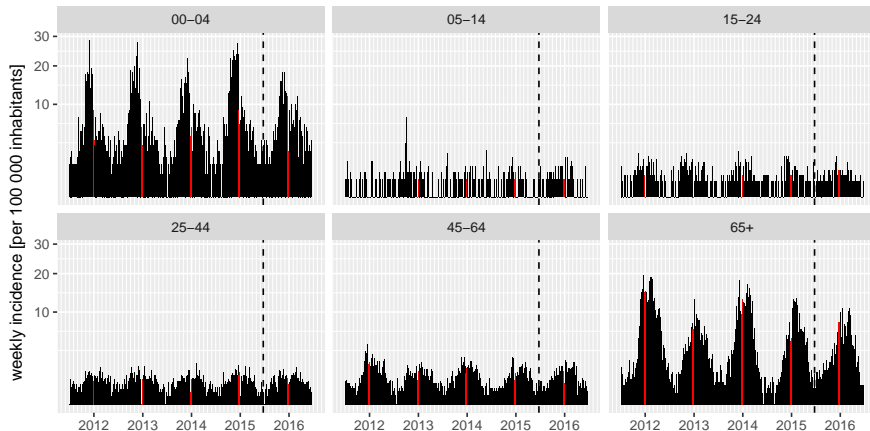
Infectious disease spread ~ time



- Occasional outbreaks
- Limited infectious period
- Seasonality (influenza, measles, norovirus gastroenteritis, ...)

Case study: norovirus gastroenteritis in Berlin, 2011–2016

```
noroBEg <- noroBE(by = "agegroups", timeRange = c("2011-w27", "2016-w26"))
```



Lab-confirmed counts from `survstat.rki.de`, stratified by 12 city districts and 6 age groups

An age-stratified, spatio-temporal model

Negative binomial likelihood for infectious disease counts Y_{grt} with endemic-epidemic mean decomposition:

$$\mu_{grt} = v_{grt} + \phi_{grt} \sum_{g',r'} c_{g'g} w_{r'r} Y_{g',r',t-1}$$

An age-stratified, spatio-temporal model

Negative binomial likelihood for infectious disease counts Y_{grt} with **endemic-epidemic** mean decomposition:

$$\mu_{grt} = v_{grt} + \phi_{grt} \sum_{g',r'} c_{g'g} w_{r'r} Y_{g',r',t-1}$$

Log-linear predictors

v_{grt} and ϕ_{grt}

- Population offsets
- Seasonality
- Group-specific susceptibility
- Covariates, e.g., vaccination coverage

An age-stratified, spatio-temporal model

Negative binomial likelihood for infectious disease counts Y_{grt} with **endemic-epidemic** mean decomposition:

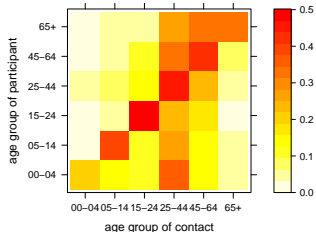
$$\mu_{grt} = v_{grt} + \phi_{grt} \sum_{g',r'} \boxed{c_{g'g}} w_{r'r} Y_{g',r',t-1}$$

Log-linear predictors

v_{grt} and ϕ_{grt}

- Population offsets
- Seasonality
- Group-specific susceptibility
- Covariates, e.g., vaccination coverage

Aggregated POLYMOD
contactmatrix() ($c_{g'g}$)



An age-stratified, spatio-temporal model

Negative binomial likelihood for infectious disease counts Y_{grt} with endemic-epidemic mean decomposition:

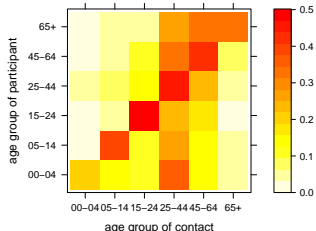
$$\mu_{grt} = v_{grt} + \phi_{grt} \sum_{g',r'} c_{g'g} w_{r'r} Y_{g',r',t-1}$$

Log-linear predictors

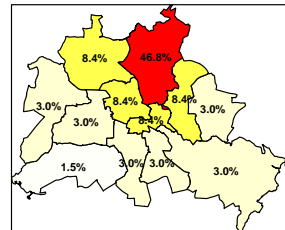
v_{grt} and ϕ_{grt}

- Population offsets
- Seasonality
- Group-specific susceptibility
- Covariates, e.g., vaccination coverage

Aggregated POLYMOD
contactmatrix() ($c_{g'g}$)

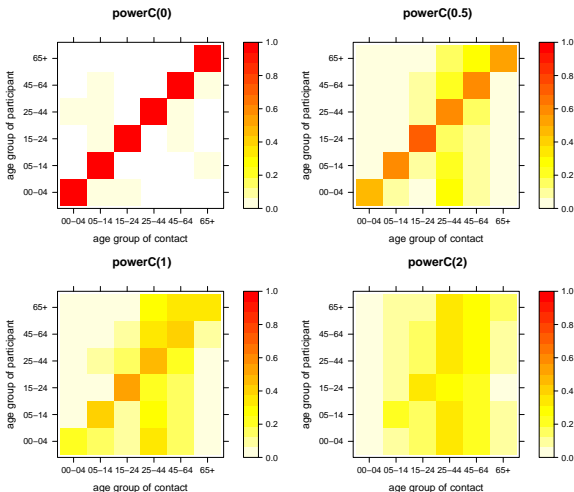


Spatial weights, e.g., power-law decay $w_{r'r} = (o_{r'r} + 1)^{-\rho}$



Power-adjustment of the contact matrix: $C^K := E\Lambda^K E^{-1}$

```
powerC <- make_powerC(contactmatrix(), normalize = TRUE)
```



Model estimation

Likelihood inference using surveillance: :hhh4() [Meyer, Held, and Höhle 2017]

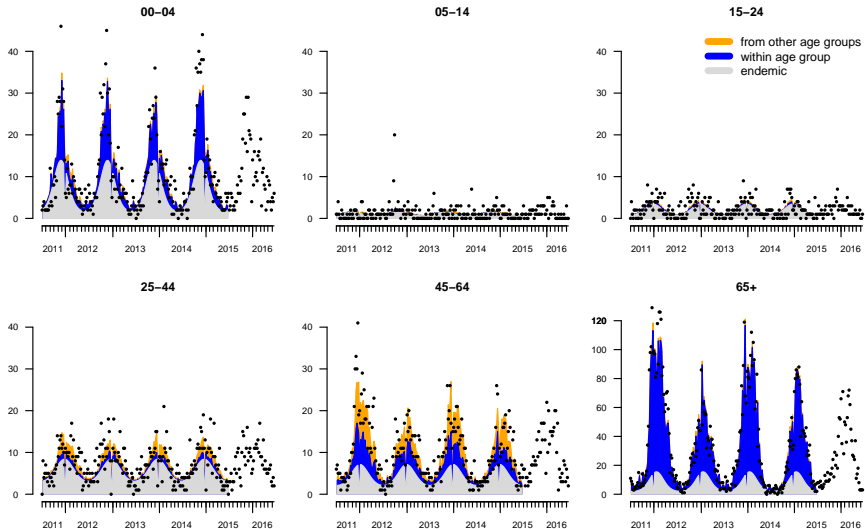
A “simple”, age-stratified, spatio-temporal model¹:

```
noroBEall <- noroBE(by = "all", flatten = TRUE, # 6 x 12 = 72 columns
                  timeRange = c("2011-w27", "2016-w26"))

fit <- hhh4(stsObj = noroBEall, control = list(
  end = list(f = addSeason2formula(~1),
             offset = prop.table(population(noroBEall), 1)),
  ne = list(f = ~1 + log(pop),
            weights = W_powerlaw(maxlag = 5, log = TRUE),
            scale = expandC(contactmatrix(), 12)),
  data = list(pop = prop.table(population(noroBEall), 1)),
  family = "NegBin1", subset = 2:(4*52)))
```

¹Full models in demo("hhh4contacts", package = "hhh4contacts")

Fitted mean by age group aggregated over districts

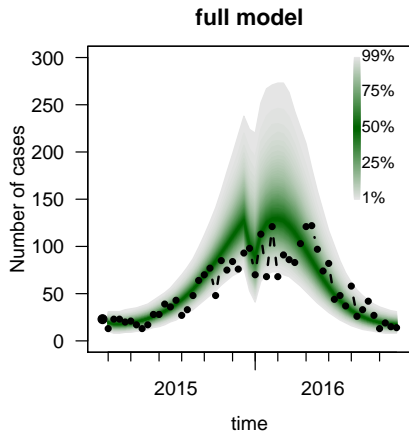
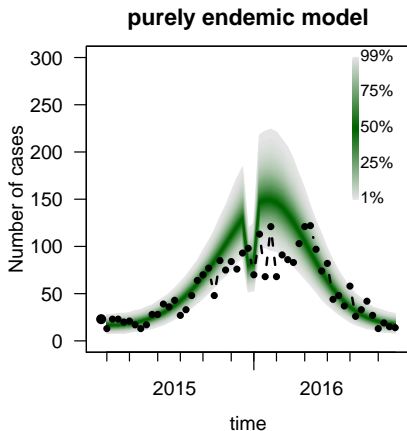


Prediction and validation

- AIC-based model comparison selects most complex model
- Is this choice also supported by predictive model assessment based on the last season?
- We can quantify **sharpness** and **calibration** of probabilistic
 - one-week-ahead forecasts (negative binomial)
 - long-term forecasts (via Monte Carlo simulation)
- **Proper scoring rules** as overall performance measures [Gneiting and Katzfuss 2014]
 - Assign penalty score based on the predictive distribution F and the actual observation y_{obs}
 - Example: Dawid-Sebastiani score

$$DSS(F, y_{obs}) = \log|\Sigma| + (y_{obs} - \mu)^\top \Sigma^{-1} (y_{obs} - \mu)$$

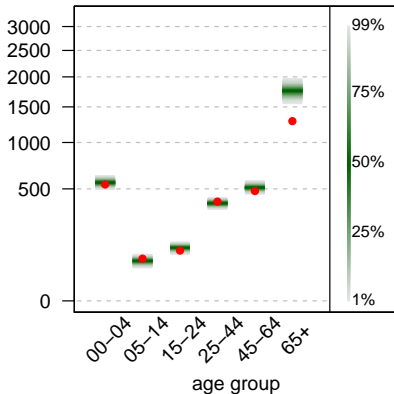
Target quantity: overall epidemic curve



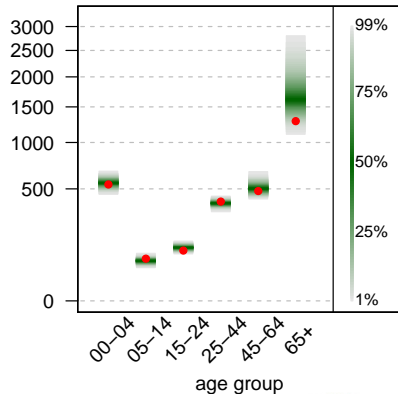
BRUSSELS
 USER!
 2017!

Target quantity: final size by age group

purely endemic model



full model



Summary and outlook

- Models do not perfectly represent individual-level disease transmission, but: still useful for prediction of aggregate-level surveillance counts
- Improved model fit and predictions by incorporating spatial weights and social contact data [Held, Meyer, and Bracher 2017]
- If the modelling goal is forecasting, use proper scoring rules to assess the quality of probabilistic forecasts
- surveillance currently implements the following univariate scores for Poisson and NegBin predictions: rps, dss, logs
- For continuous distributions: package scoringRules

Summary and outlook

- Models do not perfectly represent individual-level disease transmission, but: still useful for prediction of aggregate-level surveillance counts
- Improved model fit and predictions by incorporating spatial weights and social contact data [Held, Meyer, and Bracher 2017]
- If the modelling goal is forecasting, use proper scoring rules to assess the quality of probabilistic forecasts
- surveillance currently implements the following univariate scores for Poisson and NegBin predictions: rps, dss, logs
- For continuous distributions: package scoringRules

- User interface for multivariate scoring rules
- Binomial hhh4 models
- hhh4 add-on package (Johannes Bracher):
 - Analytical DSS of multivariate path forecasts
 - Distributed higher-order time lags

References

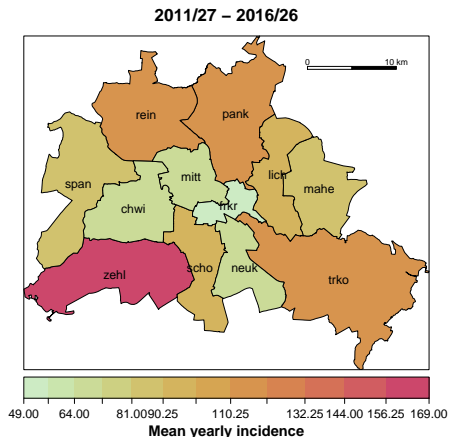
- Gneiting, Tilmann and Katzfuss, Matthias (2014). “Probabilistic forecasting”. In: *Annual Review of Statistics and Its Application* 1.1, pp. 125–151. DOI: 10.1146/annurev-statistics-062713-085831.
- Held, Leonhard, Meyer, Sebastian, and Bracher, Johannes (2017). “Probabilistic forecasting in infectious disease epidemiology: The 13th Armitage lecture”. In: *Statistics in Medicine* (in press). DOI: 10.1002/sim.7363.
- Meyer, Sebastian and Held, Leonhard (2017). “Incorporating social contact data in spatio-temporal models for infectious disease spread”. In: *Biostatistics* 18.2, pp. 338–351. DOI: 10.1093/biostatistics/kxw051.
- Meyer, Sebastian, Held, Leonhard, and Höhle, Michael (2017). “Spatio-temporal analysis of epidemic phenomena using the R package **surveillance**”. In: *Journal of Statistical Software* 77.11, pp. 1–55. DOI: 10.18637/jss.v077.i11.
- Mossong, Joël et al. (2008). “Social contacts and mixing patterns relevant to the spread of infectious diseases”. In: *PLoS Medicine* 5.3, e74. DOI: 10.1371/journal.pmed.0050074.
- World Health Organization (2014). “Anticipating epidemics”. In: *Weekly Epidemiological Record* 89.22, p. 244. URL: <http://www.who.int/wer>.

Questions? Comments? ✉ seb.meyer@fau.de

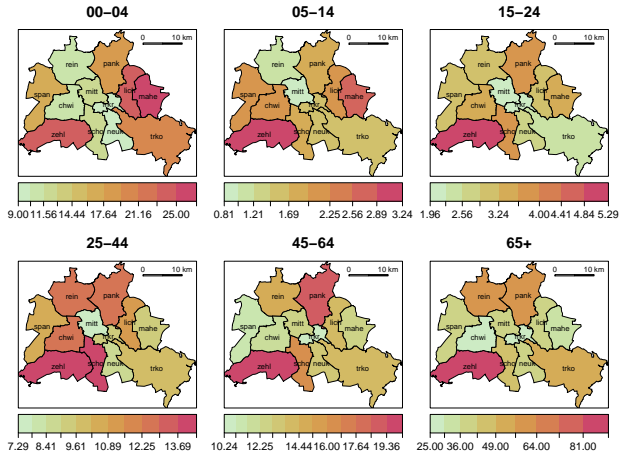
Appendix

Disease incidence map

```
noroBEr <- noroBE(by = "districts",
  timeRange=c("2011-w27", "2016-w26"))
scalebar <- layout.scalebar(noroBEr@map,
  corner = c(0.7, 0.9), scale = 10,
  labels = c(0, "10 km"), cex = 0.6,
  height = 0.02)
plot(noroBEr, type = observed ~ unit,
  sub = "Mean yearly incidence",
  population = 100000 / (
    sum(pop2011)*(nrow(noroBEr)/52)
  ), labels = list(cex = 0.8),
  sp.layout = scalebar)
```



```
noroBERbyg <- noroBE(by = "all", timeRange = c("2011-w27", "2016-w26"))
```



```
animation::saveHTML(animate(noroBERbyg[["00-04"]], tps = 1:52,  
timeplot = list(as.Date = TRUE)))
```

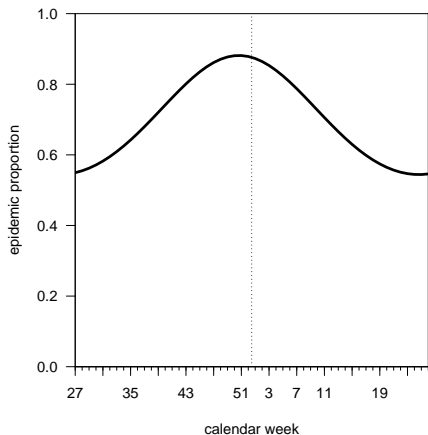
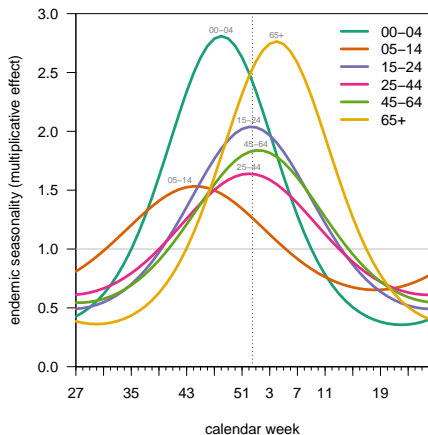
Model formulation for the norovirus data

$$\mu_{grt} = e_{gr} \exp \left\{ \alpha_g^{(v)} + \alpha_r^{(v)} + \beta x_t + \gamma_g^{(v)} \sin(\omega t) + \delta_g^{(v)} \cos(\omega t) \right\} \\ + \exp \left\{ \alpha_g^{(\phi)} + \alpha_r^{(\phi)} + \tau \log(e_{gr}) + \gamma^{(\phi)} \sin(\omega t) + \delta^{(\phi)} \cos(\omega t) \right\} \\ \sum_{g',r'} [(C^K)_{g'g} (o_{r'r} + 1)^{-\rho}] Y_{g',r',t-1}$$

- Group- and district-specific effects $\alpha_g^{(\cdot)}$ and $\alpha_r^{(\cdot)}$
- Christmas break indicator $x_t \rightarrow$ reduced reporting
- Group-specific endemic seasonality (sinusoidal log-rates, $\omega = 2\pi/52$)
- “Gravity model” $e_{gr}^\tau \rightarrow$ force of infection scales with population size
- C^K : power-adjusted contact matrix
- Power-law weights $w_{r'r} = (o_{r'r} + 1)^{-\rho}$

+ group-specific overdispersion parameters

Estimated seasonality



Estimated power-law weights

