University of Zurich[UZH]

ISPM, Division of Biostatistics

Modelling power-law spread of infectious diseases

Sebastian Meyer and Leonhard Held

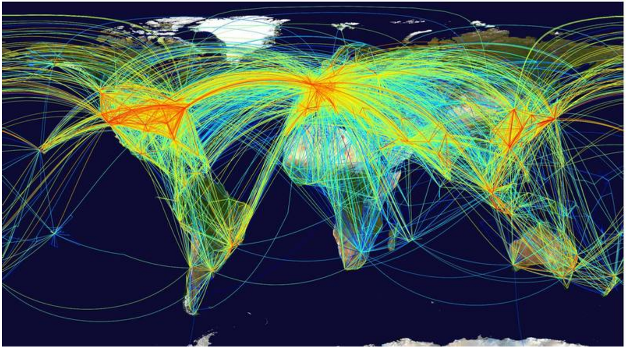
Financially supported by the Swiss National Science Foundation
(project 137919: *Statistical methods for spatio-temporal modelling and prediction of infectious diseases*)

Epidemic Modelling

- Prospective surveillance: outbreak detection (Farrington).
- This talk is concerned with retrospective surveillance:
  - Explain the spread of epidemics through statistical modelling
  - Assess influential factors, e.g., seasonality, climate, concurrent epidemics of related pathogens, contact networks

- Data basis: routine public health surveillance including temporal *as well as spatial* information
- This talk deals with two types of surveillance data:
  - individual case reports
  - aggregated counts by week and administrative district

Mobility networks determine the spread of epidemics



Source: Max Planck Institute for Dynamics and Self-Organization
(http://www.mpg.de/4406928/)

How to quantify spatial interaction between regions or individuals
in the absence of network data?

Power law! Why?

Brockmann et al., 2006:

- Analysed trajectories of 464 670 dollar bills in the USA
- Short-time travel behaviour follows a power law wrt distance

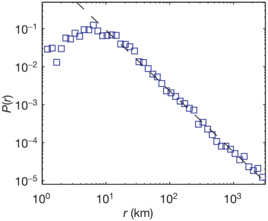


Fig. 1c: Histogram of the distance r traversed within 4 days.
Dashed line: $P(r) \propto r^{-1.59}$

- "Starting point for the development of a new class of models for the spread of infectious diseases"

Power law! Why?

Brockmann et al., 2006:

- Analysed trajectories of 464 670 dollar bills in the USA
- Short-time travel behaviour follows a power law wrt distance

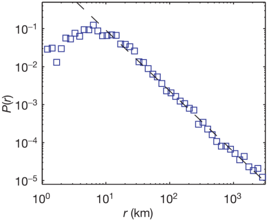


Fig. 1c: Histogram of the distance r traversed within 4 days.
Dashed line: $P(r) \propto r^{-1.59}$

- “Starting point for the development of a new class of models for the spread of infectious diseases”

Let’s do it! We use this finding to improve upon two previously established model frameworks for infectious disease spread.

Two additive components (Held et al., 2005)

$\oplus$ Endemic: seasonality, population, socio-demography, climate, ...
Epidemic: dependency on previously infected individuals

Space-time point process model for individual case reports

$$\lambda^*(t, \mathbf{s}) = \nu_{[t][\mathbf{s}]}\, \rho_{[t][\mathbf{s}]} + \sum_{j:t_j<t} \eta_j \cdot g(t - t_j) \cdot f(\|\mathbf{s} - \mathbf{s}_j\|)$$

$$\log(\nu_{[t][\mathbf{s}]}) = \beta_0 + \boldsymbol{\beta}^\top \mathbf{z}_{[t][\mathbf{s}]}\,, \qquad \log(\eta_j) = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{m}_j$$

(Meyer, Elias, and Höhle, 2012)

Multivariate time-series model for counts

$$Y_{it} | \mathbf{Y}_{\cdot,t-1} \sim \text{NegBin}(\mu_{it}, \psi)$$

$$\mu_{it} = \nu_{it}\, e_{it} + \lambda_{it}\, Y_{i,t-1} + \phi_{it} \sum_{j \neq i} w_{ji}\, Y_{j,t-1}$$

$$\log(\cdot_{it}) = \beta_0^{(\cdot)} + b_i^{(\cdot)} + \boldsymbol{\beta}^{(\cdot)\top} \mathbf{z}_{it}^{(\cdot)} \qquad \cdot \in \{\nu, \lambda, \phi\}$$

(Held and Paul, 2012, and previous work)

Two additive components (Held et al., 2005)

$\oplus$ Endemic: seasonality, population, socio-demography, climate, ...
Epidemic: dependency on previously infected individuals

Space-time point process model for individual case reports

$$\lambda^*(t,\mathbf{s}) = \nu_{[t][\mathbf{s}]}\,\rho_{[t][\mathbf{s}]} + \sum_{j:t_j<t} \eta_j \cdot g(t-t_j) \cdot \boxed{f(\|\mathbf{s}-\mathbf{s}_j\|)}$$

$$\log(\nu_{[t][\mathbf{s}]}) = \beta_0 + \boldsymbol{\beta}^\top \mathbf{z}_{[t][\mathbf{s}]}\,, \qquad \log(\eta_j) = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{m}_j$$

(Meyer, Elias, and Höhle, 2012)

Multivariate time-series model for counts

$$Y_{it}|\mathbf{Y}_{\cdot,t-1} \sim \text{NegBin}(\mu_{it},\psi)$$

$$\mu_{it} = \nu_{it}\,e_{it} + \lambda_{it}\,Y_{i,t-1} + \phi_{it}\sum_{j\neq i}\boxed{w_{ji}}\,Y_{j,t-1}$$

$$\log(\cdot_{it}) = \beta_0^{(\cdot)} + b_i^{(\cdot)} + \boldsymbol{\beta}^{(\cdot)\top}\mathbf{z}_{it}^{(\cdot)} \qquad \cdot \in \{\nu,\lambda,\phi\}$$

(Held and Paul, 2012, and previous work)

Two additive components (Held et al., 2005)

$\oplus$ Endemic: seasonality, population, socio-demography, climate, ...
Epidemic: dependency on previously infected individuals

Space-time point process model for individual case reports

$$\lambda^*(t, \mathbf{s}) = \nu_{[t][\mathbf{s}]}\, \rho_{[t][\mathbf{s}]} + \sum_{j: t_j < t} \eta_j \cdot g(t - t_j) \cdot \boxed{f(\|\mathbf{s} - \mathbf{s}_j\|)}$$

"twinstim"

$$\log(\nu_{[t][\mathbf{s}]}) = \beta_0 + \boldsymbol{\beta}^\top \mathbf{z}_{[t][\mathbf{s}]}\,, \qquad \log(\eta_j) = \gamma_0 + \boldsymbol{\gamma}^\top \mathbf{m}_j$$

Multivariate time-series model for counts

$$Y_{it} | \mathbf{Y}_{\cdot, t-1} \sim \text{NegBin}(\mu_{it}, \psi)$$

$$\mu_{it} = \nu_{it}\, e_{it} + \lambda_{it}\, Y_{i,t-1} + \phi_{it} \sum_{j \neq i} \boxed{w_{ji}}\, Y_{j,t-1}$$

"hhh4"

$$\log(\cdot_{it}) = \beta_0^{(\cdot)} + b_i^{(\cdot)} + \boldsymbol{\beta}^{(\cdot)\top} \mathbf{z}_{it}^{(\cdot)} \qquad \cdot \in \{\nu, \lambda, \phi\}$$

Power-law distance decay in `twinstim`

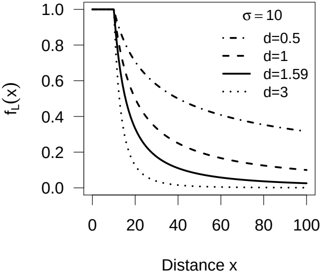
$f(x) = x^{-d}$ not suitable: pole at $x = 0 \Rightarrow$ not integrable.

Power-law distance decay in `twinstim`

$f(x) = x^{-d}$ not suitable: pole at $x = 0 \Rightarrow$ not integrable.

"Lagged" power law with uniform short-range dispersal:

$$f_L(x) = \begin{cases} 1 & \text{for } x < \sigma, \\ \left(\frac{x}{\sigma}\right)^{-d} & \text{otherwise.} \end{cases}$$

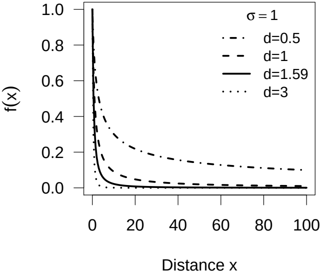


Power-law distance decay in `twinstim`

$f(x) = x^{-d}$ not suitable: pole at $x = 0 \Rightarrow$ not integrable.

Kernel of the density of the shifted Pareto distribution:

$$f(x) = (x + \sigma)^{-d}$$

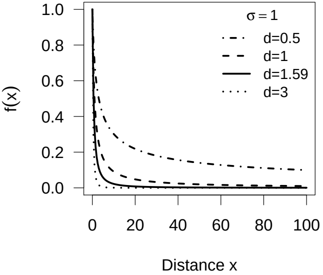


Power-law distance decay in `twinstim`

$f(x) = x^{-d}$ not suitable: pole at $x = 0 \Rightarrow$ not integrable.

Kernel of the density of the shifted Pareto distribution:

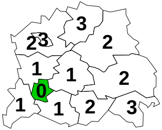
$$f(x) = (x + \sigma)^{-d}$$



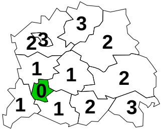
- Joint ML-inference for all model parameters
- Numerical cubature of $f_{2D}(\mathbf{s}) = f(\|\mathbf{s}\|)$ over polygonal domains in likelihood via product-Gauss cubature (Sommariva and Vianello, 2007)

Power-law weights in `hhh4`

- On which distance measure between regions should the power law act?
  $\longrightarrow$ Order of neighbourhood o_{ji}!



Power-law weights in `hhh4`

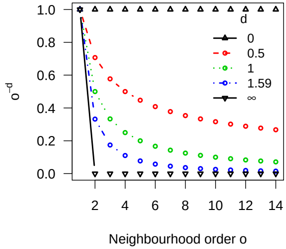


- On which distance measure between regions should the power law act?
  $\longrightarrow$ Order of neighbourhood o_{ji}!

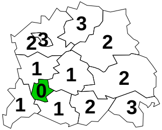
- Generalisation of previously used first-order weights w_{ji}:

first-order	power law
$\mathbb{1}(j \sim i)$	o_{ji}^{-d}

- Normalisation: $w_{ji} / \sum_k w_{jk}$



Power-law weights in hhh4



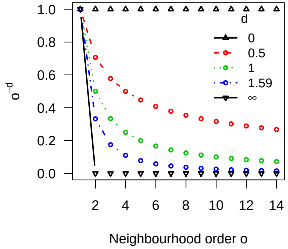
– On which distance measure between regions should the power law act?
$\longrightarrow$ Order of neighbourhood o_{ji}!

– Generalisation of previously used first-order weights w_{ji}:

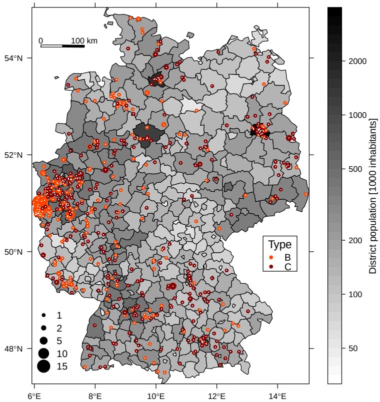
first-order	power law
$\mathbb{1}(j \sim i)$	o_{ji}^{-d}

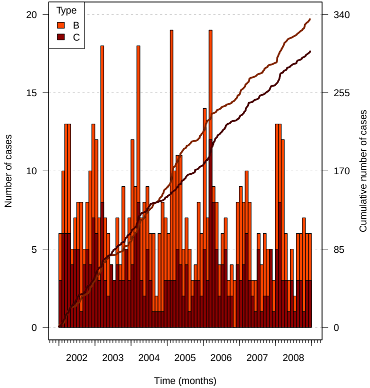
– Normalisation: $w_{ji} / \sum_k w_{jk}$

– Estimate d within the penalised likelihood framework simultaneously with all other model parameters.

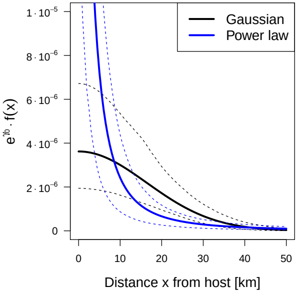


Example of individual-level surveillance data: Invasive meningococcal disease in Germany (2002–8)





▷ Estimated power law



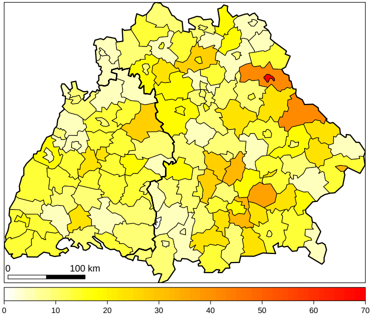
Endemic: seasonality, trend, population density as offset

Epidemic: type, age group

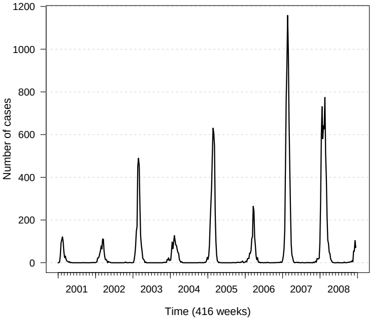
Decay parameter: $\hat{d} = 2.3$ (95% CI: $[1.74, 3.03]$)

	AIC	$\hat{R}(B)$	$\hat{R}(C)$
Gaussian	18972.04	0.22 [0.17,0.31]	0.10 [0.06,0.15]
Power law	18944.25	0.26 [0.14,0.35]	0.13 [0.06,0.19]

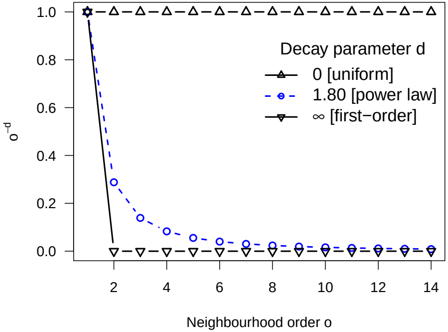
Example of aggregated surveillance data: Influenza in Southern Germany (2001–8)



Mean yearly incidence per 100 000 inhabitants
in the 140 districts of Baden-Württemberg and Bavaria



▷ Estimated power law



– population fractions as endemic offset

– seasonality, region-specific random intercepts in all three components

▷ Predictive performance

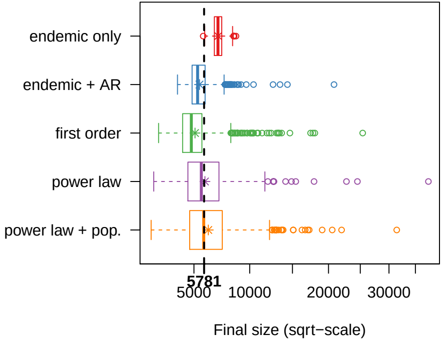
- Use strictly proper scoring rules to evaluate consistency of predictive distribution with later observed value: logarithmic score (logS) and ranked probability score (RPS) (Czado et al., 2009)
- Based on one-week-ahead predictions in the last two years
- Calculate mean scores and p-values via permutation tests

	logS	RPS
first order	0.5511	0.4194
power law	0.5448	0.4168
p-value	0.0001	0.19

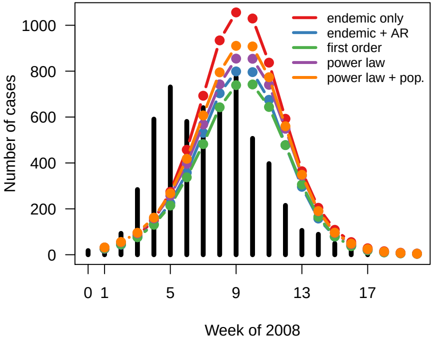
▷ Long-term predictive performance

- Simulate the 2008 wave of influenza
- Based on models fitted on 2001–2007
- Initialised by the 18 cases of the last week of 2007
- Run 1000 simulations for each model and evaluate by
  - the final size distribution
  - proper scoring rules on the empirical distribution of the simulations compared to the later reported counts
- Additional benchmark against
  - endemic-only model
  - model without neighbourhood effects
  - model with additional population effect in spatio-temporal component

▷ Long-term predictive performance
└ final size

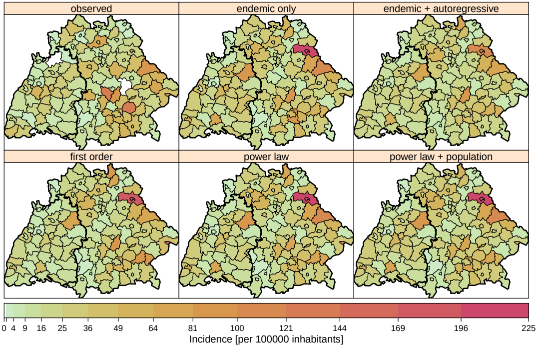


▷ Long-term predictive performance
└ time domain



DSS	RPS
27.03	149.77
31.36	112.15
26.46	**108.61**
16.41	110.2
15.49	111.86

▷ Long-term predictive performance
└ space domain



	DSS	RPS
endemic only	7.85	15.39
endemic + AR	7.59	15.04
first order	7.51	15.63
power law	7.36	14.75
power law + pop.	**7.24**	**14.3**

▷ Long-term predictive performance
└ space-time domain

	DSS	RPS
endemic only	2.91	1.31
endemic + AR	2.58	1.26
first order	2.5	1.26
power law	2.29	1.25
power law + pop.	**2.29**	**1.24**

[animation]

Discussion

- Human mobility
  - is an important driver of epidemic spread
  - follows a power law with respect to distance
- Predictive performance improves when using a power law for spatial interaction of cases
- Infectious imports increase with population size (Bartlett, 1957)
- Edge effects:
  - random intercepts account for unobserved heterogeneity
  - incorporate region-specific incoming traffic from abroad

Outlook

- Semiparametric estimate of weight function to confirm power law
- Estimate impact of traffic data on neighbourhood weights w_{ji} (Geilhufe et al., 2013)

Outlook

- Semiparametric estimate of weight function to confirm power law
- Estimate impact of traffic data on neighbourhood weights w_{ji} (Geilhufe et al., 2013)

- Further reading: arXiv:1308.5115
- Further application: all methods are implemented in the open-source R package `surveillance` for visualisation, modelling and monitoring of epidemic phenomena.

References

- Bartlett, M. S. (1957). Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 120(1):48–70.
- Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65(4):1254–1261.
- Geilhufe, M., Held, L., Skrøvseth, S. O., Simonsen, G. S., and Godtliebsen, F. (2013). Power law approximations of movement network data for modeling infectious disease spread. *Biometrical Journal*. In press.
- Held, L., Höhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5:187–199.
- Held, L. and Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54(6):824–843.
- Höhle, M., Meyer, S., and Paul, M. (2013). *surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena*.
- Meyer, S., Elias, J., and Höhle, M. (2012). A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68(2):607–616.
- Meyer, S. and Held, L. (2013). Modelling power-law spread of infectious diseases. Submitted to Annals of Applied Statistics.
- Sommariva, A. and Vianello, M. (2007). Product Gauss cubature over polygons based on Green's integration formula. *Bit Numerical Mathematics*, 47(2):441–453.