University of Zurich^UZH

Epidemiology, Biostatistics and Prevention Institute

Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package `surveillance`

Sebastian Meyer

```
> sessionInfo()
[1] ”June 30 – July 3, 2015”
[2] ”Aalborg, Denmark”
```

Epidemic phenomena

Examples:

- Earth quakes
- Riots / crimes
- Infectious diseases

Data: Surveillance systems routinely collect

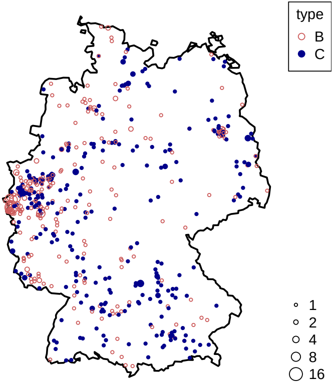
- time-stamped
- geo-referenced

case reports

Case study I: Invasive meningococcal disease

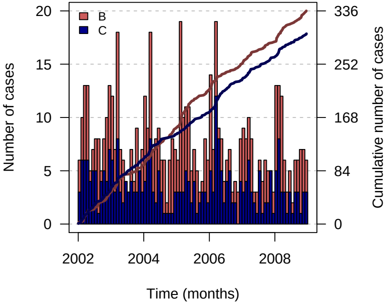
`library("surveillance"); data("imdepi")`

```
plot(imdepi, "space")
```



Dot size proportional to the number of cases (residence postcode)

```
plot(imdepi, "time")
```



Monthly and cumulative number of cases (by date of specimen sampling)

```
animate(subset(imdepi, type=="B"),
        time.spacing = 7)
```

```
animate(subset(imdepi, type=="C"),
        time.spacing = 7)
```

Does the force of infection depend on the bacterial finetype?

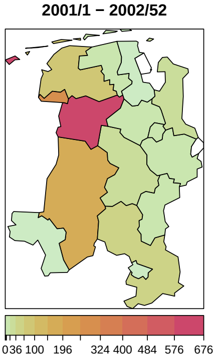
Case study II: Measles

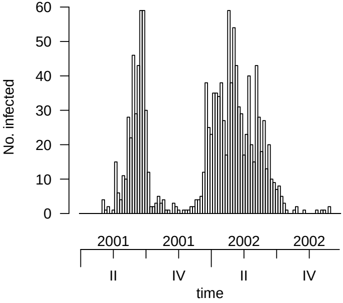
```
library("surveillance"); data("measlesWeserEms")
```

Publically available surveillance data:
time series of counts of newly reported infections by district

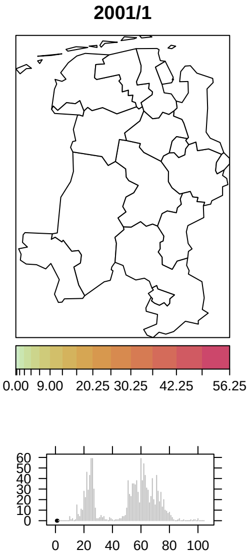
```
plot(measlesWeserEms,
     type = observed ~ unit)
```

```
plot(measlesWeserEms,
     type = observed ~ time)
```





```r
animate(measlesWeserEms)
```



Is local vaccination coverage related to disease dynamics?

Characteristics of epidemic-type data

- Low number of cases
- Seasonality
- Occassional outbreaks (“self-exciting” process)
- Dependence between areas, age groups, etc.
- Underreporting, reporting delays

Characteristics of epidemic-type data

- Low number of cases
- Seasonality
- Occassional outbreaks (“self-exciting” process)
- Dependence between areas, age groups, etc.
- Underreporting, reporting delays

Aims of `surveillance`

Monitoring (prospective): Outbreak prediction and detection (→ “Zombie Preparedness” talk by Michael Höhle)

Modelling (retrospective): Quantify epidemicity and effects of external covariates on disease dynamics

Place in the world of R packages

`surveillance` is the first and only software package dedicated to the space-time modelling and monitoring of epidemic phenomena

Related packages:

`spacetime`:	Basic classes and methods for spatio-temporal data
`spatstat`:	THE package for purely spatial point patterns
`tscount`, `EpiEstim`, `outbreaker`, `amei`:	Several packages dealing with purely temporal epidemic data
`stpp`:	Simulation & visualization of space-time point patterns

For a more complete picture:

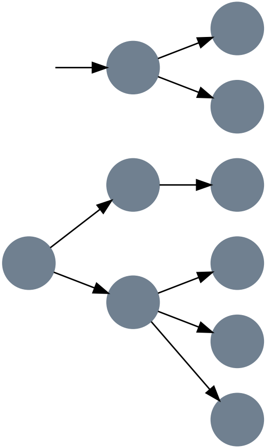
→ CRAN task view “Handling and Analyzing Spatio-Temporal Data”

Three modelling frameworks in `surveillance`

Data Resolution	Example	Model	Function
individual events in continuous space-time	cases of invasive meningococcal disease (IMD) Meyer et al., 2012	spatio-temporal point process	`twinstim()`
event counts aggregated in space & time	week×district counts of measles Meyer et al., 2014	multivariate NegBin time series	`hhh4()`
individual SIR event history of a fixed population	spread of classical swine fever among domestic pig farms Höhle, 2009	multivariate temporal point process	`twinSIR()`

Basic modelling concept

Stochastic branching process with immigration



- Decomposed disease risk:

  Endemic: seasonality, population, socio-demography, . . .

  $\oplus$

  Epidemic: force of previously infected individuals
- Ebola: R_0 of about 1.5 – 2.5
- Force of infection may depend on age and spatial/temporal distance to infective

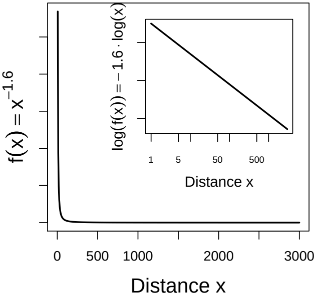
Spatial interaction

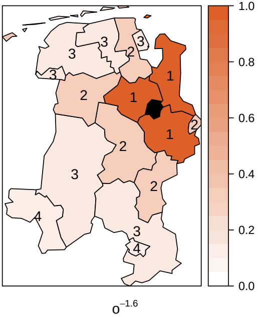
Tobler's First Law of Geography:

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

Brockmann et al., 2006 (dollar bill tracking):

> *The distribution of travelling distances*
> *decays as a power law.*





Case study I: Invasive meningococcal disease

Regression framework for the conditional intensity function

$$\lambda(\boldsymbol{s}, t) = \rho_{[\boldsymbol{s}][t]} \, \nu_{[\boldsymbol{s}][t]}$$

Endemic component

- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor $\nu_{[\boldsymbol{s}][t]}$
- Equivalent to Poisson-GLM for aggregated counts

Case study I: Invasive meningococcal disease

Regression framework for the conditional intensity function

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

Endemic component

- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor $\nu_{[\boldsymbol{s}][t]}$
- Equivalent to Poisson-GLM for aggregated counts

Force of infection

- Depends on event-specific characteristics $\boldsymbol{m}_j$ via

$$\log(\eta_j) = \gamma_0 + \boldsymbol{\gamma}^\top \boldsymbol{m}_j$$

- Decays over space/time according to parametric interaction function $f(\cdot)/g(\cdot)$

Case study I: Invasive meningococcal disease

Regression framework for the conditional intensity function

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

Endemic component

- Piecewise constant on a suitable space-time grid
- Explanatory variables in a log-linear predictor $\nu_{[\boldsymbol{s}][t]}$
- Equivalent to Poisson-GLM for aggregated counts

Force of infection

- Depends on event-specific characteristics $\boldsymbol{m}_j$ via

$$\log(\eta_j) = \gamma_0 + \boldsymbol{\gamma}^\top \boldsymbol{m}_j$$

- Decays over space/time according to parametric interaction function $f(\cdot)/g(\cdot)$

Likelihood inference

- `nlminb()` with analytical score function *and* Fisher info
- R package `polyCub` for cubature of $f(\|\boldsymbol{s}\|)$ over polygons

Case study I: Invasive meningococcal disease

Model estimation

```
imdfit <- twinstim(
    endemic = ~offset(log(popdensity)) + I(start/365 - 3.5) +
        sin(2 * pi * start/365) + cos(2 * pi * start/365),
    epidemic = ~type + agegrp,
    siaf = siaf.powerlaw(), tiaf = tiaf.constant(),
    data = imdepi, subset = !is.na(agegrp),
    start = c("e.(Intercept)"=-6.5, "e.siaf.1"=1.5, "e.siaf.2"=0.9),
    optim.args = list(fixed = "e.siaf.1"), model = TRUE, cores = 4)
```

Case study I: Invasive meningococcal disease

Model estimation

```
imdfit <- twinstim(
    endemic = ~offset(log(popdensity)) + I(start/365 - 3.5) +
        sin(2 * pi * start/365) + cos(2 * pi * start/365),
    epidemic = ~type + agegrp,
    siaf = siaf.powerlaw(), tiaf = tiaf.constant(),
    data = imdepi, subset = !is.na(agegrp),
    start = c("e.(Intercept)"=-6.5, "e.siaf.1"=1.5, "e.siaf.2"=0.9),
    optim.args = list(fixed = "e.siaf.1"), model = TRUE, cores = 4)
```

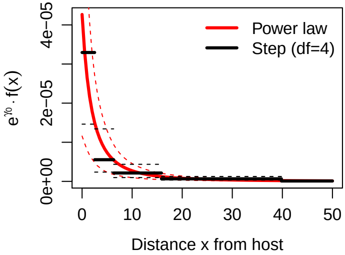
```
xtable(imdfit)
```

	RR	95% CI	p-value
h.I(start/365 - 3.5)	0.959	0.92–1.00	0.071
h.sin(2 * pi * start/365)	1.231	1.08–1.41	0.0022
h.cos(2 * pi * start/365)	1.379	1.21–1.57	<0.0001
e.typeC	0.450	0.27–0.74	0.0017
e.agegrp[3,19)	2.133	1.10–4.12	0.024
e.agegrp[19,Inf)	0.824	0.33–2.05	0.68

Case study I: Invasive meningococcal disease

Estimated spatial interaction

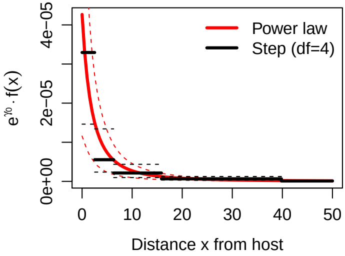
```
plot(imdfit, which = "siaf",
     xlim = c(0, 50))

imdfit_fstep <- update(imdfit,
    siaf = siaf.step(
        knots = exp((1:4)*log(100)/5),
        maxRange = 100),
    optim.args = list(fixed = NULL))
plot(imdfit_fstep, which = "siaf",
     add = TRUE, col.estimate = 1)
```



Case study I: Invasive meningococcal disease

Estimated spatial interaction

```
plot(imdfit, which = "siaf",
     xlim = c(0, 50))

imdfit_fstep <- update(imdfit,
    siaf = siaf.step(
        knots = exp((1:4)*log(100)/5),
        maxRange = 100),
    optim.args = list(fixed = NULL))
plot(imdfit_fstep, which = "siaf",
     add = TRUE, col.estimate = 1)
```



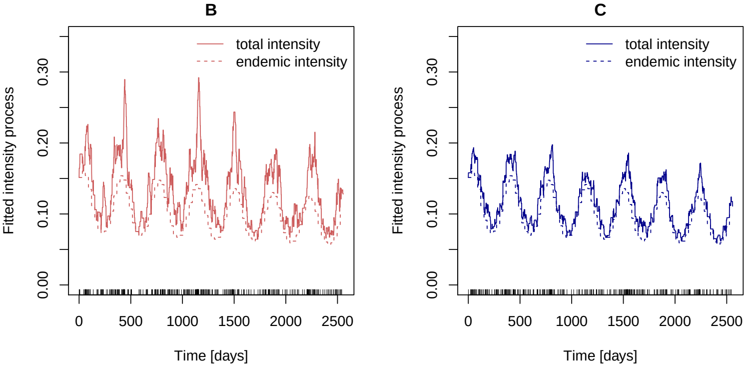
Predefined interaction functions:

Spatial (`siaf.*`)	Temporal (`tiaf.*`)
`constant`	`constant`
`gaussian`	`exponential`
`powerlaw`	`step`
`powerlawL`	
`step`	
`student`	

Case study I: Invasive meningococcal disease

Fitted ground intensity $\int \hat{\lambda}(\boldsymbol{s}, t)\,\mathrm{d}\boldsymbol{s}$

```
plot(imdfit, which = "total intensity", aggregate = "time",
     types = 1, ylim = c(0,0.3), tgrid = 2500)
```



Case study I: Invasive meningococcal disease

Methods for `"twinstim"`

Display	Extract	Modify	Other
`print`	`nobs`	`update`	`simulate`
`summary`	`vcov`	`add1`	`coeflist`
`xtable`	`logLik`	`drop1`	
`plot`	`extractAIC`	`stepComponent`	
`intensityplot`	`profile`		
`iafplot`	`residuals`		
`checkResidualProcess`	`terms`		
	`R0`		

Case study II: Measles (areal count time series)

Regression framework

Number of cases in region r at time t

$$Y_{rt} \mid \boldsymbol{Y}_{\cdot,t-1} \sim \text{NegBin}(\mu_{rt}, \psi)$$

Endemic-Epidemic decomposition of disease risk:

$$\mu_{rt} = e_{rt}\,\nu_{rt} + \lambda_{rt}\,Y_{r,t-1} + \phi_{rt} \sum_{s \neq r} w_{sr}\;Y_{s,t-1}$$

e_{rt}	population offset
$\nu_{rt}, \lambda_{rt}, \phi_{rt}$	log-linear predictors, e.g., vaccination coverage
w_{sr}	weight for s to r transmission, e.g., $w_{sr} = o_{sr}^{-d}$

Case study II: Measles (areal count time series)

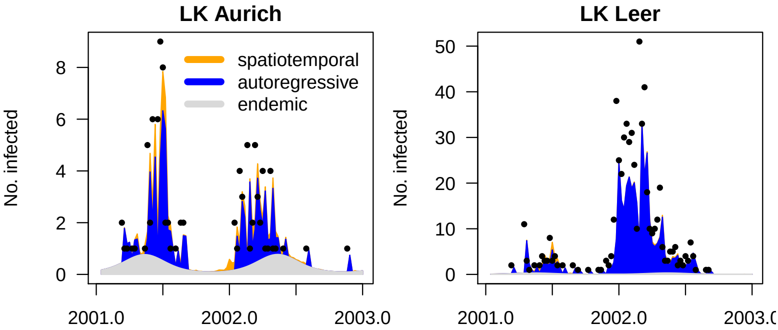
Model estimation

```
(endemic <- addSeason2formula(~log(pSusceptible) + t))

## ~log(pSusceptible) + t + sin(2 * pi * t/52) + cos(2 * pi * t/52)

measlesModel <- list(
    end = list(f = endemic, offset = population(measlesWeserEms)),
    ar = list(f = ~1),
    ne = list(f = ~1, weights = W_powerlaw(maxlag = 5)),
    family = "NegBin1", data = list(pSusceptible = 1 - pVacc))

measlesFit <- hhh4(measlesWeserEms, control = measlesModel)
```

Case study II: Measles (areal count time series)

Fitted mean components

```
plot(measlesFit, type = "fitted", units = c(7,12), hide0s = TRUE)
```



Case study II: Measles (areal count time series)

Association with vaccination coverage v_r

Endemic incidence is proportional to $(1 - v_r)^{\beta_v}$:

```
summary(measlesFit)$fixef["end.log(pSusceptible)", ]
```

```
##   Estimate Std. Error
##      2.054      0.379
```

Case study II: Measles (areal count time series)

Association with vaccination coverage v_r

Endemic incidence is proportional to $(1 - v_r)^{\beta_v}$:

```
summary(measlesFit)$fixef["end.log(pSusceptible)", ]

##   Estimate Std. Error
##      2.054      0.379
```

Other methods:

`update()`, `simulate()`, `oneStepAhead()`, ...

Conclusion

`surveillance` offers a comprehensive framework for the spatio-temporal analysis of epidemic phenomena, including visualisation, modelling, inference and simulation of:

- (multivariate) surveillance time series
- spatio-temporal point patterns
- geo-referenced SIR event histories

Key references:

- http://surveillance.r-forge.r-project.org/
- arXiv:1411.0416 (Meyer et al., 2014): a guide to the three presented endemic-epidemic model classes

Acknowledgments

Joint work with:

- Leonhard Held (University of Zurich)
- Michael Höhle (University of Stockholm)

Funding:

- Munich Center of Health Sciences (2007–2010)
- Swiss National Science Foundation (2012–2015)

References

- Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The Scaling Laws of Human Travel. *Nature*, 439(7075):462–465.
- Höhle, M. (2009). Additive-Multiplicative Regression Models for Spatio-Temporal Epidemics. *Biometrical Journal*, 51(6):961–978.
- 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. (2014). Power-Law Models for Infectious Disease Spread. *The Annals of Applied Statistics*, 8(3):1612–1639.
- Meyer, S., Held, L., and Höhle, M. (2014). Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package `surveillance`. arxiv:1411.0416.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46:234–240.

Feedback?

☞ sebastian.meyer@uzh.ch