



Modelling and monitoring of epidemic phenomena using the surveillance package in R

Sebastian Meyer Wei Wei

Part of the slides are based on presentations given by:

Michaela Paul (ISPM, University of Zurich)

Michael Höhle (Robert Koch Institute, Berlin, Germany)

Financial support by the Swiss National Science Foundation (project 137919: *Statistical methods for spatio-temporal modelling and prediction of infectious diseases*, supervised by Leonhard Held) is gratefully acknowledged.



Outline

Overview of the package

Prospective monitoring illustrated by the Farrington algorithm

Examples of model-based surveillance

Campylobacteriosis in humans and chicken

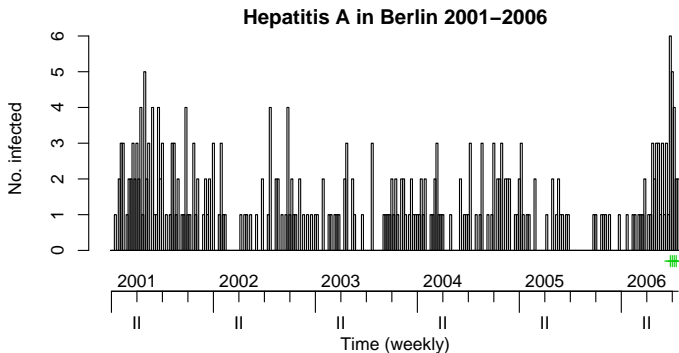
Two agents of invasive meningococcal disease

Conclusion



Example of surveillance data

```
R> library("surveillance")  
R> data("ha.sts")  
R> plot(ha.sts, type = observed ~ time, legend.opts = NULL,  
+       main = "Hepatitis A in Berlin 2001-2006", xlab = "Time (weekly)")
```







Who does surveillance? (AFAIK)

- Public Health Department of Lower Saxony ([Hulth et al., 2010](#))
- Swedish Institute for Communicable Disease Control:
Computer Assisted Search for Epidemics (CASE) project
- National Public Health Institute of Finland
- French National Reference Centre for Salmonella (Institut Pasteur)
- Infectious Disease Surveillance and Analysis System, Dept. of
Animal Production and Health, Sri Lanka ([Robertson et al., 2010](#))
- In the pipeline:
 - German Federal Institute for Public Health:
specialised analyses, e.g., “back-projection” and “now-casting” for
the 2011 EHEC outbreak ([Robert Koch Institute, 2011](#))
 - Austrian Agency for Health and Food Safety (AGES)
 - Switzerland (BAG? BVET?)



What is surveillance?

 is a free software environment for statistical computing and graphics (www.R-project.org).

surveillance is an open source  package for the **visualisation**, **modelling** and **monitoring** of routinely collected public health surveillance data.

History: Development started 2004 at the University of Munich as part of the DFG/SFB386 research project “Statistical methodology for infectious disease surveillance”

Motivation:

- Data structures and implementational framework for methodological developments
- Disease monitoring tool for epidemiologists and public health authorities

Availability: R-Forge (devel), CRAN

surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena

A package implementing statistical methods for the modeling and change-point detection in time series of counts, proportions and categorical data, as well as for the modeling of continuous-time epidemic phenomena, e.g. discrete-space setups such as the spatially enriched Susceptible-Exposed-Infectious-Recovered (SEIR) models for surveillance data, or continuous-space point process data such as the occurrence of disease or earthquakes. Main focus is on outbreak detection in count data time series originating from public health surveillance of infectious diseases, but applications could just as well originate from environmetrics, reliability engineering, econometrics or social sciences. Currently the package contains implementations of typical outbreak detection procedures such as Stroup et. al (1989), Farrington et al (1996), Rossi et al (1999), Rogerson and Yamada (2001), a Bayesian approach, negative binomial CUSUM methods and a detector based on generalized likelihood ratios. Furthermore, inference methods for the retrospective infectious disease model in Held et al (2005), Held et al (2006), Paul et al (2008) and Paul and Held (2011) are provided. A novel CUSUM approach combining logistic and multinomial logistic modelling is also included. Continuous self-exciting spatio-temporal point processes are modeled through additive-multiplicative conditional intensities as described in Höhle (2009) ("twinSIR", discrete space) and Meyer et al (2012) ("twinstim", continuous space). The package contains several real-world datasets, the ability to simulate outbreak data, visualize the results of the monitoring in temporal, spatial or spatio-temporal fashion.

Version: 1.5-1
Depends: R ($\geq 2.14.0$), grDevices, graphics, stats, utils, [xtable](#), methods, [Rcpp](#), [sp](#)
Imports: [MASS](#), [Matrix](#), [spatstat](#) ($\geq 1.17-0$)
LinkingTo: [Rcpp](#)
Suggests: [msm](#), [spc](#), [colorspace](#), [animation](#), [quadprog](#), [mvtnorm](#), [statmod](#), [memoise](#), [gpllib](#), [maptools](#), [intervals](#), [spdep](#), [numDeriv](#), [maxLik](#), [RUnit](#), [digest](#), [coda](#), [splancs](#), [gamlss](#)
Published: 2012-12-14
Author: M. Höhle, S. Meyer and M. Paul with contributions from T. Correa, L. Held, M. Hofmann, C. Lang, A. Riebler, D. Sabanés Bové, S. Steiner, M. Virtanen, and V. Wimmer. A few code segments are modified versions of code by The R Core Team (available under GPL-2).
Maintainer: Michael Höhle <Michael.Hoehle at stat.uni-muenchen.de>
License: [GPL-2](#)
URL: <http://surveillance.r-forge.r-project.org/>



Which types of surveillance? (among others)

- Prospective monitoring for univariate count data time series
 - Based on reference values:
 - `farrington` (Farrington et al., 1996)
 - `bayes` (Riebler, 2004)
 - Inspired from statistical process control:
 - `cusum` (Rossi et al., 1999 and extensions)
 - `glrnb` (Höhle and Paul, 2008)
- Model-based surveillance
 - Count data time series models:
 - `twins` (Held et al., 2006b)
 - `hhh4` (Paul and Held, 2011)
 - Spatio-temporal point process modelling:
 - `twinSIR` (Höhle, 2009) – discrete space
 - `twinstim` (Meyer et al., 2012) – continuous space



Prospective monitoring: based on reference values

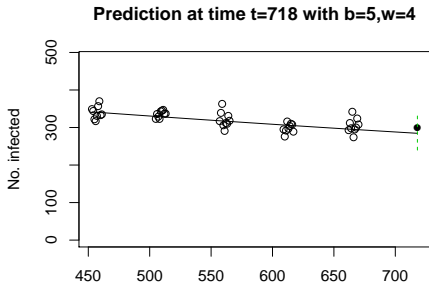
- Choose set of reference values R_t as basis for raising an alarm to deal with seasonality. Example ($w=4$ weeks, $b=3$ years):

	current week															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	...	52
current year: $b = 0$		X	X	X	X	•										
1		X	X	X	X	X	X	X	X	X						
2			X	X	X	X	X	X	X	X						
3				X	X	X	X	X	X	X						
4																



Prospective monitoring: the Farrington algorithm

- Predict number of cases at current time t based on R_t using an overdispersed Poisson GLM with intercept and trend
- If the current observation is larger than the upper limit of the $(1 - \alpha) \cdot 100\%$ prediction interval, then an alarm is generated.

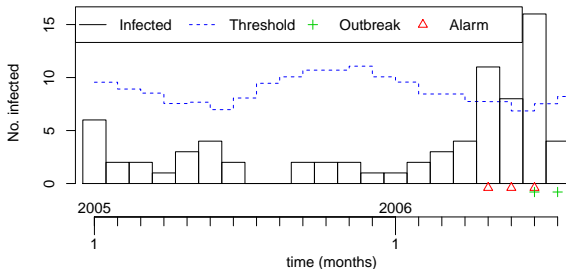




Hepatitis A example (revisited)

```
R> ha1 <- aggregate(ha.sts, by="unit")      # -> univariate time series
R> ha41 <- aggregate(ha1, nfreq=13)       # -> 4-week aggregation
R> cntrlFar <- list(range=53:73, w=2, b=3, alpha=0.01, limit54=c(0,4))
R> survha <- farrington(ha41, control=cntrlFar)
```

Surveillance using farrington(2,0,3)





Example of count data time-series modelling:
Using surveillance::hhh4 to analyse
campylobacteriosis prevalence
in humans and chicken



Campylobacteriosis

- Symptoms: nausea, diarrhoea, fever and abdominal cramps.
- Transmission: raw/undercooked chicken (main), unpasteurized milk or contaminated water, hardly spread from person-to-person.

Two kinds of analysis using hhh4

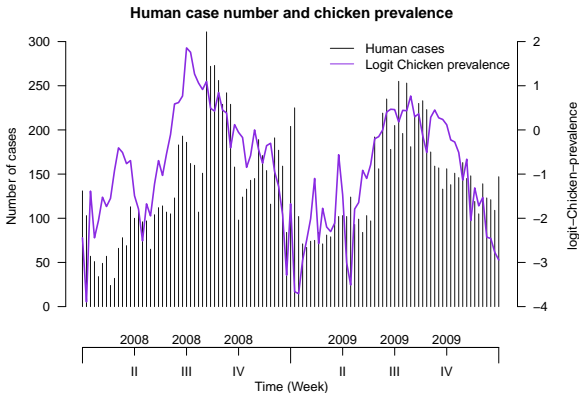
- Retrospective analysis: with seasonality and time trend
- Prospective analysis: with seasonality and time trend

Data

- **Prevalence in Human and Chicken:** Weekly number of cases on human, 2008–2009, from BAG (Federal Office of Public Health) ([Lutz et al., 2010](#)), and weekly prevalence of diseased chicken from a large Swiss chicken slaughterhouse.
- **Campylobacter in Germany:** Weekly number of cases in Germany from 2001–2012 from Robert Koch-Institut (RKI).

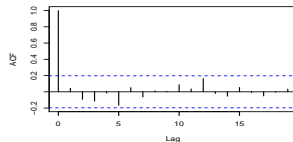
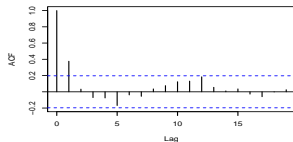
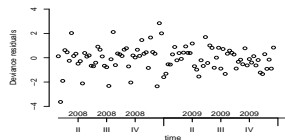
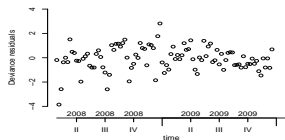
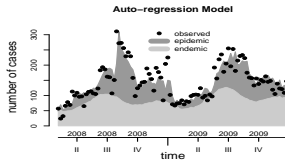
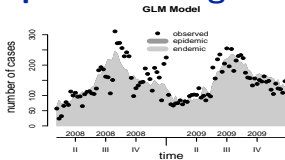


Human case number and chicken prevalence



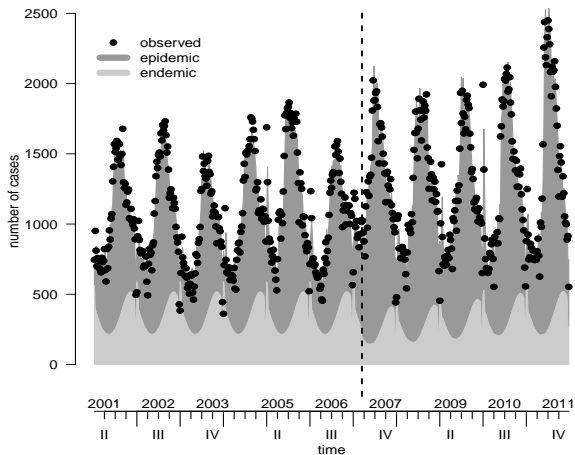


Retrospective fitting





Prospective Analysis: One-week-ahead predictions





Example of space-time point process modelling:
Using surveillance::twinstim to analyse
relative infectivity of two agents of invasive
meningococcal disease (Germany, 2002–2008)



Data representation ("epidataCS" class)

```
R> data("imdepi")  
R> print(imdepi, n=4, digits=4)
```

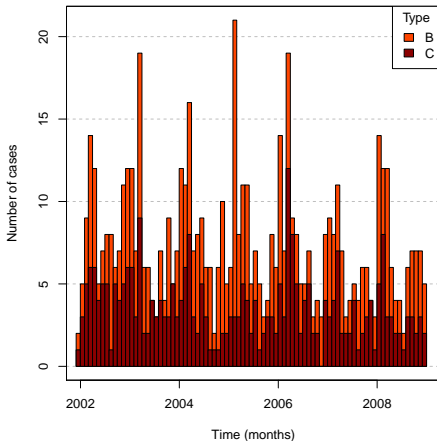
```
History of an epidemic  
Observation period: 0 -- 2557  
Observation window (bounding box): [4034, 4670] x [2687, 3543]  
Spatio-temporal grid (not shown): 84 time blocks, 413 tiles  
Types of events: 'B' 'C'  
Overall number of events: 636
```

	coordinates	ID	time	tile	type	eps.t	eps.s	sex
103	(4112.19, 3202.79)	1	0.2117	05554	B	30	200	male
402	(4122.51, 3076.97)	2	0.7124	05382	C	30	200	male
312	(4412.47, 2915.94)	3	5.5910	09574	B	30	200	female
314	(4202.64, 2879.7)	4	7.1170	08212	B	30	200	female
	agegrp	BLOCK	start	popdensity				
103	[3,19)	1	0	260.9				
402	[3,19)	1	0	519.4				
312	[19,Inf)	1	0	209.4				
314	[3,19)	1	0	1665.6				
[...]								



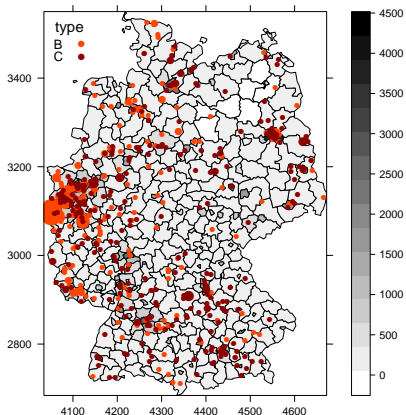
Data – aggregated over space

```
R> colTypes <- c(B="orangered",  
+               C="darkred")  
R> plot(imdepi, "time",  
+       t0.Date="2001-12-31",  
+       breaks="months", ylim=c(0,21),  
+       col=colTypes["B"])  
R> plot(imdepi, "time",  
+       t0.Date="2001-12-31",  
+       breaks="months", ylim=c(0,21),  
+       col=colTypes["C"],  
+       subset=type=="C", add=TRUE)  
R> legend("topright", bg="white",  
+       legend=names(colTypes),  
+       fill=colTypes, title="Type")
```



Data – aggregated over time

```
R> load(system.file(
+   "shapes", "districtsD.RData",
+   package="surveillance"))
R> ## [...]
R> spplot(districtsD, "POPDENSITY",
+   scales=list(draw=TRUE),
+   sp.layout = list("sp.points",
+   multloc, col=col,
+   pch=16, cex=cex),
+   legend = list(inside =
+   list(fun=slegend,
+   x=0.01, y=0.99)),
+   col.regions =
+   grey(seq(1,0,length=100)))
R>
R> ## simpler: plot(imdepi, "space")
```





Fit two-component point-process model

```
R> endemic <- addSeason2formula(
+   ~1 + offset(log(popdensity)) + I(start/365 - 3.5),
+   period=365, timevar="start")
R> fit <- twinstim(endemic = endemic,
+   epidemic = ~1 + type + agegrp, siaf = siaf.gaussian(),
+   data = imdepi, subset = !is.na(agegrp),
+   start = c("h.(Intercept)"=-20, "e.(Intercept)"=-15, "e.siaf.1"=4))
```

```
R> xtable(fit)
```

	RR	95%-CI	p-value
h.l(start/365)	0.955	0.91–1.00	0.04
h.sin(start * 2 * pi/365)	1.242	1.09–1.41	0.0008
h.cos(start * 2 * pi/365)	1.375	1.21–1.56	<0.0001
e.typeC	0.401	0.24–0.68	0.0006
e.agegrp[3,19]	1.983	1.05–3.74	0.035
e.agegrp[19,Inf]	0.757	0.31–1.87	0.55

```
R> RO(fit)
```

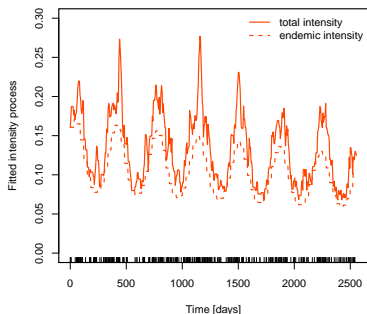
	B	C
Est.	0.25	0.10
2.5%	0.18	0.07
97.5%	0.34	0.14



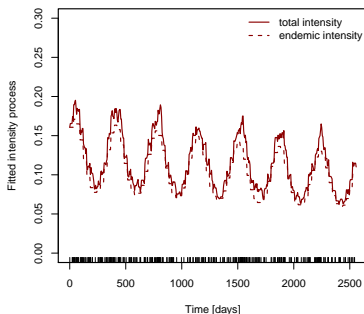
Fitted intensity process

```
R> intensityplot(fit, which = "total intensity", aggregate = "time",  
+               types = 1, col = colTypes[1], ylim = c(0,0.3))
```

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

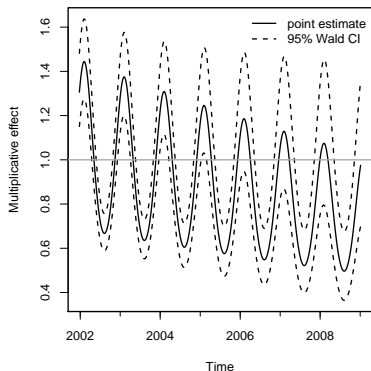


C:P1.5,2:F3-3

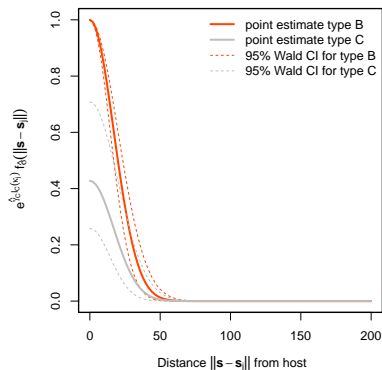




Fitted trend/seasonality, distance decay of infectivity



Typical IMD peak in late February
and minimum in August



Effective interaction range ≈ 50 km



Conclusion

- surveillance offers visualisation and modelling of (multivariate) surveillance time series and implementations of various detection algorithms.
- surveillance also provides a comprehensive framework for the modelling, inference and simulation of self-exciting spatio-temporal point processes (beyond epidemics)
- Focus of *prospective* surveillance is on outbreak detection.
- *Retrospective* surveillance tries to explain temporal and spatio-temporal patterns in the data through statistical modelling
- Current work: How to do model-based surveillance in practice
→ Journal of Statistical Software (special issue on space-time statistics)



References

- ▶ Farrington, C. P., Andrews, N. J., Beale, A. D., and Catchpole, M. A. (1996). A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 159(3):547–563.
- ▶ Held, L., Graziano, G., Frank, C., and Rue, H. (2006a). Joint spatial analysis of gastrointestinal infectious diseases. *Statistical methods in medical research*, 15(5):465–480.
- ▶ Held, L., Hofmann, M., Höhle, M., and Schmid, V. (2006b). A two-component model for counts of infectious diseases. *Biostatistics*, 7(3):422–437.
- ▶ Hulth, A., Andrews, N., Ethelberg, S., Dreesman, J., Faensen, D., van Pelt, W., and Schnitzler, J. (2010). Practical usage of computer-supported outbreak detection in five European countries. *Eurosurveillance*, 15(36):pii=19658.
- ▶ Höhle, M. (2007). surveillance: An R package for the monitoring of infectious diseases. *Computational Statistics*, 22:571–582.
- ▶ Höhle, M. (2009). Additive-multiplicative regression models for spatio-temporal epidemics. *Biometrical Journal*, 51(6):961–978.



References (cont.)

- ▶ Höhle, M., Meyer, S., and Paul, M. (2013). **surveillance**: *Temporal and spatio-temporal modeling and monitoring of epidemic phenomena*. R package version 1.5-2.
- ▶ Höhle, M. and Paul, M. (2008). Count data regression charts for the monitoring of surveillance time series. *Computational Statistics and Data Analysis*, 52(9):4357–4368.
- ▶ Lutz, A., Büttner, S., and Schüpbach, G. (2010). Vergleich der jährlichen campylobacter prävalenz von mastpoulets und menschen. Technical report, Bundesamt für Veterinärwesen, Veterinary Public Health Institute.
- ▶ Meyer, S., Elias, J., and Höhle, M. (2012). A space-time conditional intensity model for invasive meningococcal disease occurrence. *Biometrics*, 68(2):607–616.
- ▶ Paul, M. and Held, L. (2011). Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine*, 30(10):1118–1136.
- ▶ Riebler, A. (2004). Empirischer Vergleich von statistischen Methoden zur Ausbruchserkennung bei Surveillance Daten. Bachelor's thesis, Department of Statistics, University of Munich.
- ▶ Robert Koch Institute (2011). Report: Final presentation and evaluation of epidemiological findings in the EHEC O104:H4 outbreak, Germany 2011.



References (cont.)

- ▶ Robertson, C., Sawford, K., Daniel, S. L., Nelson, T. A., and Stephen, C. (2010). Mobile phone-based infectious disease surveillance system, Sri Lanka. *Emerging Infectious Diseases*, 16(10):1524–1531.
- ▶ Rossi, G., Lampugnani, L., and Marchi, M. (1999). An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine*, 18(16):2111–2122.