

# SISMID 2021: R Notes Infectious Disease Data

Jon Wakefield  
University of Washington

2021-07-09

## Lower Saxony Measles Data

As an example of the epidemic/endemic framework first described by Held et al. (2005) we examine data on measles.

These data are in the surveillance package and data consist of weekly measles counts over 2001 and 2002, for each of 17 administrative district Weser-Ems region of Lower Saxony, Germany.

Included in the dataset are a  $17 \times 17$  matrix of 0/1 entries indicating which areas share a common boundary.

There is also the population that is contained in each area, and various other data including vaccination information.

These notes are based on the vignette:

[https://cran.r-project.org/package=surveillance/vignettes/hhh4\\_spacetime.pdf](https://cran.r-project.org/package=surveillance/vignettes/hhh4_spacetime.pdf)

## Lower Saxony Measles: Reading in the Data

The data `measlesWeserEms` are of class `sts`.

The data object also contain a map of the region, as a `SpatialPolygonsDataFrame`

```
library(surveillance)
data("measlesWeserEms")
counts <- observed(measlesWeserEms)
map <- measlesWeserEms@map
populationFrac <- measlesWeserEms@populationFrac
```

# Lower Saxony Measles: Neighborhood information

```
weserems_adjmat <- poly2adjmat(map)
rowSums(weserems_adjmat)
## 03401 03402 03403 03404 03405 03451 03452 03453 03454 03455 03456 03457 03458
##    2      2      3      1      1      6      3      6      4      5      1      7      6
## 03459 03460 03461 03462
##    4      3      5      3
weserems_nbOrder <- nbOrder(weserems_adjmat, maxlag = Inf)
head(weserems_nbOrder)
##          03401 03402 03403 03404 03405 03451 03452 03453 03454 03455 03456 03457
## 03401      0      4      2      4      3      2      4      2      3      2      4      3
## 03402      4      0      3      4      3      2      1      2      2      2      3      1
## 03403      2      3      0      4      3      1      3      2      3      2      4      2
## 03404      4      4      4      0      5      3      4      2      2      4      3      3
## 03405      3      3      3      5      0      2      3      3      3      1      4      2
## 03451      2      2      1      3      2      0      2      1      2      1      3      1
##          03458 03459 03460 03461 03462
## 03401      1      3      2      1      3
## 03402      3      3      3      3      2
## 03403      1      3      2      1      3
## 03404      3      1      2      4      4
## 03405      3      4      4      2      2
## 03451      1      2      2      1      2
```

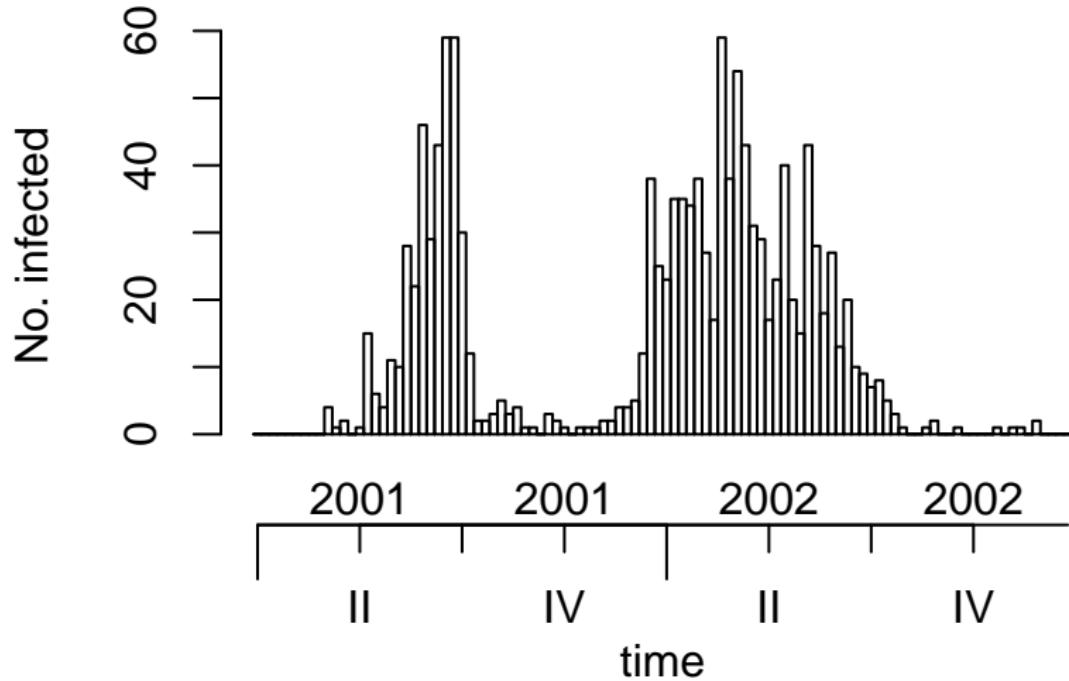
## Create a convenient data frame

Making a sts data frame helps with plotting

```
measlesWeserEms <- sts(counts, start = c(2001, 1),  
    frequency = 52, population = populationFrac, neighbourhood = weserems_nbOrder,  
    map = map)
```

## Time series of total counts

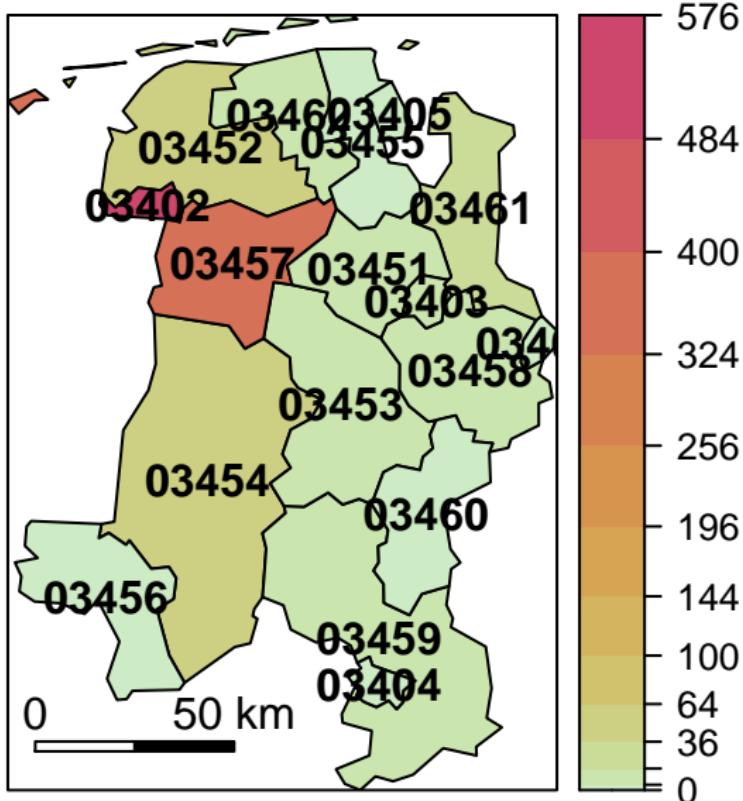
```
plot(measlesWeserEms, type = observed ~ time)
```



## Map of total counts with labels

```
plot(measlesWeserEms, type = observed ~ unit,
      population = measlesWeserEms@map$POPULATION/1e+05,
      labels = list(font = 2), colorkey = list(space = "right"),
      sp.layout = layout.scalebar(measlesWeserEms@map,
          corner = c(0.05, 0.05), scale = 50,
          labels = c("0", "50 km"), height = 0.03))
```

**2001/1 – 2002/52**

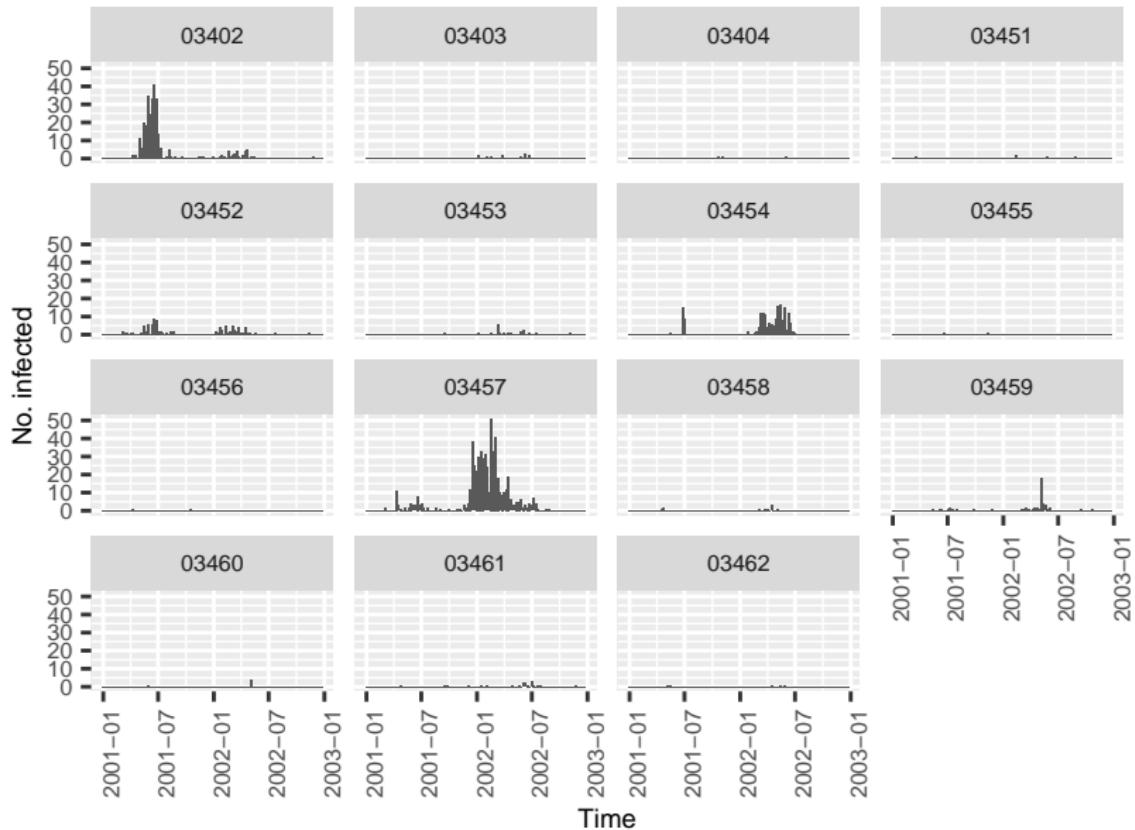


## Time series

We plot time series of counts by area

Two areas contain all zeroes, so do not plot these

```
autoplots.ts(measlesWeserEms,  
    units = which(colSums(observed(measlesWeserEms)) >  
        0))
```



# Animation

The commands

- `library(gridExtra)`
- `animate(measlesWeserEms)`

produce an animation.

Try them yourself!

## Model framework

Notation:  $Y_{it}$  are weekly counts of measles infections in area  $i$ ,  $N_i$  are population counts in area  $i$ , while  $e_i$  are population fractions

We assume a negative binomial model with

$$E[Y_{it} | \mu_{it}] = \mu_{it}$$

and

$$\text{var}[Y_{it} | \mu_{it}] = \mu_{it}(1 + \psi\mu_{it})$$

so that  $\psi = 0$  corresponds to a Poisson model.

# First model

We first fit the model

$$\mu_{it} = \underbrace{\lambda^{\text{AR}}}_{\exp(\alpha_0^{\text{AR}})} y_{i,t-1} + \underbrace{\lambda^{\text{NE}}}_{\exp(\alpha_0^{\text{NE}})} \sum_{j=1}^n w_{ji} y_{j,t-1} + e_i \lambda_t^{\text{EN}},$$

with endemic term:

$$\log(\lambda_t^{\text{EN}}) = \alpha_0^{\text{EN}} + \alpha_1 t + \gamma \sin(\omega t) + \delta \cos(\omega t)$$

## First model

In this model

- $\lambda^{\text{AR}}$  is the contribution from the susceptible's own area
- $\lambda^{\text{NE}}$  is the contribution from the susceptible's neighboring areas
- $w_{ji}$  are binary indicators of sharing a boundary
- $\lambda_t^{\text{EN}}$  is the endemic term,
- $\alpha_1$  is a slope parameter describing the large scale endemic temporal trend,
- $\gamma$  and  $\delta$  are seasonal parameters and do not vary across areas,  
 $\omega = (2\pi)/52.$

# First model

NegBin1 gives a single overdisperion parameter, i.e., common to all areas.

```
# to perform the following analysis using
# biweekly aggregated measles counts
measlesWeserEms <- aggregate(measlesWeserEms,
  by = "time", nfreq = 26)
measlesModel_basic <- list(end = list(f = addSeason2formula(~1 +
  t, period = measlesWeserEms@freq), offset = population(measlesWeserEms)),
  ar = list(f = ~1), ne = list(f = ~1,
    weights = neighbourhood(measlesWeserEms) ==
      1), family = "NegBin1")
measlesFit_basic <- hhh4(stsObj = measlesWeserEms,
  control = measlesModel_basic)
```

# Model results

```
summary(measlesFit_basic, idx2Exp = TRUE, amplitudeShift = TRUE)
##
## Call:
## hhh4(stsObj = measlesWeserEms, control = measlesModel_basic)
##
## Coefficients:
##                               Estimate Std. Error
## exp(ar.1)                  0.739254  0.105078
## exp(ne.1)                  0.008591  0.003603
## exp(end.1)                 2.701956  0.753291
## exp(end.t)                 0.987949  0.010296
## end.A(2 * pi * t/26)       1.025761  0.219054
## end.s(2 * pi * t/26)      -0.681901  0.166360
## overdisp                   1.940325  0.298668
##
## Log-likelihood: -674.09
## AIC:             1362.19
## BIC:             1395.54
##
## Number of units:      17
## Number of time points: 51
```

## Model results

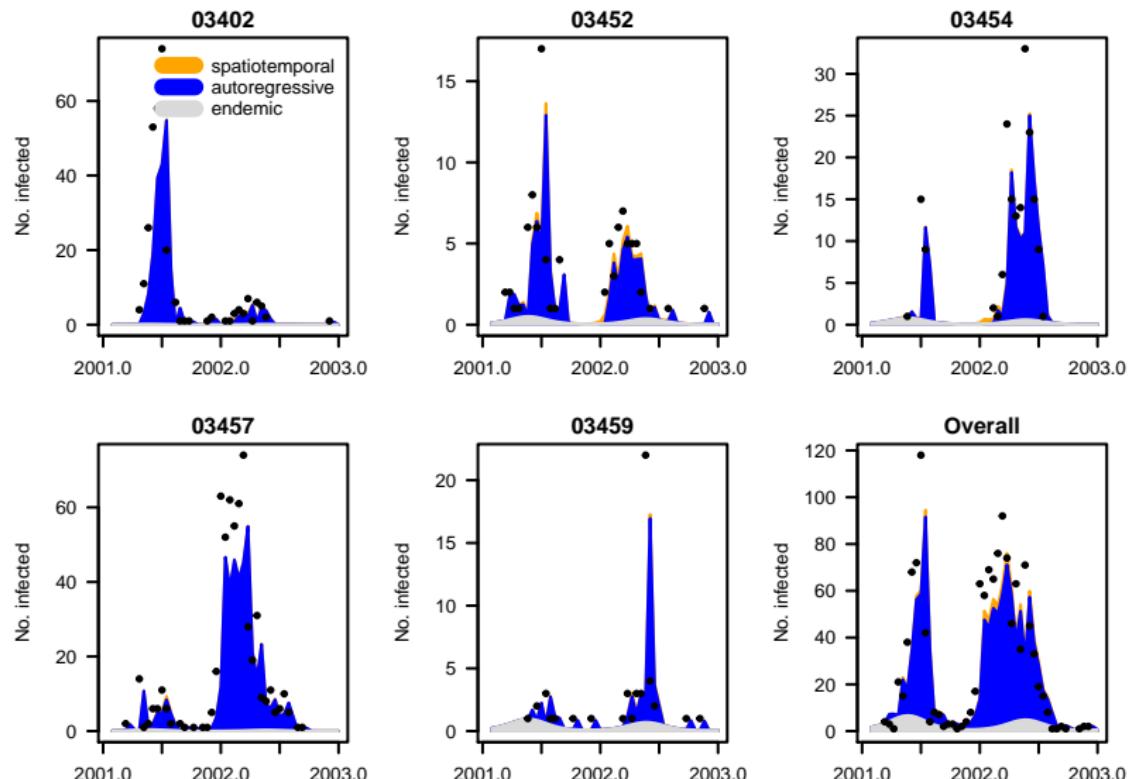
```
confint(measlesFit_basic, parm = "overdisp")
##           2.5 %   97.5 %
## overdisp 1.354946 2.525703
```

This interval suggests a Poisson model, with  $\psi = 0$  would be a poor fit.

## Results for model

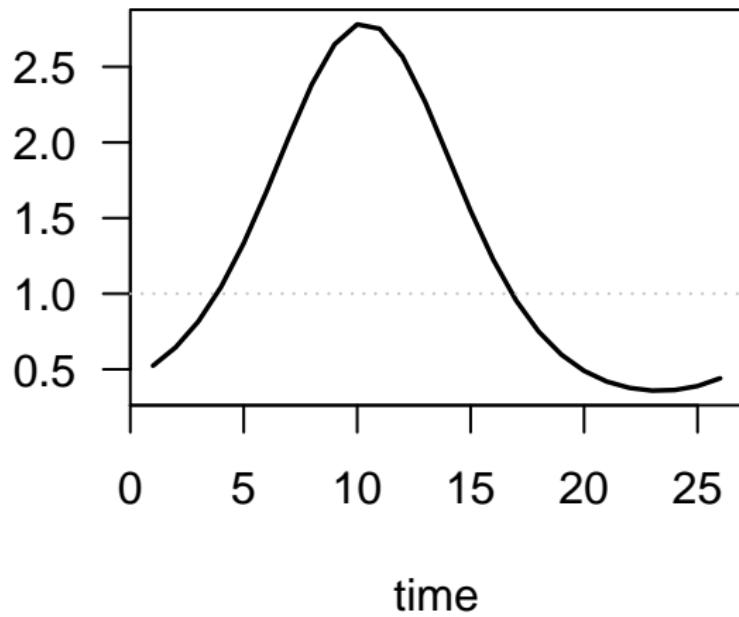
```
districts2plot <- which(colSums(observed(measlesWeserEms)) >
  50)
par(mfrow = c(2, 3), mar = c(3, 5, 2, 1), las = 1)
op <- par(cex = 0.4)
plot(measlesFit_basic, type = "fitted", units = districts2plot,
  hide0s = TRUE, par.settings = NULL, legend = 1)
fitted_components <- plot(measlesFit_basic, type = "fitted",
  total = TRUE, hide0s = TRUE, par.settings = NULL,
  legend = FALSE)
```

# Results for model



## Plot of endemic component

```
plot(measlesFit_basic, type = "season", components = "end",
     main = "")
```



## More summaries

```
AIC(measlesFit_basic, update(measlesFit_basic,
  family = "Poisson"))
##                                     df      AIC
## measlesFit_basic                  7 1362.188
## update(measlesFit_basic, family = "Poisson") 6 1954.850
Sprop <- matrix(1 - measlesWeserEms@map@data$vacc1.2004,
  nrow = nrow(measlesWeserEms), ncol = ncol(measlesWeserEms),
  byrow = TRUE)
summary(Sprop[1, ])
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.03057 0.04808 0.05808 0.06750 0.08296 0.13983
```

# Random effects

```
measlesFit_ri <- update(measlesFit_basic,
  end = list(f = update(formula(measlesFit_basic)$end,
    ~ . + ri() - 1)),
  ar = list(f = update(formula(measlesFit_basic)$ar,
    ~ . + ri() - 1)))
```

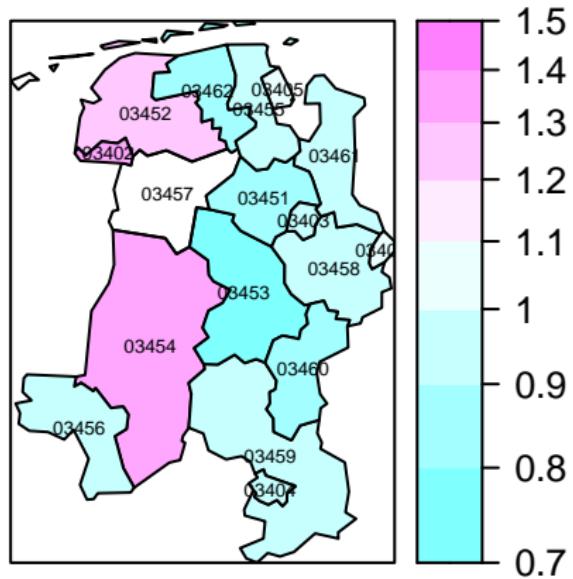
```
summary(measlesFit_ri, amplitudeShift = TRUE, maxEV = TRUE)
##
## Call:
## hhh4(stsObj = object$stsObj, control = control)
##
## Random effects:
##             Var     Corr
## ar.ri(iid) 0.1592
## end.ri(iid) 1.4424 0
##
## Fixed effects:
##                               Estimate Std. Error
## ar.ri(iid)           -0.655603  0.206697
## ne.1                -5.460159  0.649477
## end.t                 0.006881  0.010272
## end.A(2 * pi * t/26)  1.310432  0.219093
## end.s(2 * pi * t/26) -0.589227  0.144846
## end.ri(iid)          0.404900  0.414779
## overdisp              1.310582  0.213088
##
## Epidemic dominant eigenvalue: 0.79
##
## Penalized log-likelihood: -635.26
## Marginal log-likelihood:   -39.21
##
```

## Random effects

```
head(ranef(measlesFit_ri, tomatrix = TRUE), n = 3)
##           ar.ri(iid) end.ri(iid)
## 03401  0.00000000 -1.40341274
## 03402  0.29147877  2.38732785
## 03403 -0.09296041 -0.04950659
```

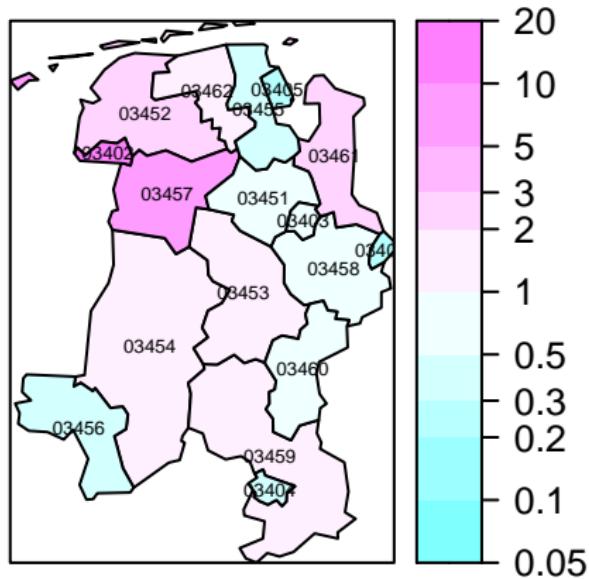
# AR random effects

```
plot(measlesFit_ri, type = "ri", component = "ar",
      exp = TRUE, labels = list(cex = 0.4))
```



# END random effects

```
plot(measlesFit_ri, type = "ri", component = "end",
      exp = TRUE, labels = list(cex = 0.4))
```



Fits to areas with more than 50 cases, and overall

```
par(mfrow = c(2, 3), mar = c(3, 5, 2, 1), las = 1)
plot(measlesFit_ri, type = "fitted", units = districts2plot,
     hide0s = TRUE, par.settings = NULL, legend = 1)
plot(measlesFit_ri, type = "fitted", total = TRUE,
     hide0s = TRUE, par.settings = NULL, legend = FALSE)
```

Blue is AR, Grey is endemic, Orange is NE

