

# SISMID 2022: R Notes on Infectious Disease Data

Jon Wakefield  
Departments of Biostatistics and Statistics  
University of Washington

2022-07-17

## Overview

We illustrate the use of the `hhh4` function that fits the Held et al. (2005) epidemic-endemic model class, within the `surveillance` package. For more details see Meyer et al (2017).

## Lower Saxony Measles Data

As an example of the epidemic/endemic framework first described by Held et al. (2005) and expanded upon in various papers, see in particular Held and Paul (2012).

We examine spatio-temporal count data on measles incidence. These data are in the `surveillance` package and data consist of weekly measles counts over 2001 and 2002, for each of 17 administrative districts in the Weser-Ems region of Lower Saxony, Germany.

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

There is also the population total 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)

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
```

Create the neighborhood information we will need later.

```
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
```

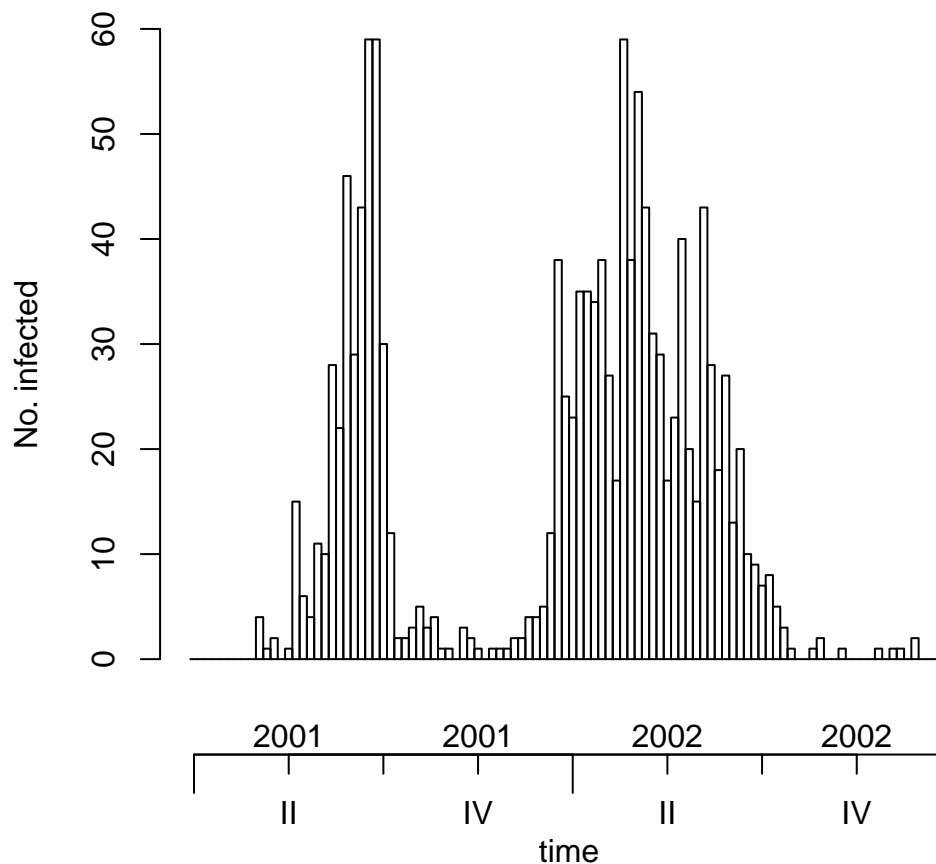
```
## 03402    4    0    3    4    3    2    1    2    2    2    3    1
## 03403    2    3    0    4    3    1    3    2    3    2    4    2
## 03404    4    4    4    0    5    3    4    2    2    4    3    3
## 03405    3    3    3    5    0    2    3    3    3    1    4    2
## 03451    2    2    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
```

We make an `sts` dataframe that will help 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)
```

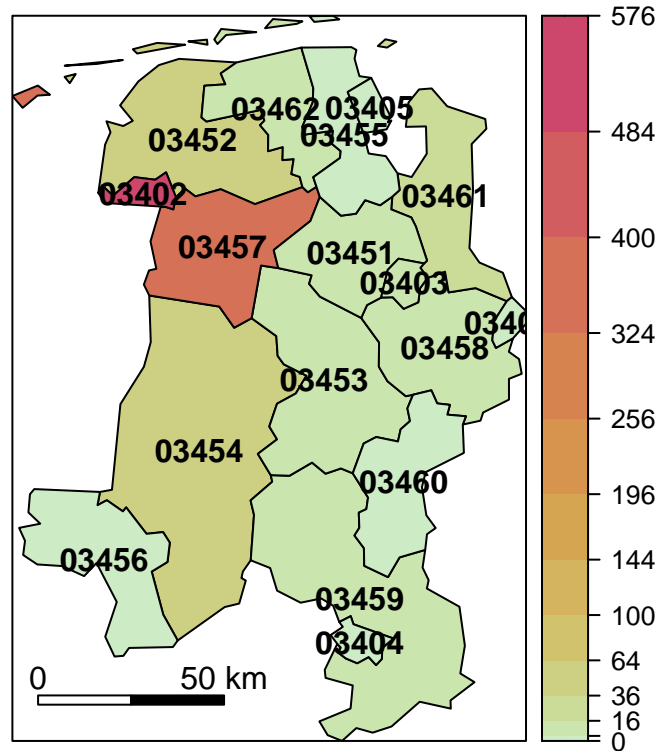


Map of total counts with labels:

```
plot(measlesWeserEms, type = observed ~ unit,
  population = measlesWeserEms@map$POPULATION/1e+05,
  labels = list(font = 2), colorkey = list(space = "right"),
```

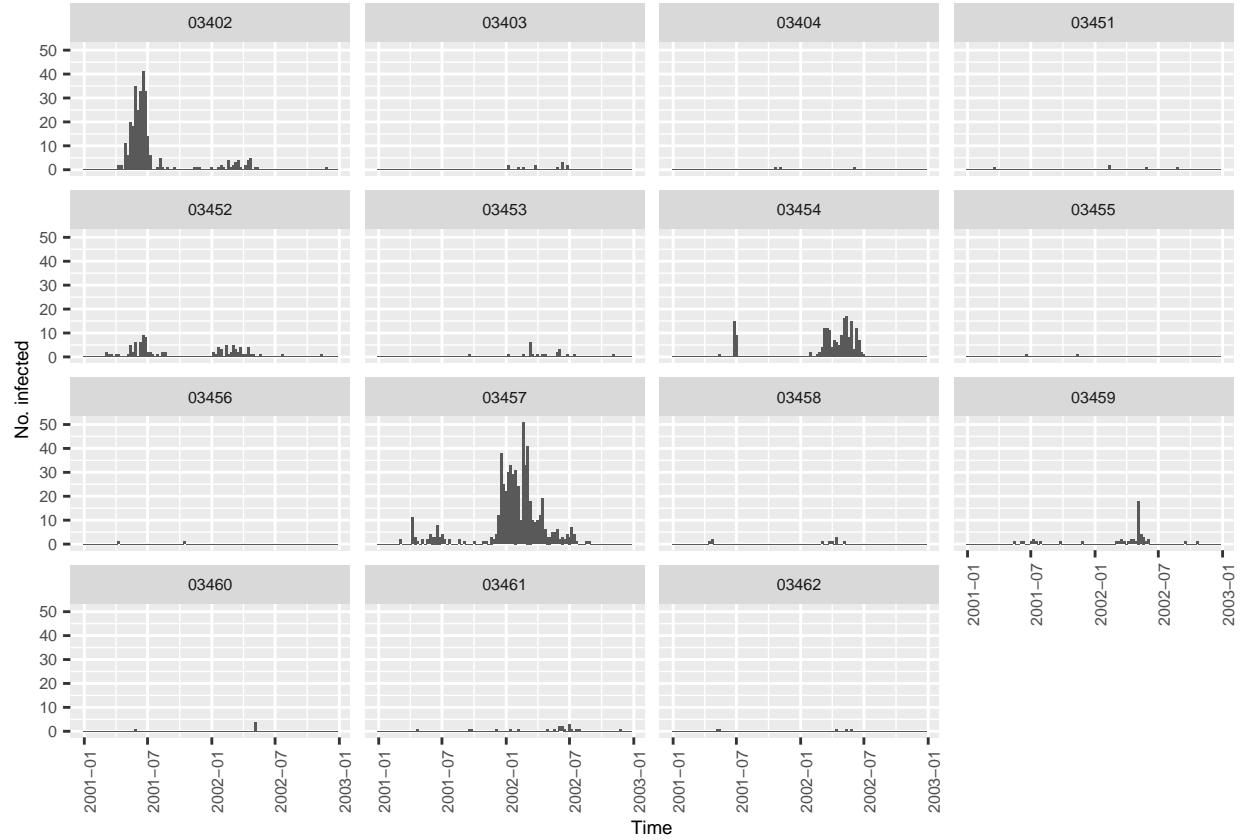
```
sp.layout = layout.scalebar(measlesWeserEms@map,
  corner = c(0.05, 0.05), scale = 50,
  labels = c("0", "50 km"), height = 0.03))
```

## 2001-W01 to 2002-W52



We plot time series of counts by area. Two areas contain all zeroes, so do not plot these.

```
autoplot.sts(measlesWeserEms, units = which(colSums(observed(measlesWeserEms)) >
  0)) + ggplot2::theme(text = ggplot2::element_text(size = 7),
  axis.text.x = ggplot2::element_text(angle = 90, hjust = 1))
```



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.

We first fit the model

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

with endemic term:

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

In this model

- $\lambda^{AR}$  is the contribution from the susceptible's own area
- $\lambda^{NE}$  is the contribution from the susceptible's neighboring areas
- $w_{ji}$  are binary indicators of sharing a boundary
- $\lambda_t^{EN}$  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,  $\omega = (2\pi)/52$ .

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

In the following analysis we use 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)

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. More confirmation of this by examining the AIC of the difference. The AIC is given by

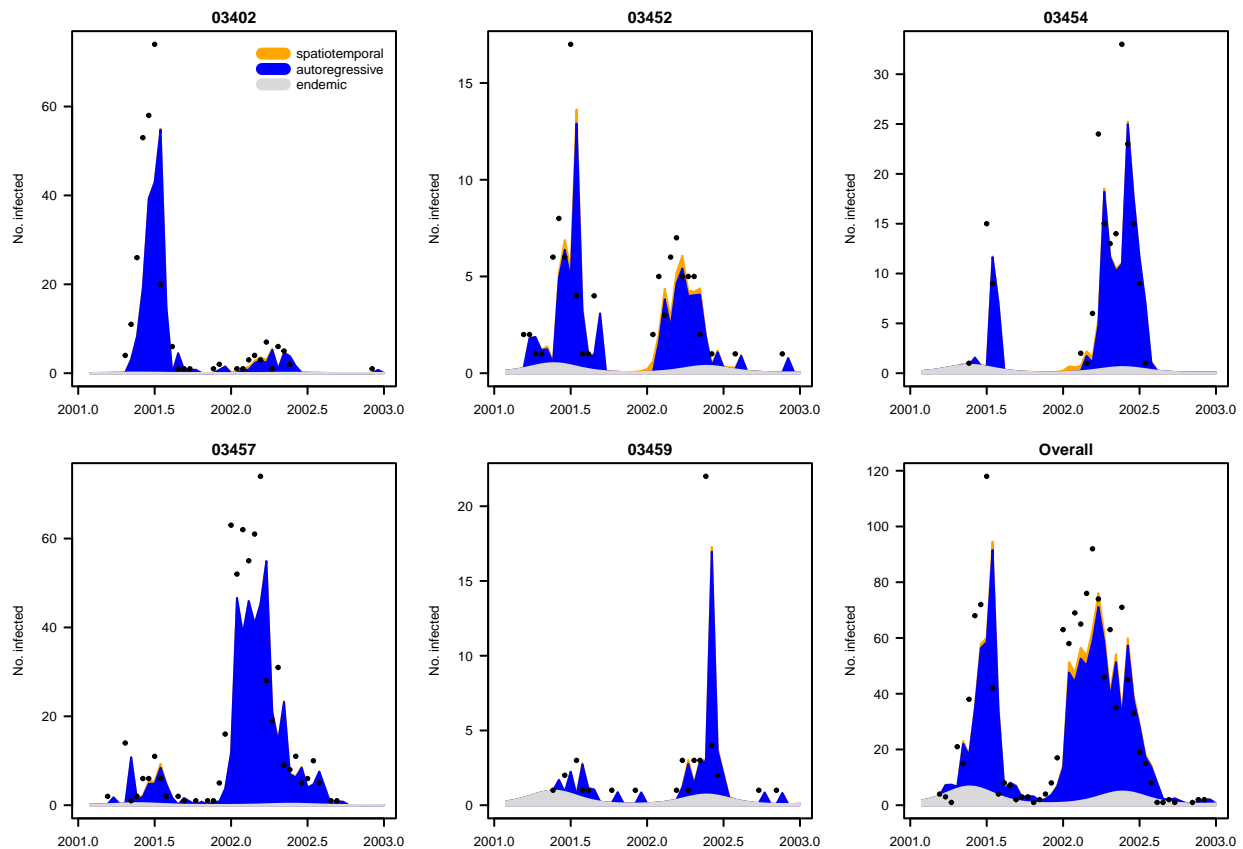
$$-2 \log\text{-likelihood} + 2p,$$

where  $p$  is the number of parameters. For a sequence of models, the one with the smallest AIC is preferred. Here, the negative binomial model is strongly preferred.

```
AIC(measlesFit_basic, update(measlesFit_basic, family = "Poisson"))
##                                df      AIC
## measlesFit_basic                7 1362.188
## update(measlesFit_basic, family = "Poisson") 6 1954.850
```

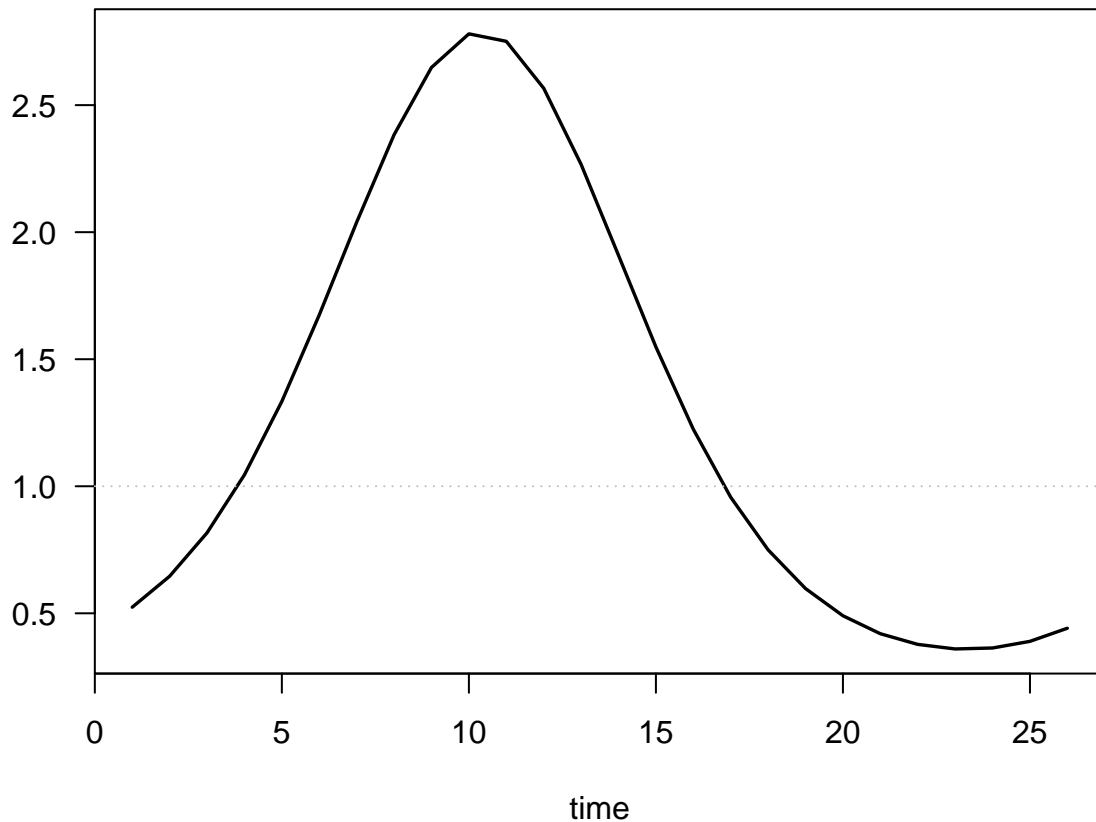
We provide summaries of the fits in those areas with more than 50 counts.

```
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 = TRUE)
plot(measlesFit_basic, type = "fitted", total = TRUE, hide0s = TRUE, par.settings = NULL,
     legend = FALSE) -> fitted_components
```



We now plot the endemic (seasonal) component.

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



## Add Random Effects

We now add area-specific random effects to the endemic and AR components.

```
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)))
```

Examine the summary of the fit.

```
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
##
## Number of units: 17
## Number of time points: 51
```

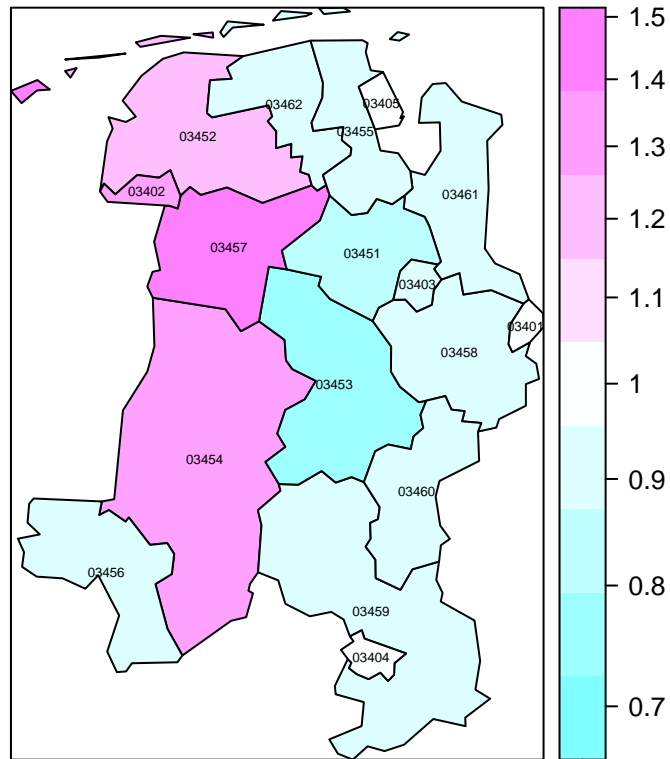
Random effect point estimates.

```
ranef(measlesFit_ri, tomatrix = TRUE)
##      ar.ri(iid) end.ri(iid)
## 03401 0.00000000 -1.40341274
## 03402 0.29147877 2.38732785
## 03403 -0.09296041 -0.04950659
## 03404 -0.04493255 -1.13587931
## 03405 0.00000000 -1.46838133
## 03451 -0.17368257 -0.51233522
## 03452 0.19401594 0.83948474
## 03453 -0.26849724 0.12984182
## 03454 0.27529339 0.51997027
## 03455 -0.08912515 -0.89986768
## 03456 -0.08860723 -0.90900208
## 03457 0.41563157 1.90205481
## 03458 -0.05945848 -0.14969753
## 03459 -0.07607612 0.25167679
## 03460 -0.11595302 -0.39490761
## 03461 -0.05072041 0.76631713
## 03462 -0.11640648 0.12631669
```

Map the AR random effects

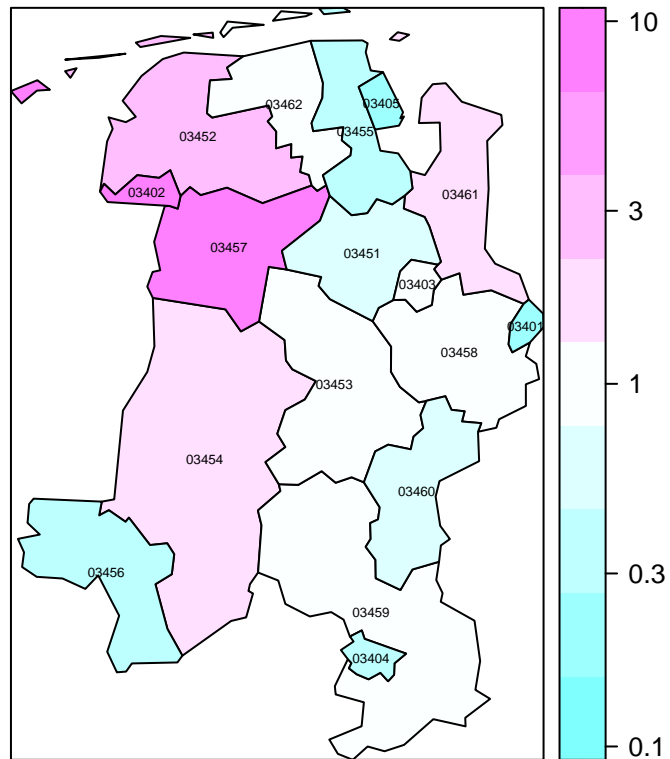
```
plot(measlesFit_ri, type = "ri", component = "ar", exp = TRUE,
     labels = list(cex = 0.4))
```





Map the END random effects

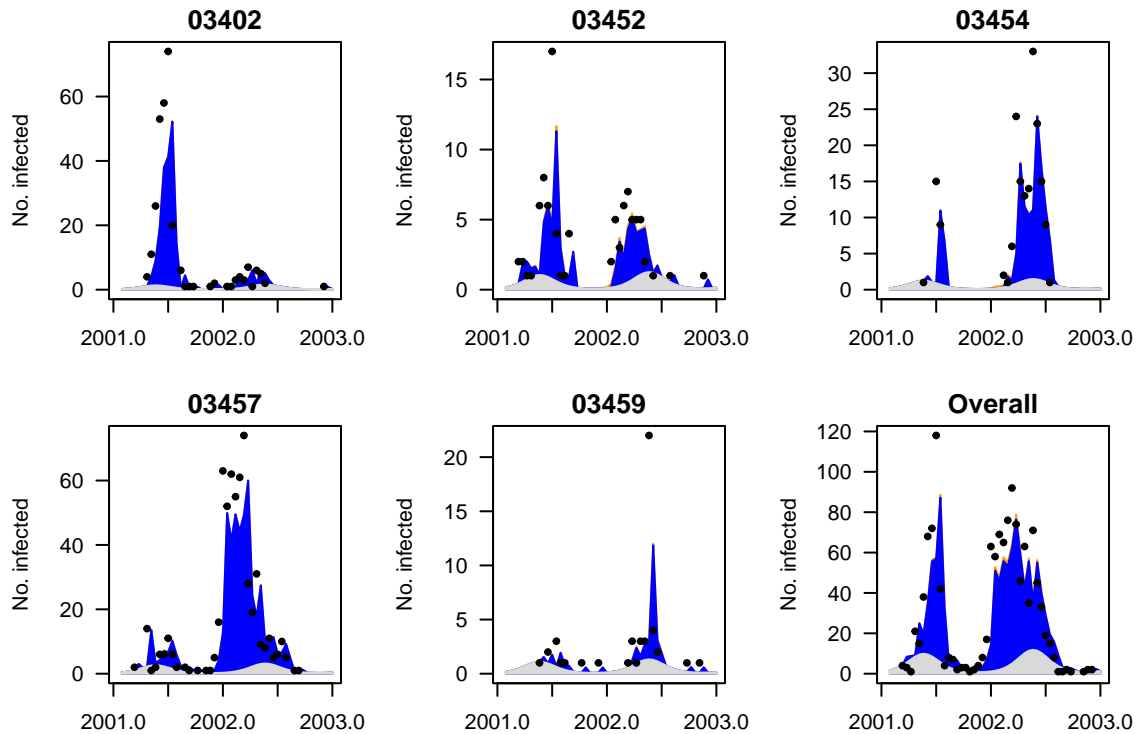
```
plot(measlesFit_ri, type = "ri", component = "end", exp = TRUE,
     labels = list(cex = 0.4))
```



We examine the fits to the areas with more than 50 cases, and overall.

Blue is the AR contribution, Grey is endemic, Orange is NE

```
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 = FALSE)
plot(measlesFit_ri, type = "fitted", total = TRUE, hide0s = TRUE,
     par.settings = NULL, legend = FALSE)
```



#### hh4 for Covid-19 modeling

- Alipour et al 2020
- Benimana et al 2021
- BerlamannHaustein2020 2020 [https://www.cesifo.org/DocDL/cesifo1\\_wp8446.pdf](https://www.cesifo.org/DocDL/cesifo1_wp8446.pdf)
- BND2021 2021 <https://rss.org.uk/news-publication/publications/journals/special-topic-meeting-on-r/>
- CelaniGiudici2021 2021 <https://doi.org/10.1016/j.spasta.2021.100528>
- Dickson2020 2020 <https://doi.org/10.1007/s11071-020-05853-7>
- FritzKauermann2020 2020 <https://arxiv.org/abs/2008.03013>
- Fronterre2020 2020 <https://doi.org/10.1101/2020.05.15.20102715>
- Giuliani2020 2020 <https://doi.org/10.2139/ssrn.3559569>
- Grimee2021 2021 <https://doi.org/10.1101/2021.05.19.21257329>
- Rui2021 2021 <https://doi.org/10.3390/ijerph18020774>
- Ssentongo2021 2021 <https://doi.org/10.1073/pnas.2026664118>