Envenoming incidence CAR model

Gerardo Martin

1st August 2019

Intro and Methods

This is the analysis of the envenoming model. I modelled envenoming incidence as a subset of snakebites. To do this I estimated the probability that a snakebite results in envenoming as a function of the snakes present weighted by envenoming severity. The I modelled the number of snakebites that result in envenoming as a negative binomial distribution, simulating the dispersion parameter r with MCMC. The model includes the effect of land cover as an intercept as I did not find any evidence of direct interaction between snakes and land cover.

The number of envenoming cases after a snakebite were used as the response variables. The first part of the model is:

$$\ln\left(\frac{P}{1-P}\right) = \text{int}_l + \sum_i^n I_i \times S_i$$

The number of envenoming cases was then the product of:

$$H_{envenomed} = P \times H_{bitten}$$

Results

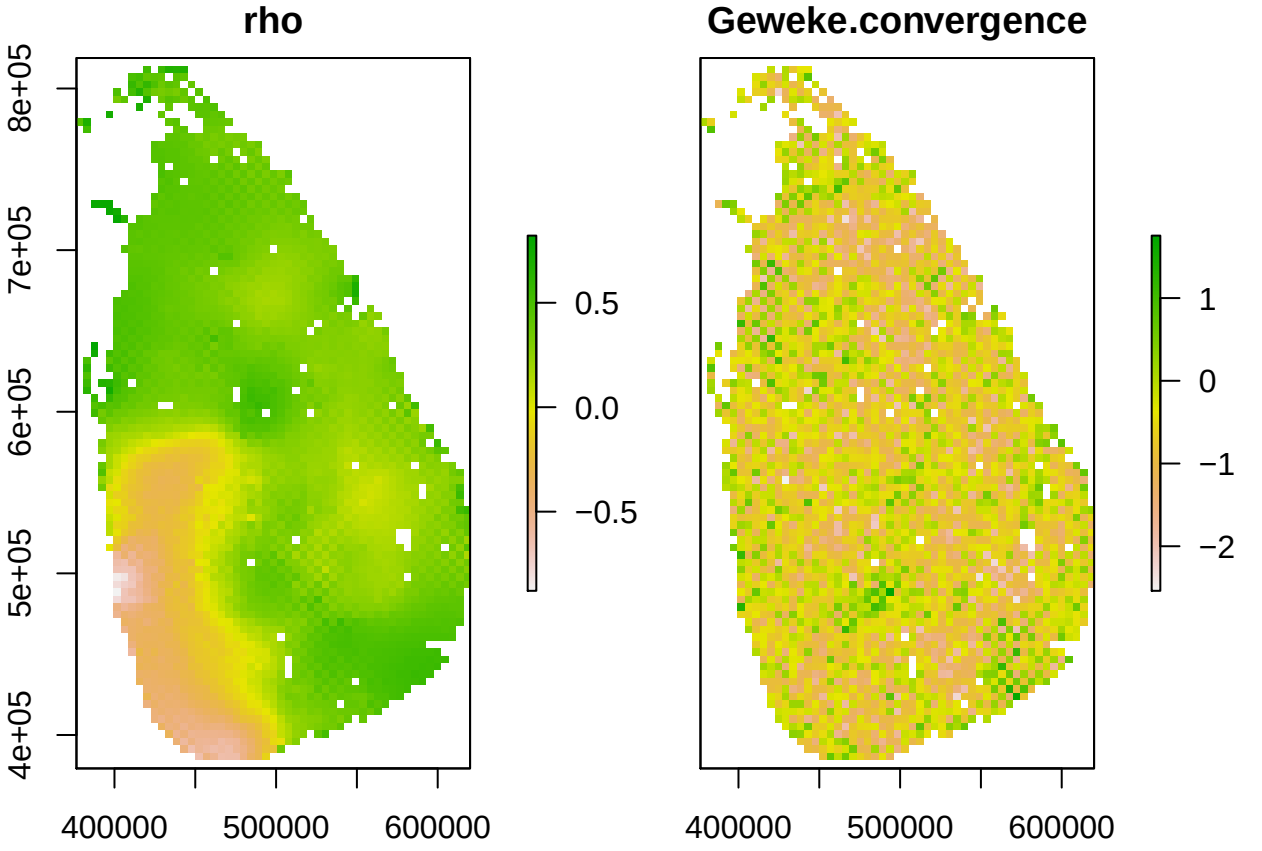
Diagnosis

First I begin to explore what are the parameter estimates in all three run chains. At the same time it is possible to see whether the chains have converged if the numbers in the `Rhat` column do not exceed 1.1. At the end of the summary the DIC appears. The model which I'm describing here had the lowest DIC. It might be possible to further reduce it but would begin to be biologically dubious.

	Median	Std.Dev	Cr.025	Cr.975	Gelman.conv
inter[1]	-0.0568716	0.0430676	-0.1416905	0.0291516	1.0032743
inter[2]	-0.3330455	0.0317759	-0.3939705	-0.2688773	1.0017263
inter[3]	-0.2375867	0.0468416	-0.3311108	-0.1486992	1.0141526
inter[4]	-0.4869941	0.0302641	-0.5531077	-0.4325493	0.9997987
inter[5]	0.5283284	0.1126123	0.3293302	0.7565075	1.0137542
s.effects[1]	197.8827458	13.4113139	171.4964980	225.2700061	1.0044031
s.effects[2]	-3.8688706	31.6877300	-63.1016445	60.5733289	1.0064001
s.effects[3]	-165.1297559	25.6144929	-214.0628846	-113.7864068	1.0090875
s.effects[4]	5.2209880	31.9818926	-50.8584790	71.1829622	1.0025758
s.effects[5]	-34.2488378	19.7749855	-70.3180729	4.4564775	1.0231450
s.effects[6]	-373.9742219	27.6675627	-422.9302326	-314.6643205	1.0090922
s.effects[7]	-1.6405372	31.5927138	-64.2461047	56.7570370	1.0009647

	Median	Std.Dev	Cr.025	Cr.975	Gelman.conv
r	49.6930835	0.4174706	48.7177017	49.9998852	1.0008273

```
random.effects <- lapply(envenom.incidence, function(x1){x1[, paste0("rho[", 1:3057, "]")]})

rho <- rbind(random.effects[[1]], random.effects[[2]], random.effects[[3]])
rho <- apply(rho, 2, median)

rho.r <- rasterFromXYZ(data.frame(incid.data[, c("x", "y")], rho))

rho.diag <- geweke.diag(random.effects)
rho.diag.r <- rasterFromXYZ(data.frame(incid.data[, c("x", "y")], rho.diag$z))

rho.stack <- stack(rho.r, rho.diag.r)
names(rho.stack) <- c("rho", "Geweke convergence")

plot(rho.stack)
```



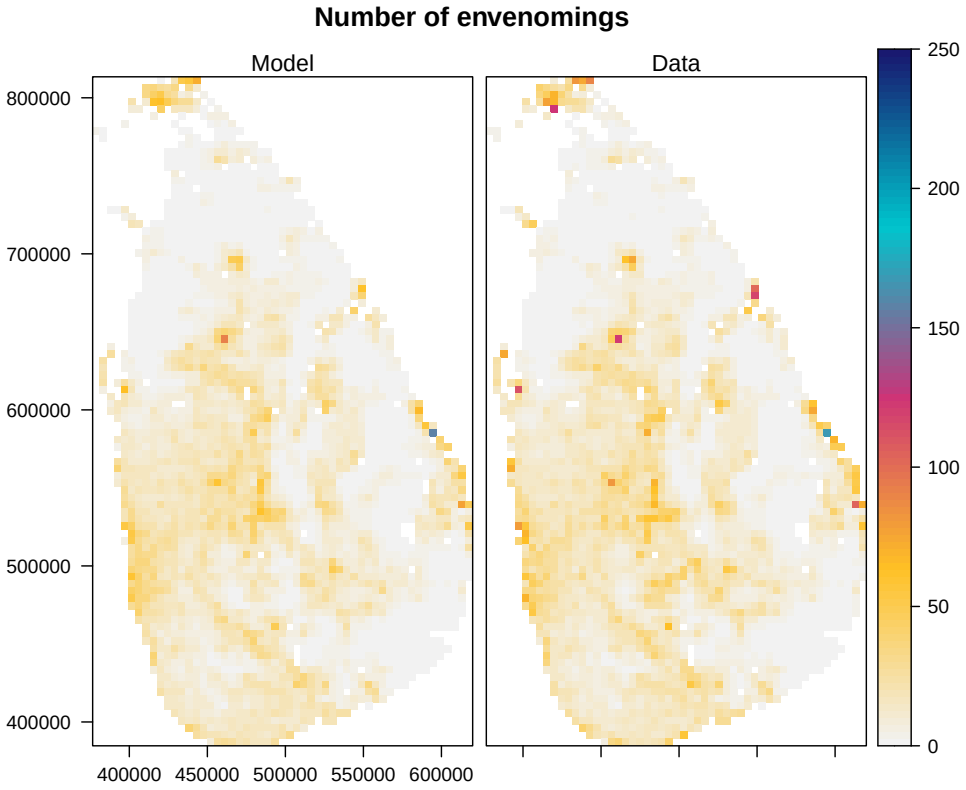
The tests show in agreement with the `Rhat` column of the model summary, that all chains mixed well and converged.

Model predictions

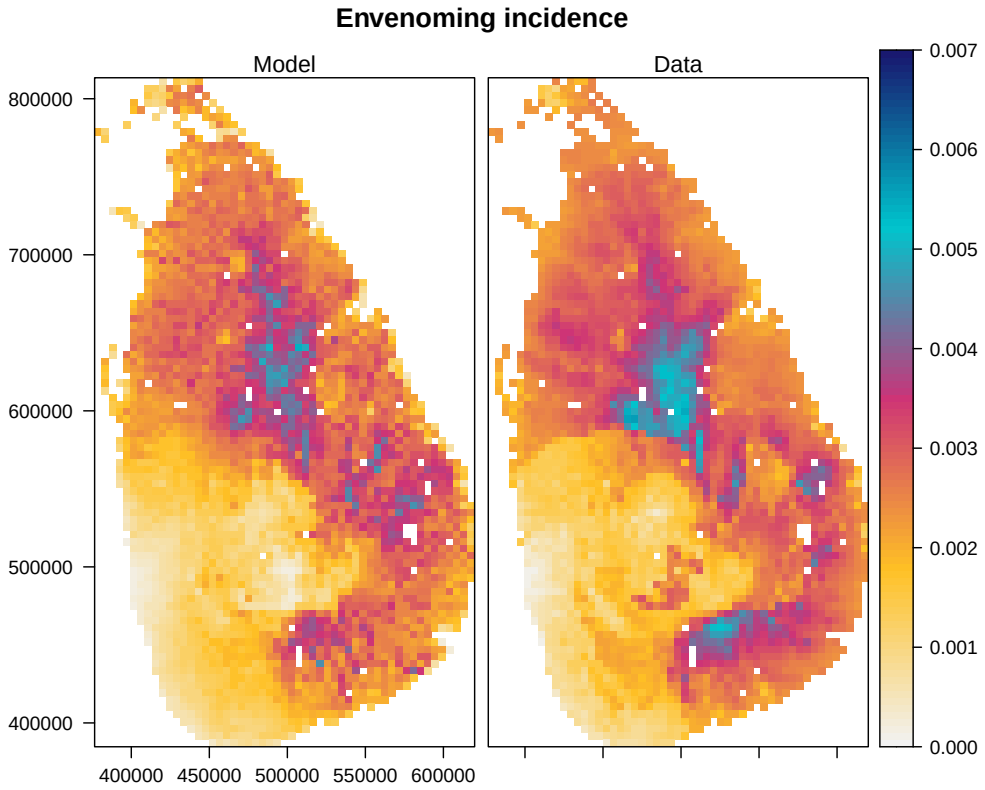
Given the structure of the model passed to JAGS, the `R` function to predict and be able to see the patterns of envenoming with the estimated parameters is:

Then by extracting the median of the posteriors and transforming them o raster format we can plot the

number of envenoming cases predicted by the model and compare it with the number of envenoming cases used as data:

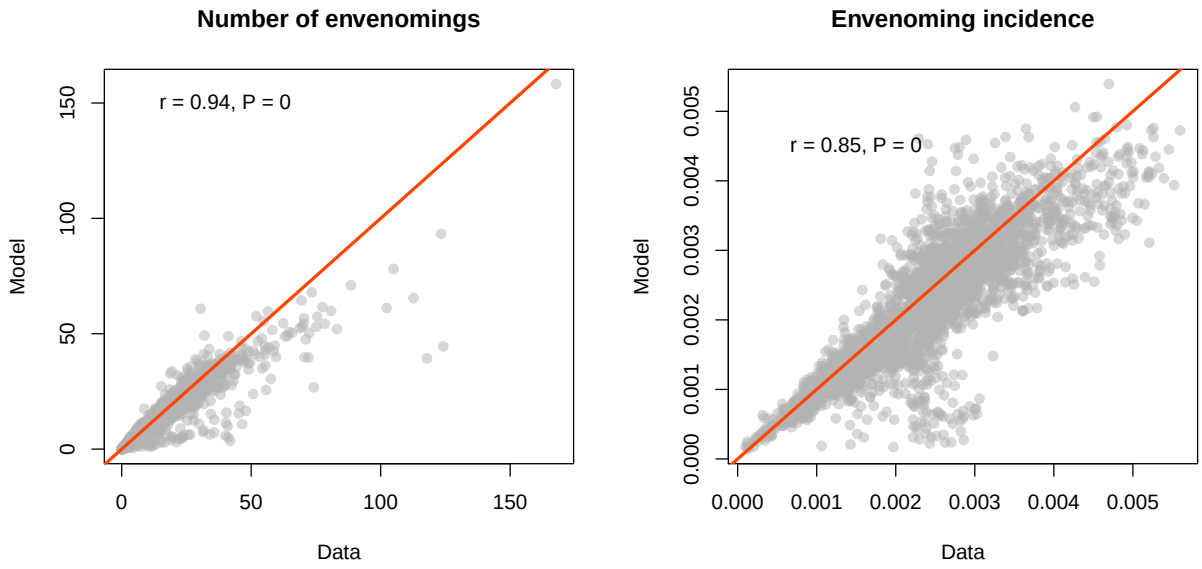


Then if the number of cases is divided by the total population density we get envenoming incidence:

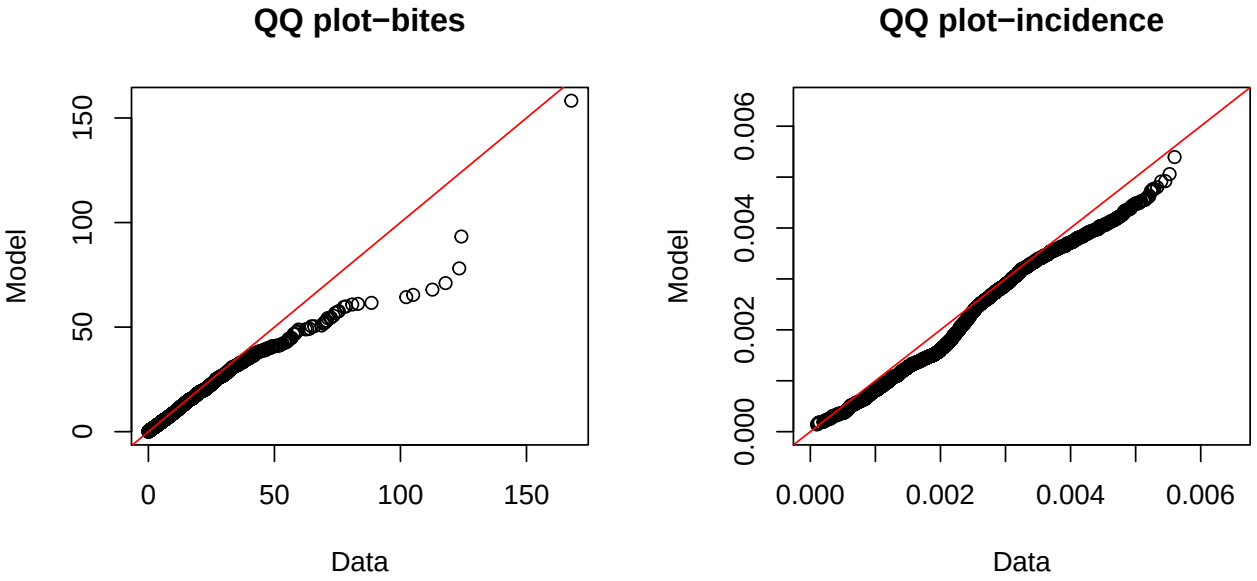


I believe the obvious visual differences between these two maps are mainly caused by the approach I took in the first place of modelling envenoming cases as a subset of snakebite cases, whereas the original analyses did not use this approach. In fact in few of the grid cells envenoming incidence was higher than snakebite incidence, which is in fact impossible. To correct this I replaced those envenoming incidence values with the snakebite incidence value (100% of the snakebite cases resulted in envenoming).

As a mode of validation I analyse the correlation between the patterns produced by the model and those of the data. The plots below show the correlation coefficient between incidence produced by the model and the data, and the number of bites. The red line represents the "gold standard", which would be a regression line with intercept equal to zero and slope equal to one. The dashed blue line is a smoothing loess regression that shows the actual trend between model and data:



The correlation in both cases is relatively high and the trend does not look to depart too much from the straight (reference) line.



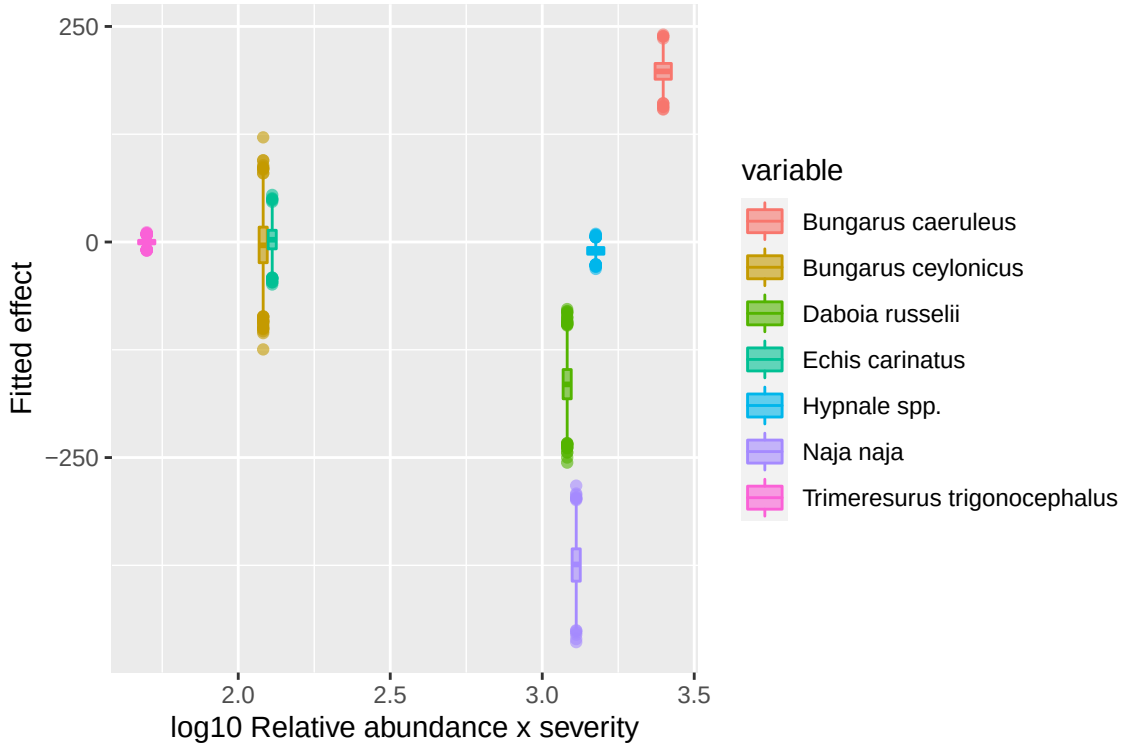
These two plots suggest that the statistical distribution of the number of envenoming cases and rates are both very similar to data. Envenoming incidence, although tens to predict higher incidence more frequently than there are in the data.

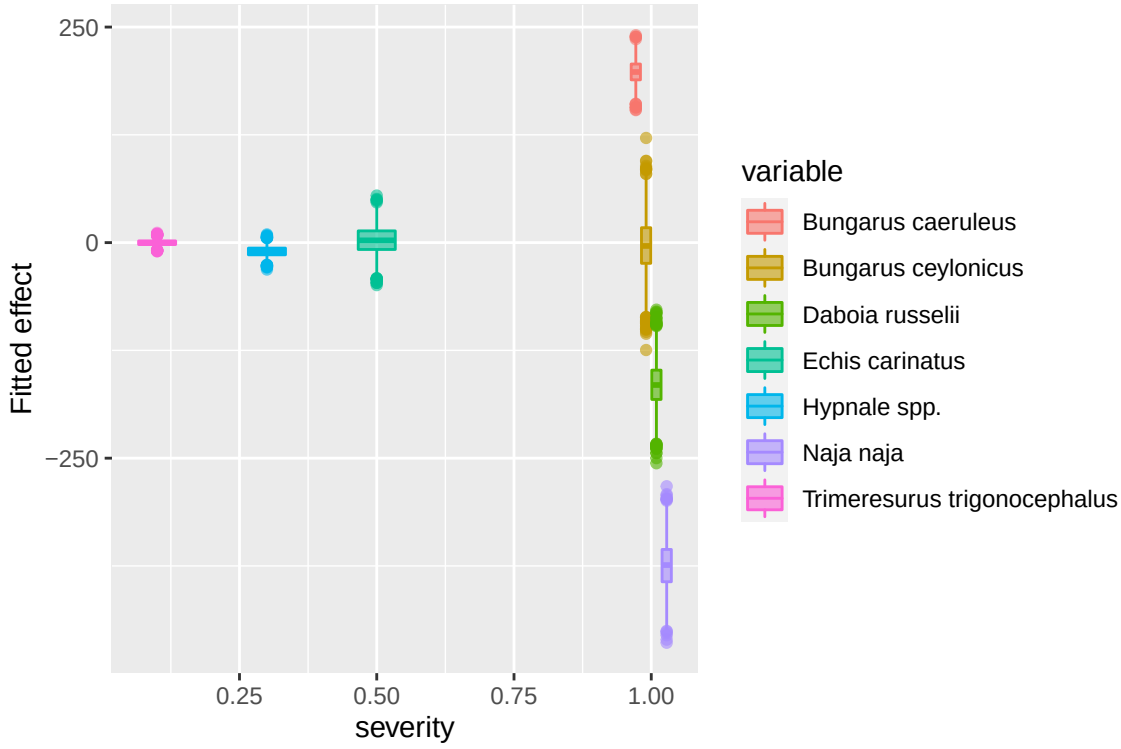
Relationship between species fitted effects and indices of severity and relative abundance.

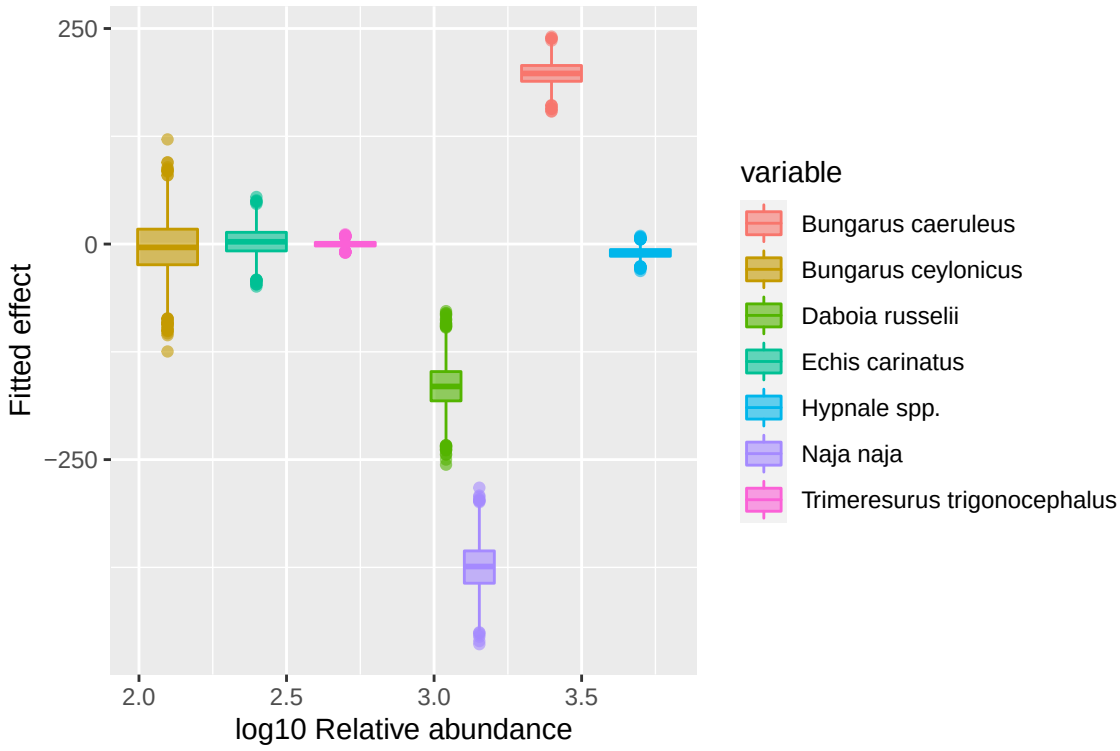
Comparing the fitted effects and the relative abundance and bite severity indices used previously to explain envenoming.

```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:latticeExtra':
```

```
##
##     layer
```

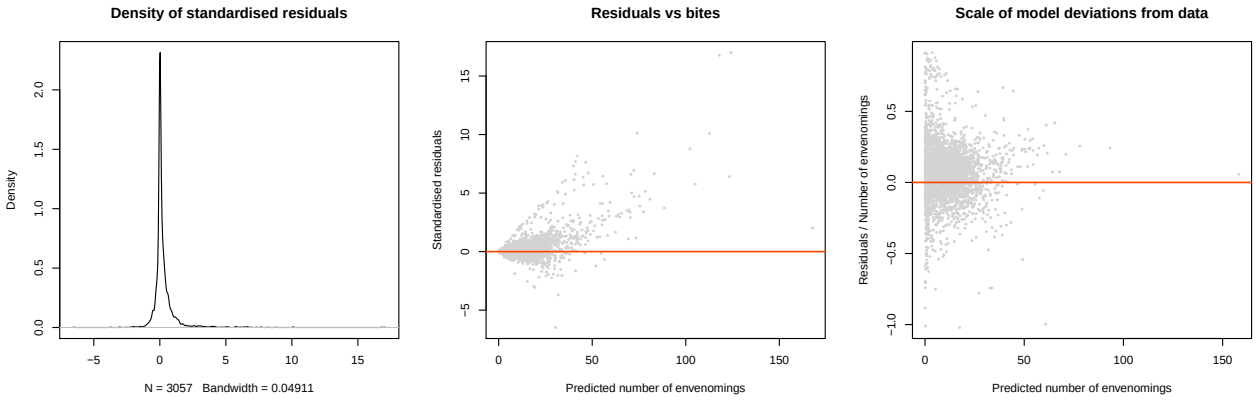

250
0
−250
Fitted effect
2.0
2.5
3.0
3.5
log10 Relative abundance x severity
variable
Bungarus caeruleus
Bungarus ceylonicus
Daboia russelii
Echis carinatus
Hypnale spp.
Naja naja
Trimeresurus trigonocephalus


250
0
−250
Fitted effect
0.25
0.50
0.75
1.00
severity
variable
Bungarus caeruleus
Bungarus ceylonicus
Daboia russelii
Echis carinatus
Hypnale spp.
Naja naja
Trimeresurus trigonocephalus



Residuals

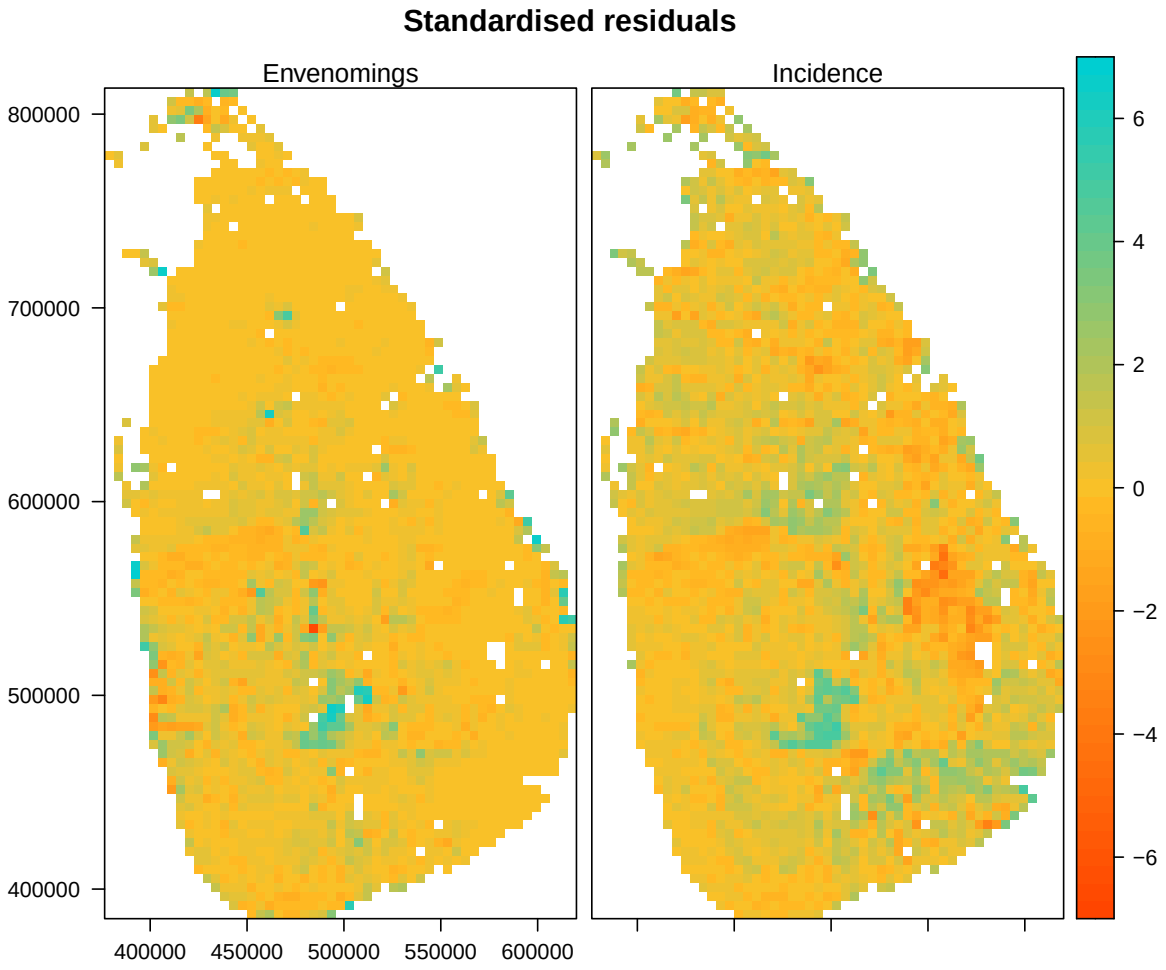
Finally here I assess how well the model complies with the assumed relationship between envenoming cases and the effect of the different snake species and human population density. To start with, the far left hand side plot shows the density of residuals. These appear to be centred at zero, and with tails at both sides, although the tail on the right hand side of the zero appears longer, which indicates that the model tends to predict more envenomings than there actually are. The centre plot shows that residuals tend to be larger at larger values of the response variable, and the largest departures from data occur at intermediate values of the data (between 5 and 150 envenoming cases). The departure of the model is as high as 10 standard deviations from the mean. This value, according to the far right hand side plot is as much as three times the target value of the data. The observed trend (blue dashed line) between x and y values in the second and third plots do not seem to depart too much from the reference horizontal red line.



Residuals in space

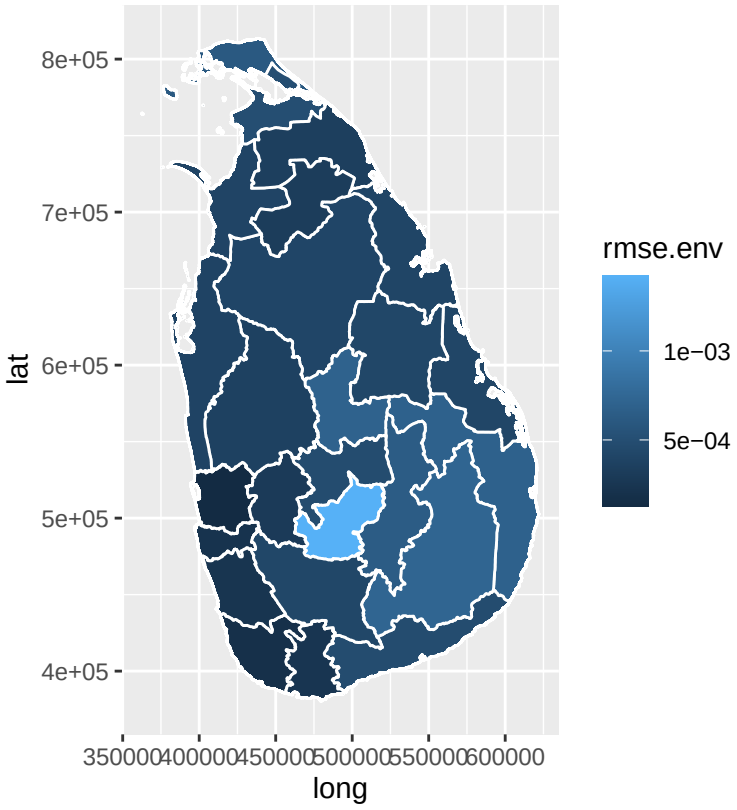
In the approach I took to model envenomings using all grid cells as data I assumed that all observations are independent. Upon analysing residuals it is possible to see that this is not true. In previous attempts I tried

to model the effect of space, but this was not possible due to the large number of data points. Regardless of the independence assumption violation here I show where the model is over or under estimating the number of envenoming cases and their incidence. First is the number of envenoming cases:



The absolute deviation between data and the model is further presented as root mean squared error in each Sri Lankan district. RMSE represents the average size of the error in the same units as the data, thus the map below shows the average incidence difference between model and data.

```
## OGR data source with driver: ESRI Shapefile
## Source: "/media/gerardo/Almacen/SB-zoonotic-transmission/Data/Sri Lanka boundaries/LKA_adm1.shp", la
## with 25 features
## It has 14 fields
## Integer64 fields read as strings:  ID_0 ID_1 CCN_1
```


8e+05
7e+05
6e+05
5e+05
4e+05
lat
350000
400000
450000
500000
550000
600000
long
rmse.env
1e−03
5e−04