

Snakebite incidence: Simple mass action model

Gerardo Martin

7th June 2019

Intro and Methods

These are the results of the *simple mass action* model, in which snakebites result from:

$$\frac{dH_{bitten}}{dt} = \beta SH$$

and where

$$S = \sum_{i=1}^n A_i I_i S_i$$

That is, the sum of all snake species' i abundance weighted by the aggressiveness indices A_i and an estimated effect I_i for estimating other effects. H is human population density and β is the contact rate.

I estimated parameters by discretising the above model, to estimate the probability that snakebites occur during the time where data was collected:

$$H_{bitten,t+1} = H_{bitten,t} + H_t \times (1 - \exp(-\beta S_t))$$

Hence $H_{bitten,t+1} - H_{bitten,t}$ is the number of snakebites that occur during one year.

To include the effect of land cover on snakebite incidence, I used the most frequent land cover classes between 2004 and 2017, and estimated the parameter β of the model for each land cover class, and transformed it into a function of human population density to correct snake abundance in relation to humans. The actual final equation was:

$$\beta(L, H) = \exp(\beta_{0,l} + \beta_{1,l} \ln(H)^2)$$

Results

Model summary

Fixed effects

```
library(foreach)
fixed.effects <- lapply(bites.incidence, function(x1){x1[, 1:22]})

gel.diag <- gelman.diag(as.mcmc.list(fixed.effects))

sum.inc.pars <- foreach(i = 1:22, .combine = rbind) %do% {
  med <- median(c(fixed.effects[[i]][, i], fixed.effects[[2]][, i], fixed.effects[[3]][, i]))
}
```

```

std.dev <- sd(c(fixed.effects[[1]][, i], fixed.effects[[2]][, i], fixed.effects[[3]][, i]))
cr.int <- HPDinterval(as.mcmc(fixed.effects[[1]][, i], x1[[2]][, i], fixed.effects[[3]][, i]), 0.95)
return(c(med, std.dev, cr.int))
}

sum.inc.pars <- data.frame(sum.inc.pars)
names(sum.inc.pars) <- c("Median", "Std.Dev", "Cr.025", "Cr.975")
sum.inc.pars$Gelman.conv <- gel.diag$psrf[, 2]

rownames(sum.inc.pars) <- colnames(fixed.effects[[1]])
knitr::kable(sum.inc.pars)

```

	Median	Std.Dev	Cr.025	Cr.975	Gelman.conv
beta0[1]	-12.9193486	0.1144861	-13.1473606	-12.6944212	1.0478827
beta0[2]	-12.8727391	0.1150190	-13.0888539	-12.6293909	1.0339168
beta0[3]	-13.1187598	0.1265627	-13.3506796	-12.8599359	1.0227626
beta0[4]	-11.8498334	0.1597204	-12.1597111	-11.5319509	1.0151924
beta0[5]	-11.3228354	0.2015742	-11.7203119	-10.9286790	1.0273697
beta1[1]	-0.0039378	0.0006796	-0.0051861	-0.0024929	1.0017653
beta1[2]	-0.0057035	0.0006668	-0.0071653	-0.0044671	1.0006342
beta1[3]	-0.0014653	0.0011869	-0.0035608	0.0010914	1.0011155
beta1[4]	-0.0186738	0.0012897	-0.0210802	-0.0161038	1.0025519
beta1[5]	-0.0274362	0.0018818	-0.0318450	-0.0240987	1.0069996
indices[1]	4.3438751	0.4823422	3.3741690	5.3637670	1.0456885
indices[2]	5.5887558	2.2654653	1.6210204	9.9441387	1.0060186
indices[3]	1.7274542	0.3912552	1.0992237	2.6872885	1.0388849
indices[4]	2.8139337	0.9783767	0.8829060	4.5733175	1.0078482
indices[5]	0.0478502	0.0152536	0.0197281	0.0791045	1.0010196
indices[6]	3.5169562	0.6952306	2.2671238	4.9612017	1.0158608
indices[7]	0.1456995	0.1138336	0.0000129	0.3632991	1.0073032
r[1]	34.4774381	2.3383585	30.1302354	38.9411160	1.0005425
r[2]	23.5624351	0.9138550	21.8383916	25.4058884	1.0038691
r[3]	13.2573368	0.9327915	11.4339173	15.0615922	0.9999053
r[4]	9.3962322	0.5619703	8.3477050	10.5265641	1.0008068
r[5]	18.8545727	7.0247510	7.8606017	33.3518514	1.0023815

Random effects

```

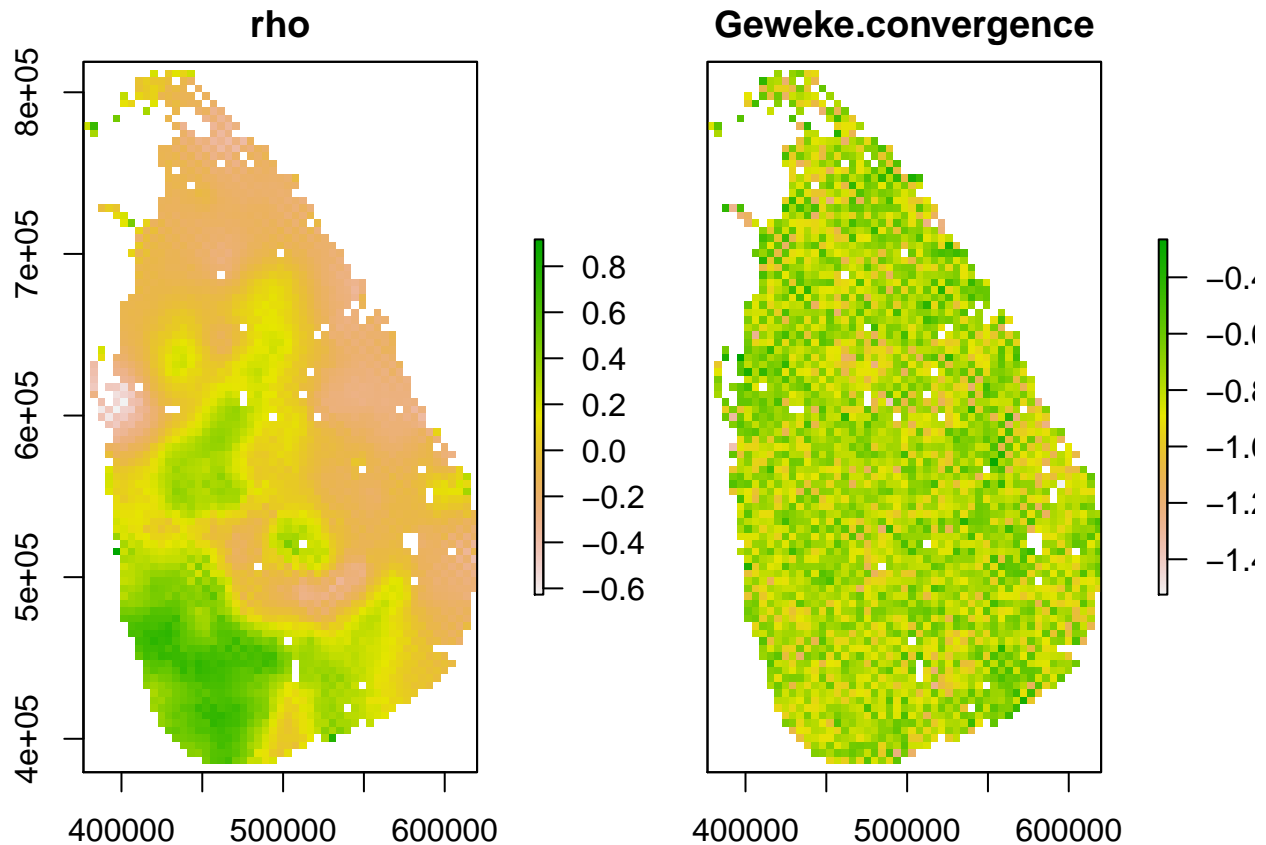
random.effects <- lapply(bites.incidence, function(x1){x1[, 23:3079]})
diag.rho <- geweke.diag(random.effects)

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 <- rasterFromXYZ(data.frame(incid.data[, c("x", "y")], diag.rho$z))
rho.stack <- stack(rho.r, rho.diag)
names(rho.stack) <- c("rho", "Geweke convergence")

plot(rho.stack)

```



Model predictions

```
x.full <- rbind(fixed.effects[[1]], fixed.effects[[2]], fixed.effects[[3]])

x.beta0 <- x.full[, paste0("beta0[", 1:5, "]")]
x.beta1 <- x.full[, paste0("beta1[", 1:5, "]")]
x.indices <- x.full[, paste0("indices[", 1:7, "]")]
x.rho <- rho <- rbind(random.effects[[1]], random.effects[[2]], random.effects[[3]])

s.preds <- apply(x.indices, 1, function(x){
  rowSums(sweep(S, MARGIN = 2, x, FUN = "*"))
})

library(foreach)
hu.ef <- foreach(i = 1:3000, .combine = cbind) %do%{
  exp(x.beta0[i, incid.data$land.cover] + x.beta1[i, incid.data$land.cover] * Hum.log^2)
}

inc.preds <- (1 - exp(-s.preds * hu.ef)) * exp(t(x.rho))

inc.sum <- data.frame(t(apply(inc.preds, 1, function(x){HPDinterval(as.mcmc(x), prob = 0.5)})))
names(inc.sum) <- c("id", "incid")
```

The result of running the function is the number of snakebite cases (H.bit) predicted by the model.

From the JAGS object and the function I transformed the median of the posterior estimates into a raster object. Then to see the spatial pattern of incidence I divided the median of posteriors by the human population

density.

First here is the total number of snakebites estimated by the model (left) and the number of snakebites used as data (right). The colourscale of both maps is the same.

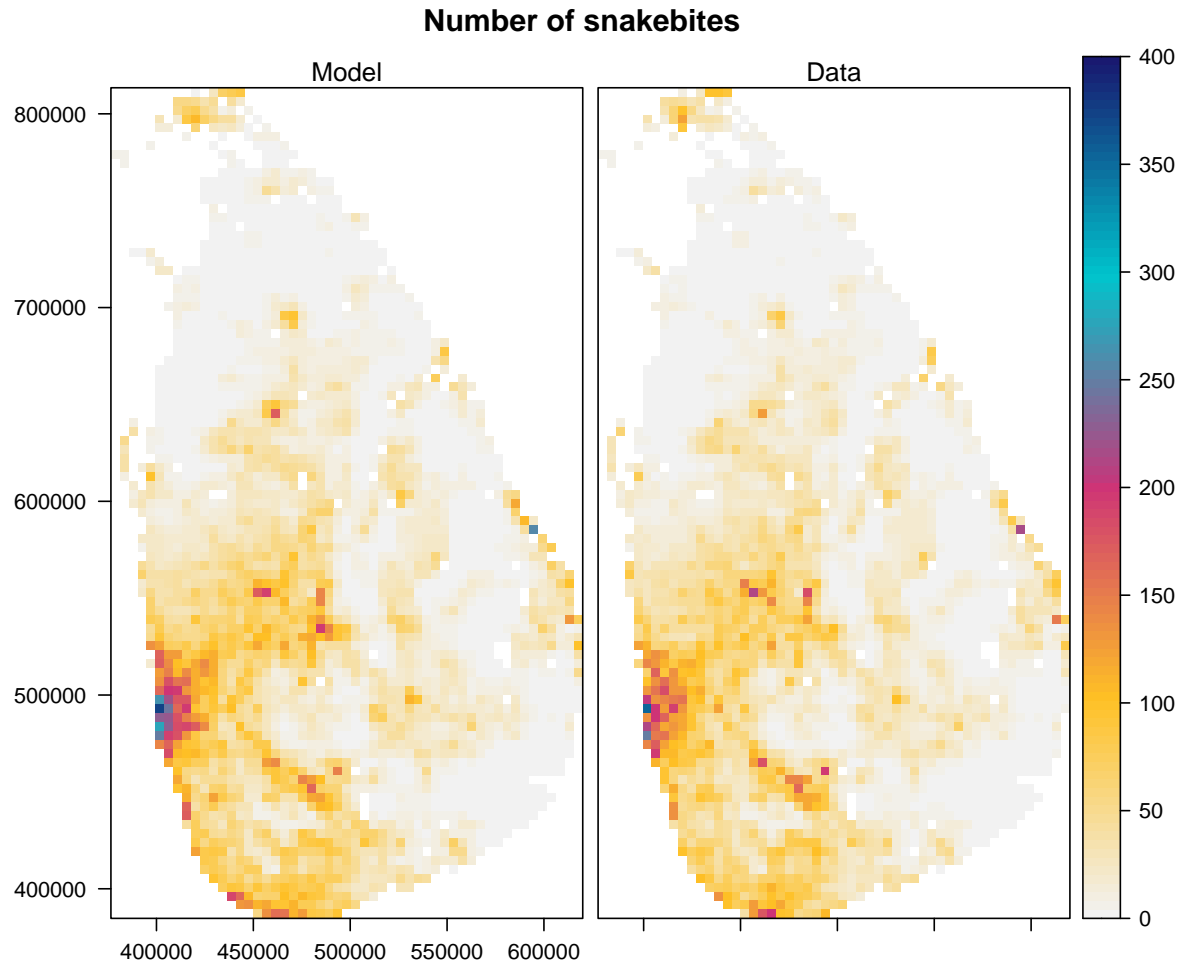


Figure 1: Number of snakebites predicted by the model and snakebites used as data to fill the model.

Then the pattern of snakebite incidence after dividing the above layers by human population density is:

There are obvious differences between the model predictions and data, but below is a more formal assessment of the discrepancies between model and data. I first did a correlation test between the number of bites of the model and the data, and between incidence prediction and observed.

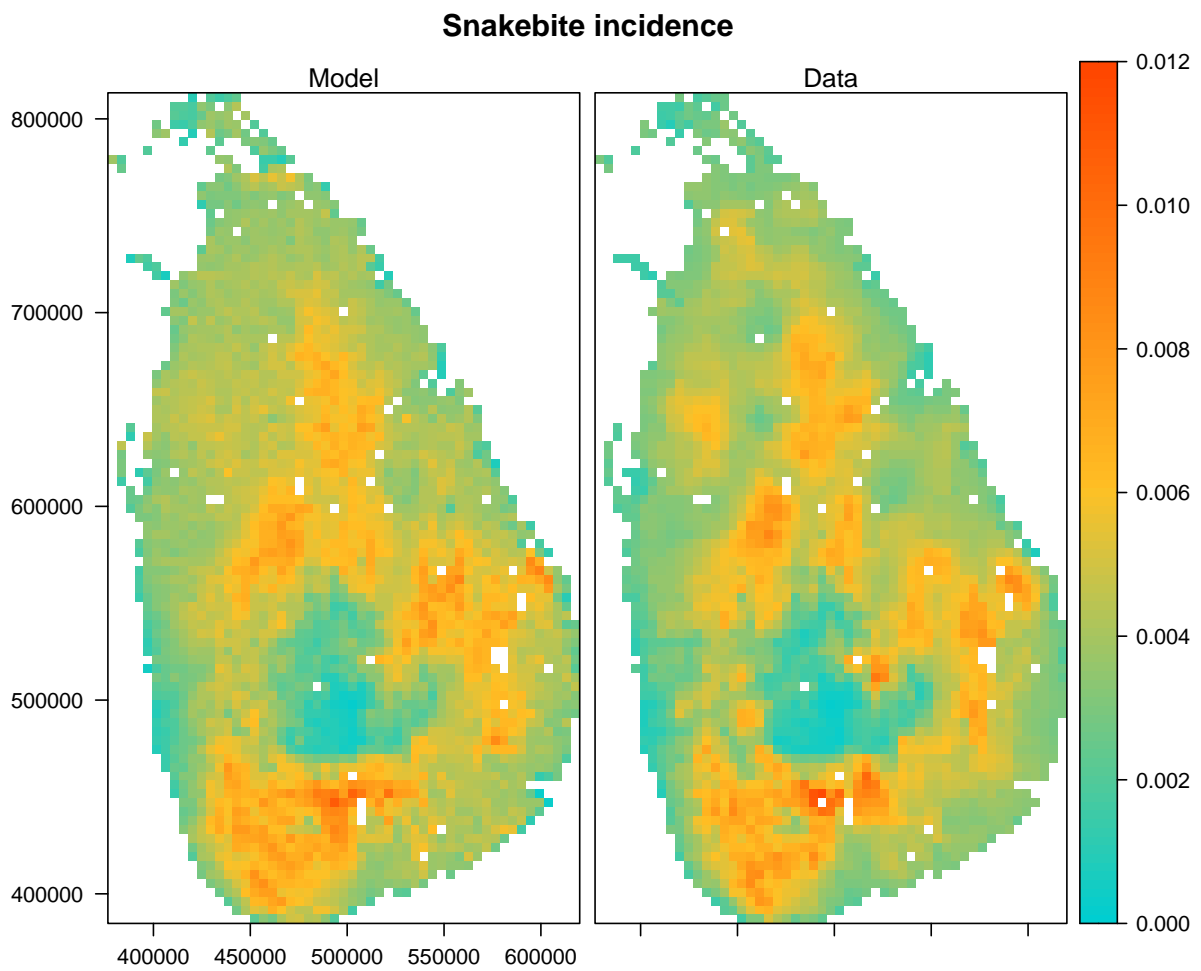
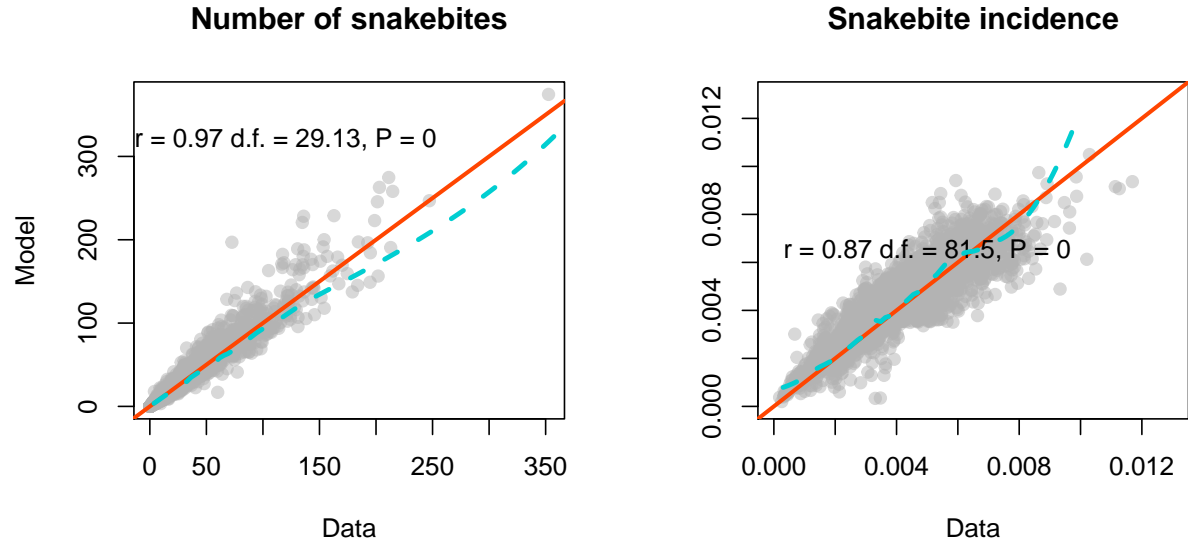


Figure 2: Snakebite incidence patterns predicted by the model and snakebite incidence used as data to fit the model.



The orange line represents the perfect regression line (intercept zero and a slope one). The dashed green line is the scatter plot smoothing with a scale (smoothing) parameter of 0.1. The numbers in each plot are the estimated correlation coefficient and the significance adjusted for spatial autocorrelation.

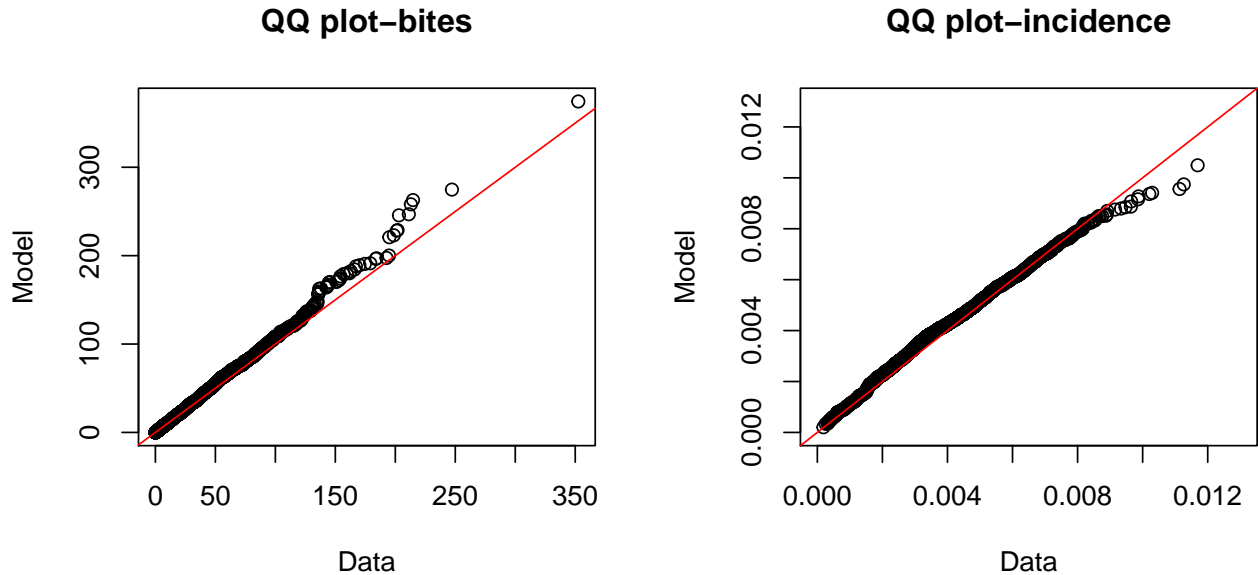


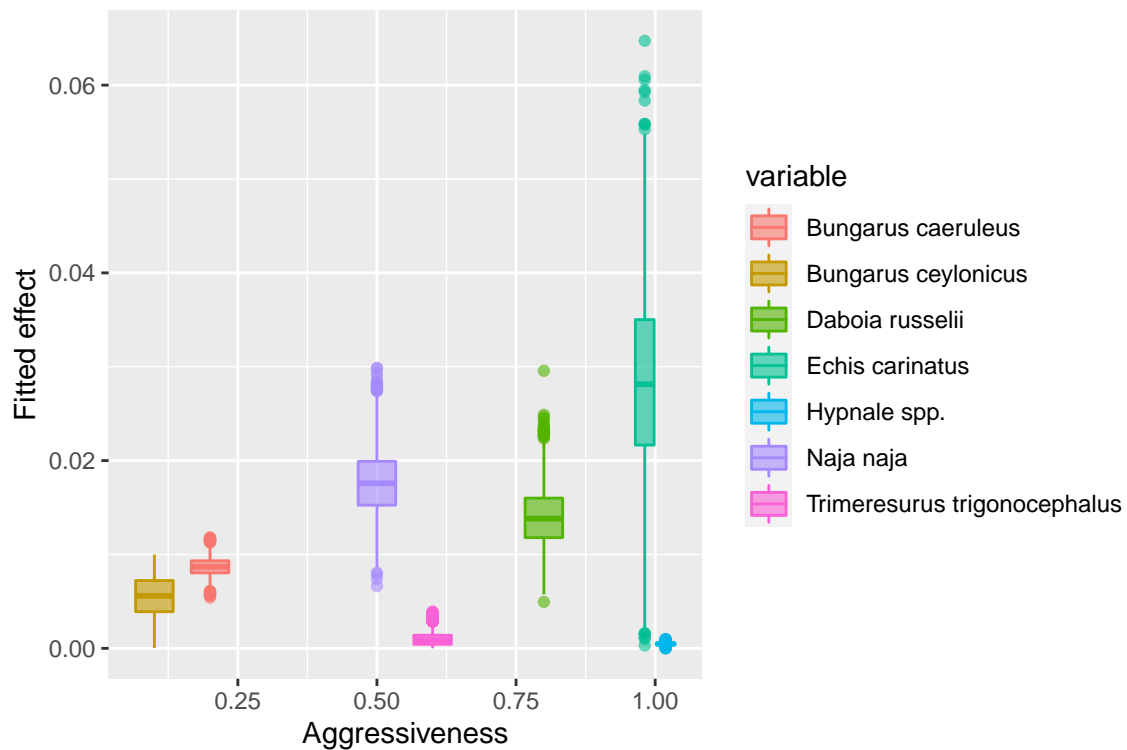
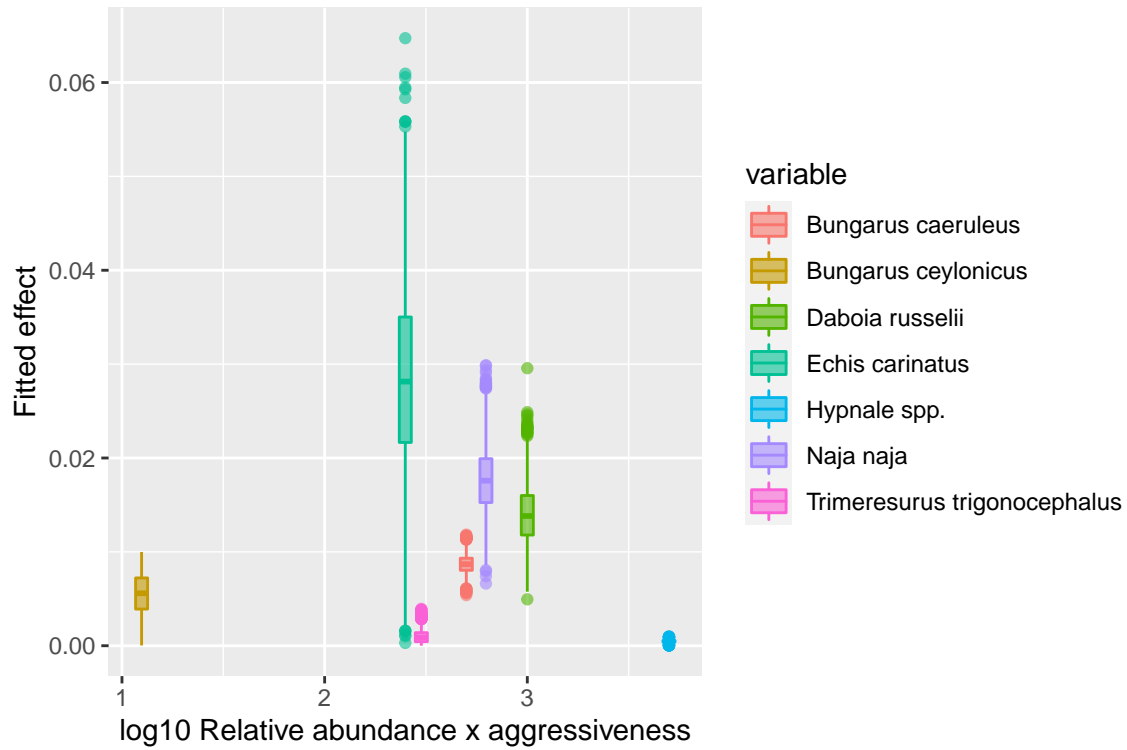
Figure 3: Comparison of the statistical distribution of the number of snakebites and snakebite incidence with the data. Adherence of quantiles to a straight line (ref reference line), indicates similar distributions.

These two plots are a comparison of the distribution of both data, that generated by the model, and the source data used to fit the model. Given that both, snakebites and incidence are close to a straight lines in both plots, the model generated a distribution that is very close to that of the source data.

Relationship of indices estimated indices and expert-derived information

This is the comparison of the fitted effects and the aggressiveness and relative abundance adjustments used previously to explain snakebites.

```
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:latticeExtra':
##
## layer
```



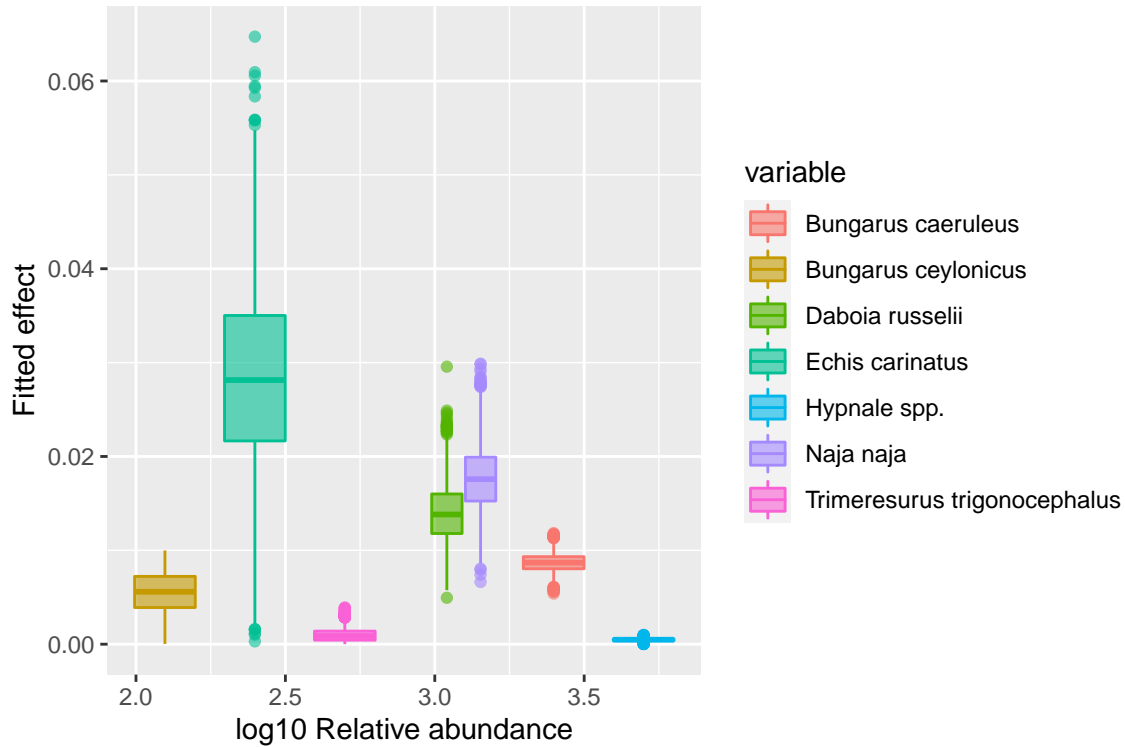
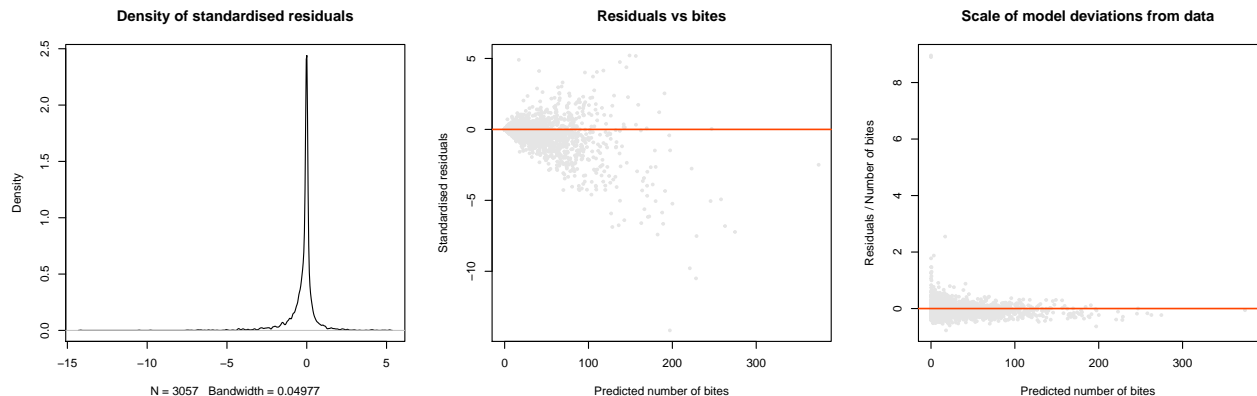


Figure 4: Relationship between relative abundance and aggressiveness weightings, aggressiveness only, and relative abundance-only weightings with estimated contact rates for each snake species.

Analysis of residuals

As with all regression exercises, residuals should be normally distributed. Here are a series of diagnostic plots.



The first one shows the density of residuals, which was obtained by subtracting the observed number of snakebites from the median of model posterior estimates. Then in the second I see whether there is some non-linearity between residuals and predicted values. As with correlation plots, the orange line represents the gold standard and the green dashed line the actual trend. For the most part residuals follow a linear pattern with predicted values, but residuals increase towards largest predicted values. This is in agreement with the longer tail on the left side of the density plot (far left). The third plot, shows the predicted values against the residuals divided by the observed number of bites. This represents the number of times that model predictions are greater or smaller than they should be. As evidenced by the green dashed line, the model tends to overestimate at smaller number of snakebites.

Residuals in space:

The first assumption in a regression procedure is that all data are independent. With spatial data this assumption is easily violated. Thus here I plotted the standardised residuals in space to see whether there is any clustering or the presence of other factors that we did not account for.

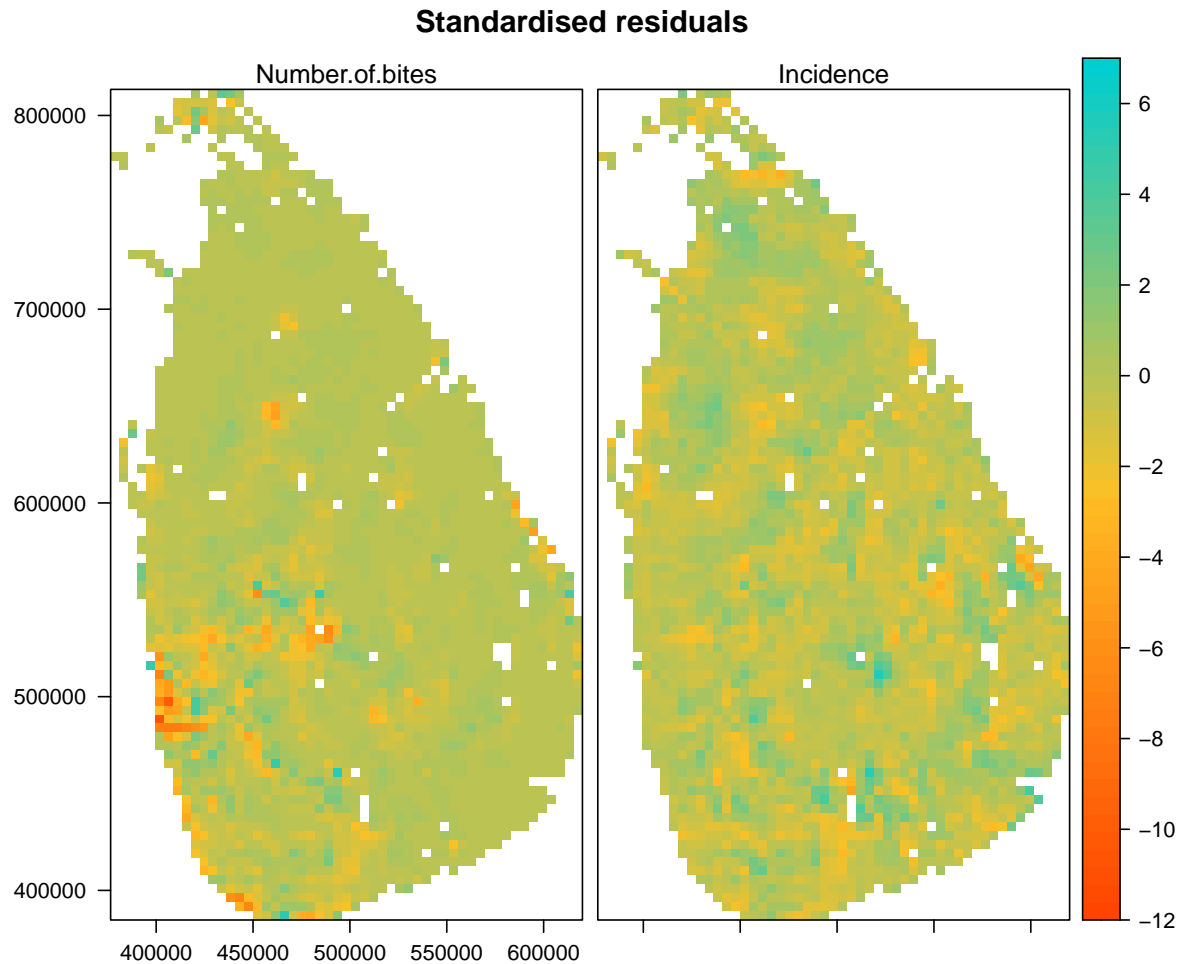
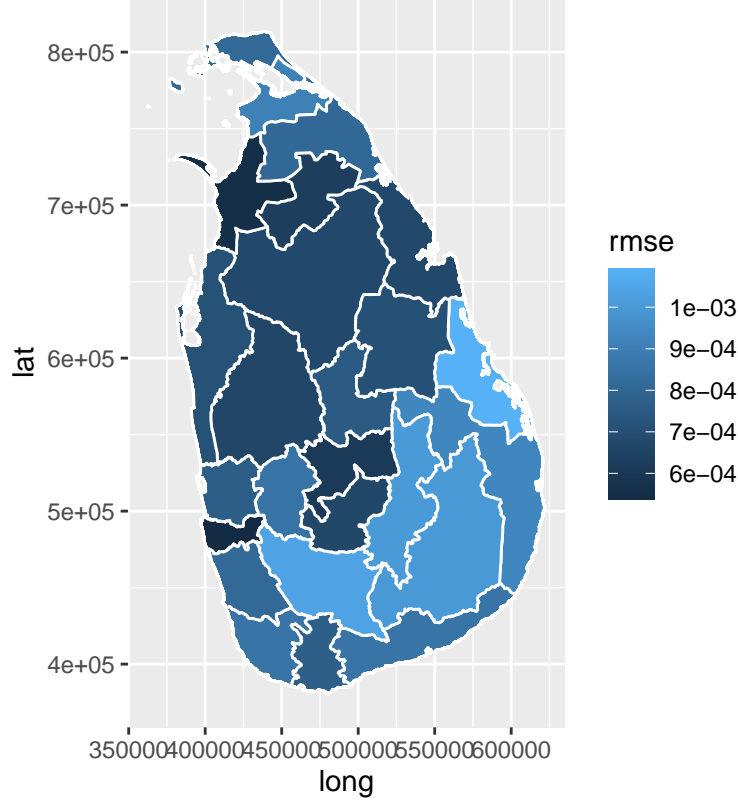


Figure 5: Spatial patterns of residuals for the number of snakebites and incidence.

The left side shows the standardised residuals of snakebite numbers, and the residuals of incidence on the right hand side. In both cases residuals are close to zero, but depart from the mean by as much as 12 standard deviations, and tend to be clustered. The predominantly negative values mean that the model underestimates both the number of snakebites and incidence, especially in areas where incidence is high.

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", layer:
## with 25 features
## It has 14 fields
## Integer64 fields read as strings: ID_0 ID_1 CCN_1
```



Responses fitted by the model

The model fitted is capable of reproducing non-linear relationships. Given that the best-performing model kept land cover as an important variable, these are the responses to both human population density and snake potential abundance in each of the land cover classes, across all values of human population density and snake potential abundance that occur in each class.

Previous analyses suggested that there is a negative effect of human population on incidence. However the number of snakebite cases increases with human population. The relationship with snakes however is non-linear. The model suggests that in certain cases there could be either less cases than expected by the potential number of snakes, or that there are less snakes than indicated by models:

When the number of cases is transformed to incidence, the model is in agreement with analyses that suggest that there is a negative effect of humans on snakebite incidence. Given the parameter combinations in each of the land cover classes, there are cases where snakes might correlate negatively with snakebite incidence (notice the negative slopes in response to snake abundance in some of the land cover classes):

The existence of the negative effect of snakes on incidence will be confirmed in the later sections of these analyses.

Contact rate and change of snake density across Sri Lanka

The contact rate β is a function of human density and two other parameters β_0 and β_1 :

$$\beta(\ln(H_{total}), l) = \exp(\beta_{0,l} + \beta_{1,l} \ln(H_{total})^2)$$

Where l are the five land cover classes. I used the function above with the parameter estimates to represent the contact rate in space. The far left and right panels represent the 2.5 and 97.5% credible intervals, whereas the central one is the median of posterior estimates.

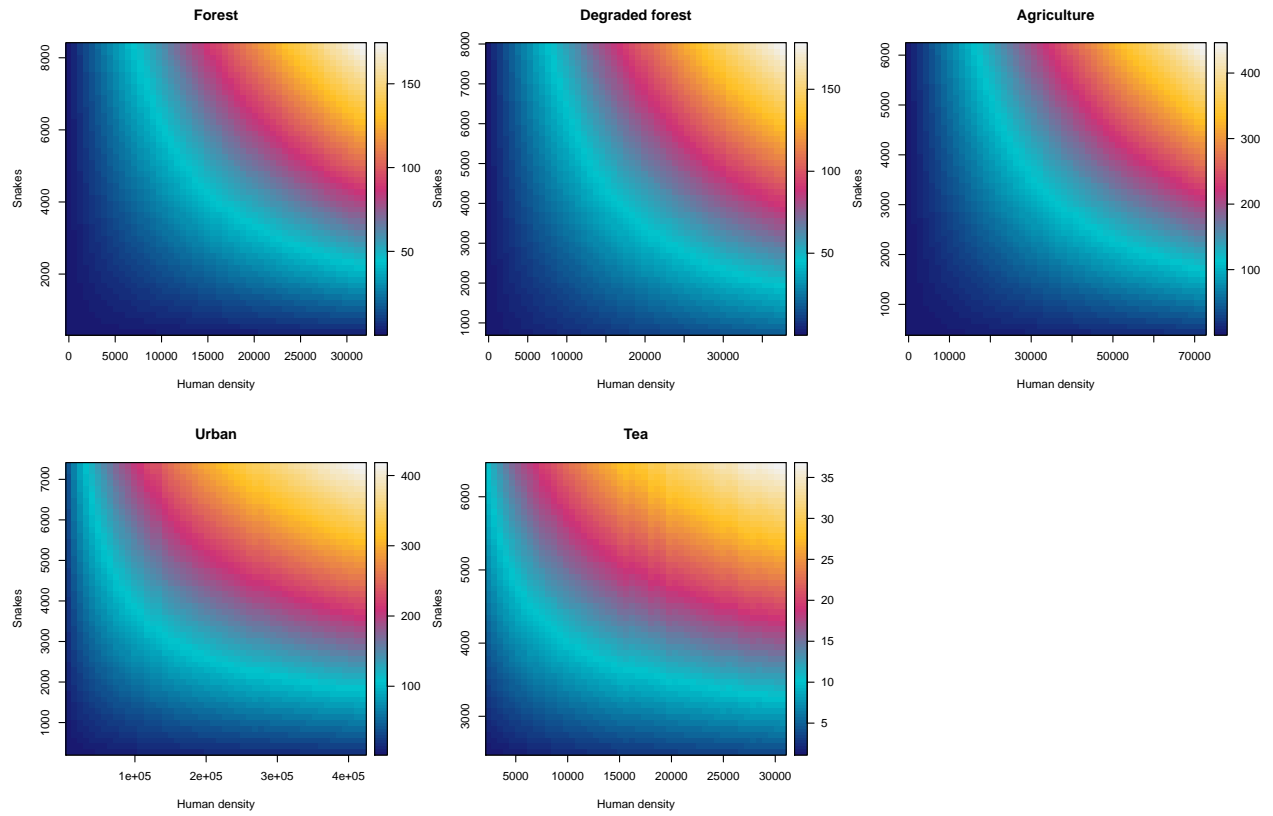


Figure 6: Partial responses of snakebites to humans and snakes.

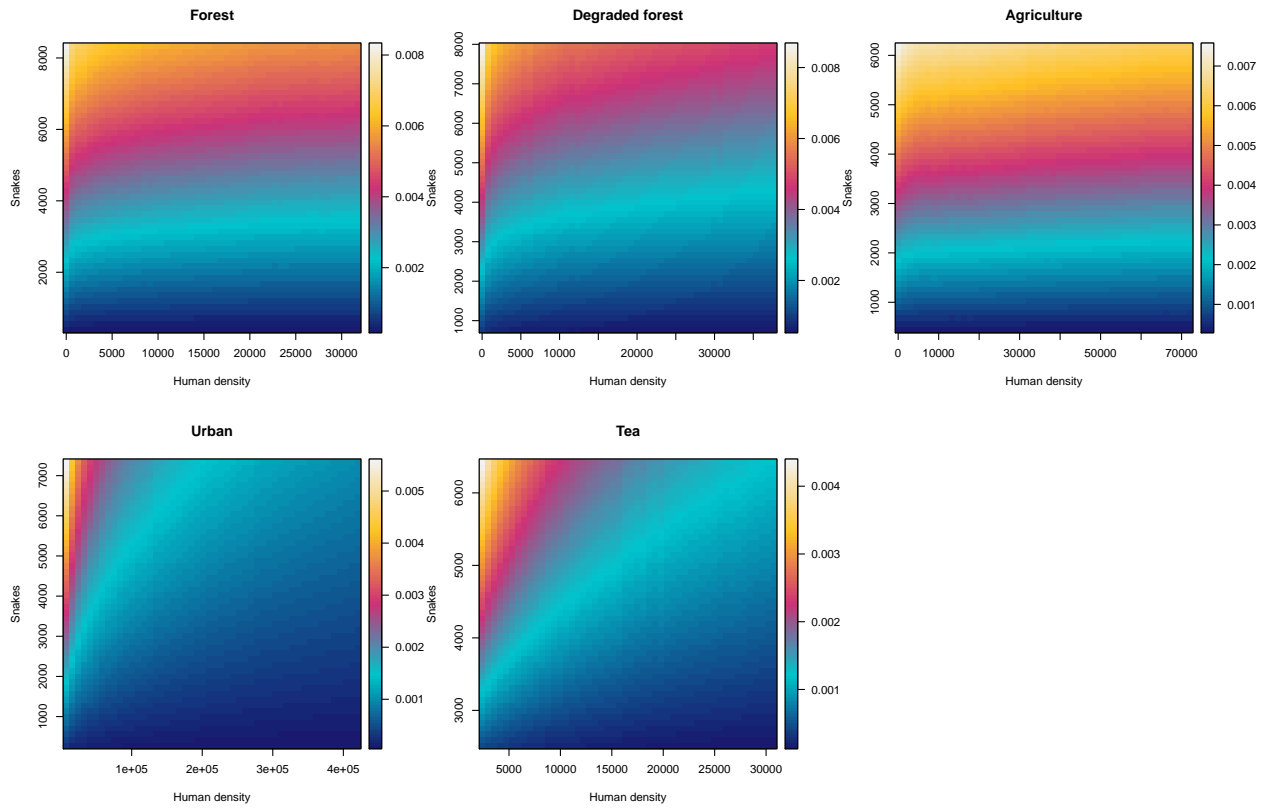


Figure 7: Partial response of snakebite incidence to snakes and humans per land cover class

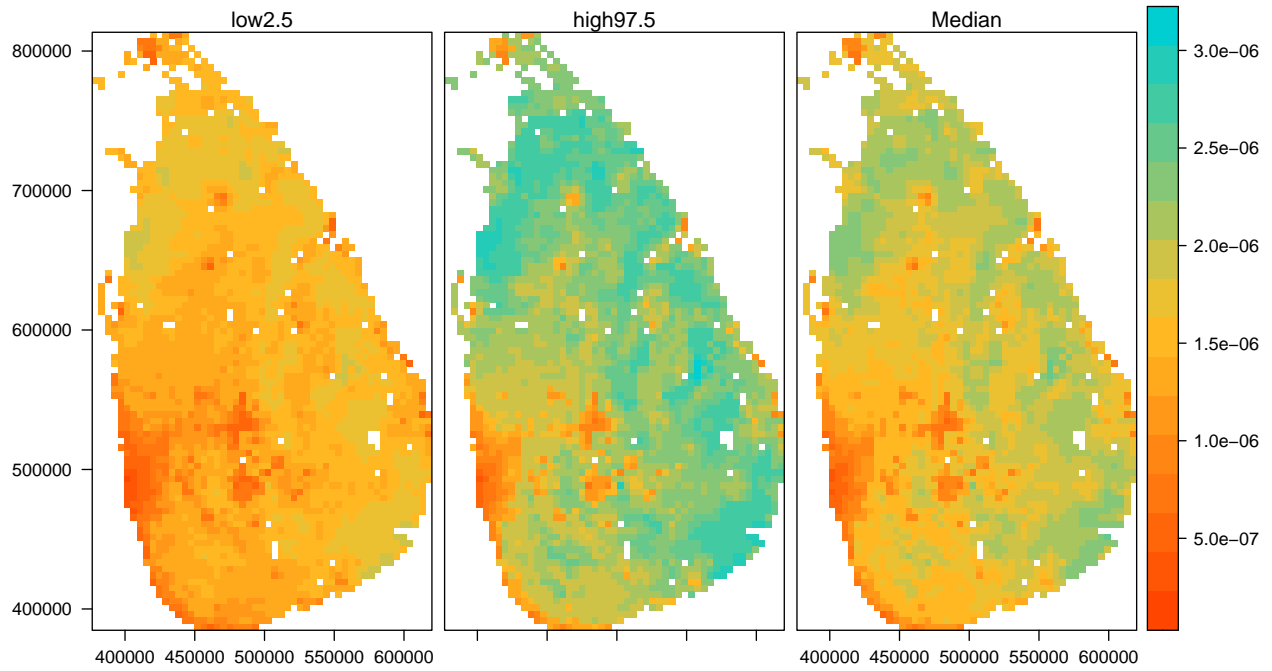


Figure 8: Spatial variability of the effect of humans on snake abundance (credible intervals and median)

To visually assess the individual responses among land cover classes here are different values that the contact rate can take given the range of human population densities observed across Sri Lanka in each of the land cover classes, in two different scales. First with human population density in log-scale:

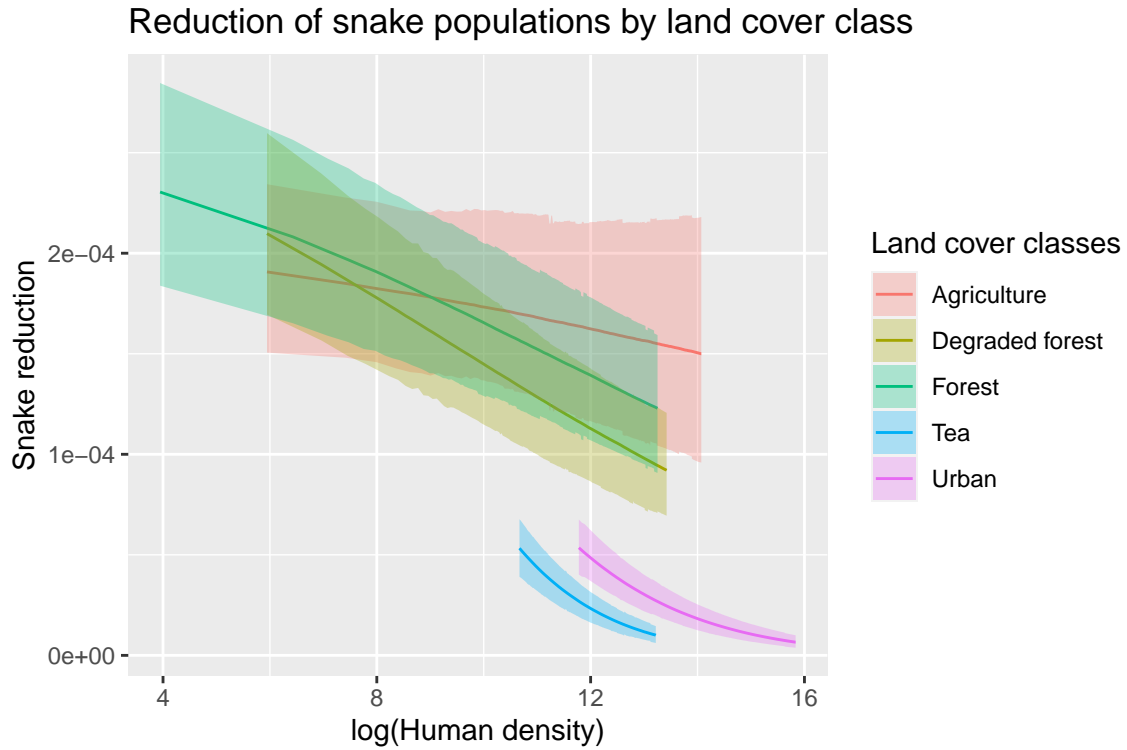


Figure 9: Effect of humans on snake abundance.

These responses and credible intervals are the result of the β_0 parameter:

And β_1 estimates:

It appears that there are two (possibly three) different effects of land cover on the contact rate. On one side, Forest and Degraded forest are largely similar, and greatly differ from Urban and Tea cover. Agricultural land is relatively similar to Forest and Degraded forest but results in the highest contact rates within the range of human population densities in which it occurs.

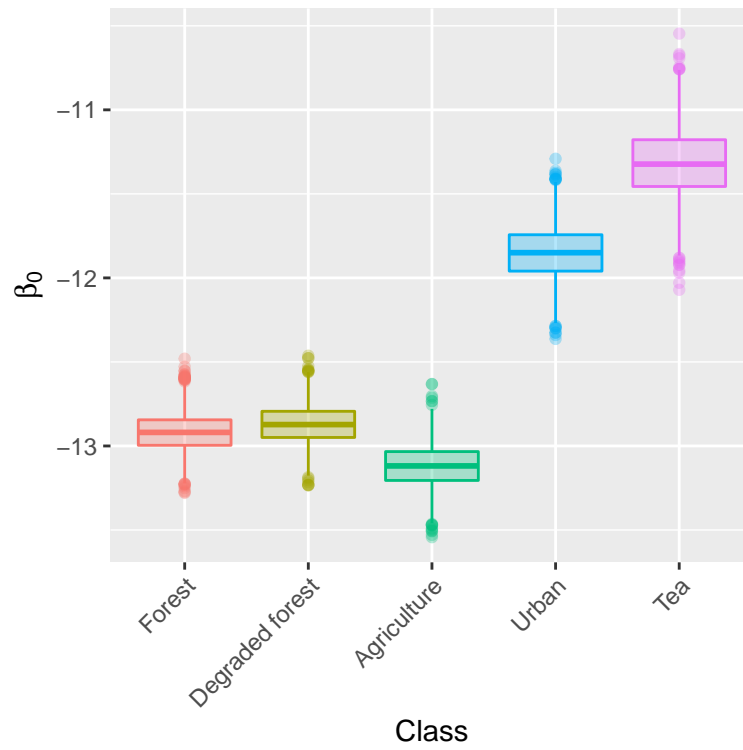


Figure 10: Distribution of posterior samples for beta 0 parameters of the function to adjust snake abundance in relation to humans.

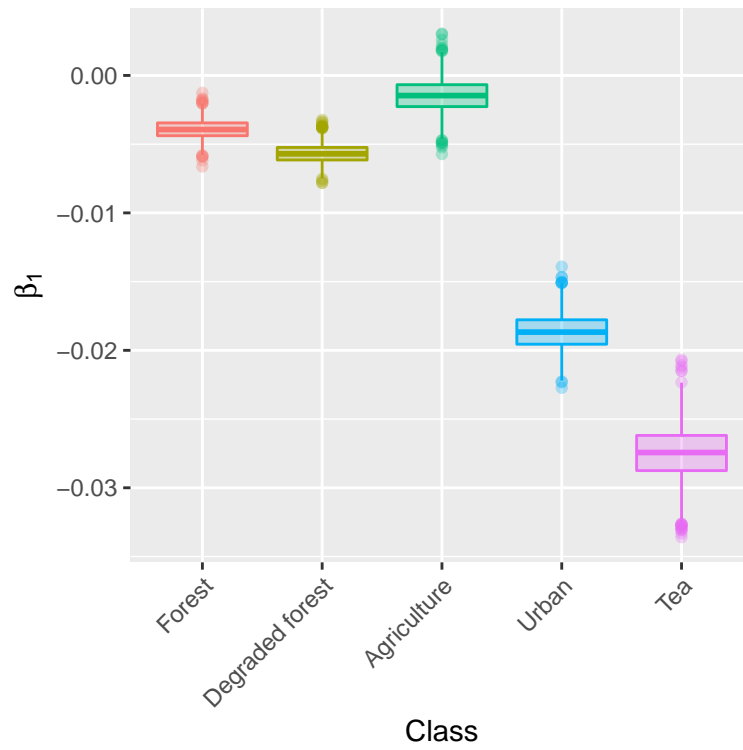


Figure 11: Distribution of posterior samples for beta 1 parameters of the function to adjust snake abundance in relation to humans.