Snakebite incidence: refuge effect model

Gerardo Martin

7th June 2019

Intro and Methods

These are the results of the *refuge effect* model, in which snakebites result from:

$$\frac{dH_{bitten}}{dt} = \beta S \left(H - \frac{S}{q_h} \right)$$

Where

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

S is the sum of all snake species' abundance weighted by aggressiveness A_i and an estimated effect I_i to accoutn for other species' effects. H is human population density per grid cell and β and q_h are the contact rate and proportion of snakes exposed to humans.

In the equation above, intead of H I used the $\ln H$ because the human density data is very biased towards zero and a few very large values. Using the log-converted human density layer made MCMC sampling and parameter inference possible.

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

$$\frac{\Delta H_{bitten}}{\Delta t} = H \times \left[1 - \exp \left(-\beta S \left(1 - \frac{S}{q_h H} \right) \right) \right]$$

And given that $\Delta t = 1$ year, ΔH_{bitten} is the number of snakebites that occur during one year.

Parameter β, was transformed into a function of human population density and land cover:

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

Then by multiplying P by the total human population density we obtain the number of bites:

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

To include land cover in the analyses, I used the most frequent land cover classes between 2004 and 2017, to estimate the parameters of the above model in each land cover class. So that we'd be able to find if there are different responses to human population density and snake abundance in the different land cover classes.

The JAGS script used for inferring parameters of this model is:

```
## JAGS model:
##
## model{
##        #Likelihood
##        for(i in 1:n){
##                Beta[i] <- exp(beta0[class[i]] + beta1[class[i]] * Hum.log[i]^2)
##                S.sum[i] <- inprod(indices[], S[i,])
##                P[i] <- 1 - exp( - (Beta[i] * S.sum[i]  -  Beta[i] * S.sum[i]^2/(qh[class[i]] * Hum.log[
##                P.thres[i] <- ifelse(P[i] < 0, 0, P[i])
##                H.bit[i] <- P.thres[i] * Hum.tot[i]
##                P.r[i] <- r[class[i]]/(r[class[i]] + H.bit[i])
##                Hum.bit[i] ~ dnegbin(P.r[i], r[class[i]])
##        }
##        #Priors
##        for(j in 1:n.classes){
##                beta0[j] ~ dnorm(0, 1.0E-3)
##                beta1[j] ~ dnorm(0, 1.0E-3)
##                qh[j] ~ dnorm(qh.min, 1.0E-3)T(qh.min,)
##                r[j] ~ dunif(0, 100)
##        }
##        for(k in 1:n.species){
##                indices[k] ~ dunif(0,10)
##        }
## }
## Fully observed variables:
##  Hum.bit Hum.log Hum.tot S class n n.classes n.species qh.min
```

Results

Model summary

Parameter estimates of JAGS run. Although results are not shown here, all parameters converged according
to the Gelman convergence diagnostic test.

```
## Inference for Bugs model at "~/Documentos/Protection-model.txt", fit using jags,
##  2 chains, each with 5e+05 iterations (first 50000 discarded), n.thin = 450
##  n.sims = 2000 iterations saved
##             mu.vect sd.vect    2.5%     25%     50%     75%    97.5%
## beta0[1]     -2.961   0.143   -3.232  -3.056  -2.965  -2.876  -2.667
## beta0[2]     -2.777   0.133   -3.017  -2.865  -2.788  -2.696  -2.496
## beta0[3]     -3.134   0.144   -3.397  -3.231  -3.143  -3.041  -2.835
## beta0[4]     -1.448   0.188   -1.803  -1.579  -1.453  -1.330  -1.054
## beta0[5]     -1.788   0.568   -2.834  -2.163  -1.819  -1.422  -0.621
## beta1[1]     -0.008   0.001   -0.009  -0.008  -0.008  -0.007  -0.006
## beta1[2]     -0.009   0.001   -0.010  -0.009  -0.009  -0.008  -0.007
## beta1[3]     -0.005   0.001   -0.008  -0.006  -0.005  -0.004  -0.002
## beta1[4]     -0.023   0.001   -0.025  -0.023  -0.023  -0.022  -0.020
## beta1[5]     -0.028   0.006   -0.041  -0.032  -0.028  -0.024  -0.017
## indices[1]    2.675   0.273    2.052   2.508   2.709   2.856   3.149
## indices[2]    2.851   2.345    0.092   0.939   2.247   4.137   8.811
## indices[3]    0.756   0.285    0.239   0.557   0.739   0.933   1.390
## indices[4]    1.827   0.825    0.427   1.235   1.768   2.334   3.625
## indices[5]    0.007   0.006    0.000   0.003   0.006   0.010   0.023
## indices[6]    9.175   0.746    7.197   8.799   9.389   9.745   9.974
```

```
## indices[7]      0.118   0.106    0.004    0.037    0.090    0.171    0.393
## qh[1]           0.111   0.021    0.079    0.097    0.107    0.122    0.162
## qh[2]           0.087   0.010    0.067    0.082    0.088    0.093    0.106
## qh[3]           0.098   0.009    0.077    0.093    0.099    0.104    0.113
## qh[4]           0.058   0.006    0.045    0.054    0.058    0.062    0.069
## qh[5]          25.735  19.546    1.406   10.430   21.796   36.862   72.780
## r[1]           28.249   3.787   21.784   25.618   27.976   30.547   36.233
## r[2]           21.035   1.485   18.319   19.970   20.934   22.034   24.098
## r[3]           13.934   1.960   10.611   12.592   13.687   15.149   18.162
## r[4]           11.131   1.122    9.057   10.367   11.078   11.839   13.413
## r[5]           15.266   6.506    7.074   10.992   13.824   17.867   32.069
## deviance    17036.791   7.217 17024.884 17031.576 17036.054 17041.298 17052.792
##               Rhat n.eff
## beta0[1]     1.001  2000
## beta0[2]     1.005   340
## beta0[3]     1.003   510
## beta0[4]     1.002   820
## beta0[5]     1.001  2000
## beta1[1]     1.001  2000
## beta1[2]     1.004   460
## beta1[3]     1.001  2000
## beta1[4]     1.001  2000
## beta1[5]     1.001  2000
## indices[1]   1.003   590
## indices[2]   1.002  1100
## indices[3]   1.017   330
## indices[4]   1.003  2000
## indices[5]   1.001  2000
## indices[6]   1.002  1800
## indices[7]   1.001  2000
## qh[1]        1.002  2000
## qh[2]        1.002   810
## qh[3]        1.002  1200
## qh[4]        1.004   460
## qh[5]        1.002  1200
## r[1]         1.001  2000
## r[2]         1.001  2000
## r[3]         1.001  2000
## r[4]         1.002   770
## r[5]         1.001  2000
## deviance     1.001  2000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 26.0 and DIC = 17062.8
## DIC is an estimate of expected predictive error (lower deviance is better).

## Potential scale reduction factors:
##
##            Point est. Upper C.I.
## beta0[1]        1.003       1.01
## beta0[2]        1.008       1.04
```

```
## beta0[3]        1.005      1.02
## beta0[4]        1.001      1.00
## beta0[5]        1.002      1.01
## beta1[1]        1.001      1.00
## beta1[2]        1.005      1.03
## beta1[3]        1.000      1.00
## beta1[4]        1.000      1.00
## beta1[5]        1.002      1.00
## deviance        1.001      1.00
## indices[1]      1.005      1.02
## indices[2]      1.003      1.02
## indices[3]      1.004      1.02
## indices[4]      1.000      1.00
## indices[5]      1.003      1.00
## indices[6]      1.003      1.01
## indices[7]      1.002      1.00
## qh[1]           1.000      1.00
## qh[2]           1.003      1.02
## qh[3]           1.002      1.01
## qh[4]           1.006      1.03
## qh[5]           1.006      1.03
## r[1]            0.999      1.00
## r[2]            0.999      1.00
## r[3]            0.999      1.00
## r[4]            1.000      1.00
## r[5]            1.008      1.01
##
## Multivariate psrf
##
## 1.02
```

Model predictions

The `R` function to predict the JAGS script above is:

```
incid.fun <- function(H.log, H.tot, Species, beta0, beta1, qh, indices){
    S <- apply(indices, 1, function(x){sum(x * Species)})
    Beta <- exp(beta0 + beta1 * H.log^2)
    P <-1 - exp( - (Beta * S -Beta * S^(2)/(qh * H.log)))
    P.thres <- ifelse(P < 0, 0, P)
    H.bit <- P.thres * H.tot
    return(H.bit)
}
```

Which uses the indices of snakes abundance weighted by their aggressivenenss in advance. The result of running the function is the number of snakebite cases (`H.bit`) predcited 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.

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

There are some obvious differences between the model predictions and data, but below is a mroe formal
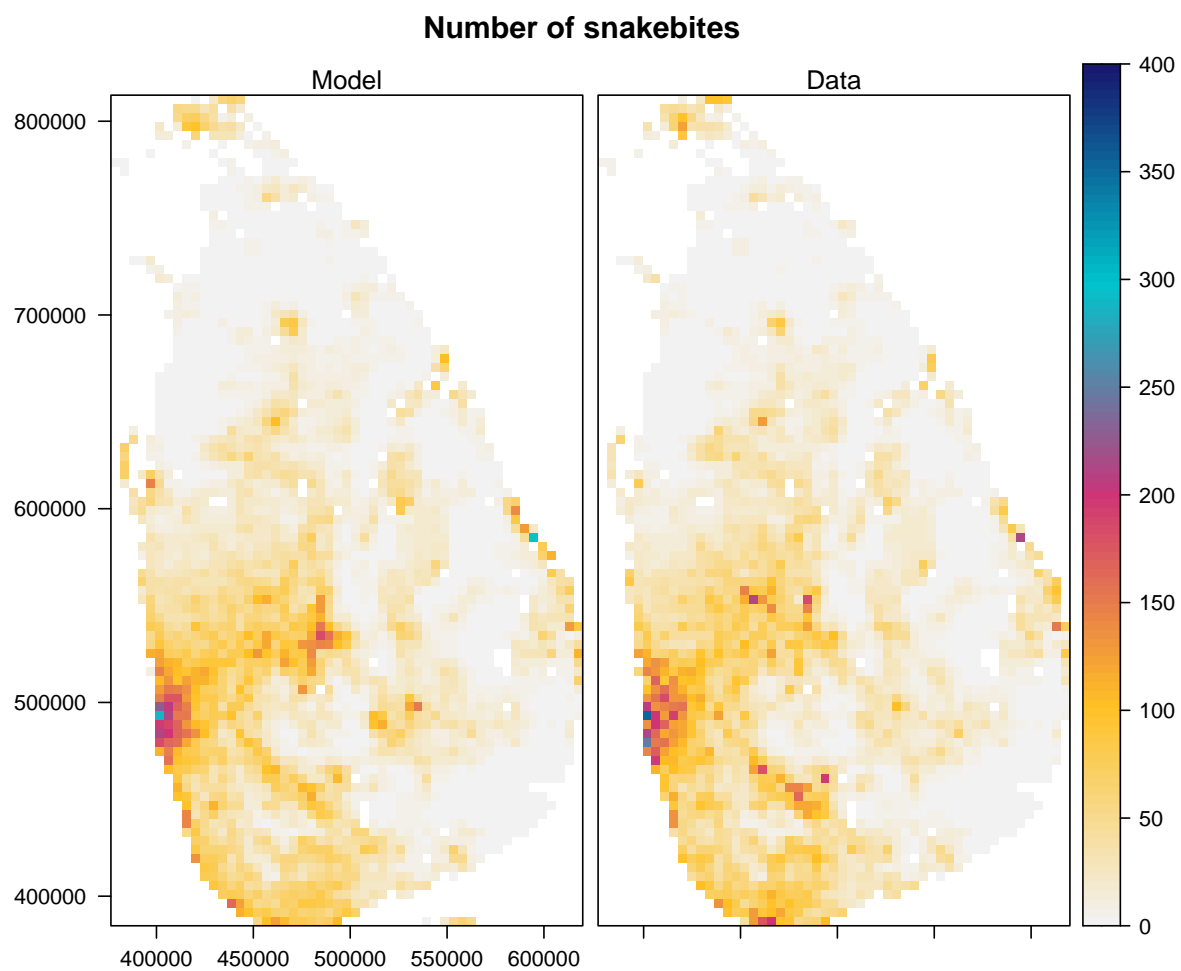
Number of snakebites

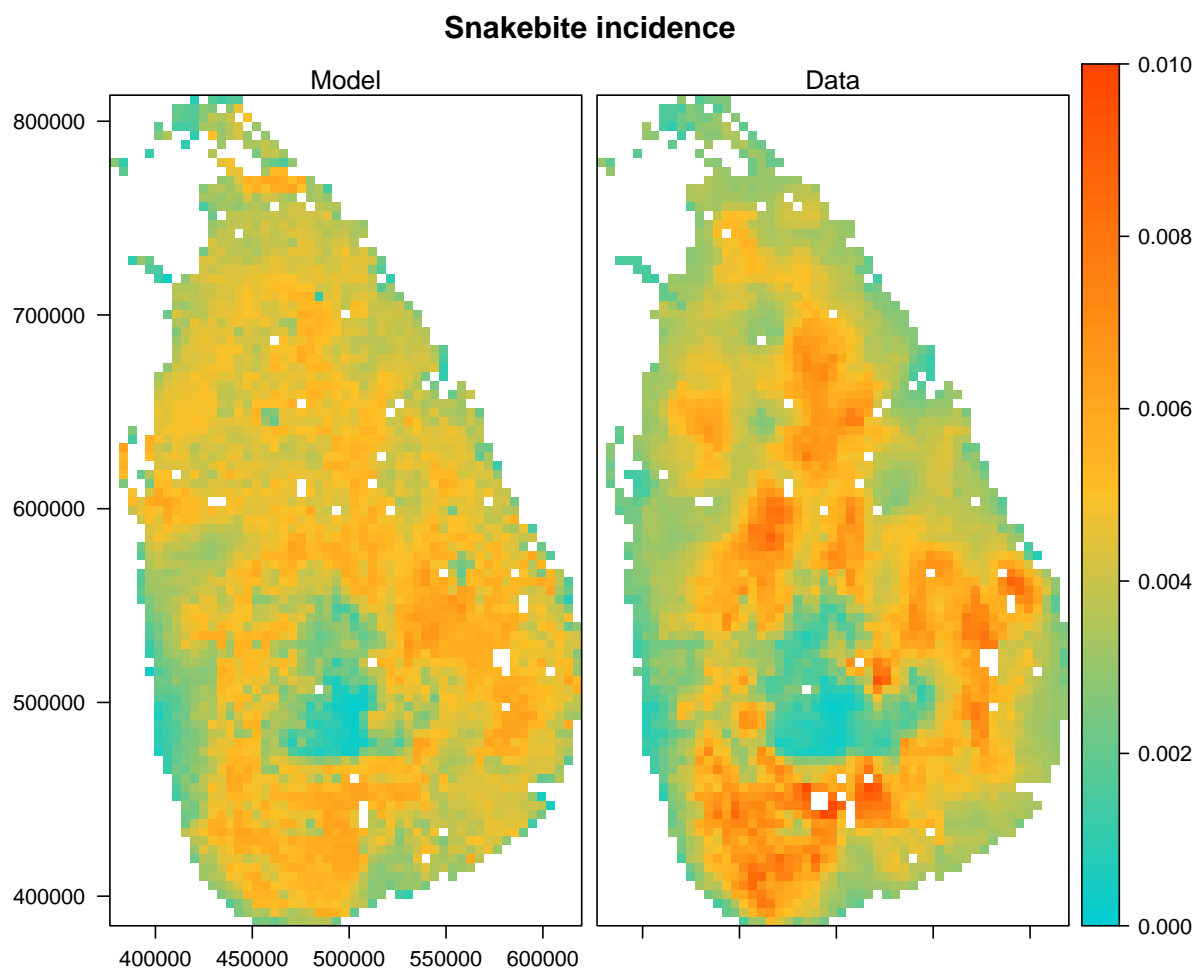Figure 1: Number of snakebites predicted by the model and from data used to fit the model

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

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.
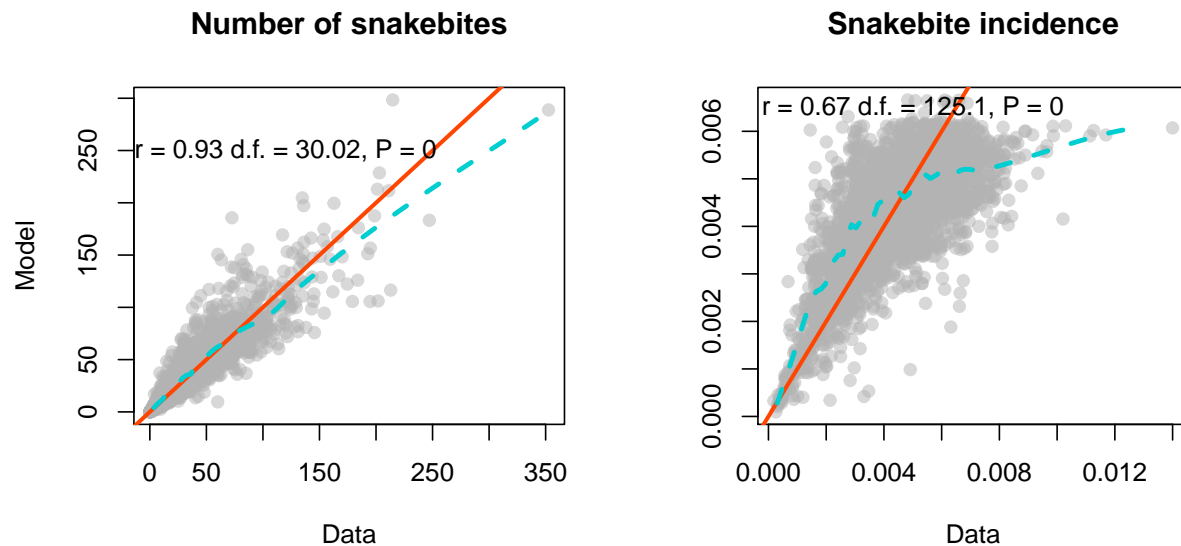

Figure 3: Correlation between model and data.

The orange line represents the gold standard, which is a line with intercept zero and a slope of one. The dashed green line is the actual trend between model and data, and was generated with a loess regression with a scale (smoothing) parameter of 0.1. The upper left corner numbers are the estimated correlation coefficient and the significance of the correlation between data and the median of model posteriors.

Relationship of estimated indices and collated information

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

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

## The following object is masked from 'package:latticeExtra':
##
##     layer
```
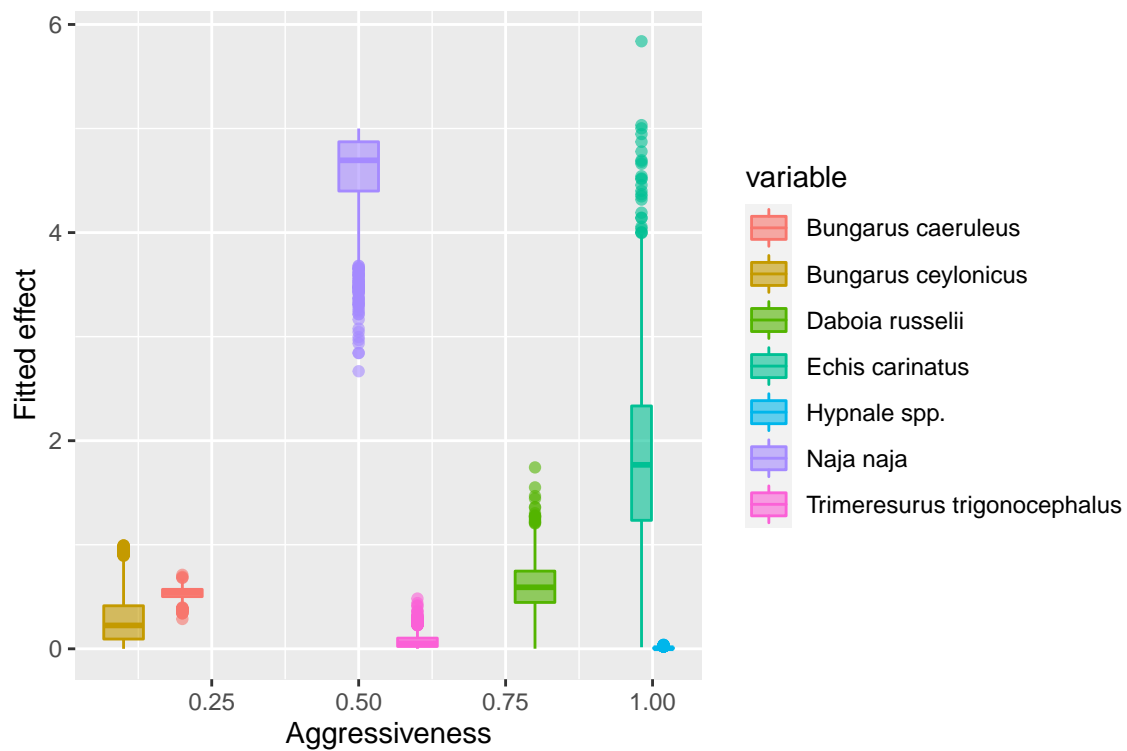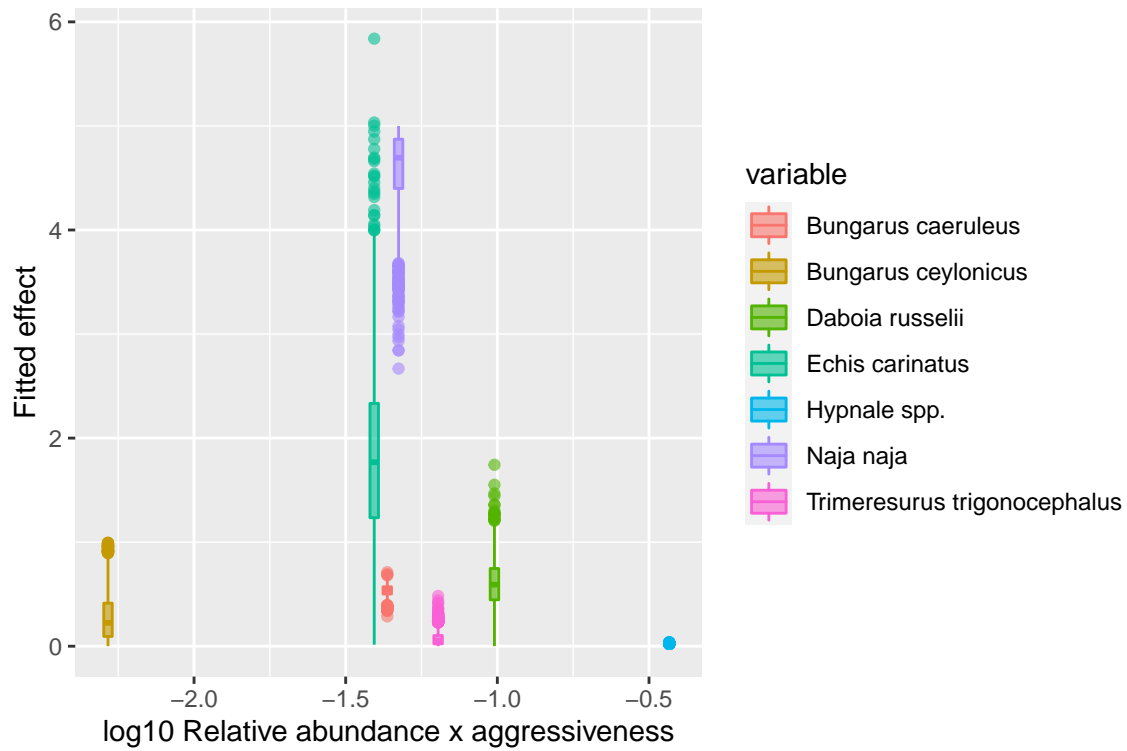
Analysis of residuals

As with all regression exercises, residuals should be normally distributed. Here are a series of diagnostic plots.
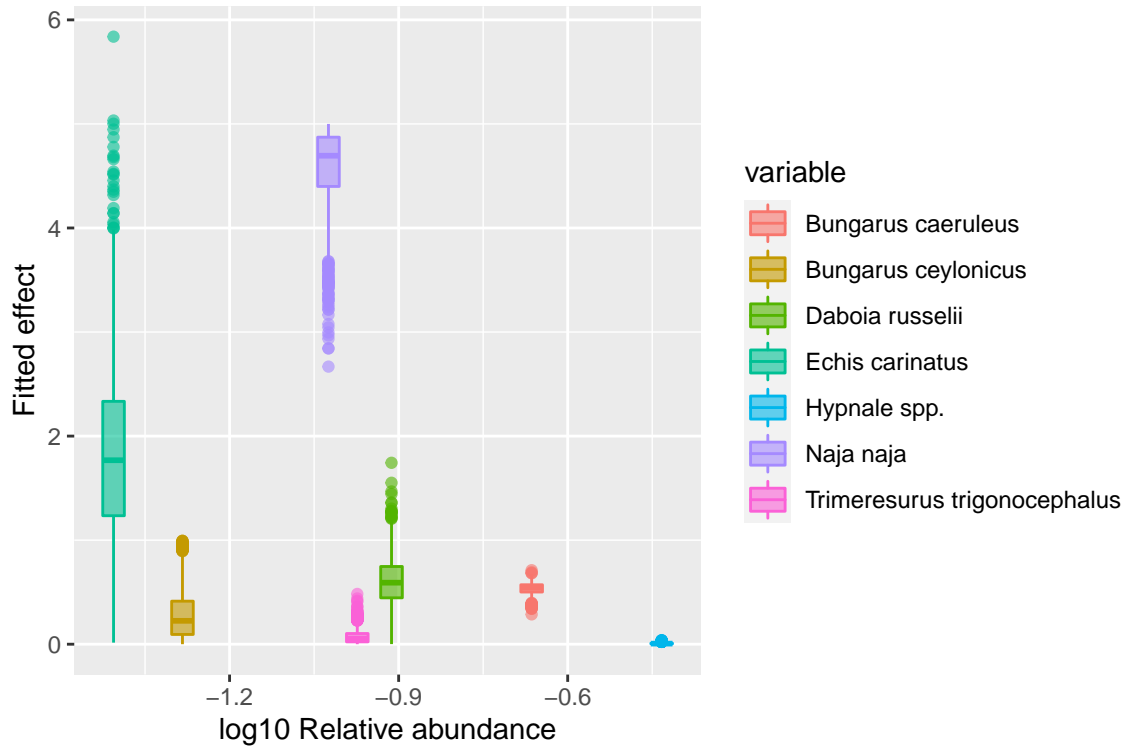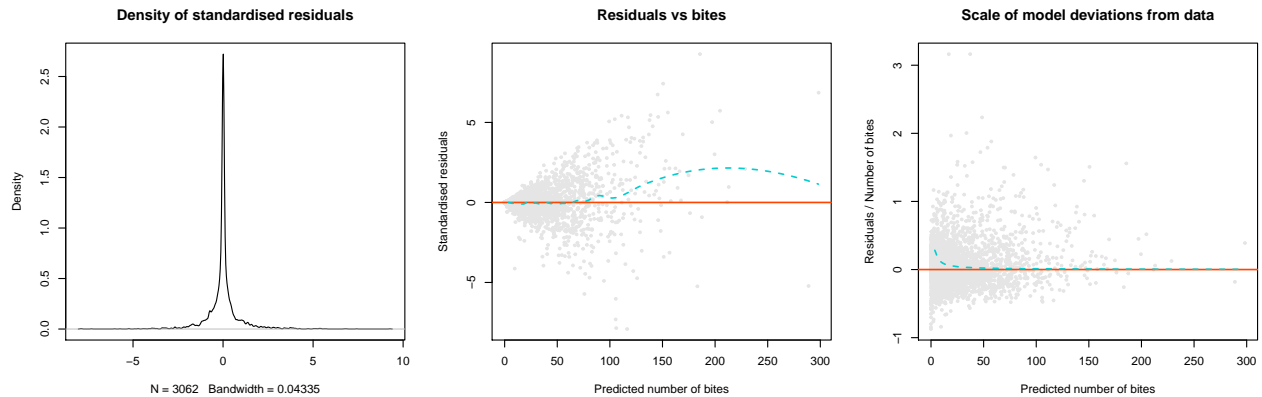
Figure 4: Relationship between individual species cotnact rates and aggressiveness and relative abundance weights, and aggressiveness weights only and relative abundance weights only.

Statistical distribution of model predictions

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.
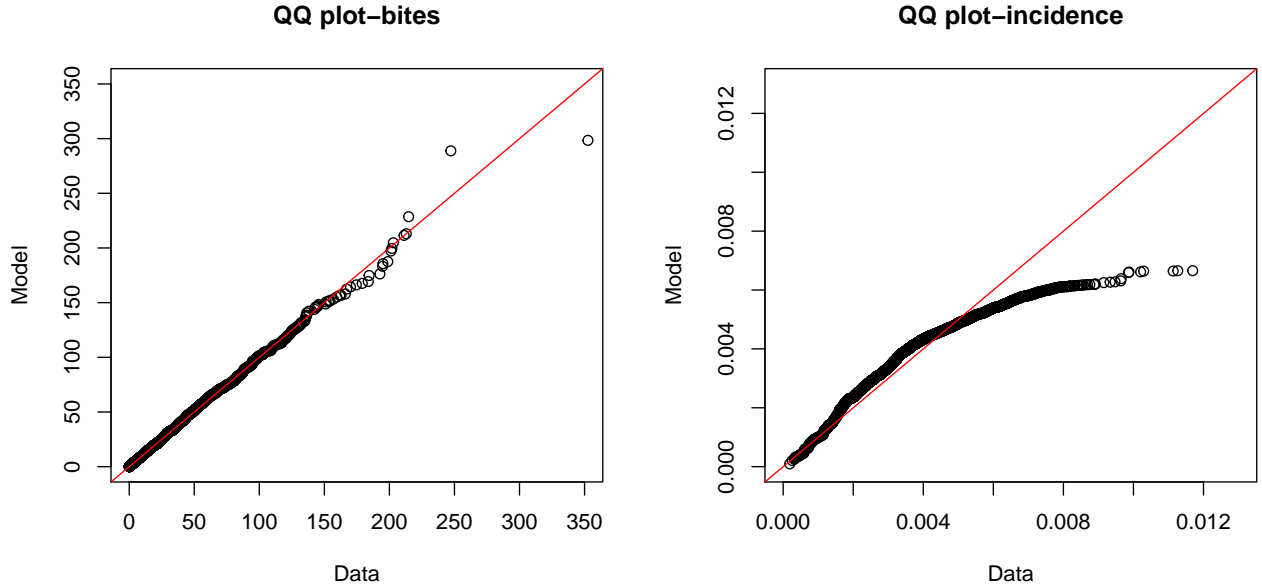
Figure 5: Statistical distribution of the number of snakebites produced by the model, compared with the statistical distribution of the data. If both data by quantile regions for a straight line both data have similar distributions. Bite incidence rates produced by the model differ from data.

Residuals in space:

The first assumption in a regression procedure is that all data re 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.

The left side shows the standardised residuals of snakebites, and teh 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.

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 sugested that there is a negative effect of human population on incidence. However the number of snakebite cases does escalate 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.
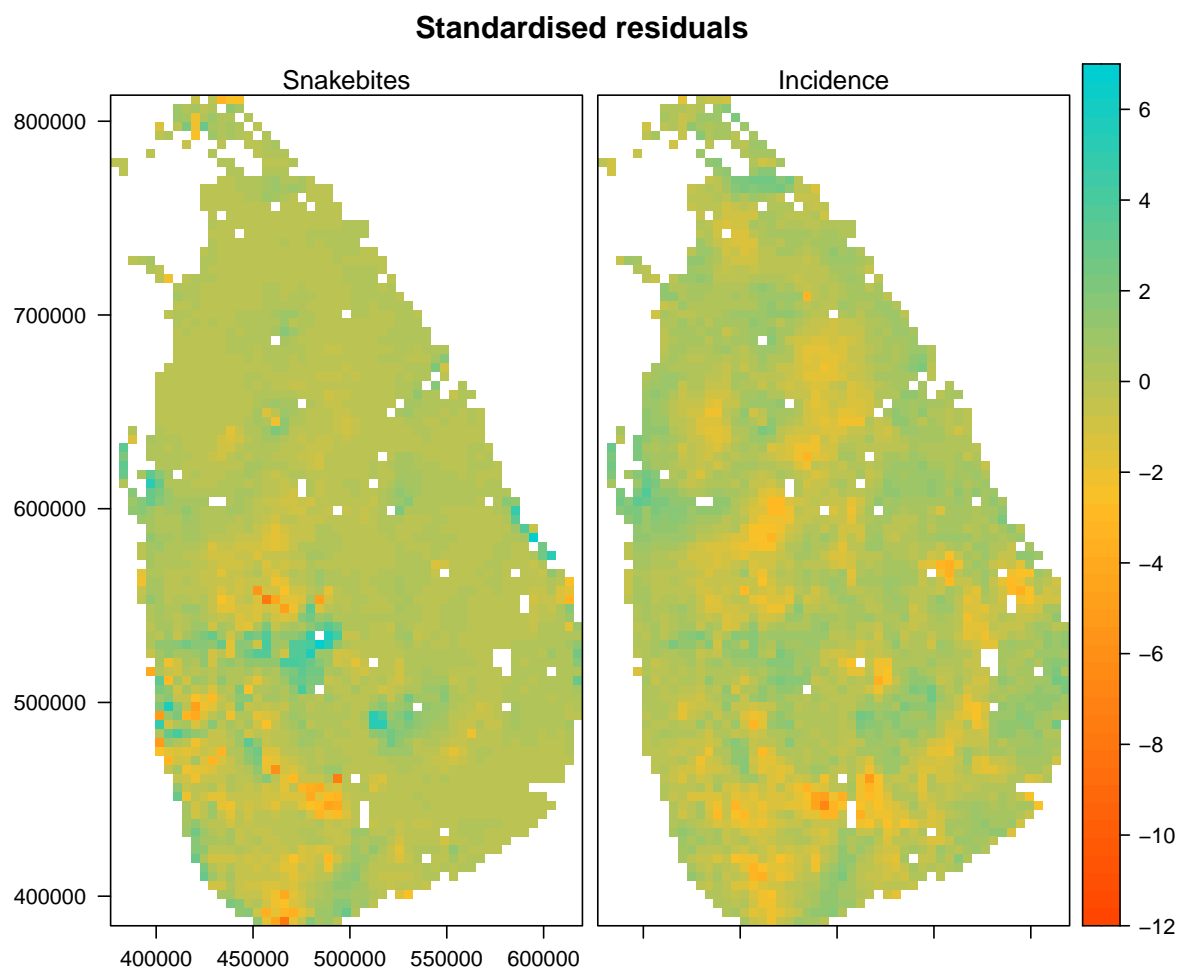
10

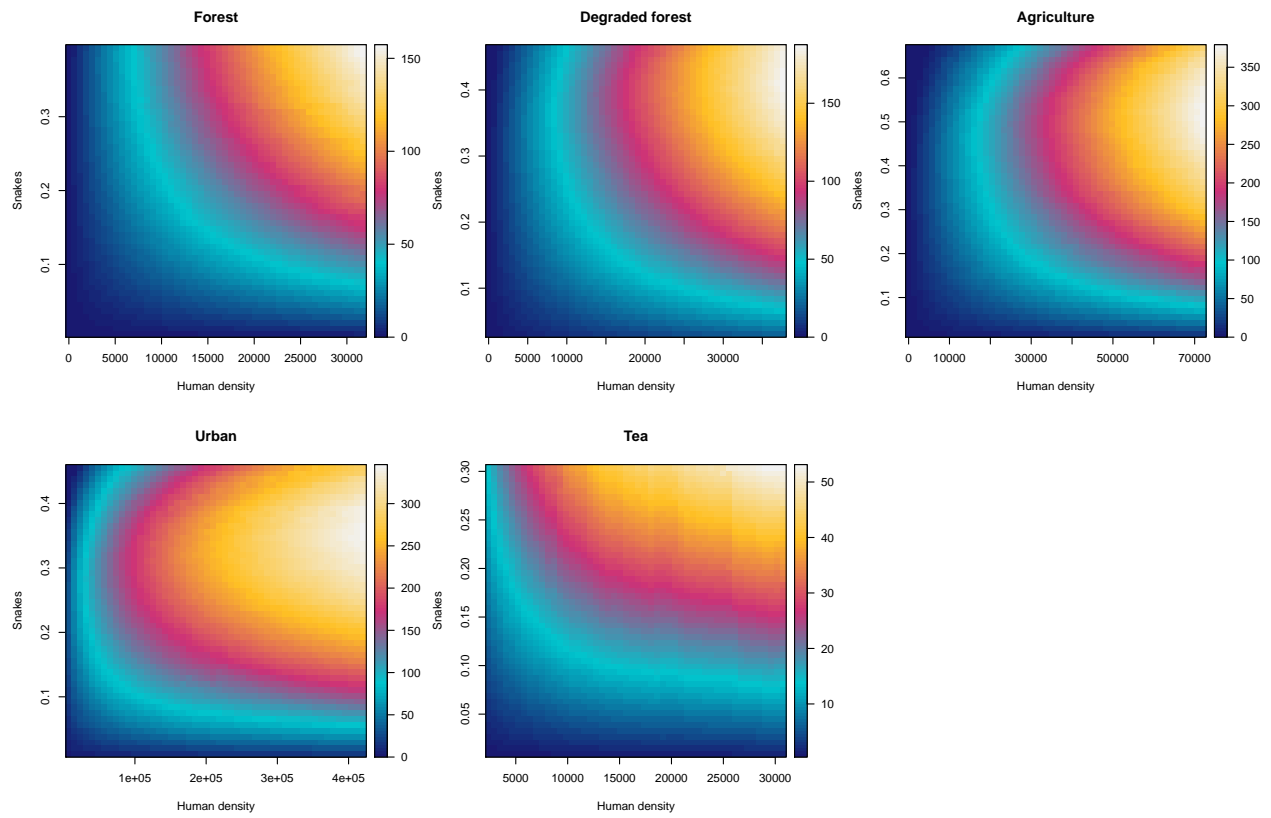Figure 6: Spatial pattern of standardised residuals

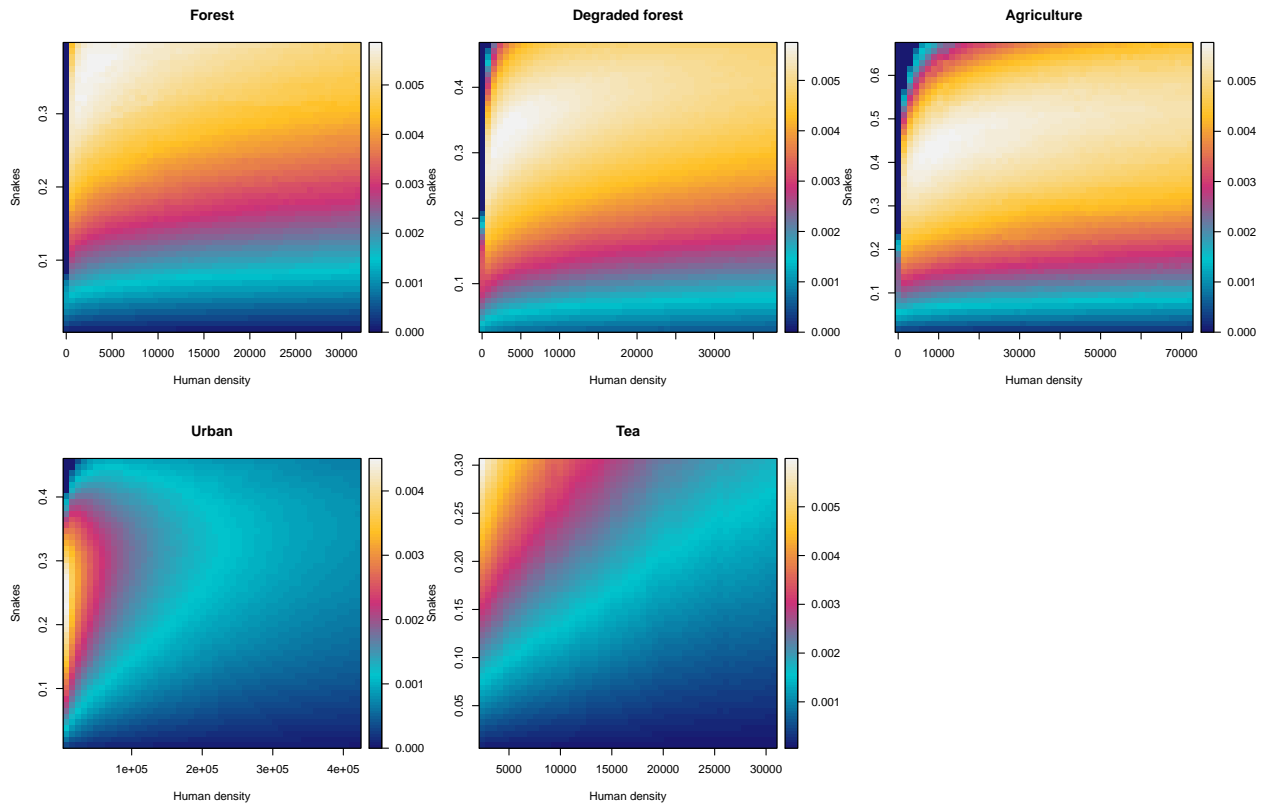Figure 7: Partial response of number of snakebites to humans and snakes per land cover class.

Figure 8: Partial responses of snakebite incidence to humans and snakes per land cover class.

Contact rate across Sri Lanka

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

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

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.

```
## Warning in matrix(values, nrow = ncell(x), ncol = nlayers(x)): la longitud de
## los datos [14760] no es un submúltiplo o múltiplo del número de filas [4929] en
## la matriz
```
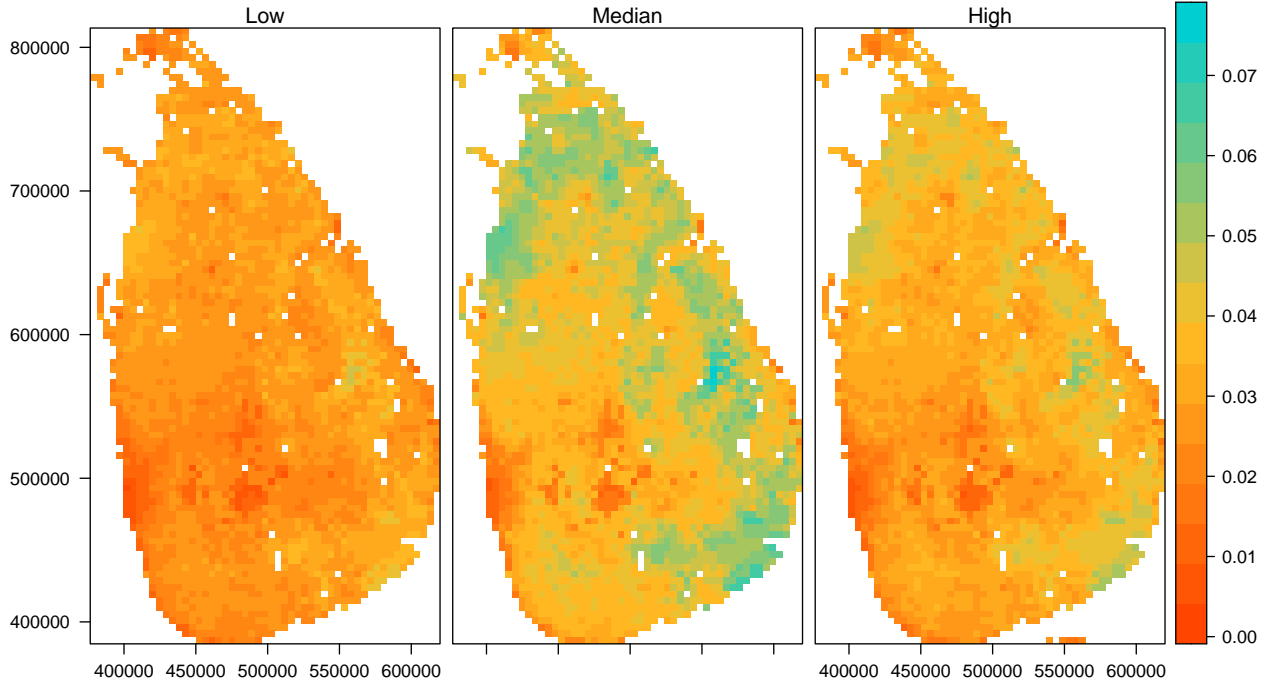

Figure 9: Spatial representation of the reduction of total snakes in response to human population density.

To visualy 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:

Second with the contact rate in logarithmic scale.

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

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.

The negative effect of snakes

As evidenced by the partial responses the number of bites could decrease after a given value of the snake abundance index. If this actually occurs in the data used to fit the mode it could indicate 1) inadequate
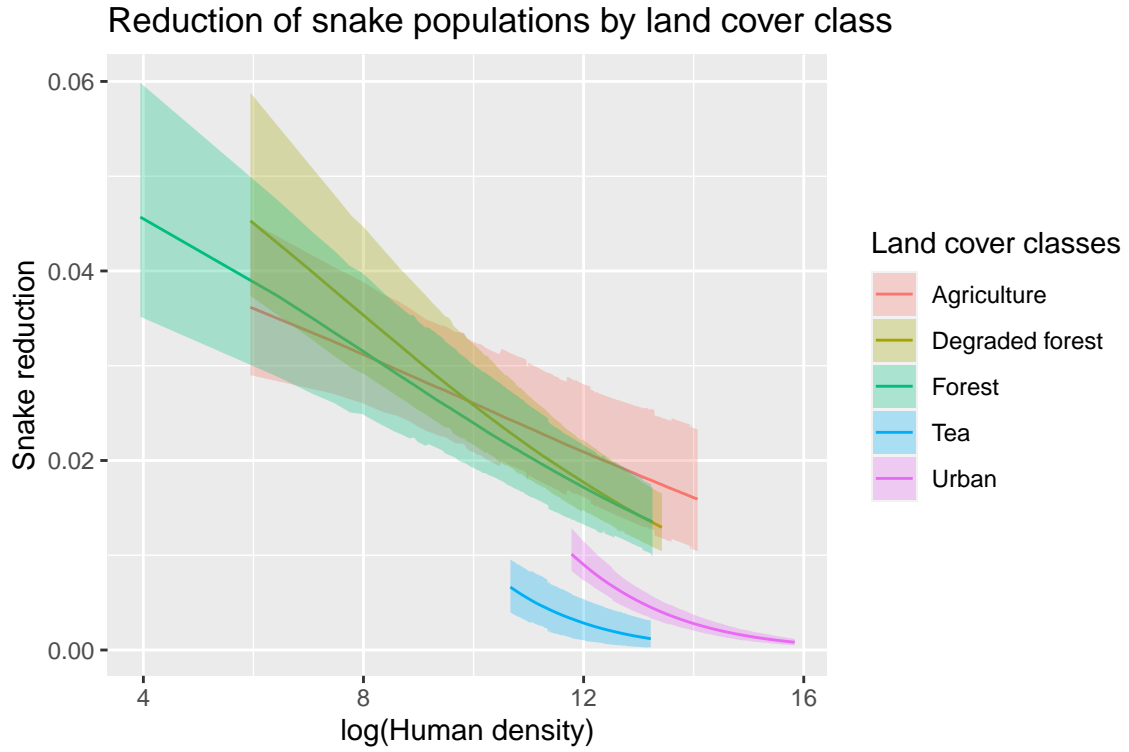
14

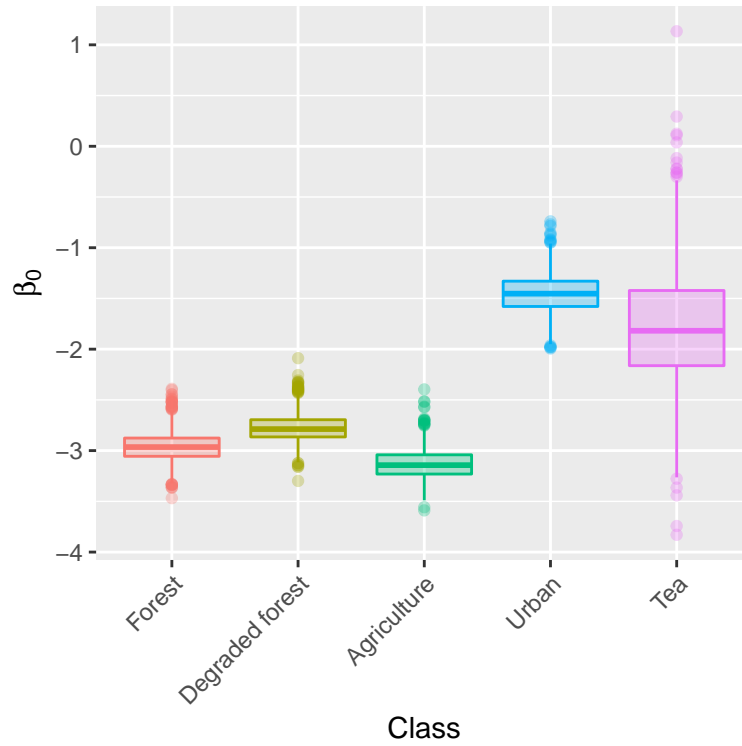Figure 10: Effect of humans on snake abundance.

Figure 11: Distribution of posterior samples for beta 0 parameters of the function to adjust snake abundance in relation to humans.
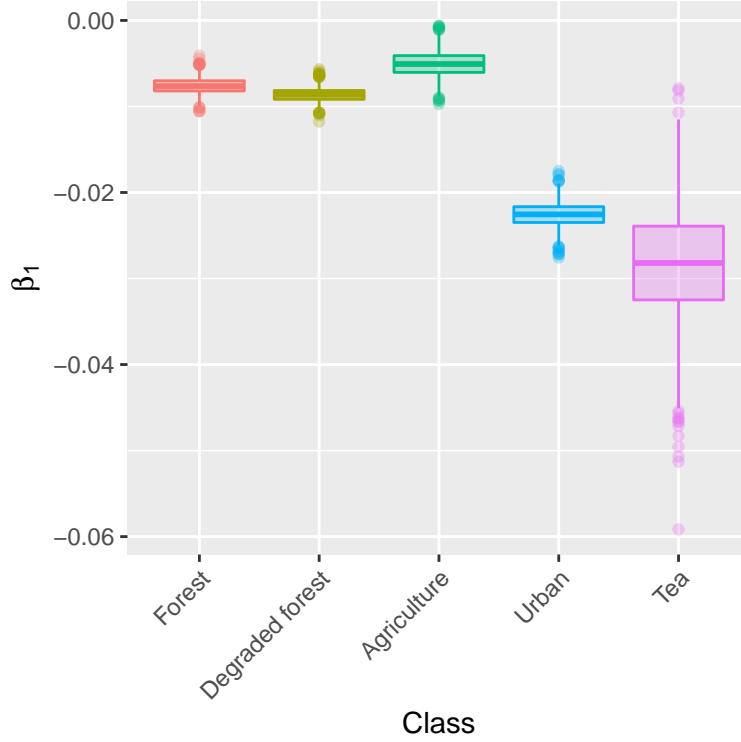
Figure 12: Distribution of posterior samples for beta 1 parameters of the function to adjust snake abundance in relation to humans.

estimates of potential snake abundance with respect to humans (similar to what the contact rate suggests), or 2) inadequate quantification of the effect of humans on snakebite.

By analysing the equilibrium conditions of equation @label(eq:bites) the number of bites starts decreasing at increasing values of the snake abundance index when:

$$S > \ln H^* q$$

Then it is possible to map the areas where the number of bites decreases with respect to the snak abundance index. By solving this inequality I found that there were no areas in the model predictions where snakes correlate negatively with the number of snakebites:
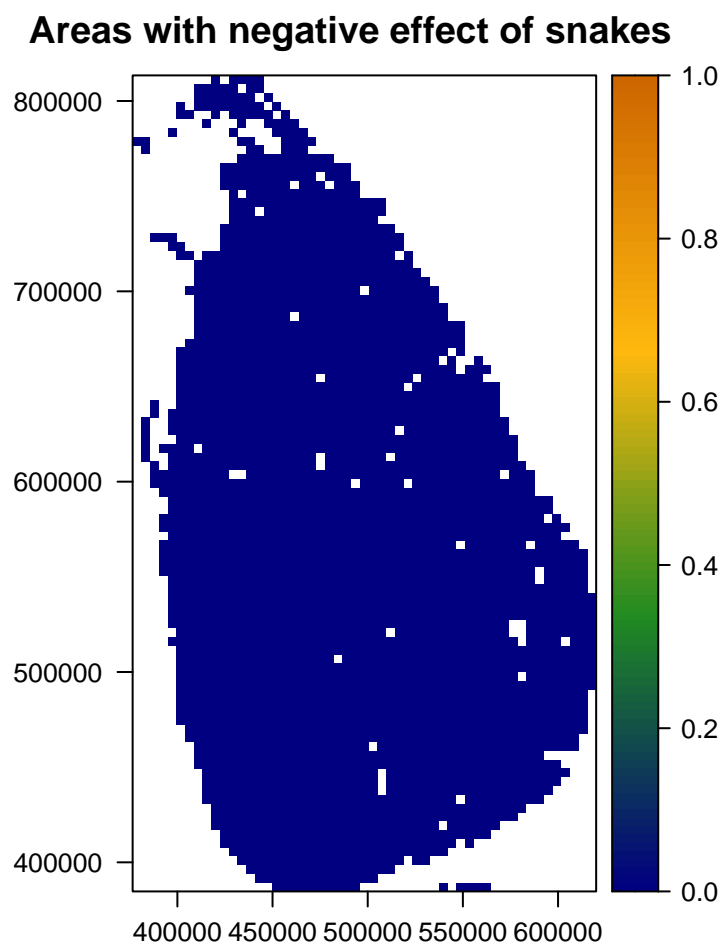
Figure 13: Probability that there is a negative effect of snakes on snakebite incidence according to the poterior samples of the refige effect model.