

Potential incidence: Refuge effect model

Gerardo Martin

7th June 2019

Intro and Methods

The model explained here was selected from a pool of models based on generic terms for modelling directly transmitted diseases. In the selected model the number of snakebite cases per unit time depends on the product of snake abundance and human population density with a refuge effect:

$$\frac{dH_{bitten}}{dt} = \beta S(\ln(H_{total}) - S/q_h)$$

Where

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

is the sum of each snake species abundance weighted by the expert generated index of aggressiveness A_i and an estimated index I_i for correcting abundance with respect to the other snake species; H_{total} is human population density per grid cell and β and q_h are the contact rate and proportion of snakes which do not actually encounter humans.

The other disease transmission terms are listed in McCallum 2001 (Trends in Ecology and Evolution).

As noted in the equation above I used the natural logarithm of human population density because the raw layer of human density had a very clustered distribution around zero with few grid cells with very large values. Using the log-converted human density layer made MCMC sampling and parameter inference.

The most succesful approach to estimate parameters I found is by transforming the above model into the probability the probability that a snakebite occurs during the time where data was collected. This method assumes that the above equation also represents the entire process of snakebite dynamics and that people bitten remain susceptible to bites. The transformed equation is:

$$P = 1 - \exp(-\beta S(\ln(H_{total}) - S/q_h))$$

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

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

To analyse the role of land cover in the incidence of snakebite, I used the most frequent land cover classes between 2004 and 2017, to estimate the parameters of the above model for each land cover. So that we'd be able to find if there are different responses to huma 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(ind.ag[], S[i,])
##     P[i] <- 1 - exp( - Beta[i] * S.sum[i] * (Hum.log[i] - S.sum[i]/qh[class[i]]))
##     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, 50)
##   }
##   for(k in 1:n.species){
##     indices[k] ~ dunif(0,1)
##     ind.ag[k] <- indices[k] * aggressiveness[k]
##   }
## }
## Fully observed variables:
## Hum.bit Hum.log Hum.tot S aggressiveness 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.

```
print(bites.incidence)
```

```

## Inference for Bugs model at "Snakebites models/Indices-aggressiveness/Land cover human-beta/Negbin-i
## 3 chains, each with 6e+05 iterations (first 60000 discarded), n.thin = 540
## n.sims = 3000 iterations saved
##
##      mu.vect sd.vect      2.5%      25%      50%      75%      97.5%
## beta0[1]   -3.882  0.139   -4.132   -3.974   -3.890   -3.802   -3.568
## beta0[2]   -3.745  0.122   -3.938   -3.828   -3.762   -3.679   -3.449
## beta0[3]   -3.996  0.149   -4.252   -4.099   -4.006   -3.908   -3.674
## beta0[4]   -2.634  0.214   -3.044   -2.778   -2.634   -2.500   -2.195
## beta0[5]   -2.517  0.566   -3.662   -2.907   -2.508   -2.155   -1.374
## beta1[1]    -0.012  0.001   -0.014   -0.013   -0.012   -0.012   -0.011
## beta1[2]    -0.013  0.001   -0.014   -0.013   -0.013   -0.012   -0.011
## beta1[3]    -0.011  0.001   -0.013   -0.011   -0.011   -0.010   -0.008
## beta1[4]    -0.024  0.002   -0.027   -0.025   -0.024   -0.023   -0.021
## beta1[5]    -0.034  0.006   -0.046   -0.038   -0.034   -0.030   -0.022
## indices[1]   0.916  0.078   0.707   0.882   0.939   0.975   0.998
## indices[2]   0.061  0.055   0.002   0.020   0.045   0.089   0.204
## indices[3]   0.112  0.035   0.046   0.088   0.110   0.135   0.185

```

```

## indices[4]      0.025   0.011   0.006   0.017   0.024   0.032   0.050
## indices[5]      0.006   0.005   0.000   0.002   0.005   0.009   0.019
## indices[6]      0.439   0.046   0.334   0.414   0.443   0.470   0.515
## indices[7]      0.016   0.013   0.001   0.006   0.013   0.023   0.049
## qh[1]           0.048   0.010   0.032   0.042   0.047   0.053   0.071
## qh[2]           0.036   0.004   0.027   0.033   0.036   0.038   0.042
## qh[3]           0.040   0.004   0.030   0.037   0.040   0.042   0.046
## qh[4]           0.023   0.002   0.018   0.022   0.023   0.025   0.027
## qh[5]           25.706  18.965   1.259  10.605  21.848  37.195  71.322
## r[1]            20.920  2.010  17.240  19.542  20.810  22.195  25.166
## r[2]            18.404  1.125  16.330  17.596  18.368  19.133  20.714
## r[3]            12.554  1.633   9.738  11.405  12.438  13.558  16.163
## r[4]             8.647  0.808   7.122   8.074   8.628   9.173  10.302
## r[5]            14.305  5.268   6.946  10.611  13.345  16.916  26.976
## deviance      18837.659   7.271 18825.726 18832.406 18836.777 18842.095 18853.859
##               Rhat n.eff
## beta0[1]      1.002  1900
## beta0[2]      1.001  3000
## beta0[3]      1.001  3000
## beta0[4]      1.001  2000
## beta0[5]      1.001  2500
## beta1[1]      1.002  1900
## beta1[2]      1.002  2600
## beta1[3]      1.001  3000
## beta1[4]      1.001  2500
## beta1[5]      1.001  2300
## indices[1]    1.004  3000
## indices[2]    1.001  3000
## indices[3]    1.007  2300
## indices[4]    1.001  3000
## indices[5]    1.001  3000
## indices[6]    1.001  3000
## indices[7]    1.001  3000
## qh[1]         1.001  3000
## qh[2]         1.002  3000
## qh[3]         1.003  3000
## qh[4]         1.002  3000
## qh[5]         1.001  3000
## r[1]          1.001  3000
## r[2]          1.001  3000
## r[3]          1.001  3000
## r[4]          1.002  1900
## r[5]          1.001  3000
## deviance      1.002  2700
##
## 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 = 26.4 and DIC = 18864.1
## DIC is an estimate of expected predictive error (lower deviance is better).

```

Model predictions

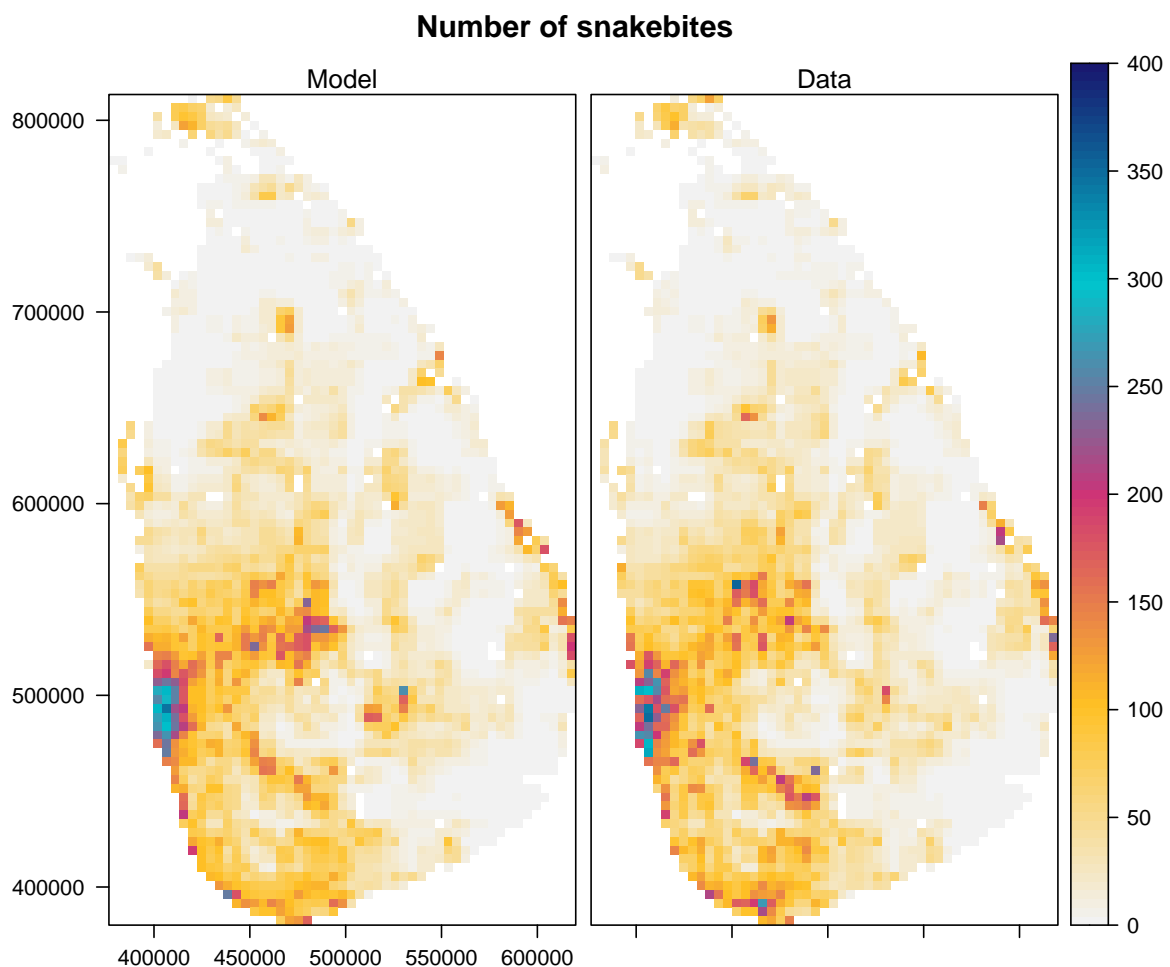
The R function to predict the JAGS script above is:

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

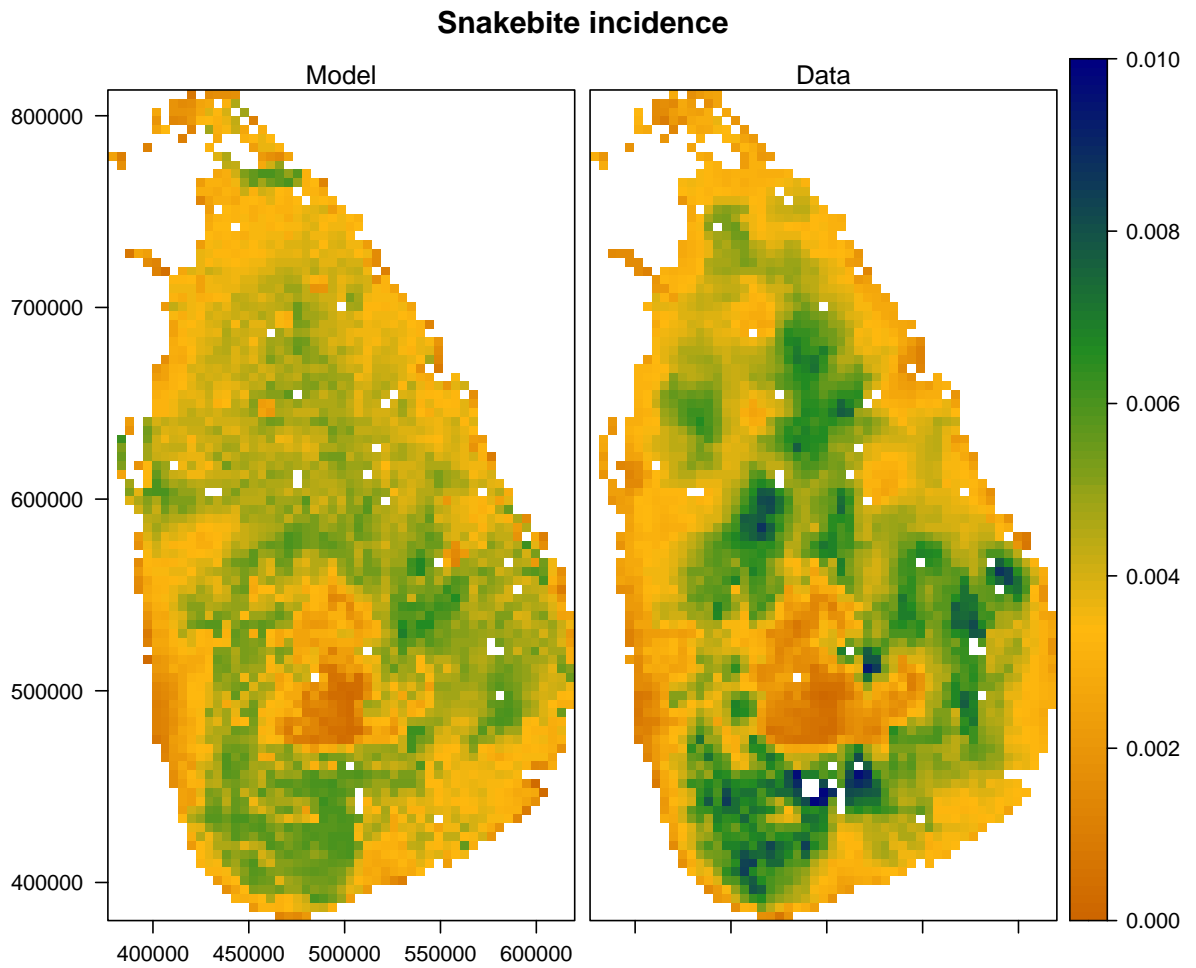
Which uses the indices of snakes abundance weighted by their aggressiveness in advance. The result of running the function is the number of snakebite cases (H.bit) predicted by the model.

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.

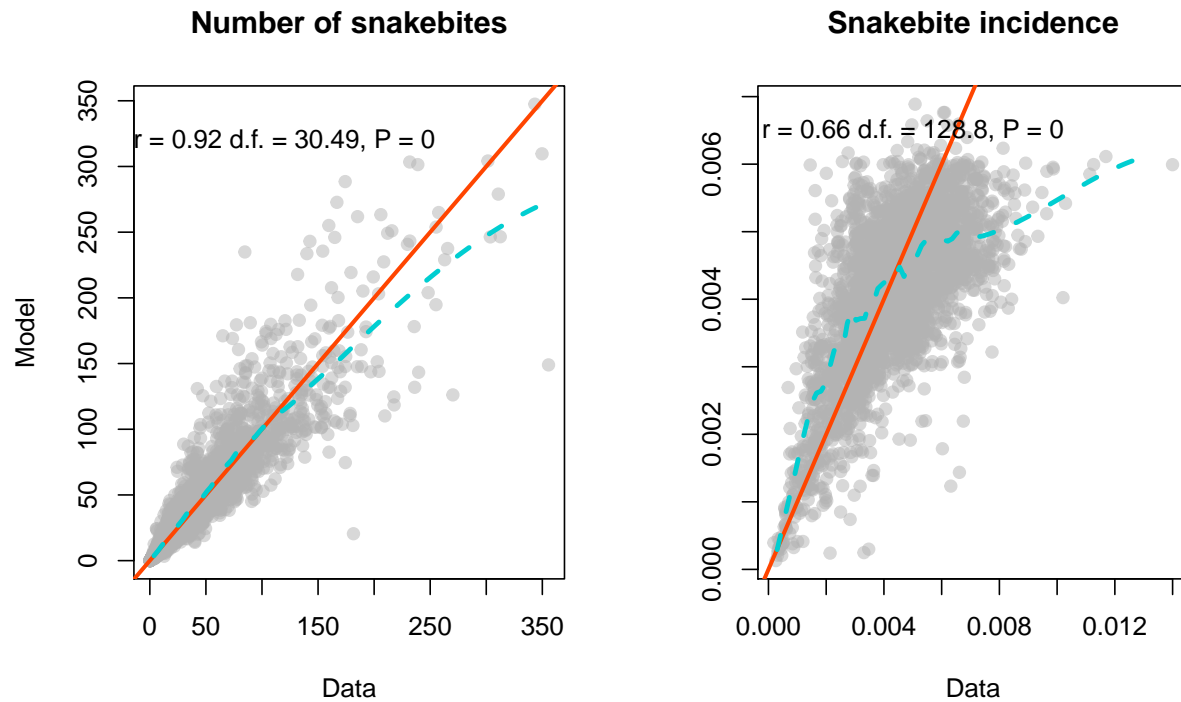
First here is the total number of snakebites estimated by the model (left) and the number of snakebites used as data (right). The colourscale of both maps is the same.



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 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.



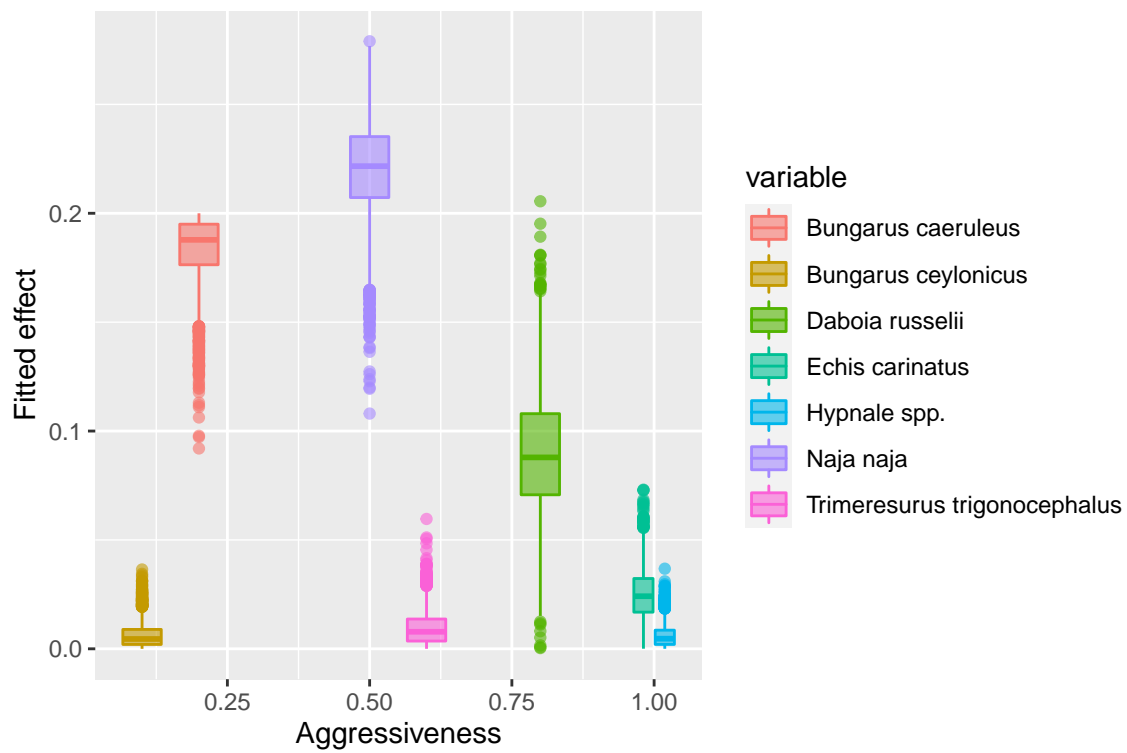
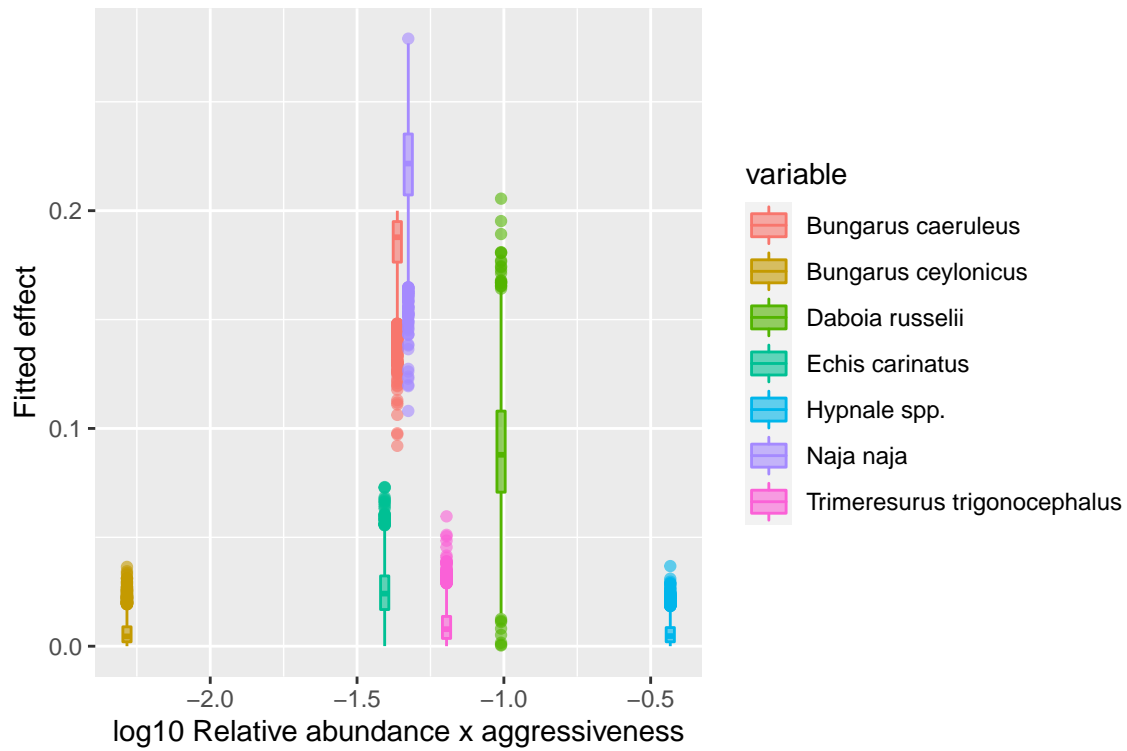
The orange line represents the gold standard, which is a line with intercept zero and a slope of one. The dashed green line is the actual trend between model and data, and was generated with a loess regression with a scale (smoothing) parameter of 0.1. The upper left corner numbers are the estimated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

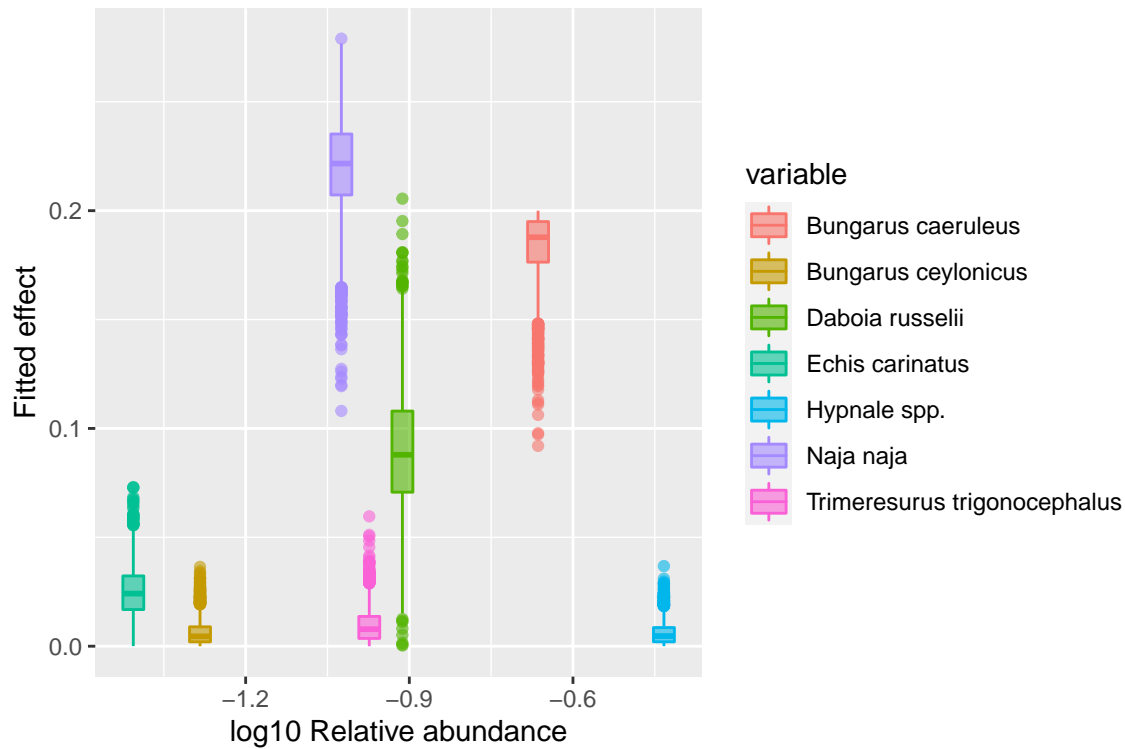
Relationship of estimated indices and collated information

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

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

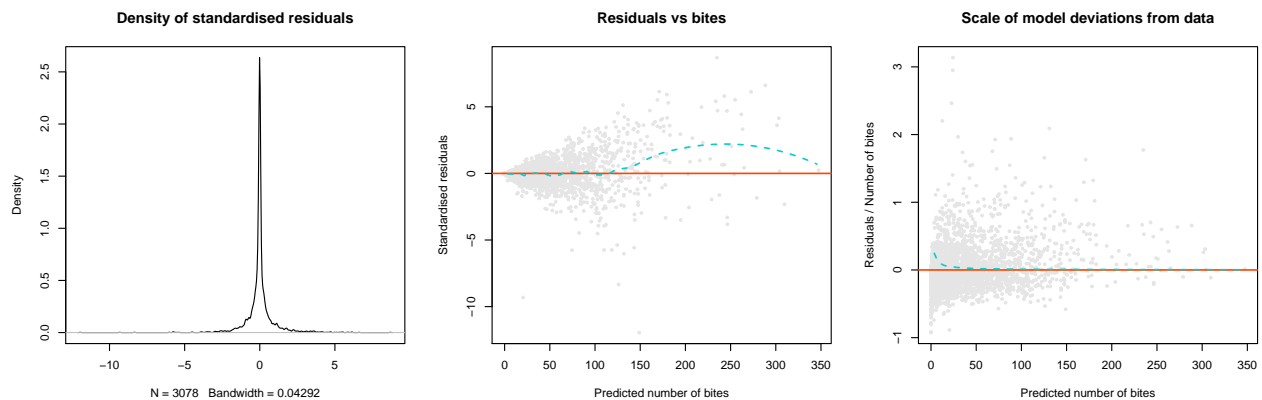
## The following object is masked from 'package:latticeExtra':
##
## layer
```





Analysis of residuals

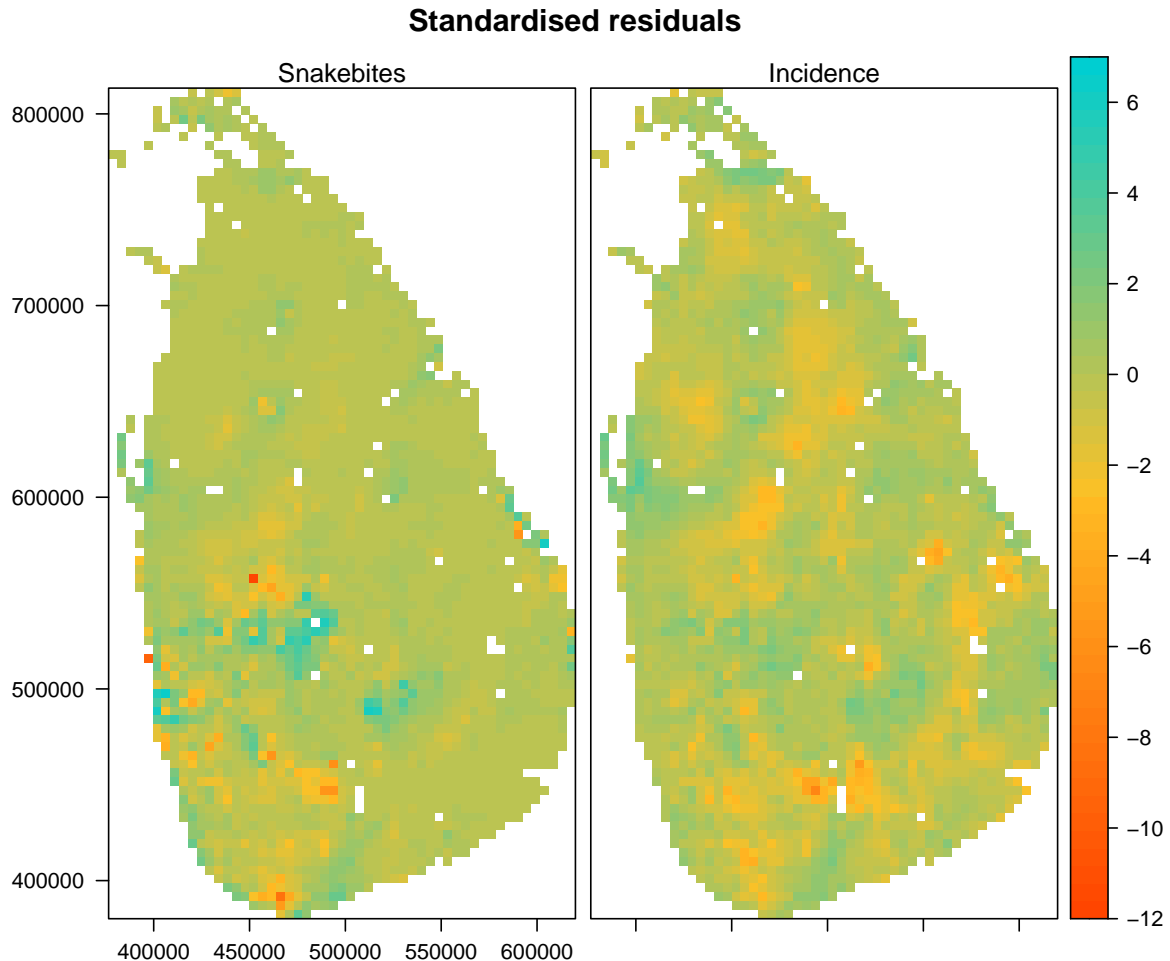
As with all regression exercises, residuals should be normally distributed. Here are a series of diagnostic plots.



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:

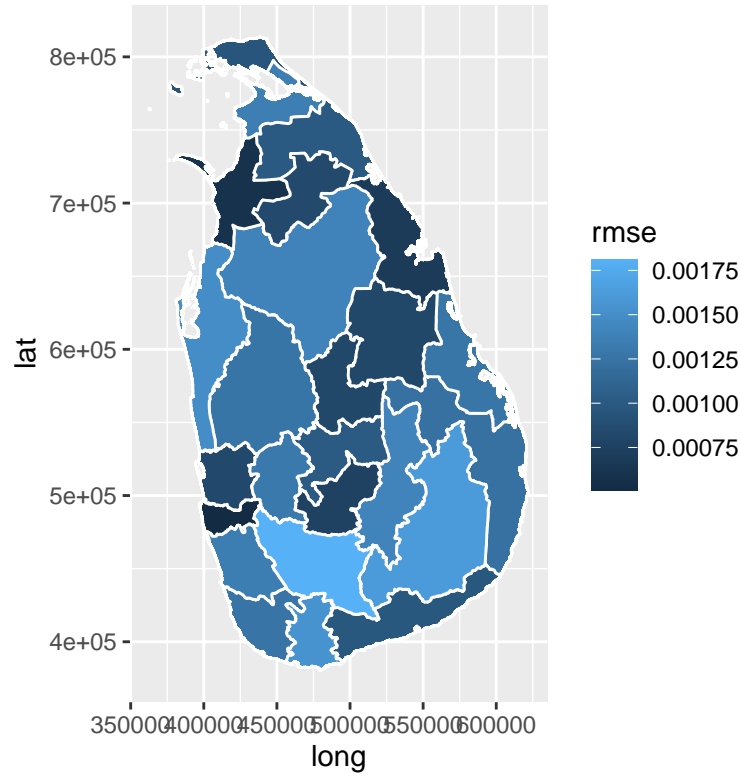
The first assumption in a regression procedure is that all data are independent. With spatial data this assumption is easily violated. Thus here I plotted the standardised residuals in space to see whether there is any clustering or the presence of other factors that we did not account for.



The left side shows the standardised residuals of snakebites, 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.

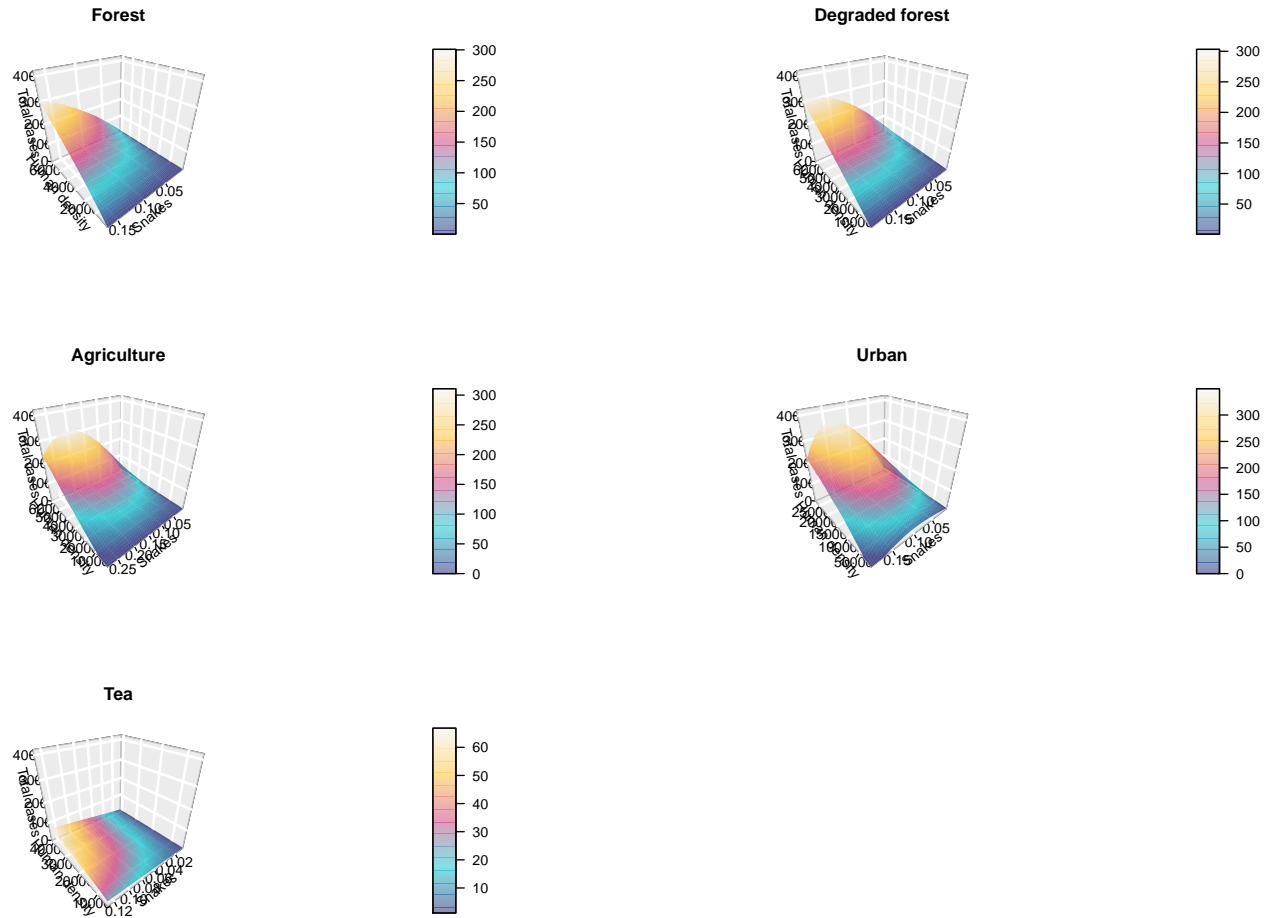
```
## OGR data source with driver: ESRI Shapefile
## Source: "/media/gerardo/Almacen/SNAKEBITE MODELLING/Popn and topo data/Sri Lanka boundaries/LKA_adm1
## with 25 features
## It has 14 fields
## Integer64 fields read as strings: ID_0 ID_1 CCN_1
```



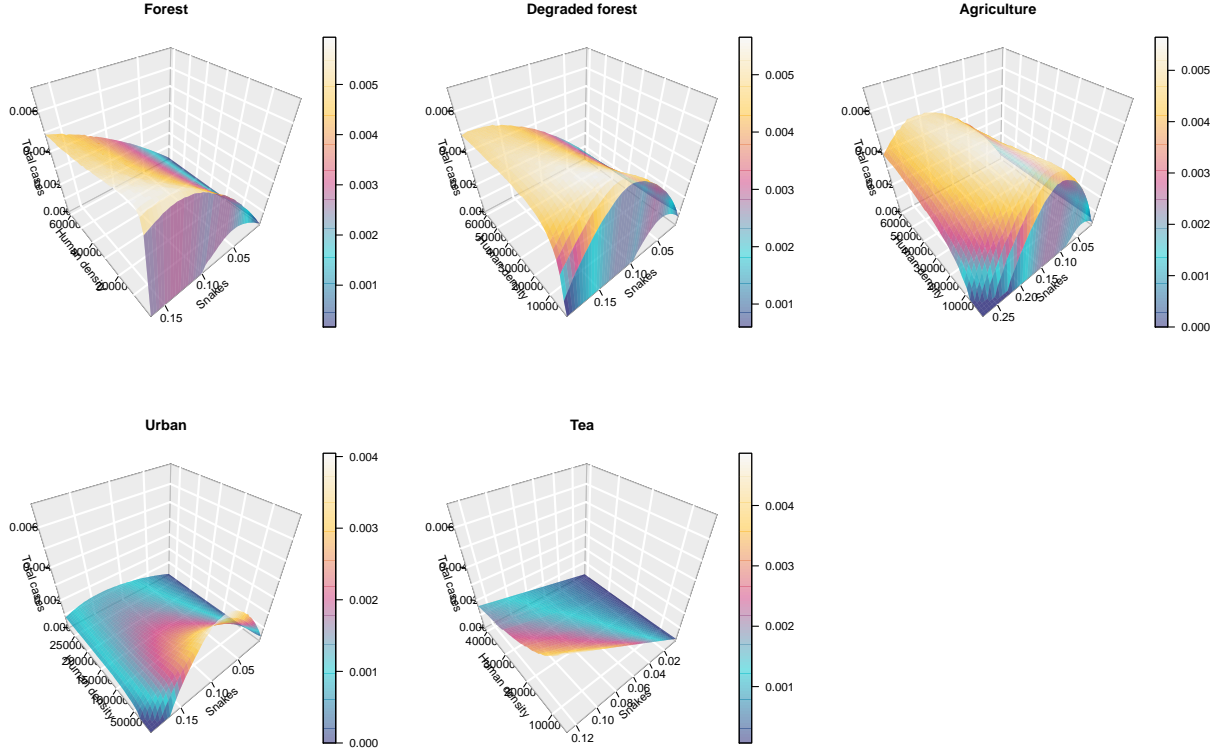
Responses fitted by the model

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 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):



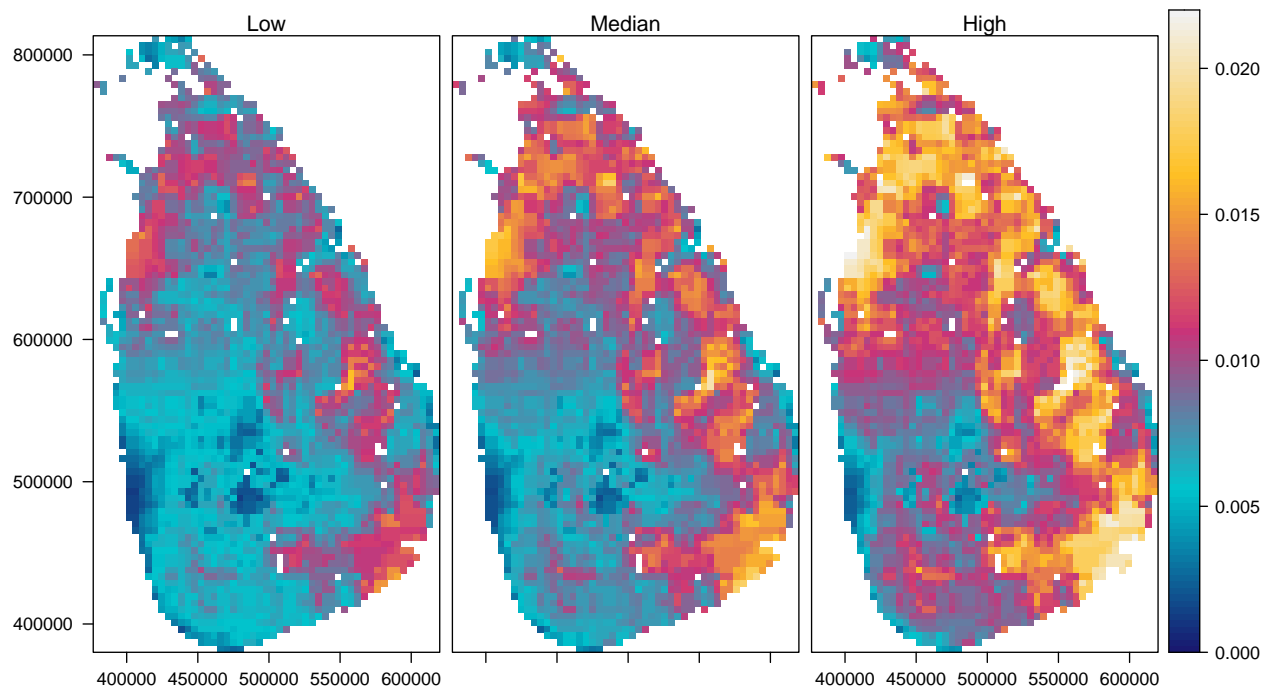
The existence of the negative effect of snakes on incidence will be confirmed in the later sections of these analyses.

Contact rate across Sri Lanka

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

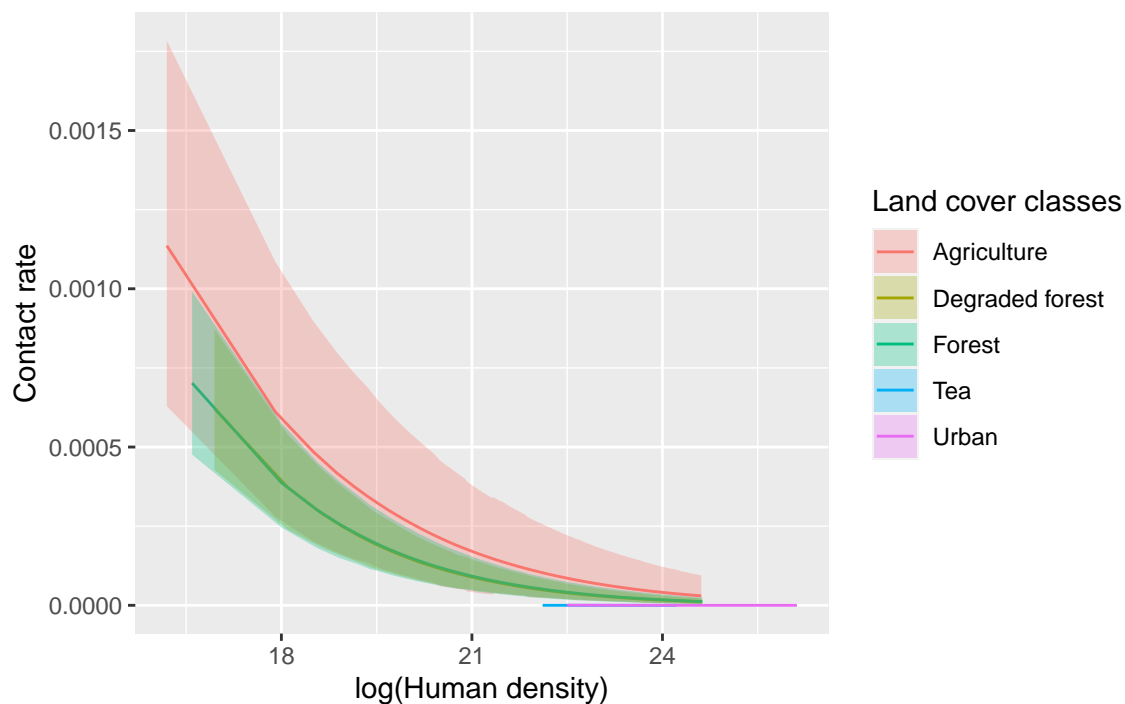
$$\beta(\ln(H_{total}), l) = \exp(\beta_{0,l} + \beta_{1,l} \ln(H_{total}))$$

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.

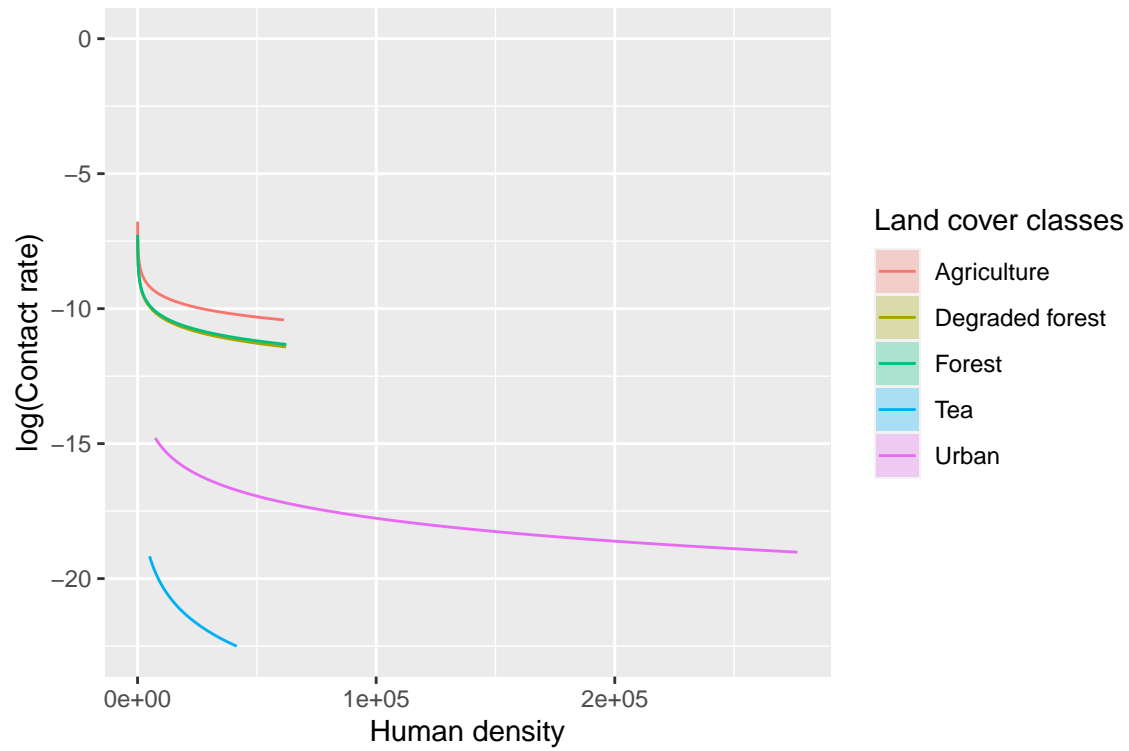


To visually assess the individual responses among land cover classes here are different values that the contact rate can take given the range of human population densities observed across Sri Lanka in each of the land cover classes, in two different scales. First with human population density in log-scale:

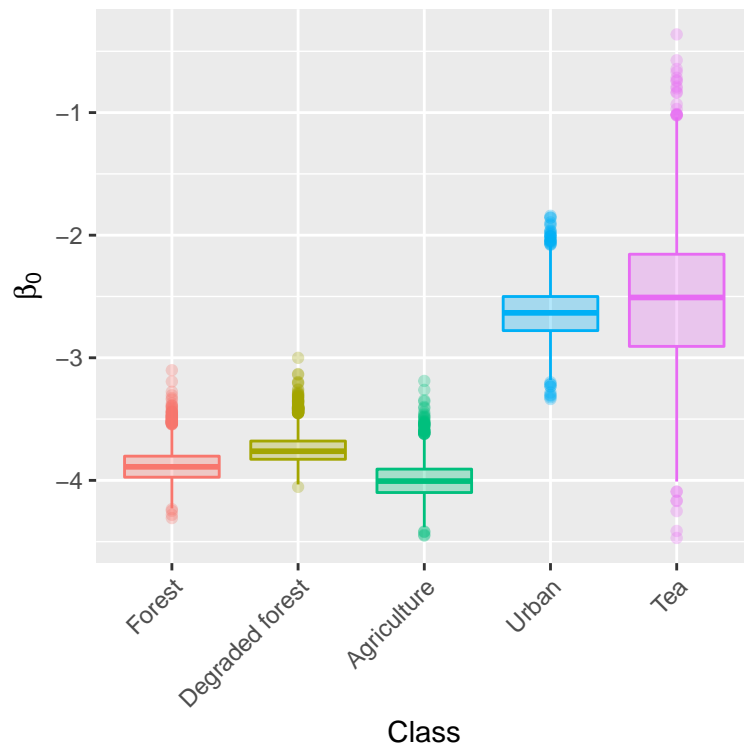
Contact rate by land cover class



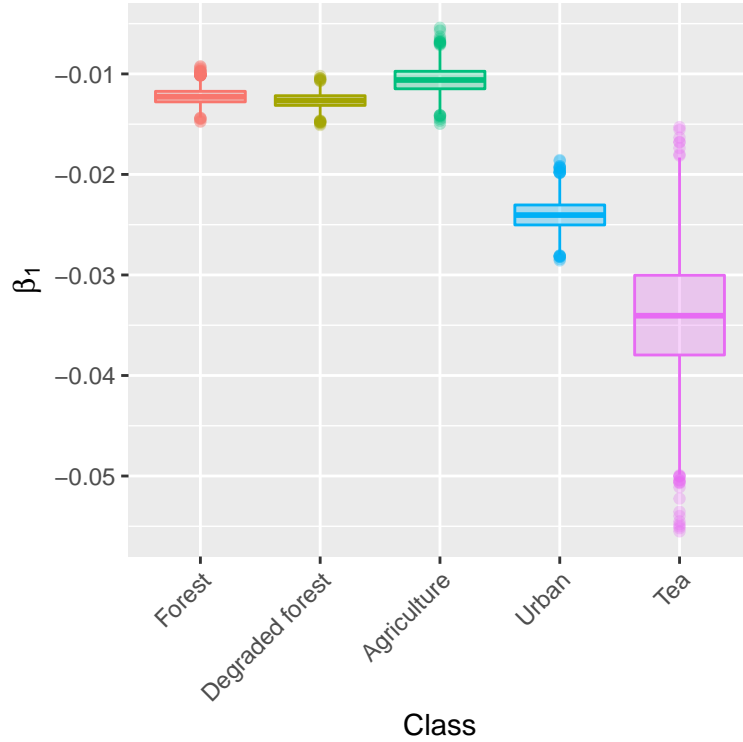
Second with the contact rate in logarithmic scale.



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

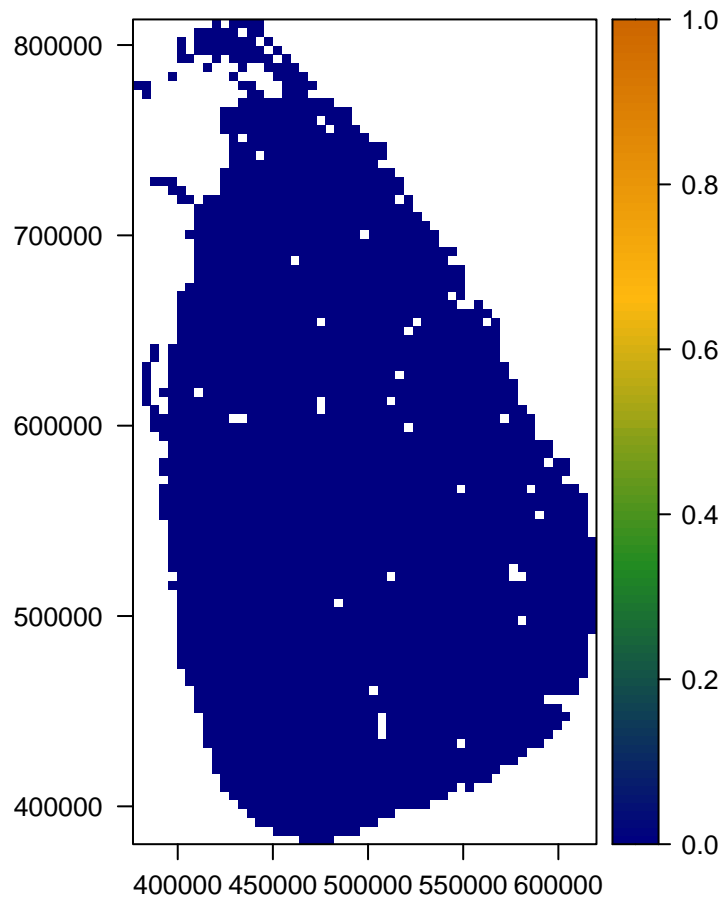
As evidenced by the partial responses the number of bites could decrease after a given value of the snake abundance index. If this actually occurs in the data used to fit the model it could indicate 1) inadequate estimates of potential snake abundance with respect to humans (similar to what the contact 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 snake 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:

Areas with negative effect of snakes



This result is encouraging and possibly indicates that improved mapping of snake abundance could be used to fully describe spatial heterogeneity of snakebite.