Lecture 20

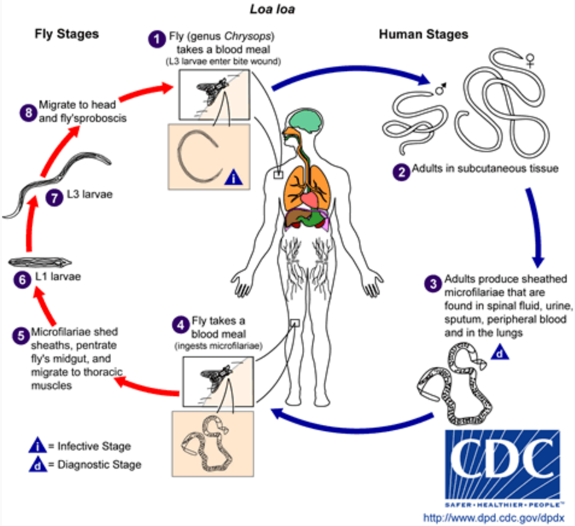
Point referenced data (pt. 2)

Colin Rundel

11/14/2018

Loa Loa Example

Loa Loa

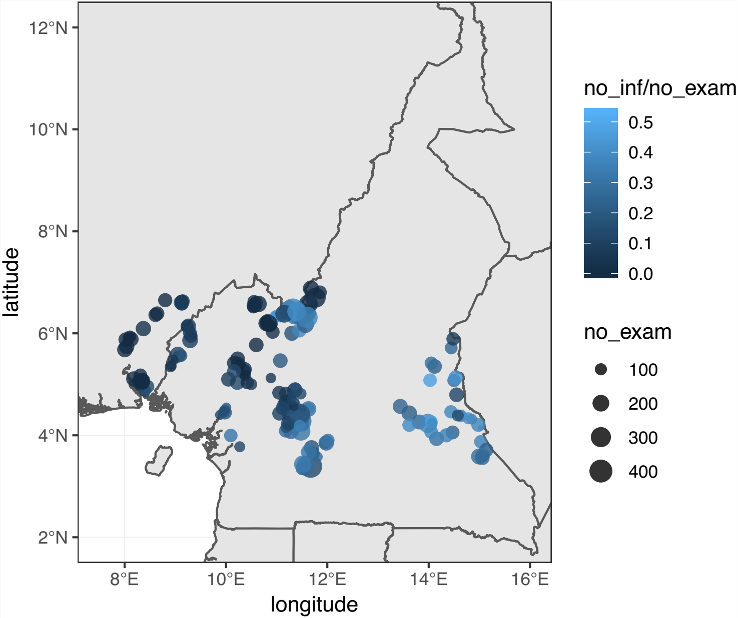


Loa loa
Fly Stages
Human Stages
1 Fly (genus Chrysops) takes a blood meal (L3 larvae enter bite wound)
2 Adults in subcutaneous tissue
3 Adults produce sheathed microfilariae that are found in spinal fluid, urine, sputum, peripheral blood and in the lungs
4 Fly takes a blood meal (ingests microfilariae)
5 Microfilariae shed sheaths, pentrate fly's midgut, and migrate to thoracic muscles
6 L1 larvae
7 L3 larvae
8 Migrate to head and fly'sproboscis
i = Infective Stage
d = Diagnostic Stage
CDC
SAFER·HEALTHIER·PEOPLE
http://www.dpd.cdc.gov/dpdx

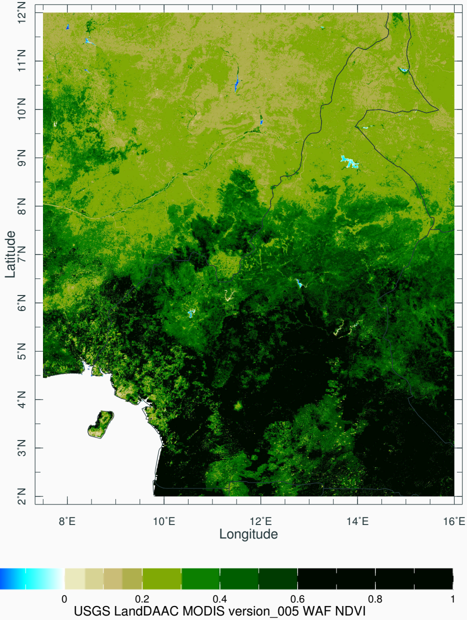
Data

```r
loaloa = tbl_df(PrevMap::loaloa) %>% setNames(., tolower(names(.))) %>%
  rename(elev=elevation)

loaloa
## # A tibble: 197 x 11
##      row villcode longitude latitude no_exam no_inf  elev mean9901 max990
##    <int>    <int>     <dbl>    <dbl>   <int>  <int> <int>    <dbl>    <dbl
##  1     1      214      8.04     5.74     162      0   108    0.439    0.6
##  2     2      215      8.00     5.68     167      1    99    0.426    0.7
##  3     3      118      8.91     5.35      88      5   783    0.491    0.7
##  4     4      219      8.10     5.92      62      5   104    0.432    0.6
##  5     5      212      8.18     5.10     167      3   109    0.415    0.8
##  6     6      116      8.93     5.36      66      3   909    0.436    0.8
##  7     7       16     11.4      4.88     163     11   503    0.502    0.7
##  8     8      217      8.07     5.90      83      0   103    0.373    0.6
##  9     9      112      9.02     5.59      30      4   751    0.481    0.8
## 10    10      104      9.31     6.00      57      4   268    0.487    0.8
## # ... with 187 more rows, and 2 more variables: min9901 <dbl>,
## #   stdev9901 <dbl>
```

Spatial Distribution



Normalized Difference Vegetation Index (NDVI)



Original paper - Diggle, et. al. (2007). *Spatial modelling and prediction of Loa loa risk: decision making under uncertainty*. Annals of Tropical Medicine and Parasitology, 101, 499-509.

- `no_exam` and `no_inf` - Collected between 1991 and 2001 by NGOs (original paper mentions 168 villages and 21,938 observations)
- `elev` - USGS gtopo30 (1km resolution)
- `mean9901` to `stdev9901` - aggregated data from 1999 to 2001 from the Flemish Institute for Technological Research (1 km resolution)

$$Y(s) \sim \text{Binom}(p(s), n(s))$$

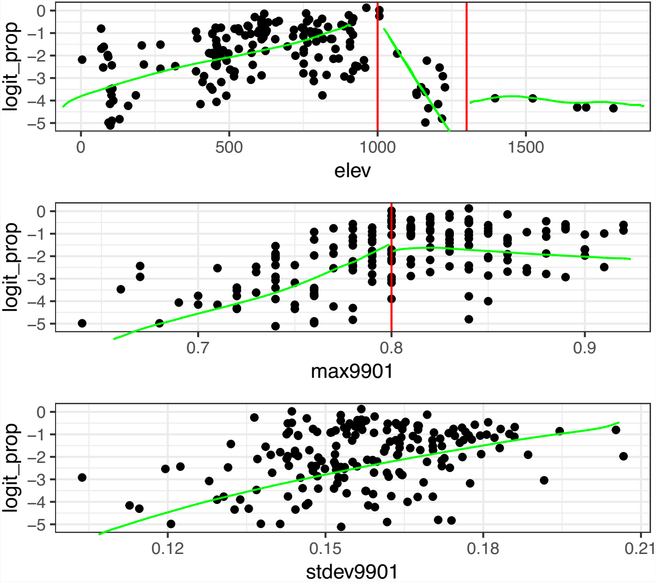
$$\begin{aligned}\log\left(\frac{p(s)}{1-p(s)}\right) = \alpha &+ f_1(\text{elev}(s)) \\ &+ f_2(\text{MAX.NDVI}(s)) \\ &+ f_3(\text{SD.NDVI}(s)) + w(s)\end{aligned}$$

where

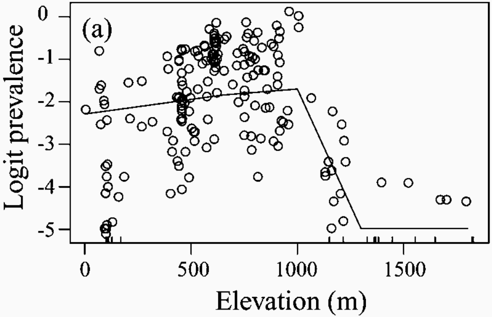
$$w(s) \sim \mathcal{N}(0, \Sigma)$$

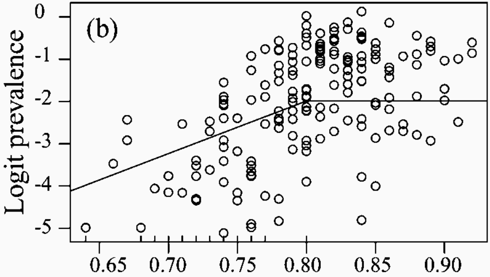
$$\{\Sigma\}_{ij} = \sigma^2 \exp(-d\,\phi)$$

EDA


logit_prop
0
−1
−2
−3
−4
−5
0
500
1000
1500
elev
max9901
0.7
0.8
0.9
stdev9901
0.12
0.15
0.18
0.21

Diggle's EDA


(a)
Logit prevalence
0
-1
-2
-3
-4
-5
0
500
1000
1500
Elevation (m)


(b)
Logit prevalence
0
-1
-2
-3
-4
-5
0.65
0.70
0.75
0.80
0.85
0.90

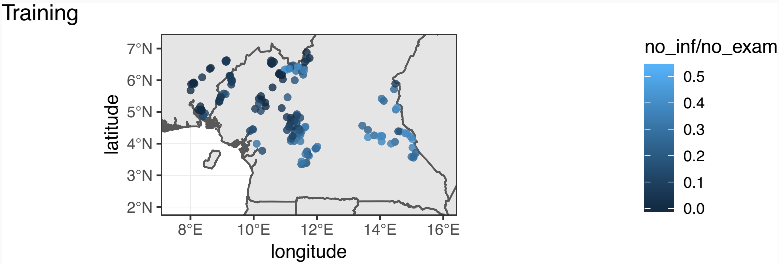
```r
loaloa = loaloa %>%
  mutate(
    elev_f = cut(elev, breaks=c(0,1000,1300,2000), dig.lab=5),
    max_f  = cut(max9901, breaks=c(0,0.8,1))
  )

loaloa %>% select(elev, elev_f, max9901, max_f)
## # A tibble: 197 x 4
##     elev elev_f   max9901 max_f
##    <int> <fct>      <dbl> <fct>
##  1   108 (0,1000]    0.69 (0,0.8]
##  2    99 (0,1000]    0.74 (0,0.8]
##  3   783 (0,1000]    0.79 (0,0.8]
##  4   104 (0,1000]    0.67 (0,0.8]
##  5   109 (0,1000]    0.85 (0.8,1]
##  6   909 (0,1000]    0.8  (0,0.8]
##  7   503 (0,1000]    0.78 (0,0.8]
##  8   103 (0,1000]    0.69 (0,0.8]
##  9   751 (0,1000]    0.8  (0,0.8]
## 10   268 (0,1000]    0.84 (0.8,1]
## # ... with 187 more rows
```

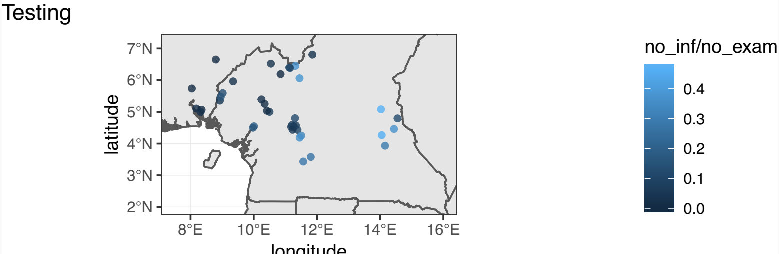
Model Matrix {.}

```
model.matrix(
  ~ elev:elev_f - 1,
  data = loaloa
) %>%
  as_data_frame()
## # A tibble: 197 x 3
##    `elev:elev_f(0,1000]` `elev:elev_f(1000,1300]` `elev:elev_f(1300,2000]`
##                    <dbl>                    <dbl>                    <dbl>
##  1                   108                        0                        0
##  2                    99                        0                        0
##  3                   783                        0                        0
##  4                   104                        0                        0
##  5                   109                        0                        0
##  6                   909                        0                        0
##  7                   503                        0                        0
##  8                   103                        0                        0
##  9                   751                        0                        0
## 10                   268                        0                        0
## # ... with 187 more rows
```

OOS Validation

```r
loaloa_test = loaloa %>% sample_frac(0.20)
loaloa = anti_join(loaloa, loaloa_test, quiet=TRUE)
```

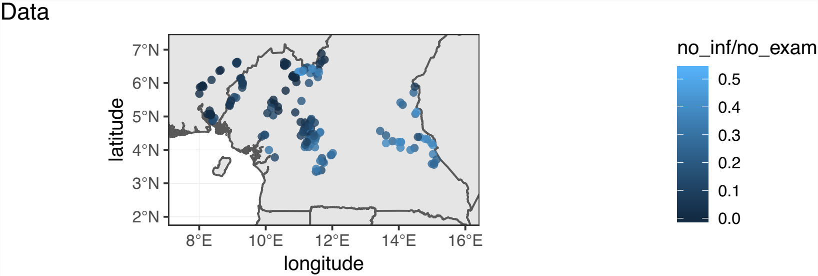


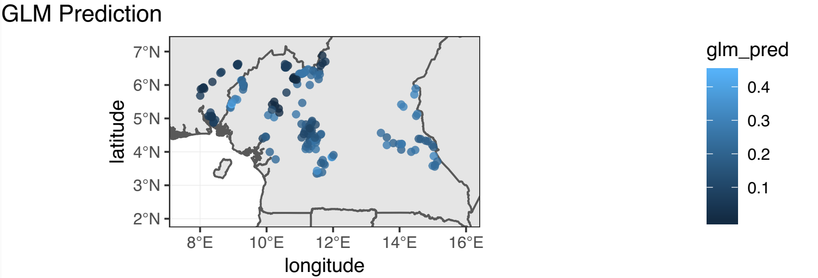


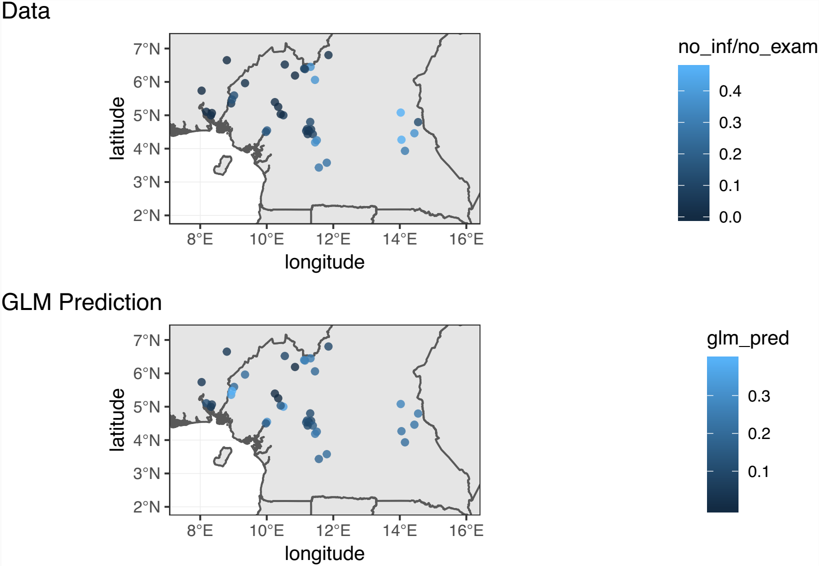
Model

```r
g = glm(no_inf/no_exam ~ elev:elev_f + max9901:max_f + stdev9901,
        data=loaloa, family=binomial, weights=loaloa$no_exam)

summary(g)
##
## Call:
## glm(formula = no_inf/no_exam ~ elev:elev_f + max9901:max_f +
##     stdev9901, family = binomial, data = loaloa, weights = loaloa$no_exam)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.9522  -2.5662  -0.4621   1.6720  10.1809
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -8.5735389  0.5333413 -16.075  < 2e-16 ***
## stdev9901               11.9141737  1.3070028   9.116  < 2e-16 ***
## elev:elev_f(0,1000]      0.0015951  0.0001018  15.660  < 2e-16 ***
## elev:elev_f(1000,1300]   0.0003343  0.0000953   3.507 0.000453 ***
## elev:elev_f(1300,2000] -0.0016964  0.0002513  -6.750 1.48e-11 ***
## max9901:max_f(0,0.8]     5.2697375  0.6918702   7.617 2.60e-14 ***
## max9901:max_f(0.8,1]     5.2632126  0.6362108   8.273  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3210.2  on 157  degrees of freedom
```

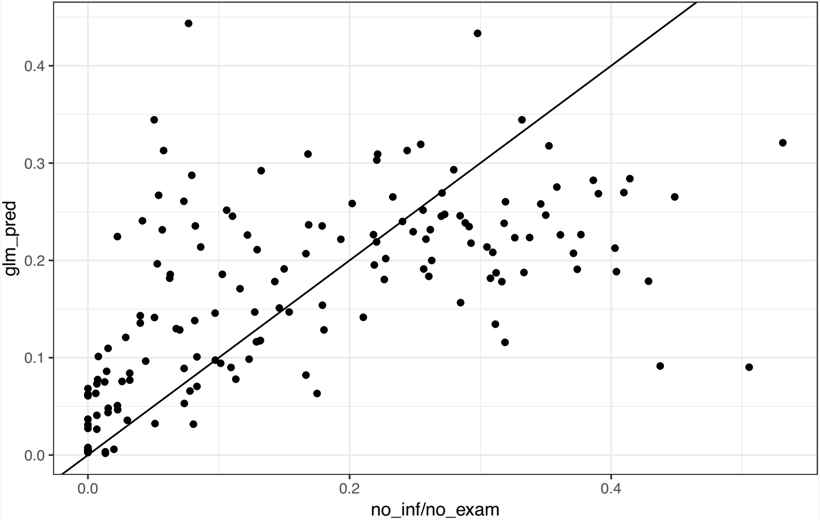
Predictions - Training



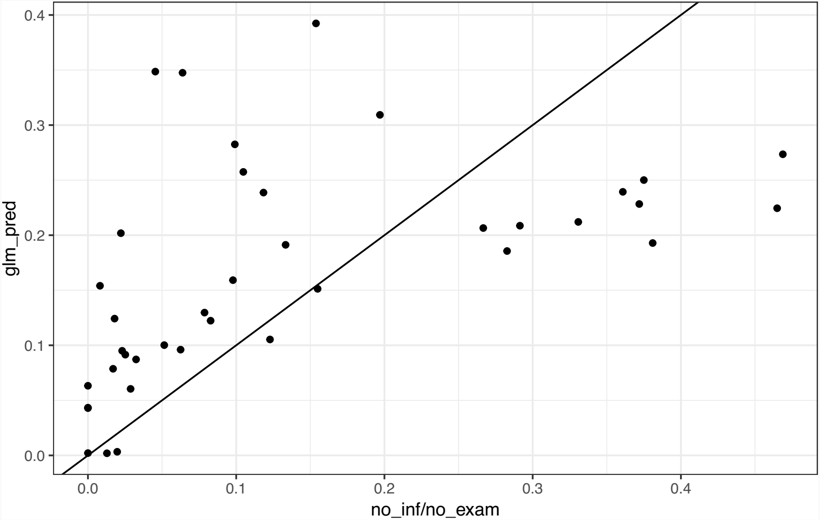



Data
7°N
6°N
5°N
4°N
3°N
2°N
latitude
8°E
10°E
12°E
14°E
16°E
longitude
no_inf/no_exam
0.4
0.3
0.2
0.1
0.0
GLM Prediction
7°N
6°N
5°N
4°N
3°N
2°N
latitude
8°E
10°E
12°E
14°E
16°E
longitude
glm_pred
0.3
0.2
0.1

Fit - Training

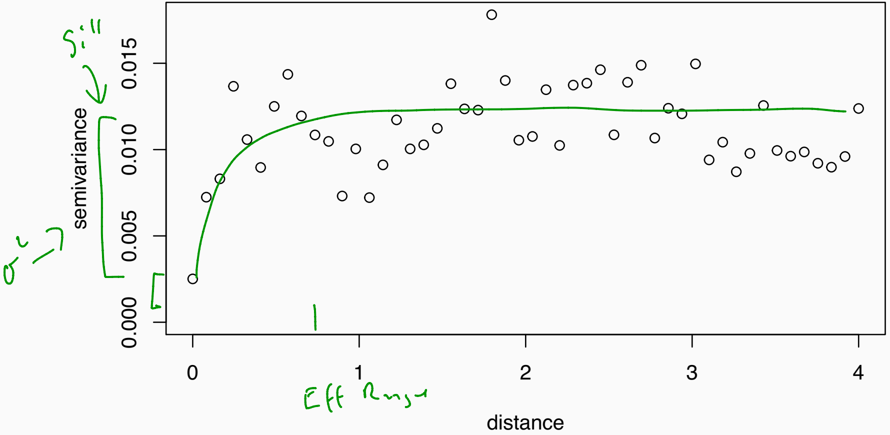


Fit - Testing



Spatial Structure?

```
geoR::variog(coords = cbind(loaloa$longitude, loaloa$latitude),
       data = loaloa$prop - loaloa$glm_pred,
       uvec = seq(0, 4, length.out = 50)) %>% plot()
## variog: computing omnidirectional variogram
```



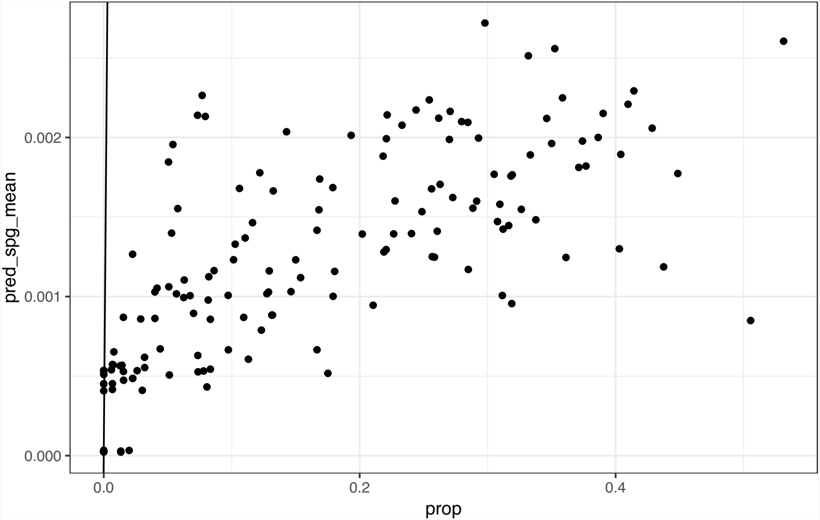
spBayes GLM Model

```
spg = spBayes::spGLM(
  no_inf/no_exam ~ elev:elev_f + max9901:max_f + stdev9901,
  data=loaloa, family="binomial", weights=loaloa$no_exam,
  coords=cbind(loaloa$longitude, loaloa$latitude),
  cov.model="exponential", n.samples=20000,
  starting=list(beta=rep(0,7), phi=3, sigma.sq=1, w=0),
  priors=list(phi.unif=c(0.1, 10), sigma.sq.ig=c(2, 2)),
  amcmc=list(n.batch=1000, batch.length=20, accept.rate=0.43))
save(spg, loaloa, file="loaloa.Rdata")
```

```
spg$p.beta.theta.samples %>%
  post_summary() %>%
  knitr::kable(digits=5)
```

param	post_mean	post_med	post_lower	post_upper
(Intercept)	-7.62467	-7.10607	-15.33201	-1.56786
stdev9901	1.77896	-0.26705	-19.15846	24.59887
elev:elev_f(0,1000]	0.00010	0.00065	-0.00780	0.00316
elev:elev_f(1000,1300]	-0.00059	-0.00035	-0.00471	0.00176
elev:elev_f(1300,2000]	-0.01448	-0.01064	-0.04942	-0.00030
max9901:max_f(0,0.8]	0.08517	-0.78200	-6.96111	9.06059
max9901:max_f(0.8,1]	0.69926	-0.25813	-5.79400	9.08833
sigma.sq	0.45277	0.39071	0.14322	1.17856
phi	2.12385	1.44856	0.12026	8.46872

Prediction

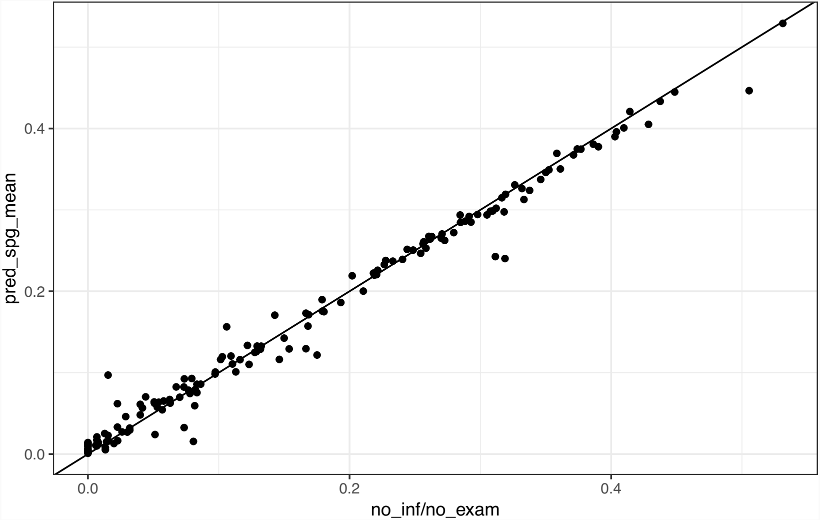


spBayes GLM Model - Fixed?

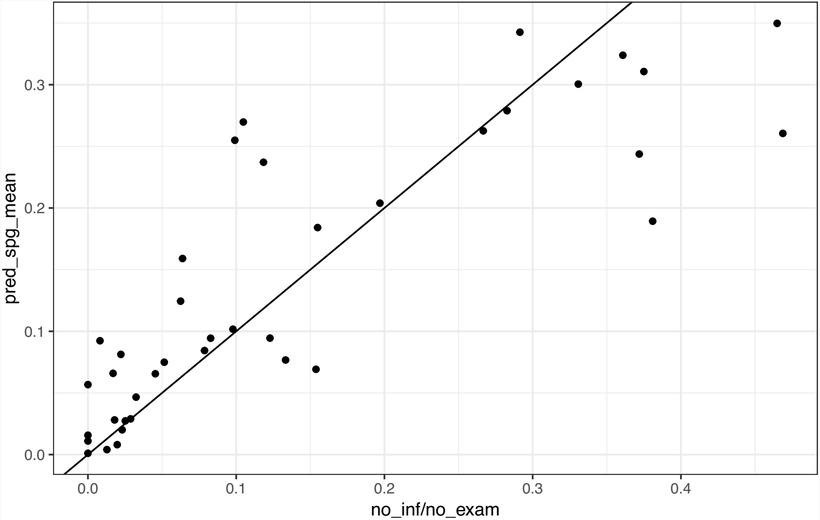
```
spg_fix = spBayes::spGLM(
  no_inf ~ elev:elev_f + max9901:max_f + stdev9901,

  data=loaloa, family="binomial", weights=loaloa$no_exam,

  coords=cbind(loaloa$longitude, loaloa$latitude),

  cov.model="exponential", n.samples=20000,

  starting=list(beta=rep(0,7), phi=3, sigma.sq=1, w=0),

  priors=list(phi.unif=c(0.1, 10), sigma.sq.ig=c(2, 2)),

  amcmc=list(n.batch=1000, batch.length=20, accept.rate=0.43)
)

save(spg_fix, loaloa, file="loaloa_fix.Rdata")
```

param	post_mean	post_med	post_lower	post_upper
(Intercept)	-3.14223	-3.43877	-4.38140	-1.01108
stdev9901	1.88811	1.02957	-5.28818	9.04674
elev:elev_f(0,1000]	0.00036	0.00048	-0.00069	0.00114
elev:elev_f(1000,1300]	-0.00036	-0.00031	-0.00127	0.00039
elev:elev_f(1300,2000]	-0.00209	-0.00206	-0.00310	-0.00131
max9901:max_f(0,0.8]	0.74129	0.55728	-0.98971	2.78417
max9901:max_f(0.8,1]	1.15469	0.92740	-0.18829	2.89406
sigma.sq	1.26052	1.21204	0.32891	2.36502
phi	2.51439	2.38441	1.08064	4.86766

Fit - Training



Fit - Testing



Diggle's Predictive Surface

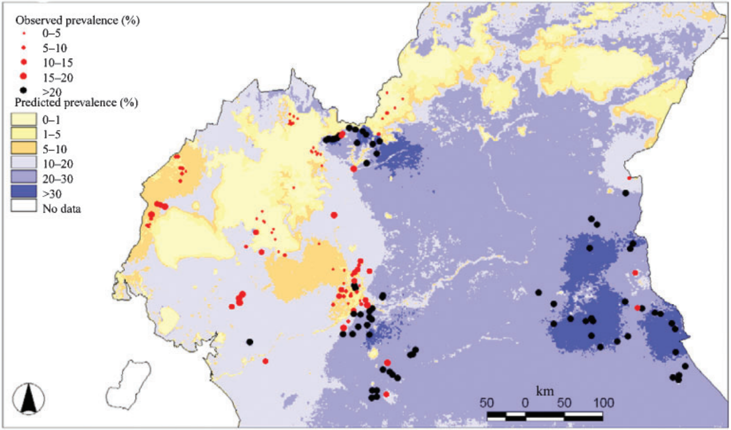
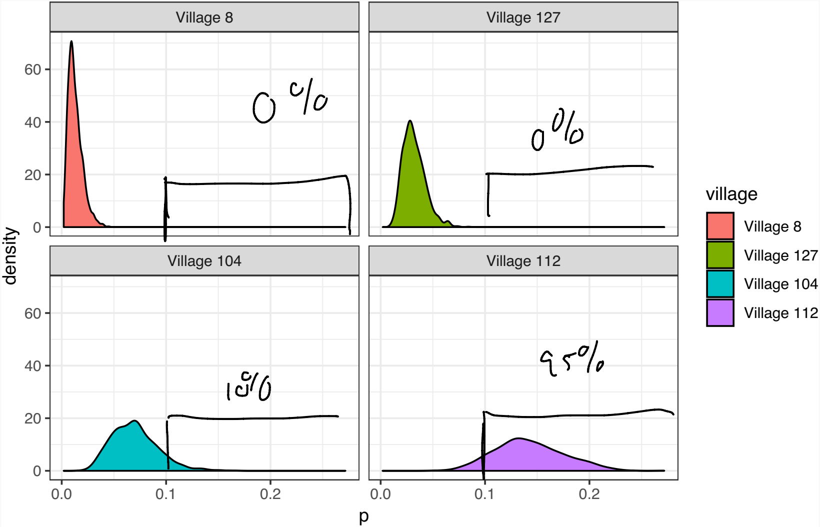


FIG. 2. Point estimates of the prevalence of *Loa loa* microfilaraemia, over-laid with the prevalences observed in field studies.

Exceedance Probability - Posterior Summary



Exceedance Probability Predictive Surface

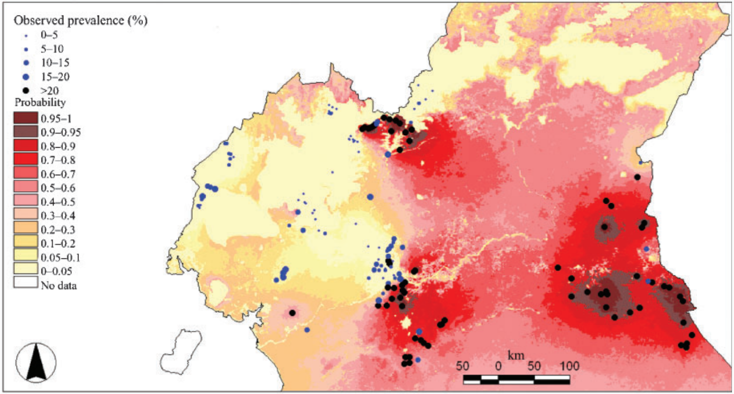


FIG. 4. A probability contour map, indicating the probability that the prevalence of *Loa loa* microfilaraemia in each area exceeds 20%, over-laid with the prevalences observed in field studies.

Spatial Assignment of Migratory Birds

Background

Using intrinsic markers (genetic and isotopic signals) for the purpose of inferring migratory connectivity.

- Existing methods are too coarse for most applications
- Large amounts of data are available (>150,000 feather samples from >500 species)
- Genetic assignment methods are based on Wasser, et al. (2004)
- Isotopic assignment methods are based on Wunder, et al. (2005)

Data - DNA microsatellites and $\delta^2\mathrm{H}$

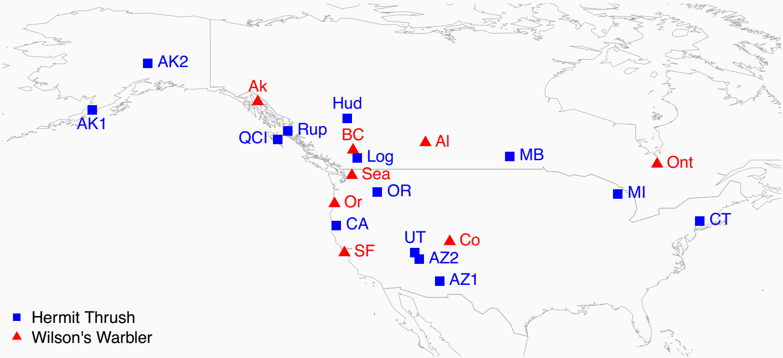
Hermit Thrush (*Catharus guttatus*)

- 138 individuals
- 14 locations
- 6 loci
- 9-27 alleles / locus

Wilson's Warbler (*Wilsonia pusilla*)

- 163 individuals
- 8 locations
- 9 loci
- 15-31 alleles / locus

Sampling Locations



Allele Frequency Model

For the allele i, from locus l, at location k

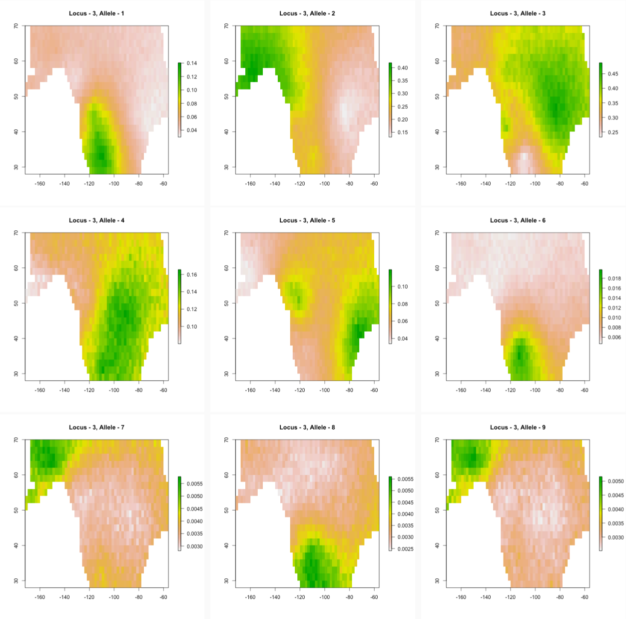
$$\mathbf{y}_{\cdot lk}|\boldsymbol{\Theta} \sim \overset{\text{MN}}{\cancel{\mathscr{M}}}\left(\textstyle\sum_i y_{ilk},\ \mathbf{f}_{\cdot lk}\right)$$

$$f_{ilk} = \frac{\exp(\Theta_{ilk})}{\sum_i \exp(\Theta_{ilk})}$$

$$\boldsymbol{\Theta}_{il}|\boldsymbol{\alpha}, \boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_{il},\ \boldsymbol{\Sigma})$$

$$\{\Sigma\}_{ij} = \sigma^2\ \exp\Big(-\left(\{d\}_{ij}\, r\right)^{\psi}\Big) + \sigma_n^2\ 1_{i=j}$$

Predictions by Allele (Locus 3)



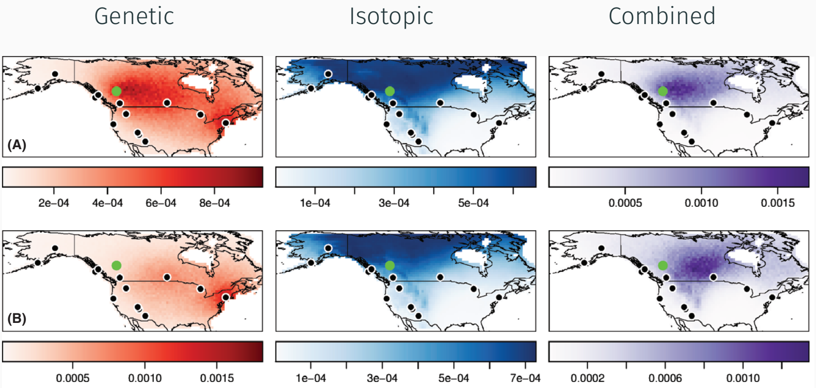
Assignment model assuming Hardy-Weinberg equilibrium and allowing for genotyping (δ) and single amplification (γ) errors.

$$P(S_G|\mathbf{f},k) = \prod_l P(i_l, j_l|\mathbf{f},k)$$

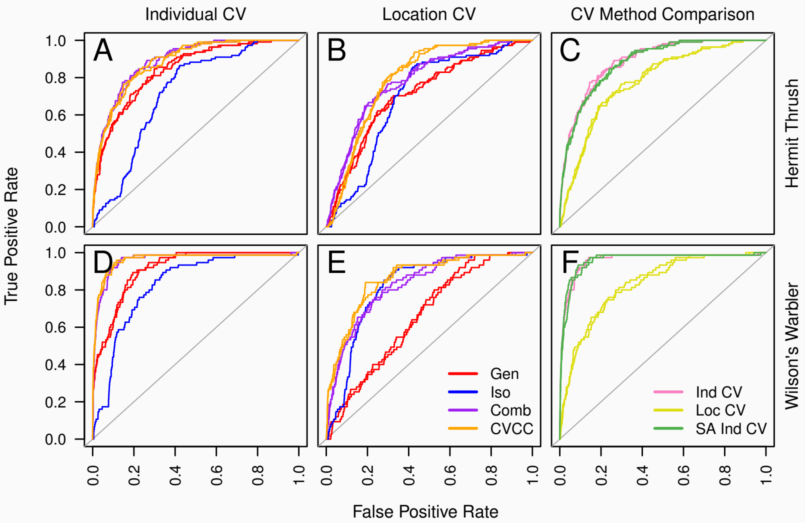
$$P(i_l, j_l|\mathbf{f},k) = \begin{cases} \gamma P(i_l|\mathbf{f},k) + (1-\gamma)P(i_l|\tilde{\mathbf{f}},k)^2 & \text{if } i = j \\ (1-\gamma)P(i_l|\mathbf{f},k)P(j_l|\mathbf{f},k) & \text{if } i \neq j \end{cases}$$

$$P(i_l|\mathbf{f},k) = (1-\delta)f_{lik} + \delta/m_l$$

Combined Model



Model Assessment



Migratory Connectivity