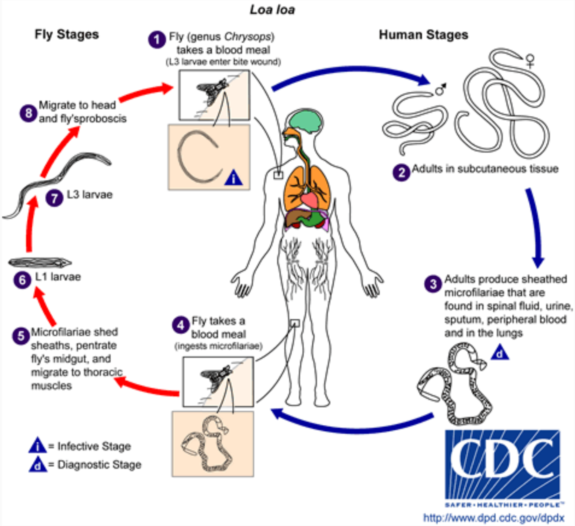
Lecture 20

Point referenced data (pt. 2)

Colin Rundel
04/05/2017

Loa Loa Example

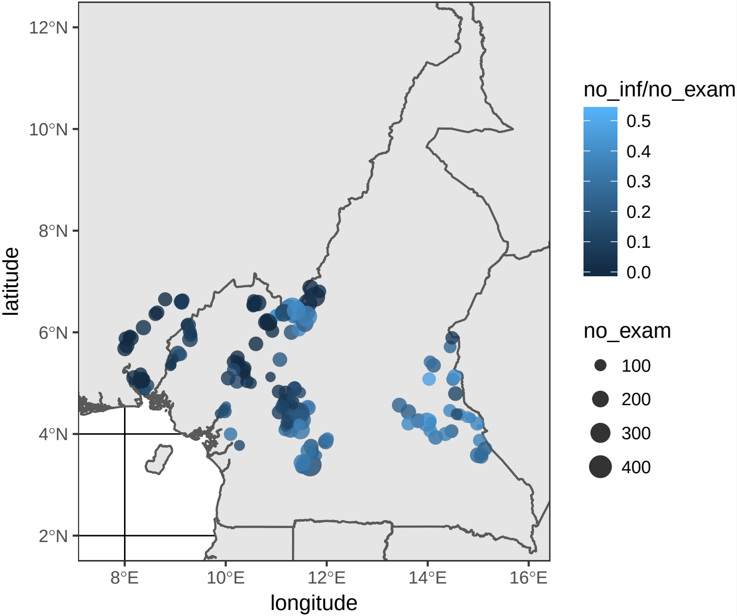
Loa Loa


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
d
i = Infective Stage
d = Diagnostic Stage
CDC
SAFER • HEALTHIER • PEOPLE™
http://www.dpd.cdc.gov/dpdx

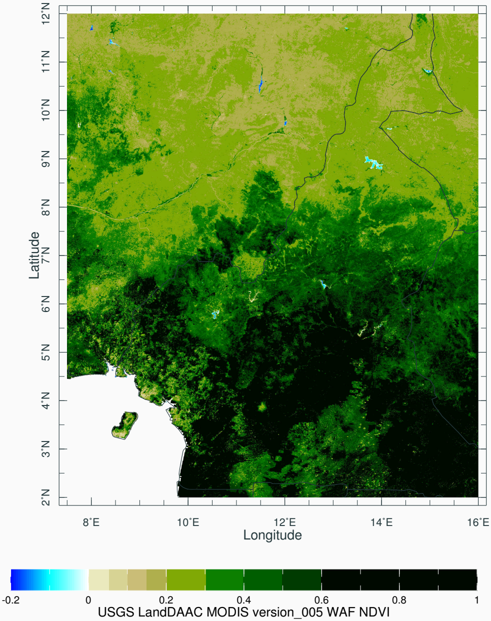
Data

```
loaloa = tbl_df(PrevMap::loaloa) %>% setNames(., tolower(names(.)))

loaloa
## # A tibble: 197 x 11
##      row villcode longitude latitude no_exam no_inf elevation mean9901
##    <int>    <int>     <dbl>    <dbl>   <int>  <int>     <int>    <dbl>
##  1     1      214      8.04     5.74     162      0       108    0.439
##  2     2      215      8.00     5.68     167      1        99    0.426
##  3     3      118      8.91     5.35      88      5       783    0.491
##  4     4      219      8.10     5.92      62      5       104    0.432
##  5     5      212      8.18     5.10     167      3       109    0.415
##  6     6      116      8.93     5.36      66      3       909    0.436
##  7     7       16     11.4      4.88     163     11       503    0.502
##  8     8      217      8.07     5.90      83      0       103    0.373
##  9     9      112      9.02     5.59      30      4       751    0.481
## 10    10      104      9.31     6.00      57      4       268    0.487
## # ... with 187 more rows, and 3 more variables: max9901 <dbl>,
## #   min9901 <dbl>, stdev9901 <dbl>
```

Spatial Distribution



Normalized Difference Vegetation Index (NDVI)



Paper / Data summary

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)
- `elevation` - USGS gtopo30 (1km resolution)
- `mean9901` to `stdev9901` - aggregated data from 1999 to 2001 from the Flemish Institute for Technological Research (1 km resolution)

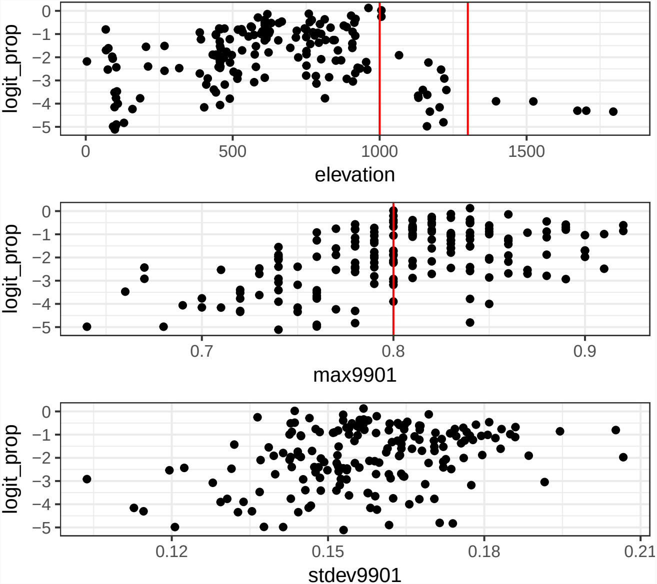
Diggle’s Model

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

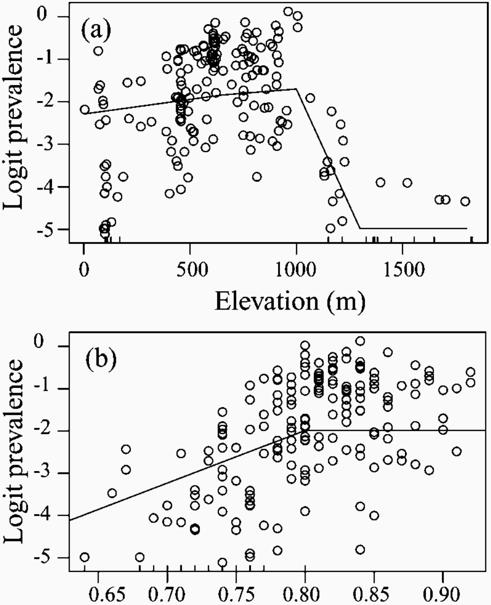
where

$$w(s) \sim \mathcal{N}(0, \Sigma)$$
$$\{\Sigma\}_{ij} = \sigma^2\ \exp(-d\,\phi)$$

EDA


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

Diggle's EDA

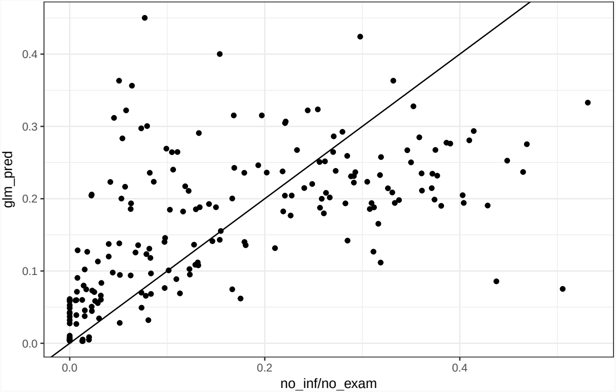

(a)
Logit prevalence
0
-1
-2
-3
-4
-5
0
500
1000
1500
Elevation (m)
(b)
Logit prevalence
0
-1
-2
-3
-4
-5
0.65
0.70
0.75
0.80
0.85
0.90

Model EDA

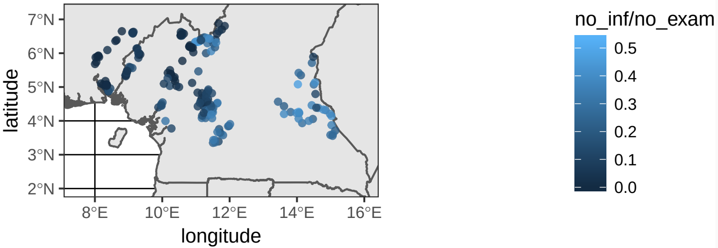
```
loaloa = loaloa %>%
  mutate(elev_factor = cut(elevation, breaks=c(0,1000,1300,2000), dig.lab=5),
         max_factor  = cut(max9901, breaks=c(0,0.8,1)))

g = glm(no_inf/no_exam ~ elevation:elev_factor + max9901:max_factor + stdev9901,
        data=loaloa, family=binomial, weights=loaloa$no_exam)

summary(g)
##
## Call:
## glm(formula = no_inf/no_exam ~ elevation:elev_factor + max9901:max_factor +
##     stdev9901, family = binomial, data = loaloa, weights = loaloa$no_exam)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.1434  -2.5887  -0.8993   1.6375  10.9052
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                         -8.343e+00  4.825e-01 -17.291  < 2e-16
## stdev9901                            8.781e+00  1.205e+00   7.288 3.14e-13
## elevation:elev_factor(0,1000]        1.606e-03  8.749e-05  18.358  < 2e-16
## elevation:elev_factor(1000,1300]     1.631e-04  8.792e-05   1.855   0.0636
## elevation:elev_factor(1300,2000]    -1.432e-03  1.887e-04  -7.588 3.25e-14
## max9901:max_factor(0,0.8]            5.511e+00  6.299e-01   8.749  < 2e-16
## max9901:max_factor(0.8,1]            5.626e+00  5.793e-01   9.711  < 2e-16
##
## (Intercept)                         ***
```

Fit

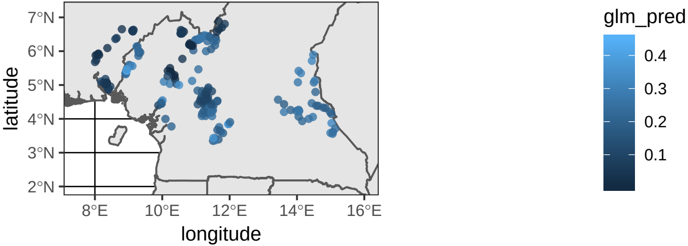
```
loaloa = loaloa %>%
  mutate(glm_pred = predict(g, type="response"))

ggplot(loaloa, aes(x=no_inf/no_exam, y=glm_pred)) +
  geom_point() +
  geom_abline(slope = 1, intercept = 0)
```



Data

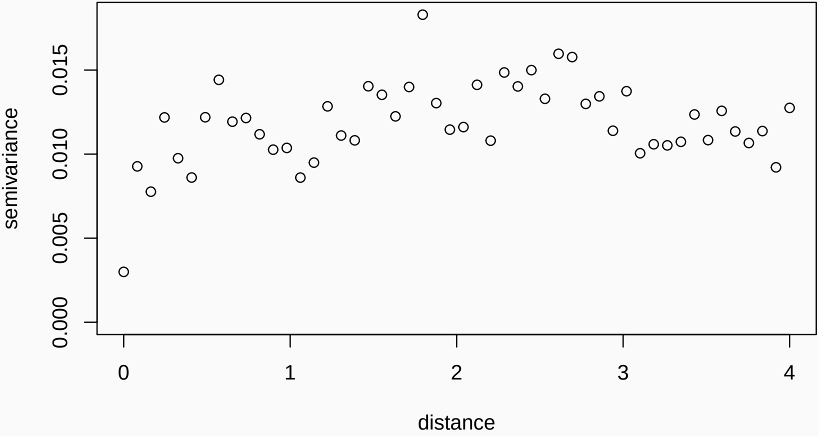


GLM Prediction



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
```



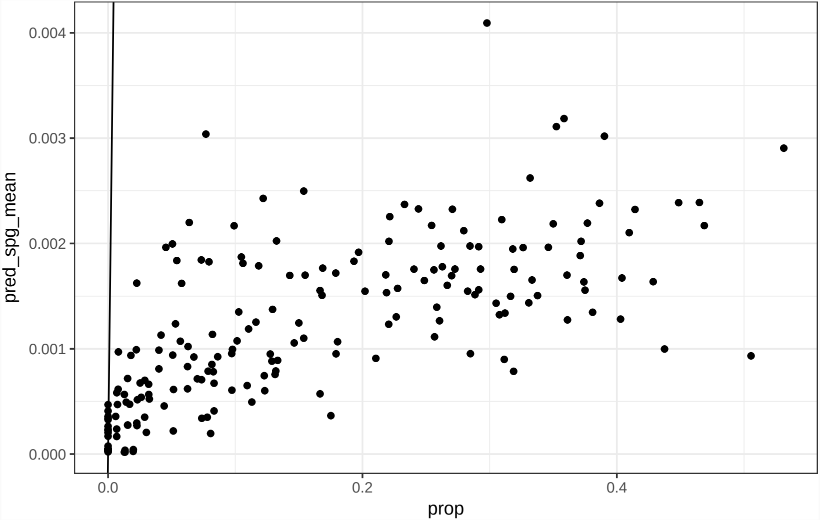
spBayes GLM Model

```
spg = spBayes::spGLM(
  no_inf/no_exam ~ elevation:elev_factor + max9901:max_factor + 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”)
```

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

param	post_mean	post_med	post_lower	post_upper
(Intercept)	-12.69885	-11.61326	-21.65388	-6.96361
stdev9901	9.24231	9.15244	-14.48649	29.76058
elevation:elev_factor(0,1000]	0.00048	0.00077	-0.00474	0.00291
elevation:elev_factor(1000,1300]	-0.00048	-0.00032	-0.00359	0.00169
elevation:elev_factor(1300,2000]	-0.00814	-0.00581	-0.02900	0.00004
max9901:max_factor(0,0.8]	4.87762	3.99492	-2.93030	15.63246
max9901:max_factor(0.8,1]	5.08690	4.44632	-2.18626	14.89011
sigma.sq	0.38088	0.34626	0.12793	0.88673
phi	6.22996	5.18205	0.69584	18.67107

Prediction

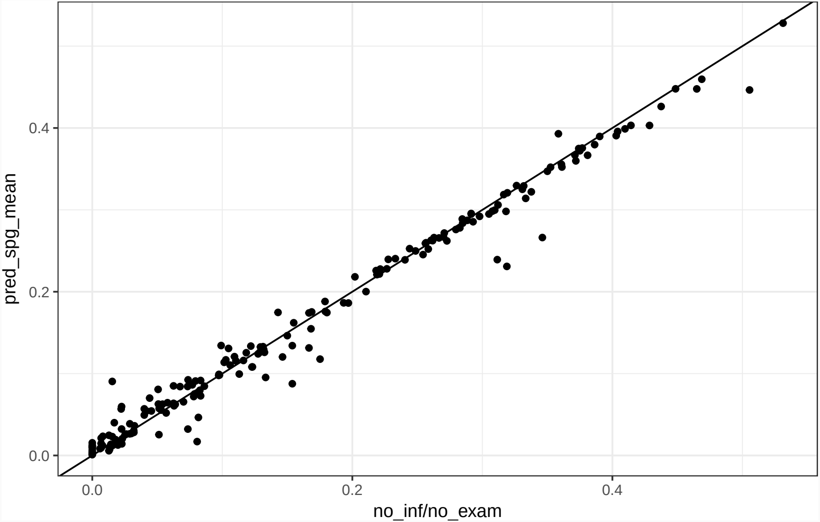


spBayes GLM Model - Fixed?

```
spg_fix = spBayes::spGLM(
  no_inf ~ elevation:elev_factor + max9901:max_factor + 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)	-2.66090	-2.13138	-6.31576	-0.80487
stdev9901	-0.12840	-0.41947	-5.86766	8.58835
elevation:elev_factor(0,1000]	0.00023	0.00024	-0.00051	0.00086
elevation:elev_factor(1000,1300]	-0.00054	-0.00055	-0.00128	0.00020
elevation:elev_factor(1300,2000]	-0.00204	-0.00200	-0.00285	-0.00127
max9901:max_factor(0,0.8]	0.88041	0.90550	-1.03795	3.63477
max9901:max_factor(0.8,1]	1.28673	1.13796	-0.26884	3.83860
sigma.sq	1.47552	1.39146	0.43359	3.05883
phi	2.22372	2.09524	0.86456	4.14663

Fit



Diggle’s Predictive Surface

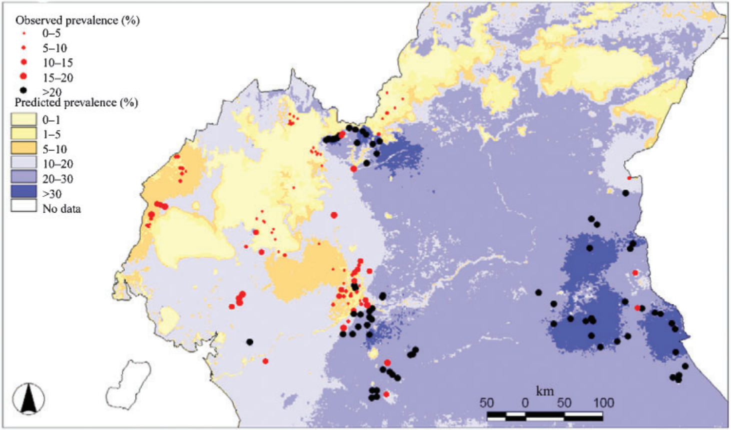
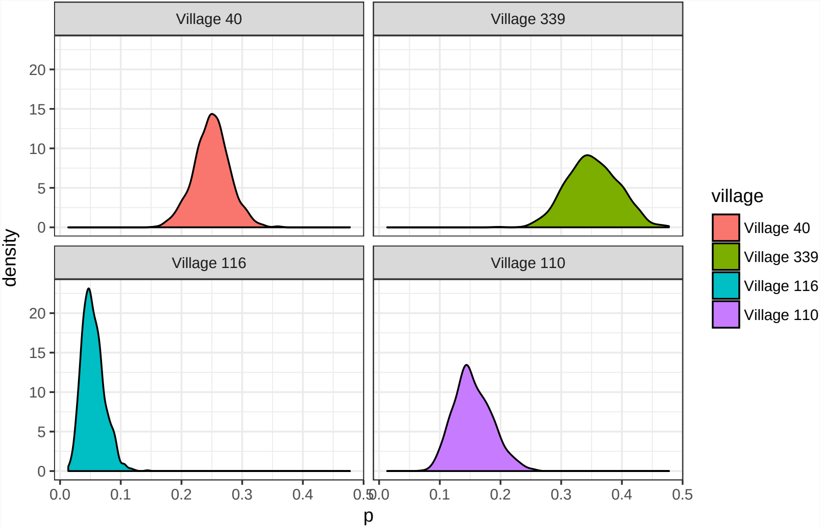


FIG. 2. Point estimates of the prevalence of *Loa loa* microfilaraemia, over-laid with the prevalences observed in field studies.

Exceedance Probability - Posterior Summary



Exceedance Probability Predictive Surface

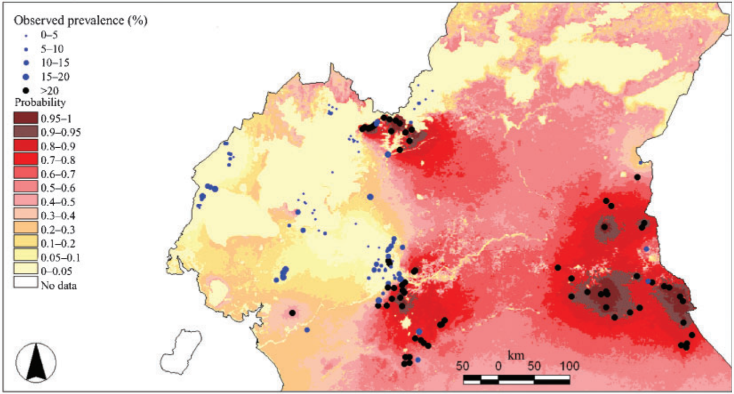


FIG. 4. A probability contour map, indicating the probability that the prevalence of *Loa loa* microfilaraemia in each area exceeds 20%, over-laid with the prevalences observed in field studies.

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

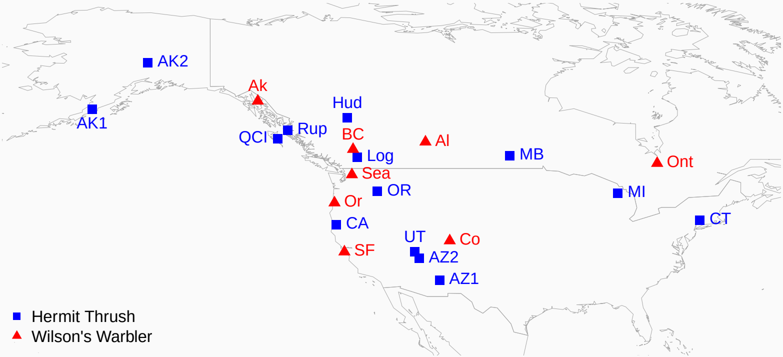
Photo by John Ingram

Wilson's Warbler (*Wilsonia pusilla*)

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

© Glenn Bartley

Sampling Locations



Allele Frequency Model

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

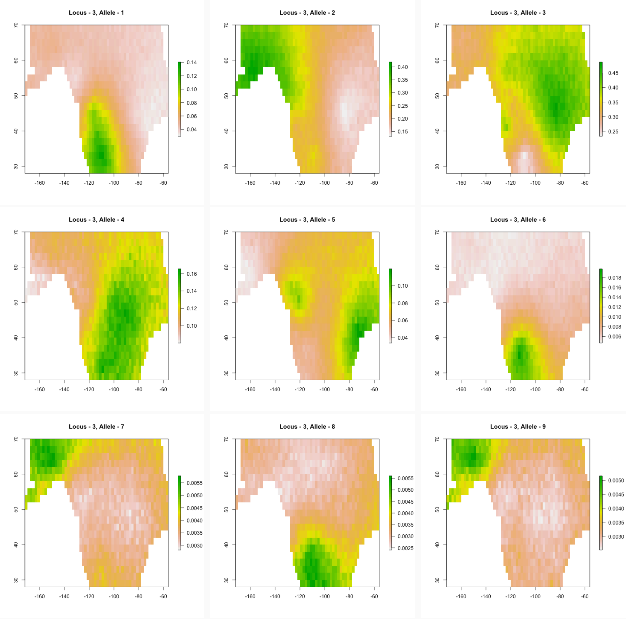
$$\mathbf{y}_{\cdot lk} | \boldsymbol{\Theta} \sim \mathcal{N}\left(\sum_i y_{ilk},\, \mathbf{f}_{\cdot lk}\right)$$

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

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

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

Predictions by Allele (Locus 3)



Genetic Assignment Model

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

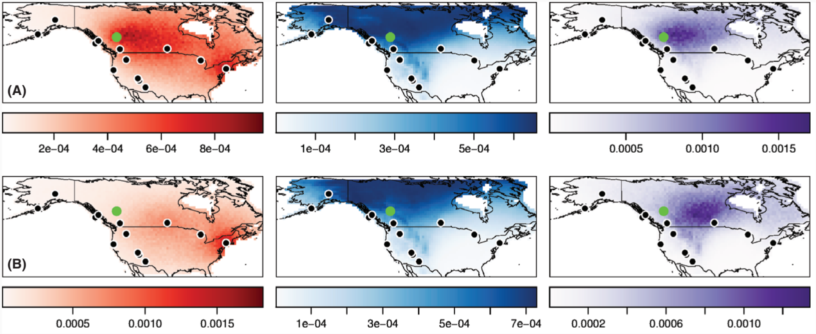
$$P(S_G|\mathbf{f},k) = \prod_l P(i_l, j_l|\mathbf{f},k)$$

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

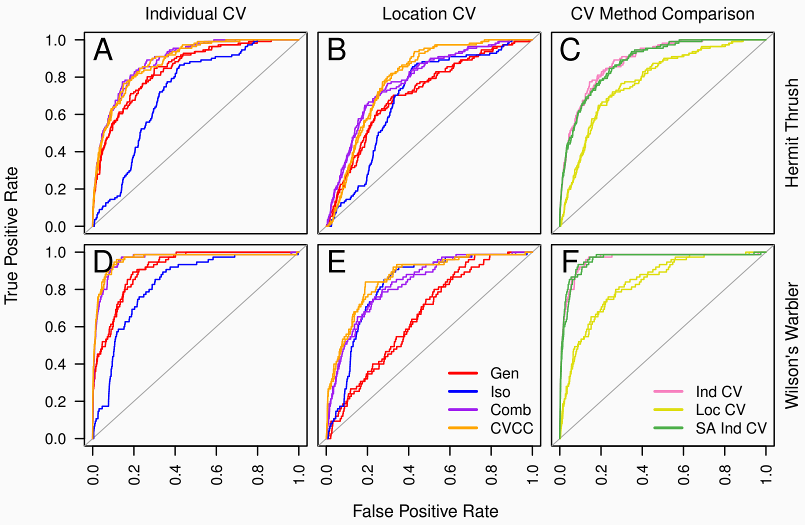
$$P(i_l|\mathbf{f},k) = (1-\delta)f_{lik} + \delta/m_l$$

Combined Model

Genetic Isotopic Combined



Model Assessment



Migratory Connectivity