

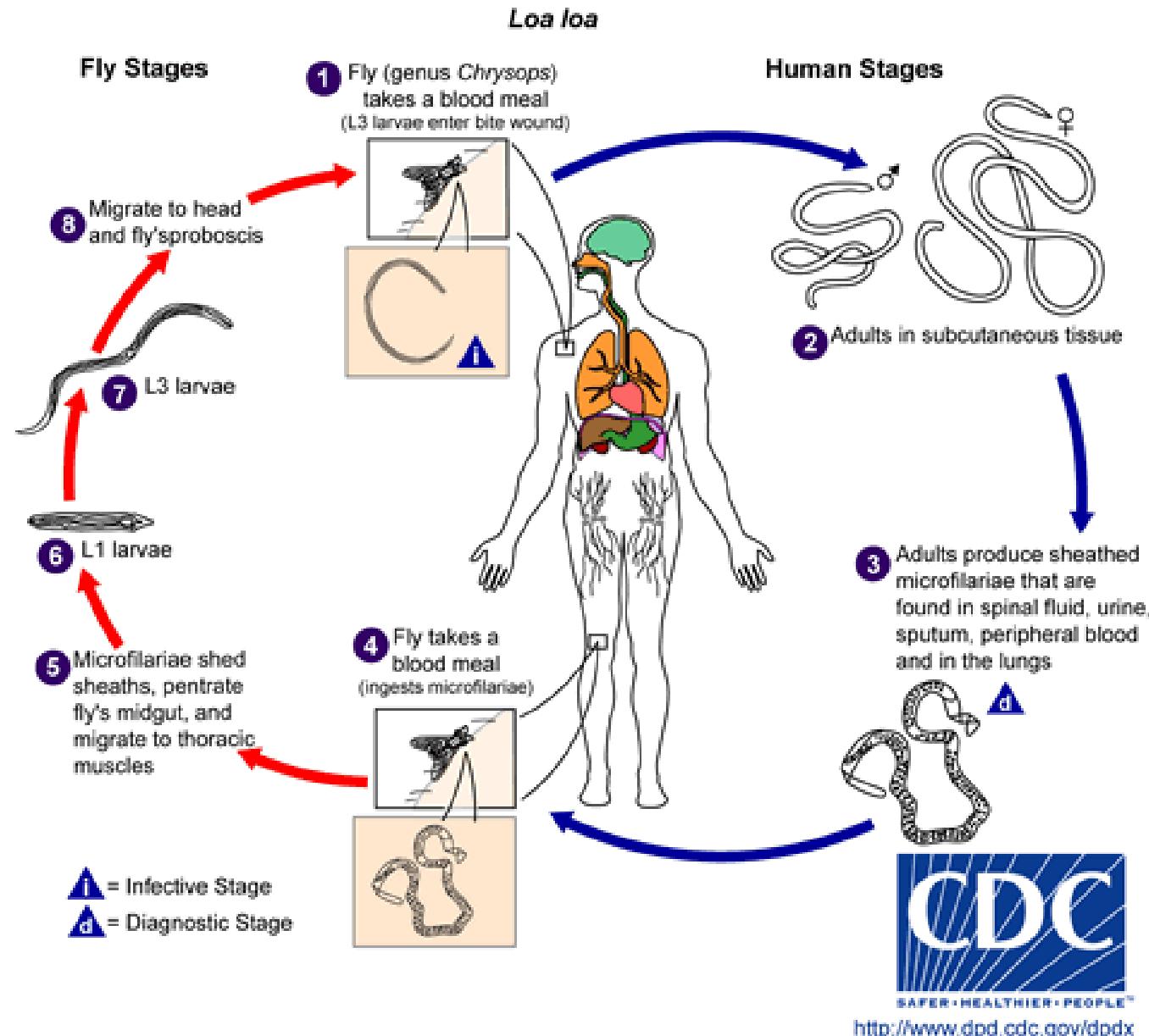
# Point referenced data (pt. 2)

Lecture 21

Dr. Colin Rundel

# Loa Loa Example

# Loa Loa



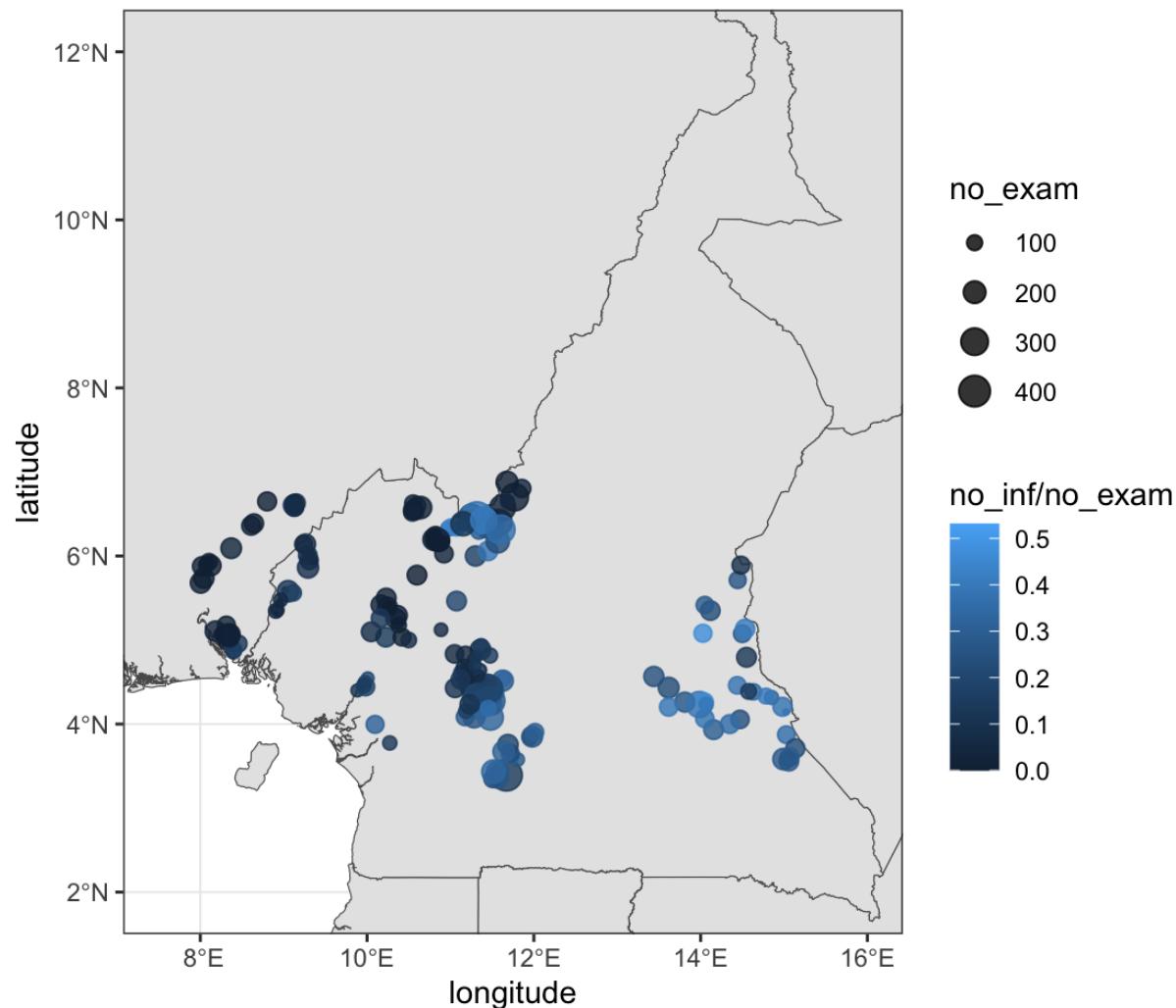
# Data

```

1 loaloa = tbl_df(PrevMap:::loaloa) %>%
2   setNames(., tolower(names(.))) %>%
3   rename(elev=elevation)
4
5 loaloa

# A tibble: 197 × 11
# ... with 187 more rows, 2 more variables: min9901 <dbl>,
#   stdev9901 <dbl>, and abbreviated variable names `longitude`,
#   `latitude`, `mean9901`
# ... with 187 more rows, 2 more variables: min9901 <dbl>,
#   stdev9901 <dbl>, and abbreviated variable names `longitude`,
#   `latitude`, `mean9901`
```

# 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)
- `elev` - 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

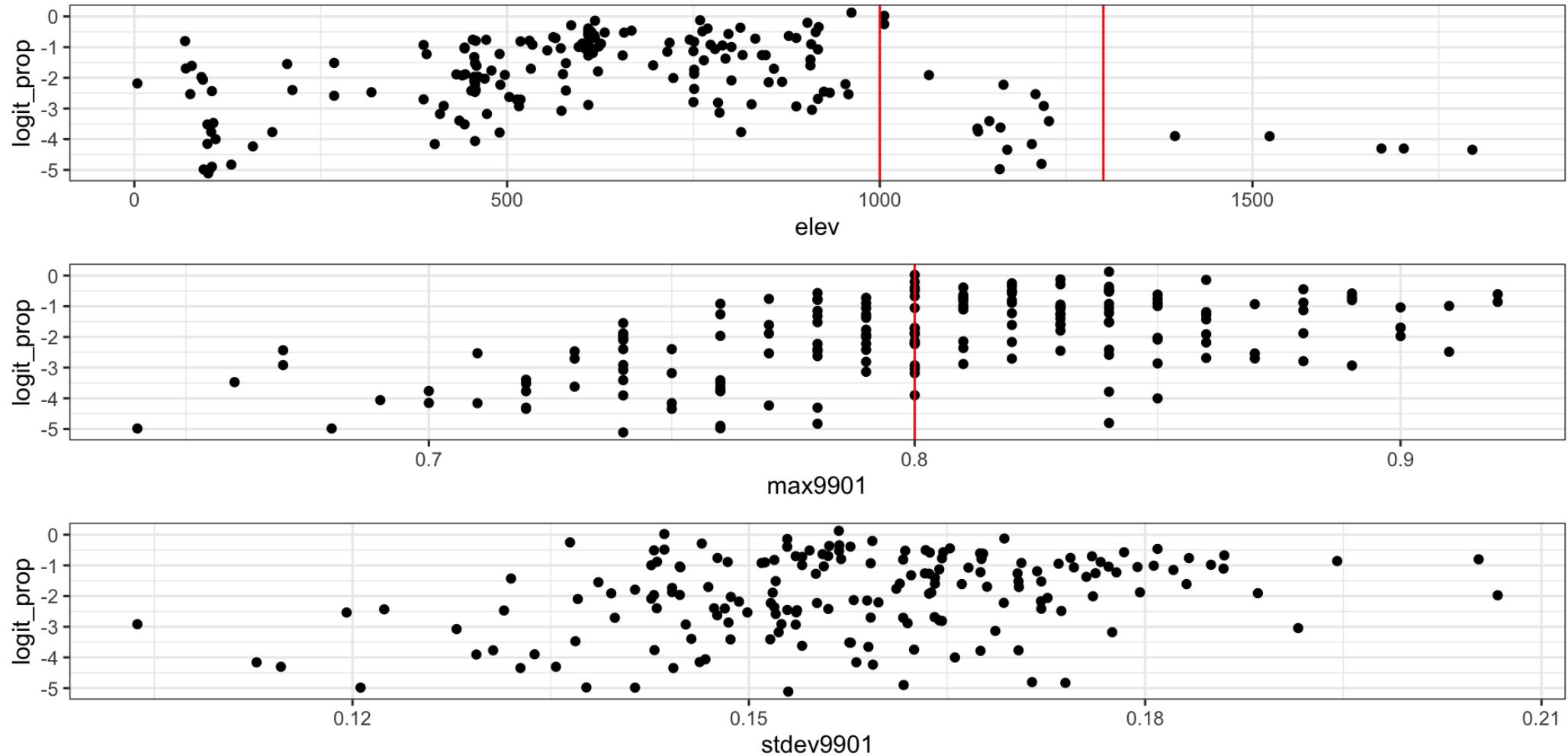
$$\begin{aligned}\log\left(\frac{p(s)}{1 - p(s)}\right) = & \alpha + f_1(\text{elev}(s)) \\ & + f_2(\text{MAX.NDVI}(s)) \\ & + f_3(\text{SD.NDVI}(s)) + w(s)\end{aligned}$$

where

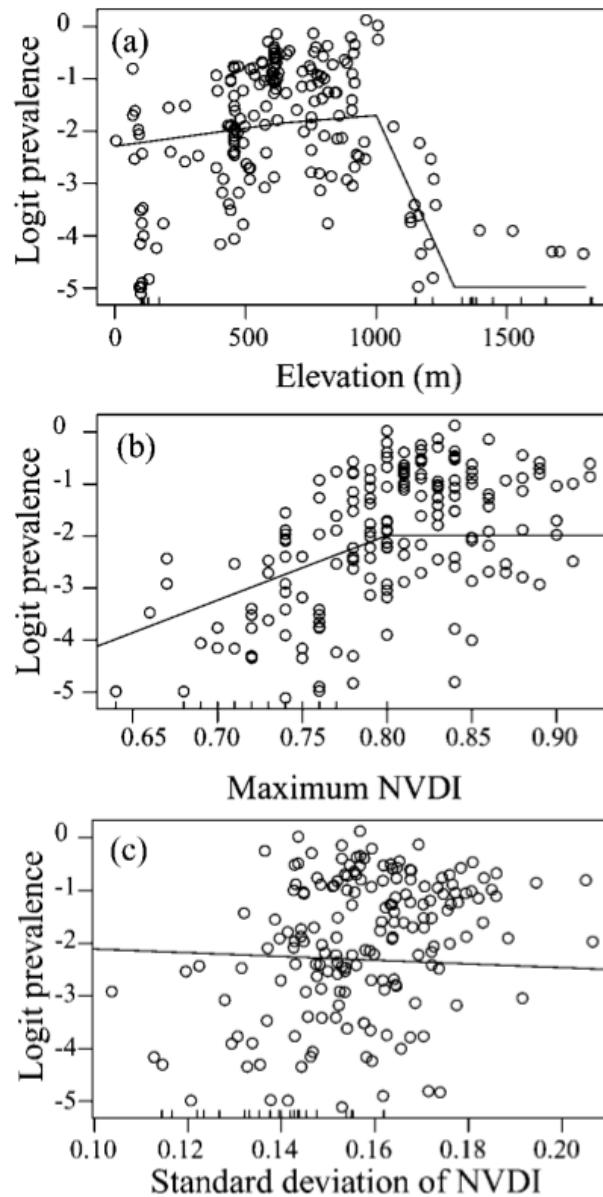
$$w(s) \sim (0, \Sigma)$$

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

# EDA



# Diggle's EDA



# Feature engineering

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

```
# A tibble: 197 × 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
```

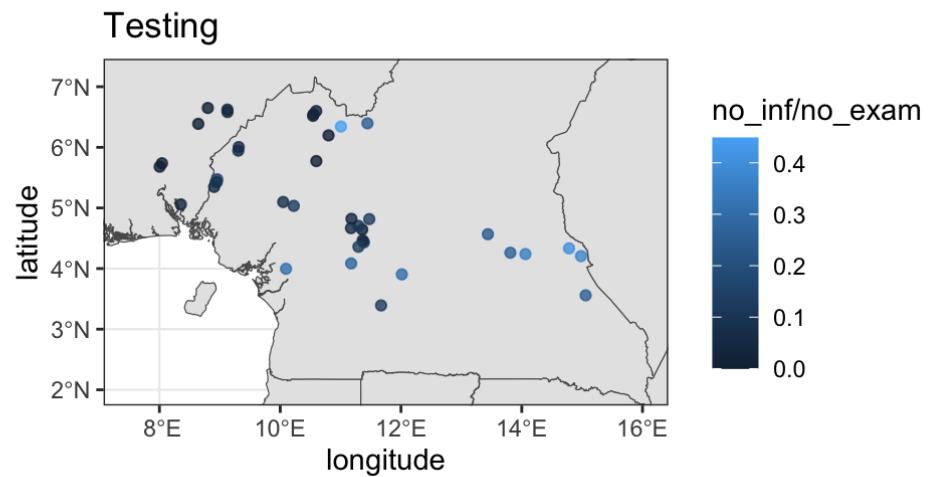
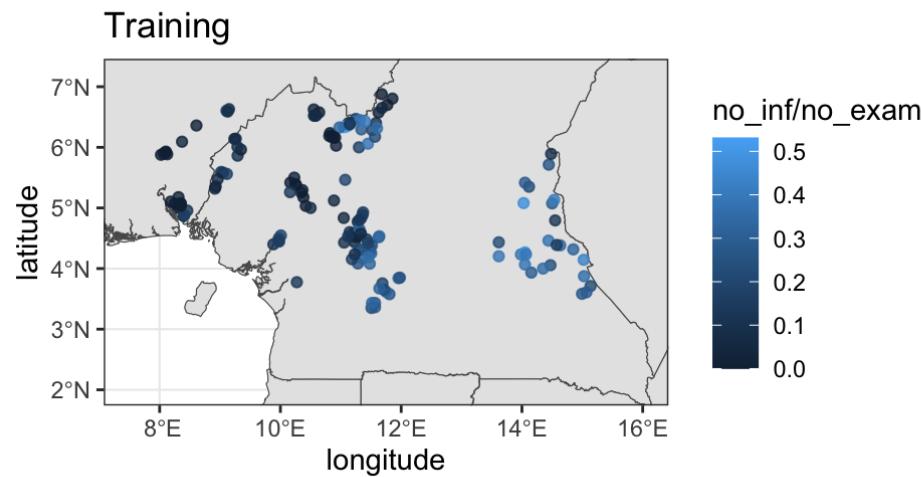
# Model Matrix

```
1 model.matrix(  
2 ~ elev:elev_f - 1,  
3 data = loaloa  
4 ) %>%  
5 as_tibble()
```

```
#> # A tibble: 197 × 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, and abbreviated variable name
#> #   `¹`elev:elev_f(1300,2000)`
```

# OOS Validation

```
1 set.seed(12345)
2 loaloa_test = loaloa %>% slice_sample(prop=0.20)
3 loaloa = anti_join(loaloa, loaloa_test, quiet=TRUE)
```



# Model

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

Call:

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

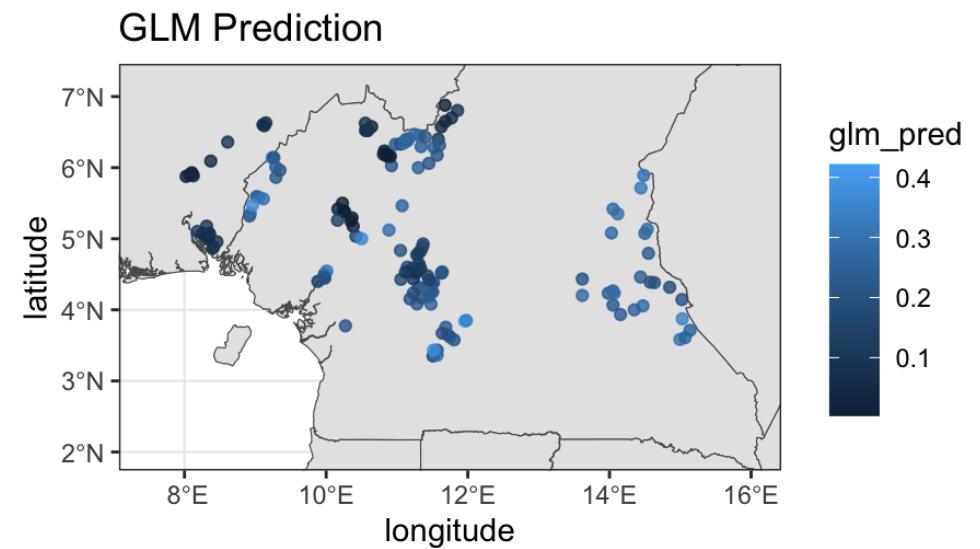
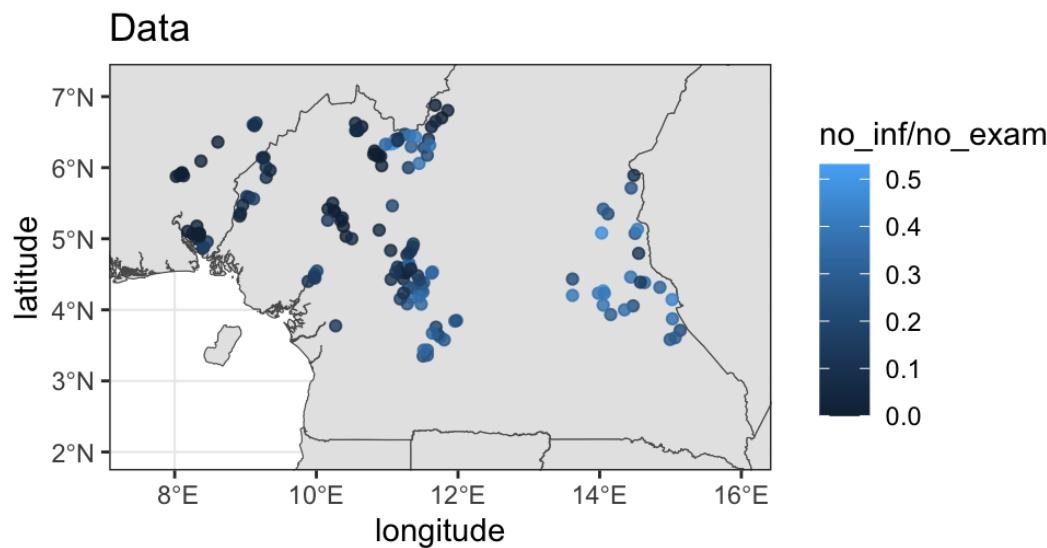
Deviance Residuals:

Min	1Q	Median	3Q	Max
-7.2205	-2.4954	-0.7776	1.6020	9.9667

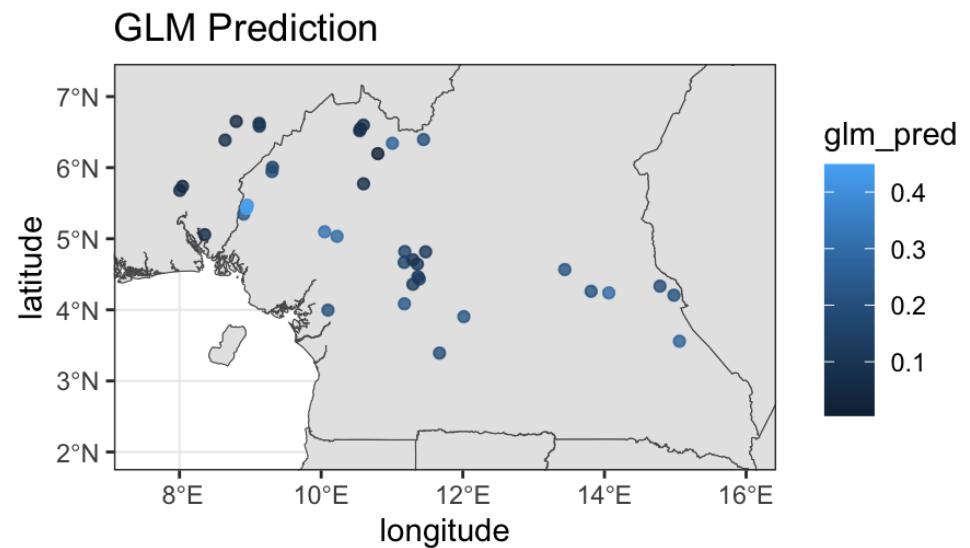
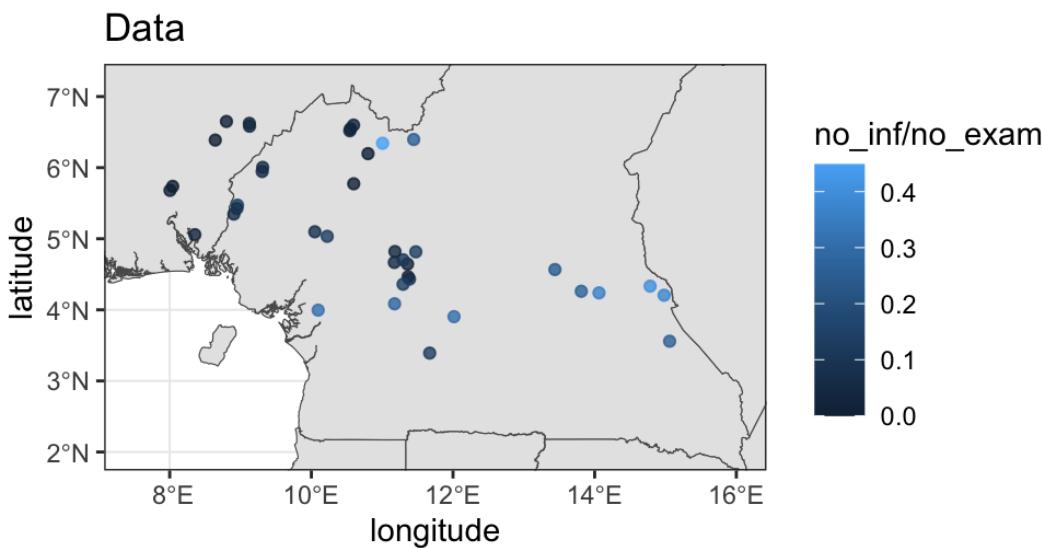
Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.537e+00	5.408e-01	-15.785	< 2e-16 ***
stdev9901	6.750e+00	1.449e+00	4.659	3.18e-06 ***
elev:elev_f(0,1000]	1.467e-03	9.481e-05	15.471	< 2e-16 ***
elev:elev_f(1000,1300]	1.940e-04	9.279e-05	2.091	0.0365 *
elev:elev_f(1300,2000)	-1.506e-03	1.912e-04	-7.880	3.29e-15 ***

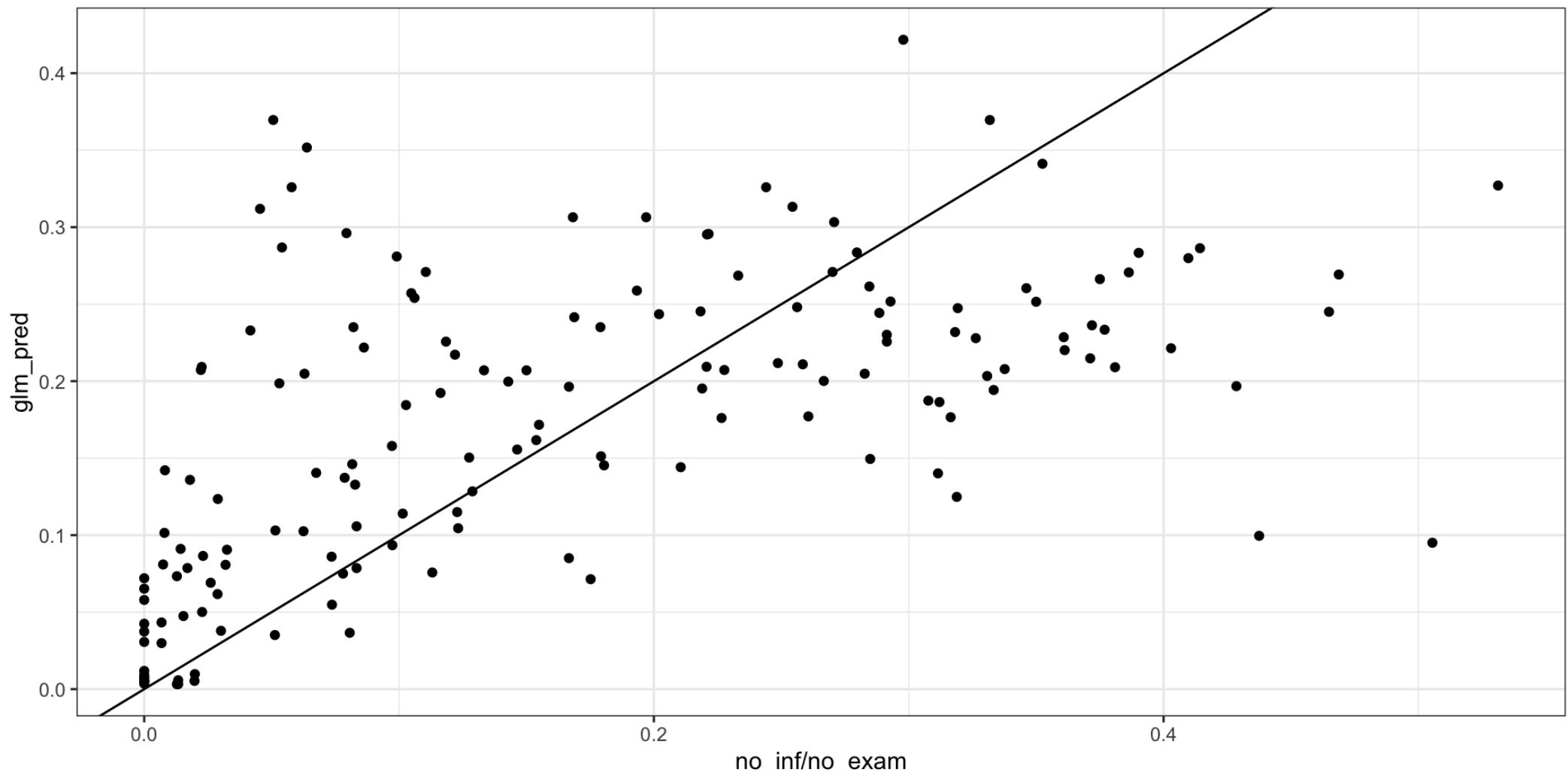
# Predictions - Training



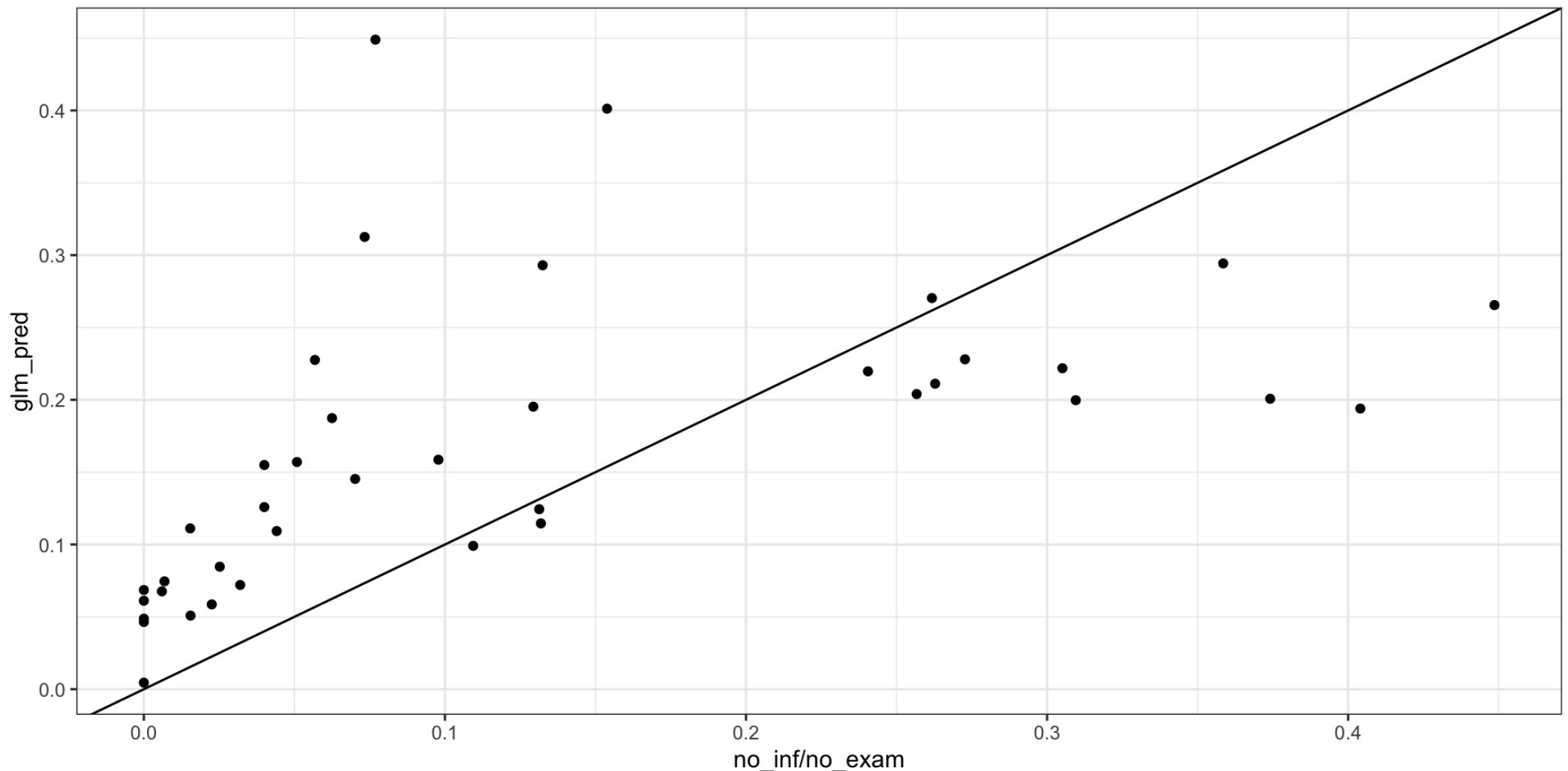
# Predictions - Testing



# Fit - Training



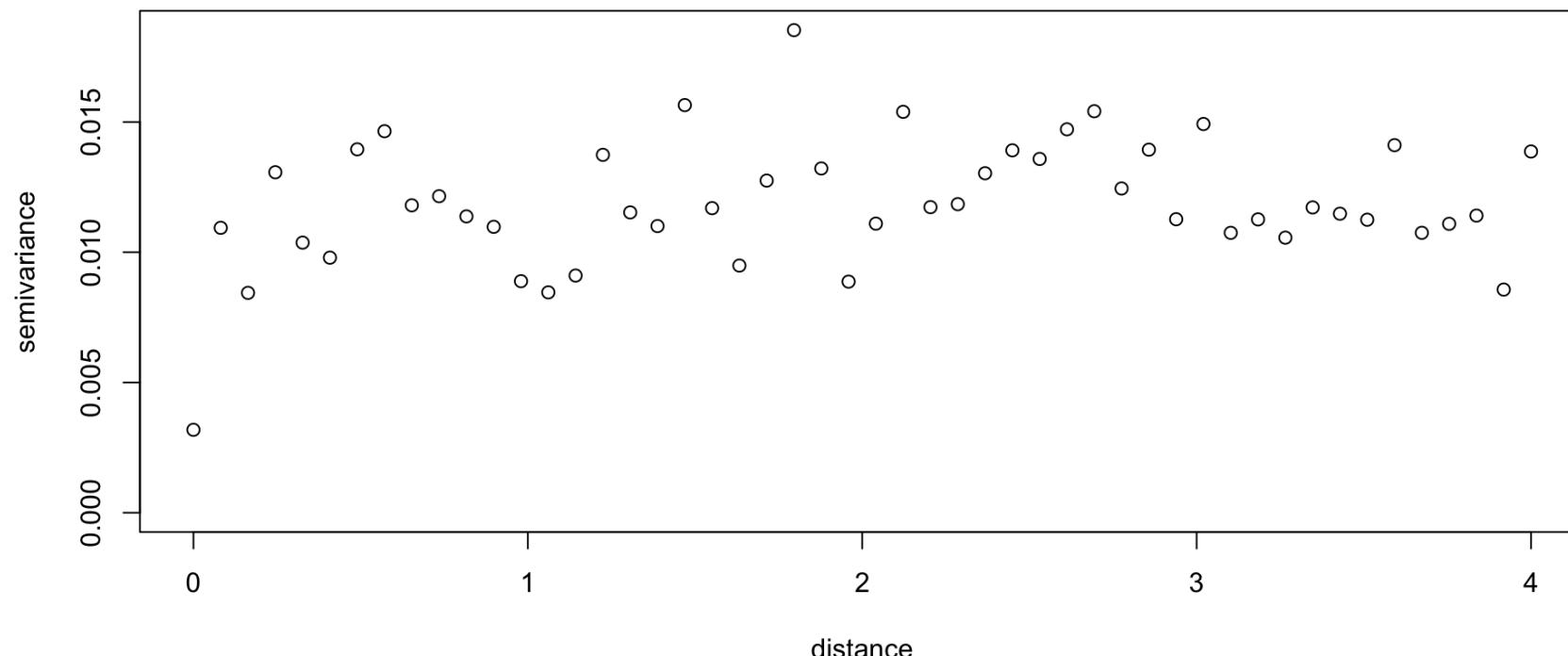
# Fit - Testing



# Spatial Structure?

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

variog: computing omnidirectional variogram



# gpglm model

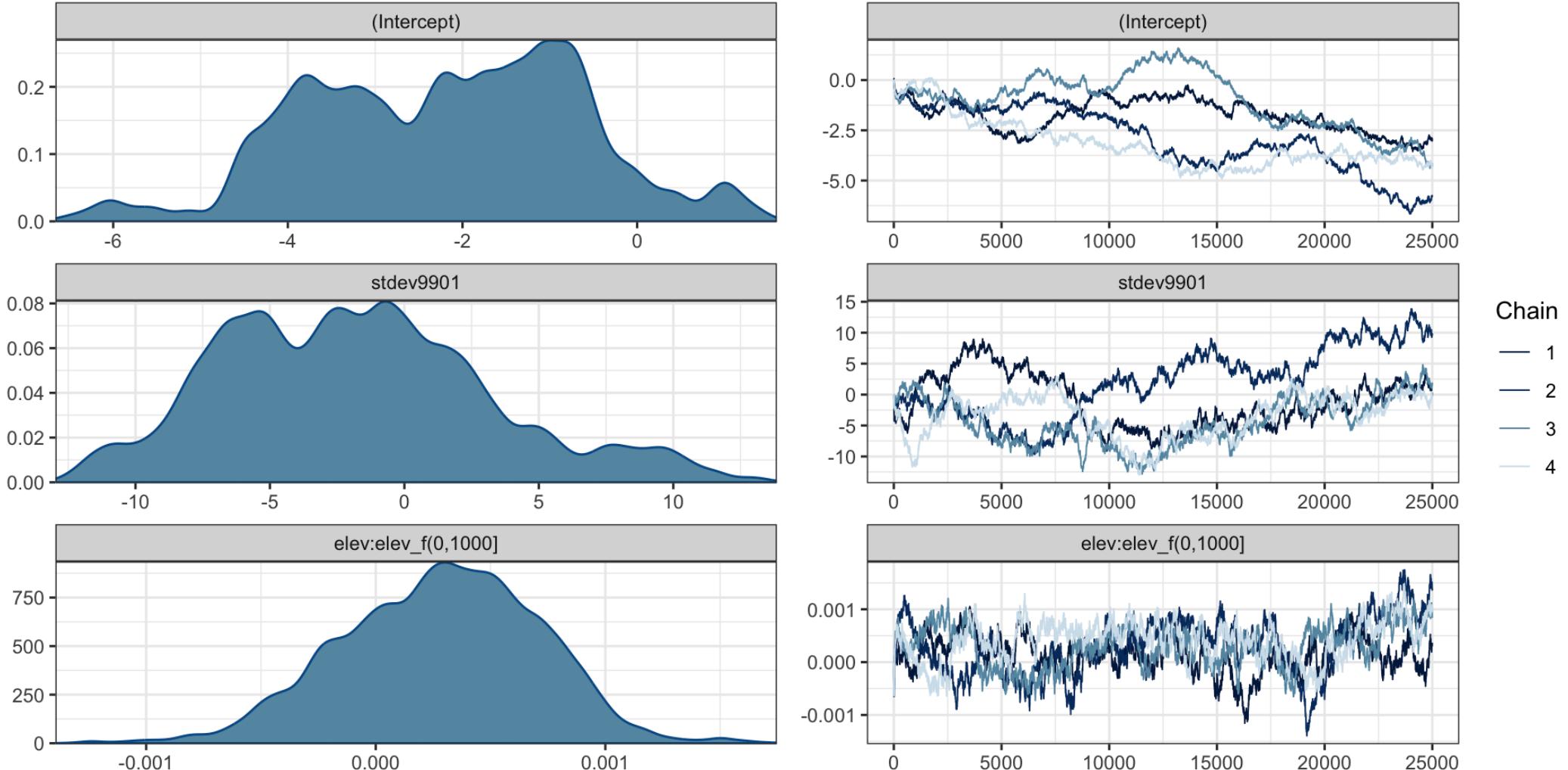
```
1 ll_gp = gpglm(  
2   no_inf ~ elev:elev_f + max9901:max_f + stdev9901,  
3   data = loaloa, family="binomial", weights=loaloa$no_exam,  
4   coords = c("longitude", "latitude"),  
5   cov_model="exponential",  
6   starting = list(  
7     beta=rep(0,7), phi=3, sigma.sq=1, w=0  
8   ),  
9   priors = list(phi.unif=c(3/4, 3/0.1), sigma.sq.sig=c(2, 2)),  
10  tuning = list(  
11    "beta"=c(0.05,0.3,0.0001,0.0003,0.0005,0.1,0.07),  
12    "phi"=0.6, "sigma.sq"=0.3, "w"=0.1  
13  ),  
14  n_batch = 500,  
15  batch_len = 50,  
16  verbose = TRUE,  
17  n_report = 10  
18 )  
19  
20  
21 saveRDS(ll_gp, file="Lec22_loaloa.rds")
```

```
1 11_gp
```

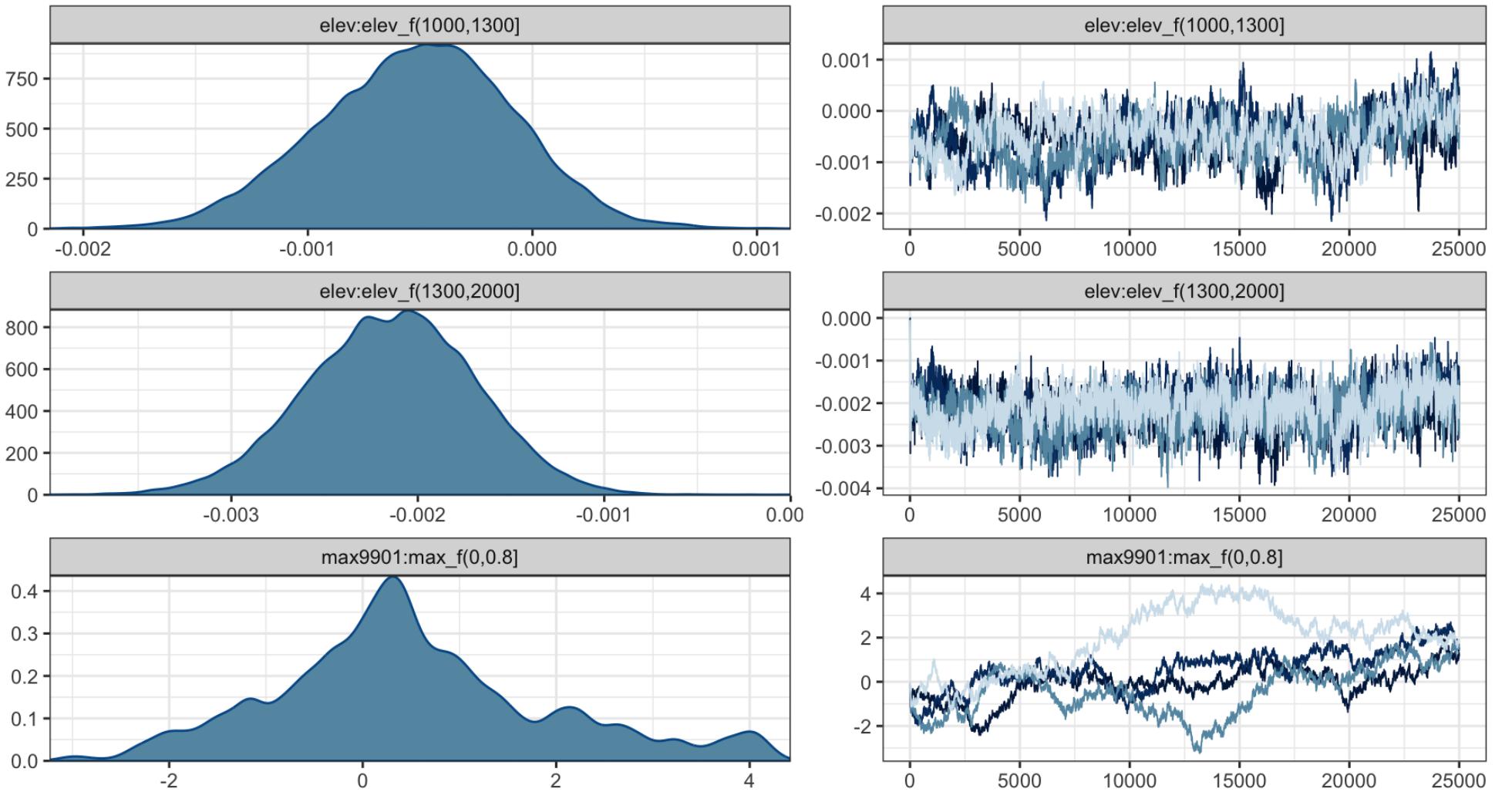
```
# A gpglm model (spBayes spGLM) with 4 chains, 9 variables, and 100000 iterations.
# A tibble: 9 × 10
  variable      mean    median     sd     mad      q5      q95   rhat
  <chr>       <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl> <dbl>
1 (Intercep... -2.21e+0 -2.10e+0 1.56e+0 1.76e+0 -4.55e+0 2.49e-1 1.70
2 stdev9901  -1.79e+0 -2.09e+0 5.05e+0 5.16e+0 -9.22e+0 8.06e+0 1.39
3 elev:elev... 2.84e-4  3.04e-4 4.31e-4 4.43e-4 -4.41e-4 9.38e-4 1.08
4 elev:elev... -5.18e-4 -5.01e-4 4.32e-4 4.35e-4 -1.25e-3 1.66e-4 1.06
5 elev:elev... -2.14e-3 -2.13e-3 4.53e-4 4.57e-4 -2.90e-3 -1.42e-3 1.05
6 max9901:m... 5.11e-1  3.58e-1 1.42e+0 1.18e+0 -1.69e+0 3.29e+0 1.57
7 max9901:m... 7.51e-1  5.85e-1 1.34e+0 1.10e+0 -1.29e+0 3.46e+0 1.59
8 sigma.sq    1.38e+0  1.29e+0 4.28e-1 3.37e-1 8.69e-1 2.19e+0 1.04
9 phi         2.49e+0  2.39e+0 8.55e-1 8.06e-1 1.26e+0 4.03e+0 1.04
# ... with 2 more variables: ess_bulk <dbl>, ess_tail <dbl>
```

# Diagnostics

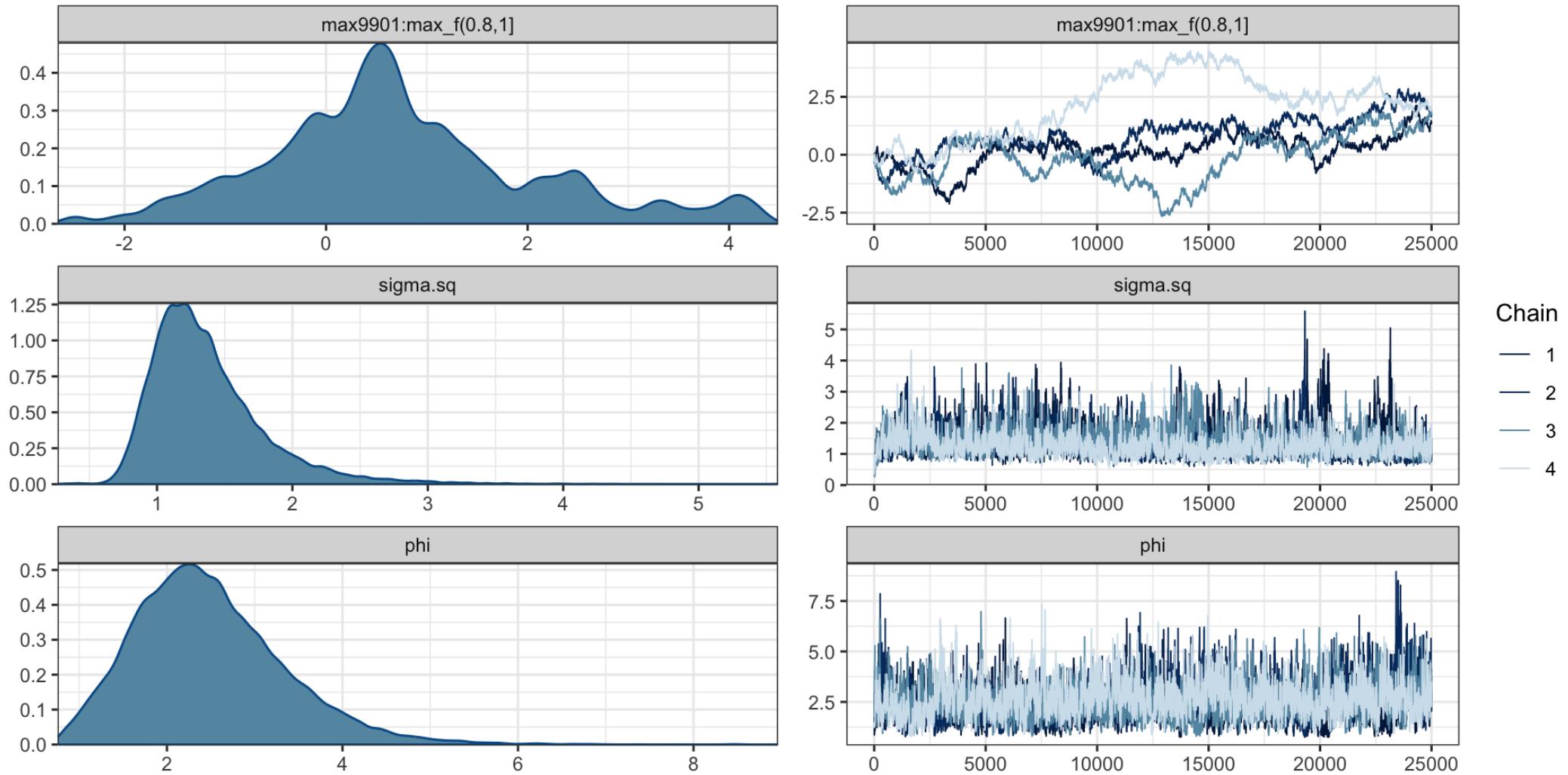
```
1 plot(ll_gp, vars=1:3)
```



```
1 plot(ll_gp, vars=4:6)
```

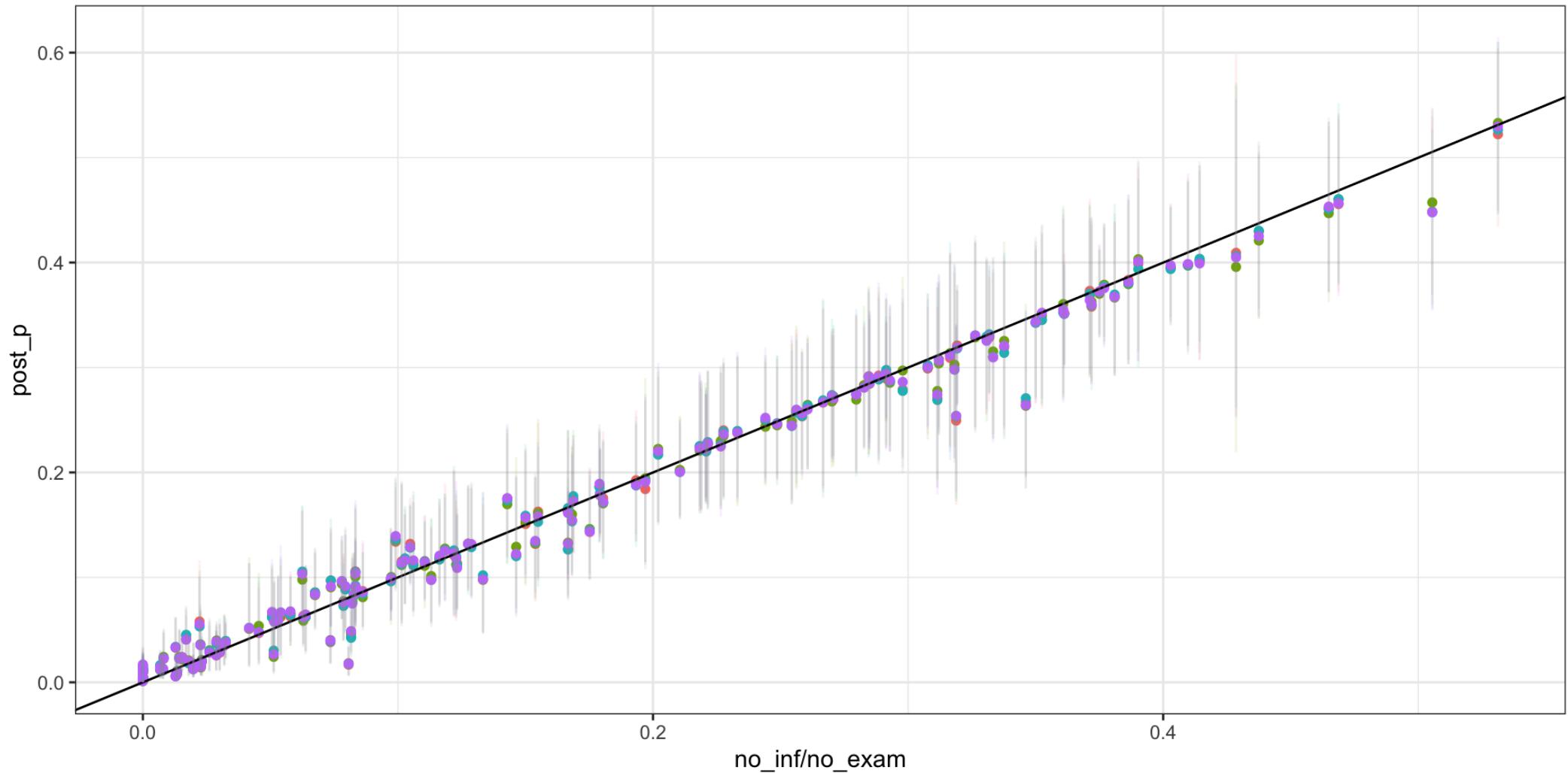


```
1 plot(ll_gp, vars=7:9)
```



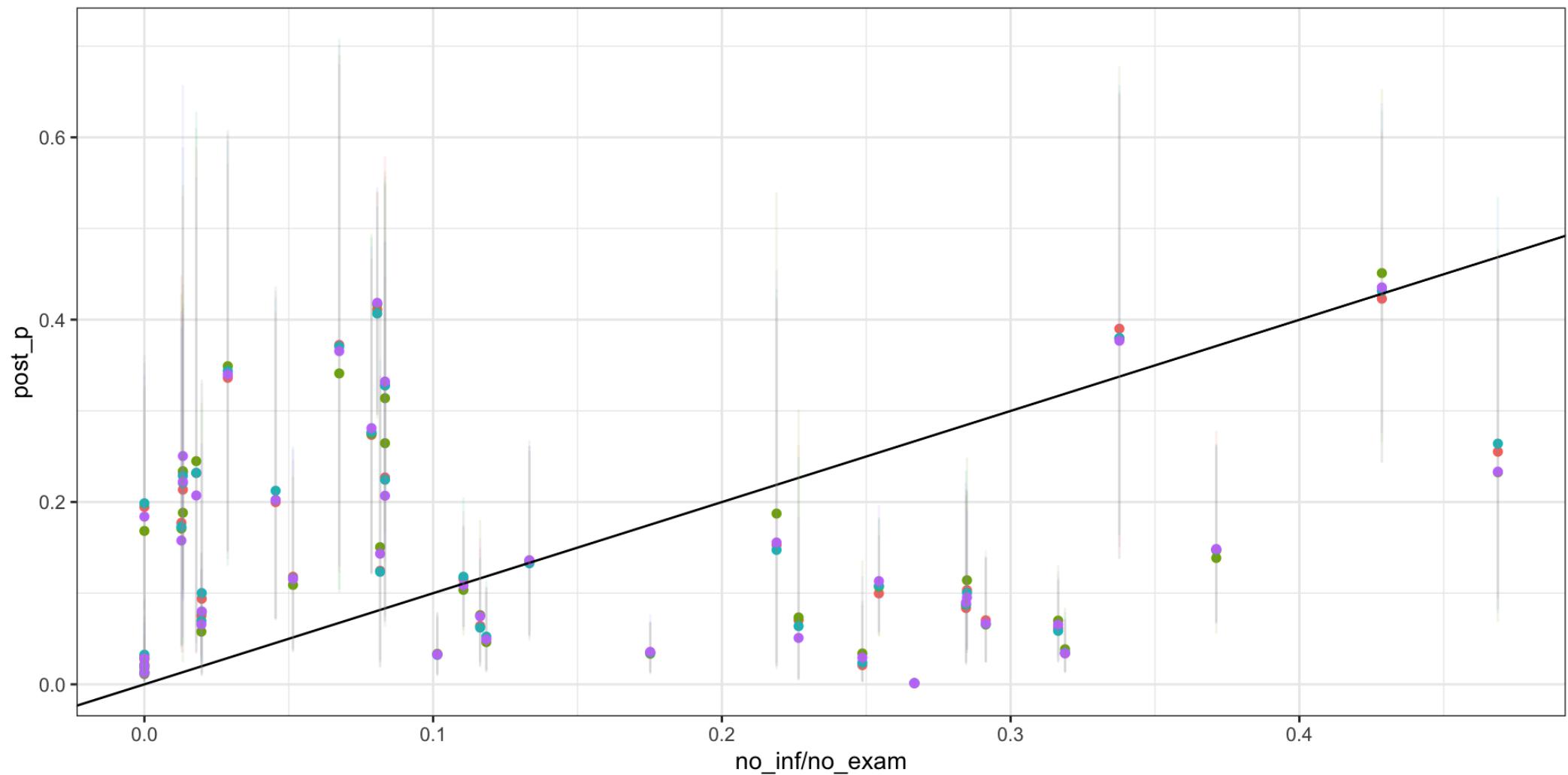
# Prediction (training)

```
1 ll_gp_pred = predict(  
2   ll_gp,  
3   newdata=loaloa,  
4   coords = c("longitude", "latitude"),  
5   thin = 25,  
6   verbose=FALSE  
7 )  
8  
9 ll_gp_pred_y = tidybayes::gather_draws(ll_gp_pred, y[i]) %>%  
10 group_by(.chain, i) %>%  
11 summarize(  
12   post_p = mean(.value),  
13   q025 = quantile(.value, 0.025),  
14   q975 = quantile(.value, 0.975)  
15 )
```



# Prediction - Testing

```
1 ll_gp_test_pred = predict(  
2   ll_gp,  
3   newdata=loaloa_test,  
4   coords = c("longitude", "latitude"),  
5   thin = 25,  
6   verbose=FALSE  
7 )  
8  
9 ll_gp_test_pred_y = tidybayes::gather_draws(ll_gp_test_pred, y[i]) %>%  
10 group_by(.chain, i) %>%  
11 summarize(  
12   post_p = mean(.value),  
13   q025 = quantile(.value, 0.025),  
14   q975 = quantile(.value, 0.975)  
15 )
```



# Diggle's Predictive Surface

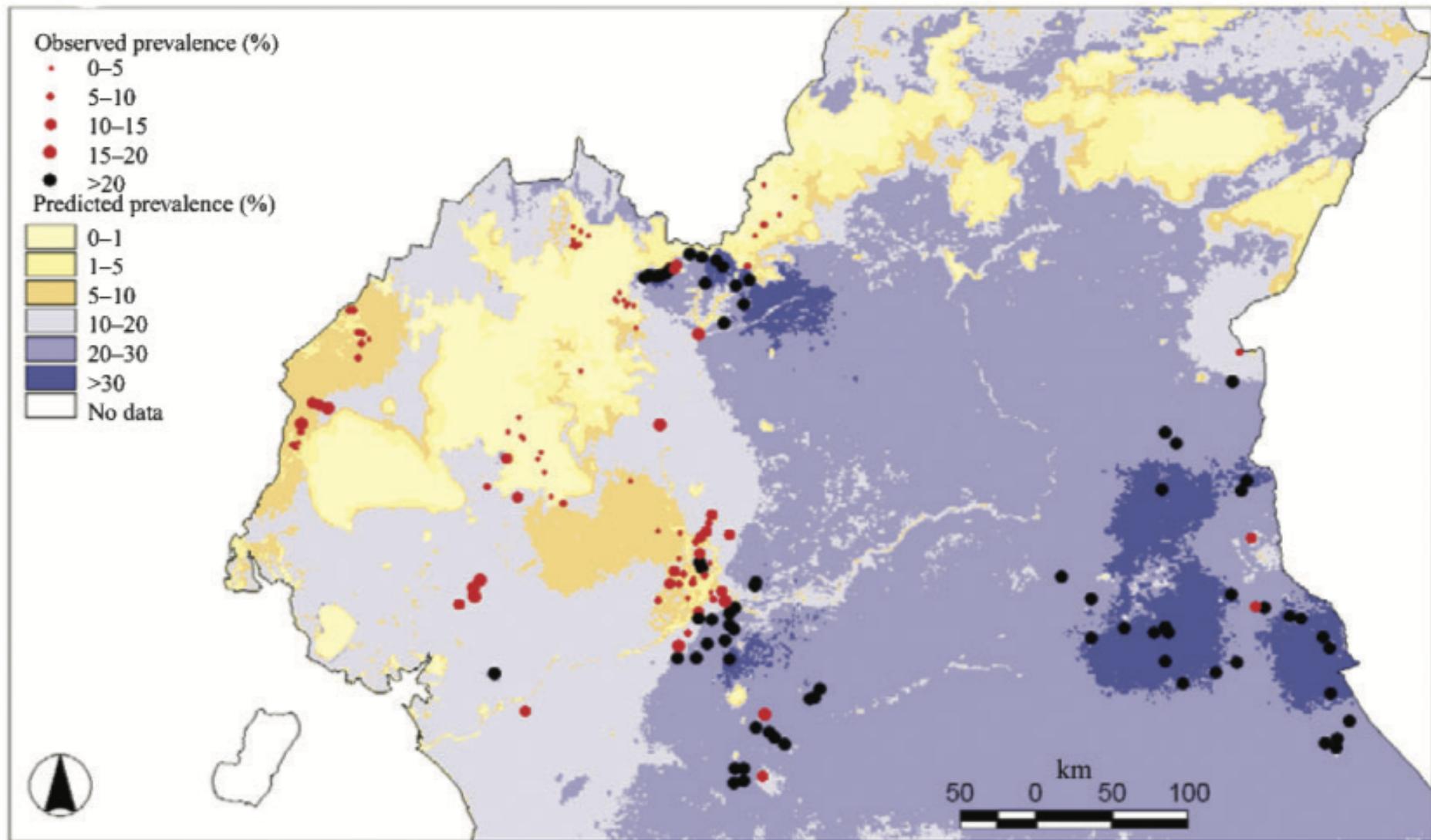
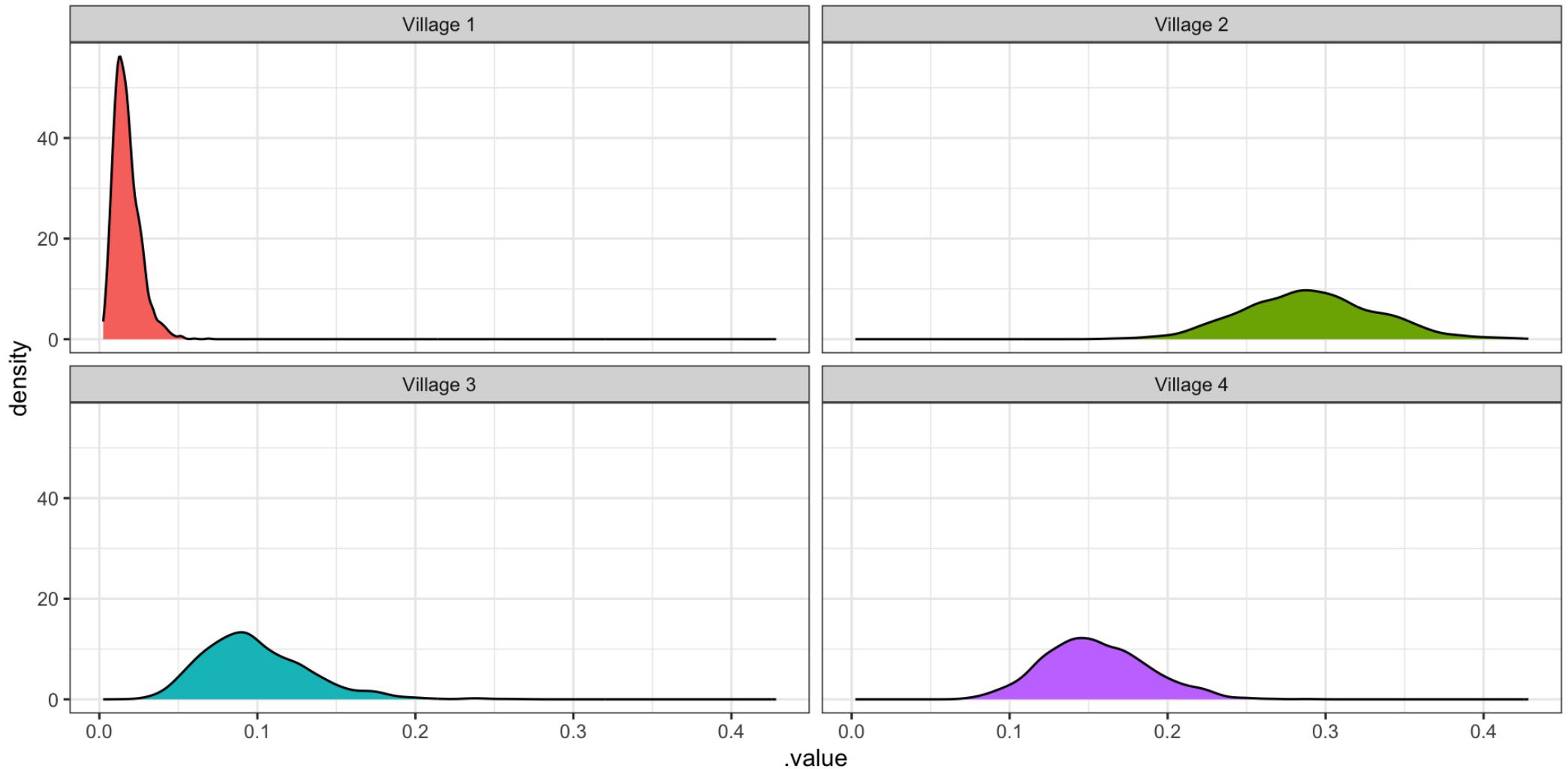


FIG. 2. Point estimates of the prevalence of *Loa loa* microfilaraemia, over-laid with the prevalences observed in field studies.

# Exceedance Probability



# Exceedance Probability Predictive Surface

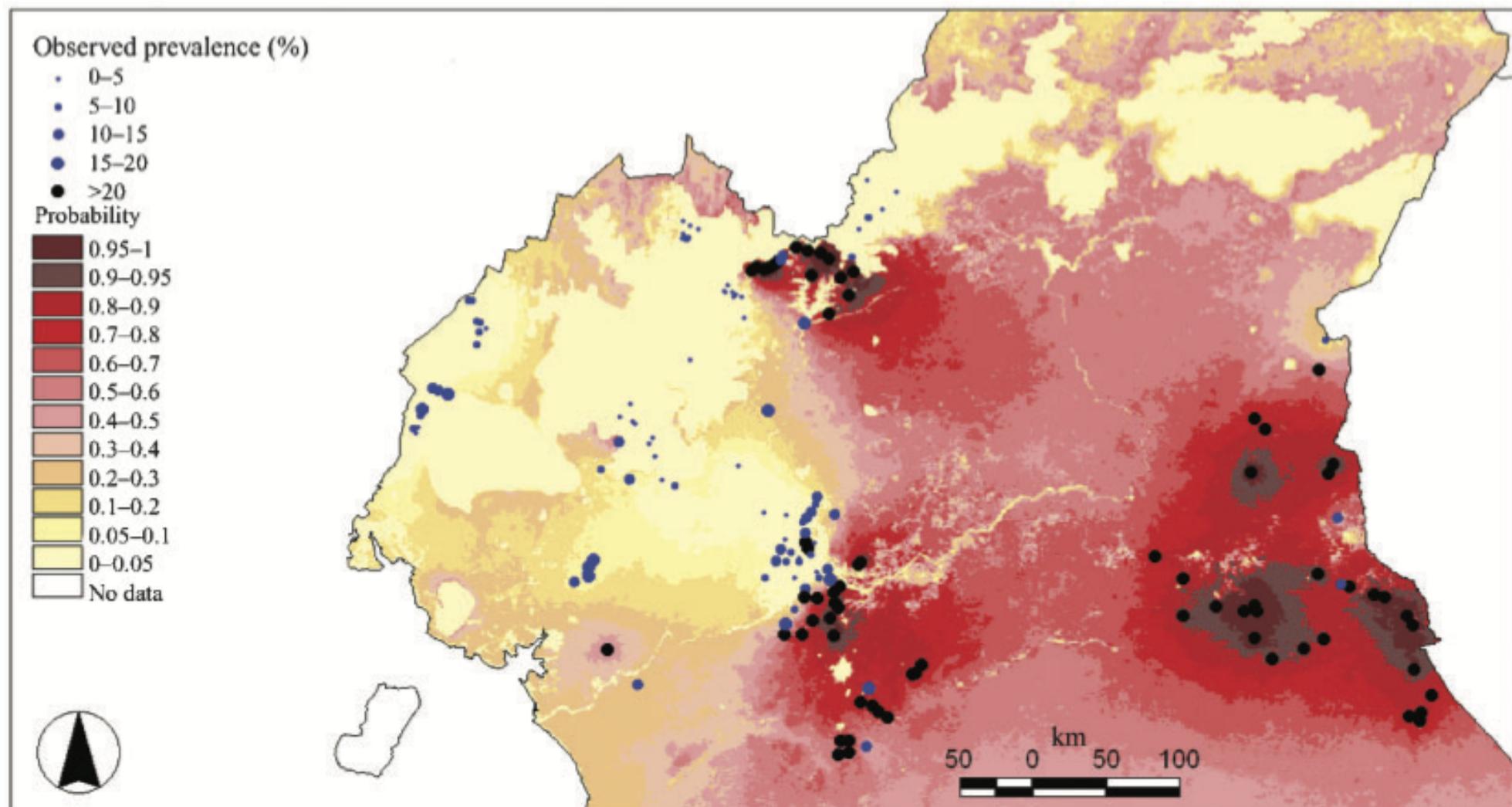


FIG. 4. A probability contour map, indicating the probability that the prevalence of *Loa loa* microfilaraemia in each area exceeds 20%, over-laid with the prevalences observed in field studies.

# 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 $\delta^2\text{H}$

Hermit Thrush  
(*Catharus guttatus*)

138 individuals

14 locations

6 loci

9-27 alleles / locus

Wilson's Warbler  
(*Wilsonia pusilla*)

163 individuals

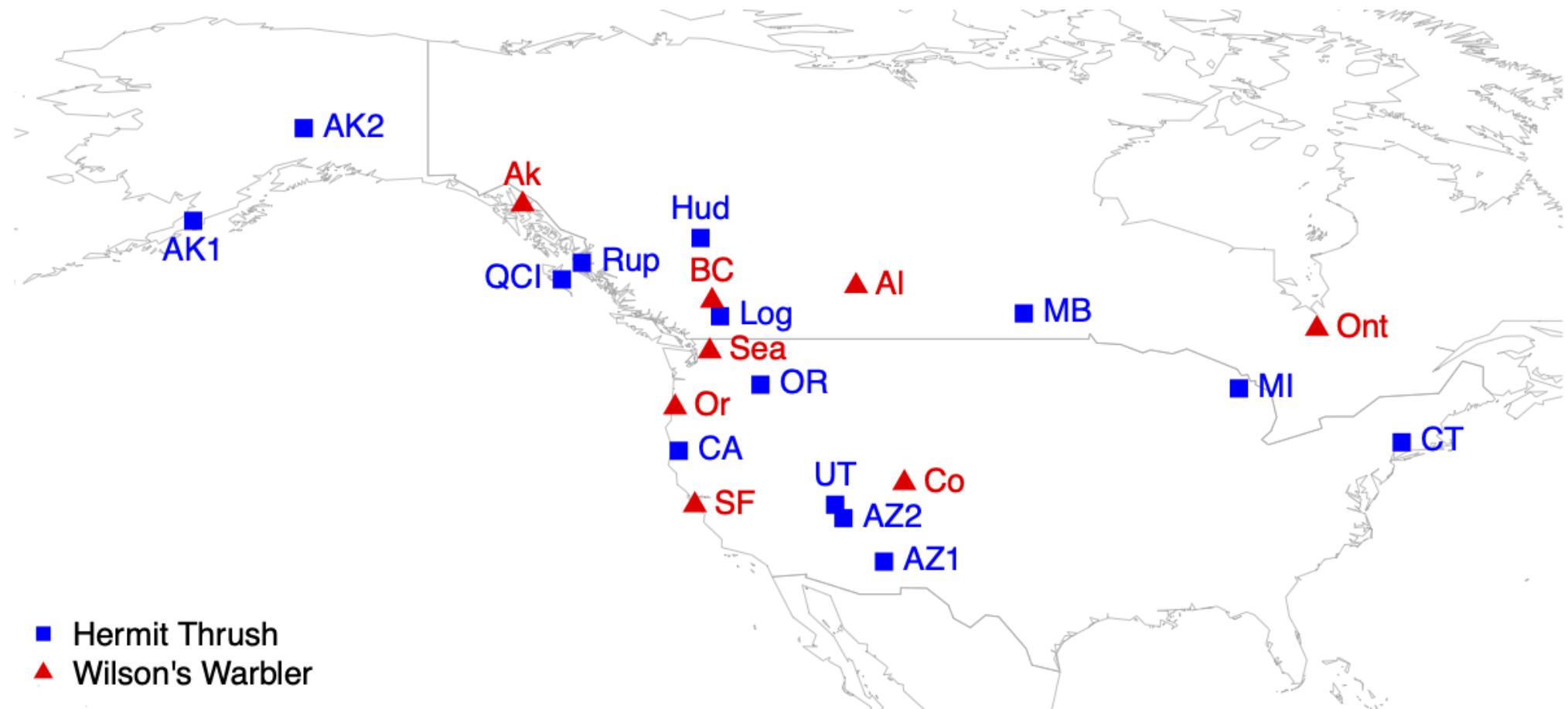
8 locations

9 loci

15-31 alleles / locus



# Sampling Locations



# Allele Frequency Model

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

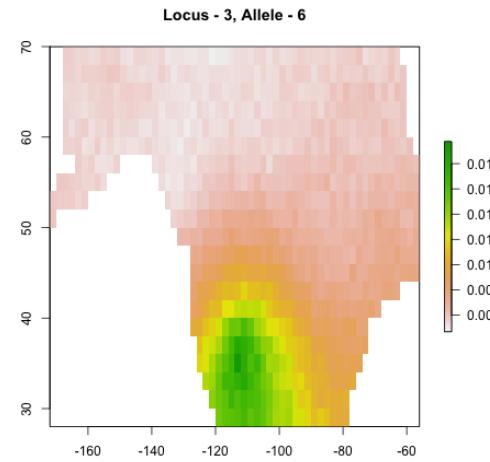
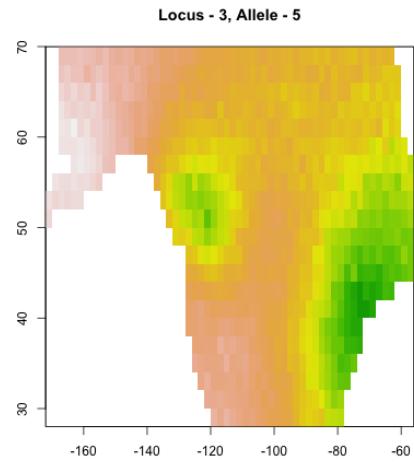
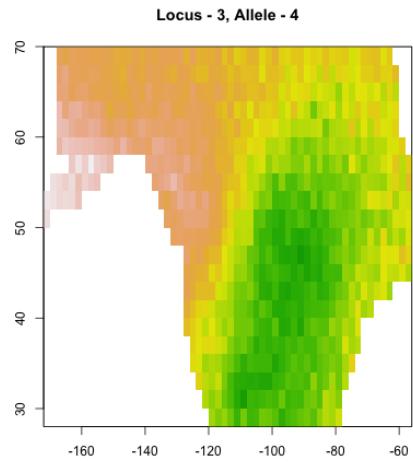
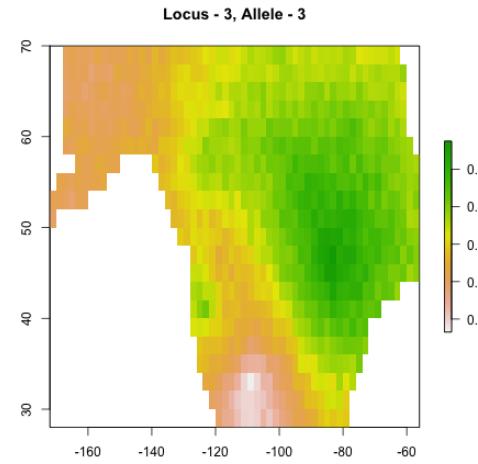
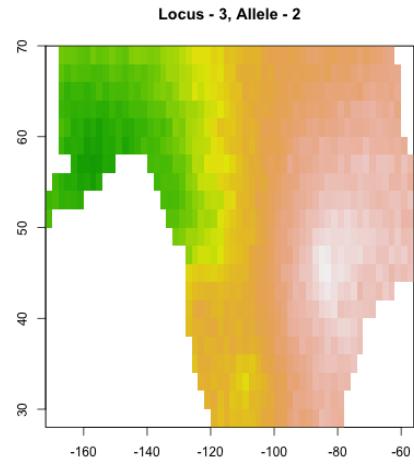
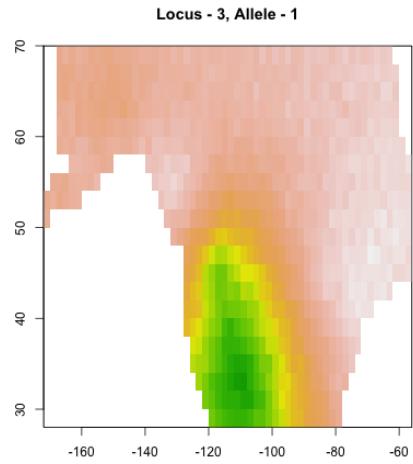
$$y_{\cdot lk} | \Theta \sim \left( \sum_i y_{ilk}, f_{\cdot lk} \right)$$

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

$$\Theta_{il} | \alpha, \mu \sim (\mu_{il}, \Sigma)$$

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

# Predictions by Allele (Locus 3)



# Genetic Assignment Model

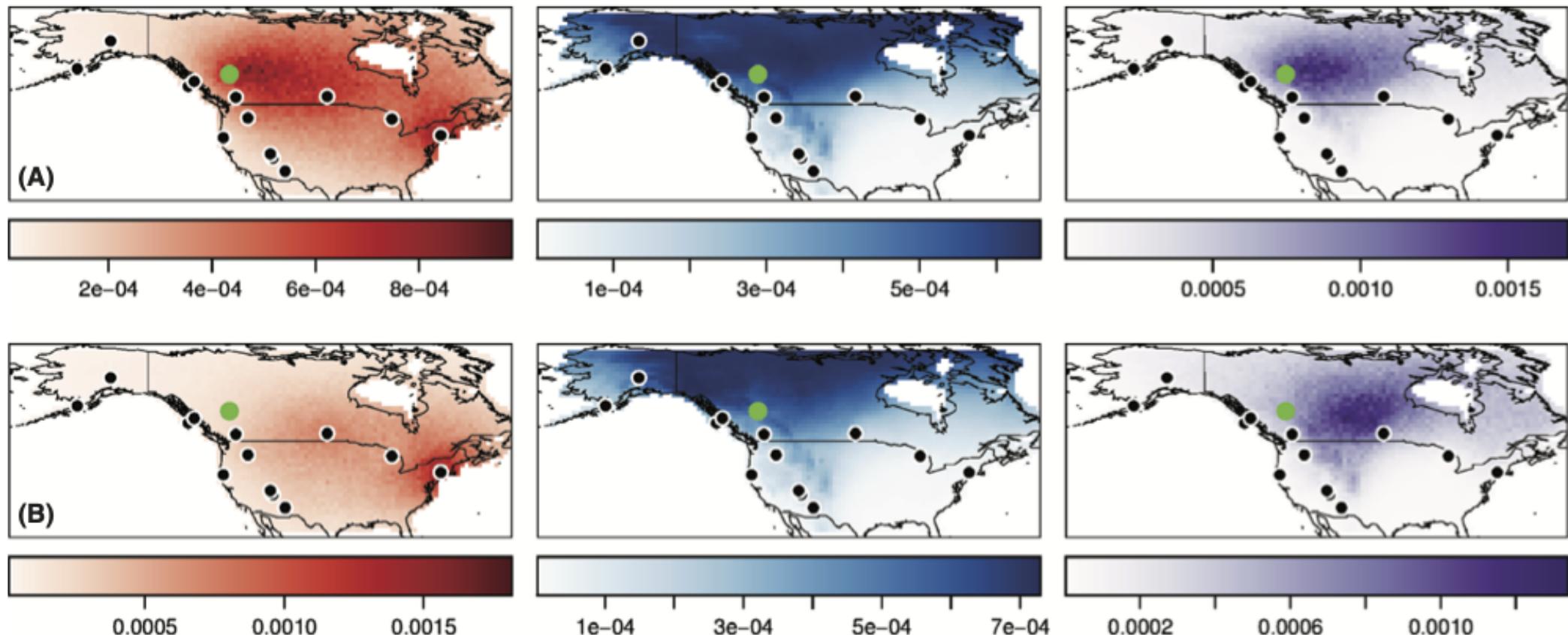
Assignment model assuming Hardy-Weinberg equilibrium and allowing for genotyping ( $\delta$ ) and single amplification ( $\gamma$ ) errors.

$$P(S_G | f, k) = \prod_1 P(i_l, j_l | f, k)$$

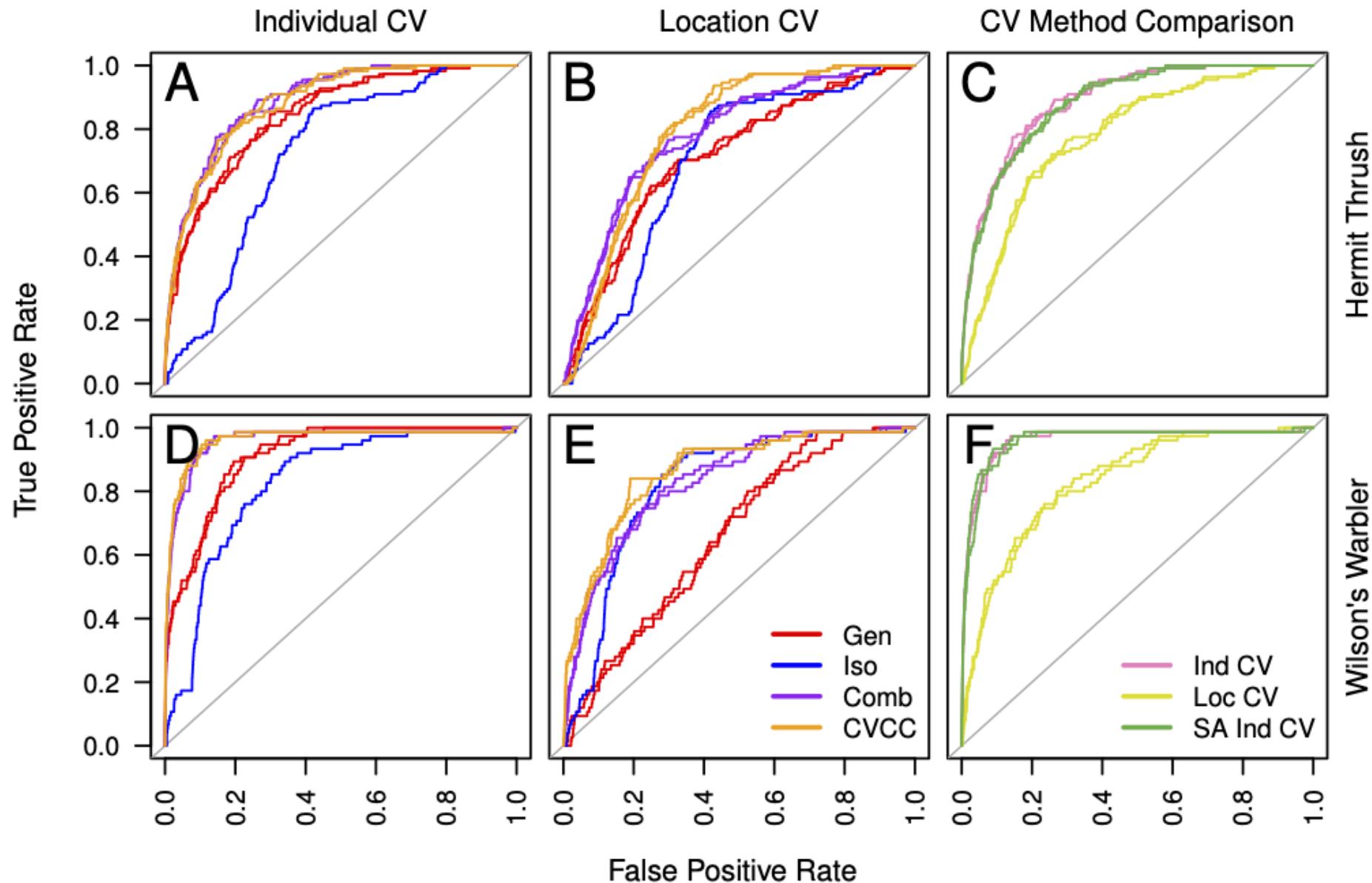
$$P(i_l, j_l | f, k) = \begin{cases} \gamma P(i_l | f, k) + (1 - \gamma) P(i_l | \tilde{f}, k)^2 & \text{if } i = j \\ (1 - \gamma) P(i_l | f, k) P(j_l | f, k) & \text{if } i \neq j \end{cases}$$

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

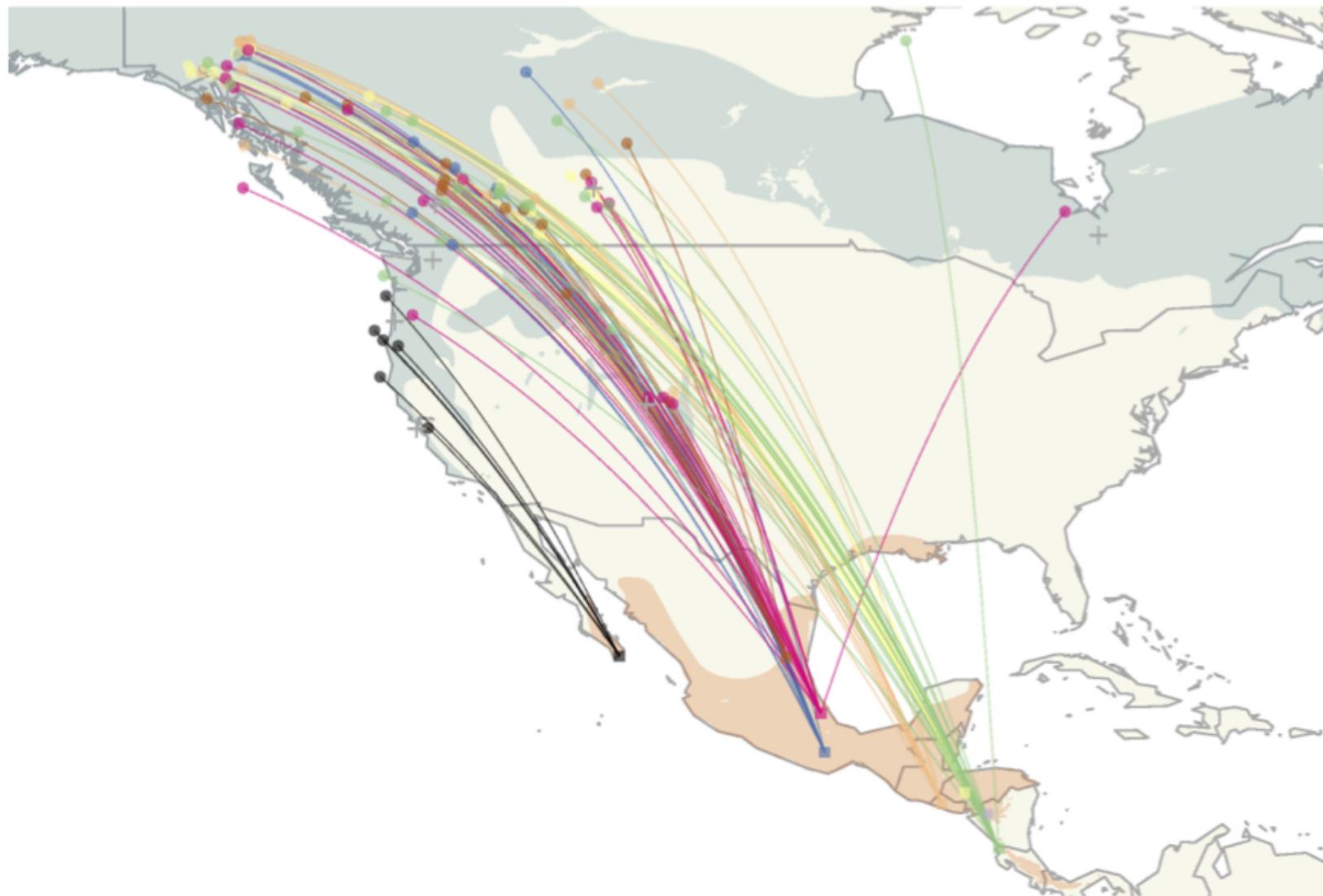
# Combined Model



# Model Assessment



# Migratory Connectivity



Sta 344 - Fall 2022