

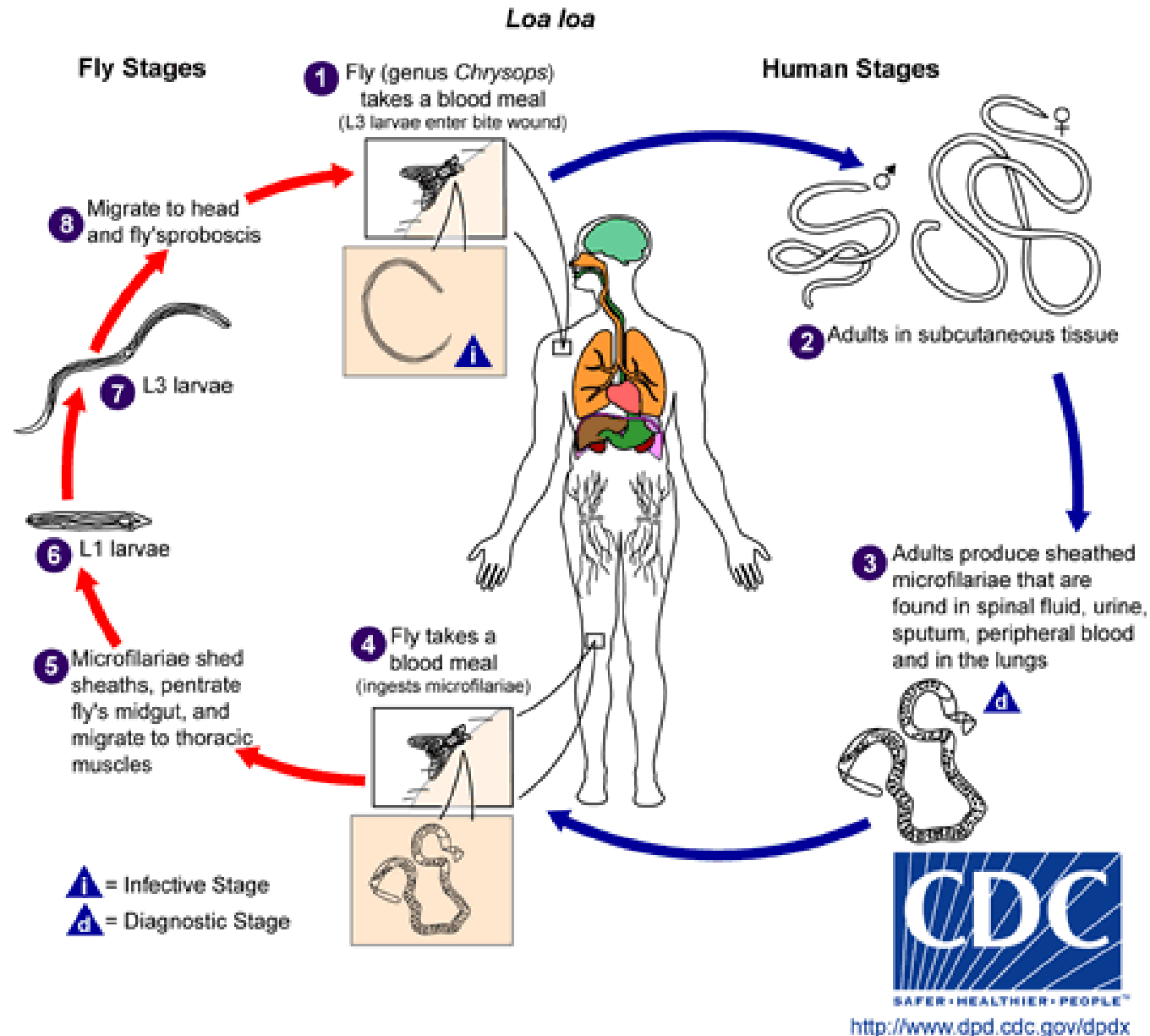
GLMs + Point referenced data

Lecture 22

Dr. Colin Rundel

Loa Loa Example

Loa Loa



Data

```
1 ( loaloe = PrevMap::loaloe |>
2   as_tibble() |>
3   rename_with(tolower) |>
4   rename(elev=elevation)
5 )
```

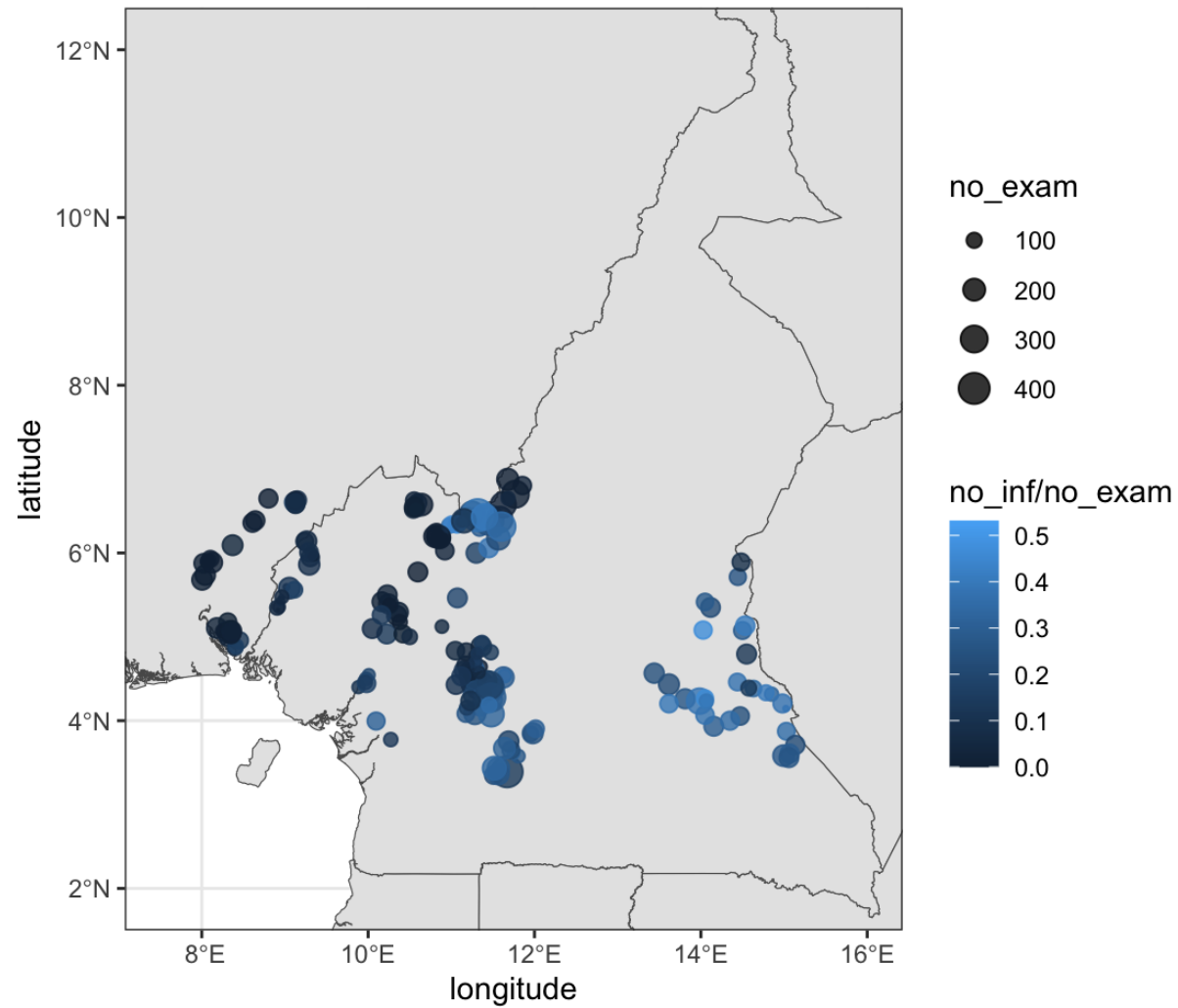
```
# A tibble: 197 × 11
```

	row	villcode	longitude	latitude	no_exam	no_inf	elev	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

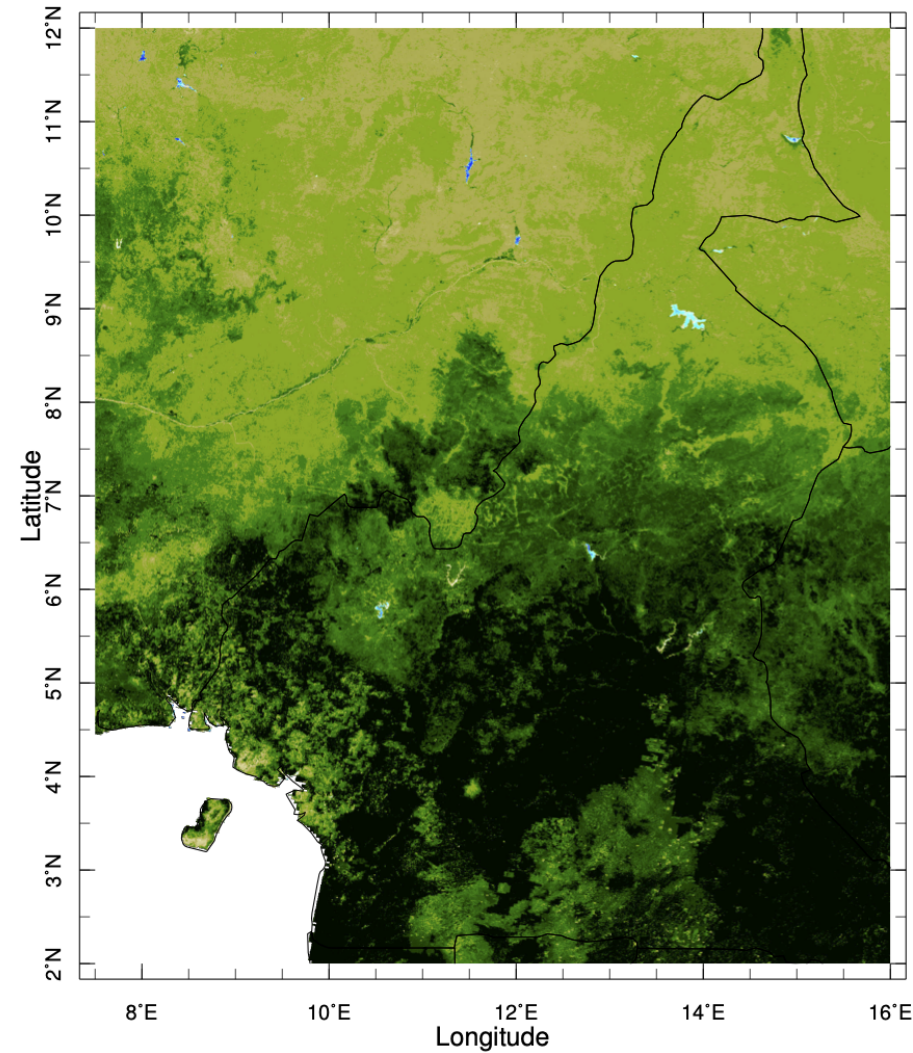
```
# i 187 more rows
```

```
# i 3 more variables: max9901 <dbl>, min9901 <dbl>, stdev9901 <dbl>
```

Spatial Distribution



Normalized Difference Vegetation Index



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`, `max9901`, `min9901`, `stdev9901` - aggregated NDVI data from 1999 to 2001 from the Flemish Institute for Technological Research (1 km resolution)

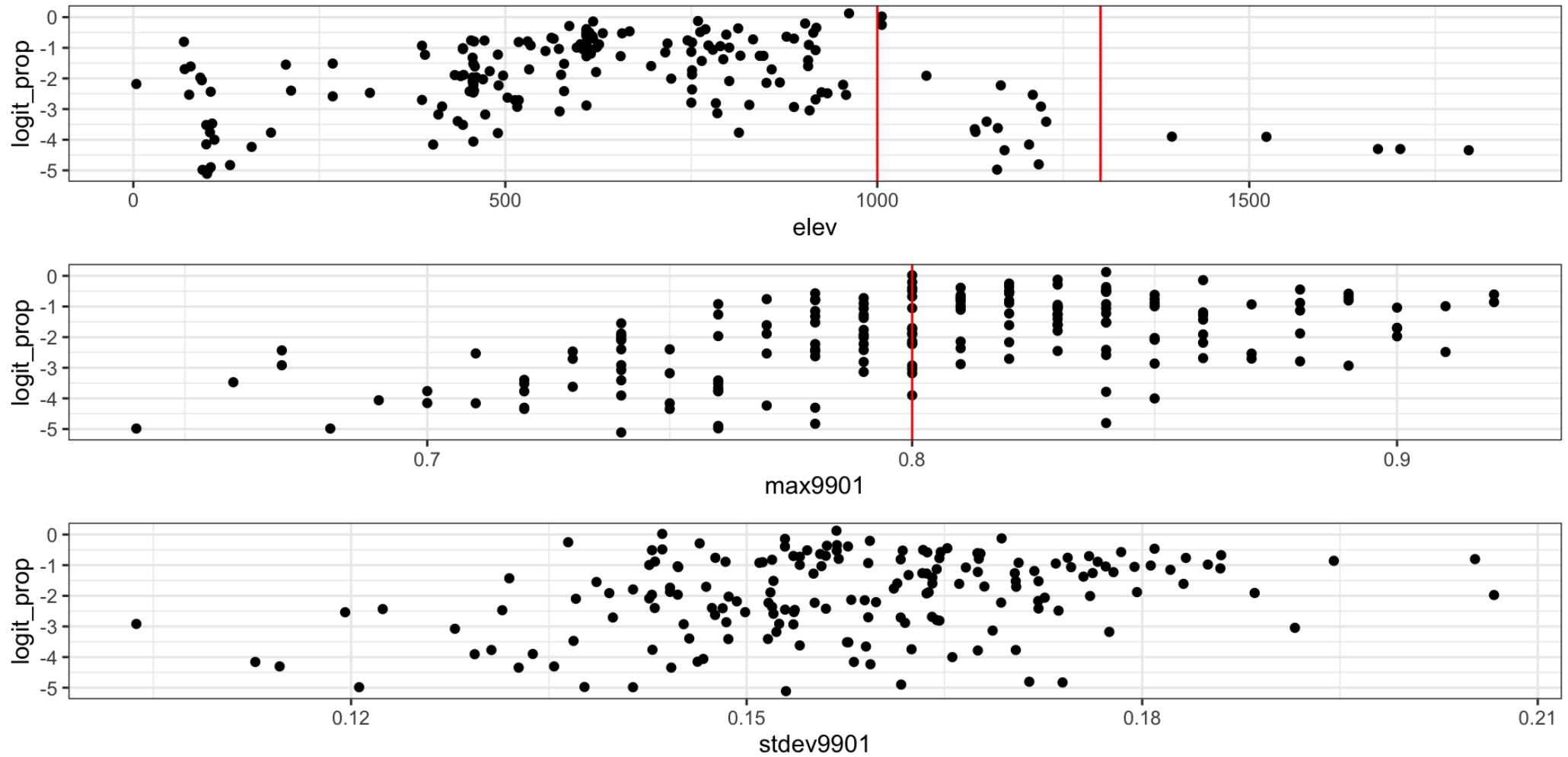
Diggle's Model

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

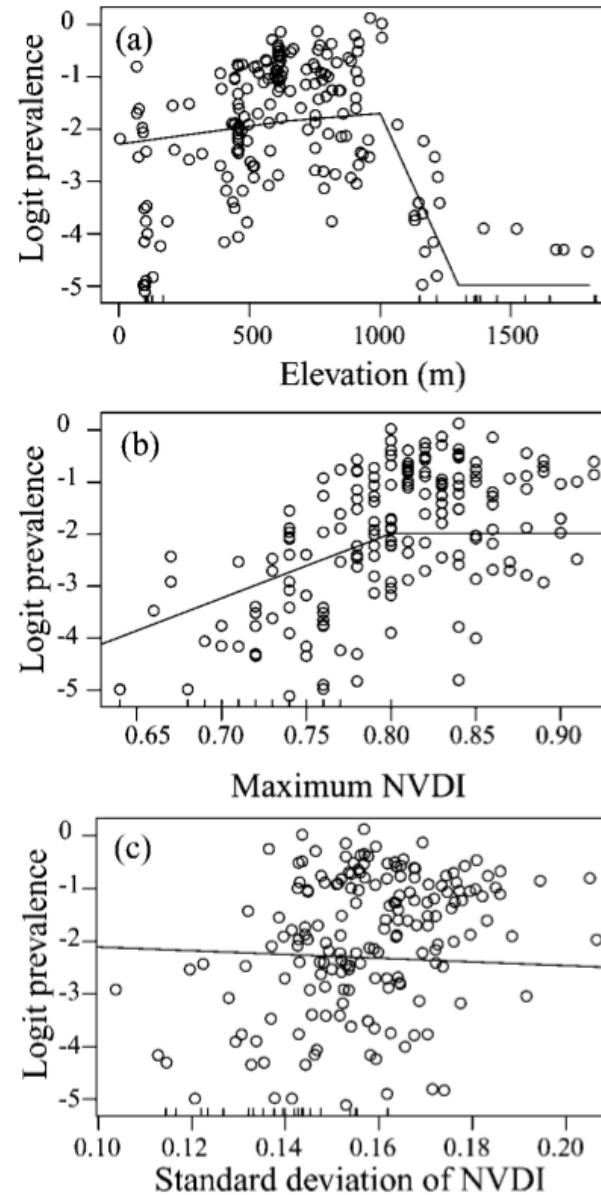
$$w(s) \sim N(0, \Sigma)$$

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

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]

i 187 more rows

Model Matrix

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

```
# A tibble: 197 × 4
```

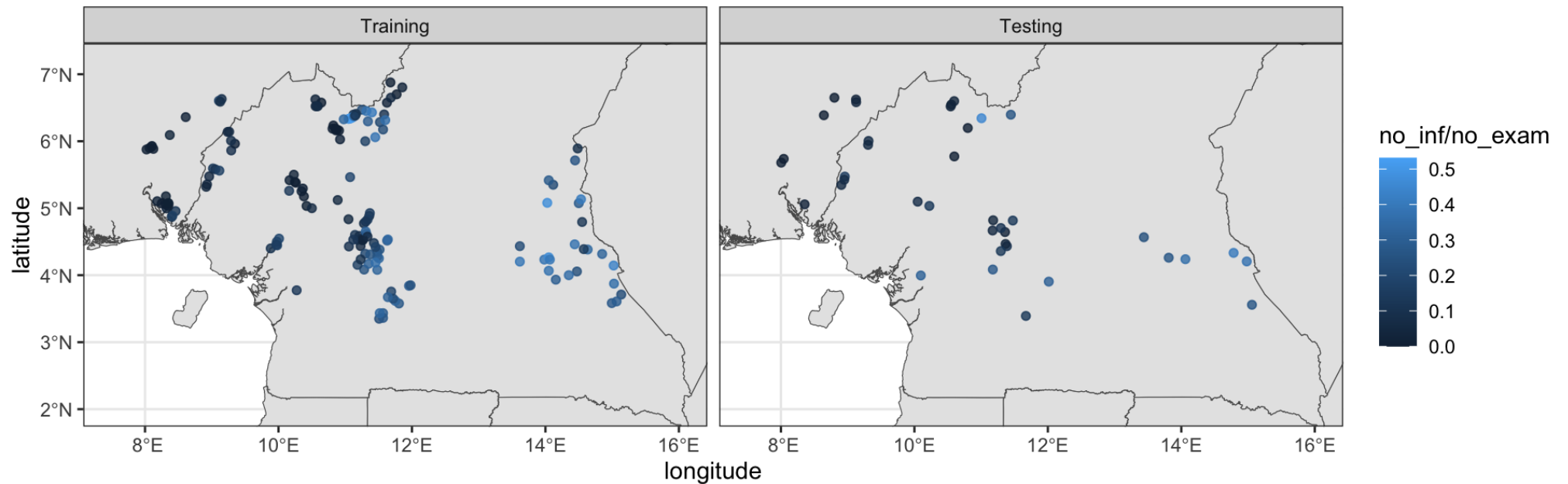
```
  `(Intercept)` `elev:elev_f(0,1000]` `elev:elev_f(1000,1300]`  
    <dbl>          <dbl>          <dbl>  
1         1         108             0  
2         1          99             0  
3         1        783             0  
4         1        104             0  
5         1        109             0  
6         1       909             0  
7         1       503             0  
8         1       103             0  
9         1       751             0  
10        1       268             0
```

```
# i 187 more rows
```

```
# i 1 more variable: `elev:elev_f(1300,2000]` <dbl>
```

OOS Validation

```
1 set.seed(12345)
2 loaloa_test = loaloa |> slice_sample(prop=0.20)
3 loaloa = anti_join(loaloa, loaloa_test, by="row")
```



Model

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

Call:

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

Coefficients:

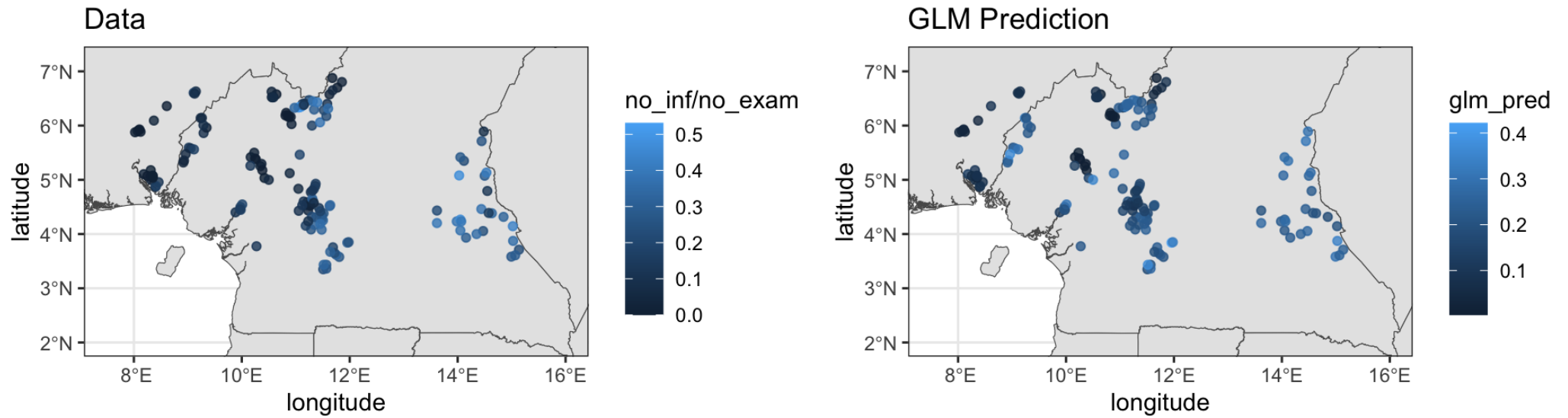
	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-8.537e+00	5.408e-01	-15.785	< 2e-16	***
stdev9901	6.750e+00	1.449e+00	4.659	3.18e-06	***
elev:elev_f(0,1000]	1.467e-03	9.481e-05	15.471	< 2e-16	***
elev:elev_f(1000,1300]	1.940e-04	9.279e-05	2.091	0.0365	*
elev:elev_f(1300,2000]	-1.506e-03	1.912e-04	-7.880	3.29e-15	***
max9901:max_f(0,0.8]	6.399e+00	6.951e-01	9.207	< 2e-16	***
max9901:max_f(0.8,1]	6.364e+00	6.387e-01	9.965	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

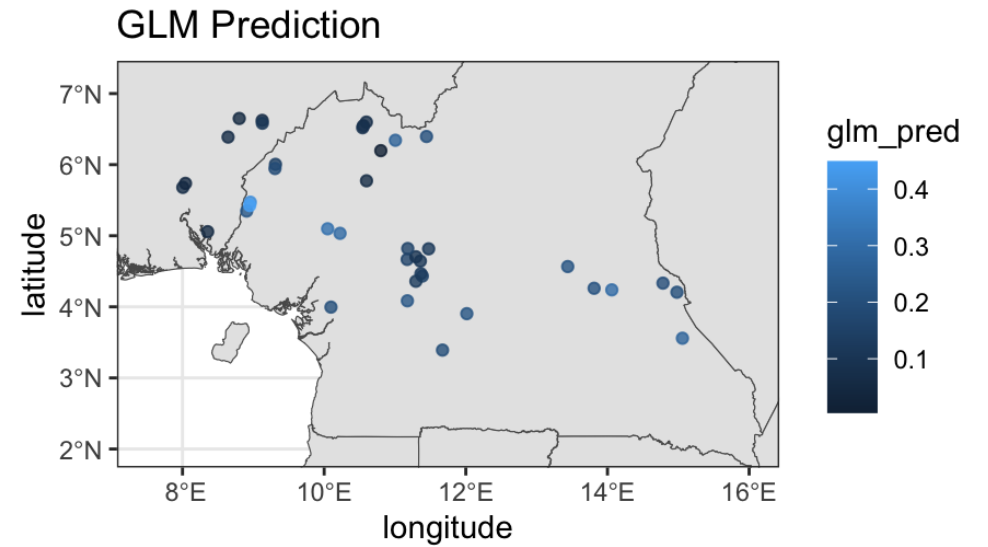
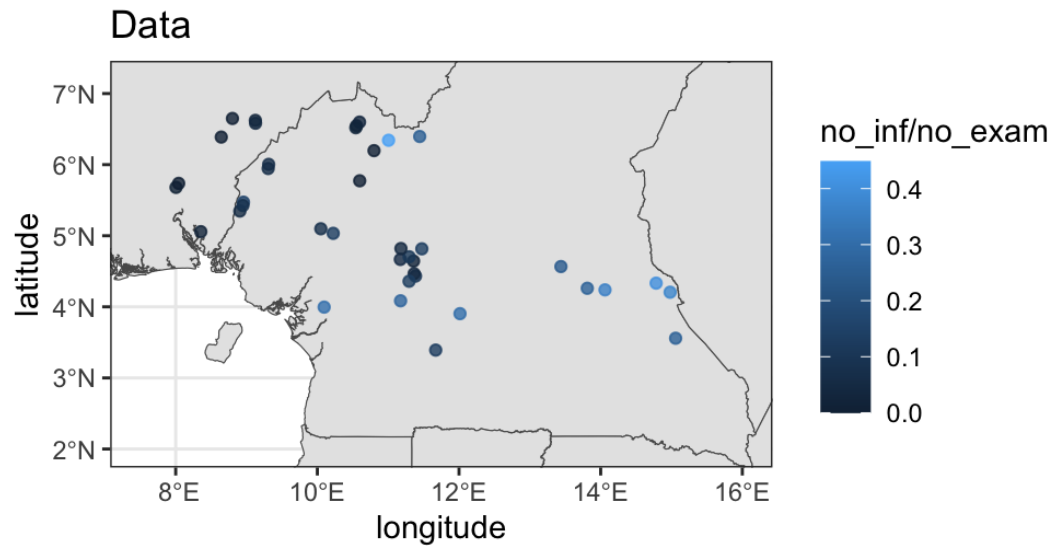
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3364.8 on 157 degrees of freedom
Residual deviance: 1630.8 on 151 degrees of freedom

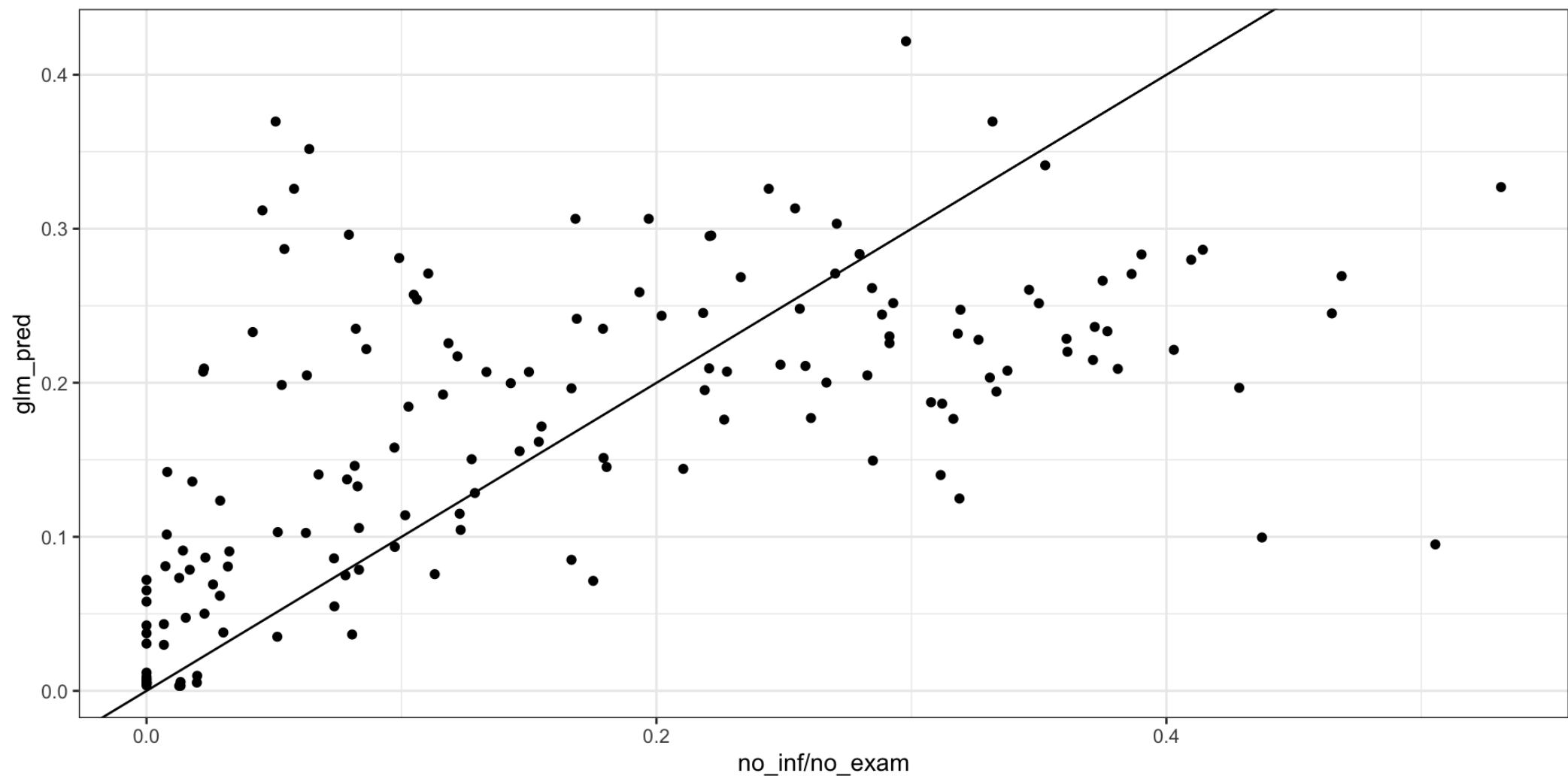
Predictions - Training



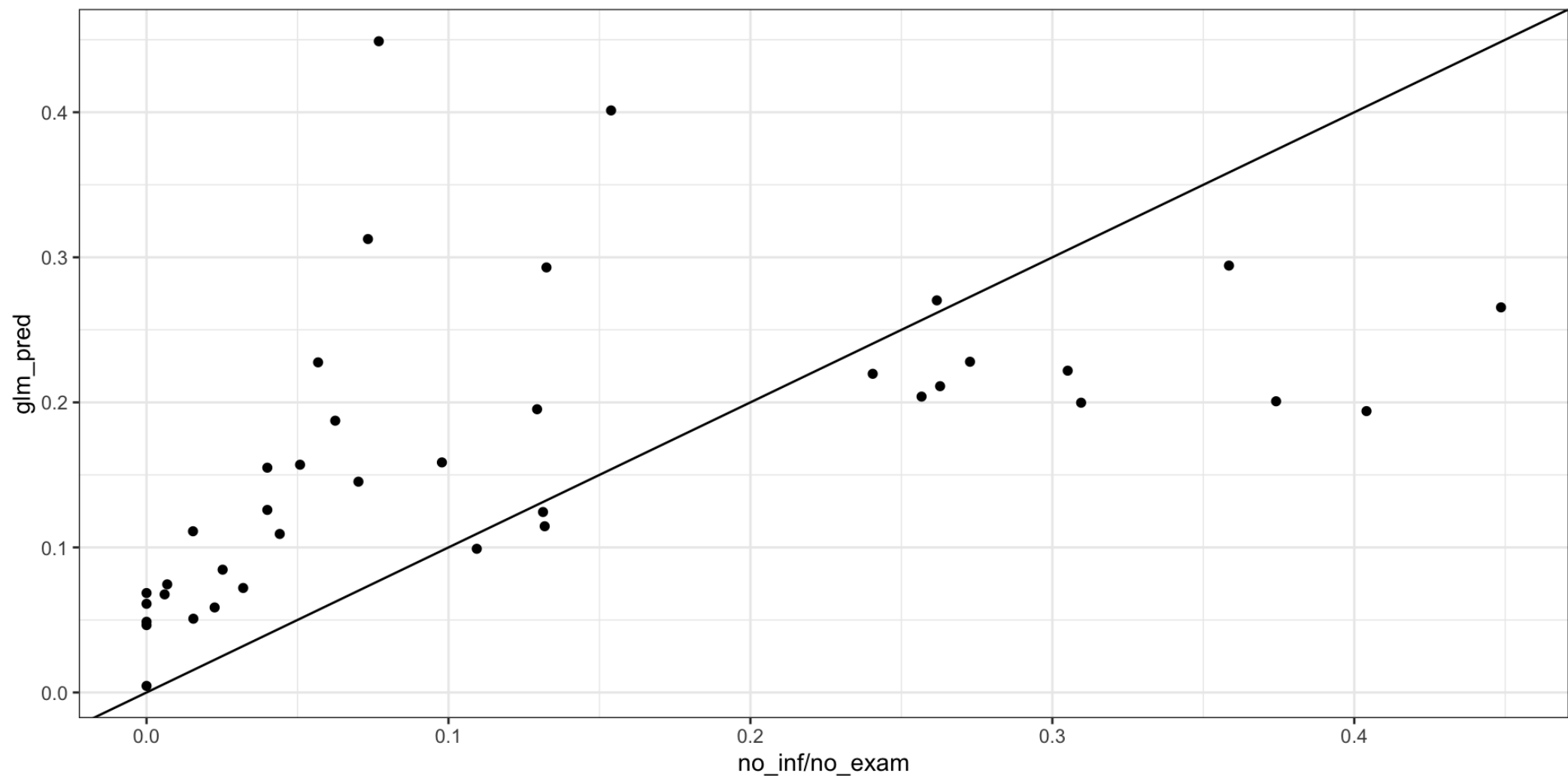
Predictions - Testing



Fit - Training



Fit - Testing



Fit - RMSE

Training

```
1 yardstick::rmse_vec(loaloe$no_inf/loaloe$no_exam, loaloe$glm_pred)
```

```
[1] 0.11176
```

Testing

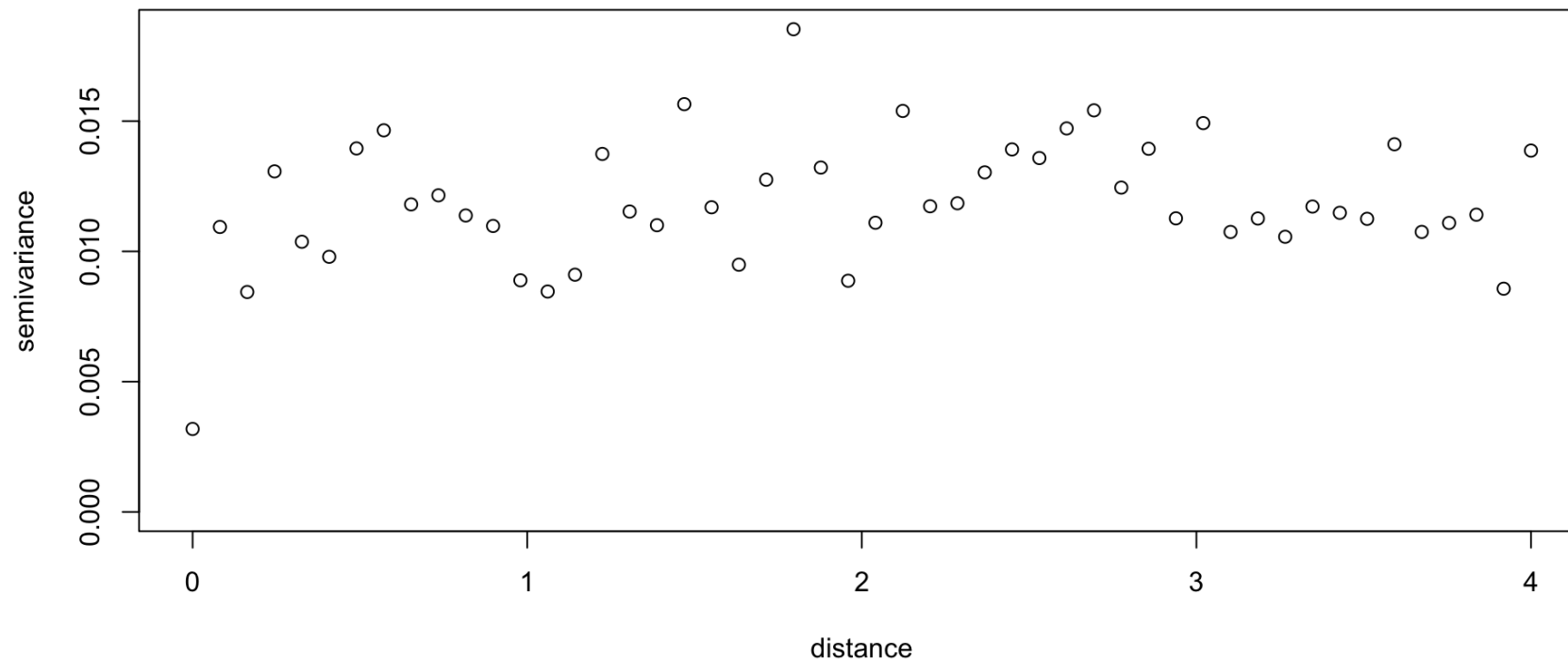
```
1 yardstick::rmse_vec(loaloe_test$no_inf/loaloe_test$no_exam, loaloe_test$
```

```
[1] 0.1192507
```

Spatial Structure?

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

variog: computing omnidirectional variogram



gpglm model

```
1 ll_gp = gpglm(  
2   no_inf ~ scale(elev):elev_f + scale(max9901):max_f + scale(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),  
8     phi=3, sigma.sq=1, w=0  
9   ),  
10  priors = list(  
11    beta.Normal=list(rep(0,7), rep(10,7)),  
12    phi.unif=c(3/4, 3/0.25), sigma.sq.ig=c(2, 2)  
13  ),  
14  tuning = list(  
15    "beta"=rep(0.1, 7),  
16    "phi"=0.6, "sigma.sq"=0.3, "w"=0.1  
17  ),  
18  n_batch = 400,  
19  batch_len = 50,  
20  verbose = TRUE,  
21  n_report = 10,  
22  chains=4
```

```
1 ll_gp
```

```
# A gpglm model (spBayes spGLM) with 4 chains, 9 variables, and 80000 iterations.
```

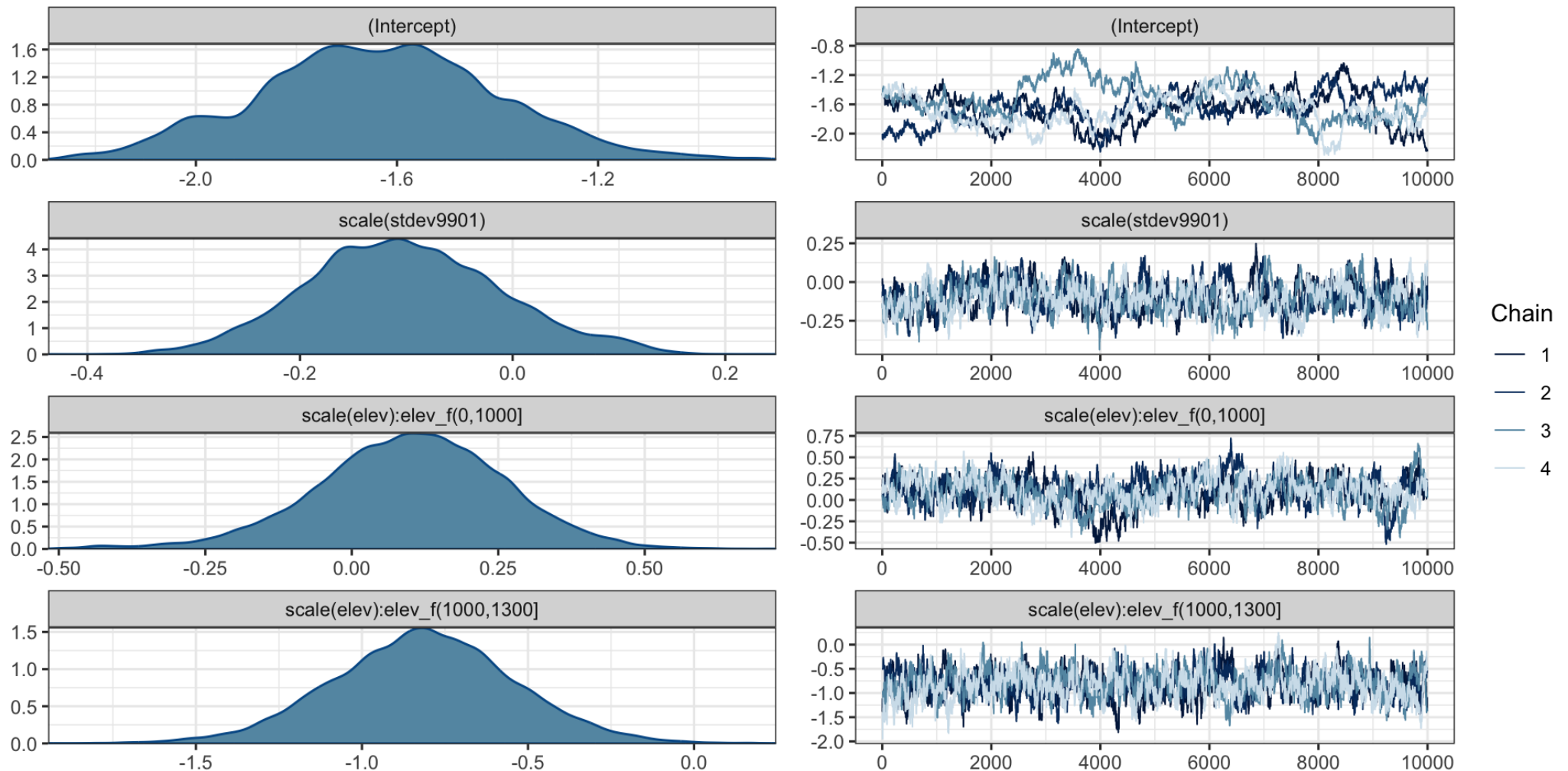
```
# A tibble: 9 × 10
```

	variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	(Interce...	-1.64	-1.64	0.237	0.234	-2.03	-1.24	1.15	22.1
2	scale(st...	-0.101	-0.105	0.0914	0.0916	-0.248	0.0573	1.01	299.
3	scale(el...	0.0992	0.104	0.160	0.154	-0.170	0.352	1.02	127.
4	scale(el...	-0.815	-0.813	0.274	0.265	-1.27	-0.363	1.01	378.
5	scale(el...	-1.51	-1.50	0.257	0.254	-1.94	-1.10	1.01	534.
6	scale(ma...	0.718	0.715	0.179	0.176	0.432	1.02	1.01	318.
7	scale(ma...	0.0990	0.100	0.139	0.137	-0.128	0.329	1.03	137.
8	sigma.sq	1.04	0.977	0.306	0.235	0.682	1.58	1.02	221.
9	phi	3.07	2.94	1.03	0.940	1.64	4.95	1.03	178.

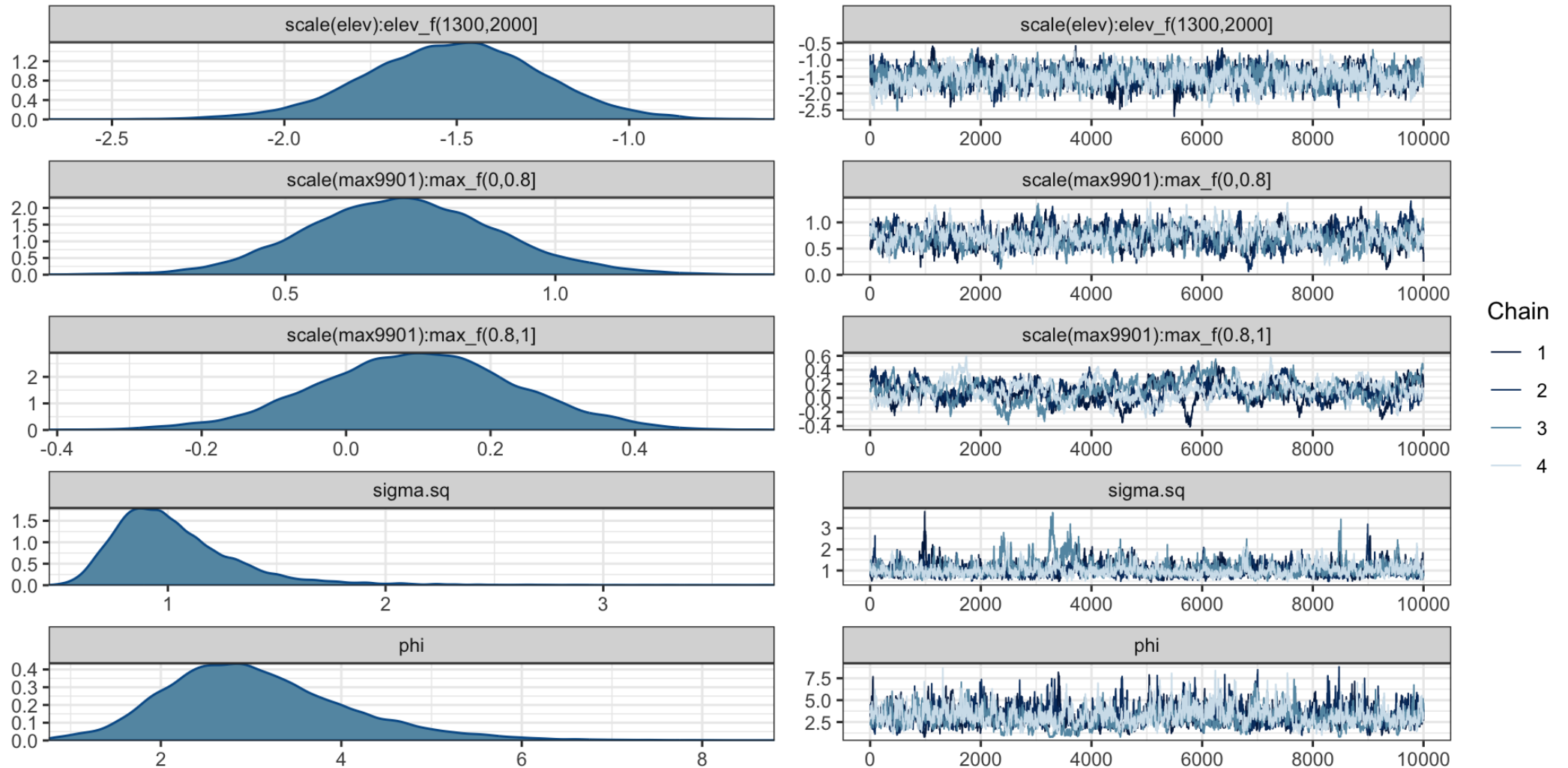
```
# i 1 more variable: ess_tail <dbl>
```

Diagnostics

```
1 vars = colnames(ll_gp$mcmc)
2 plot(ll_gp, vars=vars[1:4])
```

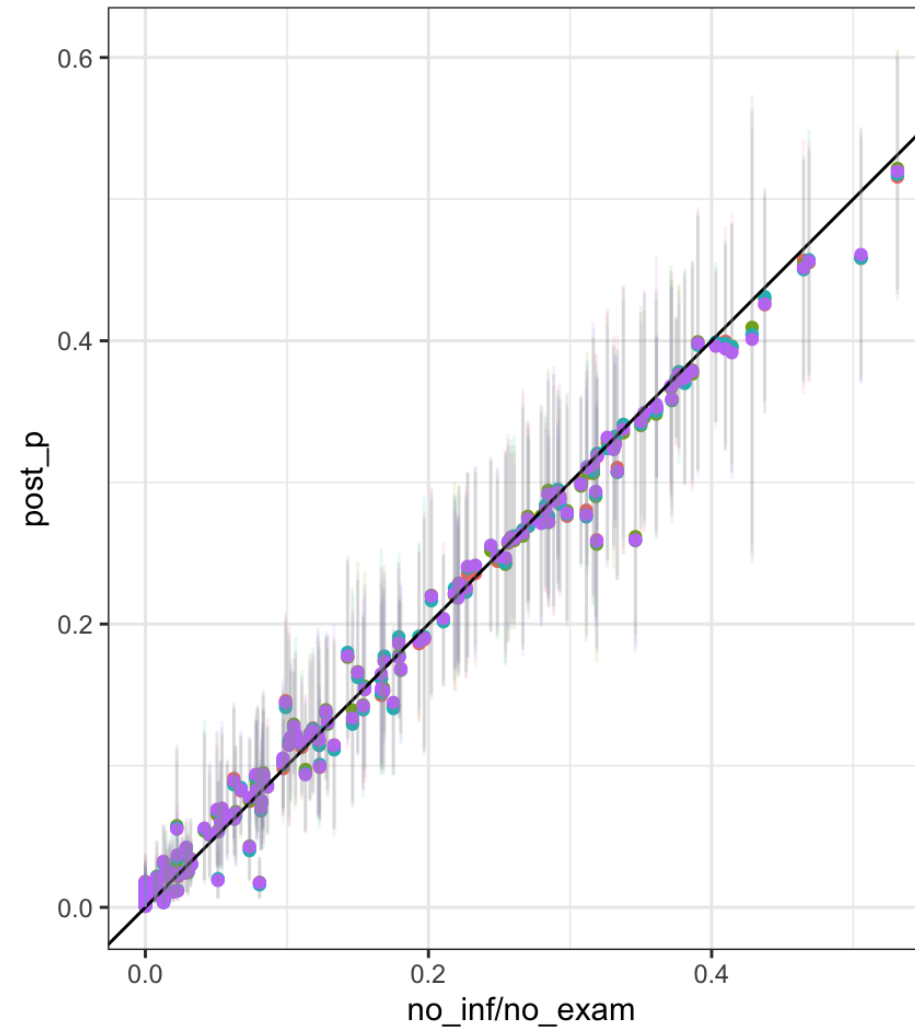


```
1 plot(ll_gp, vars=vars[5:9])
```



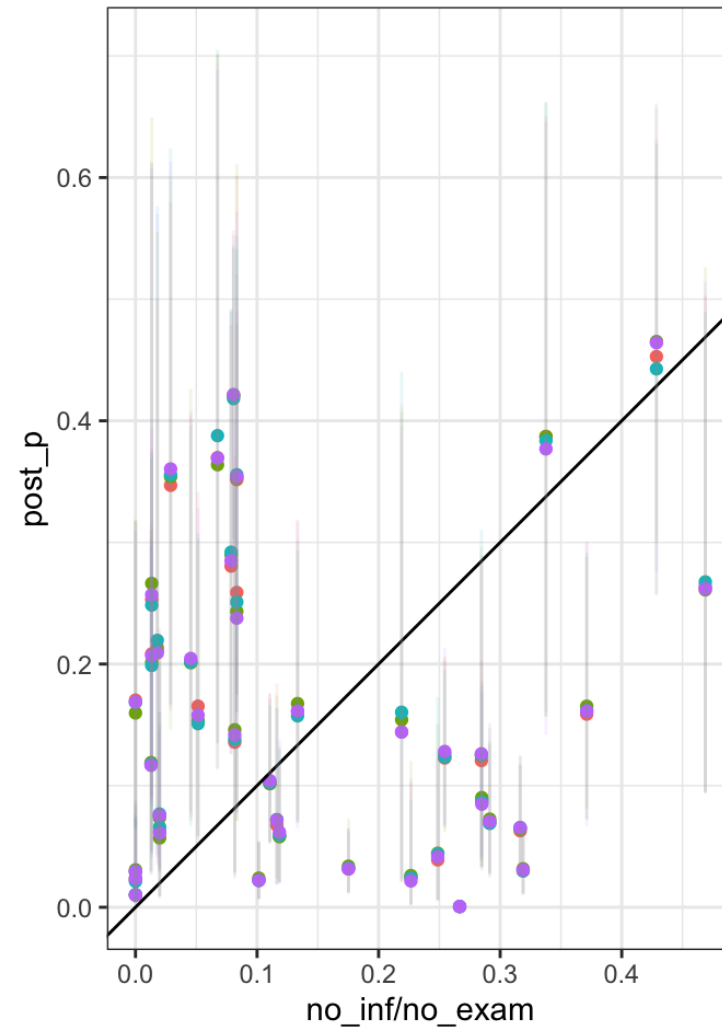
Prediction (training)

```
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

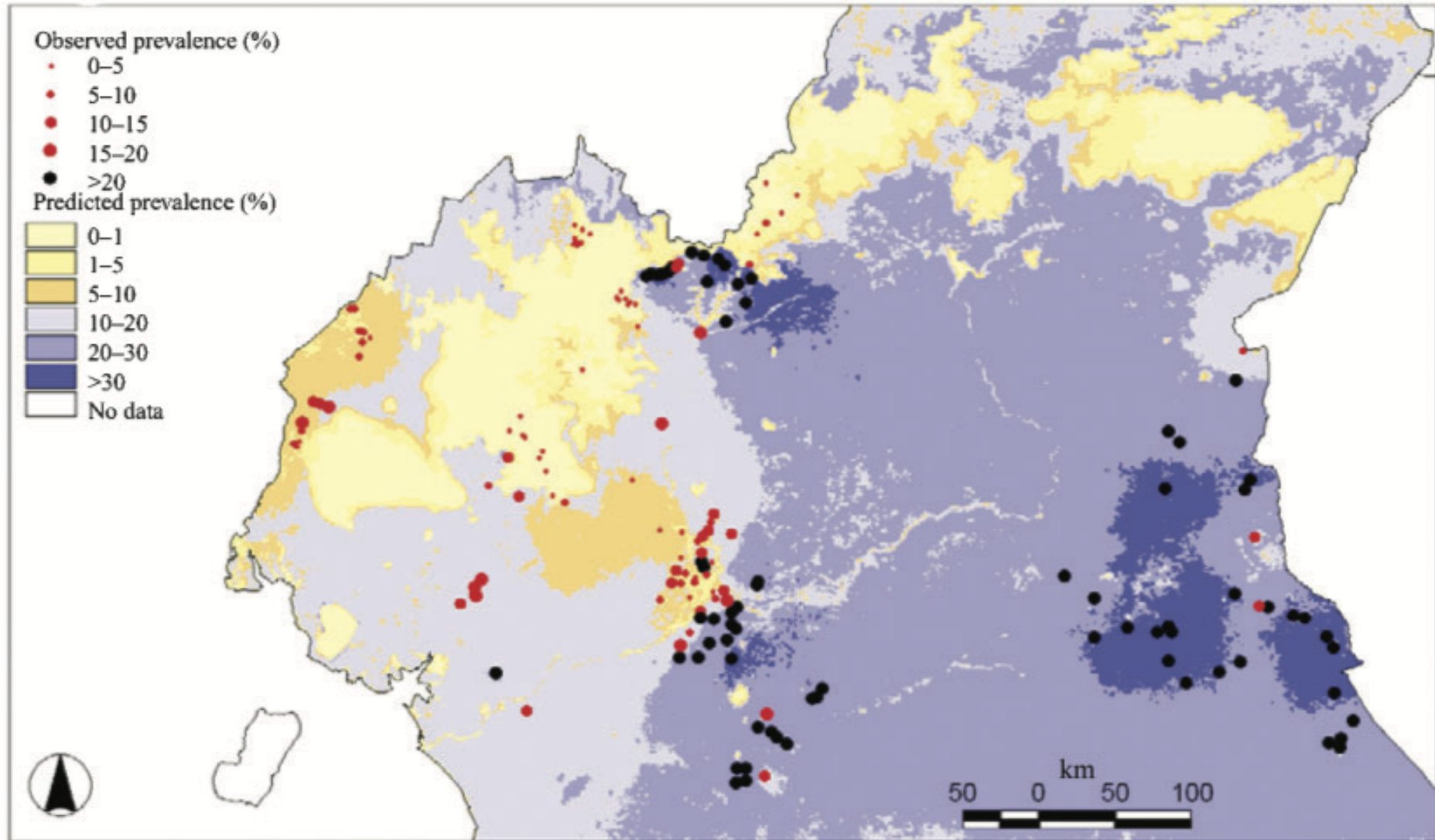
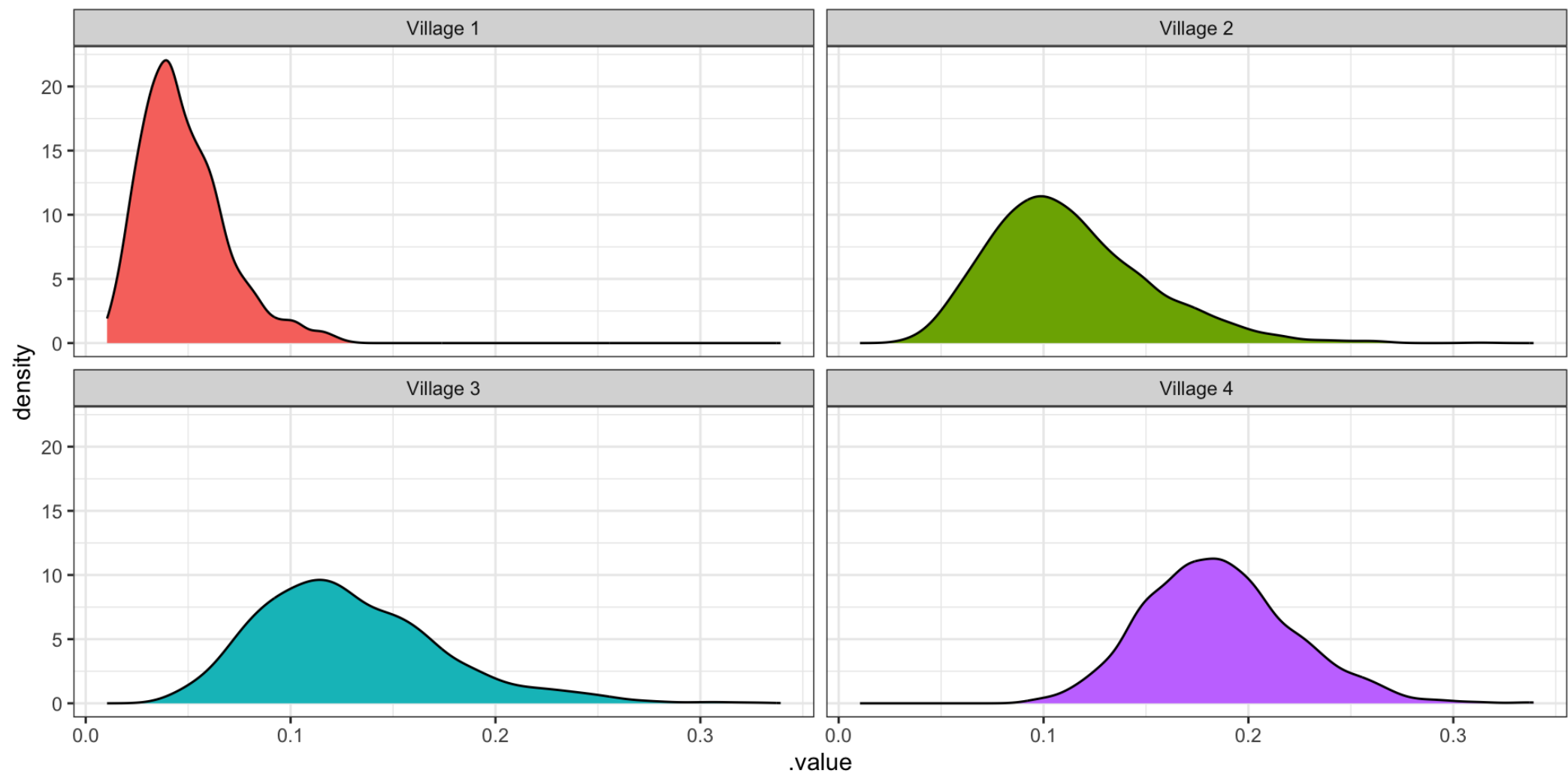


FIG. 2. Point estimates of the prevalence of *Loa loa* microfilaraemia, over-laid with the prevalences observed in field studies.

Exceedance Probability



Exceedance Probability Predictive Surface

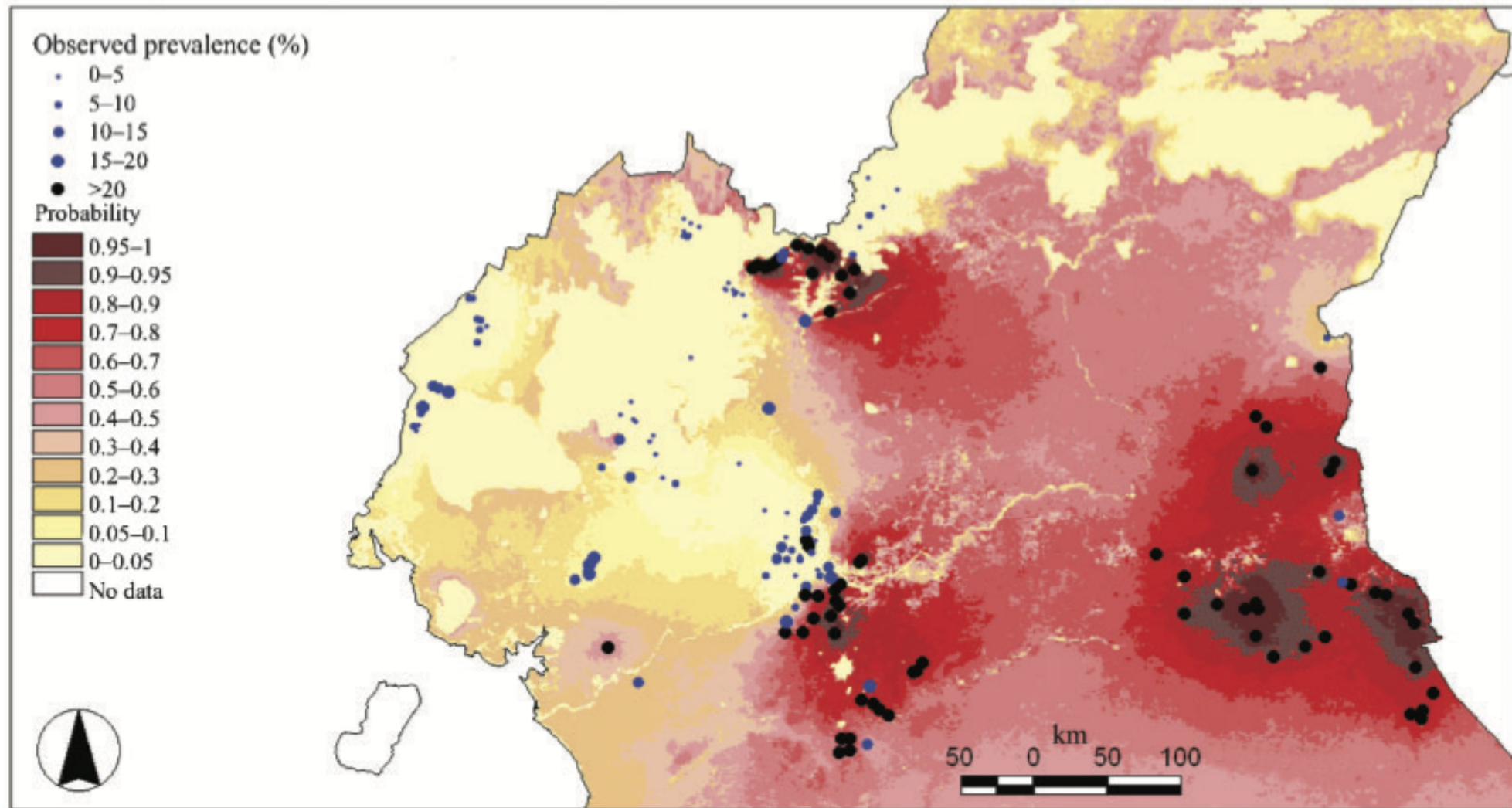


FIG. 4. A probability contour map, indicating the probability that the prevalence of *Loa loa* microfilaraemia in each area exceeds 20%, over-laid with the prevalences observed in field studies.

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*)

Wilson's Warbler
(*Wilsonia pusilla*)

138 individuals

163 individuals

14 locations

8 locations

6 loci

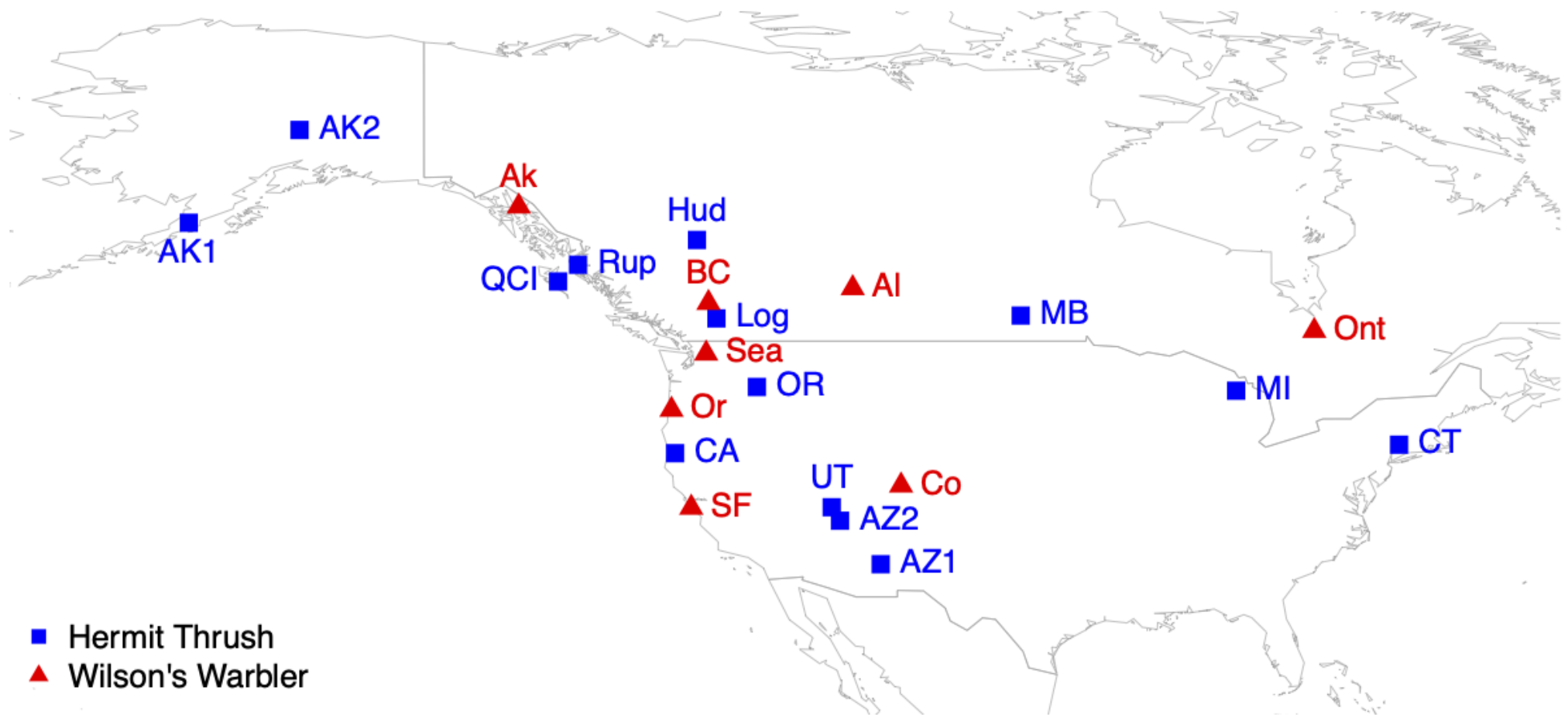
9 loci

9-27 alleles / locus

15-31 alleles / locus



Sampling Locations



Allele Frequency Model

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

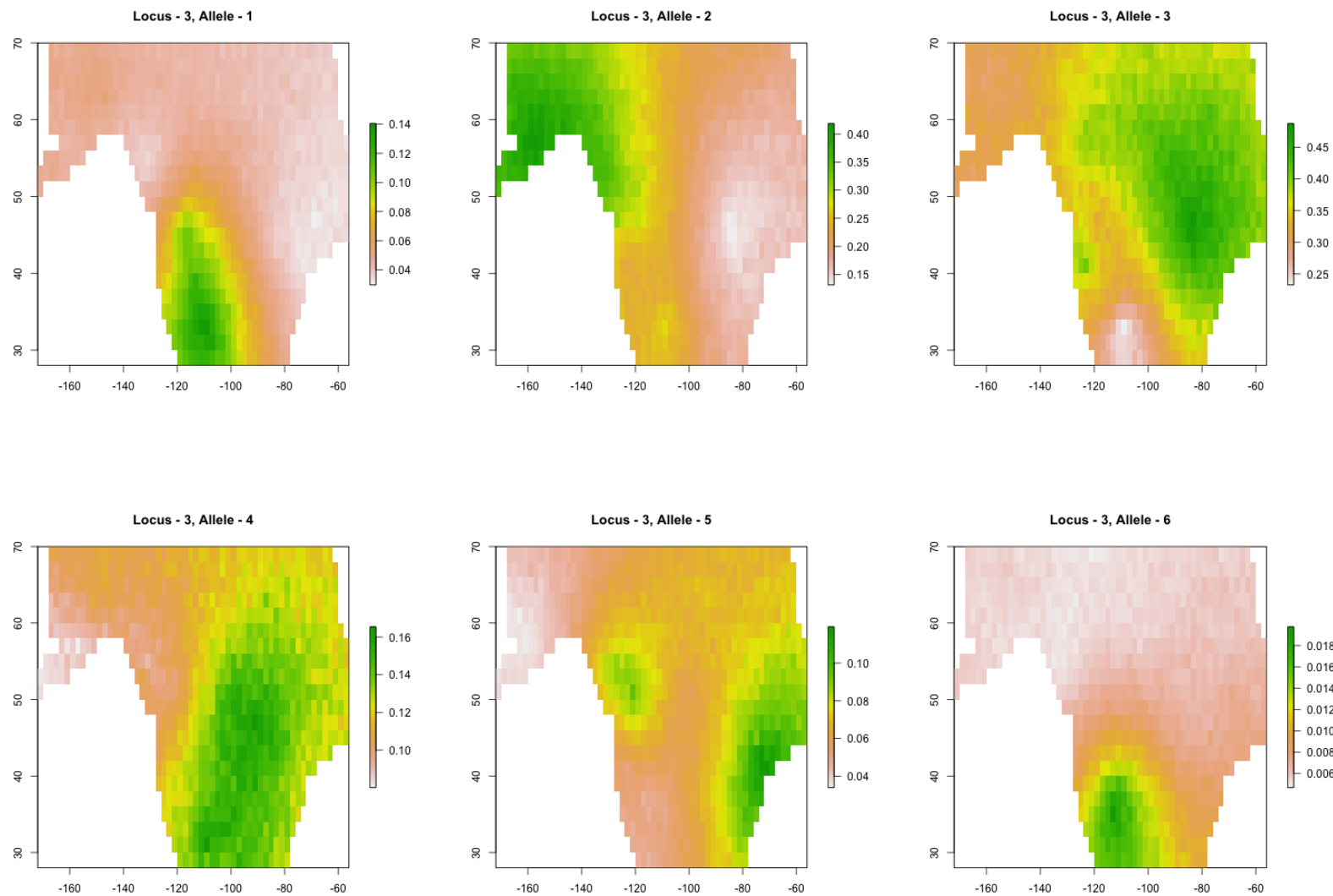
$$\mathbf{y}_{\cdot lk} | \mathbf{f}_{\cdot lk} \sim \text{Multinom}(\mathbf{f}_{\cdot lk}, n_{lk})$$

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

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

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

Predictions by Allele (Locus 3)



Genetic Assignment Model

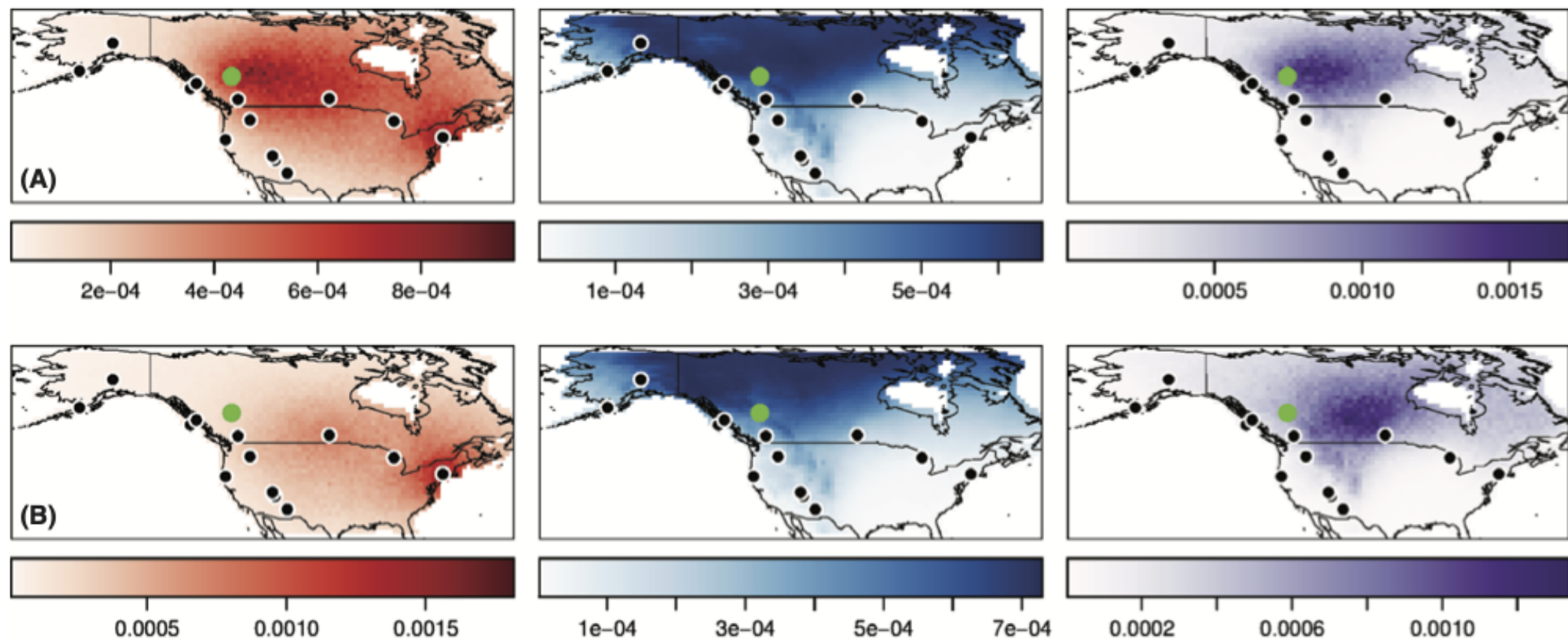
Assignment model assuming Hardy-Weinberg equilibrium and allowing for genotyping (δ) and single amplification (γ) errors.

$$P(S_G | \mathbf{f}, k) = \prod_1 P(i_1, j_1 | \mathbf{f}, k)$$

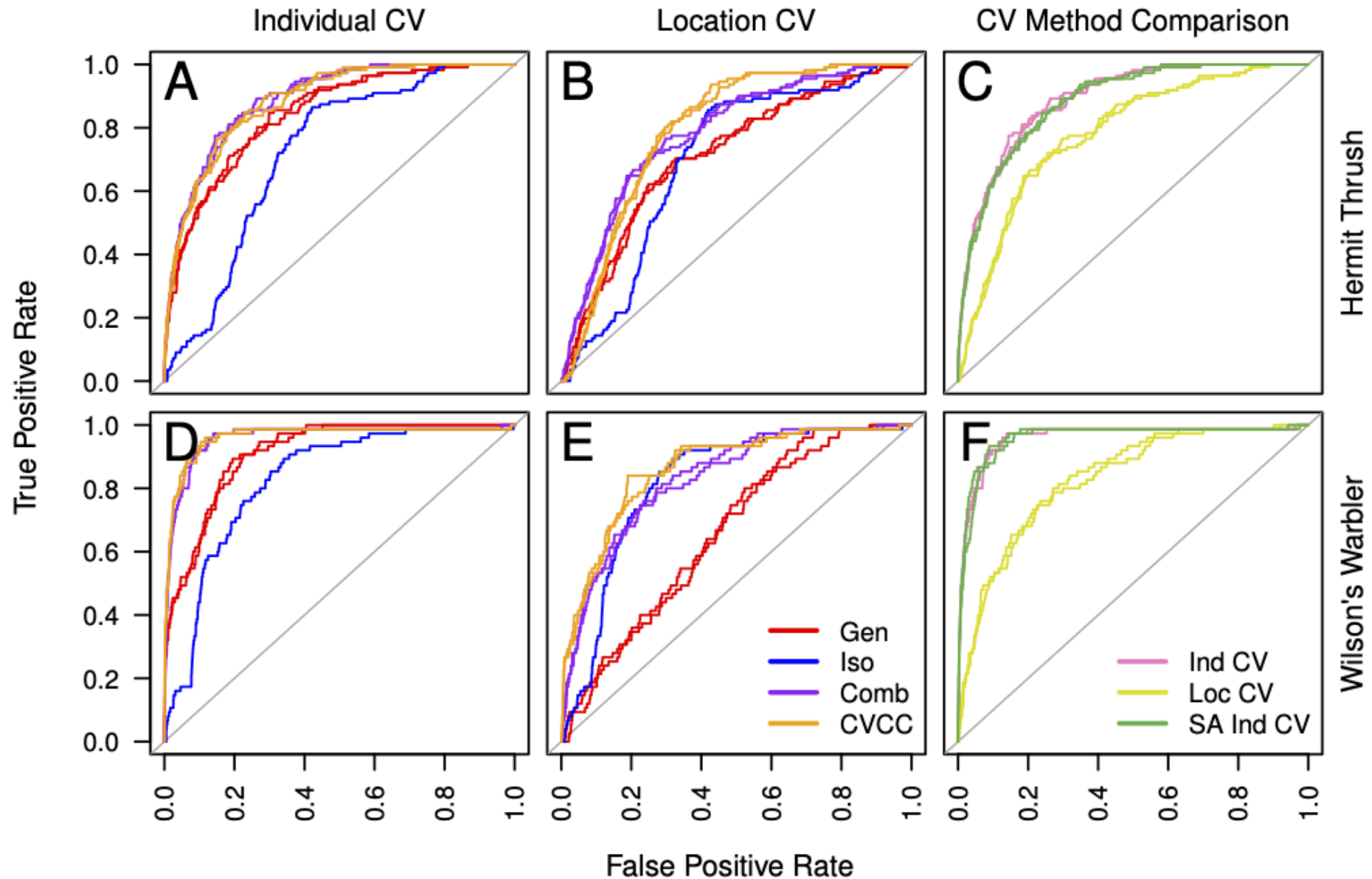
$$P(i_1, j_1 | \mathbf{f}, k) = \begin{cases} \gamma P(i_1 | \mathbf{f}, k) + (1 - \gamma) P(i_1 | \tilde{\mathbf{f}}, k)^2 & \text{if } i = j \\ (1 - \gamma) P(i_1 | \mathbf{f}, k) P(j_1 | \mathbf{f}, k) & \text{if } i \neq j \end{cases}$$

$$P(i_1 | \mathbf{f}, k) = (1 - \delta) f_{i_1 k} + \delta / m_1$$

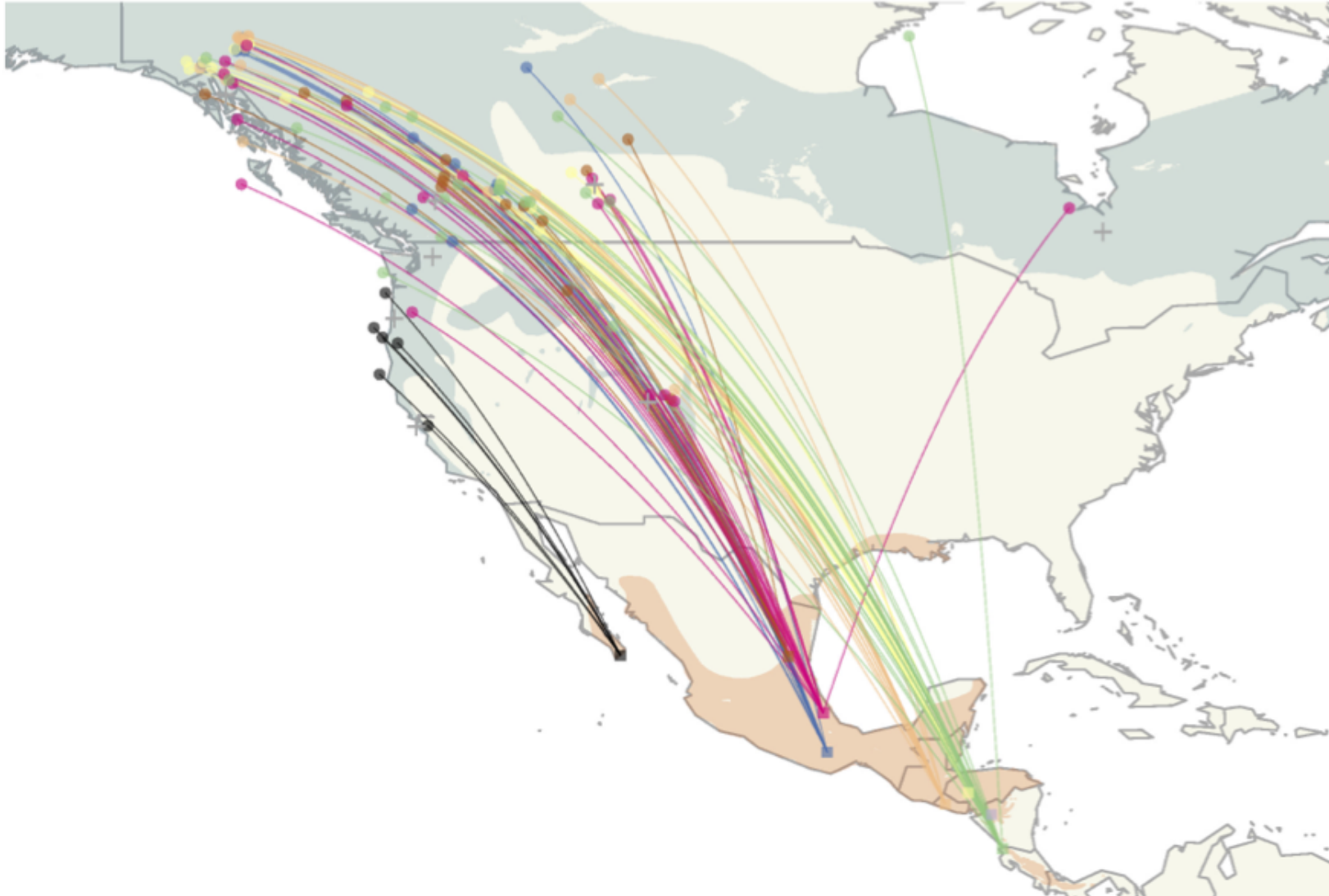
Combined Model



Model Assessment



Migratory Connectivity



Sta 344/644 - Fall 2023

