

# Occupancy detection modelling with ROccuPy

```
# Ensure everything is clean
rm(list=ls())
library(roccupy)

library(reticulate)

# If you are running this yourself, please make sure that reticulate uses the
# correct virtual environment.
# Set this to FALSE if you don't have a GPU.
roccupy::set_gpu(TRUE)
```

This document shows how on a small subset of eBird using 8,000 checklists and 32 species.

## The data format

```
# Load the example data
data(eBird)
names(eBird)
#> [1] "X_checklist"      "y_checklist"      "checklist_cell_ids"
#> [4] "X_env"
```

As you can see, the dataset consists of four different items. Let's go through these in turn.

```
head(eBird$X_checklist)
#>      last_edited_date      country country_code      state
#> S57652587 2019-06-24 12:37:13 United States      US      Montana
#> S57628599 2019-06-23 16:47:25 United States      US      Texas
#> S57370878 2019-06-14 14:33:59 United States      US      Arkansas
#> S57791363 2019-06-29 16:42:26 United States      US Massachusetts
#> S57297709 2020-07-08 16:05:34 United States      US      Virginia
#> S57116888 2019-06-05 17:01:15 United States      US      Minnesota
#>      state_code      county county_code iba_code bcr_code usfws_code
#> S57652587      US-MT      Missoula      US-MT-063      10
#> S57628599      US-TX      Montgomery      US-TX-339      25
#> S57370878      US-AR      Columbia      US-AR-027      25
#> S57791363      US-MA      Bristol      US-MA-005      30
#> S57297709      US-VA      Russell      US-VA-167      28
#> S57116888      US-MN      Steele      US-MN-147      11
#>      atlas_block      locality
#> S57652587      MPG Ranch--Baldy Draw (restricted access)
#> S57628599      Silver Leaf
#> S57370878      32 U.S. 79, Magnolia, Arkansas, US (33.273, -93.214)
#> S57791363      Stonehill College, North Easton US-MA (42.0613, -71.0805)
#> S57297709      o36081H7NW      Saltville, NW pt 6
#> S57116888      SW 24th Ave ponds
#>      locality_id locality_type observation_date time_observations_started
#> S57652587      L3293950      H      2019-06-19      20:30:00
```

```

#> S57628599      L766259      P      2019-06-22      07:49:00
#> S57370878      L9467519      P      2019-06-14      13:28:00
#> S57791363      L9556030      P      2019-06-29      07:46:00
#> S57297709      L9452300      P      2019-06-10      07:17:00
#> S57116888      L5926798      P      2019-06-05      09:58:00
#>      observer_id sampling_event_identifrier protocol_type protocol_code
#> S57652587      obs436394      S57652587      Stationary      P21
#> S57628599      obs193353      S57628599      Stationary      P21
#> S57370878      obs291133      S57370878      Stationary      P21
#> S57791363      obs342615      S57791363      Stationary      P21
#> S57297709      obs312088      S57297709      Stationary      P21
#> S57116888      obs144023      S57116888      Stationary      P21
#>      project_code duration_minutes effort_distance_km effort_area_ha
#> S57652587      EBIRD      120      NA      NA
#> S57628599      EBIRD      10      NA      NA
#> S57370878      EBIRD      10      NA      NA
#> S57791363      EBIRD      375      NA      NA
#> S57297709      EBIRD_ATL_VA      6      NA      NA
#> S57116888      EBIRD      2      NA      NA
#>      number_observers all_species_reported group_identifrier
#> S57652587      1      True
#> S57628599      1      True
#> S57370878      1      True
#> S57791363      2      True
#> S57297709      1      True
#> S57116888      1      True
#>      trip_comments      X      Y latitude longitude
#> S57652587      -2607295.8 1684608.4 46.70917 -114.01421
#> S57628599      -1566757.8 -516904.3 30.12508 -95.68785
#> S57370878      -1277362.7 -208200.5 33.27337 -93.21388
#> S57791363      697330.0 704869.6 42.06133 -71.08046
#> S57297709      -210591.6 110057.3 36.96730 -81.86510
#> S57116888      Clear with light winds -1103613.0 979505.4 44.07249 -93.25762
#>      time_to_next_sunset time_to_next_sunrise time_from_last_sunrise
#> S57652587      1.050556      9.200556      14.801944
#> S57628599      12.640556      22.555833      1.448333
#> S57370878      6.922778      16.585000      7.415833
#> S57791363      12.636111      21.429167      2.578611
#> S57297709      13.491389      22.841111      1.157500
#> S57116888      10.903333      19.548333      4.445556
#>      time_from_last_sunset is_up land_cover fold_id cell_id log_duration
#> S57652587      22.95417 True      71      3      1327      4.7874917
#> S57628599      11.36278 True      81      1      15153      2.3025851
#> S57370878      17.08361 True      22      3      13683      2.3025851
#> S57791363      11.36306 True      41      2      5981      5.9269260
#> S57297709      10.51667 True      41      3      11051      1.7917595
#> S57116888      13.10889 True      22      1      3786      0.6931472
#>      time_of_day time_of_day_fine dominant_land_cover daytimes_alt
#> S57652587      afternoon/evening      18-21      baseline      late-evening
#> S57628599      morning      6-9      baseline      early-morning
#> S57370878      afternoon/evening      12-15      developed      mid-day
#> S57791363      morning      6-9      forest      early-morning
#> S57297709      morning      6-9      forest      early-morning

```

#> S57116888	morning	9-12	developed	late-morning
--------------	---------	------	-----------	--------------

X\_checklist contains the observation-level covariates. These are things like the duration of the observations, the time of day, and so on – anything that could affect the detection process.

```
head(eBird$y_checklist)
```

```
#>      Selasphorus rufus Limosa fedoa Anas platyrhynchos Tyrannus forficatus
#> S57652587      FALSE      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      TRUE      FALSE
#>      Quiscalus mexicanus Acanthis flammea Sayornis saya Buteo jamaicensis
#> S57652587      FALSE      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      FALSE      FALSE
#>      Toxostoma crissale Sphyrapicus nuchalis Icterus parisorum
#> S57652587      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      FALSE
#>      Melanerpes erythrocephalus Contopus virens Chaetura pelagica
#> S57652587      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      TRUE
#> S57297709      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      FALSE
#>      Tringa semipalmata Myiarchus cinerascens Calcarius lapponicus
#> S57652587      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      FALSE
#>      Peucaea cassinii Spinus lawrencei Aix sponsa Vireo solitarius
#> S57652587      FALSE      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE      FALSE
#> S57116888      FALSE      FALSE      FALSE      FALSE
#>      Quiscalus quiscula Bucephala albeola Antrostomus vociferus
#> S57652587      FALSE      FALSE      FALSE
#> S57628599      FALSE      FALSE      FALSE
#> S57370878      FALSE      FALSE      FALSE
#> S57791363      FALSE      FALSE      FALSE
#> S57297709      FALSE      FALSE      FALSE
```

```

#> S57116888 TRUE FALSE FALSE
#> Calypte anna Aechmophorus clarkii Vermivora cyanoptera Columba livia
#> S57652587 FALSE FALSE FALSE FALSE
#> S57628599 FALSE FALSE FALSE FALSE
#> S57370878 FALSE FALSE FALSE FALSE
#> S57791363 FALSE FALSE FALSE FALSE
#> S57297709 FALSE FALSE FALSE FALSE
#> S57116888 FALSE FALSE FALSE FALSE
#> Perdix perdix Baeolophus inornatus Dryocopus pileatus
#> S57652587 FALSE FALSE FALSE
#> S57628599 FALSE FALSE FALSE
#> S57370878 FALSE FALSE FALSE
#> S57791363 FALSE FALSE FALSE
#> S57297709 FALSE FALSE FALSE
#> S57116888 FALSE FALSE FALSE
#> Salpinctes obsoletus
#> S57652587 FALSE
#> S57628599 FALSE
#> S57370878 FALSE
#> S57791363 FALSE
#> S57297709 FALSE
#> S57116888 FALSE

```

y\_checklist specifies whether each species was or was not observed for each checklist. X\_checklist and y\_checklist should have the same number of rows.

```

head(eBird$X_env)
#> bio1 bio2 bio3 bio4 bio5 bio6 bio7 bio8 bio9 bio10 bio12 bio13 bio14 bio15
#> 2 95 89 40 4757 220 -2 222 40 157 157 1003 142 33 45
#> 3 96 90 39 4851 222 -4 226 39 159 159 1074 149 37 43
#> 4 97 94 39 4962 230 -6 236 38 160 160 1183 159 42 41
#> 13 49 122 34 8032 248 -109 357 -48 110 151 349 40 21 21
#> 14 45 121 34 7977 244 -110 354 85 0 147 368 41 22 21
#> 18 67 119 32 8334 273 -89 362 -33 167 174 615 77 32 27
#> bio18 bio19 X0 X11 X12 X21 X22 X23 X24 X31 X41 X42 X43 X52
#> 2 119 376 0 1 0 0.0000000 0 0 0 0 0 0.0000000 0 0.0000000
#> 3 130 395 0 0 0 0.1103022 0 0 0 0 0 0.8154358 0 0.0000000
#> 4 147 422 0 0 0 0.0000000 0 0 0 0 0 0.0000000 0 0.0000000
#> 13 93 96 0 0 0 0.0000000 0 0 0 0 0 0.0000000 0 0.8944125
#> 14 100 99 0 0 0 0.0000000 0 0 0 0 0 0.8480958 0 0.1519042
#> 18 128 196 0 0 0 0.0000000 0 0 0 0 0 1.0000000 0 0.0000000
#> X71 X81 X82 X90 X95 x y
#> 2 0.0000000 0.0000000 0.0000000 0.0000000 0 -122.7500 48.91667
#> 3 0.0000000 0.0000000 0.0000000 0.07426202 0 -122.5833 48.91667
#> 4 0.0000000 0.3790737 0.6209263 0.00000000 0 -122.4167 48.91667
#> 13 0.1055876 0.0000000 0.0000000 0.00000000 0 -119.2500 48.91667
#> 14 0.0000000 0.0000000 0.0000000 0.00000000 0 -119.0833 48.91667
#> 18 0.0000000 0.0000000 0.0000000 0.00000000 0 -117.9167 48.91667
#> dominant_cover has_open_water has_deciduous_forest has_evergreen_forest
#> 2 Open Water True False False
#> 3 Evergreen Forest False False True
#> 4 Cultivated Crops False False False
#> 13 Shrub/Scrub False False False
#> 14 Evergreen Forest False False True

```

```

#> 18 Evergreen Forest      False      False      True
#>   has_mixed_forest has_shrub_or_scrub has_grassland_or_herbaceous
#> 2      False      False      False
#> 3      False      False      False
#> 4      False      False      False
#> 13     False      True      True
#> 14     False      True      False
#> 18     False      False      False
#>   has_pasture_or_hay has_cultivated_crops has_other has_developed has_wetlands
#> 2      False      False      False      False      False
#> 3      False      False      False      True      True
#> 4      True      True      False      False      False
#> 13     False      False      False      False      False
#> 14     False      False      False      False      False
#> 18     False      False      False      False      False

```

`X_env` contains the environmental covariates thought to influence whether a species is present or absent at each site. Because there are repeat visits, `X_env` will typically have fewer rows than `X_checklist`: there are fewer sites than observations.

```

head(eBird$checklist_cell_ids)
#> [1] 287 3337 3099 1445 2556 907

```

Finally, the `checklist_cell_ids` provide the link between sites and observations. Each entry specifies which site (or cell) the observation was made in. For example, in this case, the first observation was made in site 287. Please note that sites are numbered from zero, so this would correspond to `X_env[288]`, for example. Storing the data in this way is useful as some sites are visited far more frequently than others. You can see this here:

```

head(sort(table(eBird$checklist_cell_ids), decreasing = TRUE), 20)
#>
#> 889 1850 733 2100 1389 2259 1511 1433 3237 1396 2294 650 1154 1469 1542 2034
#> 46 38 28 28 26 26 25 23 23 21 21 20 20 20 20 20
#> 2268 140 2020 3153
#> 20 19 19 18

```

## Fitting a model to eBird

We'll now walk through the steps required to fit a multi-species occupancy detection model to this dataset using variational inference.

```

# To make the code a little less cluttered, we can attach the entries in "eBird":
attach(eBird, warn.conflicts = FALSE)

# We'll want to scale the continuous environment variables.
bio_cols <- colnames(X_env)[grepl('bio', colnames(X_env))]

X_env_bio <- X_env[, bio_cols]
X_env_bio_scaled <- scale(X_env_bio)

# We also want to use the "has_" covariates:
discrete_cols <- X_env[, grepl('has_', colnames(X_env))] == 'True'

full_X_env <- cbind(X_env_bio_scaled, discrete_cols)

```

```
# We need to standardise log_duration:
log_durations <- X_checklist$log_duration
log_duration_mean <- mean(log_durations)
log_duration_sd <- sd(log_durations)

X_checklist$log_duration_z <- (log_durations - log_duration_mean) / log_duration_sd
```

We've now preprocessed our environmental covariates so that the continuous covariates are scaled. You can take a look at the covariates we'll use here:

```
full_X_env <- data.frame(full_X_env)
head(full_X_env)
#>      bio1      bio2      bio3      bio4      bio5      bio6
#> 2 -0.2179925 -1.8282236  0.6832729 -1.82423887 -2.0346269  0.9577327
#> 3 -0.1977128 -1.7782151  0.5389009 -1.77478073 -1.9762140  0.9293344
#> 4 -0.1774331 -1.5781811  0.5389009 -1.71637803 -1.7425624  0.9009361
#> 13 -1.1508596 -0.1779429 -0.1829592 -0.10109616 -1.2168464 -0.5615758
#> 14 -1.2319785 -0.2279514 -0.1829592 -0.13003443 -1.3336722 -0.5757749
#> 18 -0.7858246 -0.3279684 -0.4717032  0.05780128 -0.4866852 -0.2775929
#>      bio7      bio8      bio9      bio10      bio12      bio13
#> 2 -2.22152068 -1.1990981  0.8692277 -1.3091184  0.1971732  0.5055968
#> 3 -2.15671252 -1.2104460  0.8874000 -1.2589956  0.3776577  0.6346936
#> 4 -1.99469213 -1.2217940  0.8964862 -1.2339342  0.6547395  0.8191177
#> 13 -0.03424543 -2.1977179  0.4421770 -1.4594869 -1.4653177 -1.3755285
#> 14 -0.08285155 -0.6884403 -0.5573033 -1.5597325 -1.4170190 -1.3570861
#> 18  0.04676477 -2.0274986  0.9600895 -0.8830746 -0.7891364 -0.6931595
#>      bio14      bio15      bio18      bio19 has_open_water
#> 2 -0.40699474  0.6469555 -1.0942074  0.9945195          1
#> 3 -0.26443311  0.5475352 -0.9999559  1.1124795          0
#> 4 -0.08623106  0.4481150 -0.8542945  1.2801069          0
#> 13 -0.83467965 -0.5460874 -1.3169837 -0.7438389          0
#> 14 -0.79903924 -0.5460874 -1.2570055 -0.7252136          0
#> 18 -0.44263515 -0.2478267 -1.0170926 -0.1229966          0
#>      has_deciduous_forest has_evergreen_forest has_mixed_forest
#> 2              0              0              0
#> 3              0              1              0
#> 4              0              0              0
#> 13             0              0              0
#> 14             0              1              0
#> 18             0              1              0
#>      has_shrub_or_scrub has_grassland_or_herbaceous has_pasture_or_hay
#> 2              0              0              0
#> 3              0              0              0
#> 4              0              0              1
#> 13             1              1              0
#> 14             1              0              0
#> 18             0              0              0
#>      has_cultivated_crops has_other has_developed has_wetlands
#> 2              0              0              0              0
#> 3              0              0              1              1
#> 4              1              0              0              0
#> 13             0              0              0              0
#> 14             0              0              0              0
#> 18             0              0              0              0
```

```

# We can generate a formula. You can also specify one as you like.
# It just has to be compatible with the patsy package.
env_formula <- ml_tools$patsy$create_formula(
  cov_names=bio_cols, main_effects = TRUE,
  quadratic_effects = TRUE, interactions = FALSE)
to_add <- paste(colnames(full_X_env)[grep('has_', colnames(full_X_env))],
  collapse = '+')

env_formula <- paste0(env_formula, '+', to_add)
obs_formula <- "protocol_type + daytimes_alt + log_duration_z + dominant_land_cover"

# Takes about 45 seconds on GPU; 17 minutes on CPU. So GPU definitely recommended!
start_time <- Sys.time()
fit_model <- msod_vi(env_formula, obs_formula, full_X_env, X_checklist,
  y_checklist = y_checklist,
  checklist_cell_ids = checklist_cell_ids, M=20L)
end_time <- Sys.time()
print(end_time - start_time)
#> Time difference of 30.95399 secs

# We can extract the draws for the coefficients as follows:
coef_draws <- coef(fit_model)

# This is a list:
names(coef_draws)
#> [1] "env_intercepts" "env_slopes" "obs_prior_means" "obs_prior_sds"
#> [5] "obs_slopes"

```

These are the inferences made by the model. We can go through a little bit of maths to understand what they mean.

First, we model the probability that species  $j$  is present at site  $i$ . We can write this as:

$$y_{ij} \sim \text{Bern}(\Psi_{ij}), \quad (1)$$

$$\text{logit}(\Psi_{ij}) = x_i^\top \beta_j + \gamma_j, \quad (2)$$

$$\beta_j \stackrel{iid}{\sim} \mathcal{N}(0, I), \quad (3)$$

$$\gamma_j \stackrel{iid}{\sim} \mathcal{N}(0, 10^2). \quad (4)$$

The  $\gamma_j$  here are the `env_intercepts`, and the  $\beta_j$  are the `env_slopes` reported by the model.

Next, we can take a look at the detection part of the model:

$$p(s_{ijk} = 1 \mid y_{ij} = 1) = p_{ijk}, \quad (5)$$

$$\text{logit}(p_{ijk}) = x_{ik}^{(obs)\top} \beta_j^{(obs)}, \quad (6)$$

where

$$\beta_{jl}^{(obs)} \stackrel{iid}{\sim} \mathcal{N}(\mu_l, \sigma_l^2), \quad (7)$$

$$\mu_l \stackrel{iid}{\sim} \mathcal{N}(0, 1), \quad (8)$$

$$\sigma_l \stackrel{iid}{\sim} \mathcal{H}(1). \quad (9)$$

This says that the (logit) of the probability of detecting species  $j$  at site  $i$  on the  $k$ -th visit, if it is present, is given by a linear function of the observation covariates and species-specific observation coefficients  $\beta_j^{(obs)}$ . These species coefficients are the `obs_slopes` in the model results. The next three lines specify the hierarchical prior on the observation covariates. The group means  $\mu_l$  and group standard deviations  $\sigma_l$  are reported in `obs_prior_means` and `obs_prior_sds`, respectively. The  $\mathcal{H}$  denotes the half-normal distribution.

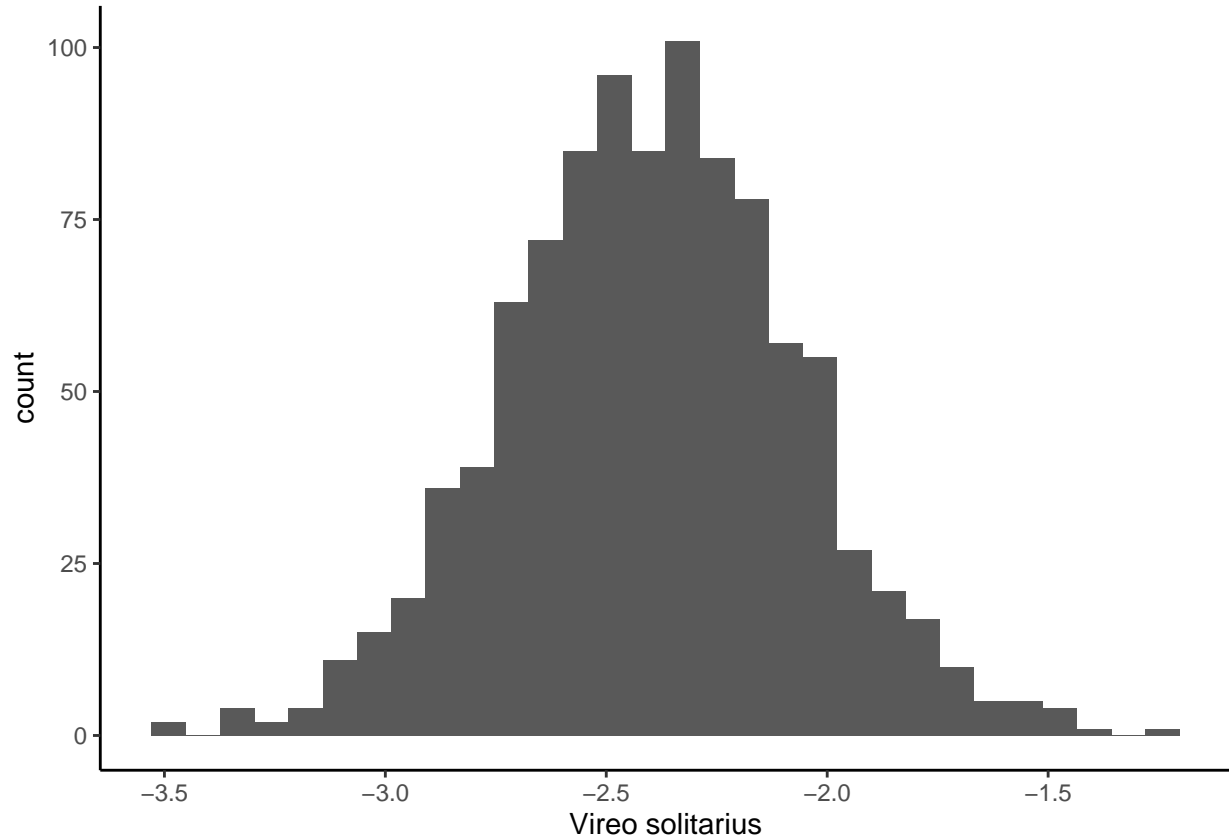
We will now quickly take a look at some of the results. First, let's look at the environment intercepts:

```
library(ggplot2)

intercept_draws <- coef_draws$env_intercepts

p <- ggplot(intercept_draws, aes(x=`Vireo solitarius`)) + geom_histogram() + theme_classic()

p
#> `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



We can see that *Vireo solitarius*, the Blue-headed Vireo, has a mean intercept of around -2.2 or so. We can also take a look at its estimates for environmental response:



```

head(coef_draws$env_slopes$`Vireo solitarius`)
#>      bio1      bio2      bio3      bio4      bio5      bio6      bio7
#> 1 -0.10486378  0.42001727  0.6818139 -0.36221573 -2.192209 -1.1969941  0.0658066
#> 2 -1.29107535 -0.20761952  1.0633597 -0.49049345 -2.035161 -1.1810886 -0.1227698
#> 3 -1.26441574 -0.05882092  0.7169656  0.22878650 -1.791161 -0.8062811 -0.1085716
#> 4  0.09298774  0.53111821  0.7715760 -0.02975783 -2.277434 -1.1348157  0.3900270
#> 5 -0.19279233 -0.26506305  0.8681947 -0.52821696 -2.268070 -0.9116366 -0.1168080
#> 6 -0.80679864 -0.53895271  0.3208386  0.10220165 -2.814021 -0.9075797 -0.9516600
#>      bio8      bio9      bio10      bio12      bio13      bio14      bio15
#> 1  1.5015209  0.29780349 -1.0867531  0.412518054  0.3529084  0.9389046 -0.0605137
#> 2  1.0866152  0.45523942 -1.5089664  0.485031933  0.5918568  1.3741454 -0.4982038
#> 3  0.9997180  0.47778189 -2.7741730  0.995091081  0.2576234  0.5909892 -0.3600232
#> 4  0.8844742  0.64706123 -1.7269028 -0.009495493  0.4149892  0.7916808 -0.3328453
#> 5  1.6359766 -0.06158514 -0.7609023  0.797966361  0.3232576  1.0653034 -0.4215962
#> 6  1.6682585  0.33342552 -1.2476581  0.504279435  0.6244470  1.2935359 -0.2144130
#>      bio18      bio19 I(bio1 ** 2) I(bio2 ** 2) I(bio3 ** 2) I(bio4 ** 2)
#> 1 -0.003550676 -0.04629546 -0.4015937 -1.457498 -0.72656262  0.71685785
#> 2 -0.635980487 -0.72637528 -0.6130591 -1.378089  0.17920132 -0.27650139
#> 3  0.079297096  0.25634286  0.2853603 -1.947917 -0.16160108  0.63295603
#> 4 -0.474355757 -0.67996174  0.3032016 -1.772970  0.44291136 -0.08907862
#> 5 -0.885753334 -0.20637357 -0.1320568 -1.325682 -0.01164698  0.61222142
#> 6 -0.506917179  0.50236696 -0.2458124 -1.209448 -0.04375888  0.43802938
#> I(bio5 ** 2) I(bio6 ** 2) I(bio7 ** 2) I(bio8 ** 2) I(bio9 ** 2)
#> 1 -0.004916164  0.8536497 -0.37873831 -0.17845592  0.251486778
#> 2 -0.244909480  1.0091422  0.03346084 -0.39483547  0.100979954
#> 3  0.195825219  0.7131346 -0.70964366 -0.70333099  0.109276921
#> 4  0.285526186  0.8503272 -0.34073558 -0.22733568 -0.084633157
#> 5  0.113791376  1.5606951 -0.10474505 -0.01200119  0.588600576
#> 6  0.413266838  0.9677999 -0.02342778 -0.67983896 -0.007551449
#> I(bio10 ** 2) I(bio12 ** 2) I(bio13 ** 2) I(bio14 ** 2) I(bio15 ** 2)
#> 1 -2.061425  0.8400455  0.3207242  0.7986960 -1.932936
#> 2 -2.115256  0.6233693 -1.1142004  0.5791370 -1.770955
#> 3 -2.161479  1.0006667 -0.5090131  0.8337358 -2.329790
#> 4 -1.633539  0.5906221 -2.0529468  1.0321314 -1.944853
#> 5 -1.478302  0.8830516 -1.4430150  0.7125186 -1.887674
#> 6 -1.946713  0.6954952 -0.8866531  0.8584982 -1.883551
#> I(bio18 ** 2) I(bio19 ** 2) has_open_water has_deciduous_forest
#> 1 -0.3238877 -1.535397  0.88045436  0.68439299
#> 2 -0.8915332 -1.061627  1.10225677 -0.05311945
#> 3 -0.8247463 -1.680576 -0.06048737 -0.58519131
#> 4 -1.2809954 -1.146700  0.34461737  0.40474352
#> 5 -0.7828191 -1.012997  1.10716581  0.20325838
#> 6 -0.9747421 -1.452904  1.13865292  0.34702125
#> has_evergreen_forest has_mixed_forest has_shrub_or_scrub
#> 1  0.8824447 -0.02476373 -0.1332018
#> 2  0.6514336  0.53232908 -1.7284856
#> 3 -0.2245940  0.85479134  0.2664584
#> 4 -0.6496049  0.08695910  1.5793399
#> 5  0.7463018  0.63738561 -0.5690021
#> 6  0.8171493 -0.11098851  0.5292687
#> has_grassland_or_herbaceous has_pasture_or_hay has_cultivated_crops
#> 1  0.4874155 -1.40105355  0.6779245
#> 2  0.1430249 -1.03263581 -0.2372007

```

```
#> 3          0.6202700          -0.76251704          0.2359560
#> 4          -0.2828451          -0.01246295          0.1210102
#> 5          0.8751560          -1.10828817          -0.3969409
#> 6          3.3052211          -0.28180158          0.7449379
#>      has_other has_developed has_wetlands
#> 1 -0.02464254  -0.03573005  -0.9258153
#> 2  0.63000447   0.50606602  -1.4878684
#> 3  0.62231028   0.38198507  -1.4287841
#> 4  1.32716250   0.15706901  -0.6281162
#> 5  0.33315670  -0.24643466   0.1585156
#> 6  0.77404428   0.17297459   0.3651171
```

These are 1000 draws. We can summarise them using their means and sds:

```
mean_slopes <- colMeans(coef_draws$env_slopes$`Vireo solitarius`)
sd_slopes <- apply(coef_draws$env_slopes$`Vireo solitarius`, 2, sd)
```

```
cbind(mean_slopes, sd_slopes)
#>      mean_slopes sd_slopes
#> bio1          -0.829004964 0.5357166
#> bio2          -0.033685147 0.4255910
#> bio3           0.539986668 0.5292552
#> bio4          -0.187308473 0.3698508
#> bio5          -2.167238673 0.4121657
#> bio6          -0.790055794 0.2943206
#> bio7          -0.216519547 0.4146649
#> bio8           0.802701833 0.3784570
#> bio9           0.271889948 0.2364641
#> bio10         -1.244003681 0.5007200
#> bio12          0.537231591 0.3778799
#> bio13          0.546059611 0.6316875
#> bio14          1.012645227 0.3366555
#> bio15         -0.219424978 0.3958828
#> bio18         -0.294348161 0.2724409
#> bio19          0.114987647 0.4220752
#> I(bio1 ** 2)    0.018623602 0.2872157
#> I(bio2 ** 2)   -1.583895997 0.3227737
#> I(bio3 ** 2)   -0.072594491 0.3739818
#> I(bio4 ** 2)    0.287255953 0.5369370
#> I(bio5 ** 2)    0.005702245 0.2673361
#> I(bio6 ** 2)    0.885659797 0.2159642
#> I(bio7 ** 2)   -0.348706578 0.2674430
#> I(bio8 ** 2)   -0.494174344 0.3379016
#> I(bio9 ** 2)    0.168826816 0.2207386
#> I(bio10 ** 2)  -1.883133493 0.3557918
#> I(bio12 ** 2)   0.823141415 0.2283174
#> I(bio13 ** 2)  -0.730311621 0.8034509
#> I(bio14 ** 2)   0.725245180 0.2255079
#> I(bio15 ** 2)  -1.891923240 0.2010848
#> I(bio18 ** 2)  -0.935789836 0.3292119
#> I(bio19 ** 2)  -0.709825644 0.6072110
#> has_open_water    0.893178930 0.5216916
#> has_deciduous_forest 0.119416605 0.4695398
#> has_evergreen_forest 0.551247700 0.7371207
```

```
#> has_mixed_forest      0.480789277 0.3852131
#> has_shrub_or_scrub    0.249171286 1.1279832
#> has_grassland_or_herbaceous 0.273161612 0.9921240
#> has_pasture_or_hay    -0.768037101 0.7020866
#> has_cultivated_crops  -0.274227702 0.4839496
#> has_other             -0.040499073 0.8590719
#> has_developed         -0.083433403 0.3228047
#> has_wetlands          -0.114492225 0.6124681
```

It appears that `bio5` seems to be associated with a decreased probability of presence for this bird. What's this one?

```
ml_tools$sdm$bioclim_lookup$bio5
#> [1] "Max Temperature of Warmest Month"
```

Indeed, looking at the range map on All About Birds, this seems plausible, as this species tends to breed in the North of the US, where it is cool.

Let's now take a look at the observation process. Summarising the group means gives:

```
colMeans(coef_draws$obs_prior_means)
#> Intercept protocol_type[T.Stationary]
#> -2.49881363 -0.42606944
#> protocol_type[T.Traveling] daytimes_alt[T.dusk]
#> 0.08784487 -1.10847239
#> daytimes_alt[T.early-evening] daytimes_alt[T.early-morning]
#> -0.13834148 0.09520711
#> daytimes_alt[T.late-evening] daytimes_alt[T.late-morning]
#> -0.24121066 -0.09719039
#> daytimes_alt[T.mid-day] daytimes_alt[T.night]
#> -0.26725315 -1.79576086
#> dominant_land_cover[T.developed] dominant_land_cover[T.forest]
#> -0.62704757 -0.34197943
#> dominant_land_cover[T.water] log_duration_z
#> -0.46095509 0.50283422
```

And the group standard deviations are:

```
colMeans(coef_draws$obs_prior_sds)
#> Intercept protocol_type[T.Stationary]
#> 1.54690767 0.48049841
#> protocol_type[T.Traveling] daytimes_alt[T.dusk]
#> 0.01802962 1.85797661
#> daytimes_alt[T.early-evening] daytimes_alt[T.early-morning]
#> 0.12741202 0.29988930
#> daytimes_alt[T.late-evening] daytimes_alt[T.late-morning]
#> 0.39341968 0.02828303
#> daytimes_alt[T.mid-day] daytimes_alt[T.night]
#> 0.06021645 2.05001878
#> dominant_land_cover[T.developed] dominant_land_cover[T.forest]
#> 0.93812872 0.72631191
#> dominant_land_cover[T.water] log_duration_z
#> 0.86305956 0.25462241
```

Let's look at how the observation slopes look for our example species:

```
means <- colMeans(coef_draws$obs_slopes$`Vireo solitarius`)
sds <- apply(coef_draws$obs_slopes$`Vireo solitarius`, 2, sd)
```

```
cbind(means, sds)
```

```
#>
#> Intercept -2.64546480 0.12684826
#> protocol_type[T.Stationary] -0.78608504 0.16186733
#> protocol_type[T.Traveling] 0.08685062 0.04049546
#> daytimes_alt[T.dusk] -1.78914672 1.90869649
#> daytimes_alt[T.early-evening] -0.18640601 0.14139676
#> daytimes_alt[T.early-morning] 0.01264768 0.17784061
#> daytimes_alt[T.late-evening] -0.62103022 0.30146692
#> daytimes_alt[T.late-morning] -0.09576107 0.04043839
#> daytimes_alt[T.mid-day] -0.26845347 0.10420017
#> daytimes_alt[T.night] -2.78459058 1.73225427
#> dominant_land_cover[T.developed] -0.76629380 0.54631619
#> dominant_land_cover[T.forest] 1.54860303 0.19084752
#> dominant_land_cover[T.water] -1.07591005 0.47419525
#> log_duration_z 0.47080436 0.17892243
```

This suggests, for example, that *Vireo solitarius* is considerably more likely to be detected in the forest than by water, which seems reasonable.

```
# How about a plot of detectability by day vs detectability by night?
```

```
obs_slopes <- coef_draws$obs_slopes
```

```
# This is a list of species names -> draws. Let's compute the means.
```

```
obs_slope_means <- lapply(obs_slopes, colMeans)
```

```
obs_slope_means <- data.frame(do.call(rbind, obs_slope_means), check.names = FALSE)
```

```
# For easier plotting:
```

```
obs_slope_means$species_name <- row.names(obs_slope_means)
```

```
head(obs_slope_means)
```

```
#> Intercept protocol_type[T.Stationary]
#> Selasphorus rufus -1.6171093 0.18111553
#> Limosa fedoa -3.5288797 -0.09087301
#> Anas platyrhynchos -1.1266217 -0.83152459
#> Tyrannus forficatus -0.5225457 -0.96869937
#> Quiscalus mexicanus -1.0061450 -0.77314862
#> Acanthis flammea -3.2946646 -0.81597892
#> protocol_type[T.Traveling] daytimes_alt[T.dusk]
#> Selasphorus rufus 0.09228801 -2.8591822
#> Limosa fedoa 0.08510109 -0.6040012
#> Anas platyrhynchos 0.09001710 0.4778010
#> Tyrannus forficatus 0.08435139 -0.4512533
#> Quiscalus mexicanus 0.09840906 -0.5381712
#> Acanthis flammea 0.08846861 -1.5039558
#> daytimes_alt[T.early-evening] daytimes_alt[T.early-morning]
#> Selasphorus rufus -0.15061968 0.201210870
#> Limosa fedoa -0.10161734 0.017603353
#> Anas platyrhynchos -0.03022044 -0.050746654
#> Tyrannus forficatus -0.07296833 -0.004024776
```

```

#> Quiscalus mexicanus -0.08072197 0.199406425
#> Acanthis flammea -0.12787787 0.043756914
#> daytimes_alt[T.late-evening] daytimes_alt[T.late-morning]
#> Selasphorus rufus -0.3186496 -0.10458377
#> Limosa fedoa -0.2906004 -0.09793923
#> Anas platyrhynchos -0.1012053 -0.11827971
#> Tyrannus forficatus -0.1656572 -0.09137659
#> Quiscalus mexicanus -0.2321436 -0.10015466
#> Acanthis flammea -0.2355961 -0.10074227
#> daytimes_alt[T.mid-day] daytimes_alt[T.night]
#> Selasphorus rufus -0.2604732 -2.090174
#> Limosa fedoa -0.2800858 -2.161803
#> Anas platyrhynchos -0.2670495 -2.197799
#> Tyrannus forficatus -0.2736187 -2.363187
#> Quiscalus mexicanus -0.2471762 -3.589080
#> Acanthis flammea -0.2707369 -2.777670
#> dominant_land_cover[T.developed]
#> Selasphorus rufus -1.3616655
#> Limosa fedoa -1.6009207
#> Anas platyrhynchos 0.2231158
#> Tyrannus forficatus -1.1599967
#> Quiscalus mexicanus 0.5773594
#> Acanthis flammea -0.7311354
#> dominant_land_cover[T.forest] dominant_land_cover[T.water]
#> Selasphorus rufus -0.2155620 -0.2990547
#> Limosa fedoa -0.8706316 1.0217136
#> Anas platyrhynchos -0.7848296 0.5053704
#> Tyrannus forficatus -0.8345924 -0.8333077
#> Quiscalus mexicanus -0.6857267 0.2933161
#> Acanthis flammea -0.2146915 -0.5906985
#> log_duration_z species_name
#> Selasphorus rufus 0.7341822 Selasphorus rufus
#> Limosa fedoa 0.3079680 Limosa fedoa
#> Anas platyrhynchos 0.5105932 Anas platyrhynchos
#> Tyrannus forficatus -0.1585740 Tyrannus forficatus
#> Quiscalus mexicanus 0.3329168 Quiscalus mexicanus
#> Acanthis flammea 0.5405778 Acanthis flammea

```

```

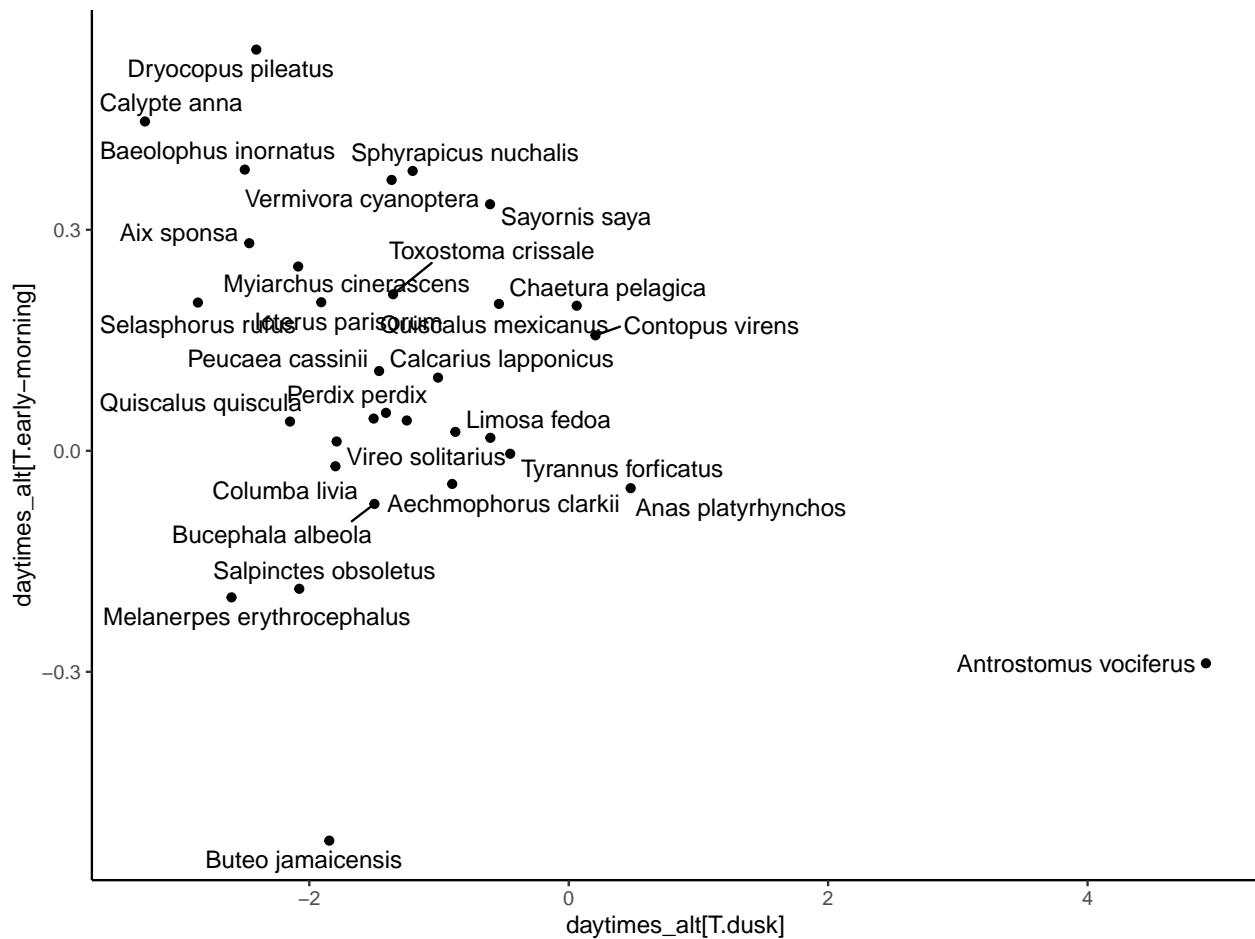
# You should see Antrostomus vociferus being much more likely to be detected at dusk than at dawn (the
library(ggrepel)

```

```

ggplot(obs_slope_means, aes(x=`daytimes_alt[T.dusk]`, y=`daytimes_alt[T.early-morning]`, label=species_
#> Warning: ggrepel: 3 unlabeled data points (too many overlaps). Consider
#> increasing max.overlaps

```



*# We can also predict. Here, let's just use the training data.*

*# If we want to predict the probabilities of presence, we can use:*

```
env_preds <- predict(fit_model, full_X_env, type='env')
```

*# If we want the probability of detection, we can use:*

```
obs_preds <- predict(fit_model, full_X_env, X_checklist, type='obs')
```

A last noteworthy feature of the package is the ability to save and restore models. You can do this as follows:

*# Save the model*

```
save_model(fit_model, 'save_test')
```

*# Restore it:*

```
restored_model <- restore_model('./save_test/')
```