Occupancy detection modelling with ROccuPy

```r
library(roccupy)
```

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

The data format

```r
# 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.

```r
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_identifier 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_identifier
#> 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 0.00000000   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:

```r
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
```

Let's make sure reticulate works as needed. Note that this will depend on your setup. I'm using reticulate with conda and an environment called `pymc3`, but you will need to change this to your own version.

```r
library(reticulate)

# You may have a different setup.
# I installed the package into my "pymc3" conda environment, so:
use_condaenv('pymc3')
```

```r
library(roccupy)

# Set this to FALSE if you don't have a GPU.
roccupy::set_gpu(TRUE)
```

```r
# 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
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"
```

```r
# 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 28.29048 secs
```

```r
# 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:

$$
\begin{aligned}
y_{ij} &\sim \mathrm{Bern}(\Psi_{ij}), && (1)\\
\mathrm{logit}(\Psi_{ij}) &= x_i^\intercal \beta_j + \gamma_j, && (2)\\
\beta_j &\overset{iid}{\sim} \mathcal{N}(0, I), && (3)\\
\gamma_j &\overset{iid}{\sim} \mathcal{N}(0, 10^2). && (4)
\end{aligned}
$$

The γ_j here are the `env_intercepts`, and the β_j are the `env_slopes` reported by the model.

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

$$
\begin{aligned}
p(s_{ijk} = 1 \mid y_{ij} = 1) &= p_{ijk}, && (5)\\
\mathrm{logit}(p_{ijk}) &= x_{ik}^{(obs)\intercal} \beta_j^{(obs)}, && (6)
\end{aligned}
$$

where

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

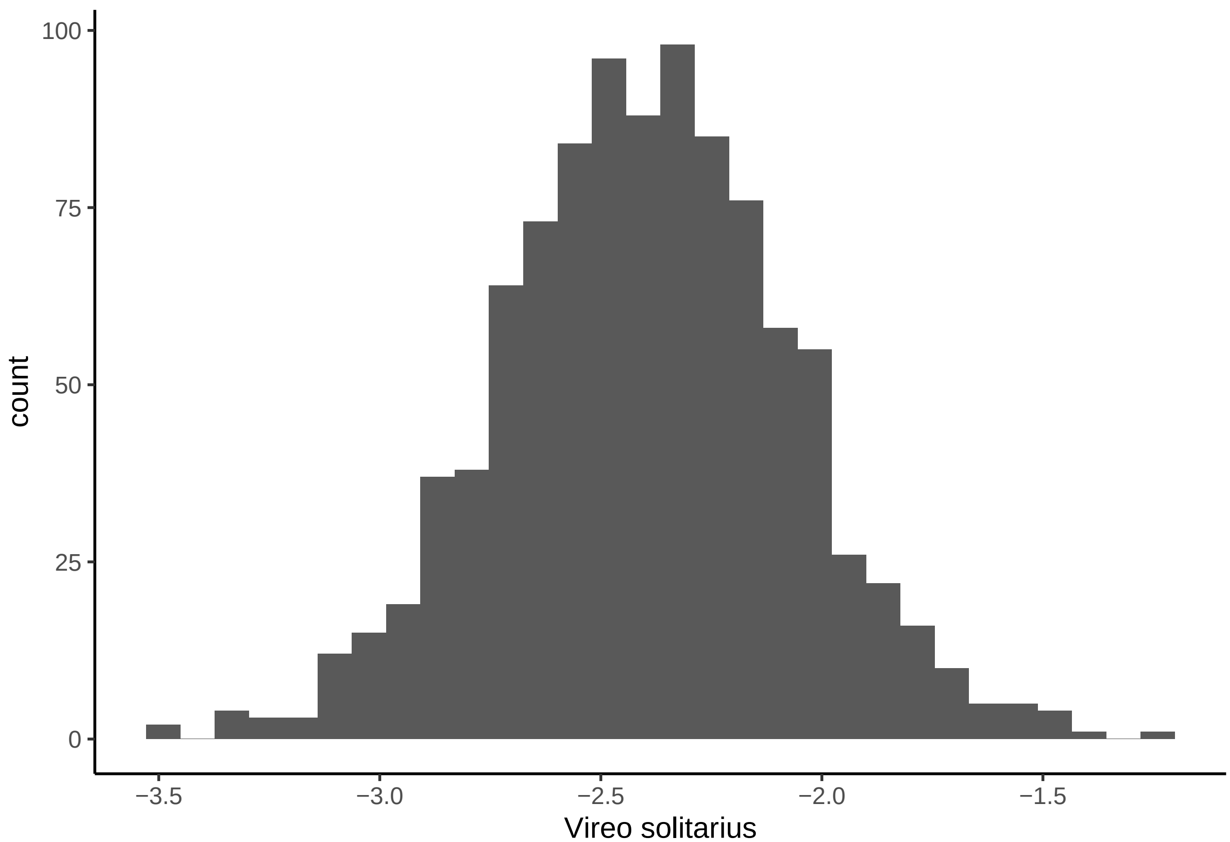
$$\mu_l \overset{iid}{\sim} \mathcal{N}(0, 1), \tag{8}$$

$$\sigma_l \overset{iid}{\sim} \mathcal{H}(1). \tag{9}$$

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 μ_l and group standard deviations σ_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:

```r
head(coef_draws$env_slopes$`Vireo solitarius`)
#>          bio1        bio2      bio3        bio4      bio5       bio6
#> 1 -0.10474803  0.42008582 0.6818631 -0.36228758 -2.192126 -1.1968877
#> 2 -1.29099059 -0.20764181 1.0634518 -0.49058217 -2.035044 -1.1809818
#> 3 -1.26433027 -0.05882168 0.7170187  0.22879228 -1.790993 -0.8061631
#> 4  0.09310865  0.53120285 0.7716353 -0.02978602 -2.277369 -1.1347076
#> 5 -0.19267888 -0.26509365 0.8682649 -0.52831060 -2.268003 -0.9115217
#> 6 -0.80670124 -0.53902298 0.3208472  0.10219079 -2.814070 -0.9074647
#>          bio7      bio8        bio9      bio10        bio12     bio13     bio14
#> 1  0.06574225 1.5017718  0.29798591 -1.0866607  0.412460387 0.3528955 0.9387513
#> 2 -0.12283526 1.0868113  0.45543918 -1.5089333  0.484970123 0.5918346 1.3739645
#> 3 -0.10863703 0.9999026  0.47798413 -2.7743180  0.995000184 0.2576142 0.5908579
#> 4  0.38996464 0.8846436  0.64728212 -1.7269005 -0.009529044 0.4149739 0.7915367
#> 5 -0.11687349 1.6362451 -0.06144226 -0.7607639  0.797886670 0.3232458 1.0651422
#> 6 -0.95173049 1.6685314  0.33361188 -1.2475883  0.504216552 0.6244236 1.2933601
#>         bio15        bio18       bio19 I(bio1 ** 2) I(bio2 ** 2) I(bio3 ** 2)
#> 1 -0.06054338 -0.003590927 -0.04637074   -0.4016776    -1.457542  -0.72659391
#> 2 -0.49815613 -0.636123240 -0.72641438   -0.6131783    -1.378136   0.17925532
#> 3 -0.35999990  0.079270273  0.25625148    0.2853909    -1.947945  -0.16157919
#> 4 -0.33282679 -0.474472314 -0.68000335    0.3032352    -1.773004   0.44299021
#> 5 -0.42156205 -0.885936558 -0.20644034   -0.1320957    -1.325731  -0.01161095
#> 6 -0.21441549 -0.507039011  0.50226253   -0.2458704    -1.209500  -0.04372588
#>   I(bio4 ** 2) I(bio5 ** 2) I(bio6 ** 2) I(bio7 ** 2) I(bio8 ** 2) I(bio9 ** 2)
#> 1   0.71702021 -0.005003318    0.8536834  -0.37858137  -0.17841674   0.25157690
#> 2  -0.27642497 -0.244981483    1.0091714   0.03363771  -0.39478686   0.10110113
#> 3   0.63311112  0.195725396    0.7131722  -0.70950276  -0.70326895   0.10939638
#> 4  -0.08898599  0.285420716    0.8503609  -0.34057683  -0.22729437  -0.08447366
#> 5   0.61237472  0.113696732    1.5607088  -0.10457486  -0.01196928   0.58862108
#> 6   0.43816760  0.413153321    0.9678303  -0.02325367  -0.67977792  -0.00740787
#>   I(bio10 ** 2) I(bio12 ** 2) I(bio13 ** 2) I(bio14 ** 2) I(bio15 ** 2)
#> 1     -2.061442     0.8398412     0.3202951     0.7983984     -1.933055
#> 2     -2.115271     0.6231714    -1.1143248     0.5788265     -1.771081
#> 3     -2.161492     1.0004576    -0.5092661     0.8334402     -2.329893
#> 4     -1.633576     0.5904251    -2.0528719     1.0318475     -1.944972
#> 5     -1.478346     0.8828460    -1.4430696     0.7122160     -1.887795
#> 6     -1.946736     0.6952951    -0.8868259     0.8582040     -1.883673
#>   I(bio18 ** 2) I(bio19 ** 2) has_open_water has_deciduous_forest
#> 1    -0.3238955     -1.535474      0.8804438           0.68448412
#> 2    -0.8915223     -1.061755      1.1022487          -0.05349492
#> 3    -0.8247376     -1.680637     -0.0605080          -0.58590341
#> 4    -1.2809716     -1.146819      0.3446011           0.40465769
#> 5    -0.7828118     -1.013131      1.1071577           0.20304510
#> 6    -0.9747285     -1.452990      1.1386452           0.34689891
#>   has_evergreen_forest has_mixed_forest has_shrub_or_scrub
#> 1            0.8825191      -0.02477264         -0.1332140
#> 2            0.6515251       0.53227353         -1.7283885
#> 3           -0.2244378       0.85470885          0.2664188
#> 4           -0.6494172       0.08694084          1.5792104
#> 5            0.7463863       0.63732129         -0.5689844
#> 6            0.8172285      -0.11099020          0.5292112
#>   has_grassland_or_herbaceous has_pasture_or_hay has_cultivated_crops
#> 1                   0.4875011        -1.40104008            0.6782973
#> 2                   0.1431040        -1.03257751           -0.2369649
```

```
#> 3                   0.6203580        -0.76242584           0.2362628
#> 4                  -0.2827740        -0.01228047           0.1212997
#> 5                   0.8752488        -1.10823905          -0.3967290
#> 6                   3.3053594        -0.28165188           0.7453209
#>     has_other has_developed has_wetlands
#> 1 -0.02465825   -0.03578883   -0.9259712
#> 2  0.62999076    0.50599062   -1.4881793
#> 3  0.62229663    0.38191348   -1.4290788
#> 4  1.32715106    0.15700433   -0.6281899
#> 5  0.33314210   -0.24648699    0.1586592
#> 6  0.77403104    0.17290941    0.3653177
```

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                        -0.828908141 0.5357306
#> bio2                        -0.033682265 0.4256526
#> bio3                         0.540019897 0.5293147
#> bio4                        -0.187357363 0.3698994
#> bio5                        -2.167150248 0.4122531
#> bio6                        -0.789937281 0.2943294
#> bio7                        -0.216585618 0.4146674
#> bio8                         0.802860413 0.3785070
#> bio9                         0.272069526 0.2364901
#> bio10                       -1.243933328 0.5007904
#> bio12                        0.537166818 0.3778584
#> bio13                        0.546039215 0.6316629
#> bio14                        1.012487211 0.3366342
#> bio15                       -0.219426551 0.3958128
#> bio18                       -0.294435543 0.2724851
#> bio19                        0.114903789 0.4220528
#> I(bio1 ** 2)                 0.018609754 0.2872635
#> I(bio2 ** 2)                -1.583936147 0.3227630
#> I(bio3 ** 2)                -0.072564204 0.3740171
#> I(bio4 ** 2)                 0.287381132 0.5369835
#> I(bio5 ** 2)                 0.005614421 0.2673193
#> I(bio6 ** 2)                 0.885692532 0.2159581
#> I(bio7 ** 2)                -0.348548194 0.2674559
#> I(bio8 ** 2)                -0.494121393 0.3378869
#> I(bio9 ** 2)                 0.168933987 0.2206930
#> I(bio10 ** 2)               -1.883159077 0.3557754
#> I(bio12 ** 2)                0.822937596 0.2283107
#> I(bio13 ** 2)               -0.730517562 0.8032802
#> I(bio14 ** 2)                0.724943248 0.2255211
#> I(bio15 ** 2)               -1.892044324 0.2010767
#> I(bio18 ** 2)               -0.935777507 0.3292010
#> I(bio19 ** 2)               -0.709992044 0.6071452
#> has_open_water               0.893168543 0.5216972
#> has_deciduous_forest         0.119150289 0.4698368
#> has_evergreen_forest         0.551346560 0.7370662
```

```
#> has_mixed_forest              0.480738052 0.3851808
#> has_shrub_or_scrub            0.249132910 1.1279060
#> has_grassland_or_herbaceous   0.273243190 0.9921425
#> has_pasture_or_hay           -0.767946601 0.7021721
#> has_cultivated_crops         -0.273997390 0.4840221
#> has_other                    -0.040514832 0.8590746
#> has_developed                -0.083490733 0.3227949
#> has_wetlands                 -0.114424088 0.6126372
```

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]
#>                      -2.49604346                       -0.42660055
#>       protocol_type[T.Traveling]             daytimes_alt[T.dusk]
#>                       0.08728682                       -1.10834650
#>    daytimes_alt[T.early-evening]    daytimes_alt[T.early-morning]
#>                      -0.13805917                        0.09540099
#>     daytimes_alt[T.late-evening]     daytimes_alt[T.late-morning]
#>                      -0.24113312                       -0.09696256
#>          daytimes_alt[T.mid-day]            daytimes_alt[T.night]
#>                      -0.26699359                       -1.79574962
#> dominant_land_cover[T.developed]    dominant_land_cover[T.forest]
#>                      -0.62711306                       -0.34104239
#>     dominant_land_cover[T.water]                    log_duration_z
#>                      -0.46127171                        0.50306904
```

And the group standard deviations are:

```
colMeans(coef_draws$obs_prior_sds)
#>                        Intercept       protocol_type[T.Stationary]
#>                       1.54673058                        0.48112204
#>       protocol_type[T.Traveling]             daytimes_alt[T.dusk]
#>                       0.01802652                        1.85743991
#>    daytimes_alt[T.early-evening]    daytimes_alt[T.early-morning]
#>                       0.12744951                        0.30007928
#>     daytimes_alt[T.late-evening]     daytimes_alt[T.late-morning]
#>                       0.39375893                        0.02827871
#>          daytimes_alt[T.mid-day]            daytimes_alt[T.night]
#>                       0.06023677                        2.04961004
#> dominant_land_cover[T.developed]    dominant_land_cover[T.forest]
#>                       0.93796329                        0.72747625
#>     dominant_land_cover[T.water]                    log_duration_z
#>                       0.86291937                        0.25499845
```

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

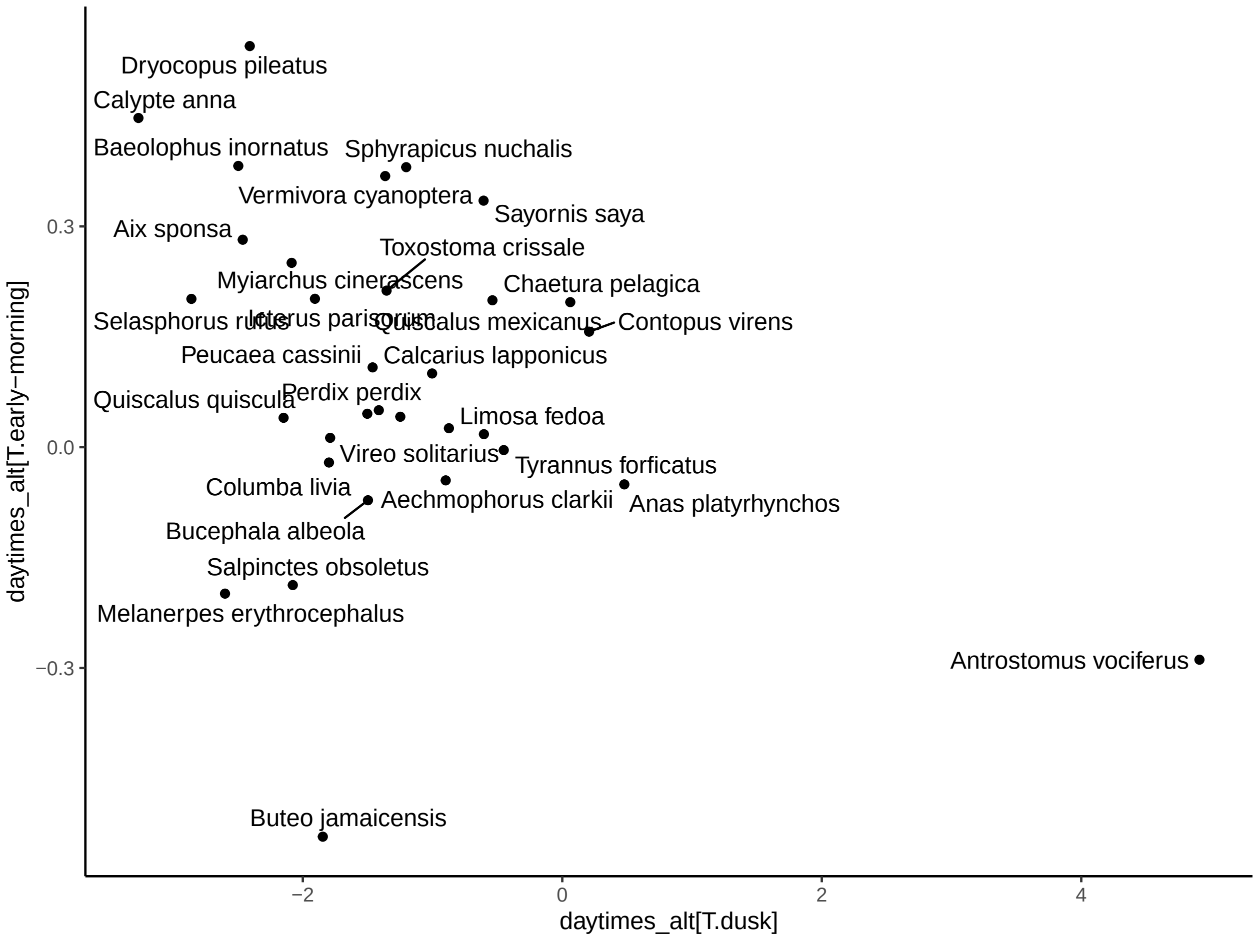
```r
means <- colMeans(coef_draws$obs_slopes$`Vireo solitarius`)
sds <- apply(coef_draws$obs_slopes$`Vireo solitarius`, 2, sd)

cbind(means, sds)
#>                                      means        sds
#> Intercept                      -2.64669575 0.12696226
#> protocol_type[T.Stationary]    -0.78721992 0.16197800
#> protocol_type[T.Traveling]      0.08627985 0.04061941
#> daytimes_alt[T.dusk]           -1.78902266 1.90805735
#> daytimes_alt[T.early-evening]  -0.18613942 0.14145635
#> daytimes_alt[T.early-morning]   0.01268119 0.17787616
#> daytimes_alt[T.late-evening]   -0.62134026 0.30171195
#> daytimes_alt[T.late-morning]   -0.09553588 0.04042068
#> daytimes_alt[T.mid-day]        -0.26819303 0.10420764
#> daytimes_alt[T.night]          -2.78439491 1.73181521
#> dominant_land_cover[T.developed] -0.76618349 0.54625790
#> dominant_land_cover[T.forest]   1.55064834 0.19125336
#> dominant_land_cover[T.water]   -1.07566542 0.47414347
#> log_duration_z                  0.47097110 0.17910196
```

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

```r
# 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   -1.616753                  0.18104408
#> Limosa fedoa        -3.528348                 -0.09095683
#> Anas platyrhynchos  -1.126452                 -0.83264669
#> Tyrannus forficatus -0.521862                 -0.96981863
#> Quiscalus mexicanus -1.005612                 -0.77442361
#> Acanthis flammea    -3.191666                 -0.81432160
#>                      protocol_type[T.Traveling] daytimes_alt[T.dusk]
#> Selasphorus rufus                    0.09169168           -2.8584917
#> Limosa fedoa                         0.08456477           -0.6039925
#> Anas platyrhynchos                   0.08947736            0.4781776
#> Tyrannus forficatus                  0.08379330           -0.4511776
#> Quiscalus mexicanus                  0.09787205           -0.5383457
#> Acanthis flammea                     0.08790266           -1.5027776
#>                      daytimes_alt[T.early-evening] daytimes_alt[T.early-morning]
#> Selasphorus rufus                      -0.15034320                   0.201464046
#> Limosa fedoa                           -0.10133261                   0.017675932
#> Anas platyrhynchos                     -0.02989096                  -0.050511002
#> Tyrannus forficatus                    -0.07266511                  -0.003912035
```

```
#> Quiscalus mexicanus                   -0.08042112                  0.199602230
#> Acanthis flammea                      -0.12753417                  0.045416394
#>                      daytimes_alt[T.late-evening] daytimes_alt[T.late-morning]
#> Selasphorus rufus                     -0.3186966                 -0.10435521
#> Limosa fedoa                          -0.2905974                 -0.09771207
#> Anas platyrhynchos                    -0.1010211                 -0.11804625
#> Tyrannus forficatus                   -0.1655200                 -0.09115050
#> Quiscalus mexicanus                   -0.2320998                 -0.09992374
#> Acanthis flammea                      -0.2350901                 -0.10050917
#>                      daytimes_alt[T.mid-day] daytimes_alt[T.night]
#> Selasphorus rufus                 -0.2602140             -2.090000
#> Limosa fedoa                      -0.2798333             -2.161809
#> Anas platyrhynchos                -0.2667908             -2.197635
#> Tyrannus forficatus               -0.2733609             -2.363010
#> Quiscalus mexicanus               -0.2469070             -3.588744
#> Acanthis flammea                  -0.2704629             -2.776975
#>                      dominant_land_cover[T.developed]
#> Selasphorus rufus                          -1.3614383
#> Limosa fedoa                               -1.6013004
#> Anas platyrhynchos                          0.2236564
#> Tyrannus forficatus                        -1.1592709
#> Quiscalus mexicanus                         0.5764878
#> Acanthis flammea                           -0.7288794
#>                      dominant_land_cover[T.forest] dominant_land_cover[T.water]
#> Selasphorus rufus                       -0.2151525                   -0.2992622
#> Limosa fedoa                            -0.8708493                    1.0215015
#> Anas platyrhynchos                      -0.7848765                    0.5053283
#> Tyrannus forficatus                     -0.8344822                   -0.8337116
#> Quiscalus mexicanus                     -0.6854435                    0.2933743
#> Acanthis flammea                        -0.2021231                   -0.5945284
#>                      log_duration_z        species_name
#> Selasphorus rufus         0.7346655   Selasphorus rufus
#> Limosa fedoa              0.3078453        Limosa fedoa
#> Anas platyrhynchos        0.5108821  Anas platyrhynchos
#> Tyrannus forficatus      -0.1593873 Tyrannus forficatus
#> Quiscalus mexicanus       0.3329129 Quiscalus mexicanus
#> Acanthis flammea          0.5425788    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
```




```r
# 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')
```

```r
# 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:

```r
# Save the model
save_model(fit_model, 'save_test')
```

```r
# Restore it:
restored_model <- restore_model('./save_test/')
```