Rufous Hummingbird habitat covariates

Model structure

The model is an elaboration of the iCAR route-level trend model, where the route-level intercepts and slopes are additive combinations of a component that is a function of a route-level predictor and a residual component, estimated either with the iCAR structure or a non-spatial random effect. The route-level predictors are derived from a habitat modeling study for Rufous Hummingbirds (details below). The mean habitat suitability within a buffer of the BBS route-path is used as a predictor on the intercept (i.e., the mean relative abundance on a given route). The rate of change in habitat suitability over time within the same buffer was used as a predictor on the slope (i.e., the trend in the species' abundance). This model structure relies on relatively simple assumptions that the amount of habitat around a BBS route should predict the number of individual birds, and that the change in the amount of habitat should predict the change in the number of birds.

The basic model is the same as all of other models in the main paper.

$$C_{r,j,t} = Negative\ Binomial\left(\lambda_{r,j,t}, \phi\right)$$

$$log\left(\lambda_{r,j,t}\right) = \alpha_r + \beta_r * (t - t_m) + \eta \mathrm{I}j, t + \omega j$$

We modeled the observed counts $(C_{r,j,t})$ of Rufous Hummingbirds on route-r, in year-t, by observer-j as as realizations of a negative binomial distribution, with mean $\lambda_{r,j,t}$ and inverse dispersion parameter ϕ. The log of the mean $(\lambda_{r,j,t})$ of the negative binomial distribution was modeled as an additive combination of route-level intercepts (α_r), observer-effects (ω_j), and a first-year observer-effect $(\eta I[j,t])$, and route-level slope parameters (β_r) for the continuous effect of year (t) centered on the mid-year of the time-series (t_m).

We estimated the route-level intercepts and slopes as an additive combination of a mean species-level intercept or slope $(\alpha'$ or $\beta')$, a varying intercept or slope that was a function of the mean habitat suitability on the route (α_r''') or rate of change in habitat suitability on the slope (β_r'''), and spatially varying effects for the residual variation in relative abundance (α_r'') and slope (β_r'') that were not explained by habitat.

$$\alpha_r = \alpha' + \alpha_r'' + \alpha_r'''$$
$$\beta_r = \beta' + \beta_r'' + \beta_r'''$$

This partitioning of the intercept and slope parameter allows the model to generate two alternative estimates of the mean abundance and trend on each route. The full trend $\beta' + \beta_r'' + \beta_r'''$ represents the full estimated trend on a given route, including the effects of habitat-change. The residual trend $\beta' + \beta_r''$ represents a counter-factual trend that would have been expected if the habitat had stayed constant on a given route. Similarly, the full relative abundance $\alpha' + \alpha_r'' + \alpha_r'''$ represents the full estimated relative abundance on a given route, including the effects of habitat. The residual relative abundance $\alpha' + \alpha_r''$ represents a counter-factual abundance that would have been expected if the habitat suitability was the same across all routes.

We estimated the effect of mean habitat suitability on the route-level intercept as a simple product of a route-specific coefficient (ρ_{α_r}) and the mean (over all years) of the annual habitat suitabilities in a buffer surrounding each route-path ($\alpha_r'' = \rho_{\alpha_r} * \mu_{habitatsuitability_r}$).The annual habitat suitability values are scaled from $0-1$, so that as scaled, the estimate of ρ_{α_r} represents the maximum possible change in suitability. However, the realized range in values was from 0.2 to 0.7, and so a more relevant interpretation is that it represents twice the maximum change in abundance due to habitat. To model the effects of habitat-change on population trend, we estimated the effect of the rate of change in habitat suitability on each route as a product of a route-specific coefficient (ρ_{β_r}) and an estimate of the average rate of change in habitat suitability on each route ($\delta_{habitatsuitability_r}$). We estimated the rate of change in habitat suitability as the slope of a simple linear regression through the annual estimates of habitat suitability measured within a buffer surrounding each route-path ($\beta_r'' = \rho_{\beta_r} * \delta_{habitatsuitability}$). We multiplied the slopes of the suitability over time by 100, so that they had a standard deviation of approximately 0.5, and so the estimate of ρ_{β_r} represents the change in the log-scale slope parameter associated with the difference between a route on which habitat has been stable and a route where the habitat has increased a lot (i.e., 2 standard deviations from the mean). The habitat suitability predictors were centered to improve convergence. The route-specific coefficients for the effects of habitat suitablility on the intercept and slope were allowed to vary among routes, but were centered on a hyperparameter mean effects across routes $\rho_{\alpha_r} \sim Normal\left(P_\alpha, \sigma_{\rho_\alpha}\right)$ and $\rho_{\beta_r} \sim Normal\left(P_\beta, \sigma_{\rho_\beta}\right)$. As such, the hyperparameters for the effect of mean habitat suitability on the intercept (P_α) and the effect of change in habitat suitablility on slope (P_β), represent a clear species-level estimate of the overall effects of habitat on abundance and trend, after adjusting for the species mean abundance and trend, as well as the residual spatially dependent variation in abundance and trend.

In the fully spatial implementation of the model, we estimated the residual component of the intercepts and slopes using an intrinsic iCAR structure, where the parameter for route-r is drawn from a normal distribution, centered on the mean of that parameter's values in all neighbouring routes, with an estimated standard deviation that is proportional to the inverse

2

of the number of neighbours for that route (Morris et al. 2019). Specifically, the component of the intercept that represents the residual spatially dependent relative abundance (α_r''') was drawn from a normal distribution centered on the mean of the intercepts for all neighbouring routes.

$$\alpha_r''' \sim Normal\left(\frac{\sum_{n \in N_r} \alpha_n'''}{N_r}, \frac{\sigma_{\alpha'''}}{N_r}\right)$$

The spatially varying component of the slope (β_r''') was estimated similarly as random route-level terms from a normal distribution centered on the mean of the slopes for all neighbouring routes using the same iCAR structure.

$$\beta_r''' \sim Normal\left(\frac{\sum_{n \in N_r} \beta_n'''}{N_r}, \frac{\sigma_{\beta'''}}{N_r}\right)$$

Alternative non-spatial residual term on intercepts

In the fully spatial version of the model, there was a relatively strong spatial autocorrelation in both the habitat suitability and the mean abundance of the species. As a result, the spatial iCAR component of the intercept abosrbed much of the variation in abundance among routes, leaving relatively little variation explained by habitat.

Since the spatial component of habitat suitability could reasonably be considered a cause of the spatial dependency in abundance, we drew our final inference on the effect of habitat suitability on abundance from a model that estimated the residual component of the intercept term with a non-spatial varying effect (i.e., a simple random effect). Specifically, the component of the intercept that represents the residual relative abundance (α_r''') was drawn from a normal distribution centered at zero with an estimated standard deviation ($\alpha_r''' \sim Normal(0, \sigma_{\alpha'''})$).

```
library(bbsBayes2)
library(tidyverse)
library(sf)
library(cmdstanr)
library(patchwork)

output_dir <- "output"
species <- "Rufous Hummingbird"


species_f <- gsub(gsub(species,pattern = " ",replacement = "_",fixed = T),pattern = "'",re

spp <- "_habitat_"
```

```r
exp_t <- function(x){
  y <- (exp(x)-1)*100
}

firstYear <- 2006
lastYear <- 2021


out_base <- paste0(species_f,spp,firstYear,"_",lastYear)



sp_data_file <- paste0("Data/",species_f,"_",firstYear,"_",lastYear,"_covariate_stan_data.


load(sp_data_file)

  mod.file = paste0("models/slope",spp,"route_NB.stan")

# trend habitat effects are not changed, but the intercept effect is
# removes the optional spatial components for intercepts
stan_data[["fit_spatial"]] <- 0 # this sets an option in the model
# to estimate the residual intercept component using a simple random
# effect, instead of a spatial one. This allows the model to estimate
# variation in abundance that is not predicted by local habitat suitability
# but does not fit an inherently spatial residual structure
# setting this fit_spatial value to 1 uses the iCAR structure to model
# a spatially explicit residual term
```

The `stan_data[["fit_spatial"]] <- 0` line sets a `false` conditional statement in the data list that allows the model to estimate the residual intercept component using a simple random effect, instead of a spatial one. This allows the model to estimate variation in abundance that is not predicted by local habitat suitability but does not fit an inherently spatial residual structure setting this `stan_data[["fit_spatial"]] <- 1` results in a true conditional statement and uses the iCAR structure to model a spatially explicit residual term on the intercept.

```r
slope_model <- cmdstan_model(mod.file, stanc_options = list("Oexperimental"))
```

```
stanfit <- slope_model$sample(
  data=stan_data,
  refresh=400,
  iter_sampling=2000,
  iter_warmup=2000,
  parallel_chains = 4)

summ <- stanfit$summary()
print(paste(species, stanfit$time()[["total"]]))

saveRDS(stanfit,
        paste0(output_dir,"/",out_base,"_stanfit.rds"))

saveRDS(summ,
        paste0(output_dir,"/",out_base,"_summ_fit.rds"))

summ %>% arrange(-rhat)
```

Fitting the model

Before fitting the model, we prepared the BBS counts, the neighbourhood structures necessary
to estimate the iCAR residual spatial component, and joined them to the habitat suitablity
predictors. The full code and data necessary to replicate the data-preparation is available in
the online supplement. In brief, we selected all routes on which the species had been observed
during and for which we had GIS route-path information that would allow us to estimate the
route-specific annual habitat suitability values.

We fit the model using the probablistic programming language Stan (Stan Development Team
2022), accessed through the R-package **cmdstanr** (Gabry and Češnovar 2022). We used a
warm-up of 2000 iterations, and **cmdstanr** default settings for other arguments, followed by a
draw of 2000 samples from which we estimated the posterior distributions. All parameters in
all models converged based on Rhat < 1.02 and bulk effective sample sizes > 500.

Results

During the 15-years from 2006-2021, the species overall population declined steeply. The model estimated an overall change in the population of approximately -43% [-52–33]. Trends were negative across the species' range, but most negative in the coastal regions where the species is also most abundant (Figure 1). The effect of habitat suitability on mean relative abundance was strong and positive ($P_\alpha = 3$ [2.2-3.8]), and this effect was robust, whether the residual abundance component was spatially autocorrellated or random. There was a clear positive effect of change in the habitat suitability on trends, such that routes with habitat-loss had more negative population trends $P_\beta = 0.025$ [0.0026-0.047]. The greater loss of habitat in the coastal region accounts for most of the increased rates of decline in that region (Figure 2), the residual trend component alone (Figure 2, right panel) does not show the same coastal-decline pattern.
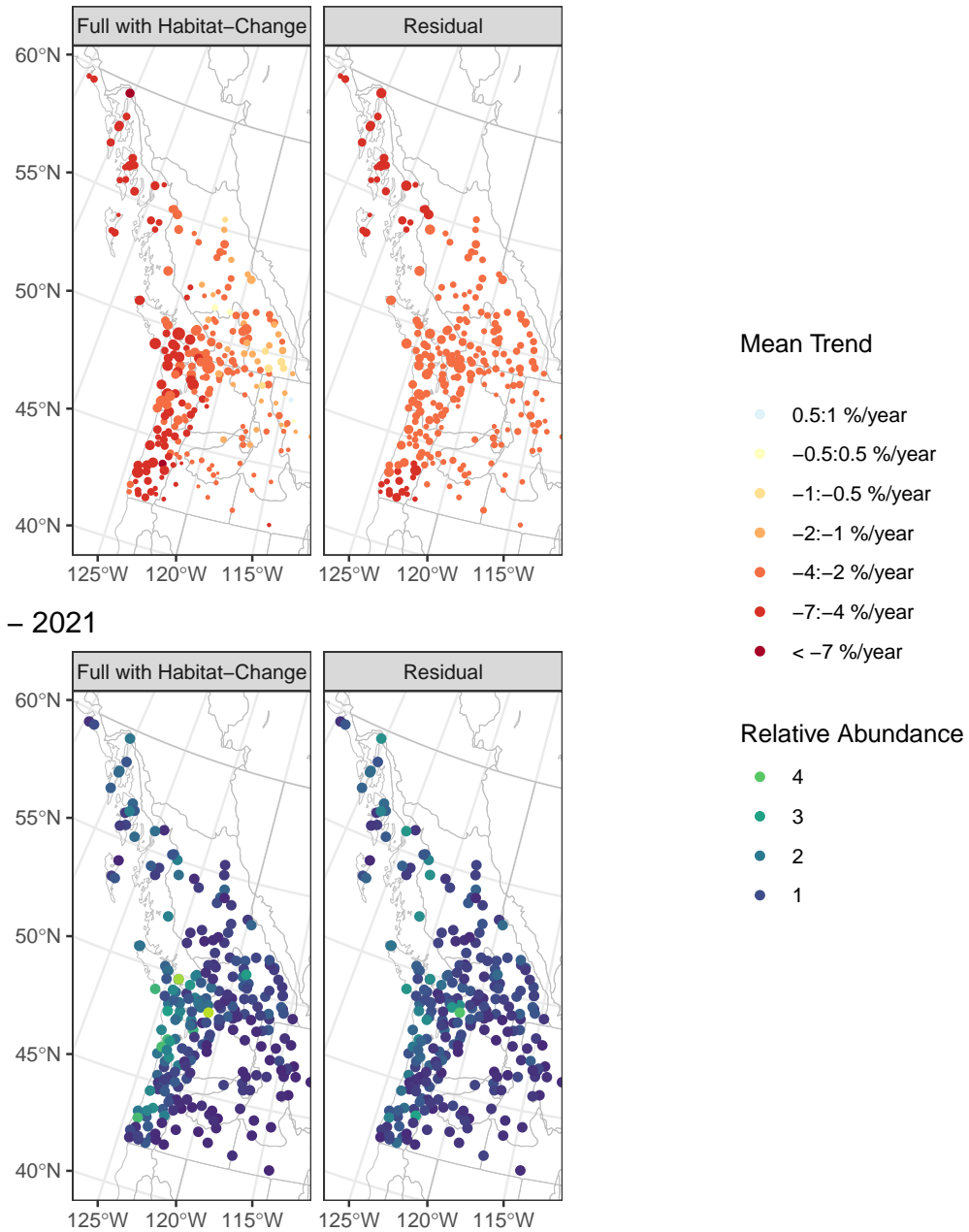
Figure 1: Map of the trends for Rufous Hummingbird from 2006-2021 The colours represent the trends in the uppper panel and the relative abundance in the lower panel. The left panel represents the full estimated trends and abundance on each route, including both the effect of habitat-suitability and the residual component not related to habitat. The right panel represents the trends and relative abundances after removing the effect of habitat-suitability. In the top-left panel, the greater declines in coastal regions are evident from the darker red points compared to the top-right panel. In the bottom-left panel, the higher abundance near the coast is evident from the lighter colours. The bottom-right panel shows much more even relative abundance across the species' range, showing that habitat suitability accounts for much of the variation in abundance

References

Gabry, Jonah, and Rok Cešnovar. 2022. *Cmdstanr: R Interface to 'CmdStan'*. https://mc-stan.org/cmdstanr.

Morris, Mitzi, Katherine Wheeler-Martin, Dan Simpson, Stephen J. Mooney, Andrew Gelman, and Charles DiMaggio. 2019. "Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan." *Spatial and Spatio-Temporal Epidemiology* 31 (November): 100301. https://doi.org/10.1016/j.sste.2019.100301.

Stan Development Team, 2.29. 2022. *Stan Modeling Language Users Guide and Reference Manual*. https://mc-stan.org.