Appendix S4 Model assessment and evaluation

Prior simulations for the spatial variance in GAM parameters

The stratum-level population trajectories are fit using a spatially explicit, hierarchical GAM smooth. This spatial hierarchical structure allows the model to share information on the shape of the population trajectory among neighbouring strata and across the full range of the sites monitored for a given species. The variation among strata in the shape of the population trajectory is estimated using an intrinsic conditional autoregressive (iCAR) structure. This iCAR component of the model estimates the parameters of the GAM smooth as drawn from a normal distribution centered on the means of the parameter values in the neighbouring strata and with an estimated standard deviation (SD). The value of the SD controls the amount of variation in the shape of the trajectories among strata.

An appropriate prior for the SD parameter is not intuitive, because the GAM parameters do not directly reflect a biological process for which prior knowledge might provide an informative prior. Priors on variance parameters can be unintentionally informative if they put substantial prior mass at improbable levels of variation (Banner, Irvine, and Rodhouse 2020). Therefore, we conducted a prior simulation to translate the SD priors into a variation in long-term population trends, for which we do have some biological intuition. The simulation was designed to ensure that the prior on the spatial SD was largely uninformative, allowing for a reasonably broad range of long-term (1980-2019) population trend estimates across a species migration range, and that it did not contain substantial prior mass at highly improbable levels of variation (Lemoine 2019).

Simulation

We simulated population trajectories from the model, using estimated parameter values for Semipalmated Plover for all parameters except the stratum-level GAM parameters. We chose to use Semipalmated Plover as the example species in the simulation so that the simulated data would closely match real data and because the Semipalmated Plover mean population trajectory is largely stable through time. Simulating a stable population allowed us to more intuitively assess the variation in population trends in the simulated data.

```
sp1 = "Semipalmated Plover"


    spf1 = "Semipalmated_Plover"
    sp_file_name1 <- "Semipalmated_Plover-gamma-t"

    #loading the original data
    load(paste0("data_local/data",sp1,"_cmdstanr_data.RData"))
    #loading the objects saved after model run
    load(paste0(output_dir,"/",sp_file_name1,"_fit_add.RData"))

# commented out section run in separate session to avoid loading the large cmdstan output
#     ## load the stan csv files that include the model output for SEPL
#     cmdstanfit <- as_cmdstan_fit(csvfl)
#
#
# # Gather parameter estimates from fitted model --------------------------------------
#
#     #posterior samples of site-level effects
#     alpha_samples <- posterior_samples(fit = cmdstanfit,
#                                        parm = "alpha",
```

```
#                                             dims = c("s"))
#
#     #posterior samples of the overall intercept
#     ALPHA1_samples <- posterior_samples(fit = cmdstanfit,
#                                 parm = "ALPHA1",
#                                 dims = NULL)
#
#
#     #posterior samples of overdispersion variance
#     sdnoise <- posterior_samples(fit = cmdstanfit,
#                                   parm = "sdnoise",
#                                   dims = NULL) %>%
#       summarise(mean = mean(.value)) %>%
#       as.numeric()
#
#     #posterior samples of variance in year-effects
#     sdyear <- posterior_samples(fit = cmdstanfit,
#                                   parm = "sdyear",
#                                   dims = NULL) %>%
#       summarise(mean = mean(.value)) %>%
#       as.numeric()
#
#
#
#
#     #posterior samples of hyperparameter GAM parameters (survey-wide mean population trajectory)
#     B_samples <- posterior_samples(fit = cmdstanfit,
#                                   parm = "B",
#                                   dims = c("k"))
#
#
#     save(list = c("stan_data",
#                   "B_samples",
#                   "sdyear",
#                   "sdnoise",
#                   "alpha_samples",
#                   "ALPHA1_samples"),
#          file = "data_local/temp_SEPL_output.RData")


    load("data_local/temp_SEPL_output.RData")
```

Generate the simulated stratum-level trajectories using the prior distributions

We used the realised adjacency matrix for SEPL monitoring data to generate the distribution of stratum-level
GAM parameters.

```
#neighbourhood matrix from the original species data
    nmat = matrix(0,nrow = stan_data$nstrata,ncol = stan_data$nstrata)
    for(j in 1:stan_data$N_edges){
      ss = stan_data$node1[j]
      ss2 = stan_data$node2[j]
```

```
    nmat[ss,ss2] <- 1
    nmat[ss2,ss] <- 1


}


# blank object to hold simulated trends
Trends_comp_out <- NULL
#number of simulations = number of posterior draws from the fitted model
niter = max(B_samples$.draw)
```

We used gamma priors as weakly informative, boundary-avoiding priors on the SD. All gamma priors had a shape parameter = 2, so that the mode of the prior distribution = 1/rate, following (Chung et al., n.d.). Using a gamma distribution excludes values of 0 from the prior on the SD. We assessed four alternative rate parameters c(0.5,1,2,4). This selection of parameter values allow the priors to span a wide range of distributions, where the prior modes range from 0.25 (4) to 2 (0.5) and the upper 95 percentile of prior distributions range from 1.2 (4) to 10 (0.5).

```
rates = c(0.5,2,1,4)
```

Then we simulated the stratum-level trajectories using the real data from each posterior draw of the SEPL model output and the proposed prior distribution of SD values for the GAM parameters in each stratum.

```
for(prior_rate in rates){
   #samples from the prior
 sdyear_gam_sim <- rgamma(niter,2,prior_rate)

 #blank array to hold the simulated population trajectories from which the trends are generated
 nsmooth <- array(NA,dim = c(niter,stan_data$nstrata,stan_data$nyears))

 for(i in 1:niter){

   #blank matrix to hold the yearly predictions of the smooths
   year_pred <- matrix(NA,nrow = stan_data$nyears,ncol = stan_data$nstrata)
   # blank matrix to hold the simulated stratum-level GAM parameters
   b <- matrix(NA,nrow = stan_data$nstrata,ncol = stan_data$nknots_year)
   #scaling the neighbourhood matrix into a covariance matrix
   sig_mat = (nmat*(sdyear_gam_sim[i]^2))
   diag(sig_mat) <- 1

   # minimal modifications to ensure the
   sig_mat = Matrix::nearPD(sig_mat,keepDiag = FALSE)

   B <- B_samples %>% filter(.draw == i)
   B <- B$.value

   ALPHA1 <-  ALPHA1_samples %>% filter(.draw == i)
   ALPHA1 <- ALPHA1$.value

   alpha <-  alpha_samples %>% filter(.draw == i)
   alpha <- alpha$.value
```

```
    #generate the simulated GAM coefficients using the mvrnorm function from MASS package.
    for(k in 1:stan_data$nknots_year){
    b[1:stan_data$nstrata,k] <- MASS::mvrnorm(mu = rep(0,stan_data$nstrata),Sigma = sig_mat$mat) + B[

    }
#generate the smooths using simulated coefficients
for(s in 1:stan_data$nstrata){
  year_pred[,s] = (stan_data$year_basispred %*% (b[s,]))
}

for(s in 1:stan_data$nstrata){


  for(y in 1:stan_data$nyears){
    #temporary vector to hold the site-level predictions for each year, stratum, and site
    # addin gin the site-level variation is not actually necessary,
    # we've retained it here just to make a clear link to the quantities estimated in the model
    atmp_smo <- vector(mode = "numeric",length = stan_data$nsites_strat[s])


    for(j in 1:stan_data$nsites_strat[s]){
      atmp_smo[j] = exp(ALPHA1 + year_pred[y,s] + 0.5*(sdyear^2) + 0.5*(sdnoise^2) + alpha[stan_data
    }
    nsmooth[i,s,y] = mean(atmp_smo)
  }

}




}#i end of iterations

#temporary matrix to contain the trend calcualtions
Trends <- matrix(NA,nrow = niter,ncol = stan_data$nstrata)
# additional matrix to contain trends scaled as the ratio of end/start values of the trajectories
# will be used to calculate %/change of the population - a sometimes more intuitive metric of
# long-term trend
p_change <- Trends
for(i in 1:niter){
    for(s in 1:stan_data$nstrata){
      Trends[i,s] <- (((nsmooth[i,s,stan_data$nyears]/nsmooth[i,s,1])^(1/stan_data$nyears))-1)*100
      p_change[i,s] <- ((nsmooth[i,s,stan_data$nyears]/nsmooth[i,s,1]))
    }
}

  # function to calculate the span i.e., the difference between two percentiles of a vector
span = function(x,q = 1){
  quantile(x,q)-quantile(x,1-q)
}


# creating a dataframe to store the summaries of the simulations for this prior
```

```
    Trends_df <- as.data.frame(Trends)
    names(Trends_df) <- gsub(names(Trends_df),pattern = "V",replacement = "Strat_")
    Trends_df$prior_scale = paste0("gamma(2,",prior_rate,")")
    Trends_df$span95 = apply(Trends,1,FUN = span,q = 0.95)
    Trends_df$span = apply(Trends,1,FUN = span,q = 1)
    Trends_df$p_ch_1 = apply(p_change,1,FUN = quantile,p = 0.1)
    Trends_df$p_ch_9 = apply(p_change,1,FUN = quantile,p = 0.9)

    # stacking the summaries across prior values
    Trends_comp_out <- bind_rows(Trends_comp_out,Trends_df)

    }#prior_rate
```

The different priors capture a large range of variation in the stratum level trend estimates.

The gamma(2,0.5) and gamma(2,1) include some prior mass at ranges of variation in the trends that are not reasonable to expect in a natural population. They each include substantial prior mass where 90% of the 39-year population trends, vary by more than 30% for an overall stable population. E.g., they may include regional populations that have extreme declines (more than 15%/year) and regional populations that have similarly extreme increases. The gamma(2,2) prior suggest fewer of these extreme levels of variation, although it still includes a long-tail allowing for unlikely levels of variation.

```
br = c(0,5,10,20,30,50,100,150)
lb = paste0(br,"%") #x-axis labels
ggh_t = ggplot(data = Trends_comp_out,aes(x = span95))+
  geom_histogram(bins = 50)+
  scale_x_continuous(breaks = br,labels = lb,limits = c(0,NA))+
  xlab("Difference between 95th and 5th percentiles of stratum-level long-term trend estimates (%/year
  theme(text = element_text(family = "serif"))+
  facet_wrap(~prior_scale,scales = "free")

print(ggh_t)
```

```
## Warning: Removed 4 rows containing missing values (`geom_bar()`).
```

Difference between 95th and 5th percentiles of stratum–level long–term trend estimates (%/year)
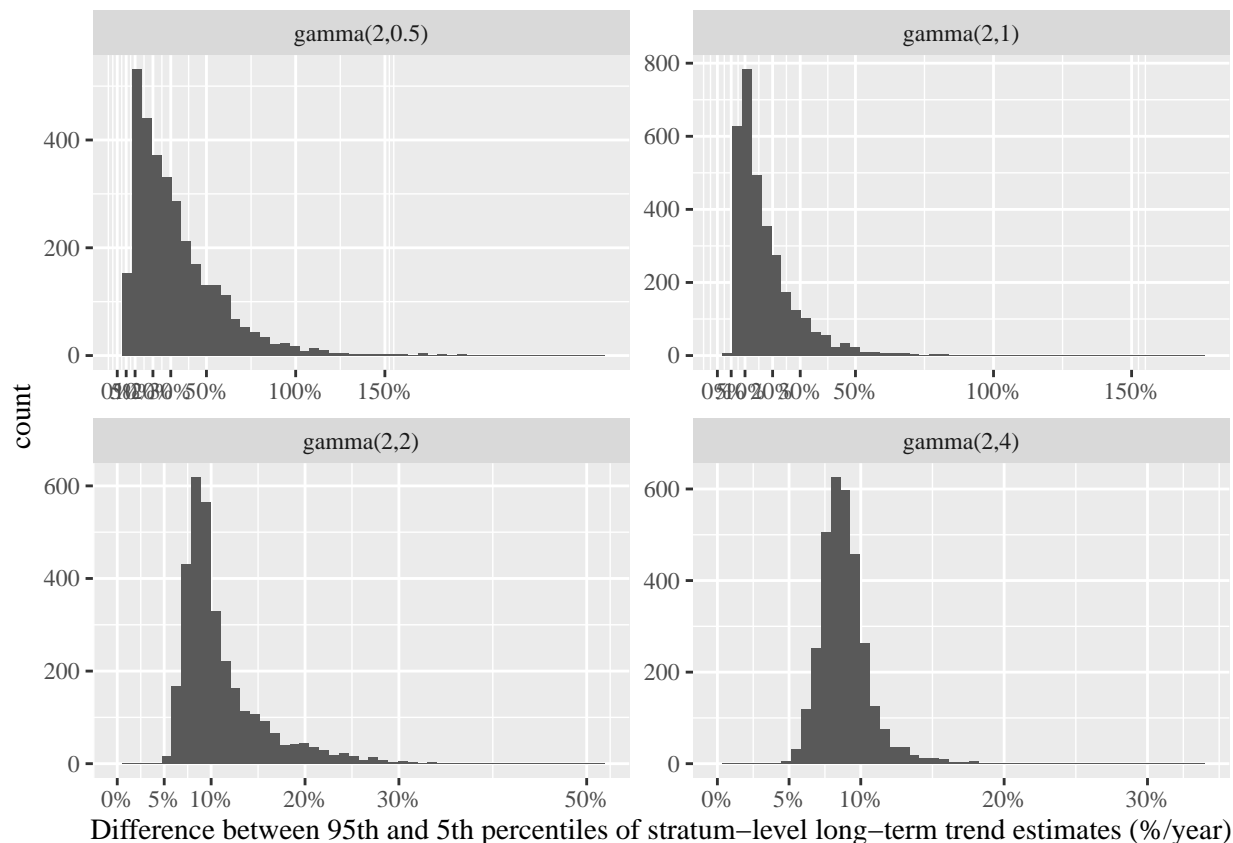
The gamma(2,4) prior allows for high levels of variation in the range of long-term trends, but the 10th and 90th percentiles of the stratum-level estimates of %-change in the population are largely limited to feasible values of local population change. Whereas the other priors include substantial prior mass at extremely unlikely levels of local population change (i.e., populations that have increased and decreased by more than 4 orders of magnitude)

```r
ch_stack <- Trends_comp_out %>% select(prior_scale,p_ch_1,p_ch_9) %>%
  pivot_longer(.,cols = starts_with("p_ch"),
                                    names_to = "quantile_population_change",
                                    values_to = "Percent_change") %>%
  mutate(quantile_population_change = ifelse(quantile_population_change == "p_ch_1","10th percentile",


br = c(c(1/1000000000,1/1000000,1/10000,1/100),c(1,100,10000,1000000,1000000000))
lb = paste0(c((br-1)*100),"%")
ggh_p = ggplot(data = ch_stack,aes(x = Percent_change,fill = quantile_population_change))+
  geom_histogram()+
  scale_x_log10(breaks = br,labels = lb)+
  scale_fill_viridis_d(end = 0.8)+
  xlab("10th and 90th percentile of stratum-level population change")+
  facet_wrap(~prior_scale,scales = "free")+
  theme(text = element_text(family = "serif"),
        legend.position = "bottom")

print(ggh_p)
```

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


10th and 90th percentile of stratum−level population change

quantile_population_change ■ 10th percentile ■ 90th percentile
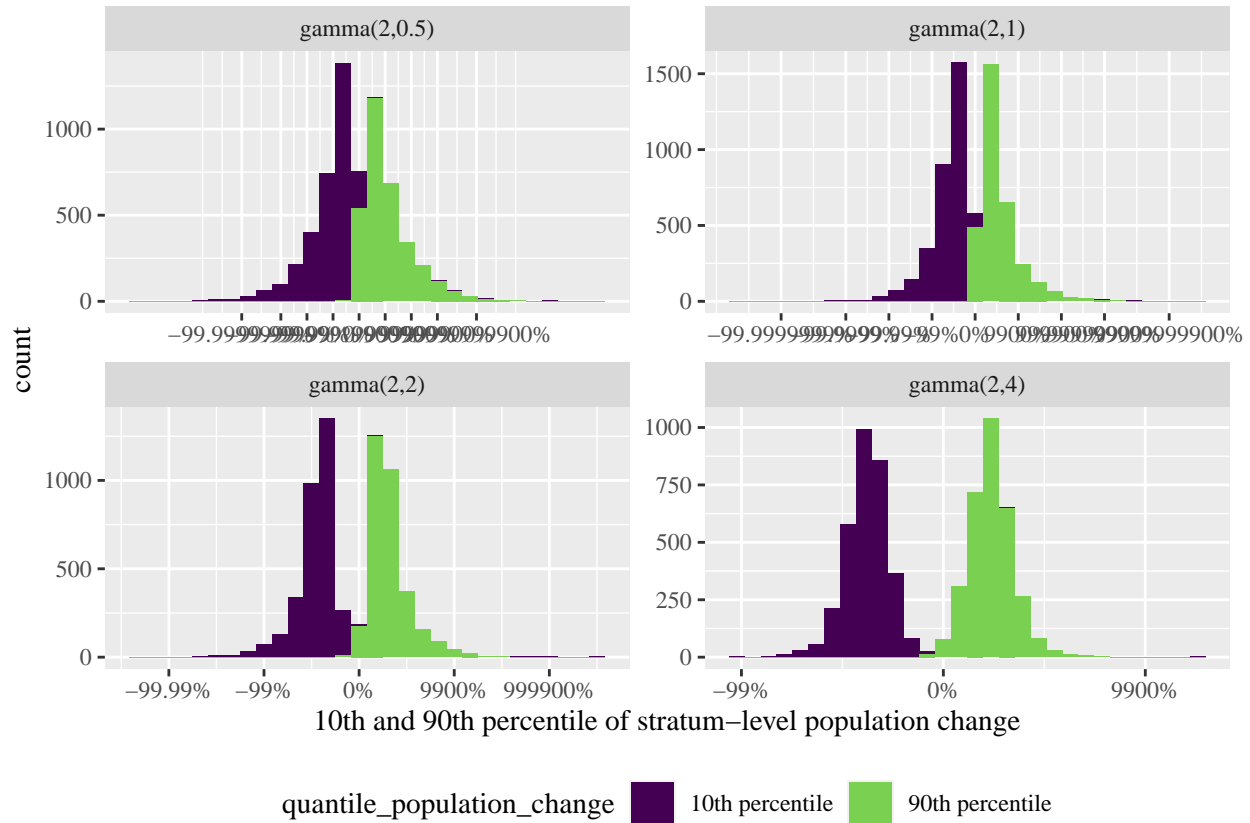
Banner, Katharine M., Kathryn M. Irvine, and Thomas J. Rodhouse. 2020. "The Use of Bayesian Priors in Ecology: The Good, the Bad and the Not Great." *Methods in Ecology and Evolution* 11 (8): 882–89. https://doi.org/10.1111/2041-210X.13407.

Chung, Yeojin, Sophia Rabe-Hesketh, Vincent Dorie, Andrew Gelman, and Jingchen Liu. n.d. "A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models." *PSYCHOMETRIKA* 78 (4): 685709. https://doi.org/10.1007/S11336-013-9328-2.

Lemoine, Nathan P. 2019. "Moving Beyond Noninformative Priors: Why and How to Choose Weakly Informative Priors in Bayesian Analyses." *Oikos* 128 (7): 912–28. https://doi.org/10.1111/oik.05985.