Prior Simulations for status and trend models

Adam C. Smith

30/03/2023

```
## Warning: package 'ggplot2' was built under R version 4.2.2

## Warning: package 'dplyr' was built under R version 4.2.2

## Warning: package 'stringr' was built under R version 4.2.2

## Warning: package 'cmdstanr' was built under R version 4.2.1

## Warning: package 'patchwork' was built under R version 4.2.1

## Warning: package 'kableExtra' was built under R version 4.2.3
```

Prior Simulation of parameters that control the population trajectories in spatial status and trend models

Priors for the annual differences in a first-difference model

Priors for the SD of these parameters among strata

Priors for the spline parameters in a thin-plate regression spline basis

Priors for the SD of these parameters among strata

```
for(i in 1:nrow(tb_sims)){
  mod <- tb_sims[i,"model"]
  prior_B <- tb_sims[i,"prior_time"]
  prior_b <- tb_sims[i,"prior_sd_time"]
  prior_y <- tb_sims[i,"prior_yeareffects"]

  if(mod == "gamye"){
    stan_data <- data_gamye
    stan_data[["prior_scale_y"]] <- prior_y

  }else{
    stan_data <- data_diff

  }
```

```
stan_data[["prior_scale_b"]] <- prior_b

if(tb_sims[i,"hierarchical"]){
  hier <- ""
  stan_data[["prior_scale_B"]] <- prior_B

  if(tb_sims[i,"spatial"]){
    spat <- "spatial"

  }else{
  spat <- "non_spatial"
    stan_data[["N_edges"]] <- NULL
    stan_data[["node1"]] <- NULL
    stan_data[["node2"]] <- NULL

  }

}else{
  hier <- "non_hierarchical"
  spat <- ""
   stan_data[["N_edges"]] <- NULL
    stan_data[["node1"]] <- NULL
    stan_data[["node2"]] <- NULL

}


out_base <- paste("prior_sim",mod,spat,hier,sep = "_")


mod_file <- paste0("models/",tb_sims[i,"model_file"])

    # Fit model ----------------------------------------------------------------

    print(paste("beginning",out_base,Sys.time()))


    ## compile model
    model <- cmdstan_model(mod_file)


    # Initial Values -----------------------------------------------------------


    # init_def <- function(){ list(sdbeta = runif(nyears_m1,0.01,0.1),
    #                              beta_raw = matrix(rnorm(nyears_m1*nstrata,0,0.01),nrow = nstrata,n
    #                              sdBETA = runif(1,0.01,0.1),
    #                              BETA_raw = rnorm(nyears_m1,0,0.01))}

    stanfit <- model$sample(
      data=stan_data,
      refresh=100,
      chains=2, iter_sampling=1000,
```

```
        iter_warmup=500,
        parallel_chains = 2,
        #pars = parms,
        adapt_delta = 0.8,
        max_treedepth = 10,
        seed = 123)


      stanfit$save_object(file = paste0("output/",out_base,".rds"))







}
```

We then summarized the estimated trajectories as well as the 1, 5, 10, 20, and 50-year trends simulated from the alternative priors.

```
source("Functions/posterior_summary_functions.R")

n_out <- NULL
trends_out <- NULL
summ_out <- NULL




for(i in 1:nrow(tb_sims)){

  trends_out_tmp <- NULL

  mod <- tb_sims[i,"model"]
  prior_B <- tb_sims[i,"prior_time"]
  prior_b <- tb_sims[i,"prior_sd_time"]
  prior_y <- tb_sims[i,"prior_yeareffects"]


  if(tb_sims[i,"hierarchical"]){
    hier <- ""

    if(tb_sims[i,"spatial"]){
      spat <- "spatial"

    }else{
    spat <- "non_spatial"

    }

  }else{
    hier <- "non_hierarchical"
    spat <- ""
```

```
  }


  out_base <- paste("prior_sim",mod,spat,hier,sep = "_")

  stanfit <- readRDS(paste0("output/",out_base,".rds"))


summ = stanfit$summary()

summ <- summ %>%
  mutate(model = mod,
         spatial = spat,
         hierarchical = hier)

n_samples <- posterior_samples(stanfit,
                               parm = "n",
                               dims = c("strat","Year_Index"))


if(mod == "gamye"){
nsmooth_samples <- posterior_samples(stanfit,
                               parm = "nsmooth",
                               dims = c("strat","Year_Index"))


# N_samples <- posterior_samples(stanfit,
#                                  parm = "NSMOOTH",
#                                  dims = c("Year_Index"))

}




nyears = max(n_samples$Year_Index)
# function to calculate a %/year trend from a count-scale trajectory
trs <- function(y1,y2,ny){
  tt <- (((y2/y1)^(1/ny))-1)*100
}


for(tl in c(2,6,11,21,51)){ #estimating all possible 1-year, 10-year, and full trends
  ny = tl-1
  yrs1 <- seq(1,(nyears-ny),by = ny)
  yrs2 <- yrs1+ny
  for(j in 1:length(yrs1)){
    y2 <- yrs2[j]
    y1 <- yrs1[j]

nyh2 <- paste0("Y",y2)
nyh1 <- paste0("Y",y1)
```

```
trends <- n_samples %>%
  filter(Year_Index %in% c(y1,y2)) %>%
  select(.draw,.value,Year_Index,strat) %>%
  group_by(.draw,Year_Index) %>%
  summarise(.value = mean(.value),
            .groups = "keep") %>%
  pivot_wider(.,names_from = Year_Index,
              values_from = .value,
              names_prefix = "Y") %>%
  rename_with(.,~gsub(pattern = nyh2,replacement = "YE", .x)) %>%
  rename_with(.,~gsub(pattern = nyh1,replacement = "YS", .x)) %>%
  group_by(.draw) %>%
  summarise(trend = trs(YS,YE,ny),
            .groups = "keep")%>%
  mutate(model = mod,
         spatial = spat,
         hierarchical = hier,
         first_year = y1,
         last_year = y2,
         nyears = ny,
         scale = "Survey_Wide",
         type = "full")



trends_out_tmp <- bind_rows(trends_out_tmp,trends)

trends <- n_samples %>%
  filter(Year_Index %in% c(y1,y2)) %>%
  select(.draw,.value,Year_Index,strat) %>%
  pivot_wider(.,names_from = Year_Index,
              values_from = .value,
              names_prefix = "Y") %>%
  rename_with(.,~gsub(pattern = nyh2,replacement = "YE", .x)) %>%
  rename_with(.,~gsub(pattern = nyh1,replacement = "YS", .x)) %>%
  group_by(.draw,strat,
            .groups = "keep") %>%
  summarise(trend = trs(YS,YE,ny),
            .groups = "keep")%>%
  mutate(model = mod,
         spatial = spat,
         hierarchical = hier,
         first_year = y1,
         last_year = y2,
         nyears = ny,
         scale = "Regional",
         type = "full")

trends_out_tmp <- bind_rows(trends_out_tmp,trends)


if(mod == "gamye"){
  trends <- nsmooth_samples %>%
```

```
  filter(Year_Index %in% c(y1,y2)) %>%
  select(.draw,.value,Year_Index,strat) %>%
  group_by(.draw,Year_Index) %>%
  summarise(.value = mean(.value),
            .groups = "keep") %>%
  pivot_wider(.,names_from = Year_Index,
              values_from = .value,
              names_prefix = "Y") %>%
  rename_with(.,~gsub(pattern = nyh2,replacement = "YE", .x)) %>%
  rename_with(.,~gsub(pattern = nyh1,replacement = "YS", .x)) %>%
  group_by(.draw) %>%
  summarise(trend = trs(YS,YE,ny),
            .groups = "keep")%>%
  mutate(model = mod,
         spatial = spat,
         hierarchical = hier,
         first_year = y1,
         last_year = y2,
         nyears = ny,
         scale = "Survey_Wide",
         type = "smooth")



trends_out_tmp <- bind_rows(trends_out_tmp,trends)

trends <- nsmooth_samples %>%
  filter(Year_Index %in% c(y1,y2)) %>%
  select(.draw,.value,Year_Index,strat) %>%
  pivot_wider(.,names_from = Year_Index,
              values_from = .value,
              names_prefix = "Y") %>%
  rename_with(.,~gsub(pattern = nyh2,replacement = "YE", .x)) %>%
  rename_with(.,~gsub(pattern = nyh1,replacement = "YS", .x)) %>%
  group_by(.draw,strat) %>%
  summarise(trend = trs(YS,YE,ny),
            .groups = "keep")%>%
  mutate(model = mod,
         spatial = spat,
         hierarchical = hier,
         first_year = y1,
         last_year = y2,
         nyears = ny,
         scale = "Regional",
         type = "smooth")
trends_out_tmp <- bind_rows(trends_out_tmp,trends)


}

}
}
save(file = paste0("output/prior_sim_summary",out_base,".RData"),
```

```
     list = c("trends_out_tmp",
              "summ"))

summ_out <- bind_rows(summ_out,summ)
trends_out <- bind_rows(trends_out,trends_out_tmp)
  }#prior_scale

saveRDS(trends_out,file = "output/prior_sim_trends.rds")
saveRDS(summ_out,file = "output/prior_sim_summaries.rds")
```

```
# loading the stored realised BBS and CBC trend estimates
 all_trends <- readRDS(file = "data/all_trends_bbs_cbc.rds")

all_sw_trends <- all_trends %>%
  filter(Region == "Survey_Wide")
all_politic_trends <- all_trends %>%
  filter(Region != "Survey_Wide")

mxabs = 2000#upper limit on the absolute trend estimates for the density plots below

## loading the prior simulated trends for all models
trends_out <- readRDS("output/prior_sim_trends.rds")


#summarising the trends for the political regions (not survey wide)
prior_trends_politic <- trends_out %>%
  filter(scale != "Survey_Wide") %>%
  mutate(abs_trend = abs(trend),
         t_years = paste(nyears,"year trends",sep = "-"),
         t_years = factor(t_years,
                          levels = c("1-year trends",
                                     "5-year trends",
                                     "10-year trends",
                                     "20-year trends",
                                     "50-year trends"),
                          ordered = TRUE),
         model_type = paste(model,spatial,hierarchical,sep = " "))

#setting up plotting label names for the models
mod_types <- unique(prior_trends_politic$model_type)
names(mod_types) <- gsub(mod_types,pattern = "(_)|[[:space:]]{2}",
                         replacement = " ")
realised_all_politic_freq <- vector(mode = "list",length = length(mod_types))
names(realised_all_politic_freq) <- mod_types

#looping through each model to generate the realised vs prior trend density plots
for(i in 1:length(mod_types)){
  mm = mod_types[i]
  mlab = names(mod_types)[i]

     tmp_sim <- prior_trends_politic %>%
    filter(model_type == mm)
```

```
tmp <- ggplot(data = all_politic_trends,
                                  aes(abs_trend,after_stat(density),
                                      groups = Survey))+
  geom_freqpoly(breaks = c(0,seq(0.5,mxabs,0.5)),center = 0,
                colour = grey(0.5))+
  geom_freqpoly(data = tmp_sim,
                aes(abs_trend,after_stat(density),
                colour = model_type),
                inherit.aes = FALSE,
                breaks = c(0,seq(0.5,mxabs,0.5)),center = 0)+
  scale_colour_viridis_d(begin = 0.8)+
  xlab("Absolute value of state/province trends USGS and Audubon models")+
  ylab("")+
  labs(title = paste0("Simulated state/province prior trends from ",mlab,"model"))+
  theme_bw()+
  theme(legend.position = "none")+
  coord_cartesian(ylim = c(0,0.7),
                  xlim = c(0,40))

  if(grepl(pattern = "gam",mm)){ # gamye models need to plots to represent the full and smooth only tre
   tmp <- tmp+facet_wrap(vars(type,t_years),
              nrow = 2,ncol = 5)
}else{
 tmp <- tmp+facet_wrap(vars(t_years),
              ncol = 5)
}

realised_all_politic_freq[[mm]] <- tmp

}


# same plots and trend summaries as above, but for the survey-wide estimates
prior_trends_sw <- trends_out %>%
  filter(scale == "Survey_Wide") %>%
  mutate(abs_trend = abs(trend),
         t_years = paste(nyears,"year trends",sep = "-"),
         t_years = factor(t_years,
                          levels = c("1-year trends",
                                     "5-year trends",
                                     "10-year trends",
                                     "20-year trends",
                                     "50-year trends"),
                          ordered = TRUE),
         model_type = paste(model,spatial,hierarchical,sep = " "))

realised_all_sw_freq <- vector(mode = "list",length = length(mod_types))
names(realised_all_sw_freq) <- mod_types

# looping through models to generate plots
for(i in 1:length(mod_types)){
  mm = mod_types[i]
  mlab = names(mod_types)[i]
```

```
  tmp_sim <- prior_trends_sw %>%
    filter(model_type == mm)
tmp <- ggplot(data = all_sw_trends,
                                  aes(abs_trend,after_stat(density),
                                      groups = Survey))+
  geom_freqpoly(breaks = c(0,seq(0.5,mxabs,0.5)),center = 0,
                colour = grey(0.5))+
  geom_freqpoly(data = tmp_sim,
                aes(abs_trend,after_stat(density),
                colour = model_type),
                inherit.aes = FALSE,
                breaks = c(0,seq(0.5,mxabs,0.5)),center = 0)+
  scale_colour_viridis_d(begin = 0.8)+
  xlab("Absolute value of survey wide trends USGS and Audubon models")+
  ylab("")+
  labs(title = paste0("Simulated survey wide prior trends from ",mlab,"model"))+
  theme_bw()+
  theme(legend.position = "none")+
  coord_cartesian(ylim = c(0,0.7),
                  xlim = c(0,40))

  if(grepl(pattern = "gam",mm)){
   tmp <- tmp+facet_wrap(vars(type,t_years),
               nrow = 2,ncol = 5)
 }else{
  tmp <- tmp+facet_wrap(vars(t_years),
               ncol = 5)
}

realised_all_sw_freq[[mm]] <- tmp

}

saveRDS(realised_all_sw_freq,
        "output/realised_all_sw_freq.rds")

saveRDS(realised_all_politic_freq,
        "output/realised_all_politic_freq.rds")
```

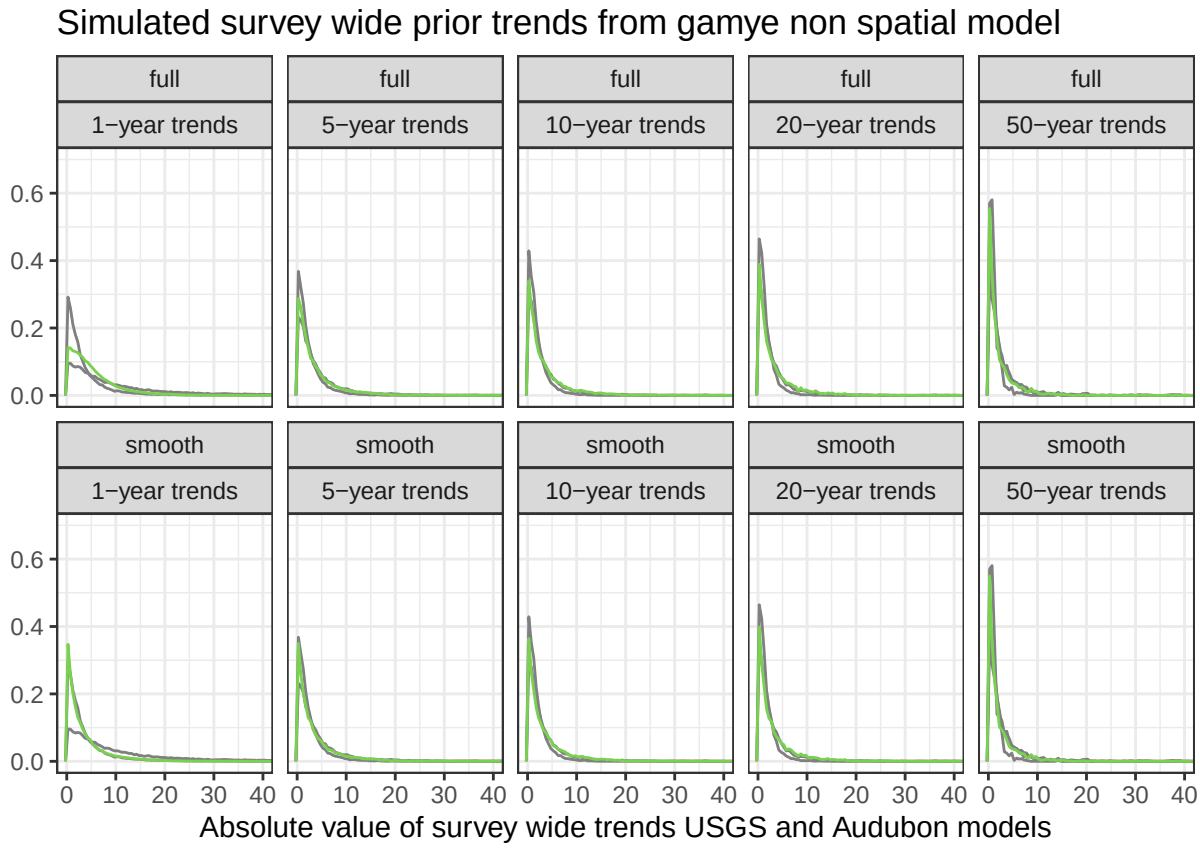
The code above plots the realised distribution of trends across the full time series of the BBS and CBC (black lines) along with the distribution of trends using the various priors used in the models in this paper.

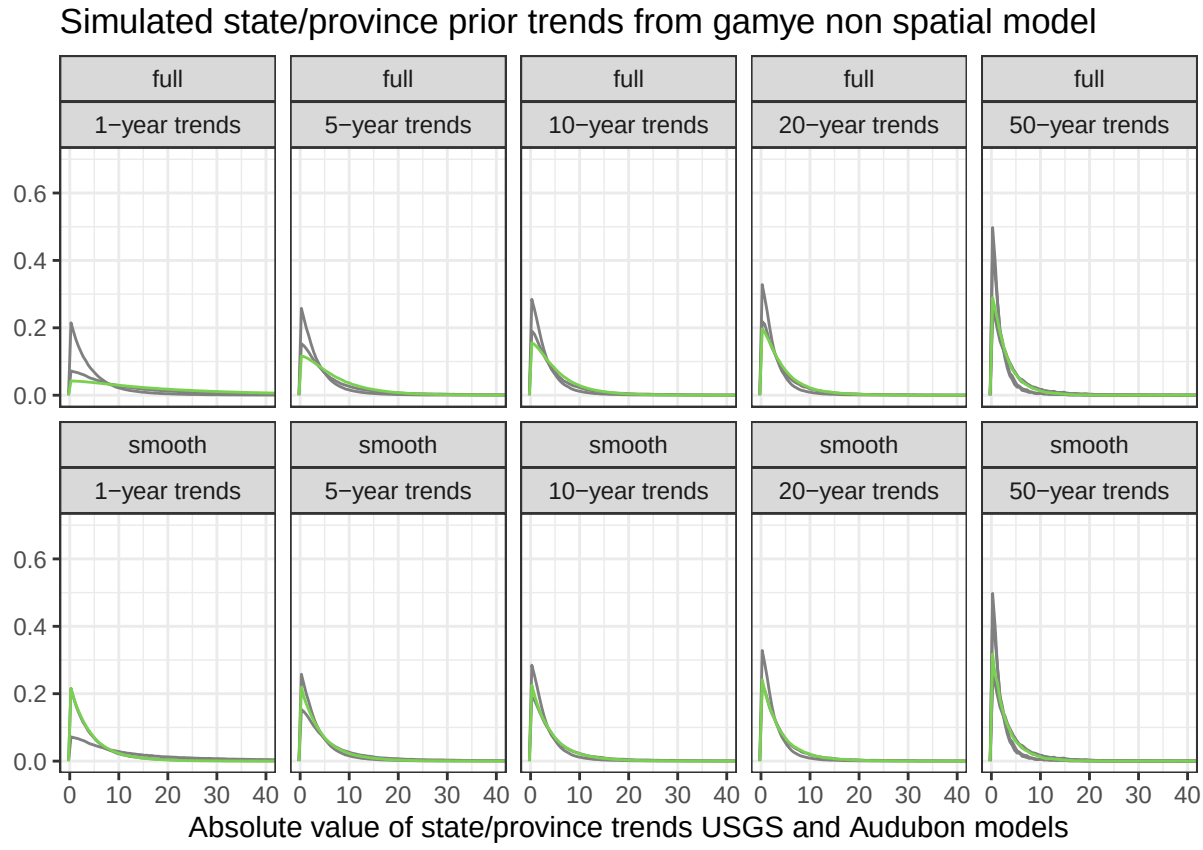
gamye non spatial

The survey wide trends, both full trajectory and the smooth only, from the gamye non-spatial model using half t-distribution with 0 mean, sd = 1, and 3 degrees of freedom for the standard deviation of the spline parameters that control the wiggliness of the mean population smooth, and a standard normal prior for wiggliness of the strata-level smooths. The standard deviation of the year-effects were given a gamma prior with shape parameter = 2 and scale parameter = 10.

```
## Warning: Combining variables of class <ordered> and <factor> was deprecated in ggplot2
## 3.4.0.
## i Please ensure your variables are compatible before plotting (location:
##   `combine_vars()`)
```

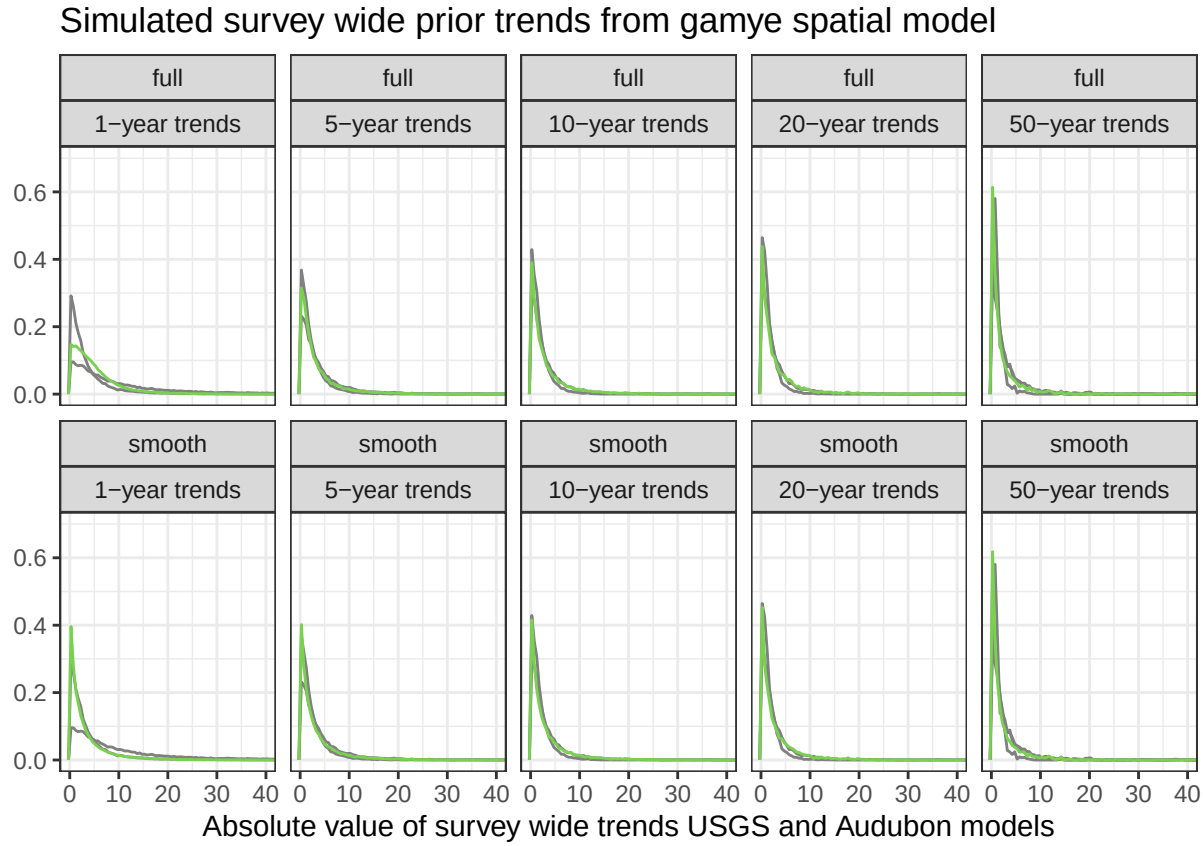
```
## Warning: Combining variables of class <ordered> and <factor> was deprecated in ggplot2
## 3.4.0.
## i Please ensure your variables are compatible before plotting (location:
##   `join_keys()`)
```

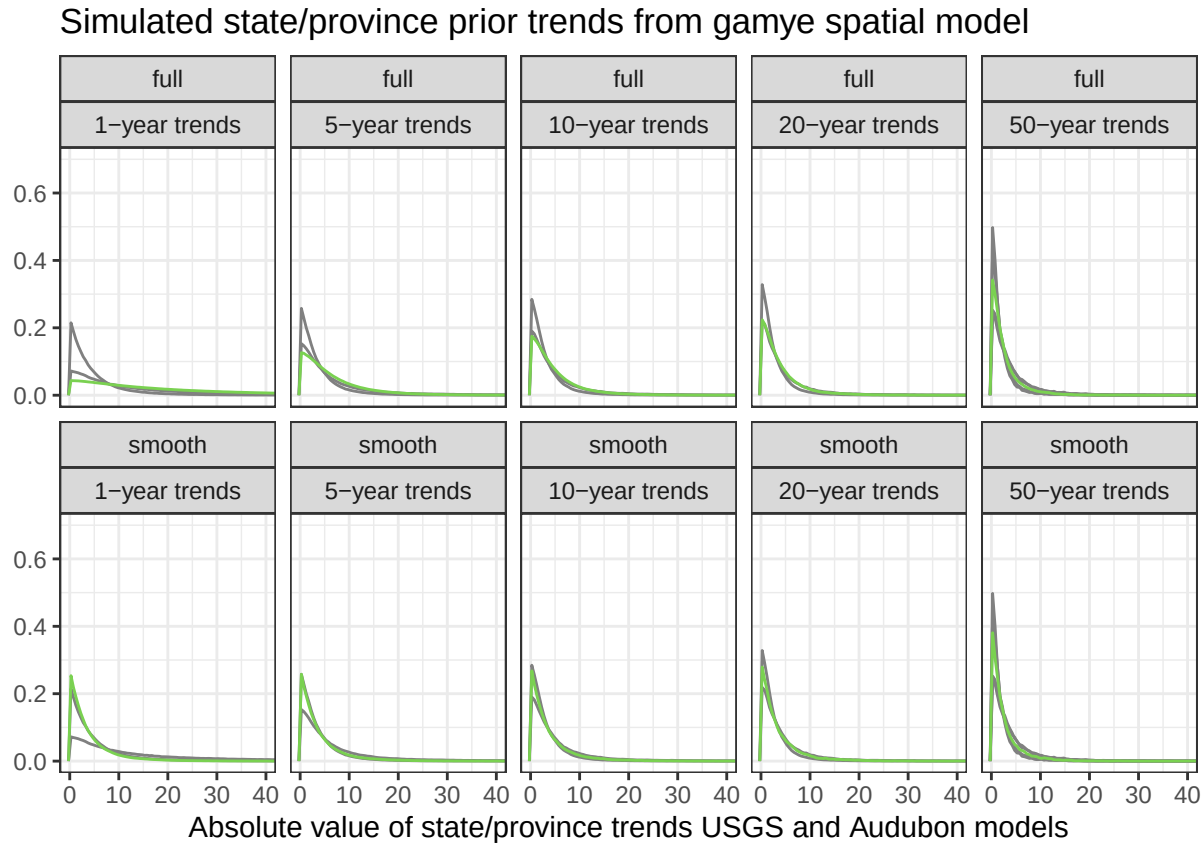



Simulated state/province prior trends from gamye non spatial model
full
1−year trends
full
5−year trends
full
10−year trends
full
20−year trends
full
50−year trends
smooth
1−year trends
smooth
5−year trends
smooth
10−year trends
smooth
20−year trends
smooth
50−year trends
0.0
0.2
0.4
0.6
0
10
20
30
40
Absolute value of state/province trends USGS and Audubon models

gamye spatial

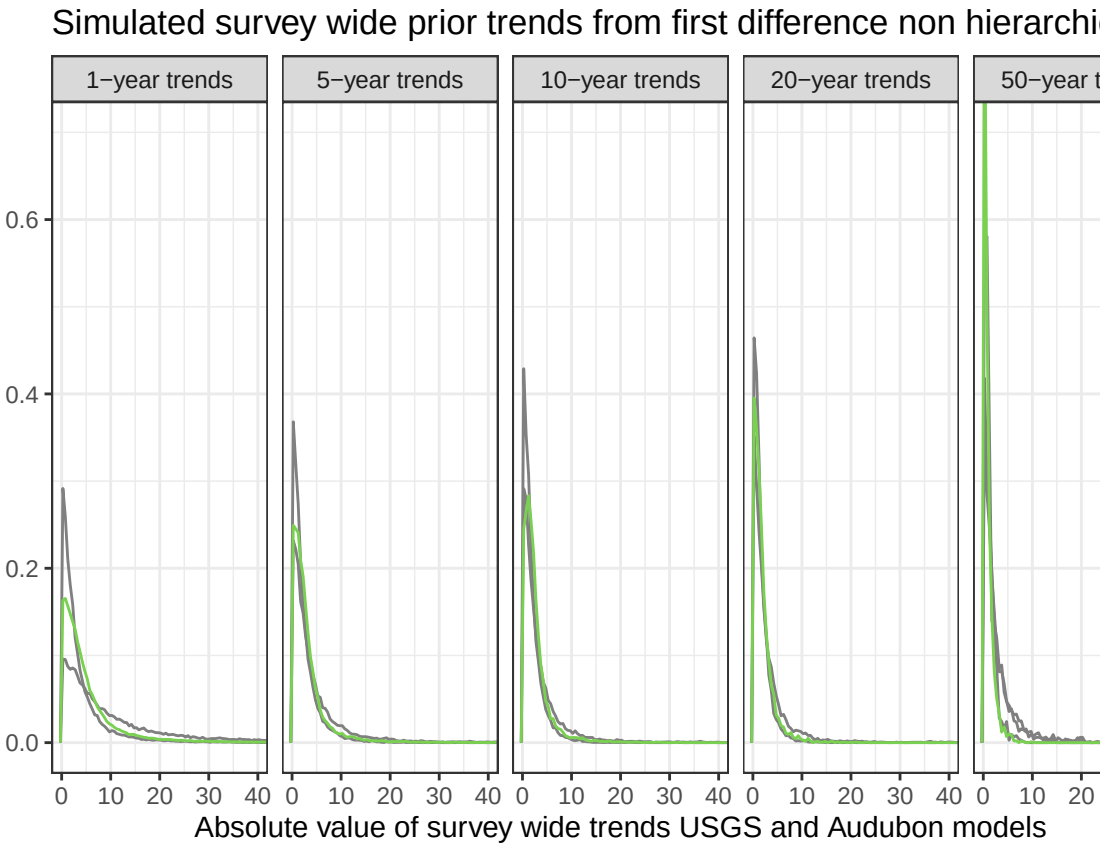
The survey wide trends, both full trajectory and the smooth only, from the spatial version of the gamye model using half t-distribution with 0 mean, sd = 1, and 3 degrees of freedom for the standard deviation of the spline parameters that control the wiggliness of the mean population smooth, and a standard normal prior for the spatial variation among strata on the spline parameters. The standard deviation of the year-effects were given a gamma prior with shape parameter = 2 and scale parameter = 10.

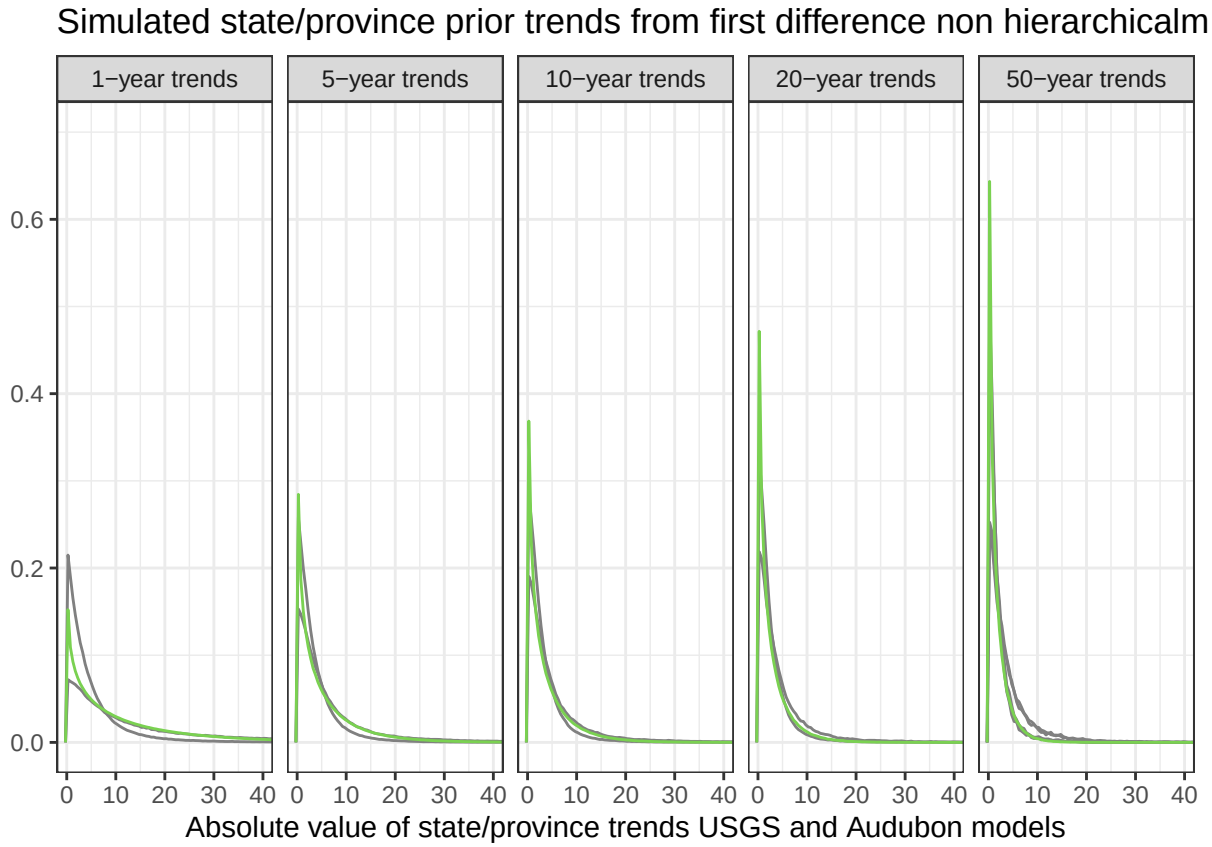



Simulated state/province prior trends from gamye spatial model
full
1−year trends
full
5−year trends
full
10−year trends
full
20−year trends
full
50−year trends
smooth
1−year trends
smooth
5−year trends
smooth
10−year trends
smooth
20−year trends
smooth
50−year trends
0.0
0.2
0.4
0.6
0
10
20
30
40
Absolute value of state/province trends USGS and Audubon models

first difference non hierarchical

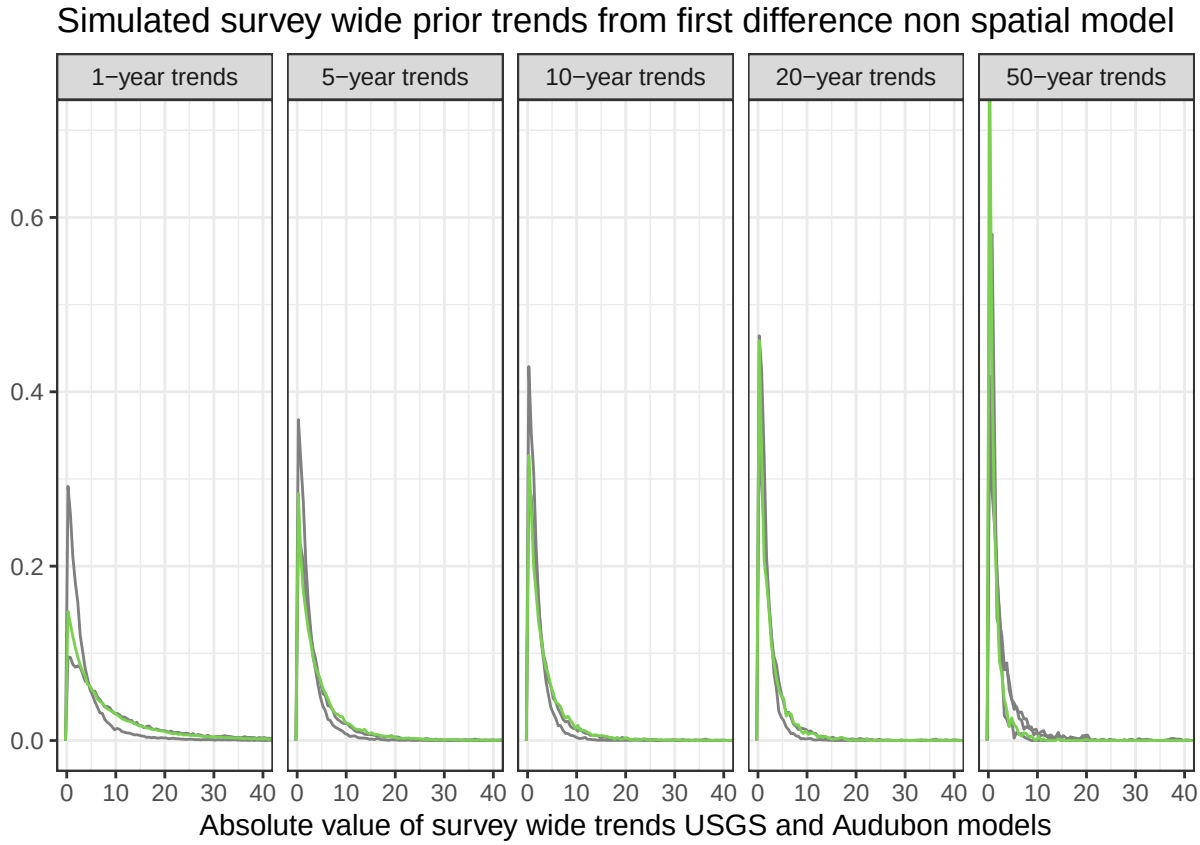
The survey wide trends, from the non hierarchical version of the first difference model using half t-distribution with 0 mean, sd = 0.2, and 3 degrees of freedom for the standard deviation of the annual differences in abundance in a given stratum.

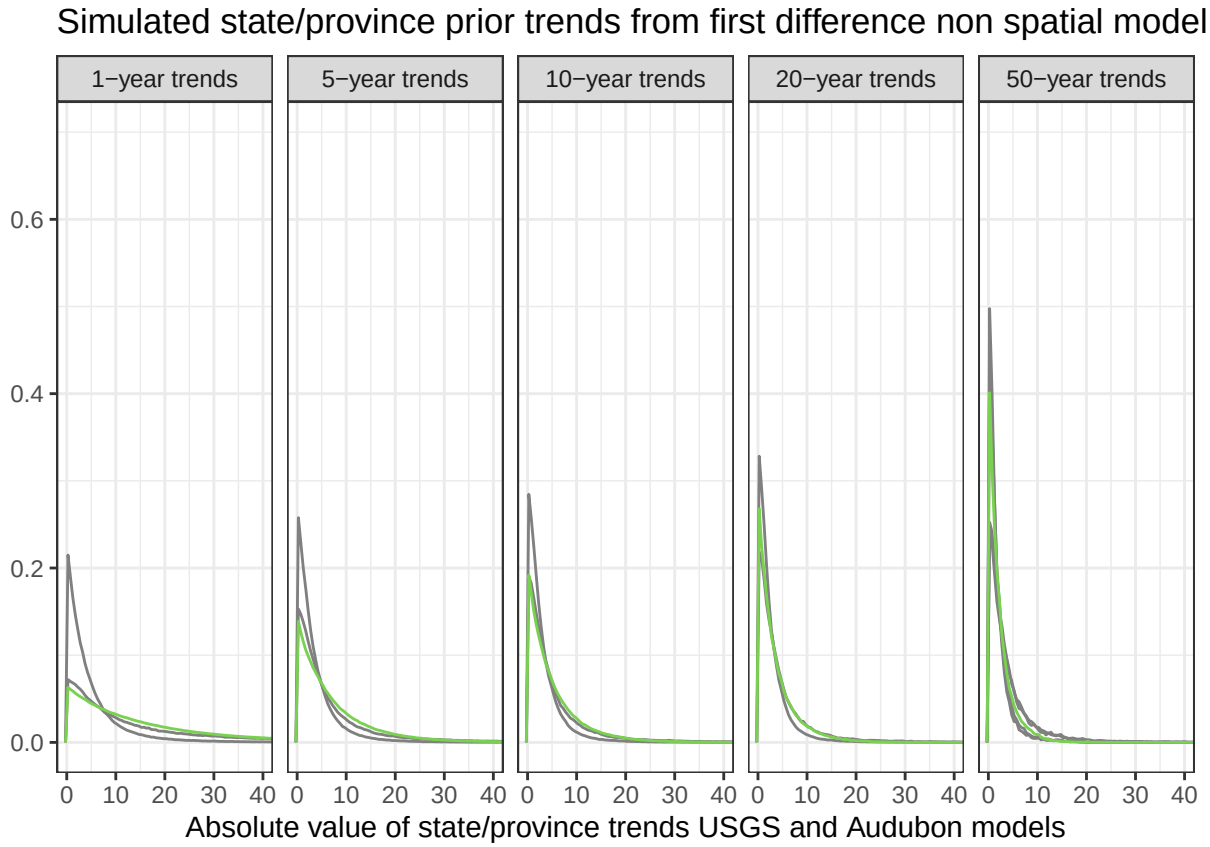



Simulated state/province prior trends from first difference non hierarchicalm
1−year trends
5−year trends
10−year trends
20−year trends
50−year trends
0.6
0.4
0.2
0.0
0
10
20
30
40
Absolute value of state/province trends USGS and Audubon models

first difference non spatial

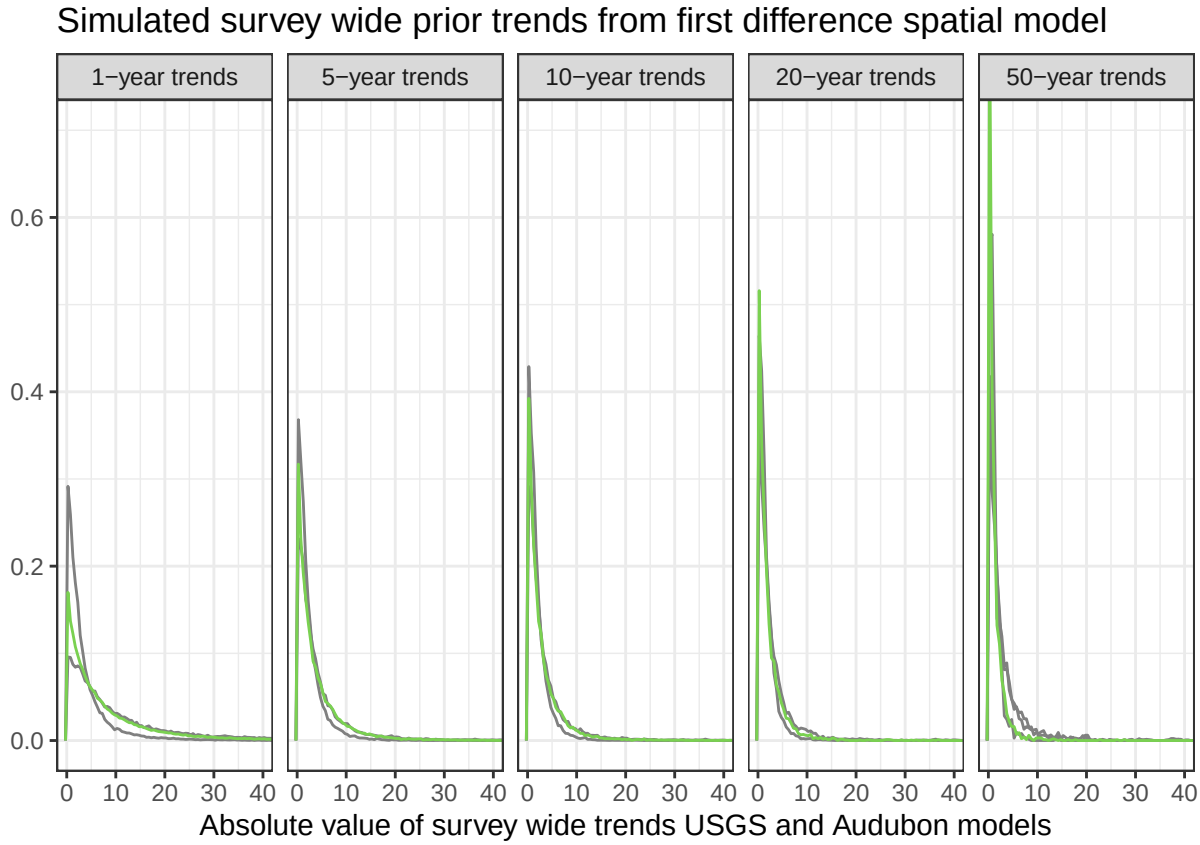
The survey wide trends, from the non spatial version of the first difference model using half t-distribution with 0 mean, sd = 0.1, and 3 degrees of freedom for the standard deviation of the mean overall annual differences in abundance, and a half t-distribution with a sd = 0.2 on the variation among strata in the differences.

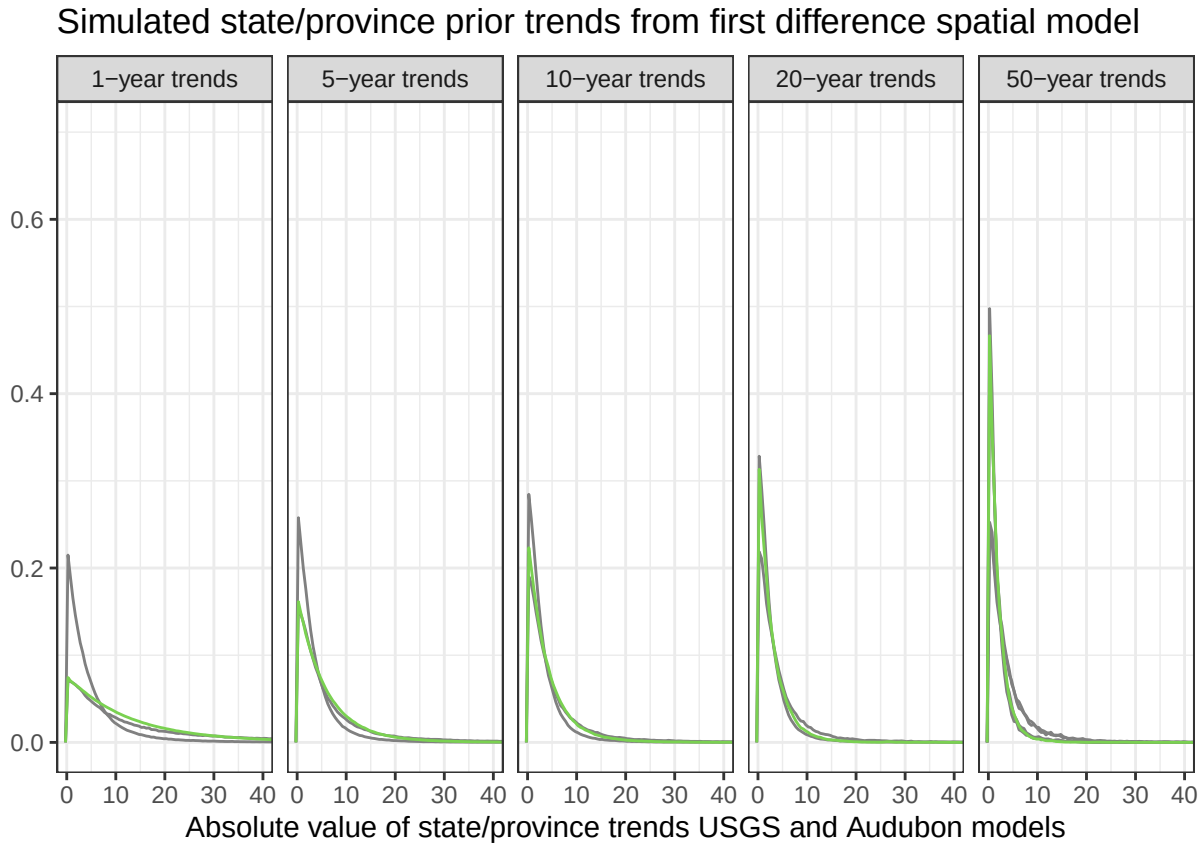



Simulated state/province prior trends from first difference non spatial model
1−year trends
5−year trends
10−year trends
20−year trends
50−year trends
0.6
0.4
0.2
0.0
0
10
20
30
40
Absolute value of state/province trends USGS and Audubon models

first difference non spatial

The survey wide trends, from the non spatial version of the first difference model using half t-distribution with 0 mean, sd = 0.1, and 3 degrees of freedom for the standard deviation of the mean overall annual differences in abundance, and a half t-distribution with a sd = 0.2 on the spatial variation among strata on the differences.



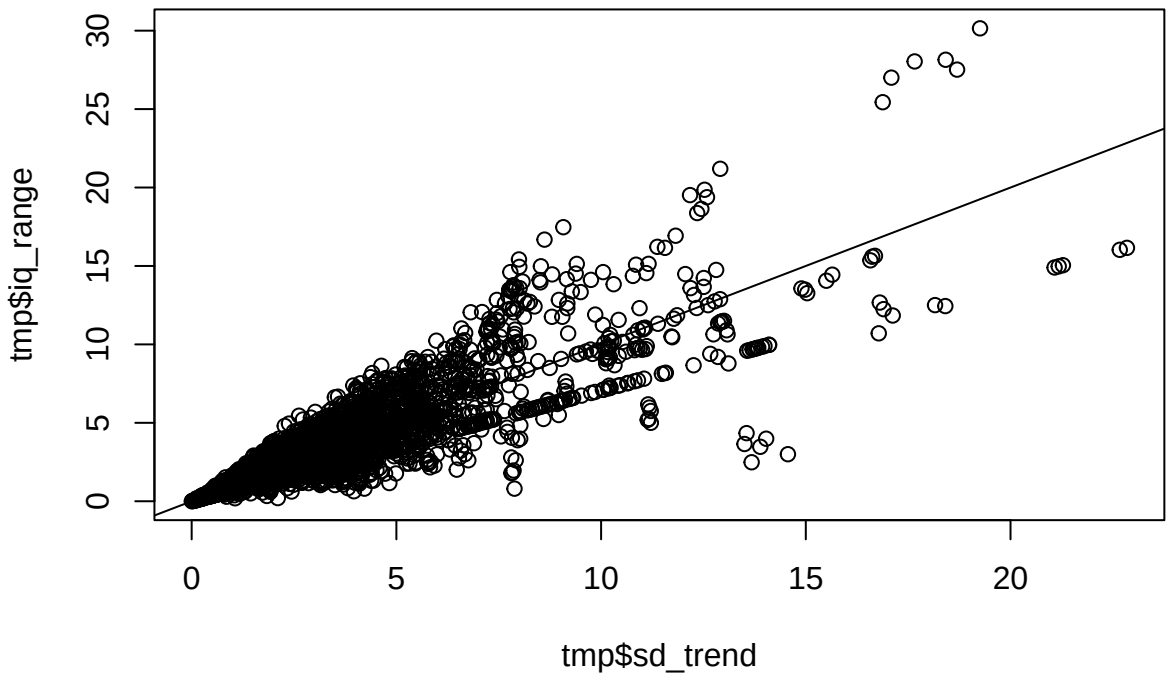

Simulated state/province prior trends from first difference spatial model
1−year trends
5−year trends
10−year trends
20−year trends
50−year trends
0.6
0.4
0.2
0.0
0
10
20
30
40
Absolute value of state/province trends USGS and Audubon models

Exploring the standard deviation of trends among regions

```
# loading the stored realised BBS and CBC trend estimates
 all_trends <- readRDS(file = "data/all_trends_bbs_cbc.rds")

all_sw_trends <- all_trends %>%
  filter(Region == "Survey_Wide")
all_politic_trends <- all_trends %>%
  filter(Region != "Survey_Wide")

realised_all_sd_freq <- vector(mode = "list",length = length(mod_types))
names(realised_all_sd_freq) <- mod_types

iq_func <- function(x,q = 0.5){
  q1 <- (1-q)/2
  q2 <- 1-q1
  iq <- quantile(x,q2) - quantile(x,q1)
  return(iq)
}

## summarizing the sd of realised trends
all_politic_sdtrends <- all_politic_trends %>%
  filter(Survey %in% c("BBS","CBC")) %>%
  group_by(Survey,AOU,first_year,last_year,nyears) %>%
  summarise(sd_trend = sd(trend),
            iq_range = iq_func(trend)) %>%
  mutate(t_years = paste(nyears,"year trends",sep = "-"),
         t_years = factor(t_years,
                          levels = c("1-year trends",
                                     "5-year trends",
                                     "10-year trends",
                                     "20-year trends",
                                     "50-year trends"),
                          ordered = TRUE))
```

```
## `summarise()` has grouped output by 'Survey', 'AOU', 'first_year', 'last_year'.
## You can override using the `.groups` argument.
```

```
tmp <- all_politic_sdtrends %>%
  filter(t_years == "50-year trends")

plot(tmp$sd_trend,tmp$iq_range)
abline(0,1)
```




```
## loading the prior simulated trends for all models
trends_out <- readRDS("output/prior_sim_trends.rds")

#summarising the trends for the political regions (not survey wide)
prior_sdtrends_politic <- trends_out %>%
  filter(scale != "Survey_Wide") %>%
  group_by(.draw,model,spatial,hierarchical,first_year,last_year,nyears,type) %>%
  summarise(sd_trend = sd(trend),
            iq_range = iq_func(trend)) %>%
  mutate(t_years = paste(nyears,"year trends",sep = "-"),
         t_years = factor(t_years,
                          levels = c("1-year trends",
                                     "5-year trends",
                                     "10-year trends",
                                     "20-year trends",
                                     "50-year trends"),
                          ordered = TRUE),
         model_type = paste(model,spatial,hierarchical,sep = " "))
```

```
## `summarise()` has grouped output by '.draw', 'model', 'spatial',
## 'hierarchical', 'first_year', 'last_year', 'nyears'. You can override using the
## `.groups` argument.
```

```
# looping through models to generate plots
for(i in 1:length(mod_types)){
```

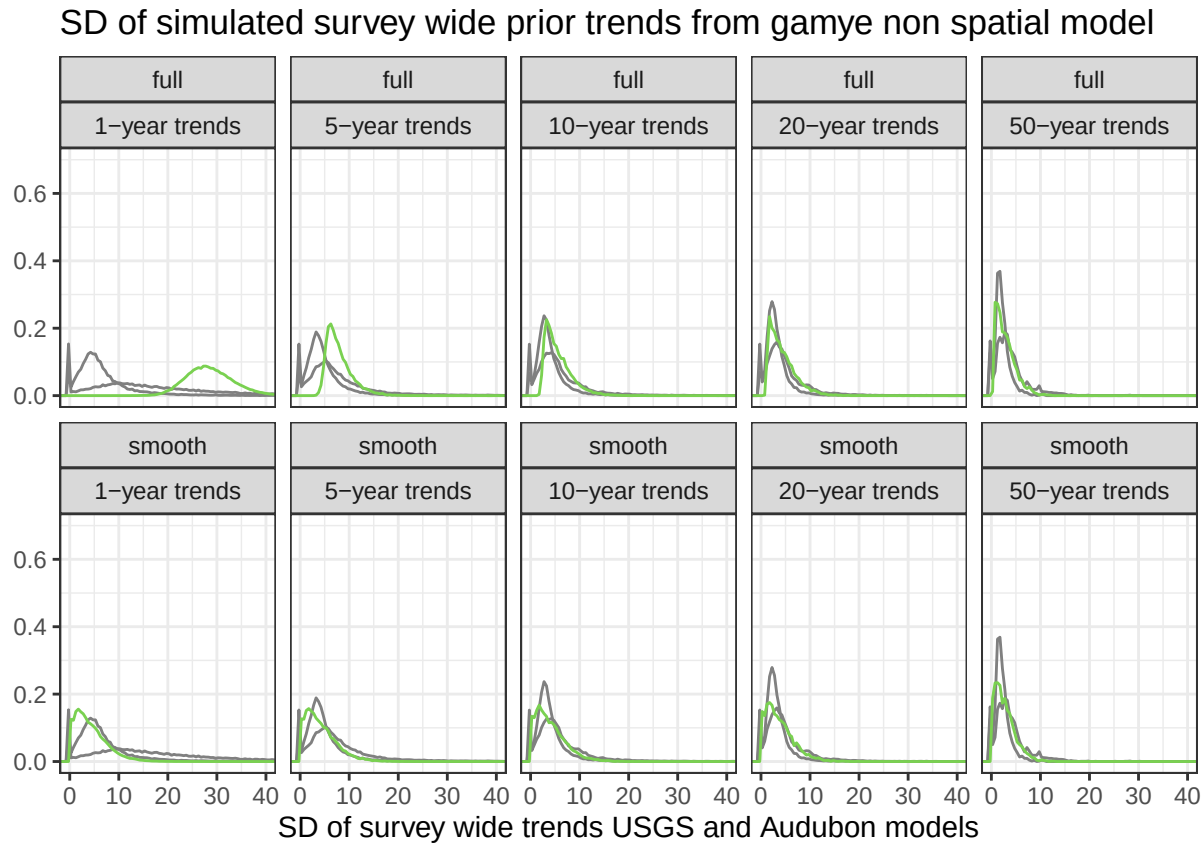
```
  mm = mod_types[i]
  mlab = names(mod_types)[i]
  tmp_sim <- prior_sdtrends_politic %>%
    filter(model_type == mm)
tmp <- ggplot(data = all_politic_sdtrends,
                                  aes(iq_range,after_stat(density),
                                      groups = Survey))+
  geom_freqpoly(breaks = c(seq(-mxabs,mxabs,0.5)),center = 0,
                colour = grey(0.5))+
  geom_freqpoly(data = tmp_sim,
    aes(iq_range,after_stat(density),
    colour = model_type),
    inherit.aes = FALSE,
    breaks = c(seq(-mxabs,mxabs,0.5)),center = 0)+
  scale_colour_viridis_d(begin = 0.8)+
  xlab("SD of survey wide trends USGS and Audubon models")+
  ylab("")+
  labs(title = paste0("SD of simulated survey wide prior trends from ",mlab,"model"))+
  theme_bw()+
  theme(legend.position = "none")+
  coord_cartesian(ylim = c(0,0.7),
                  xlim = c(0,40))

  if(grepl(pattern = "gam",mm)){
   tmp <- tmp+facet_wrap(vars(type,t_years),
              nrow = 2,ncol = 5)
 }else{
  tmp <- tmp+facet_wrap(vars(t_years),
              ncol = 5)
}

realised_all_sd_freq[[mm]] <- tmp

}
```

The code above plots the realised distribution of the standard deviation of trends among regions within species for all possible time-windows across the full time series of the BBS and CBC (black lines) along with the distribution of standard deviation of trends from the various priors used in the models in this paper.

gamye non spatial

The survey wide trends, both full trajectory and the smooth only, from the gamye non-spatial model using half t-distribution with 0 mean, sd = 1, and 3 degrees of freedom for the standard deviation of the spline parameters that control the wiggliness of the mean population smooth, and the same prior for wiggliness of the strata-level smooths. The standard deviation of the year-effects were given a gamma prior with shape parameter = 2 and scale parameter = 10.

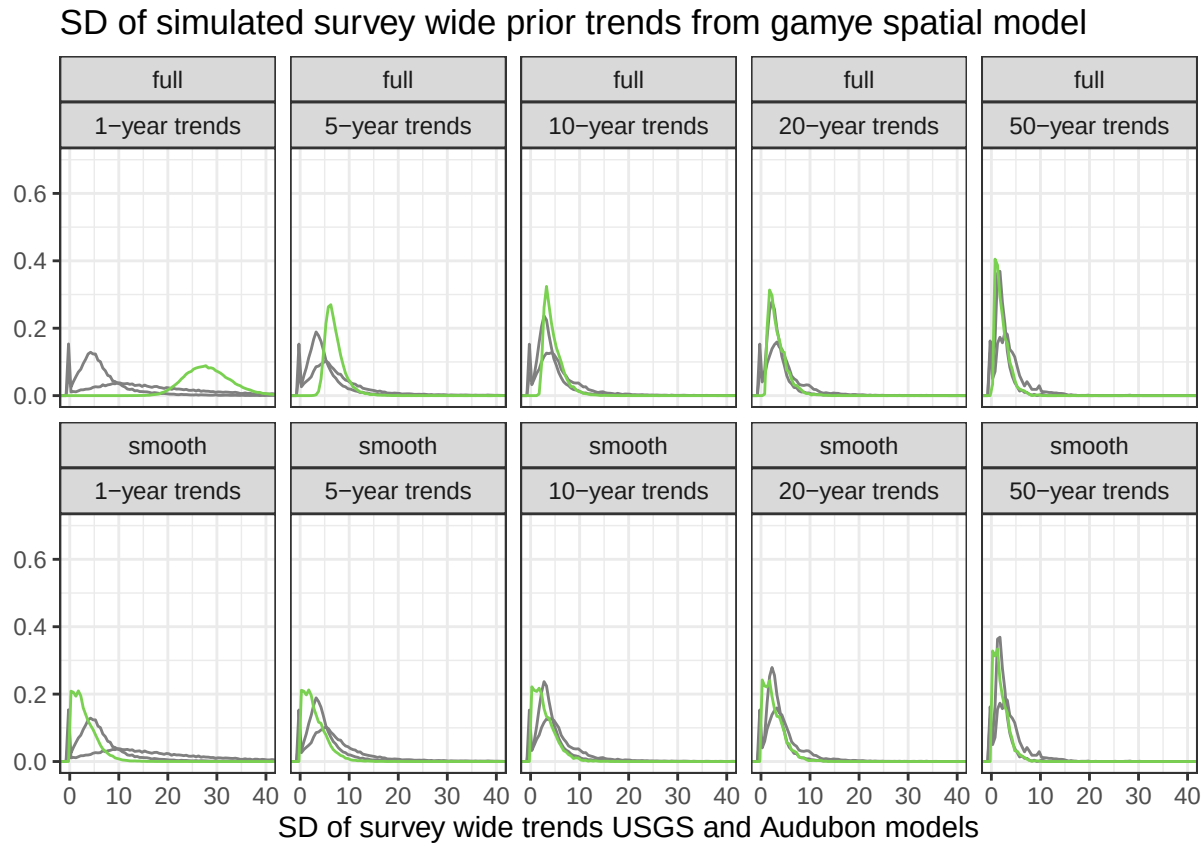
```
print(realised_all_sd_freq[[1]])
```



gamye spatial

The survey wide trends, both full trajectory and the smooth only, from the spatial version of the gamye model using half t-distribution with 0 mean, sd = 1, and 3 degrees of freedom for the standard deviation of the spline parameters that control the wiggliness of the mean population smooth, and the same prior for the spatial variation among strata on the spline parameters. The standard deviation of the year-effects were given a gamma prior with shape parameter = 2 and scale parameter = 10.

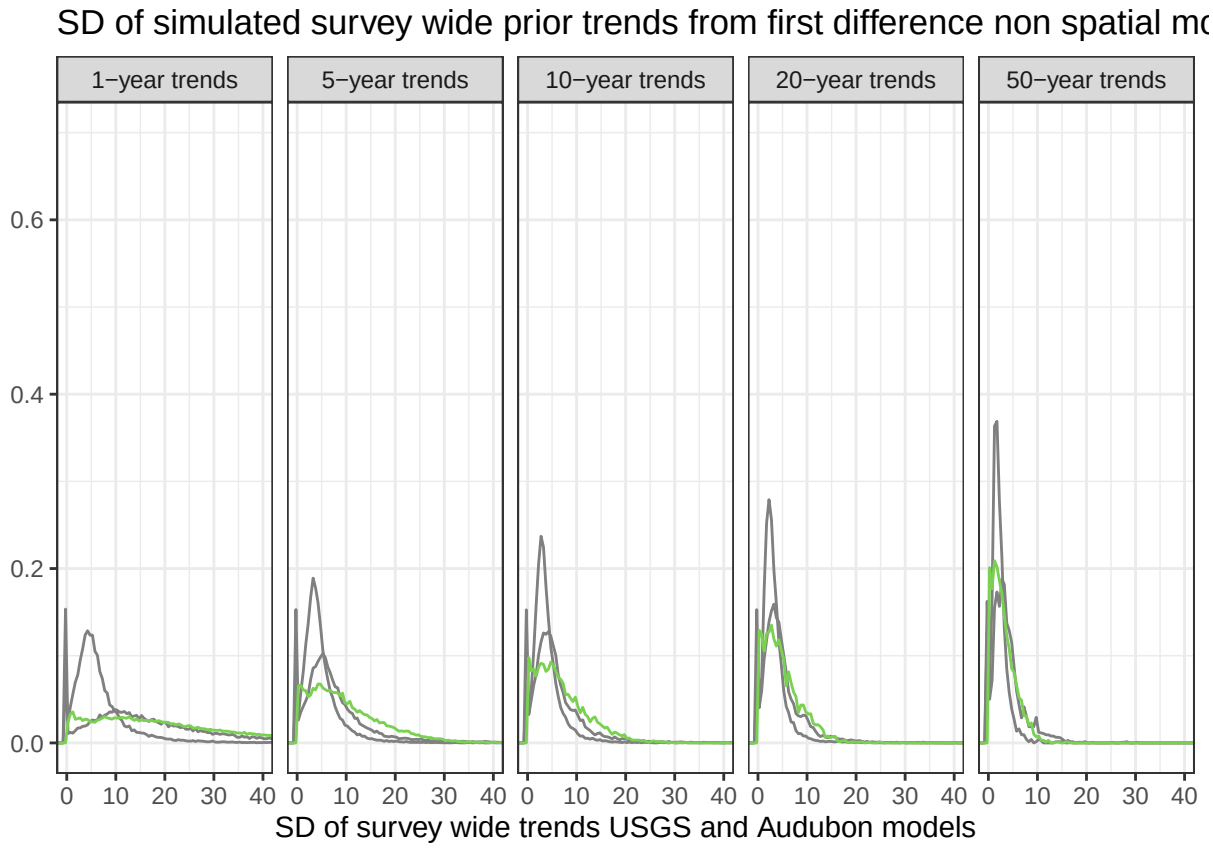
```
print(realised_all_sd_freq[[2]])
```



first difference non spatial

The survey wide trends, from the non spatial version of the first difference model using half t-distribution with 0 mean, sd = 0.1, and 3 degrees of freedom for the standard deviation of the mean overall annual differences in abundance, and a half t-distribution with a sd = 0.2 on the variation among strata in the differences.

```
print(realised_all_sd_freq[[4]])
```



first difference non spatial

The survey wide trends, from the non spatial version of the first difference model using half t-distribution with 0 mean, sd = 0.1, and 3 degrees of freedom for the standard deviation of the mean overall annual differences in abundance, and a half t-distribution with a sd = 0.2 on the spatial variation among strata on the differences.

```
print(realised_all_sd_freq[[5]])
```