SURE Project

September 08, 2022

Objective

Compare spatial predictions of kelp to "in situ" survey data. Compare each year and location for 150, 300, 600, 900 resolutions.

Extraction

Extract the predicted *log* kelps density of every year (2004 - 2021) for each site in the North Coast.

```
# set a directory
w.dir <- here()
d.dir <- here('data')
r.dir <- here('spatial_data/sp_predictions_5.1.1_V2')
r1.dir <- here('spatial_data/sp_predictions_900m_resolution')
r2.dir <- here('spatial_data/sp_predictions_150m_resolution')
r3.dir <- here('spatial_data/sp_predictions_600m_resolution')

# read and transform the observed data to the log scale
df <- read.csv(paste(d.dir,
                     'RCCA_kelp_inverts_NC_depth-zones_wave_clim_temp_nit_subs_orbvel_npp.csv',
                     sep = '/')) %>%
  dplyr::select(site_name, year, transect, zone, latitude, longitude, den_NERLUE) %>%
  mutate_at(vars(year, transect, zone, site_name), list(as.factor)) %>%
  mutate(log_den_NERLUE = log(den_NERLUE))
head(df)
```

```
##   site_name year transect  zone latitude longitude den_NERLUE log_den_NERLUE
## 1    Caspar 2018        1 INNER 39.36173  -123.822          0           -Inf
## 2    Caspar 2018        2 INNER 39.36173  -123.822          0           -Inf
## 3    Caspar 2018        3 INNER 39.36173  -123.822          0           -Inf
## 4    Caspar 2018        4 OUTER 39.36173  -123.822          0           -Inf
## 5    Caspar 2018        5 OUTER 39.36173  -123.822          0           -Inf
## 6    Caspar 2018        6 OUTER 39.36173  -123.822          0           -Inf
```

Note that log(0) returns -Inf. How to deal with log(0)?

```
df$log_den_NERLUE <- replace(df$log_den_NERLUE, df$log_den_NERLUE == -Inf, 0)
```

Calculate the mean and standard error of kelps density of every year for each site by zone (INNER/OUTER).

```
obs <- df %>%
  group_by(site_name, year, zone) %>%
  summarise_at(vars(log_den_NERLUE), list(mean = mean, se = std.error), na.rm = TRUE) %>%
  pivot_wider(names_from = zone, values_from = c(mean, se))
head(obs)
```

```
## # A tibble: 6 x 6
## # Groups:   site_name, year [6]
##   site_name year  mean_INNER mean_OUTER se_INNER se_OUTER
##   <fct>     <fct>      <dbl>      <dbl>    <dbl>    <dbl>
## 1 Caspar    2008       4.38       3.03   0.150    0.996
## 2 Caspar    2010       4.37       4.17   0.0664   0.586
## 3 Caspar    2014       0.799      0      0.799    0
## 4 Caspar    2015       0          0      0        0
## 5 Caspar    2016       0          0      0        0
## 6 Caspar    2017       0          0      0        0
```

Extract the predicted *log* kelps density of every year for each site at different resolutions.

```
# kelp density predictions at 300m resolution
pred_300m <- read.csv(paste(d.dir, 'NC_kelp_density_predictions.csv', sep ='/')) %>%
  mutate_at(vars(year, site_name), list(as.factor))
head(pred_300m)
```

```
##        site_name year       fit longitude latitude
## 1         Caspar 2006 0.5090384 -123.8220 39.36173
## 2   Caspar North 2006 0.5003366 -123.8213 39.36443
## 3     Dark Gulch 2006 0.5309656 -123.7762 39.24030
## 4 Flat Iron Rock 2006 0.7382968 -124.1578 41.05942
## 5      Fort Ross 2006        NA -123.2450 38.51060
## 6     Frolic Cove 2006 0.8228083 -123.8239 39.35503
```

```
# kelp density predictions at 900m resolution

# read the .csv file
site <- read.csv(paste(d.dir, 'RCCA_North_Coast_sites.csv', sep = '/'))
# convert from .csv to .shp
site_shp <- st_as_sf(site, coords = c('longitude', 'latitude'), crs = 'EPSG:4326')

# declaring an empty data frame
pred <- data.frame(site_name = character(),
                   year = numeric(),
                   fit = numeric())

for (i in c(2006:2021)) {
  rast <- rast(paste0(r1.dir, paste0('/', i, '_Log_Nereo_NC.tif')))
  ext <- terra::extract(rast, vect(site_shp$geometry)) %>%
    mutate(site_name = site$site_name, year = as.factor(i), .before = fit) %>%
    dplyr::select(-ID)
  pred <- rbind(pred, ext)
}

head(pred)
```

```
##        site_name year         fit
## 1         Caspar 2006 0.07573552
## 2   Caspar North 2006 0.07573552
## 3     Dark Gulch 2006 0.32145628
## 4 Flat Iron Rock 2006 0.17760107
## 5      Fort Ross 2006         NaN
## 6    Frolic Cove 2006 0.24542649
```

```r
# write to cvs
merge_df <- left_join(pred,
                      site %>% dplyr::select(c(site_name, longitude, latitude)),
                      by = 'site_name')

# write.csv(merge_df, file.path(d.dir, 'NC_kelp_density_predictions_900m_resolution.csv'), row.names = .
```

```r
# kelp density predictions at 150m resolution

# read the .csv file
site <- read.csv(paste(d.dir, 'RCCA_North_Coast_sites.csv', sep = '/'))
# convert from .csv to .shp
site_shp <- st_as_sf(site, coords = c('longitude', 'latitude'), crs = 'EPSG:4326')

# declaring an empty data frame
pred <- data.frame(site_name = character(),
                   year = numeric(),
                   fit = numeric())

for (i in c(2006:2021)) {
  rast <- rast(paste0(r2.dir, paste0('/', i, '_Log_Nereo_NC.tif')))
  ext <- terra::extract(rast, vect(site_shp$geometry)) %>%
    mutate(site_name = site$site_name, year = as.factor(i), .before = fit) %>%
    dplyr::select(-ID)
  pred <- rbind(pred, ext)
}

head(pred)
```

```
##        site_name year         fit
## 1         Caspar 2006 0.08525774
## 2   Caspar North 2006 0.19914755
## 3     Dark Gulch 2006 0.17771323
## 4 Flat Iron Rock 2006 0.17733660
## 5      Fort Ross 2006         NaN
## 6    Frolic Cove 2006 0.12495851
```

```r
# write to cvs
merge_df <- left_join(pred,
                      site %>% dplyr::select(c(site_name, longitude, latitude)),
                      by = 'site_name')

# write.csv(merge_df, file.path(d.dir, 'NC_kelp_density_predictions_150m_resolution.csv'), row.names = .
```

```r
# kelp density predictions at 600m resolution

# read the .csv file
site <- read.csv(paste(d.dir, 'RCCA_North_Coast_sites.csv', sep = '/'))
# convert from .csv to .shp
site_shp <- st_as_sf(site, coords = c('longitude', 'latitude'), crs = 'EPSG:4326')

# declaring an empty data frame
pred <- data.frame(site_name = character(),
                   year = numeric(),
                   fit = numeric())

for (i in c(2006:2021)) {
  rast <- rast(paste0(r3.dir, paste0('/', i, '_Log_Nereo_NC.tif')))
  ext <- terra::extract(rast, vect(site_shp$geometry)) %>%
    mutate(site_name = site$site_name, year = as.factor(i), .before = fit) %>%
    dplyr::select(-ID)
  pred <- rbind(pred, ext)
}

head(pred)
```

```
##         site_name year        fit
## 1          Caspar 2006 0.06638455
## 2   Caspar North 2006 0.23112571
## 3      Dark Gulch 2006 0.22723143
## 4 Flat Iron Rock 2006 0.15889785
## 5       Fort Ross 2006        NaN
## 6     Frolic Cove 2006 0.17989336
```

```r
# write to cvs
merge_df <- left_join(pred,
                      site %>% dplyr::select(c(site_name, longitude, latitude)),
                      by = 'site_name')

write.csv(merge_df, file.path(d.dir, 'NC_kelp_density_predictions_600m_resolution.csv'), row.names = FAl
```

Comparison

```r
kelp_data_300m <- left_join(pred_300m, obs, by = c('site_name', 'year')) %>%
  group_by(site_name) %>%
  arrange(year, .by_group = TRUE) %>%
  relocate(fit, .after = last_col())
head(kelp_data_300m)
```

```
## # A tibble: 6 x 9
## # Groups:   site_name [1]
##   site_name year  longitude latitude mean_INNER mean_OUTER se_IN~1 se_OU~2   fit
##   <fct>     <fct>     <dbl>    <dbl>      <dbl>      <dbl>   <dbl>   <dbl> <dbl>
## 1 Caspar    2006      -124.     39.4         NA         NA      NA      NA 0.509
```

```
## 2 Caspar      2007      -124.      39.4       NA        NA    NA       NA   0.827
## 3 Caspar      2008      -124.      39.4      4.38      3.03  0.150    0.996 1.83
## 4 Caspar      2009      -124.      39.4       NA        NA    NA       NA   1.62
## 5 Caspar      2010      -124.      39.4      4.37      4.17  0.0664   0.586 0.554
## 6 Caspar      2011      -124.      39.4       NA        NA    NA       NA   0.857
## # ... with abbreviated variable names 1: se_INNER, 2: se_OUTER
```

```r
# kelp density predictions at 900m resolution
pred_900m <- read.csv(paste(d.dir, 'NC_kelp_density_predictions_900m_resolution.csv', sep ='/')) %>%
  mutate_at(vars(year, site_name), list(as.factor))
head(pred_900m)
```

```
##            site_name year        fit longitude latitude
## 1            Caspar 2006 0.07573552 -123.8220 39.36173
## 2      Caspar North 2006 0.07573552 -123.8213 39.36443
## 3        Dark Gulch 2006 0.32145628 -123.7762 39.24030
## 4 Flat Iron Rock 2006 0.17760107 -124.1578 41.05942
## 5          Fort Ross 2006         NA -123.2450 38.51060
## 6        Frolic Cove 2006 0.24542649 -123.8239 39.35503
```

```r
kelp_data_900m <- left_join(pred_900m, obs, by = c('site_name', 'year')) %>%
  group_by(site_name) %>%
  arrange(year, .by_group = TRUE) %>%
  relocate(fit, .after = last_col())
head(kelp_data_900m)
```

```
## # A tibble: 6 x 9
## # Groups:   site_name [1]
##    site_name year  longitude latitude mean_INNER mean_OU~1 se_IN~2 se_OU~3    fit
##    <fct>     <fct>     <dbl>    <dbl>      <dbl>     <dbl>   <dbl>   <dbl>  <dbl>
## 1 Caspar     2006     -124.     39.4       NA        NA    NA       NA   0.0757
## 2 Caspar     2007     -124.     39.4       NA        NA    NA       NA   1.50
## 3 Caspar     2008     -124.     39.4      4.38      3.03  0.150    0.996 3.58
## 4 Caspar     2009     -124.     39.4       NA        NA    NA       NA   5.59
## 5 Caspar     2010     -124.     39.4      4.37      4.17  0.0664   0.586 0.348
## 6 Caspar     2011     -124.     39.4       NA        NA    NA       NA   1.45
## # ... with abbreviated variable names 1: mean_OUTER, 2: se_INNER, 3: se_OUTER
```

```r
# kelp density predictions at 150m resolution
pred_150m <- read.csv(paste(d.dir, 'NC_kelp_density_predictions_150m_resolution.csv', sep ='/')) %>%
  mutate_at(vars(year, site_name), list(as.factor))
head(pred_150m)
```

```
##            site_name year        fit longitude latitude
## 1            Caspar 2006 0.08525774 -123.8220 39.36173
## 2      Caspar North 2006 0.19914755 -123.8213 39.36443
## 3        Dark Gulch 2006 0.17771323 -123.7762 39.24030
## 4 Flat Iron Rock 2006 0.17733660 -124.1578 41.05942
## 5          Fort Ross 2006         NA -123.2450 38.51060
## 6        Frolic Cove 2006 0.12495851 -123.8239 39.35503
```

```
kelp_data_150m <- left_join(pred_150m, obs, by = c('site_name', 'year')) %>%
  group_by(site_name) %>%
  arrange(year, .by_group = TRUE) %>%
  relocate(fit, .after = last_col())
head(kelp_data_150m)
```

```
## # A tibble: 6 x 9
## # Groups:   site_name [1]
##   site_name year  longitude latitude mean_INNER mean_OU~1 se_IN~2 se_OU~3    fit
##   <fct>     <fct>     <dbl>    <dbl>      <dbl>     <dbl>   <dbl>   <dbl>  <dbl>
## 1 Caspar    2006      -124.     39.4         NA        NA      NA      NA 0.0853
## 2 Caspar    2007      -124.     39.4         NA        NA      NA      NA 1.69
## 3 Caspar    2008      -124.     39.4       4.38      3.03   0.150   0.996 2.84
## 4 Caspar    2009      -124.     39.4         NA        NA      NA      NA 5.79
## 5 Caspar    2010      -124.     39.4       4.37      4.17  0.0664   0.586 0.446
## 6 Caspar    2011      -124.     39.4         NA        NA      NA      NA 1.66
## # ... with abbreviated variable names 1: mean_OUTER, 2: se_INNER, 3: se_OUTER
```

```
# kelp density predictions at 600m resolution
pred_600m <- read.csv(paste(d.dir, 'NC_kelp_density_predictions_600m_resolution.csv', sep ='/')) %>%
  mutate_at(vars(year, site_name), list(as.factor))
head(pred_600m)
```

```
##          site_name year        fit longitude latitude
## 1           Caspar 2006 0.06638455 -123.8220 39.36173
## 2     Caspar North 2006 0.23112571 -123.8213 39.36443
## 3       Dark Gulch 2006 0.22723143 -123.7762 39.24030
## 4   Flat Iron Rock 2006 0.15889785 -124.1578 41.05942
## 5        Fort Ross 2006         NA -123.2450 38.51060
## 6       Frolic Cove 2006 0.17989336 -123.8239 39.35503
```

```
kelp_data_600m <- left_join(pred_600m, obs, by = c('site_name', 'year')) %>%
  group_by(site_name) %>%
  arrange(year, .by_group = TRUE) %>%
  relocate(fit, .after = last_col())
head(kelp_data_600m)
```

```
## # A tibble: 6 x 9
## # Groups:   site_name [1]
##   site_name year  longitude latitude mean_INNER mean_OU~1 se_IN~2 se_OU~3    fit
##   <fct>     <fct>     <dbl>    <dbl>      <dbl>     <dbl>   <dbl>   <dbl>  <dbl>
## 1 Caspar    2006      -124.     39.4         NA        NA      NA      NA 0.0664
## 2 Caspar    2007      -124.     39.4         NA        NA      NA      NA 1.21
## 3 Caspar    2008      -124.     39.4       4.38      3.03   0.150   0.996 3.30
## 4 Caspar    2009      -124.     39.4         NA        NA      NA      NA 5.49
## 5 Caspar    2010      -124.     39.4       4.37      4.17  0.0664   0.586 0.273
## 6 Caspar    2011      -124.     39.4         NA        NA      NA      NA 1.15
## # ... with abbreviated variable names 1: mean_OUTER, 2: se_INNER, 3: se_OUTER
```

Plotting

Plot log of kelps density vs year for each site at different resolutions.

```
sites <- unique(kelp_data_300m$site_name)

kelp_longer_300m <- kelp_data_300m %>%
  dplyr::select(-c(longitude, latitude)) %>%
  pivot_longer(
    -c('site_name', 'year', 'fit'),
    names_to = c('.value', 'zone'),
    names_sep = '_'
    ) %>%
  mutate(resolution = as.factor(300))

kelp_longer_900m <- kelp_data_900m %>%
  dplyr::select(-c(longitude, latitude)) %>%
  pivot_longer(
    -c('site_name', 'year', 'fit'),
    names_to = c('.value', 'zone'),
    names_sep = '_'
    ) %>%
  mutate(resolution = as.factor(900))

kelp_longer_150m <- kelp_data_150m %>%
  dplyr::select(-c(longitude, latitude)) %>%
  pivot_longer(
    -c('site_name', 'year', 'fit'),
    names_to = c('.value', 'zone'),
    names_sep = '_'
    ) %>%
  mutate(resolution = as.factor(150))

kelp_longer_600m <- kelp_data_600m %>%
  dplyr::select(-c(longitude, latitude)) %>%
  pivot_longer(
    -c('site_name', 'year', 'fit'),
    names_to = c('.value', 'zone'),
    names_sep = '_'
    ) %>%
  mutate(resolution = as.factor(600))

kelp_longer <- rbind(kelp_longer_150m, kelp_longer_300m,
                     kelp_longer_600m, kelp_longer_900m)

for (i in sites) {
  plot <- kelp_longer %>%
    filter(site_name == i) %>%
    ggplot() +
    geom_pointrange(aes(
      x = year, y = mean, group = zone, color = zone,
      ymin = mean - se, ymax = mean + se
      ), alpha = 0.5, size = 0.3) +
    geom_bar(aes(x = year, y = fit,
                 fill = ifelse(!is.na(fit) & fit >= 6.6, 'YES', 'NO')),
             stat = 'identity', position = 'dodge', alpha = 0.2) +
    facet_wrap(. ~ resolution, nrow = 2) +
```

```
    theme_bw() +
    theme(axis.text.x = element_text(angle = 90, size = 8),
          plot.title = element_text(hjust = 0.5),
          panel.grid.major = element_blank(),
          legend.title = element_text(size = 9),
          legend.text = element_text(size = 7)) +
    labs(y = 'log of kelp density', title = i, fill = 'fit >= 6.6')
  print(plot)
}
```

Caspar
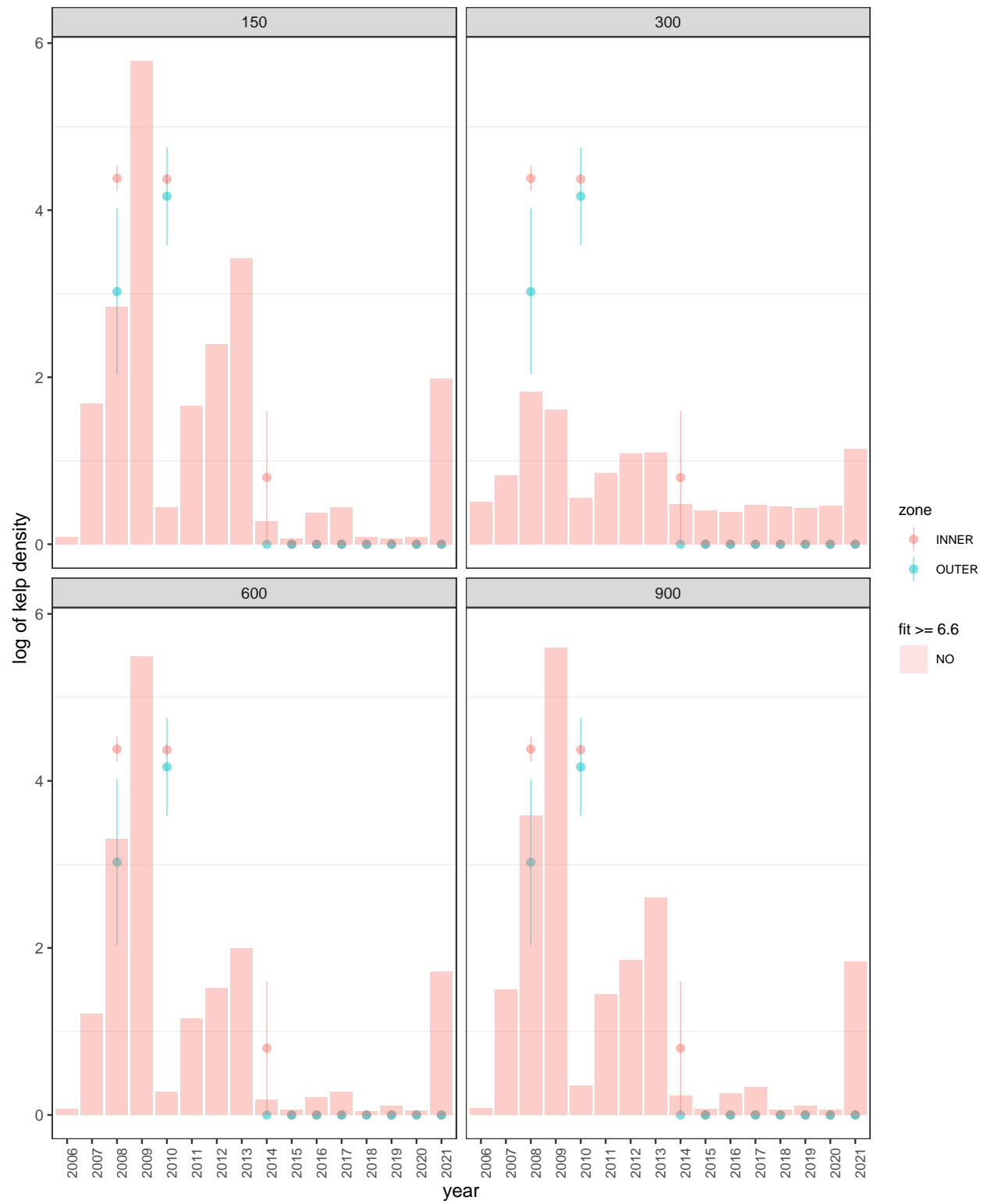
Caspar North
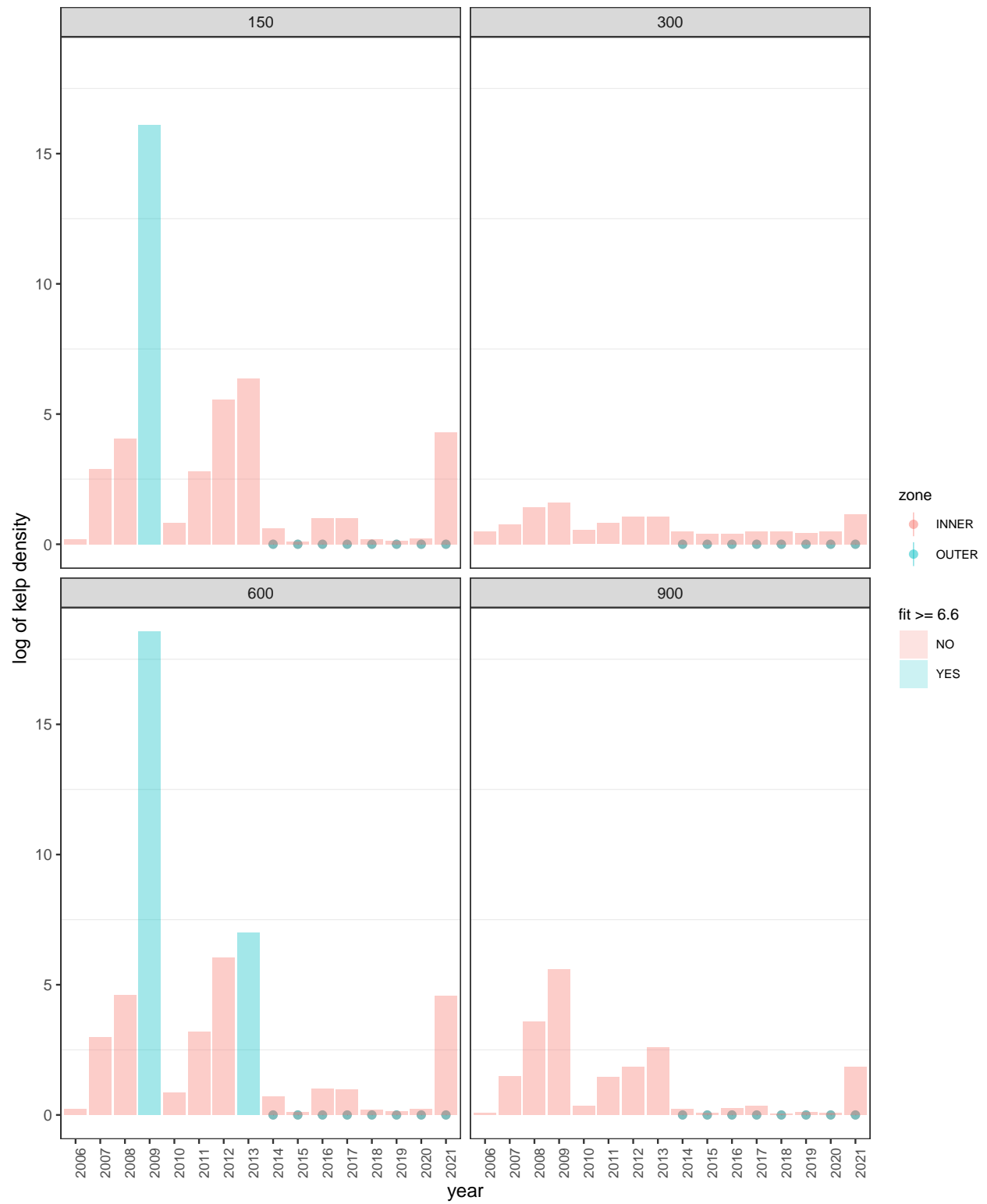
Dark Gulch
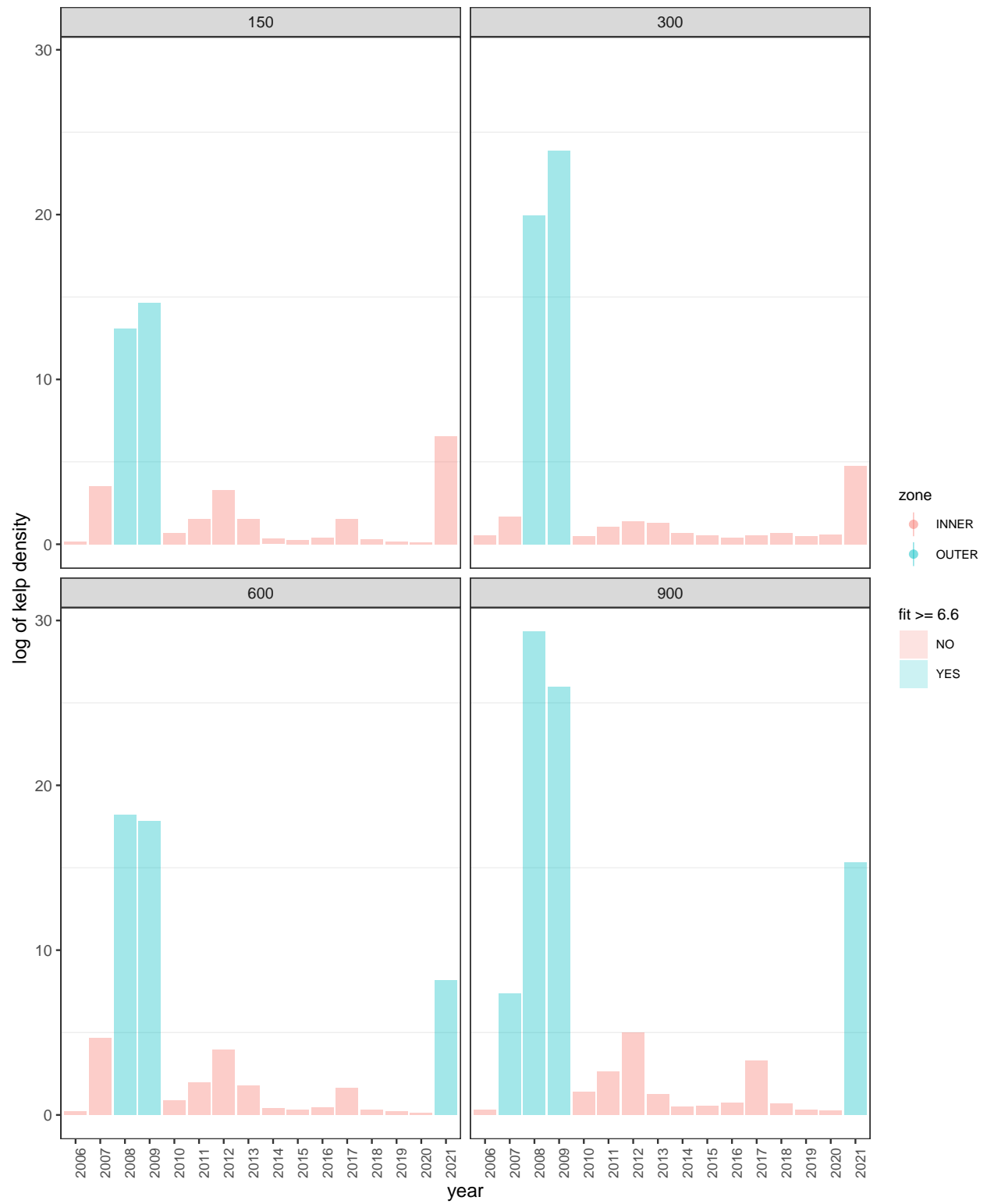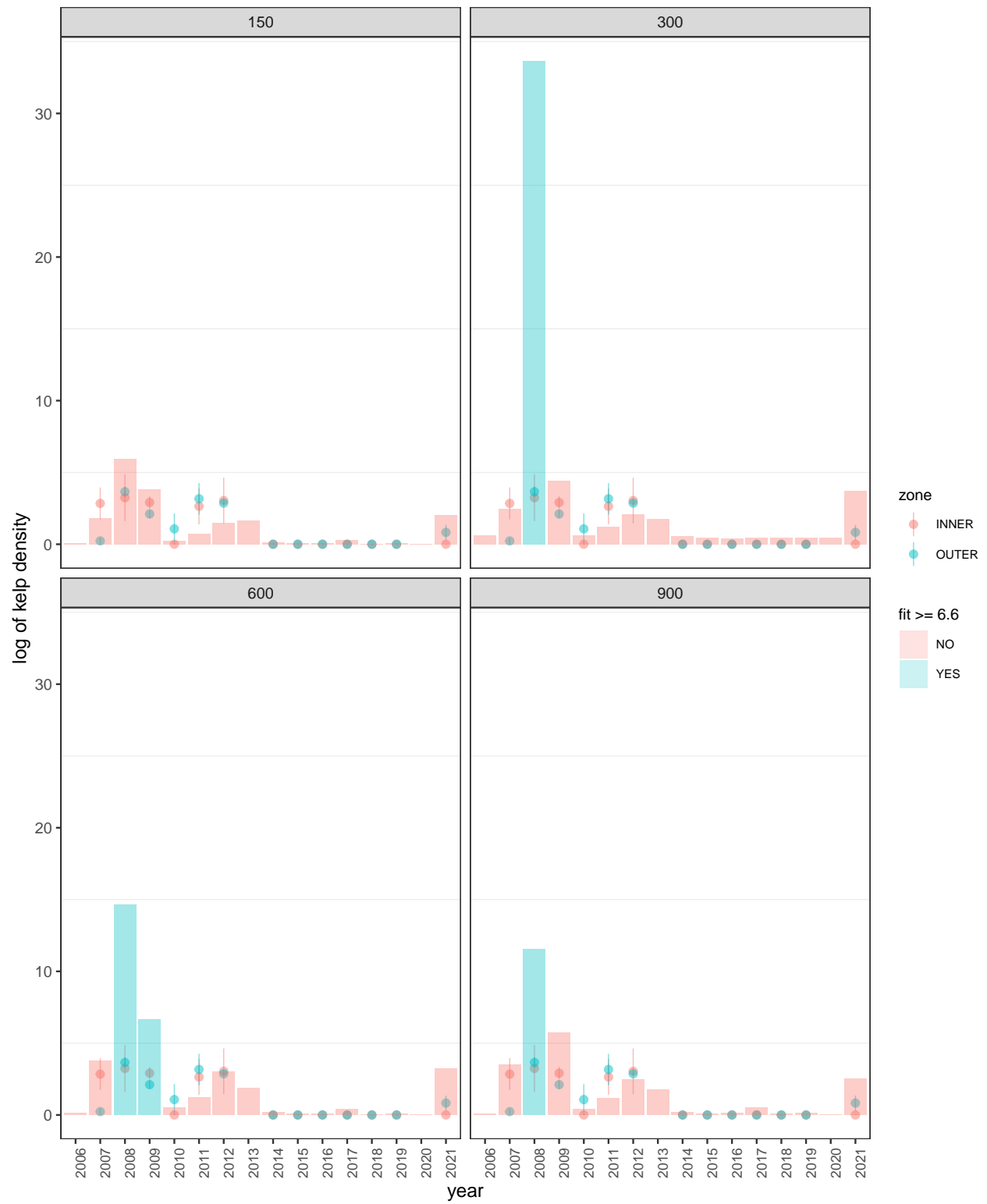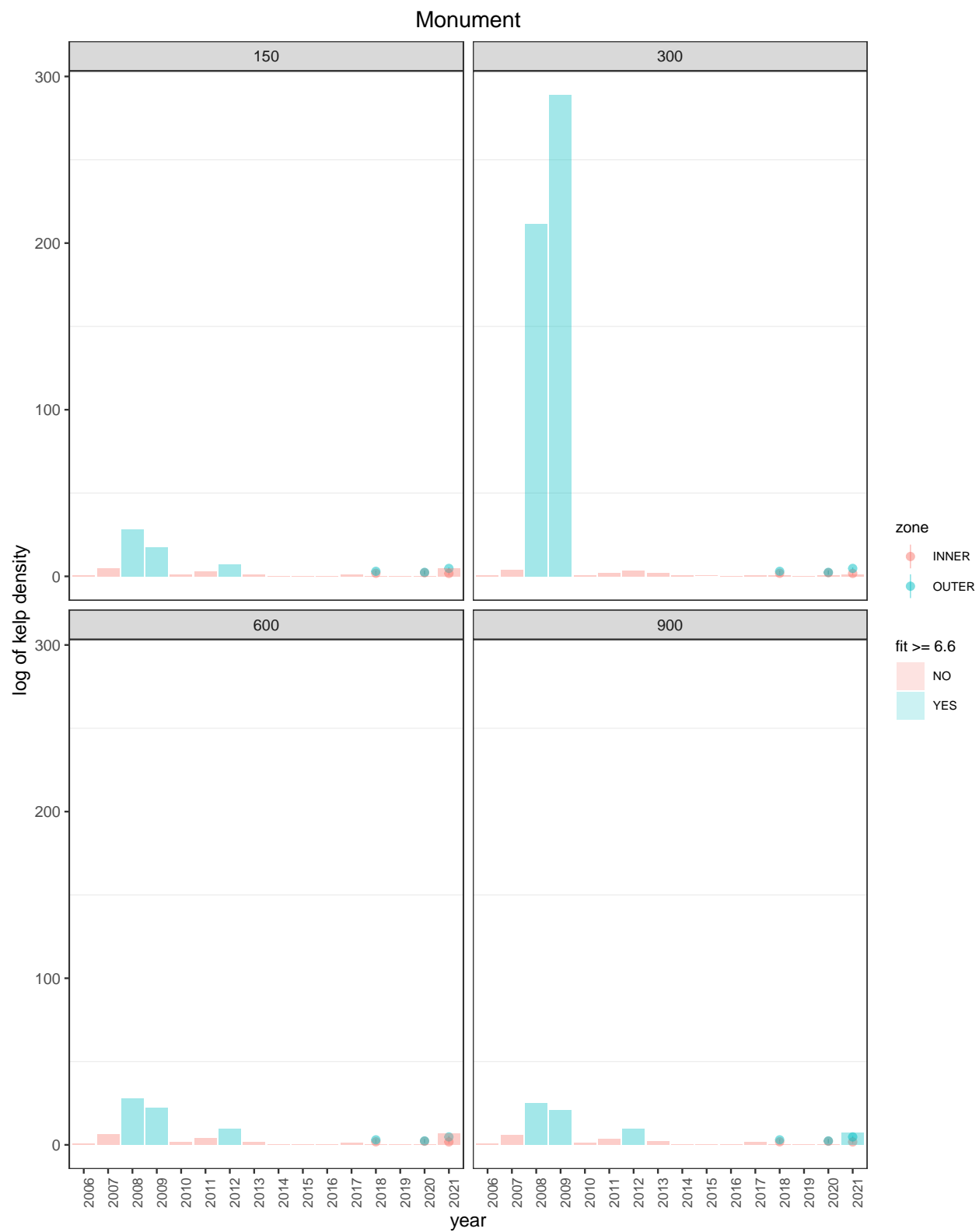
Flat Iron Rock
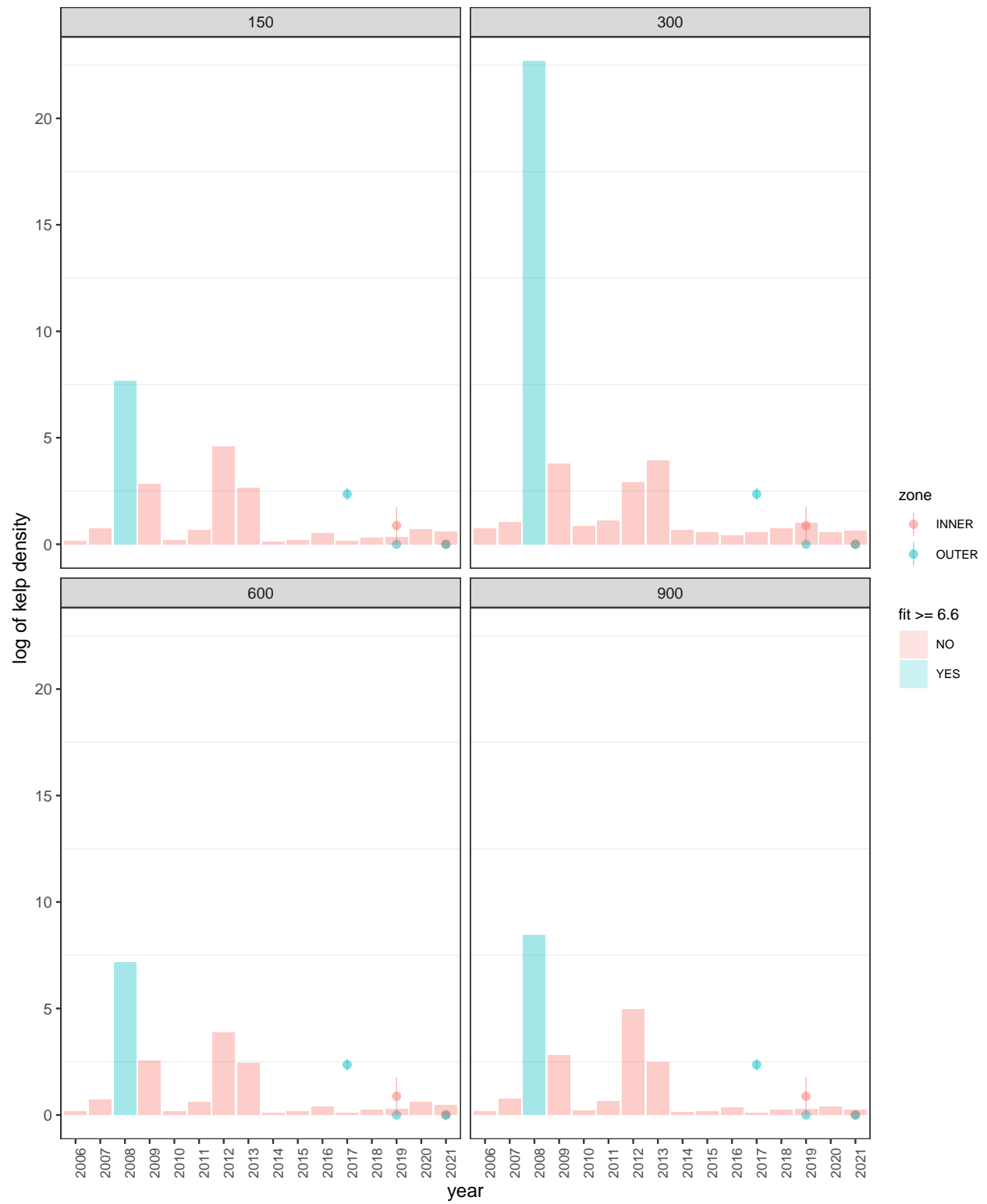
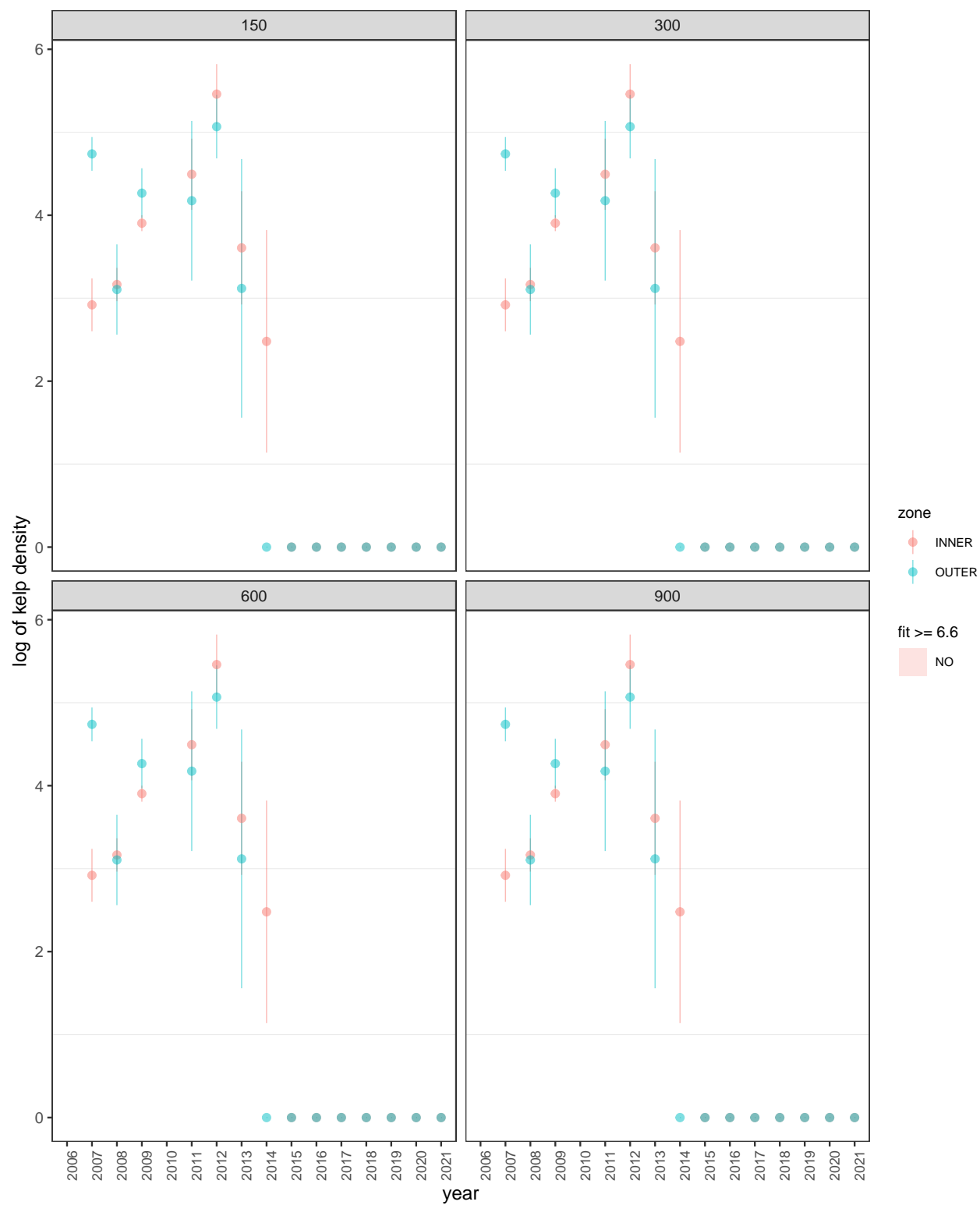Fort Ross
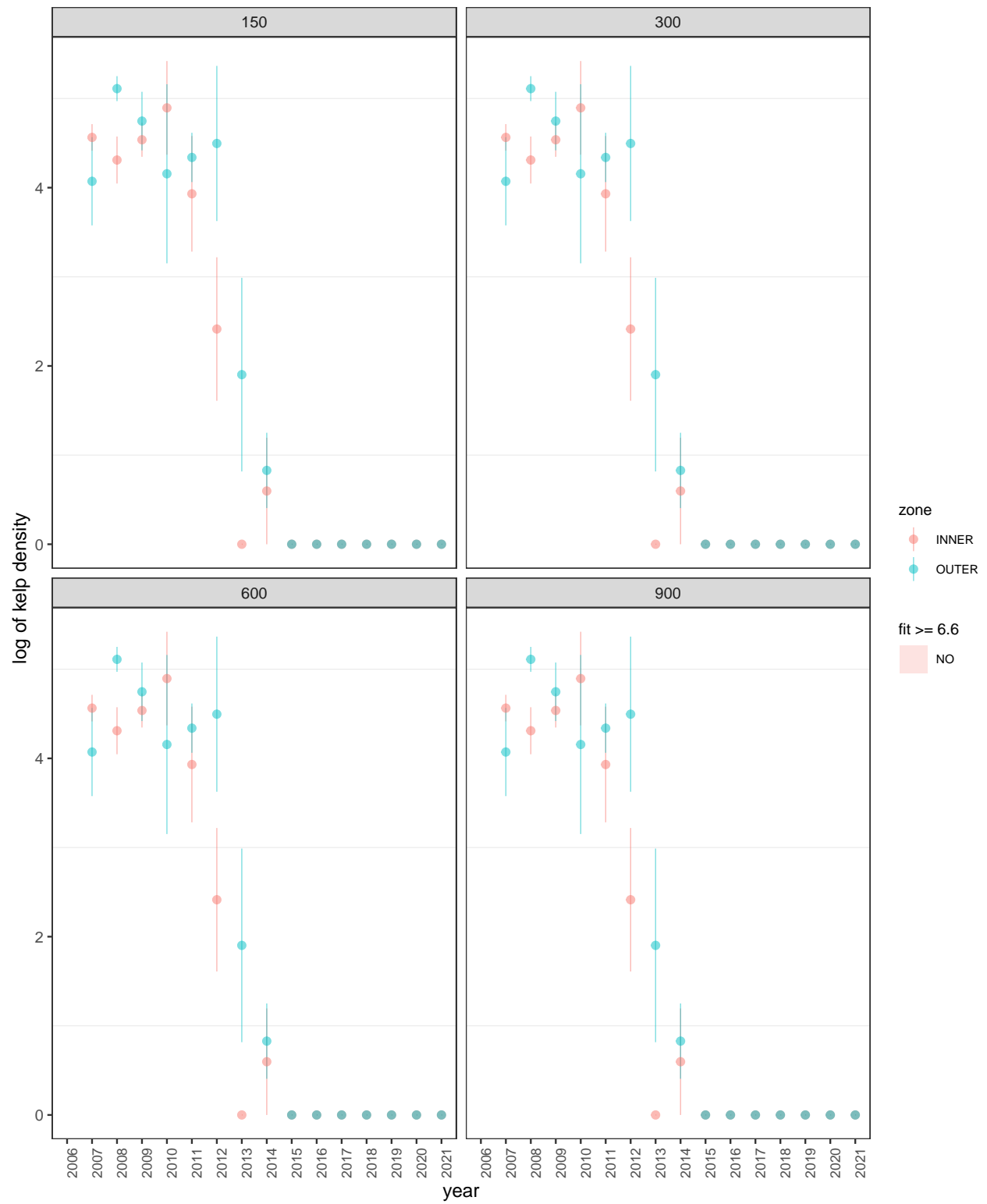
Frolic Cove
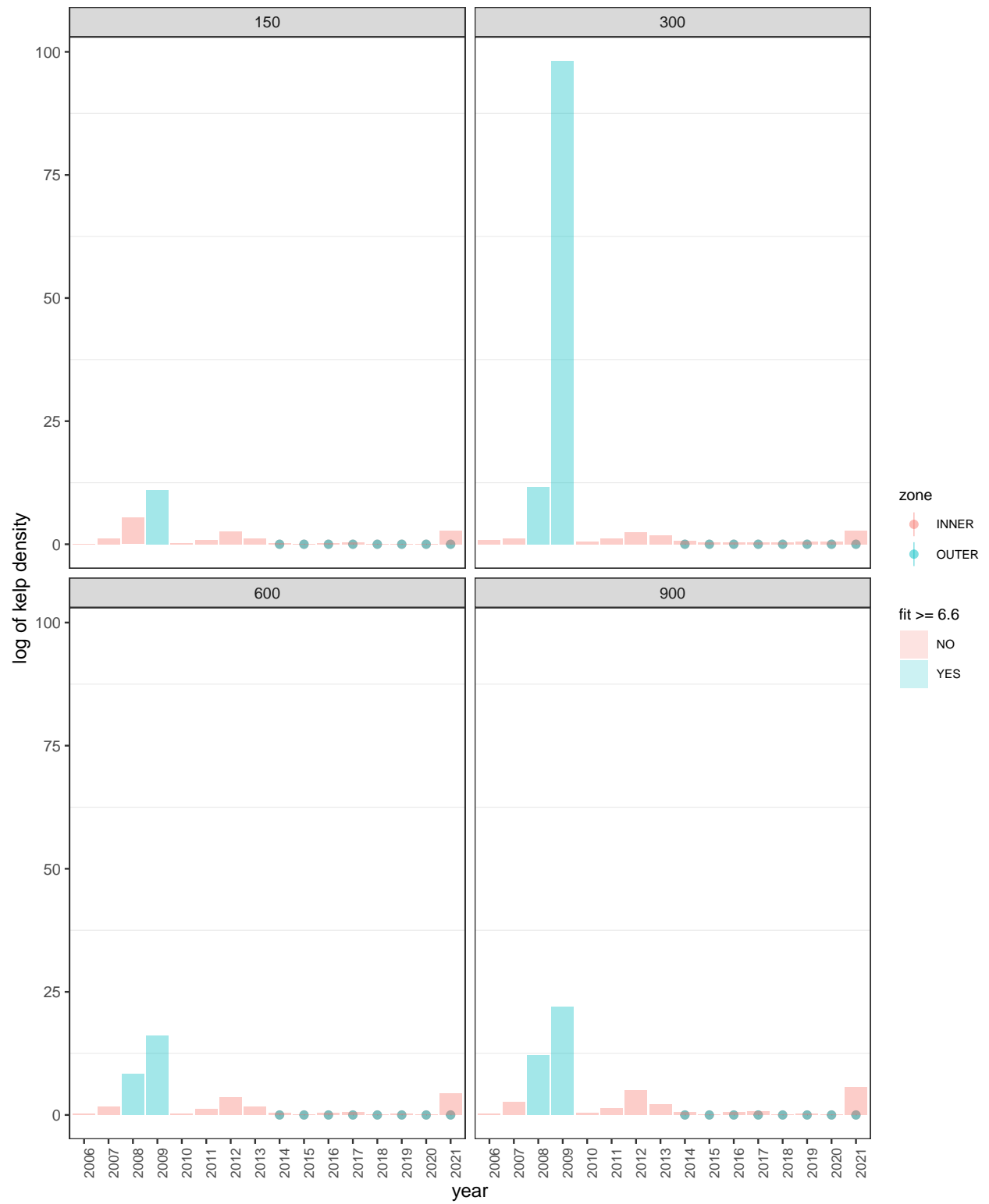
Gerstle Cove
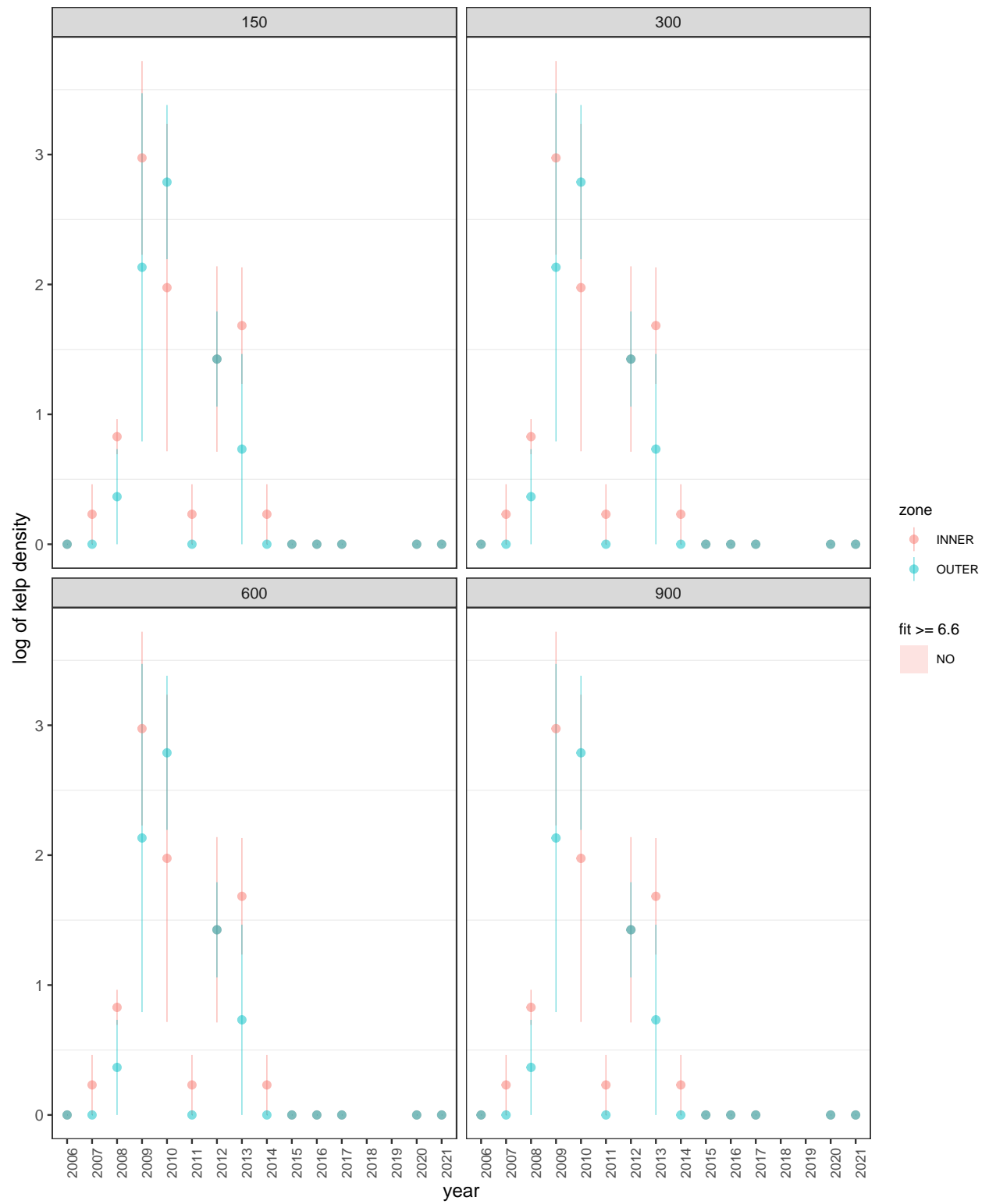
Glass Beach
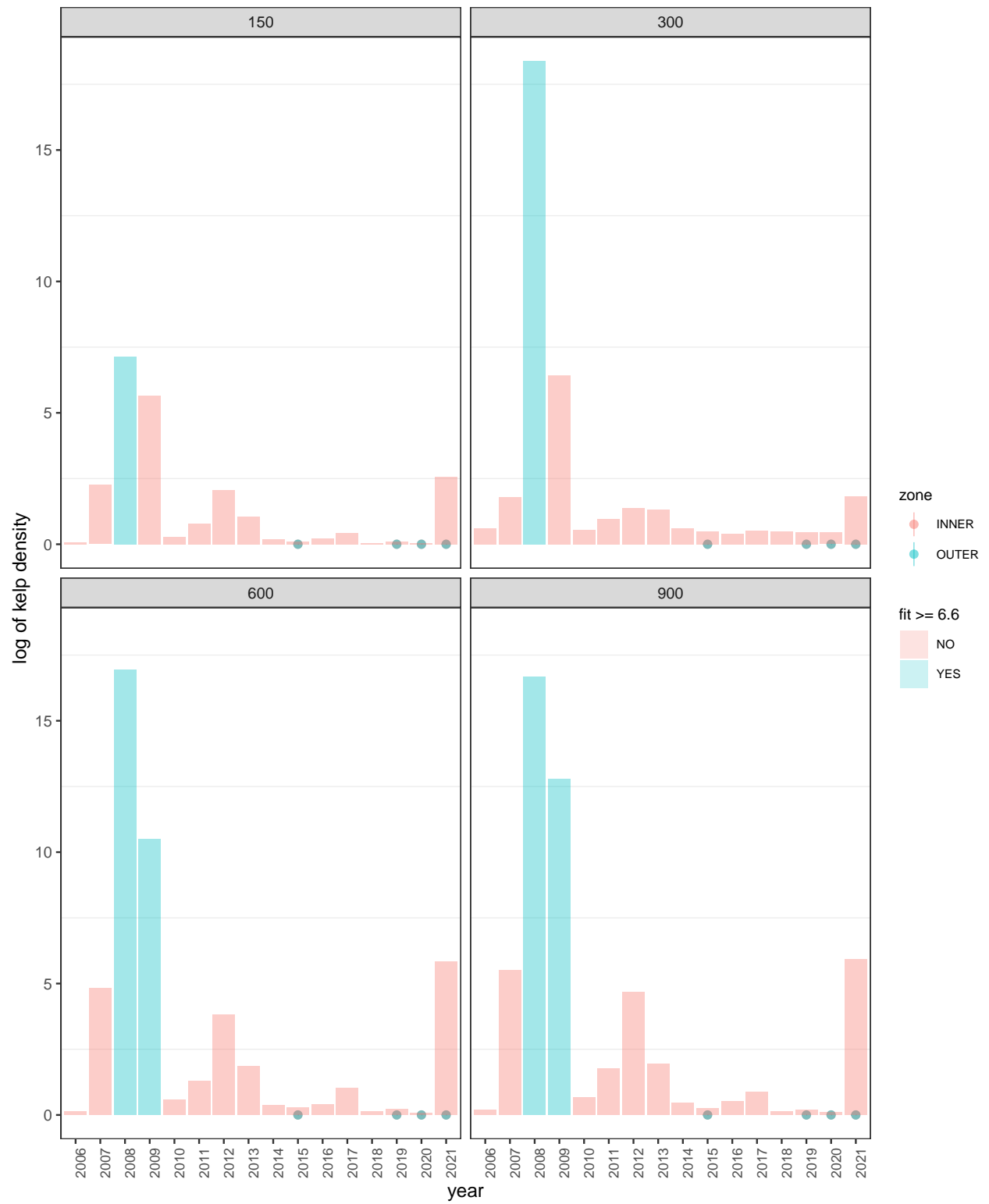
MacKerricher North
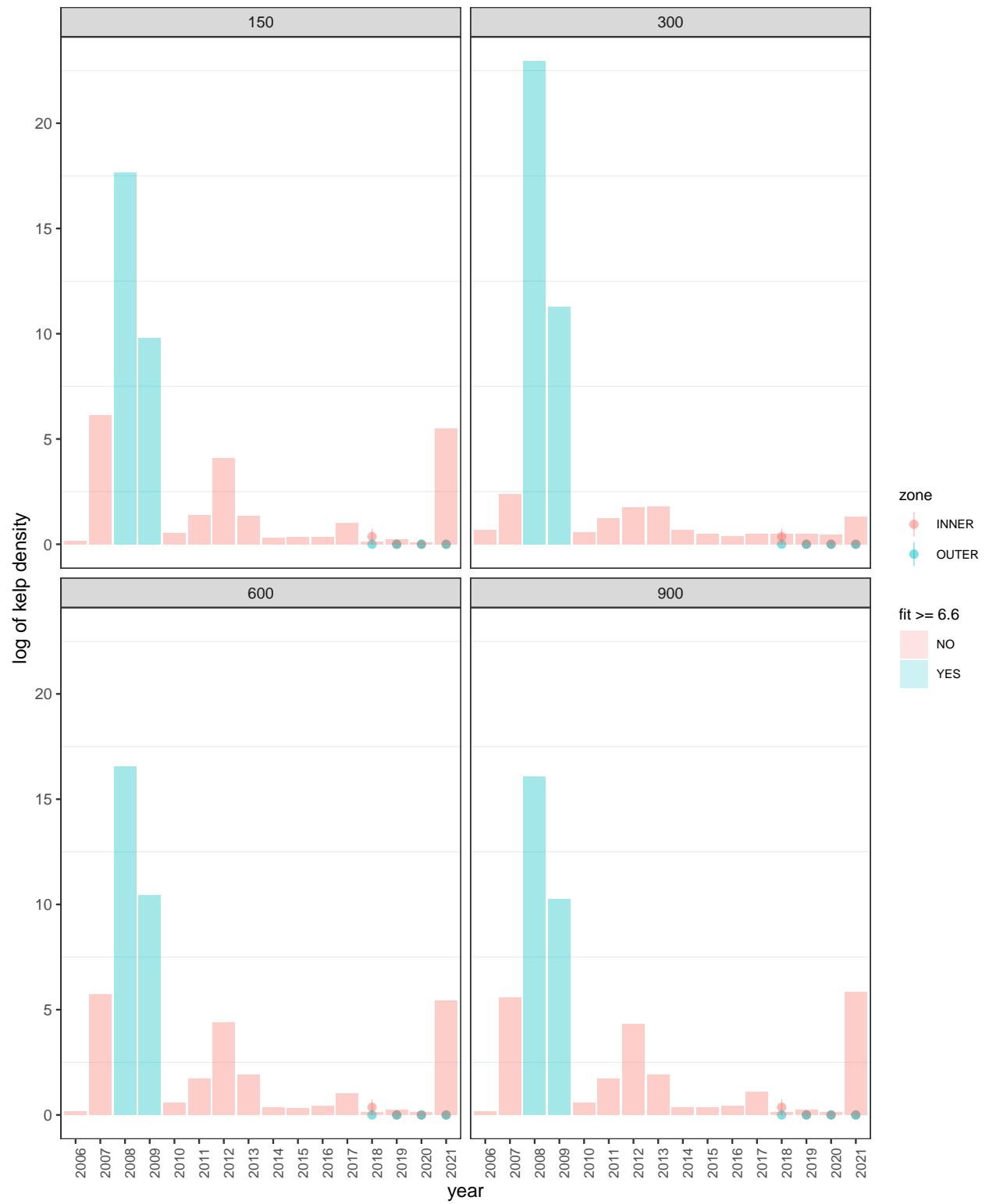
Mendocino Headlands

Monument

Ocean Cove

Pebble Beach
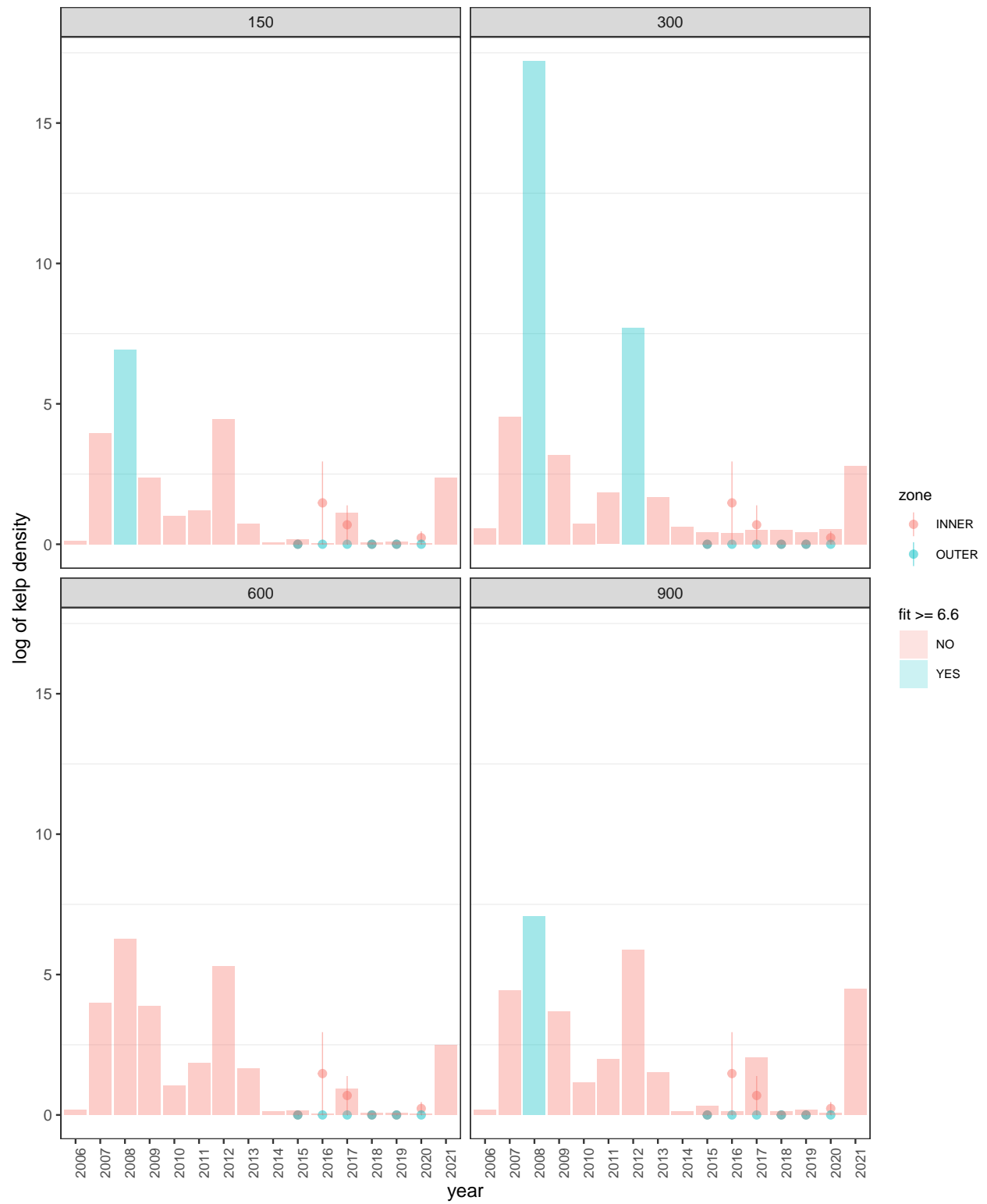
Point Arena Lighthouse
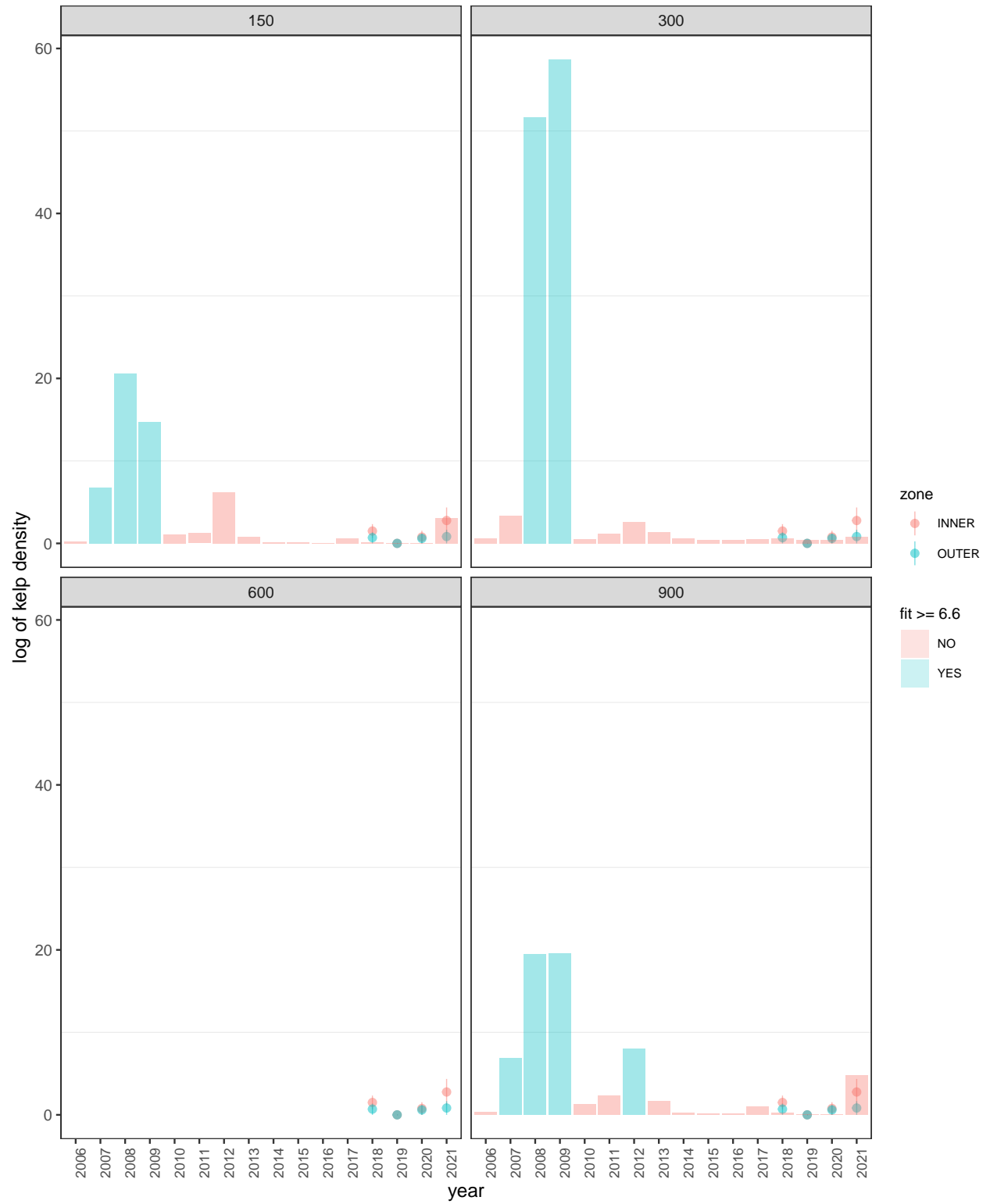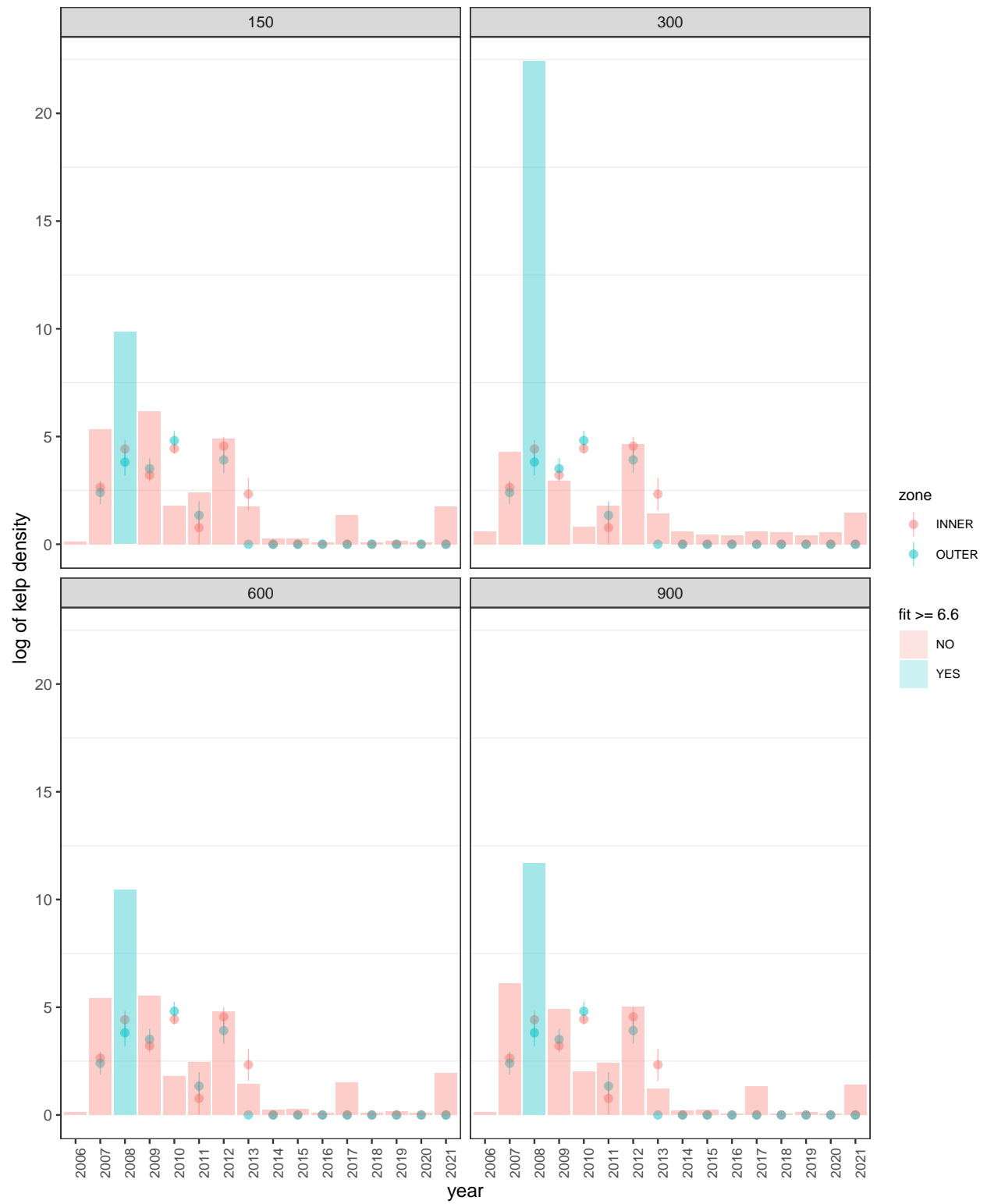
Point Arena MPA (M2)
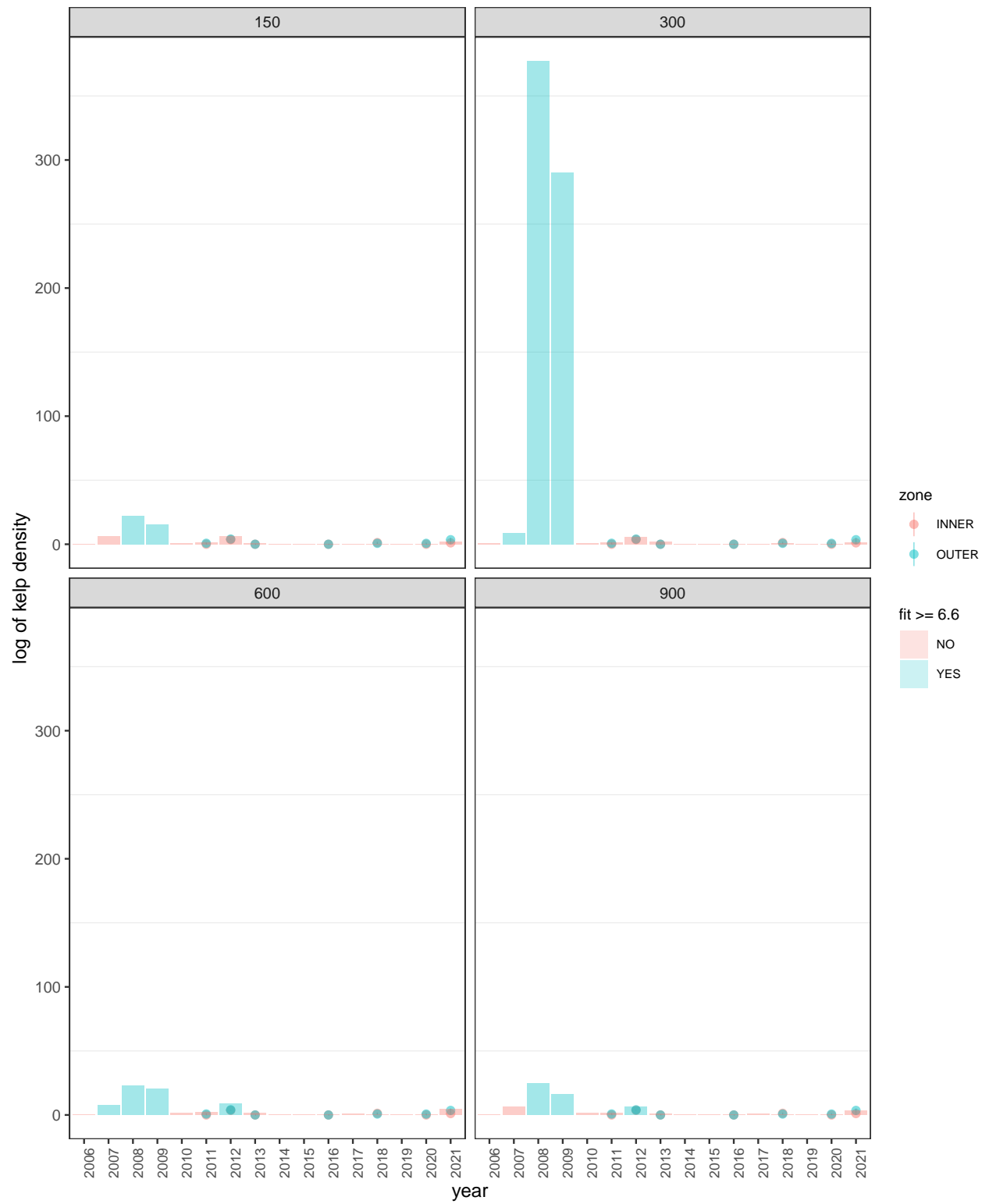
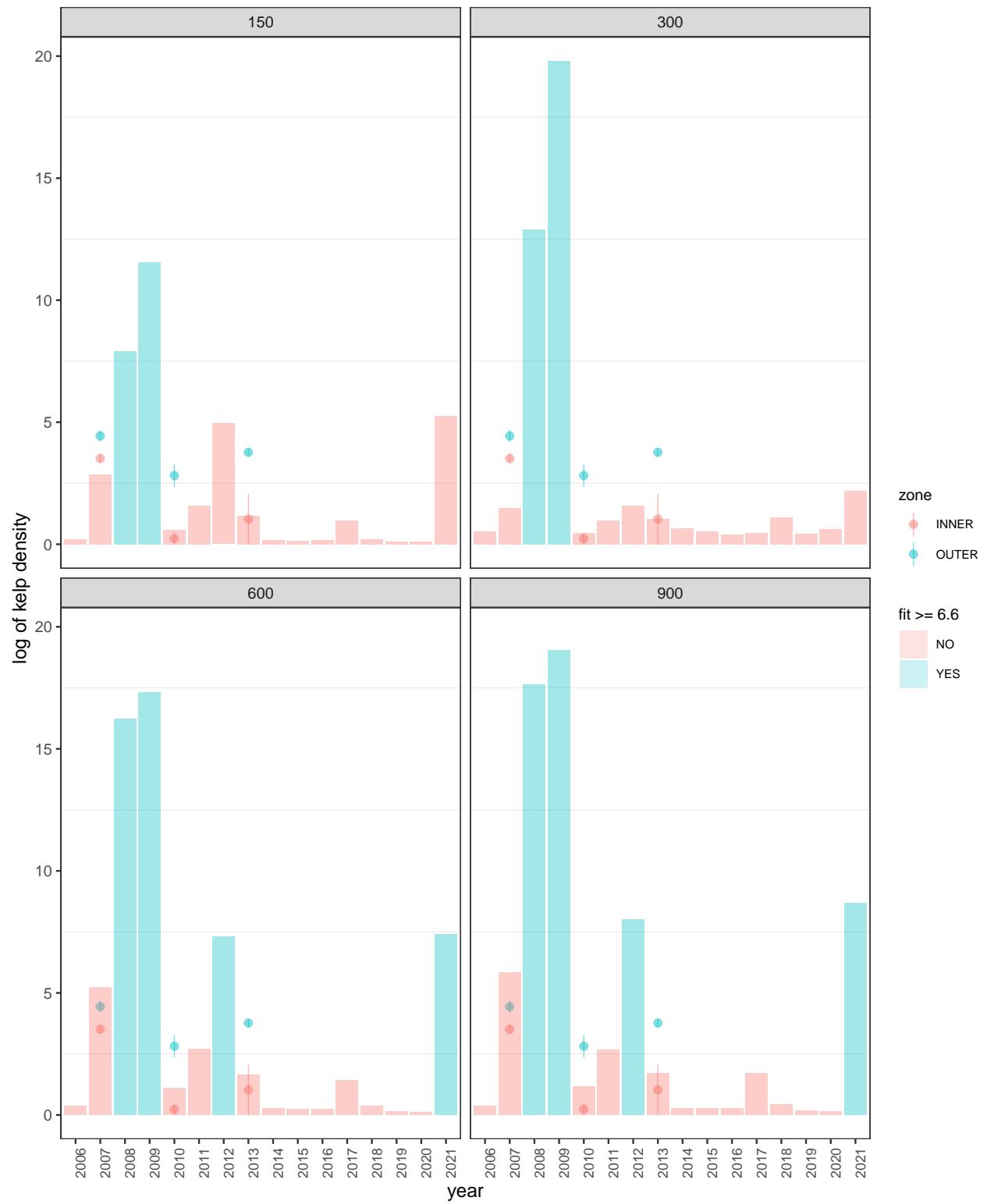Point Arena Ref
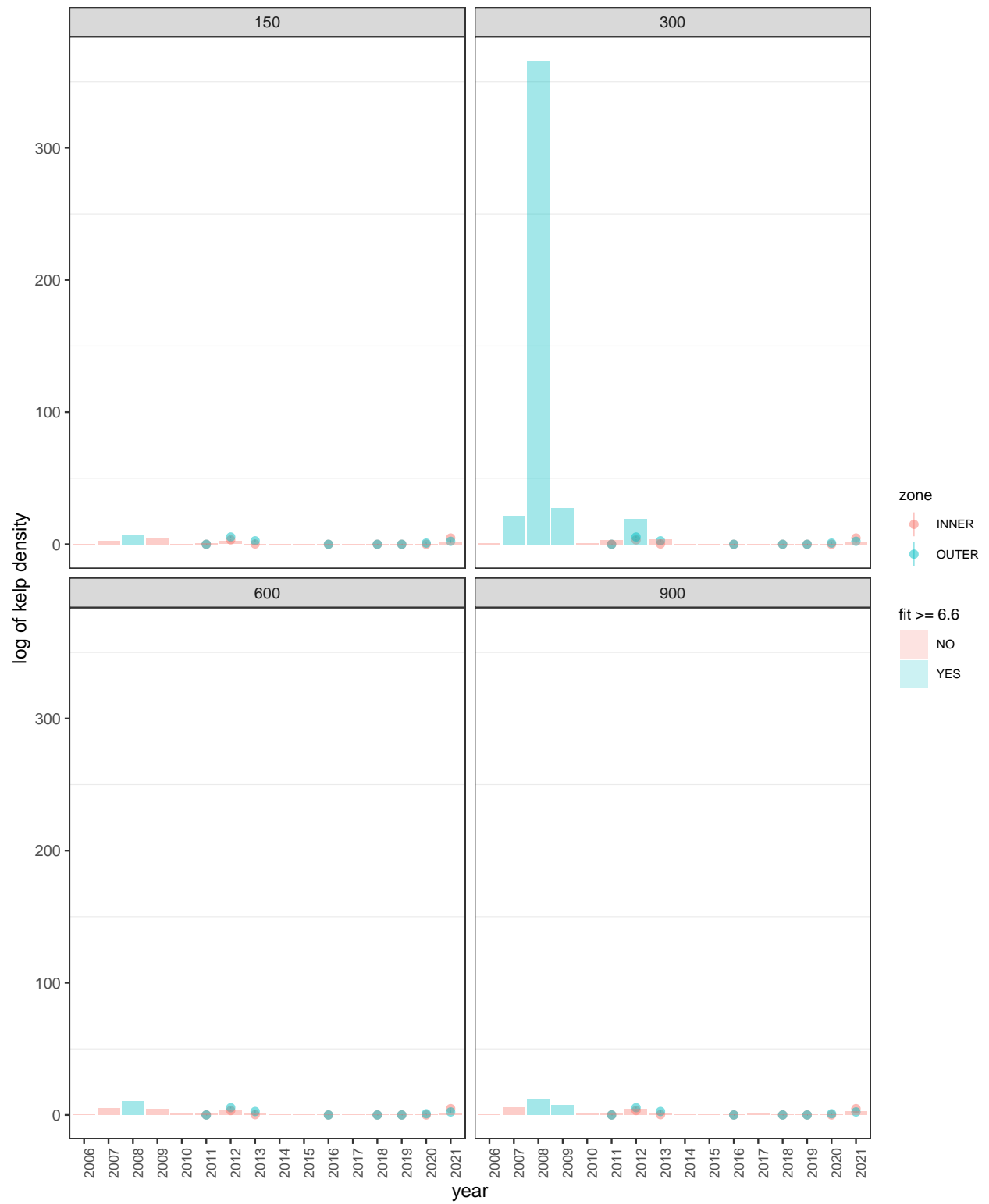
Portuguese Beach
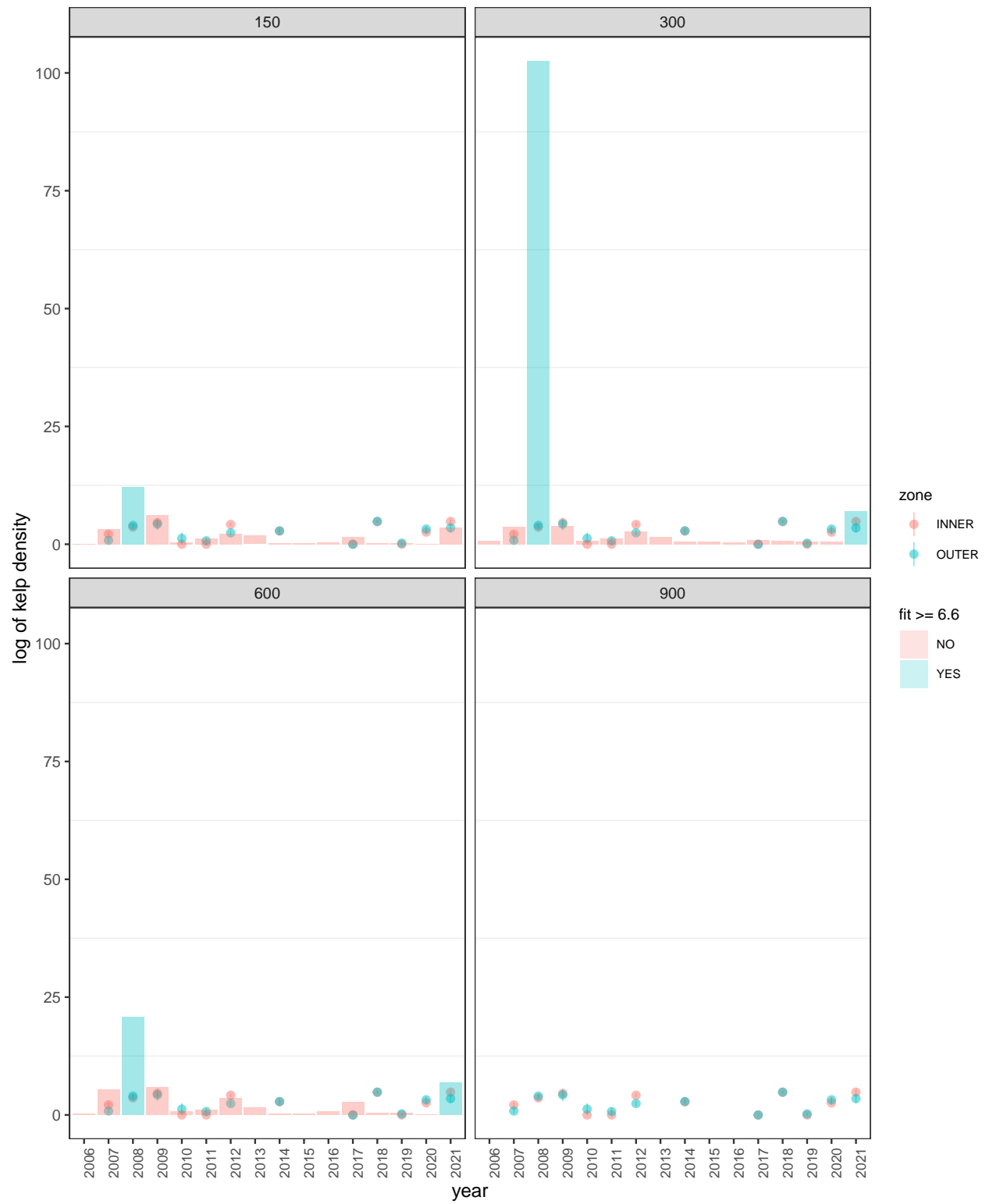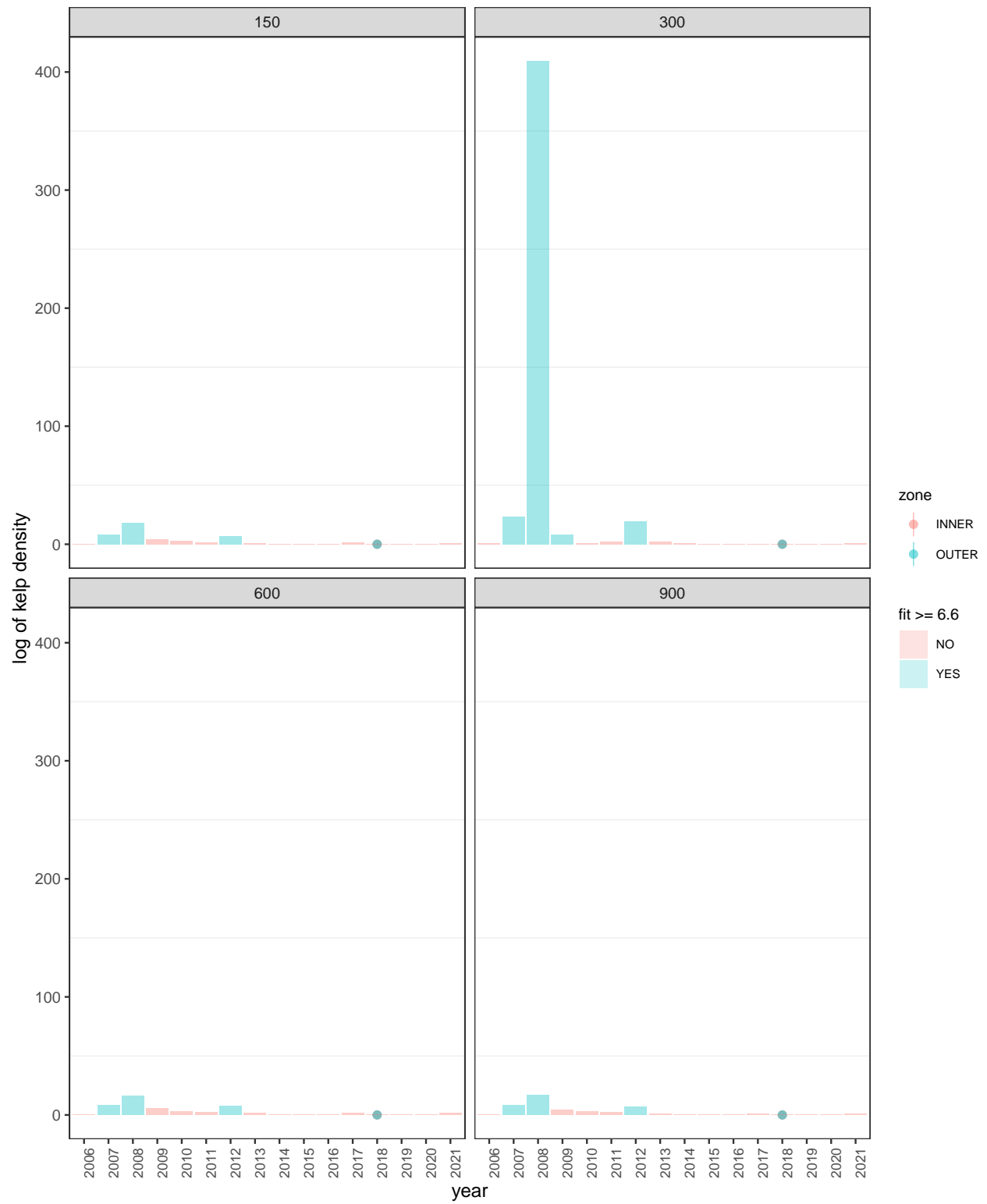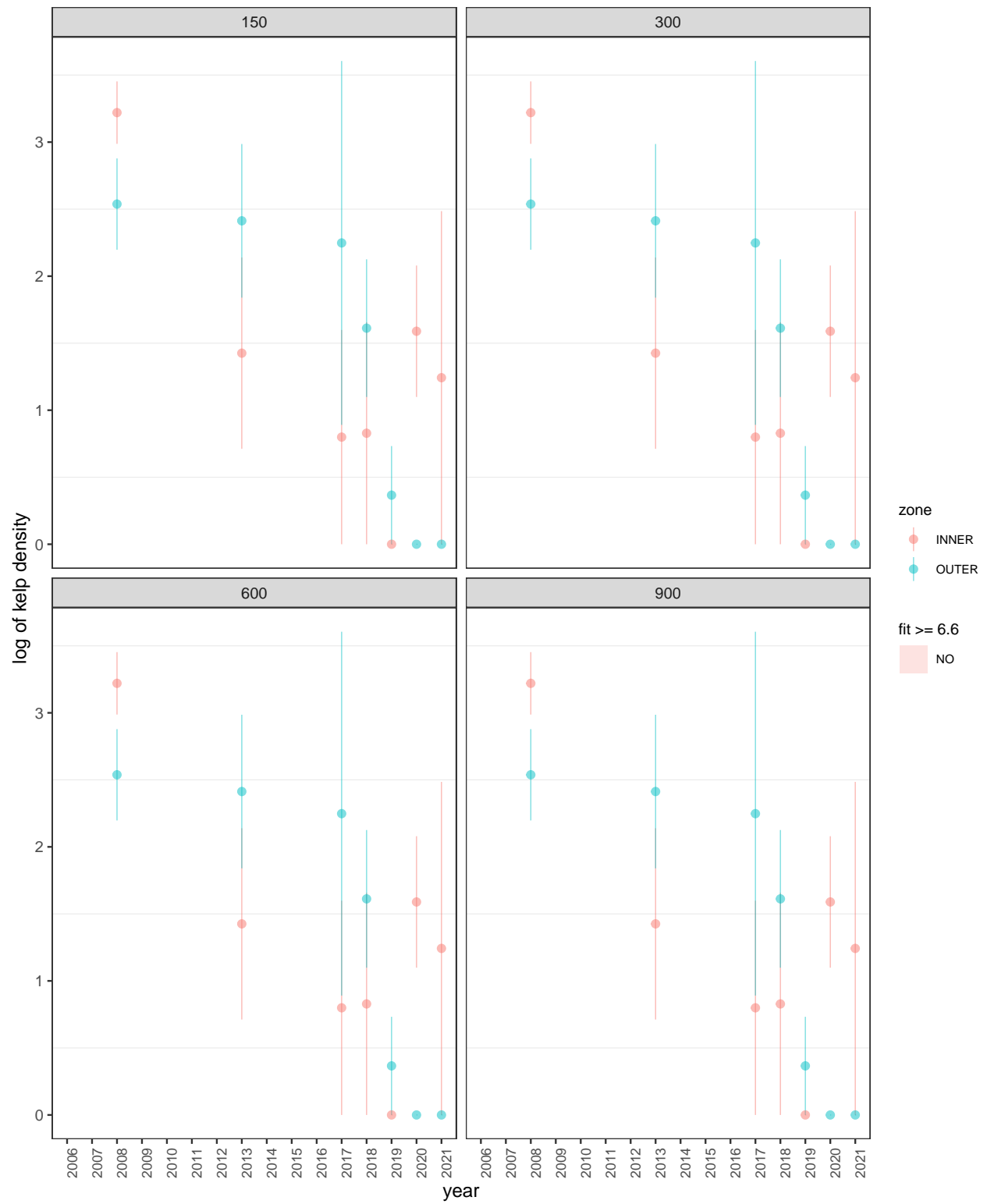
Pyramid Point

Russian Gulch

Salt Point
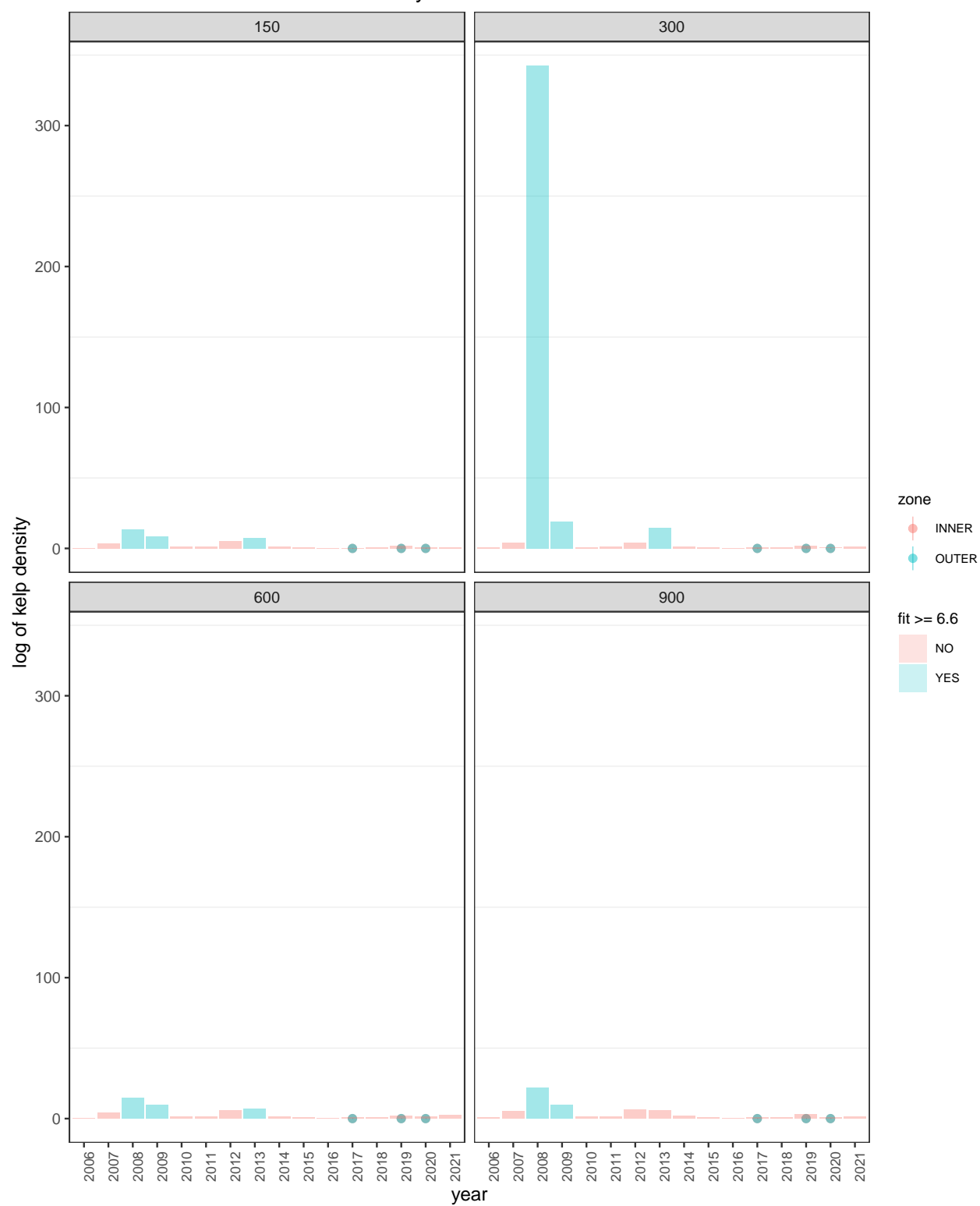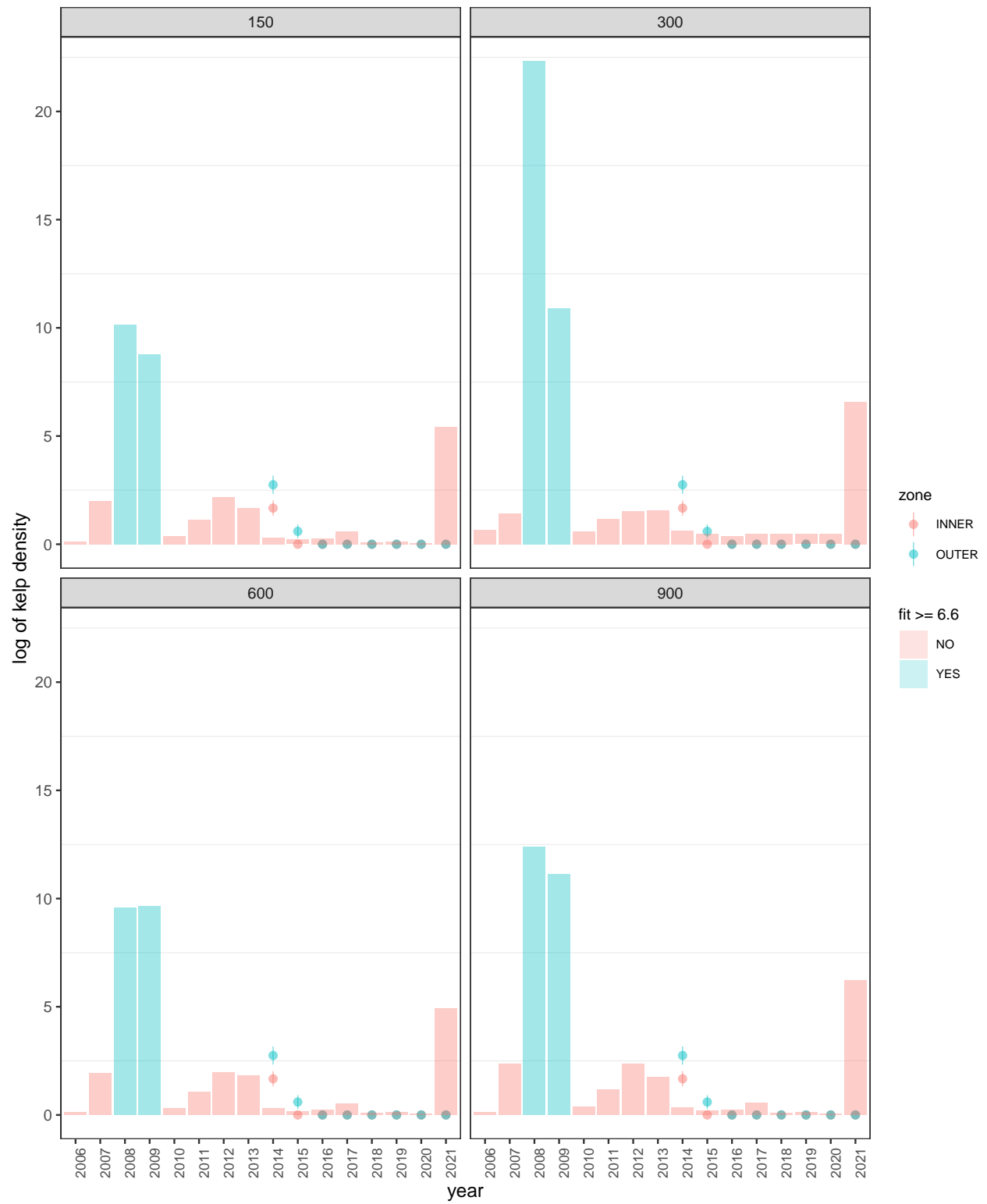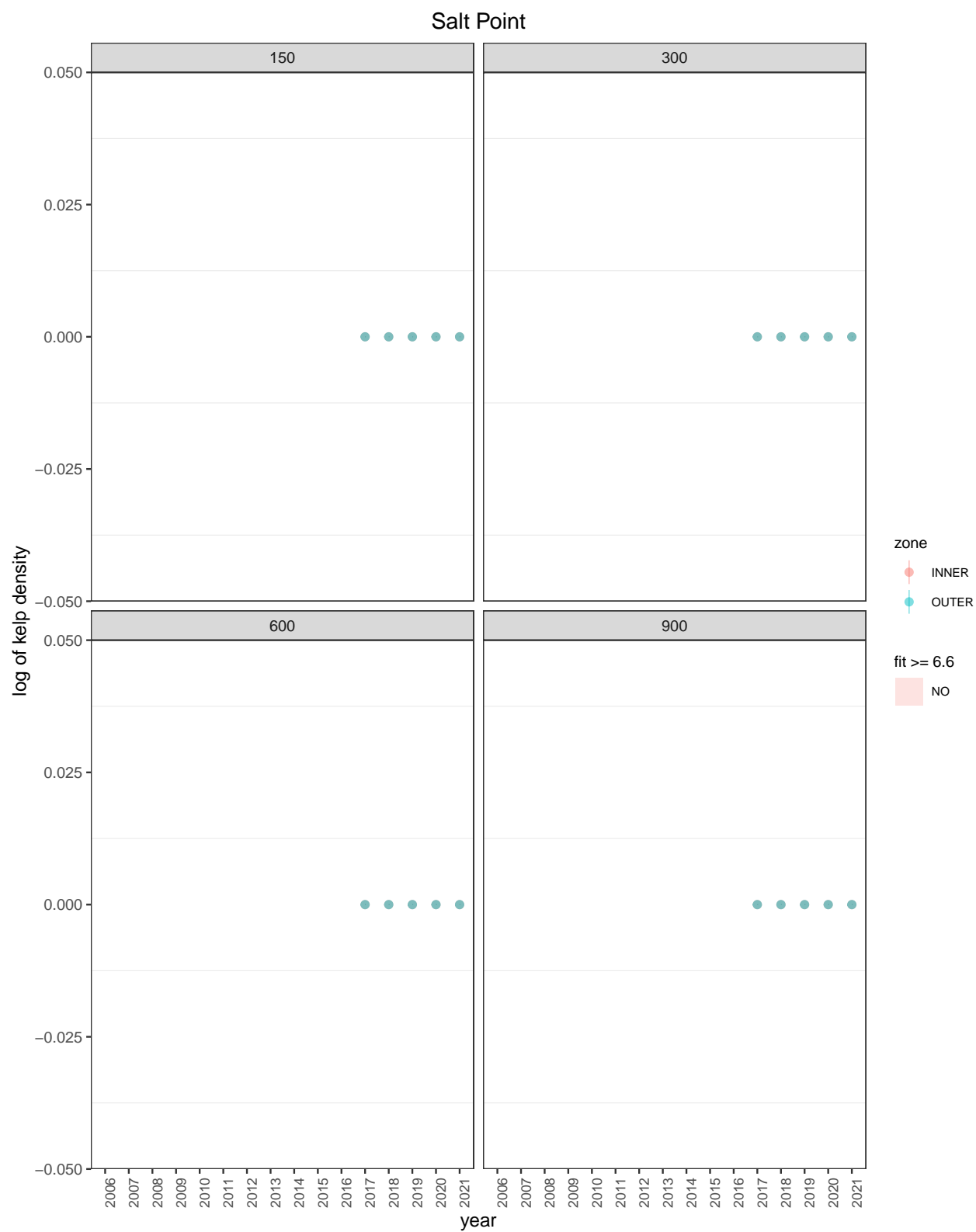
Stillwater Sonoma

Stornetta

Timber Cove

Trinidad

Van Damme