SURE Project

August 23, 2022

Objective

Compare spatial predictions of kelp to "in situ" survey data. Compare each year and location.

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')
# rock.dir <- here('spatial_data/sp_predictions_5.1.1_V2_rock')

# 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.

```
filter(df, den_NERLUE == 0) %>% count() # 719 0's
```

```
##     n
## 1 719
```

How to deal with log(0)?

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

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

```
# 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')
# write the file
st_write(site_shp, paste0(d.dir, '/RCCA_North_Coast_sites.shp'), append = FALSE)
```

```
## Deleting layer 'RCCA_North_Coast_sites' using driver 'ESRI Shapefile'
## Writing layer 'RCCA_North_Coast_sites' to data source
##   '/Users/chuntingzheng/Desktop/Git_Repositories/Chunting_Spatial_Analyses/data/RCCA_North_Coast_site
## Writing 25 features with 5 fields and geometry type Point.
```

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

```
for (i in c(2006:2021)) {
  rast <- rast(paste0(r.dir, paste0('/', i, '_Nereo_preds_NC_V4_5.1.1_V2.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 0.5090384
## 2   Caspar North 2006 0.5003366
## 3     Dark Gulch 2006 0.5309656
## 4 Flat Iron Rock 2006 0.7382968
## 5      Fort Ross 2006       NaN
## 6    Frolic Cove 2006 0.8228083
```

```
# sites and years have predicted values higher than 6.6 (log of density)
filter(pred, fit >= 6.6)
```

```
##                 site_name year        fit
## 1    Point Arena MPA (M2) 2007   8.662991
## 2         Point Arena Ref 2007  21.406689
## 3             Timber Cove 2007  23.488674
## 4              Dark Gulch 2008  19.940689
## 5          Flat Iron Rock 2008  22.692287
## 6             Frolic Cove 2008  11.673522
## 7             Glass Beach 2008  18.386805
## 8      MacKerricher North 2008  22.953793
## 9     Mendocino Headlands 2008  33.649105
## 10               Monument 2008 211.603195
## 11           Pebble Beach 2008  17.207937
## 12 Point Arena Lighthouse 2008  51.701321
## 13   Point Arena MPA (M2) 2008 377.264008
## 14        Point Arena Ref 2008 365.663879
## 15       Portuguese Beach 2008 102.522873
## 16          Pyramid Point 2008 342.526245
## 17          Russian Gulch 2008  22.330360
## 18      Stillwater Sonoma 2008  22.428894
## 19              Stornetta 2008  12.890643
## 20            Timber Cove 2008 409.307465
## 21              Van Damme 2008   9.219850
## 22             Dark Gulch 2009  23.877867
## 23            Frolic Cove 2009  98.134544
## 24     MacKerricher North 2009  11.300843
## 25               Monument 2009 288.878967
## 26 Point Arena Lighthouse 2009  58.657776
## 27   Point Arena MPA (M2) 2009 290.192169
## 28        Point Arena Ref 2009  27.217886
## 29          Pyramid Point 2009  19.214815
## 30          Russian Gulch 2009  10.897983
```

```
## 31              Stornetta 2009  19.797838
## 32           Timber Cove 2009   8.495423
## 33          Pebble Beach 2012   7.714611
## 34       Point Arena Ref 2012  19.095858
## 35           Timber Cove 2012  19.386978
## 36         Pyramid Point 2013  14.741710
## 37      Portuguese Beach 2021   7.111959
```

```
# write to cvs
merge_df <- left_join(pred,
                      site %>% select(c(site_name, longitude, latitude)),
                      by = 'site_name')
write.csv(merge_df, file.path(d.dir, 'NC_kelp_density_predictions.csv'), row.names = FALSE)
```

Comparison

Compare the predicted data to observed data.

```
dim(obs)
```

```
## [1] 193   6
```

```
dim(pred)
```

```
## [1] 400   3
```

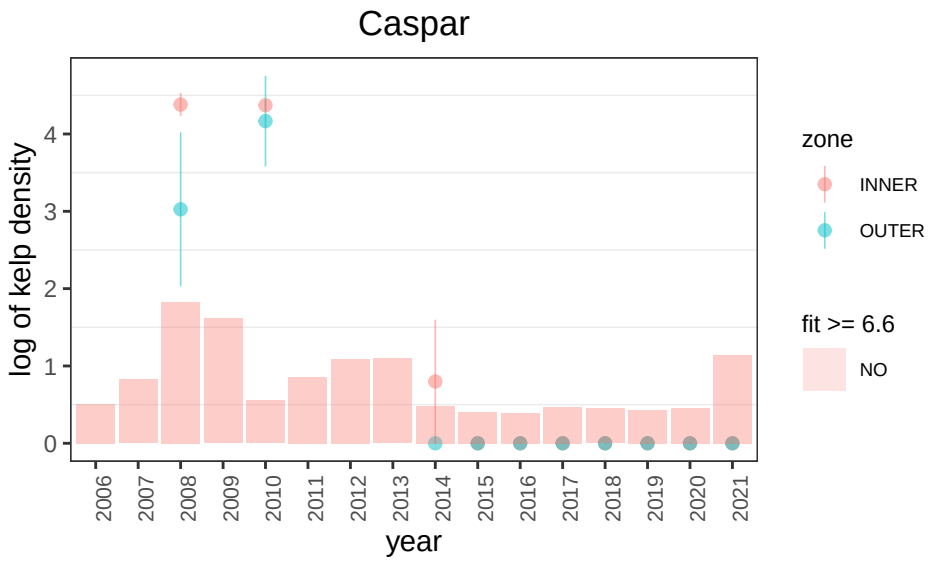
```
kelp_data <- left_join(pred, obs, by = c('site_name', 'year')) %>%
  group_by(site_name) %>%
  arrange(year, .by_group = TRUE) %>%
  relocate(fit, .after = last_col())
head(kelp_data)
```

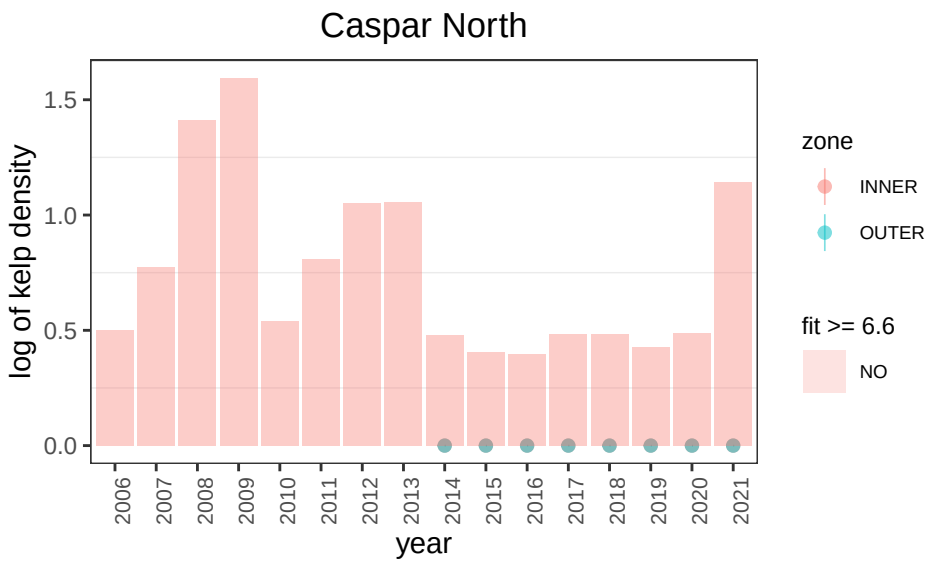
```
## # A tibble: 6 x 7
## # Groups:   site_name [1]
##   site_name year  mean_INNER mean_OUTER se_INNER se_OUTER   fit
##   <chr>     <fct>      <dbl>      <dbl>    <dbl>    <dbl> <dbl>
## 1 Caspar    2006       NA         NA     NA         NA     0.509
## 2 Caspar    2007       NA         NA     NA         NA     0.827
## 3 Caspar    2008        4.38       3.03   0.150     0.996 1.83
## 4 Caspar    2009       NA         NA     NA         NA     1.62
## 5 Caspar    2010        4.37       4.17   0.0664    0.586 0.554
## 6 Caspar    2011       NA         NA     NA         NA     0.857
```

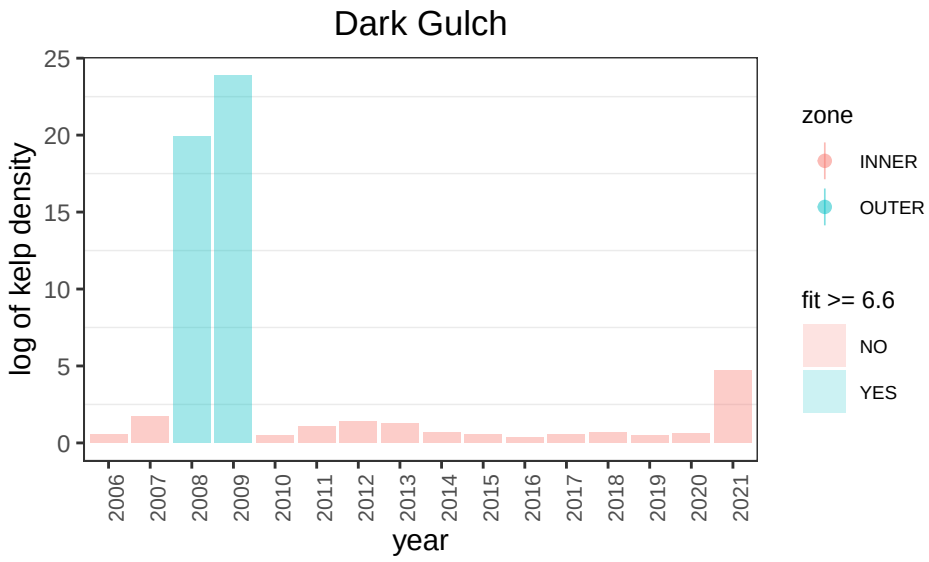
Plotting

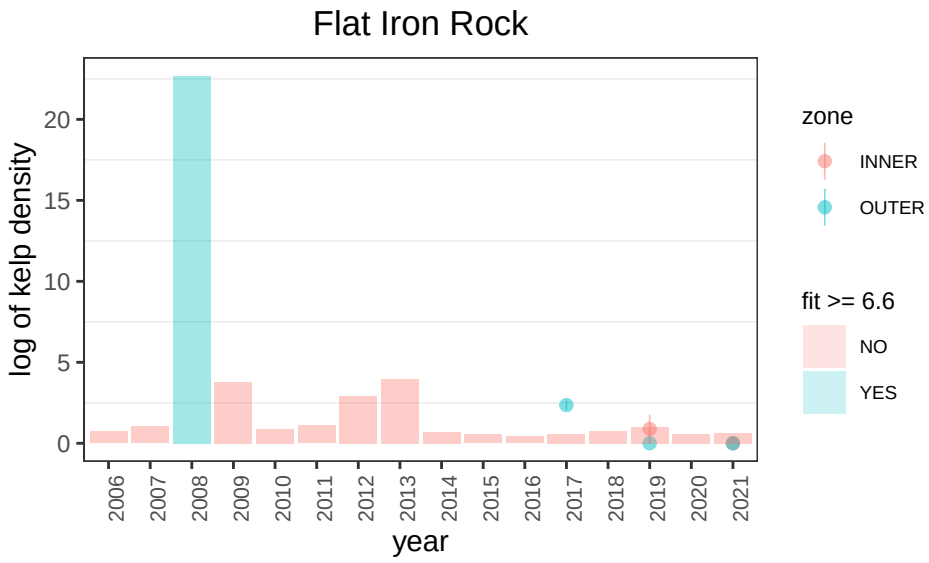
Plot log of kelps density vs year for each site.

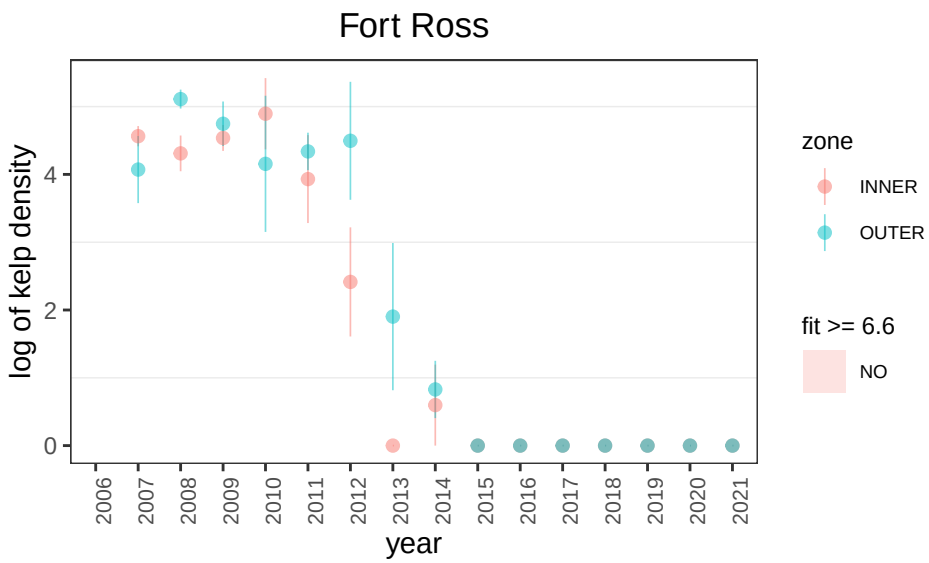
```r
# kelp_data %>%
#   pivot_longer(
#     -c('site_name', 'year', 'fit'),
#     names_to = c('.value', 'zone'),
#     names_sep = '_'
#     ) %>%
#   filter(site_name == 'Caspar') %>%
#   ggplot(aes(x = year, y = mean, fill = zone)) +
#   geom_bar(position = 'dodge', stat = 'identity')

sites <- unique(kelp_data$site_name)
kelp_longer <- kelp_data %>%
    pivot_longer(
      -c('site_name', 'year', 'fit'),
      names_to = c('.value', 'zone'),
      names_sep = '_'
      )

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) +
    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)
}
```

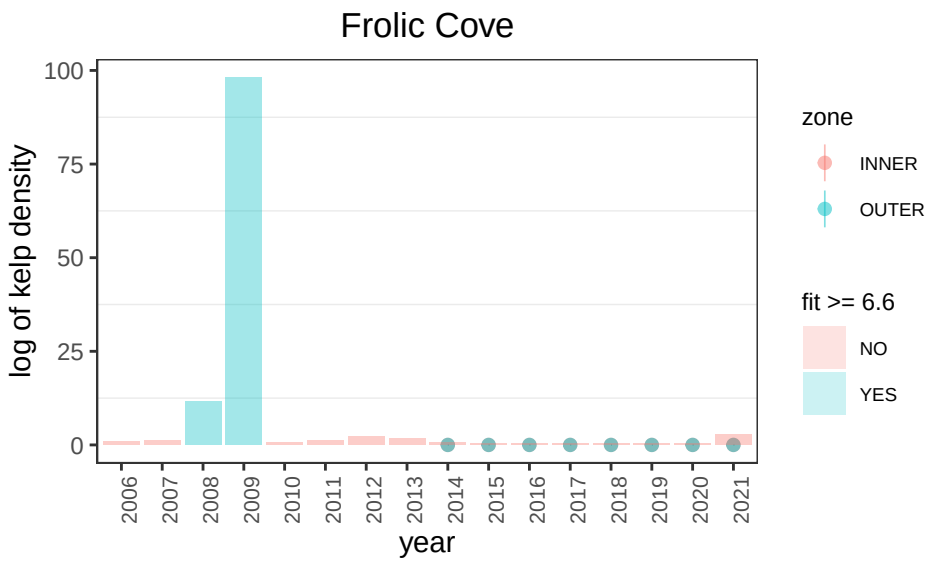

Caspar
log of kelp density
0
1
2
3
4
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO

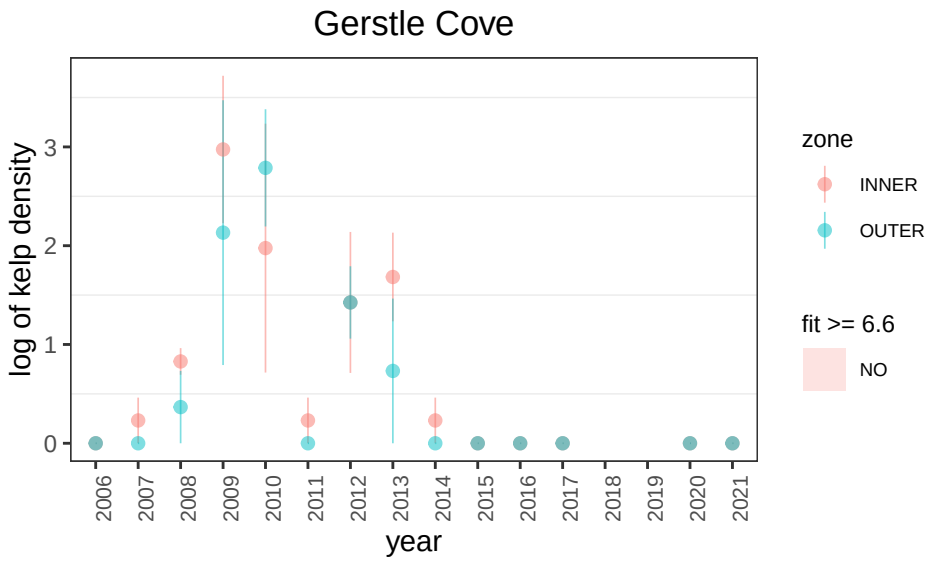

Caspar North
log of kelp density
0.0
0.5
1.0
1.5
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO

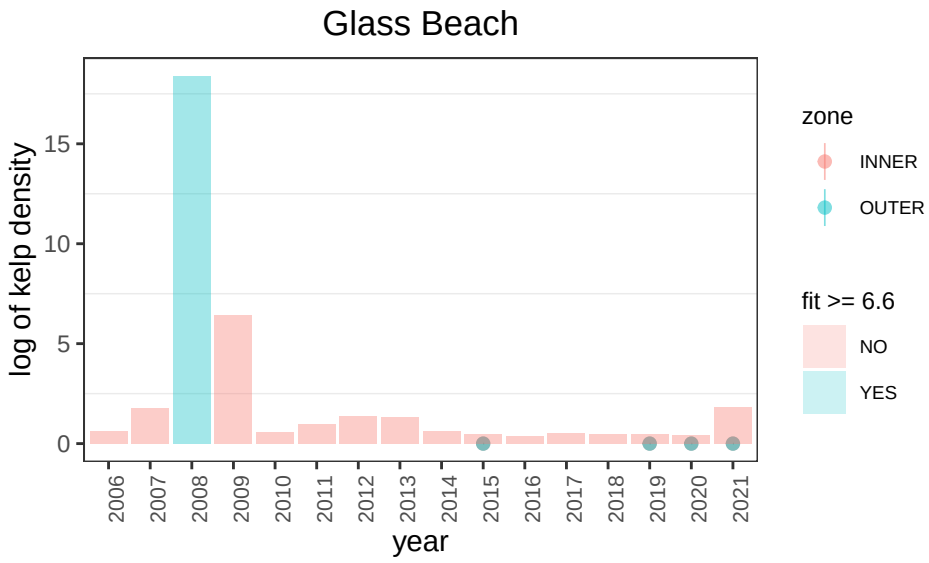

Dark Gulch
log of kelp density
0
5
10
15
20
25
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

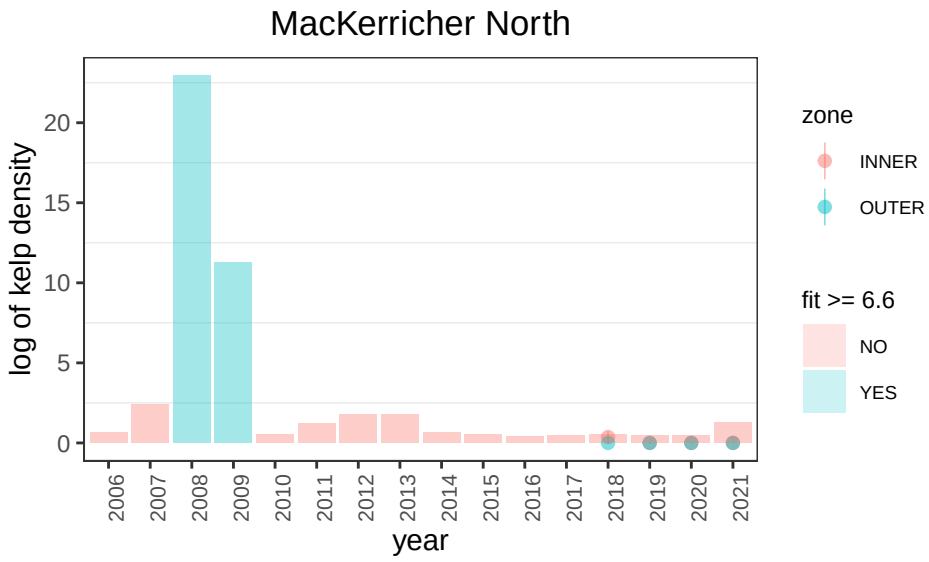

Flat Iron Rock
log of kelp density
20
15
10
5
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

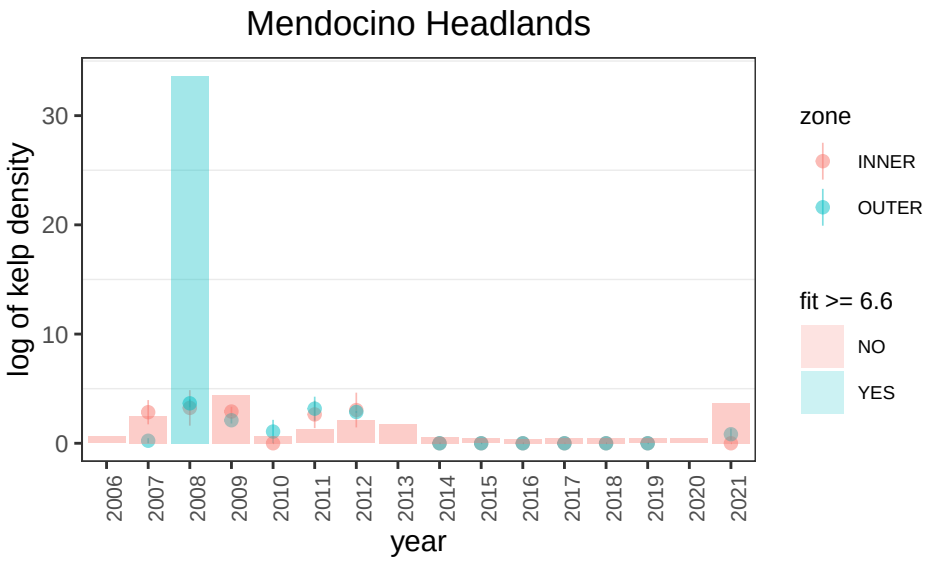

Fort Ross
log of kelp density
4
2
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO

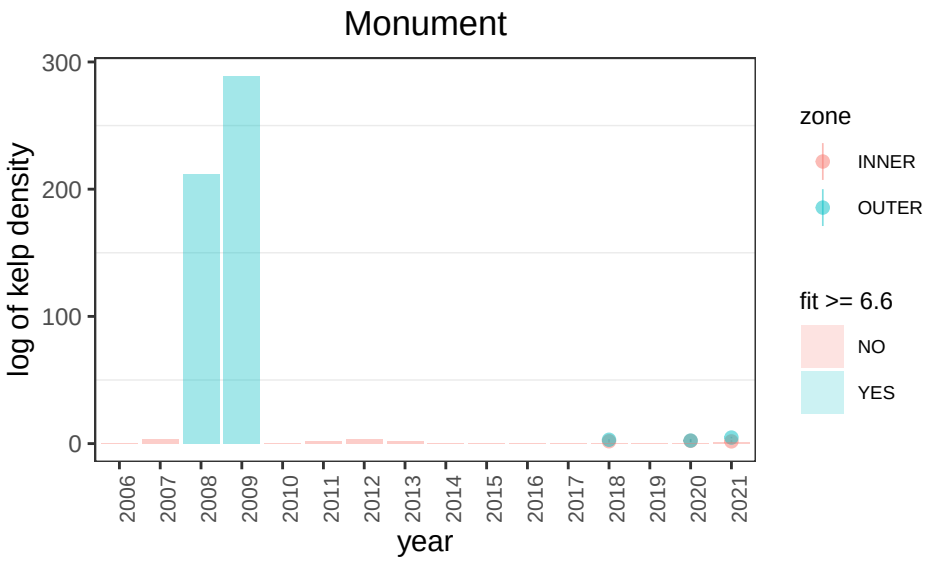

Frolic Cove
log of kelp density
100
75
50
25
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

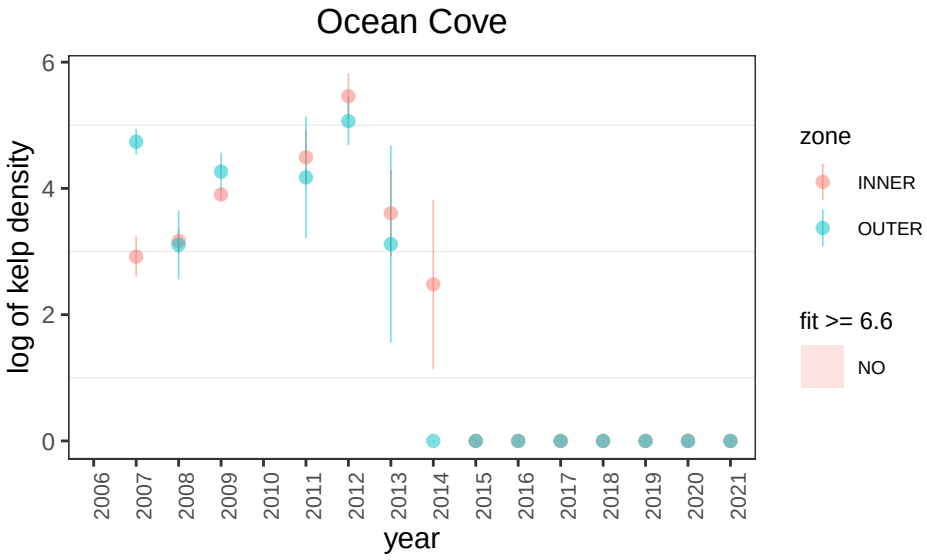

Gerstle Cove
log of kelp density
0
1
2
3
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO

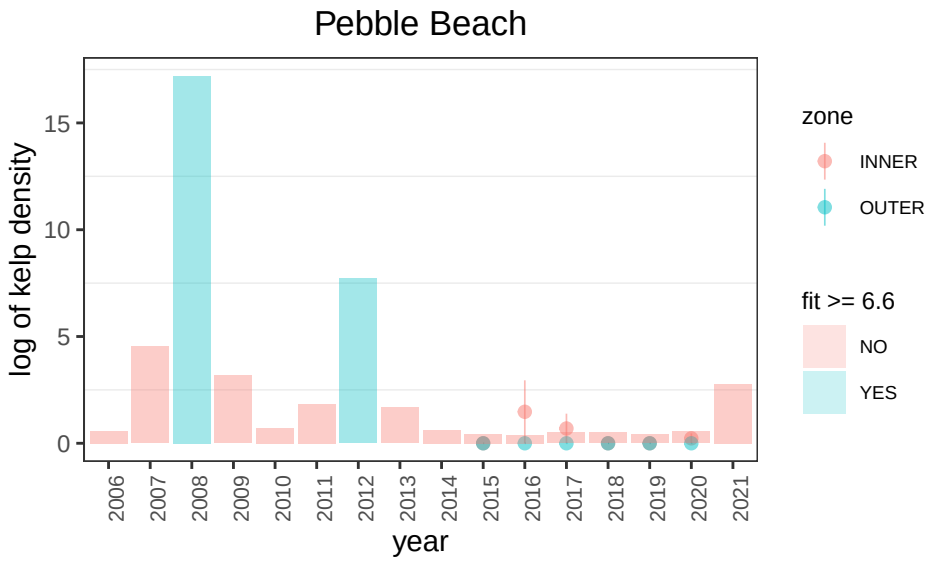

Glass Beach
log of kelp density
0
5
10
15
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

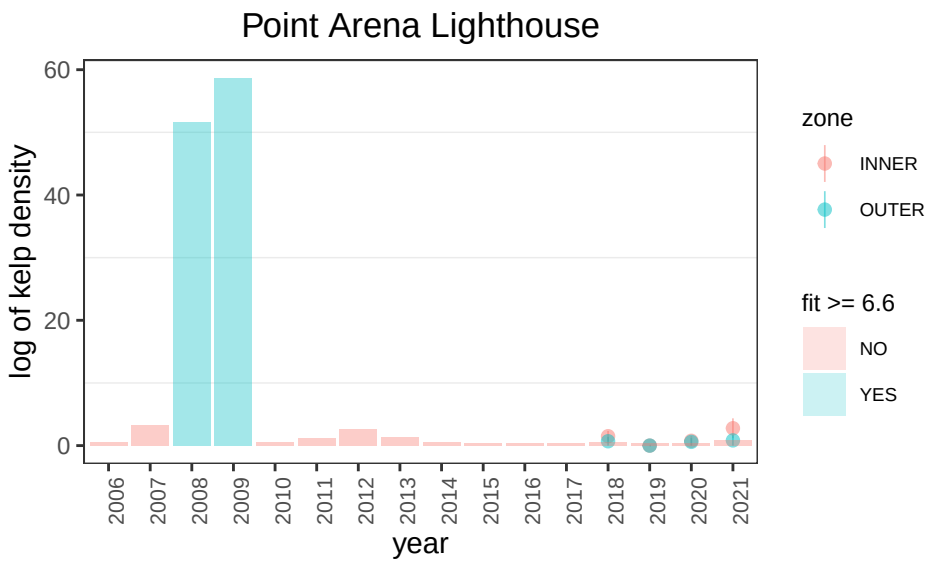

MacKerricher North
log of kelp density
0
5
10
15
20
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

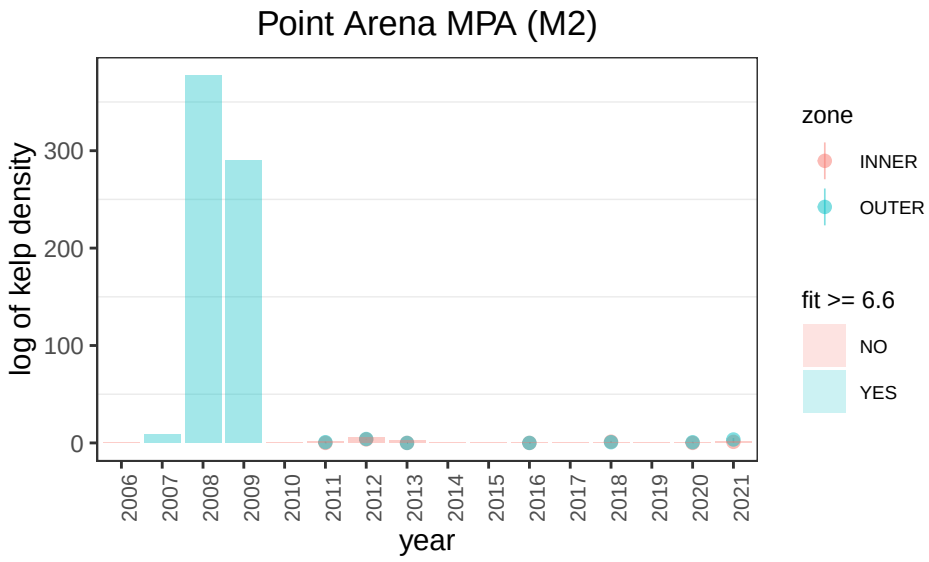

Mendocino Headlands
log of kelp density
30
20
10
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

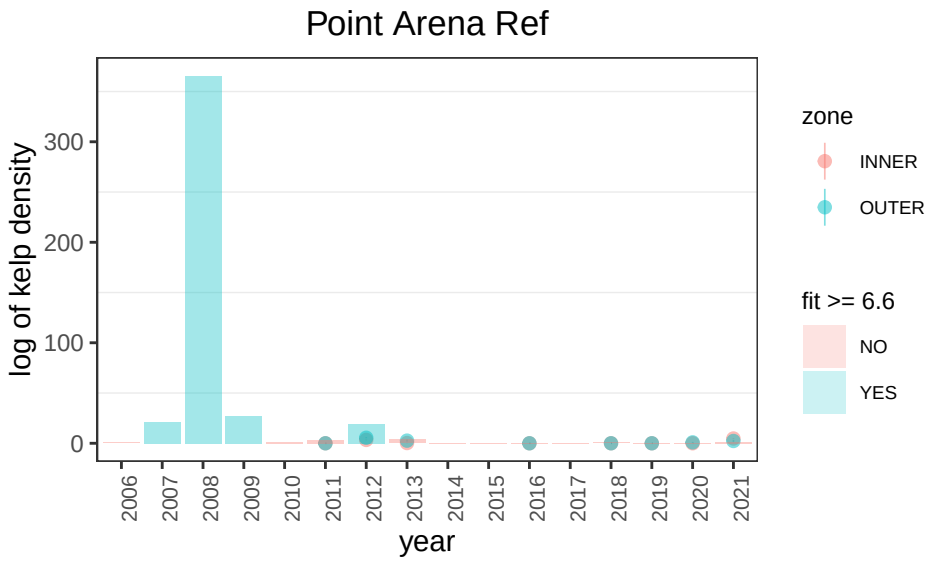

Monument
log of kelp density
300
200
100
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

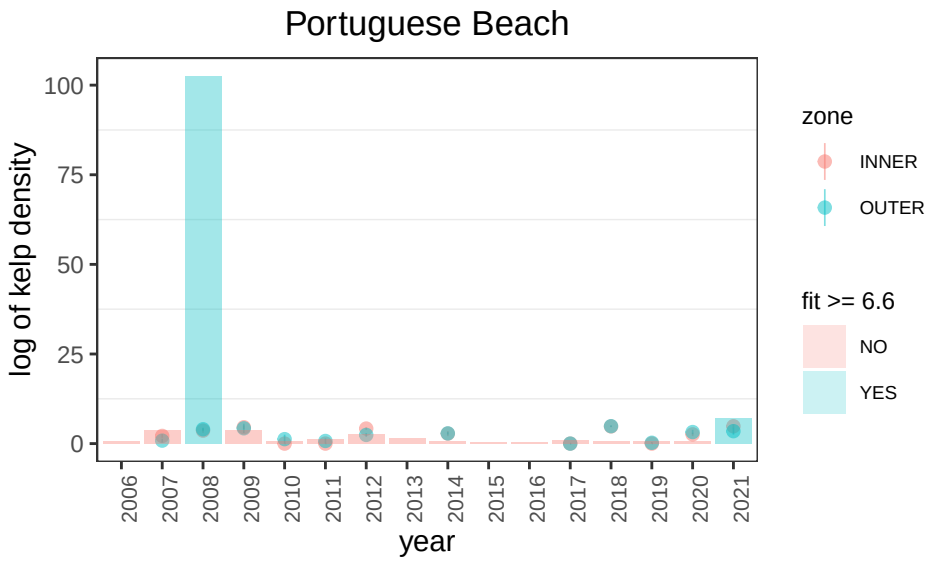

Ocean Cove
log of kelp density
6
4
2
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO

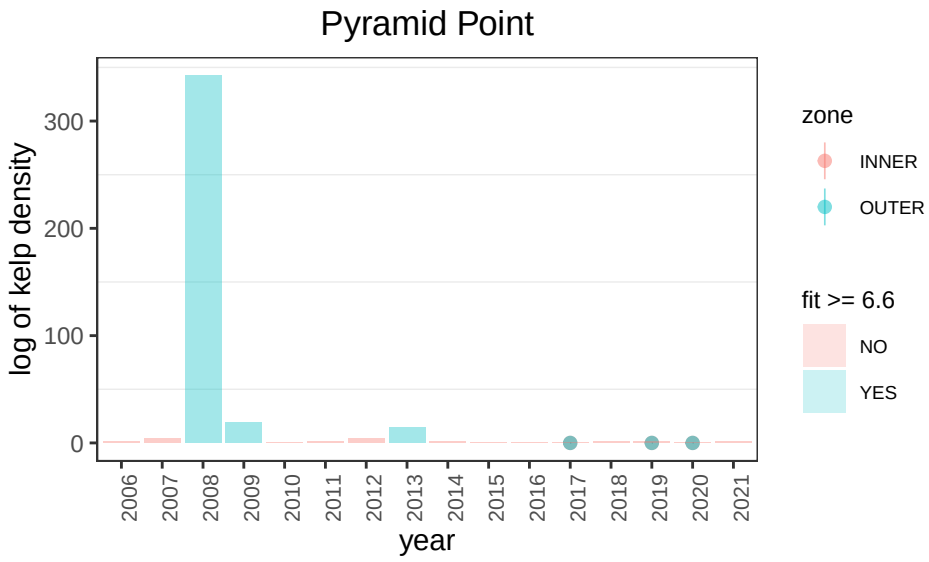

Pebble Beach
log of kelp density
15
10
5
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

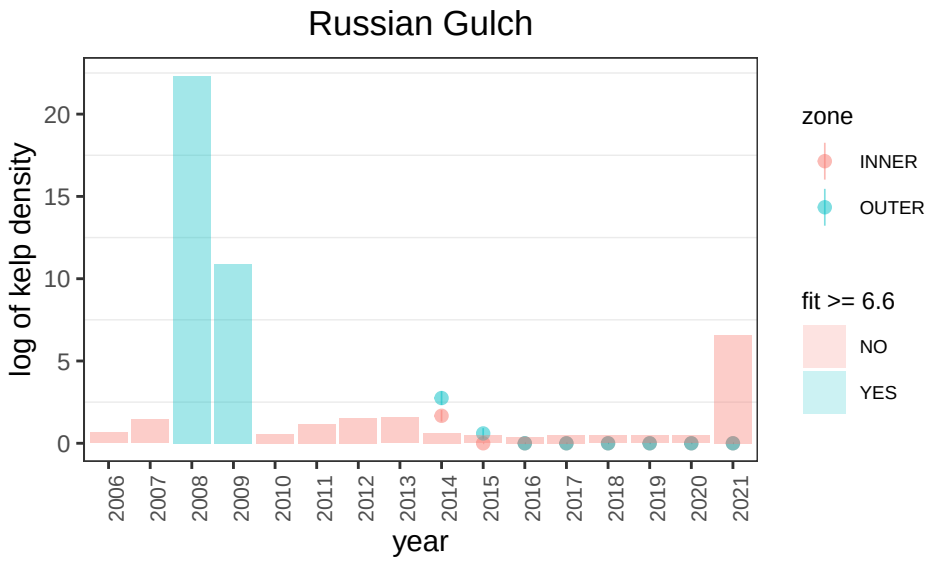

Point Arena Lighthouse
log of kelp density
60
40
20
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

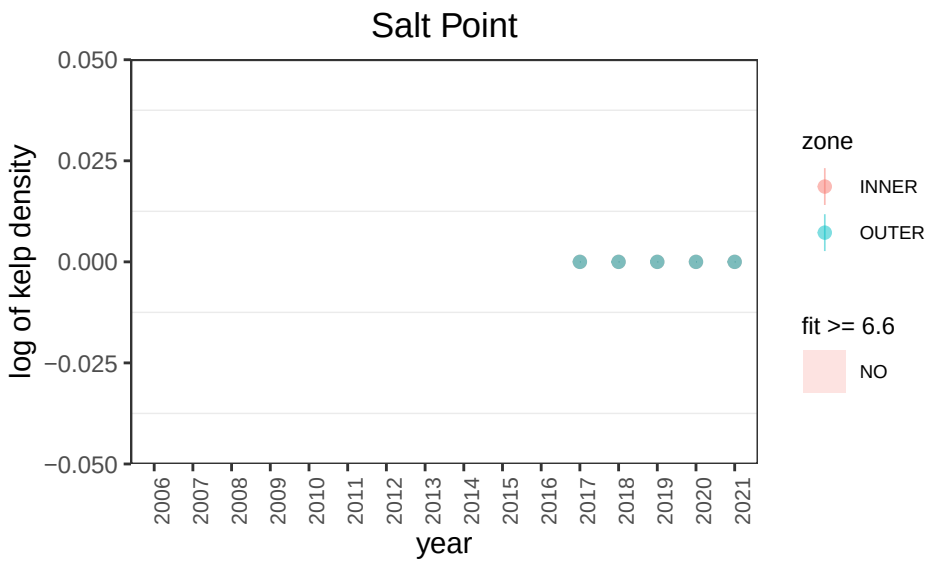

Point Arena MPA (M2)
log of kelp density
300
200
100
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

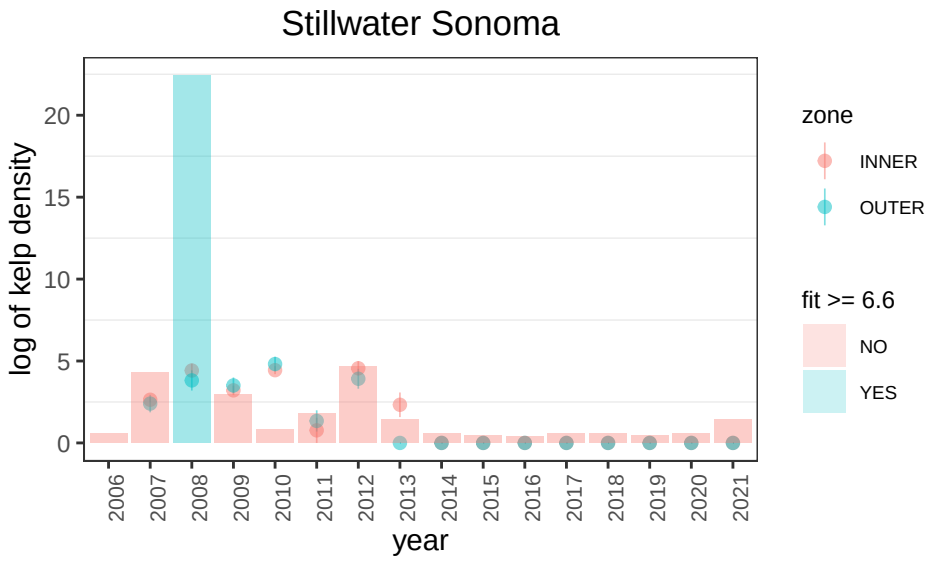

Point Arena Ref
log of kelp density
300
200
100
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

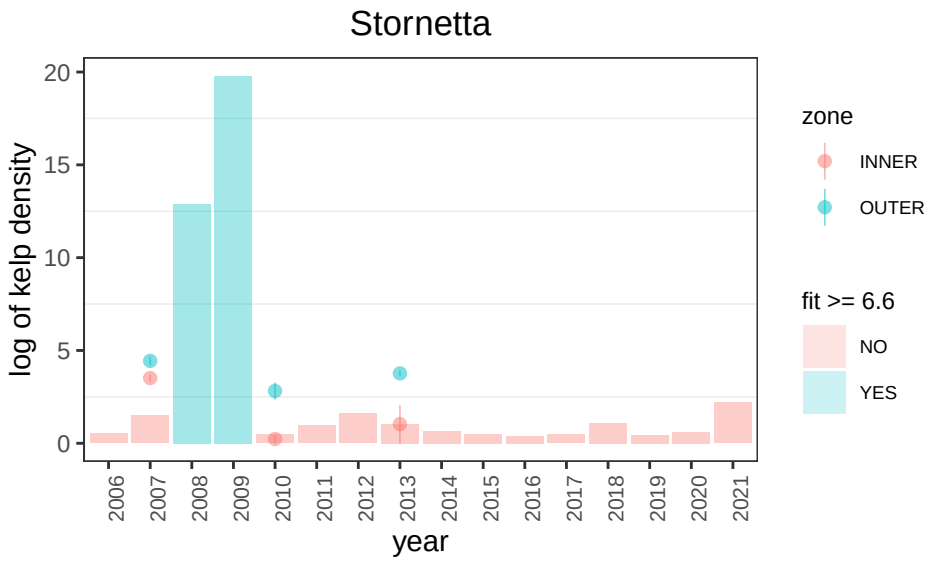

Portuguese Beach
log of kelp density
100
75
50
25
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

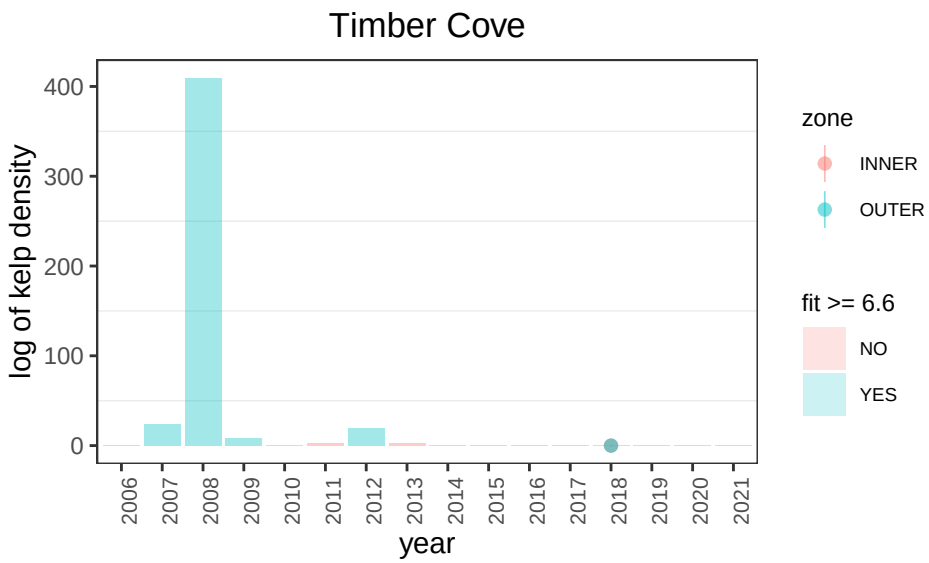

Pyramid Point
log of kelp density
300
200
100
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

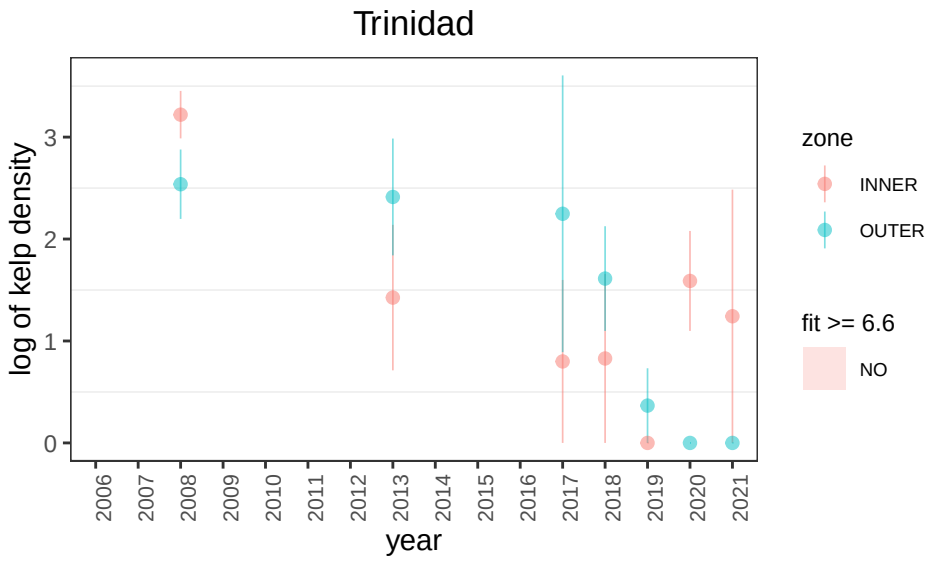

Russian Gulch
log of kelp density
20
15
10
5
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

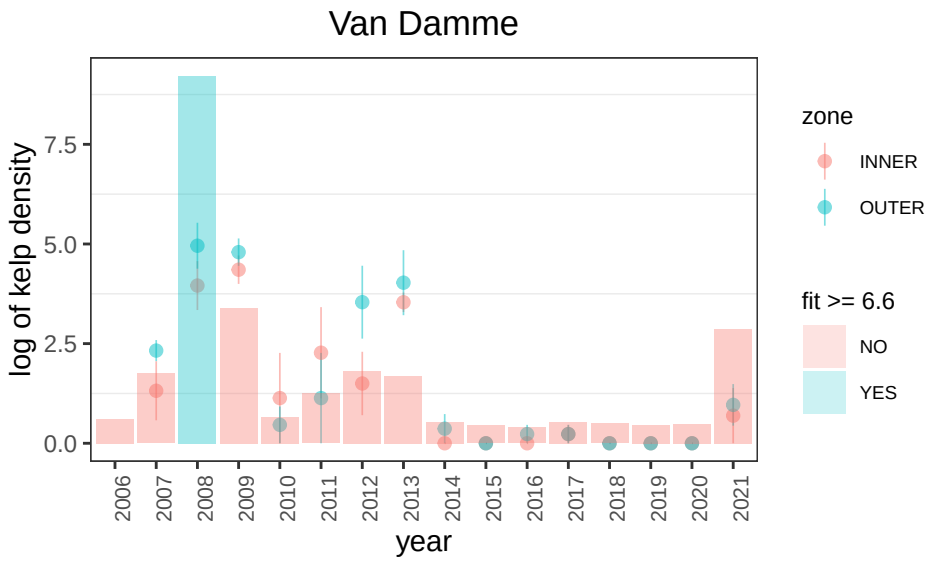

Salt Point
log of kelp density
0.050
0.025
0.000
−0.025
−0.050
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO


Stillwater Sonoma
log of kelp density
20
15
10
5
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES


Stornetta
log of kelp density
20
15
10
5
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES

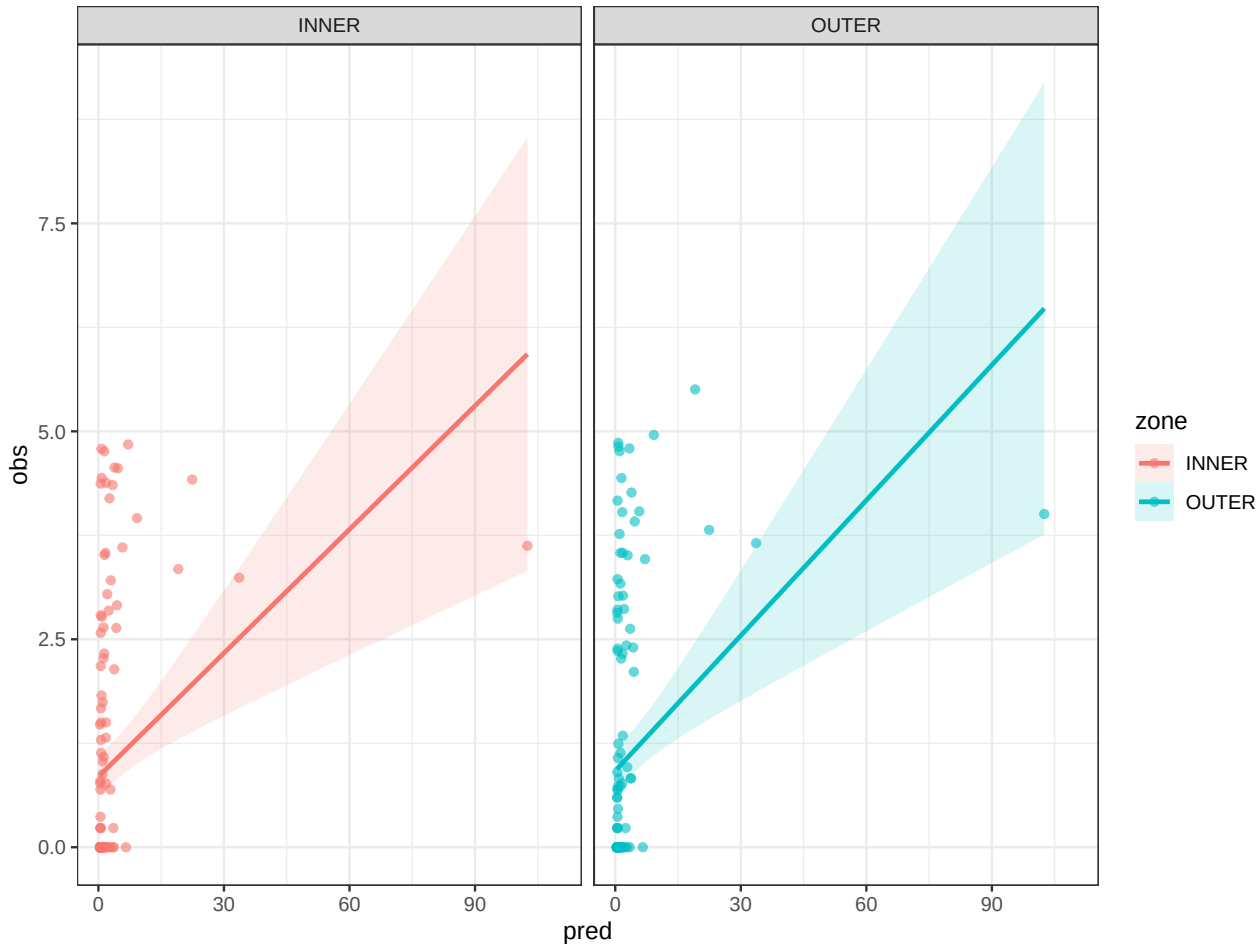

Timber Cove
log of kelp density
400
300
200
100
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO
YES


Trinidad
log of kelp density
3
2
1
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
year
zone
INNER
OUTER
fit >= 6.6
NO



Plot obs. vs pred.

```
kelp_longer %>%
  ggplot(aes(x = fit, y = mean, color = zone)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = 'lm', alpha = 0.15, aes(fill = zone)) +
  xlim(NA, 110) +
  labs(x = 'pred', y = 'obs') +
  facet_wrap(~zone, ncol = 2) +
  theme_bw()
```


INNER
OUTER
obs
pred
7.5
5.0
2.5
0.0
0
30
60
90
zone
INNER
OUTER