Assessing Pesticide Contamination in Illinois Water Supply and the Impact of Agricultural Practices

Code

Carson Edwards

1. Load in libraries

Hide

```
library(purrr)
library(readr)
library(corrplot)
library(forecast)
library(tidyr)
library(randomForest)
library(xgboost)
library(changepoint)
library(urbnmapr)
library(ggplot2)
library(dplyr)
library(data.table)
```

```
TRI PROGRAM DATASET ANALYSIS SECTION
```

2. Load and format TRI program data

Hide

```
folder_path <- "C:/Users/carso/Downloads/il_chemicals/"
file_list <- list.files(path = folder_path, pattern = "\\d{4}_il\\.csv$", full.names = TRUE)

combined_data <- map_df(file_list, ~ read_csv(.x, col_types = cols(.default = "c")))

# Clean up column names
cleaned_colnames <- colnames(combined_data) %>%
  gsub("^\\d+\\.\\s*", "", .) %>%
  gsub("\\s+", "_", .) %>%
  tolower()
colnames(combined_data) <- cleaned_colnames

# Mutate total releases to numeric, convert yes/no to logical, and as numeric
combined_data <- combined_data %>%
  mutate(
    total_releases = as.numeric(total_releases),
    across(where(~ is.character(.) & all(. %in% c("YES", "NO"))), ~ . == "YES"),
    across(where(is.logical), as.numeric)
  )

# Get sums of county, chemical, and year combinations
chemical_by_county <- combined_data %>%
  group_by(county, chemical, year) %>%
  summarise(total_releases = sum(total_releases, na.rm = TRUE))
```

```
`summarise()` has grouped output by 'county', 'chemical'. You can override using the `.groups` argument.
```

Hide

```
# Load in our county map data
illinois_counties <- get_urbn_map("counties", sf = FALSE) %>%
  filter(state_abbv == "IL")

head(combined_data)
```

y... <chr>	trifd <chr>	frs_id <chr>	facility_name <chr>	street_address <chr>	city <chr>
2001	60039TCNDSRT31X	110006165746	TC INDUSTRIES INC	3703 S RT 31	CRYSTAL LAKE
2001	62201THYLPMONSA	110000438884	AFTON CHEMICAL CORP	501 MONSANTO AVE	SAUGET
2001	60628SHRWN11541	110008457295	SHERWIN-WILLIAMS CO	11700 S COTTAGE GROVE	CHICAGO
2001	60120HNKLD1345G	110035808837	HENKEL US OPERATIONS CORP.	1345 GASKET DR	ELGIN
2001	61132BRBRC1354C	110066942474	SCHNEIDER-ELECTRIC LLC	1354 CLIFFORD AVE	LOVES PA
2001	60901HNKLCSKENS	110043972207	KENSING LLC	2525 S KENSINGTON AVE	KANKAKEI

6 rows | 1-6 of 122 columns

Hide

```
head(chemical_by_county)
```

county <chr>	chemical <chr>	year <chr>	total_releases <dbl>
ADAMS	2-Phenylphenol	2015	0.0
ADAMS	2-Phenylphenol	2016	0.0
ADAMS	Aluminum oxide (fibrous forms)	2004	72939.0
ADAMS	Aluminum oxide (fibrous forms)	2005	72967.0
ADAMS	Aluminum oxide (fibrous forms)	2006	98633.0
ADAMS	Aluminum oxide (fibrous forms)	2007	86443.2

6 rows

3. Correlation modeling

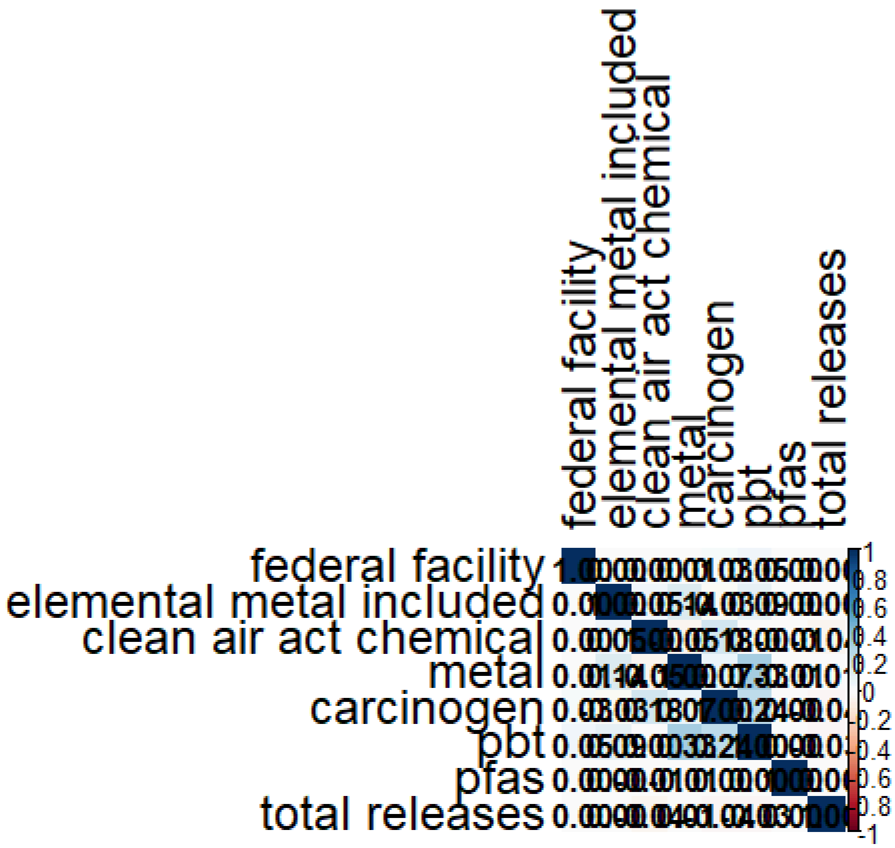
Hide

```
# Selecting our numeric data, we have already converted true/false to 1/0
numeric_data <- combined_data %>%
  select_if(is.numeric)
correlation_matrix <- cor(numeric_data, use = "complete.obs")

correlation_matrix
```

```
                         federal_facility elemental_metal_included clean_air_act_chemical        metal    ca
rcinogen           pbt          pfas total_releases
federal_facility             1.0000000000             -0.004400299           2.756273e-03  0.010069210  0.0
33442971  4.546254e-02 -0.0004614205  -0.0033693730
elemental_metal_included    -0.0044002990              1.000000000           4.703926e-02  0.141343327 -0.0
29909940  9.252047e-02 -0.0011087767   0.0037716602
clean_air_act_chemical       0.0027562734              0.047039260           1.000000e+00 -0.048969995  0.1
77330690  6.052090e-07 -0.0115923471  -0.0419258675
metal                        0.0100692104              0.141343327          -4.897000e-02  1.000000000  0.0
65808622  3.302266e-01 -0.0074324551   0.0112175256
carcinogen                   0.0334429714             -0.029909940           1.773307e-01  0.065808622  1.0
00000000  2.413876e-01  0.0023449760  -0.0397761881
pbt                          0.0454625433              0.092520466           6.052090e-07  0.330226619  0.2
41387601  1.000000e+00 -0.0037715181  -0.0328457127
pfas                        -0.0004614205             -0.001108777          -1.159235e-02 -0.007432455  0.0
02344976 -3.771518e-03  1.0000000000  -0.0008549982
total_releases              -0.0033693730              0.003771660          -4.192587e-02  0.011217526 -0.0
39776188 -3.284571e-02 -0.0008549982   1.0000000000
```

Hide

```
colnames(correlation_matrix) <- gsub("_", " ", colnames(correlation_matrix))
rownames(correlation_matrix) <- gsub("_", " ", rownames(correlation_matrix))

# Paste into console to see a readable plot in plot window
corrplot(correlation_matrix,
         method = "color",
         tl.cex = 1.5,
         tl.col = "black",
         addCoef.col = "black")
```



4. Random forest model for feature importance

Hide

```
# Only selecting fields of interest
model_data <- combined_data %>%
  select(total_releases, year, federal_facility, elemental_metal_included, clean_air_act_chemical, metal, c
arcinogen, pbt, pfas)

set.seed(123)
rf_model <- randomForest(total_releases ~ ., data = model_data, importance = TRUE, ntree = 500)

print(rf_model)
```

```
Call:
 randomForest(formula = total_releases ~ ., data = model_data,      importance = TRUE, ntree = 500)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 2

          Mean of squared residuals: 73827638192
                    % Var explained: 0.48
```

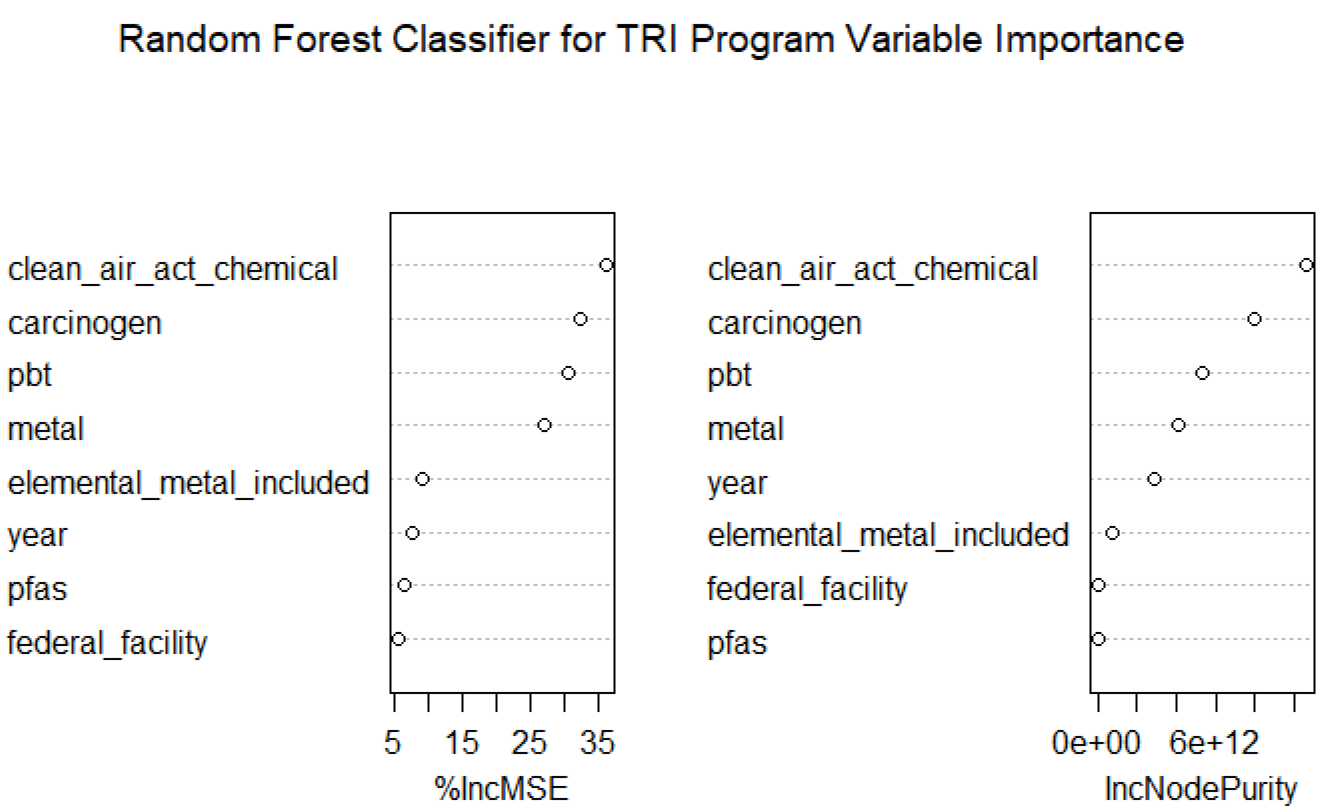
Hide

```
importance(rf_model)
```

```
                           %IncMSE  IncNodePurity
year                      7.674713   2.830637e+12
federal_facility          5.625889   3.394976e+10
elemental_metal_included  9.285051   7.679550e+11
clean_air_act_chemical   36.188221   1.065696e+13
metal                    26.927291   4.120005e+12
carcinogen               32.283238   7.966650e+12
pbt                      30.601349   5.337563e+12
pfas                      6.646834   2.565066e+09
```

Hide

```
varImpPlot(rf_model,
           main = "Random Forest Classifier for TRI Program Variable Importance")
```



5. Change point analysis for TRI program data

Hide

```
get_change_points <- function(df) {
  df$year <- as.numeric(df$year)
  df <- df[order(df$year), ]
  if (any(is.na(df$total_releases)) || !is.numeric(df$total_releases)) {
    return(NULL)
  }

  releases <- as.numeric(df$total_releases)
  max_Q <- length(releases) - 1

  # The max change points to search for. Taking either length of county chemical combinations or total numb
er of rows. Want less change points rather than more
  Q <- min(3, max_Q)
  if (Q < 1) {
    return(NULL)
  }

  # Was getting too many warnings with the output
  suppressWarnings({
    cpt <- cpt.mean(releases, method = "BinSeg", penalty = "BIC", Q = Q)
  })

  return(cpts(cpt))
}

change_points <- list()

# Only want combinations with at least 3 reported years
valid_data <- chemical_by_county %>%
  group_by(county, chemical) %>%
  filter(n() >= 3) %>%
  ungroup()

# Need unique pairs
unique_combinations <- unique(valid_data[, c("county", "chemical")])

# For each unique combination of county and chemical, take the subset of data and run get change points fun
ction
for (i in 1:nrow(unique_combinations)) {
  county1 <- unique_combinations$county[i]
  chemical1 <- unique_combinations$chemical[i]
  subset_data <- subset(valid_data, county == county1 & chemical == chemical1)

  result <- get_change_points(subset_data)

  if (!is.null(result)) {
    change_points[[paste(county1, chemical1, sep = "_")]] <- result
  } else {
    change_points[[paste(county1, chemical1, sep = "_")]] <- "No valid data or change points"
  }
}
```

6. Functions for plotting change points and getting year counts

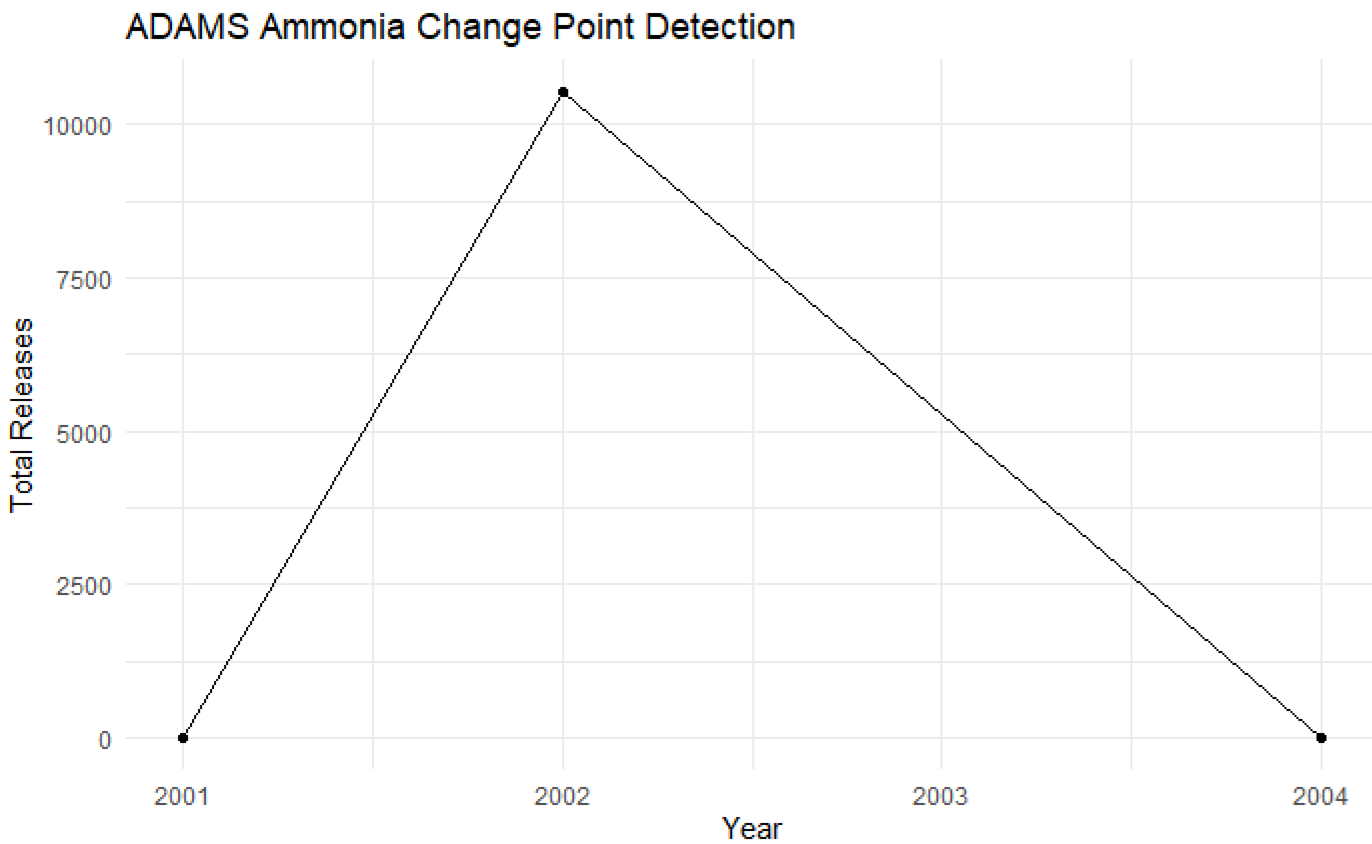
Hide

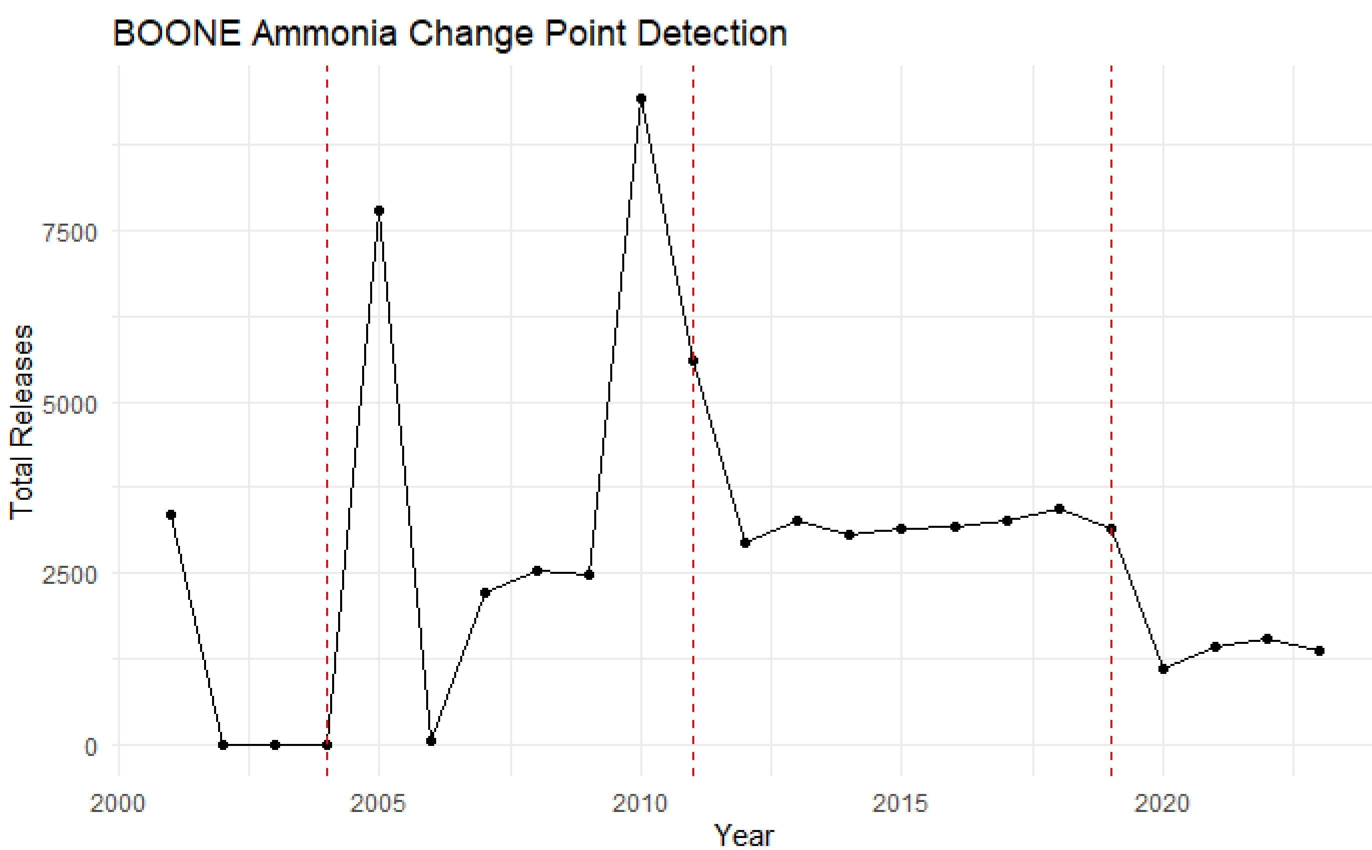
```
plot_change_points <- function(df, change_points) {
  df$year <- as.numeric(df$year)

  # Order by year
  df <- df[order(df$year), ]

  # Create change point plot of total releases
  change_point_plot <- ggplot(df, aes(x = year, y = total_releases)) +
    geom_line() +
    geom_point() +
    labs(title = paste(df$county[1], df$chemical[1], "Change Point Detection"),
         x = "Year", y = "Total Releases") +
    theme_minimal()

  # Plotting change points using a vertical red line
  if (!is.null(change_points) && length(change_points) > 0) {
    change_point_plot <- change_point_plot + geom_vline(xintercept = df$year[change_points], color = "red",
linetype = "dashed")
  }
  print(change_point_plot)
}

analyze_change_points <- function(chemical_name, chemical_by_county, change_points) {
  unique_counties <- unique(chemical_by_county$county[chemical_by_county$chemical == chemical_name])
  change_point_years <- c()

  # For each county chemical combination, get the change point counts
  for (county1 in unique_counties) {
    subset_data <- subset(chemical_by_county, county == county1 & chemical == chemical_name)
    change_point_result <- change_points[[paste(county1, chemical_name, sep = "_")]]

    if (!is.null(change_point_result)) {
      plot_change_points(subset_data, change_point_result)
      change_point_years <- c(change_point_years, change_point_result)  # Collect change point years
    }
  }

  change_point_year_counts <- table(change_point_years)

  # Plot the totals
  barplot(change_point_year_counts,
          main = paste("Frequency of Change Point Years for", chemical_name),
          xlab = "Year",
          ylab = "Frequency",
          col = "skyblue",
          las = 2,
          names.arg = as.character(as.numeric(names(change_point_year_counts)) + 2000)) # Have years displa
y correctly

}
```

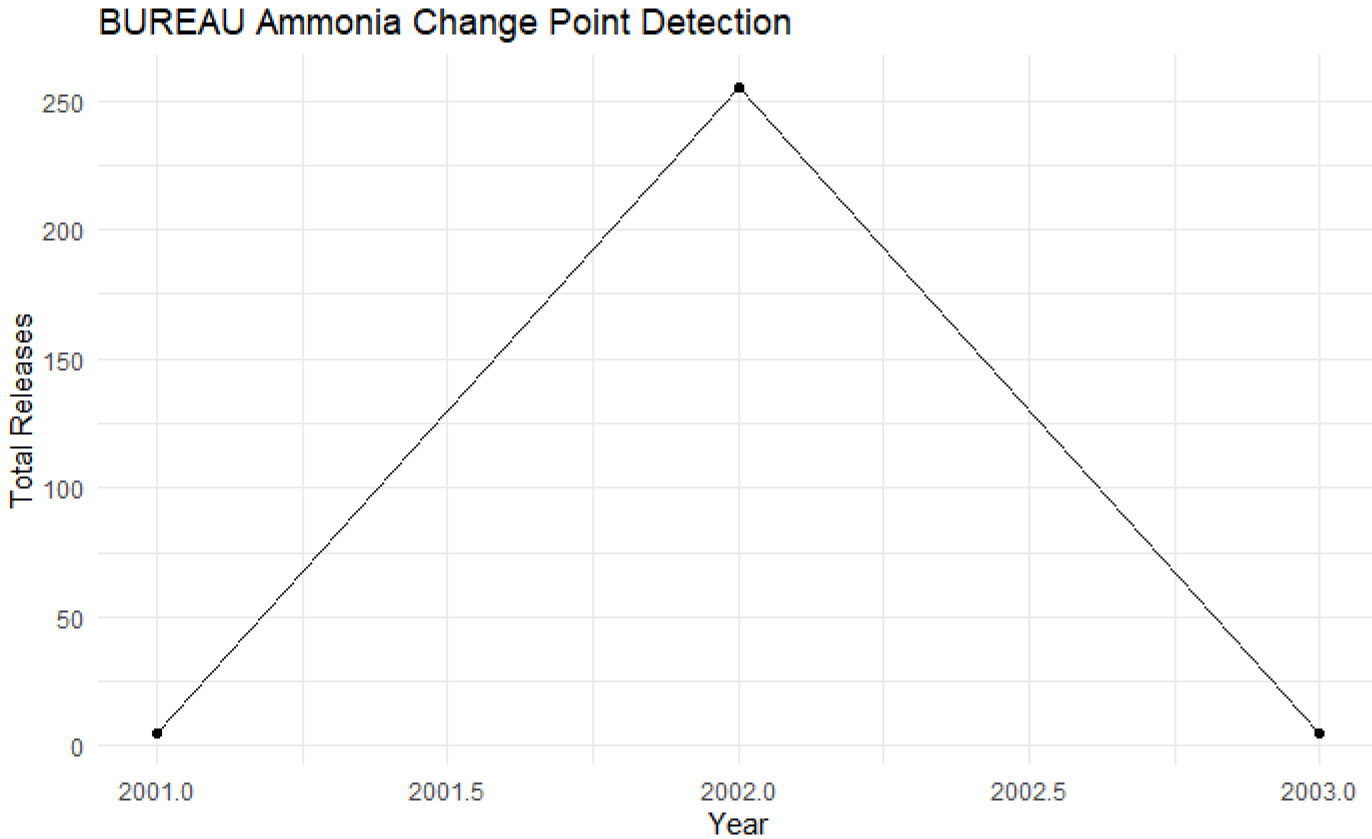
Ammonia example

Hide

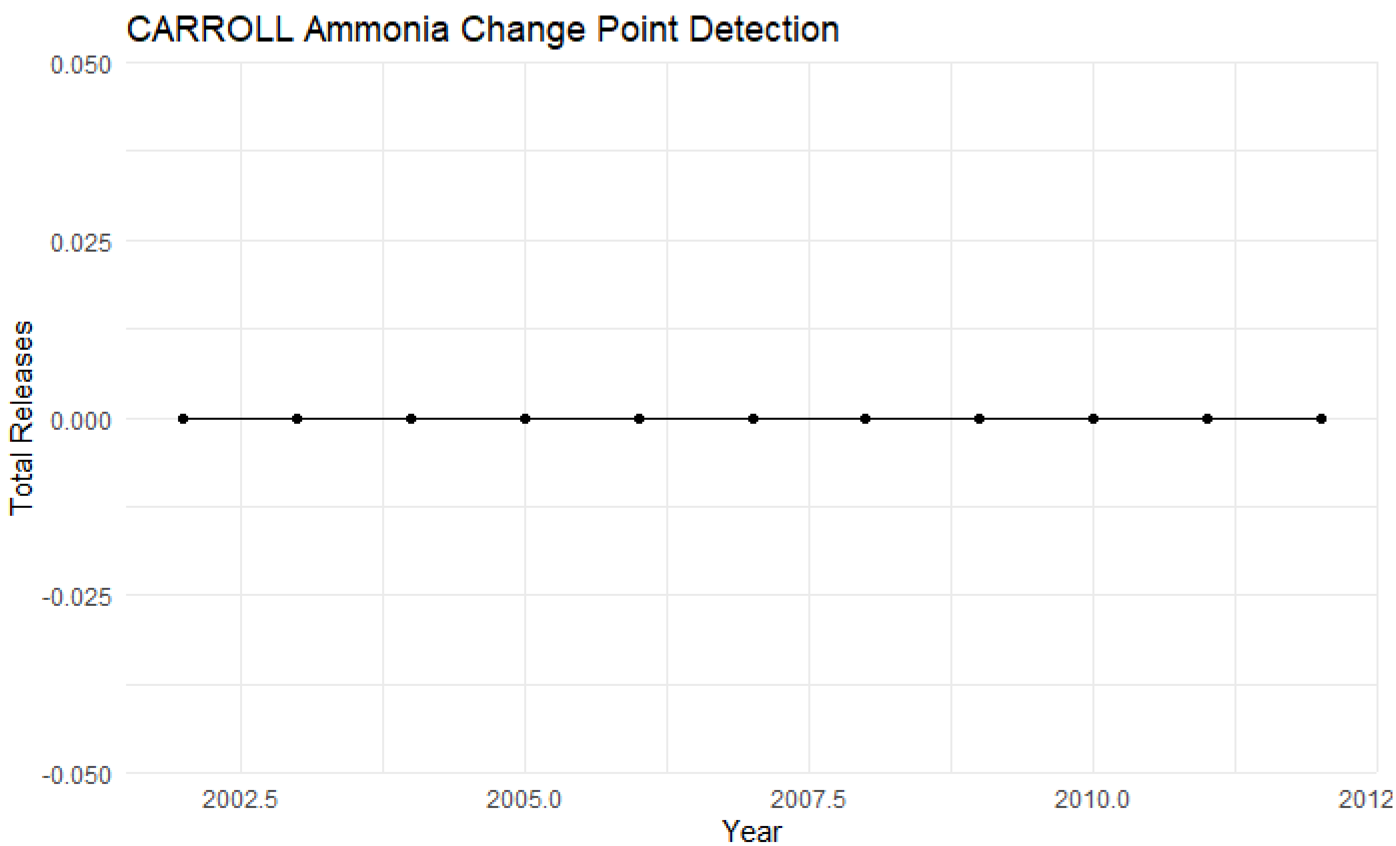
```
analyze_change_points("Ammonia", chemical_by_county, change_points)
```



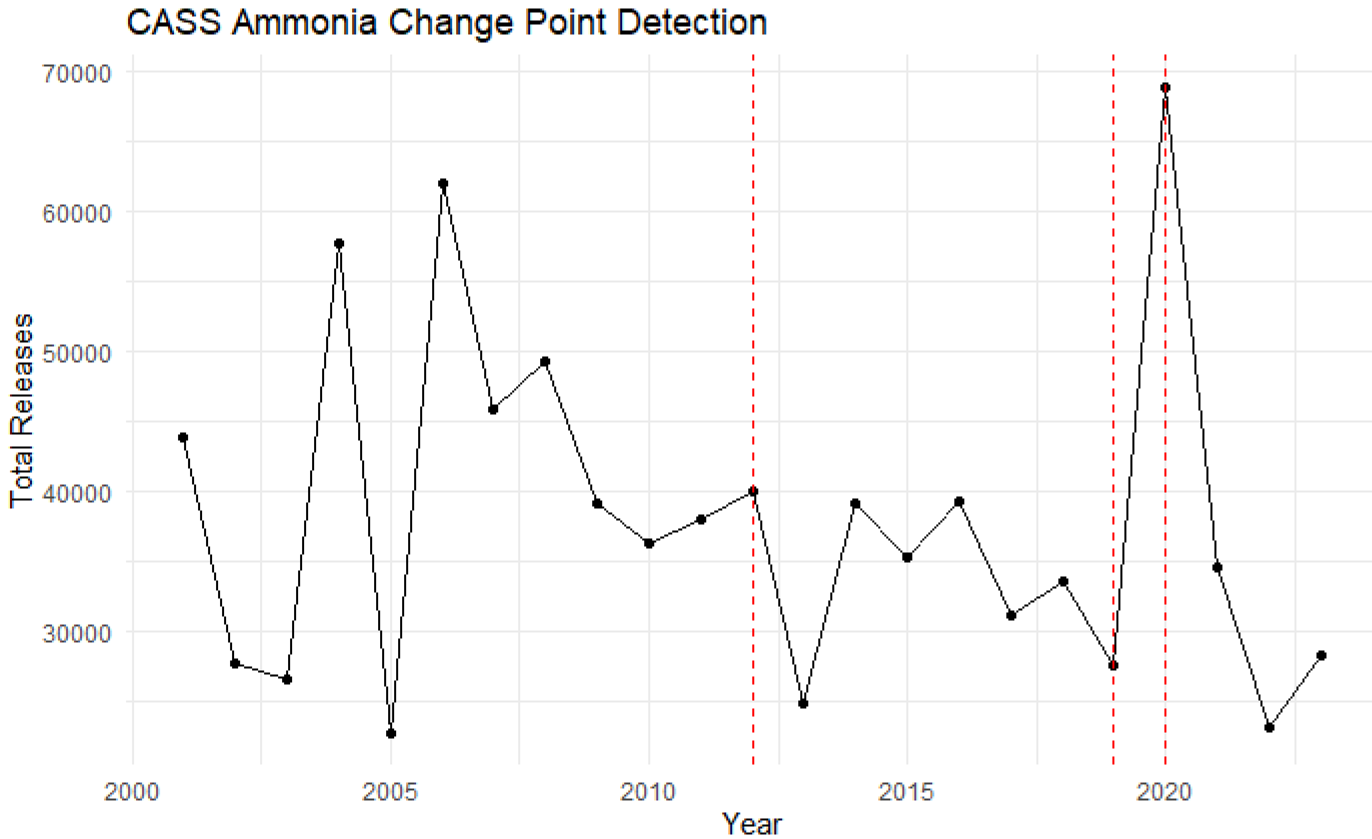


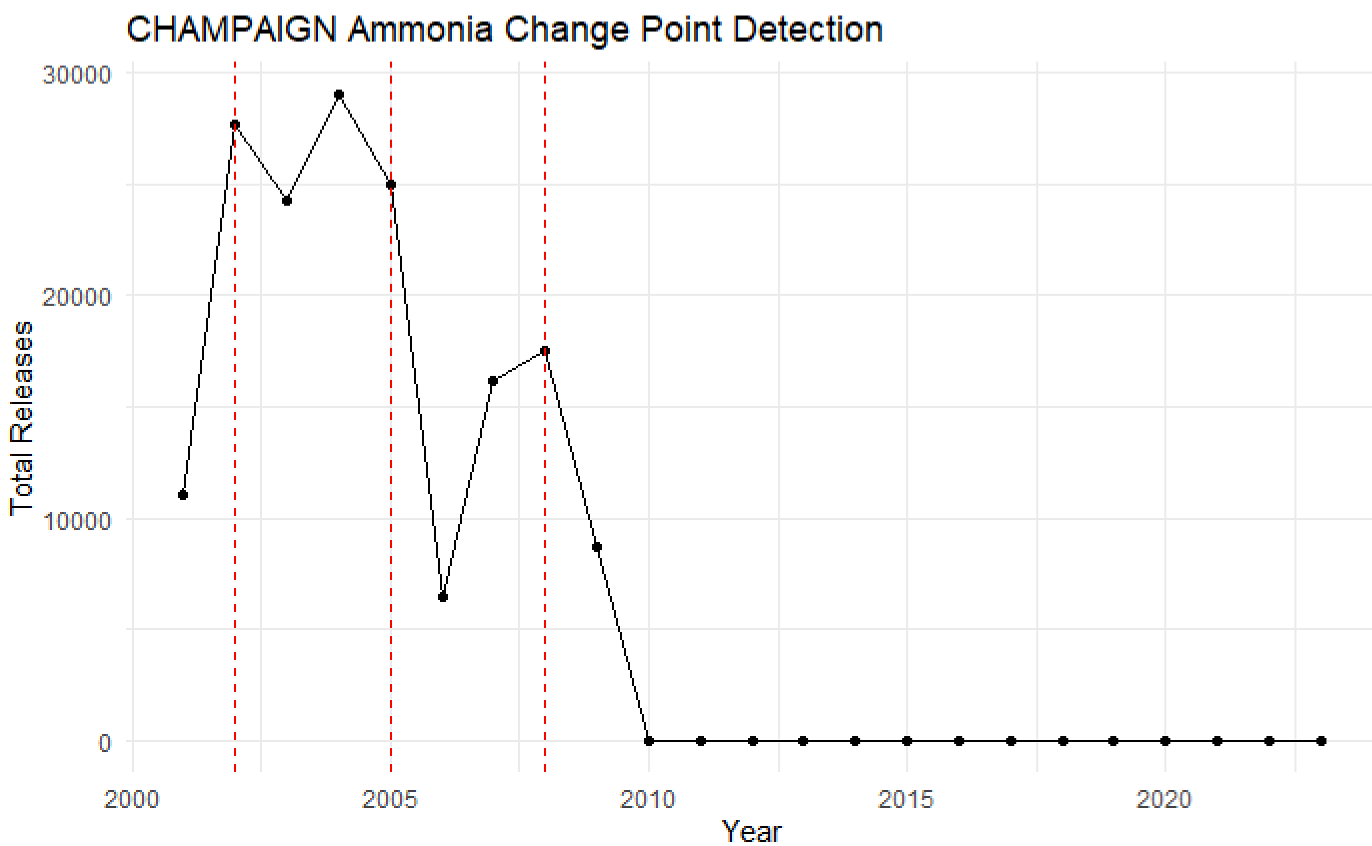


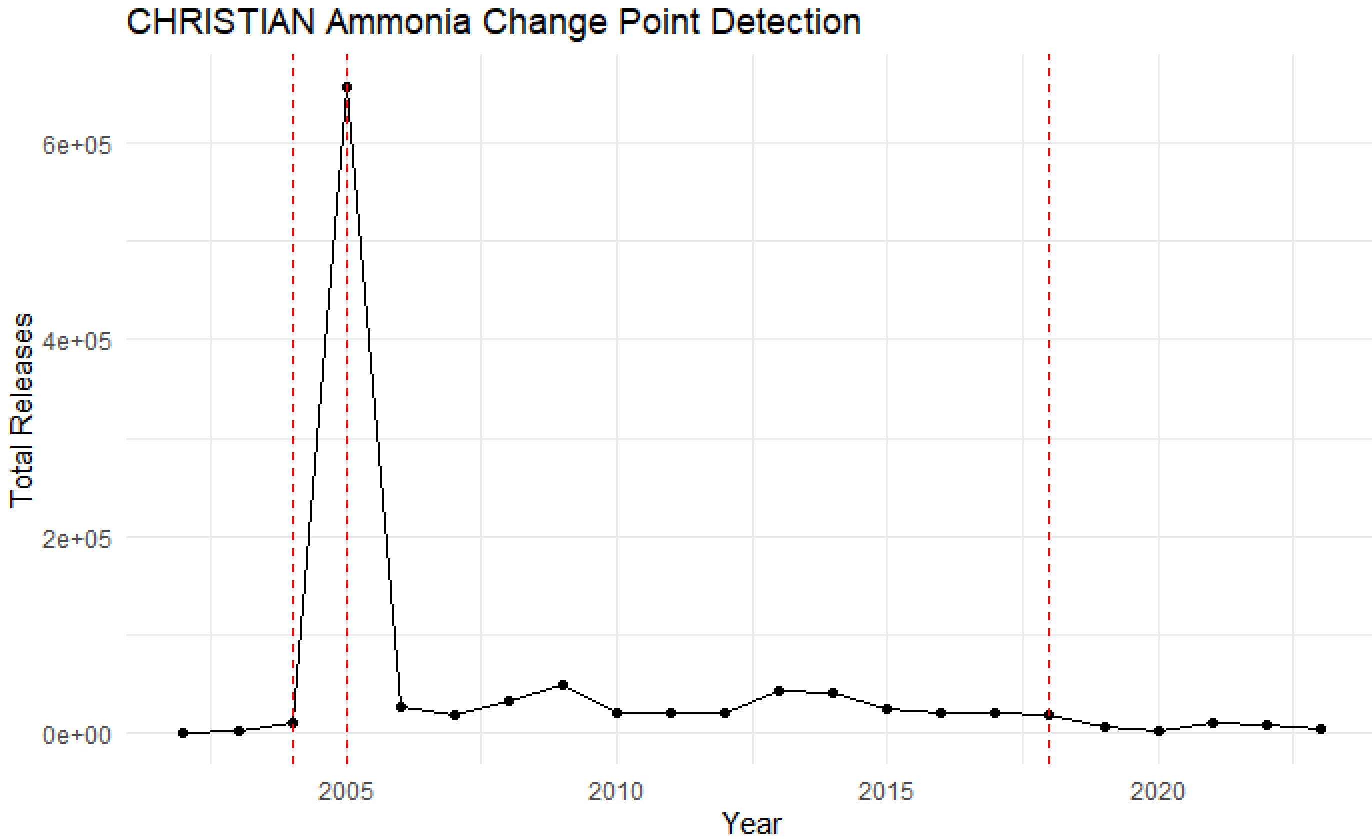
BUREAU Ammonia Change Point Detection
Total Releases
Year
0
50
100
150
200
250
2001.0
2001.5
2002.0
2002.5
2003.0



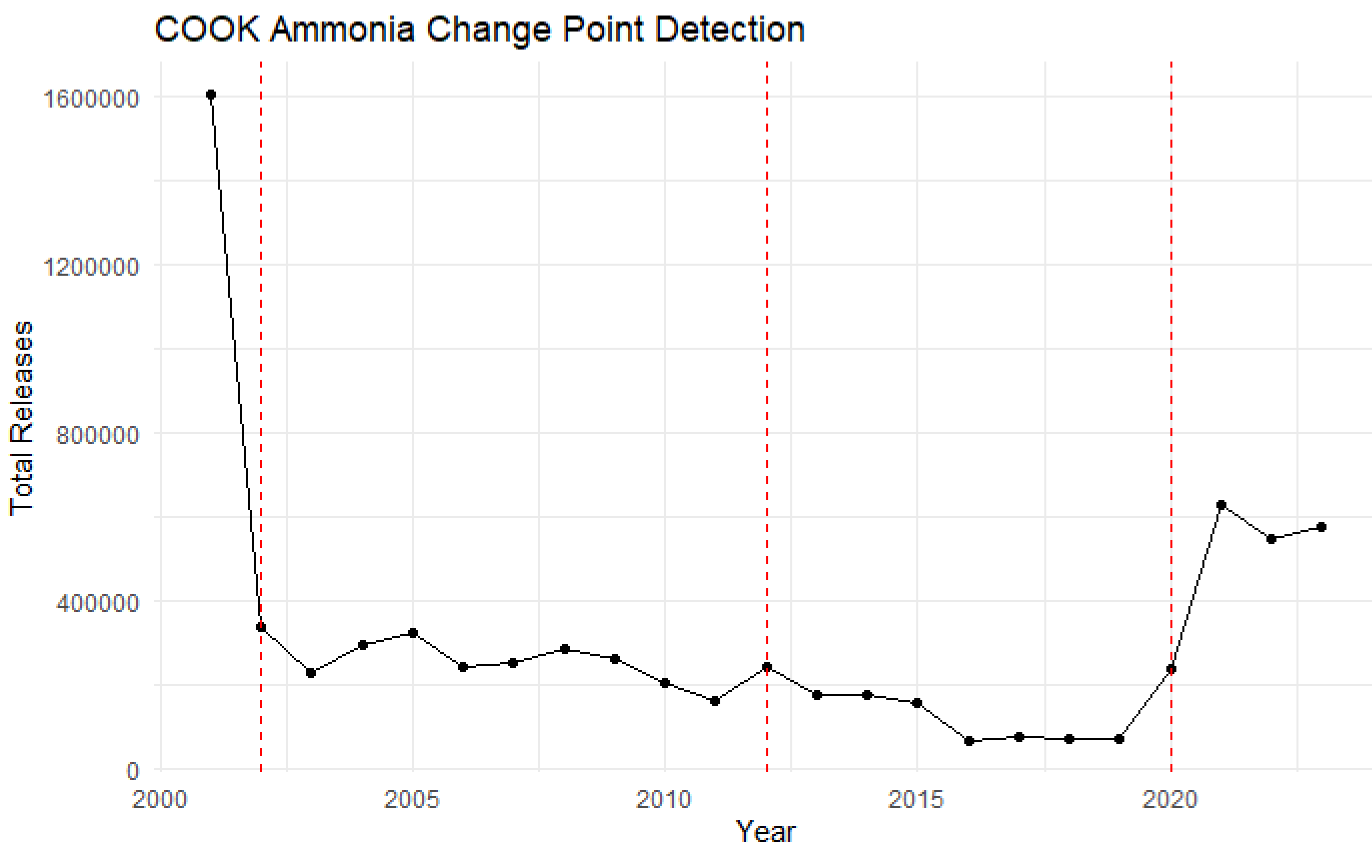
CARROLL Ammonia Change Point Detection
Total Releases
Year
-0.050
-0.025
0.000
0.025
0.050
2002.5
2005.0
2007.5
2010.0
2012


CASS Ammonia Change Point Detection
70000
60000
50000
40000
30000
Total Releases
2000
2005
2010
2015
2020
Year

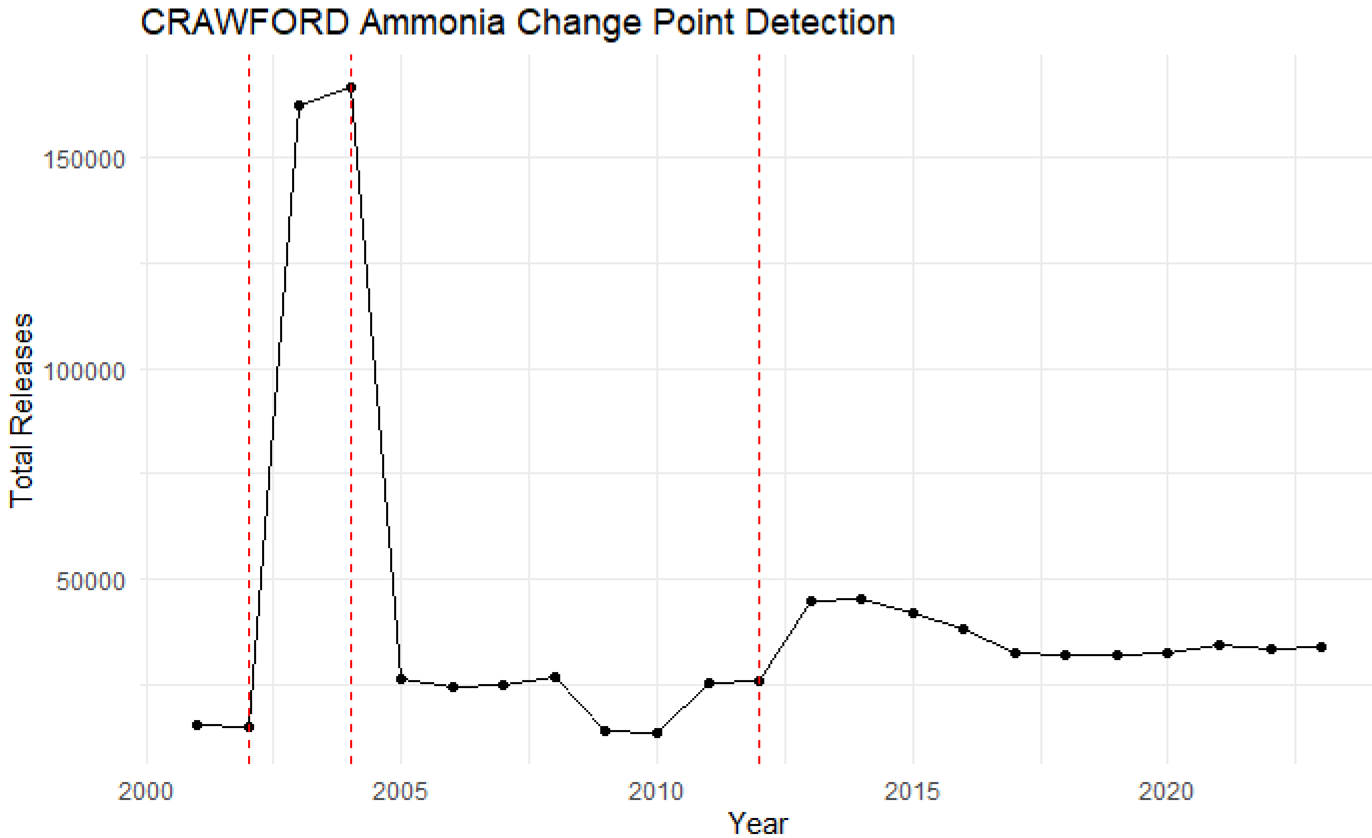

CHAMPAIGN Ammonia Change Point Detection
30000
20000
10000
0
Total Releases
2000
2005
2010
2015
2020
Year

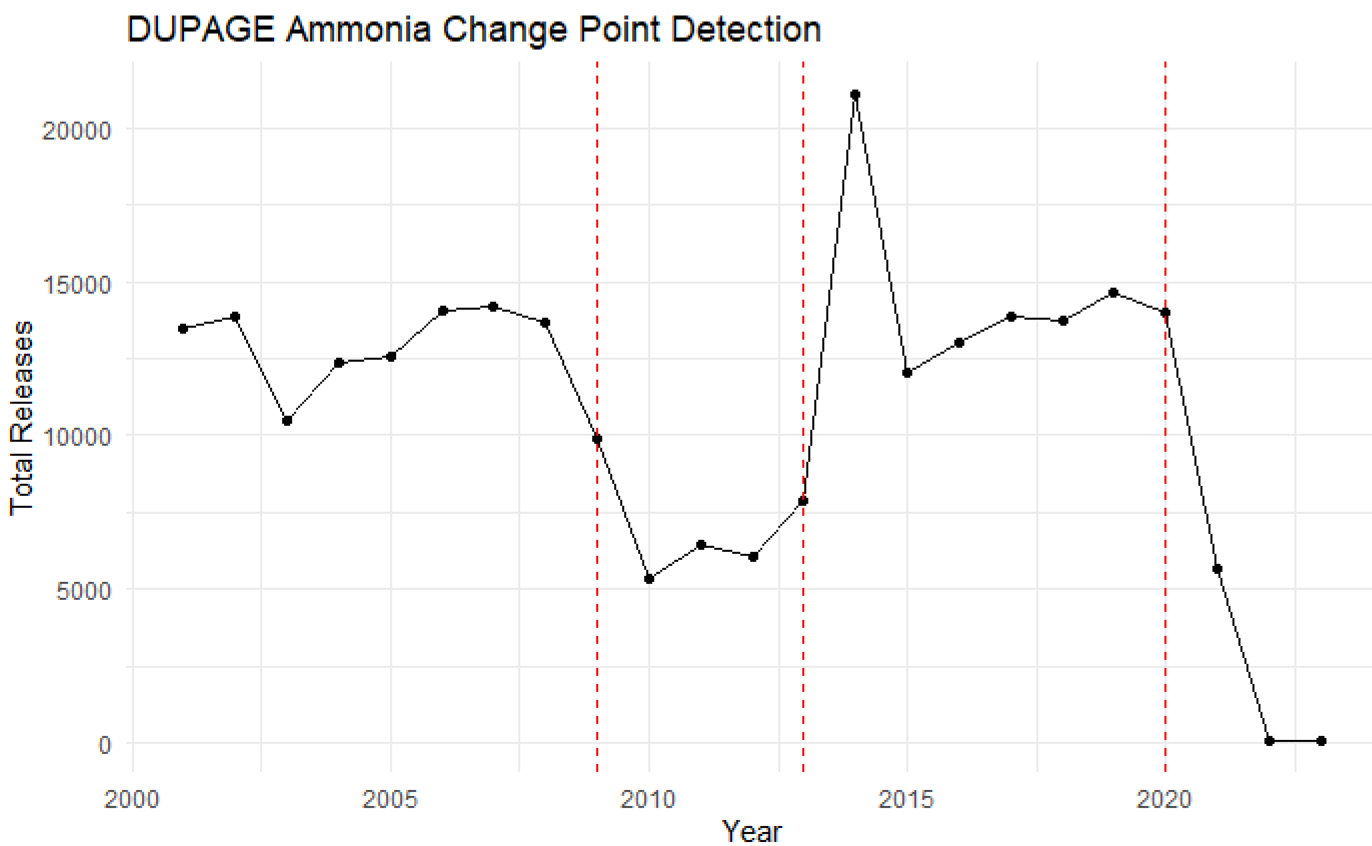


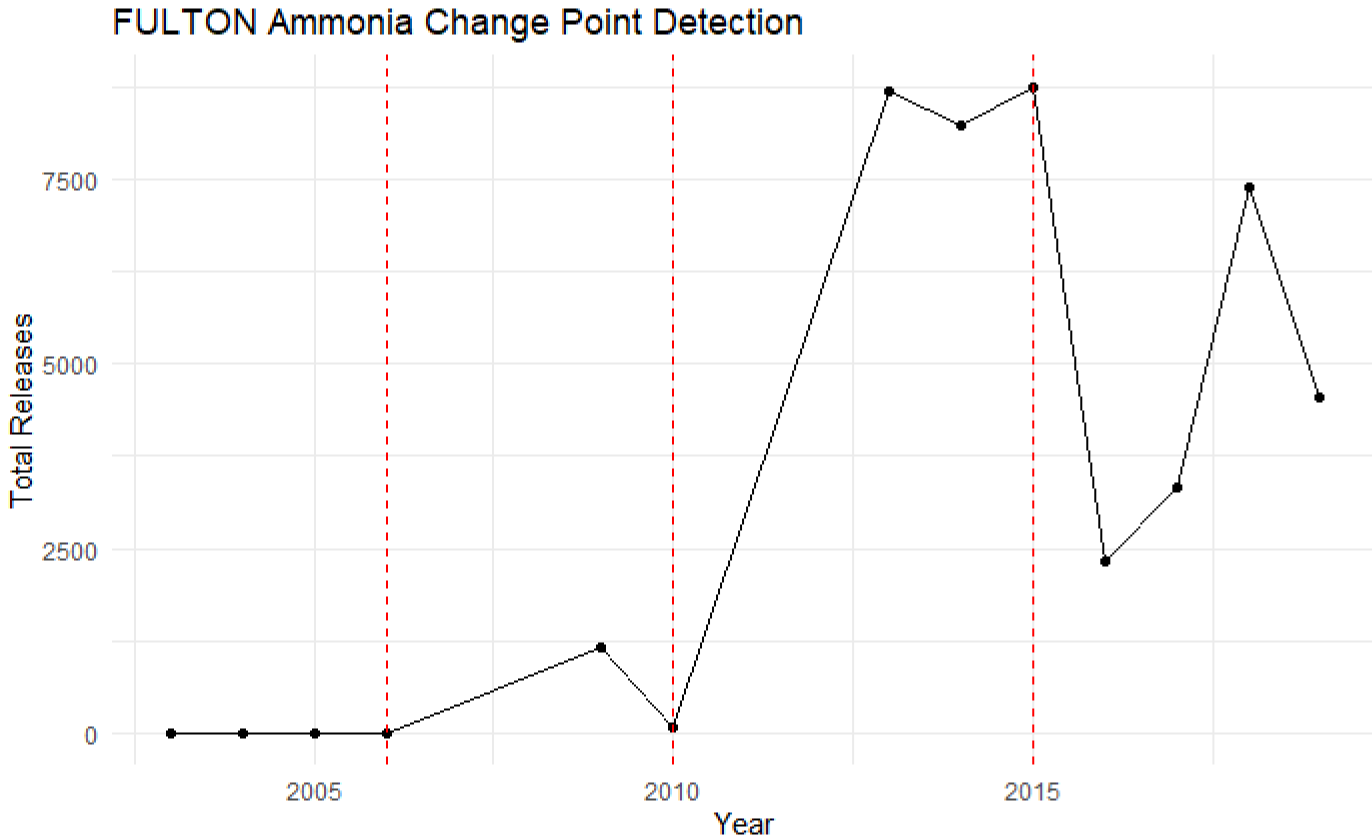
CHRISTIAN Ammonia Change Point Detection
Total Releases
6e+05
4e+05
2e+05
0e+00
2005
2010
2015
2020
Year

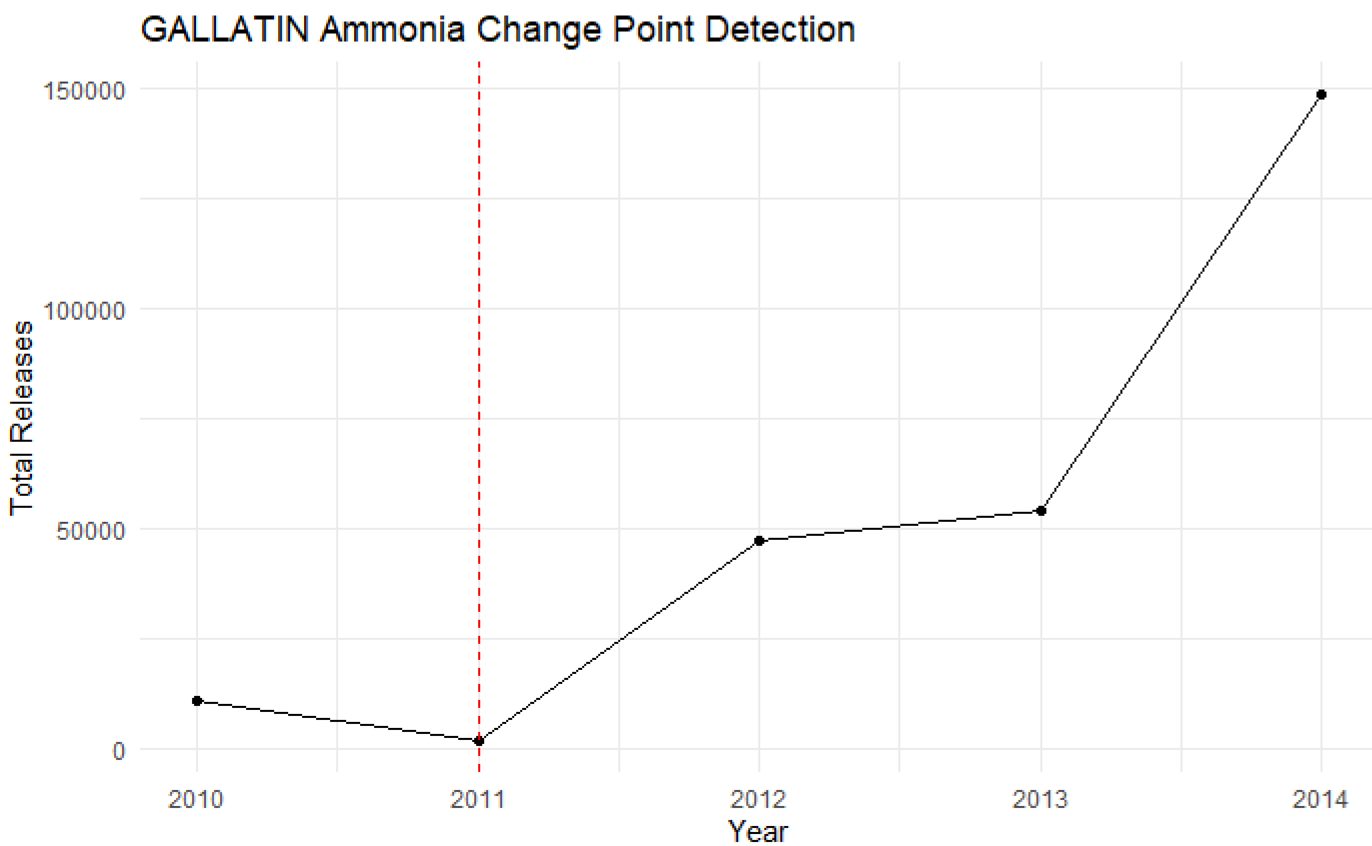


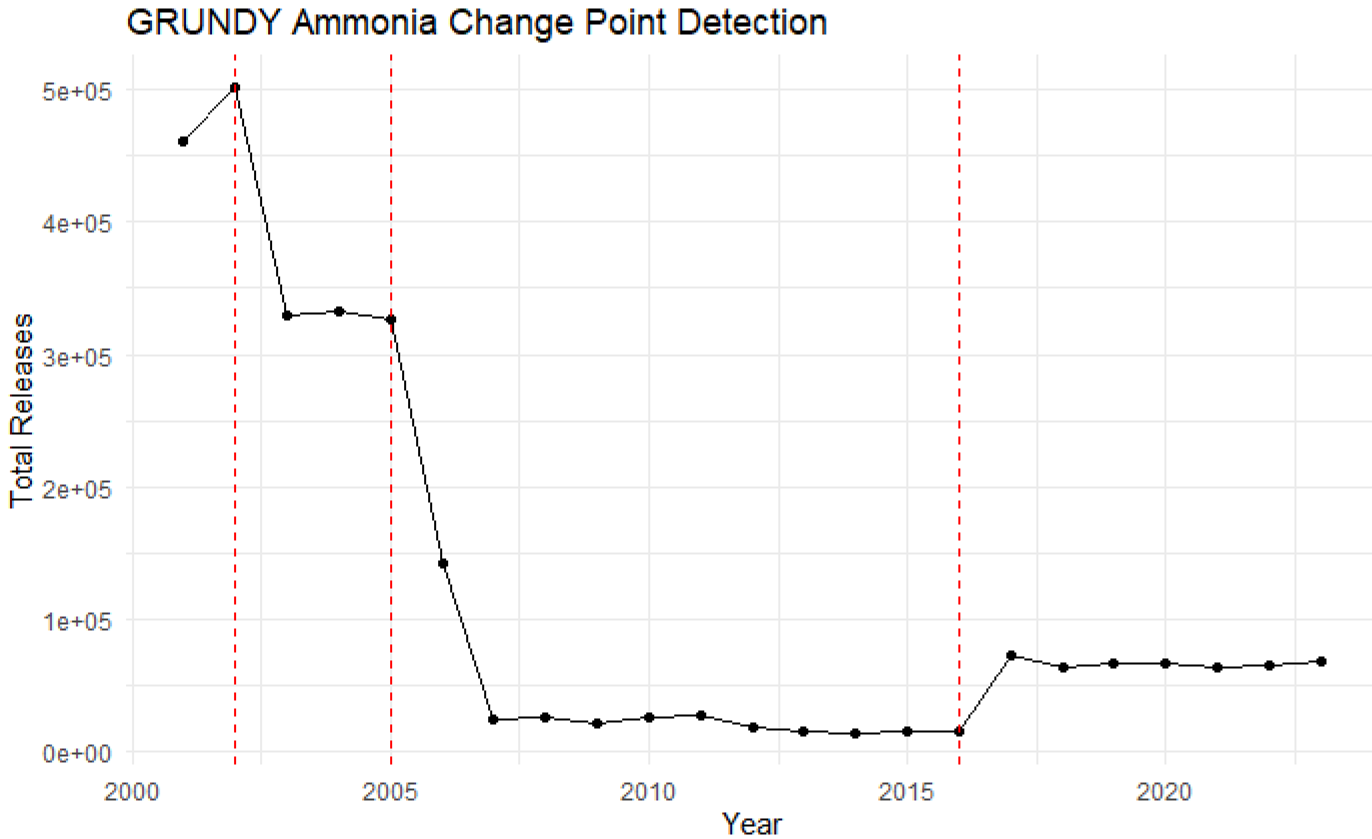
COOK Ammonia Change Point Detection
Total Releases
1600000
1200000
800000
400000
0
2000
2005
2010
2015
2020
Year

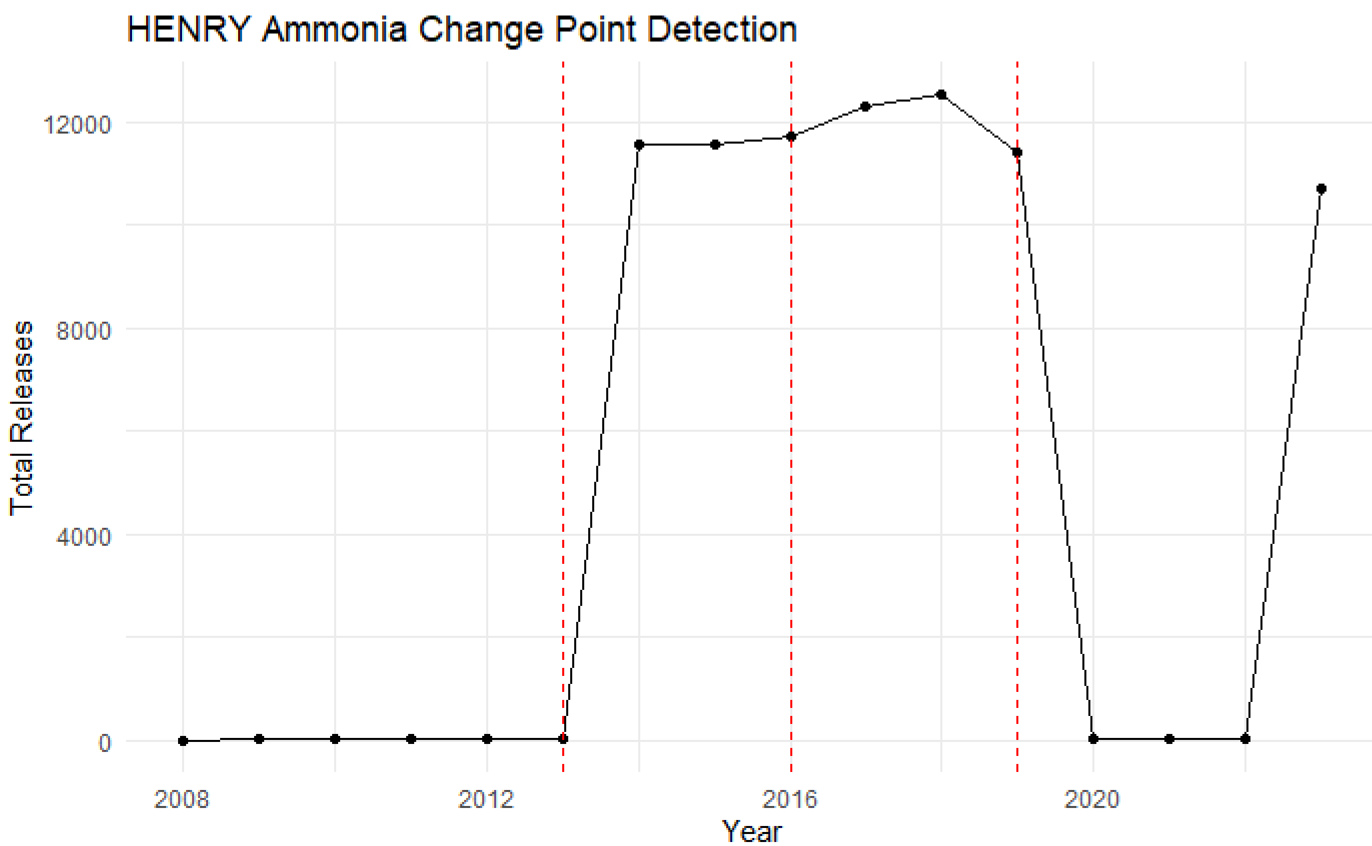

CRAWFORD Ammonia Change Point Detection
Total Releases
150000
100000
50000
2000
2005
2010
2015
2020
Year

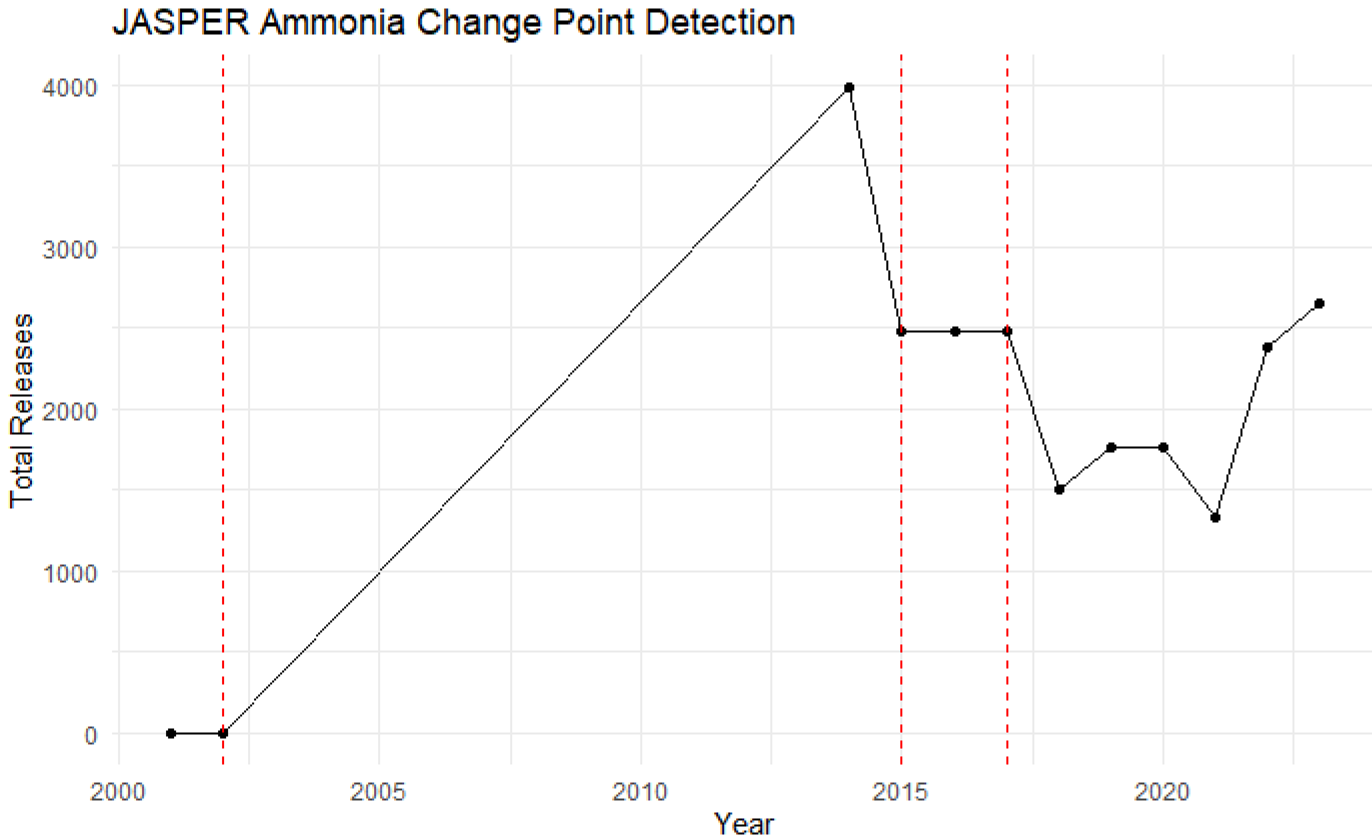

DUPAGE Ammonia Change Point Detection
Total Releases
20000
15000
10000
5000
0
2000
2005
2010
2015
2020
Year


FULTON Ammonia Change Point Detection
Total Releases
7500
5000
2500
0
2005
2010
2015
Year

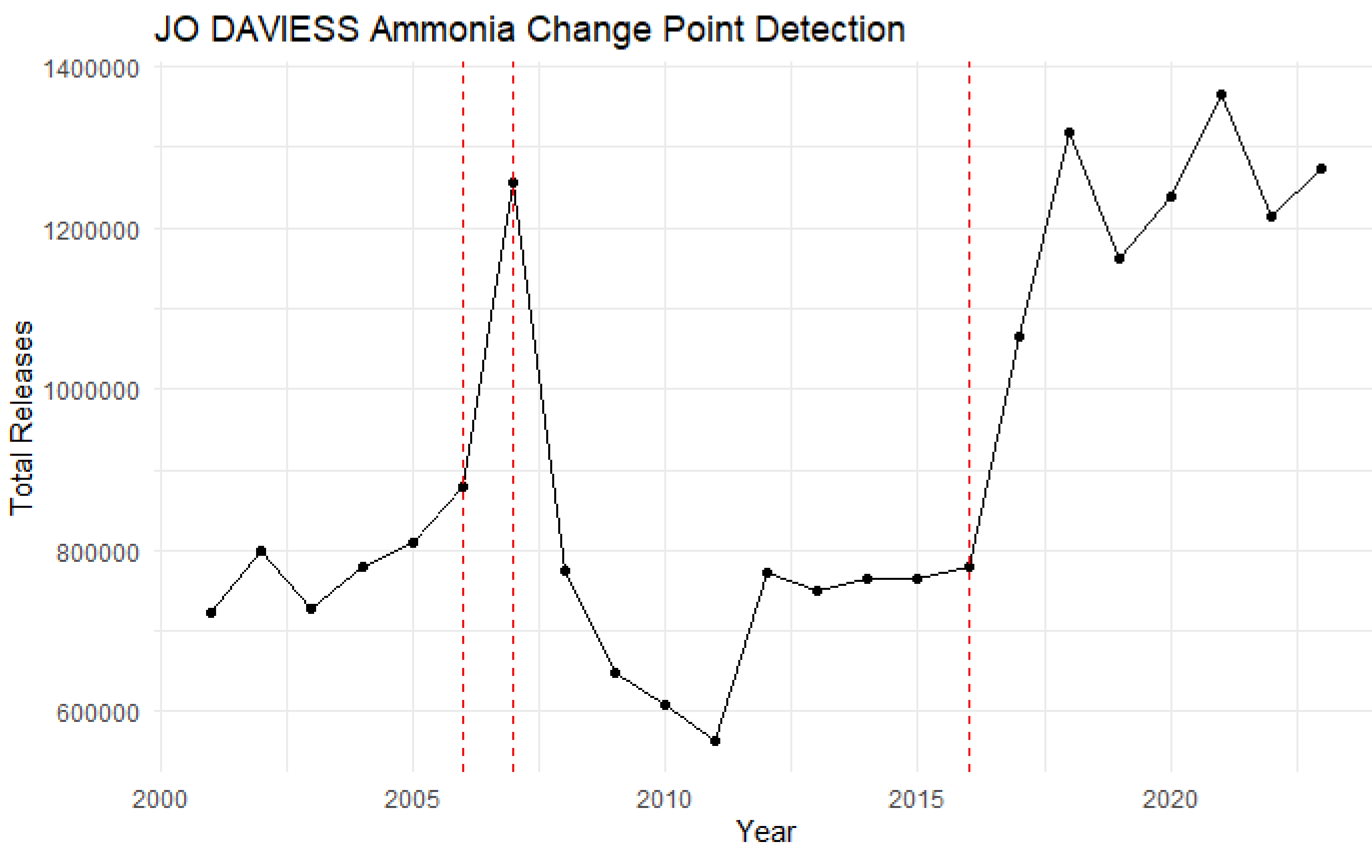

GALLATIN Ammonia Change Point Detection
Total Releases
150000
100000
50000
0
2010
2011
2012
2013
2014
Year


GRUNDY Ammonia Change Point Detection
Total Releases
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
2000
2005
2010
2015
2020
Year

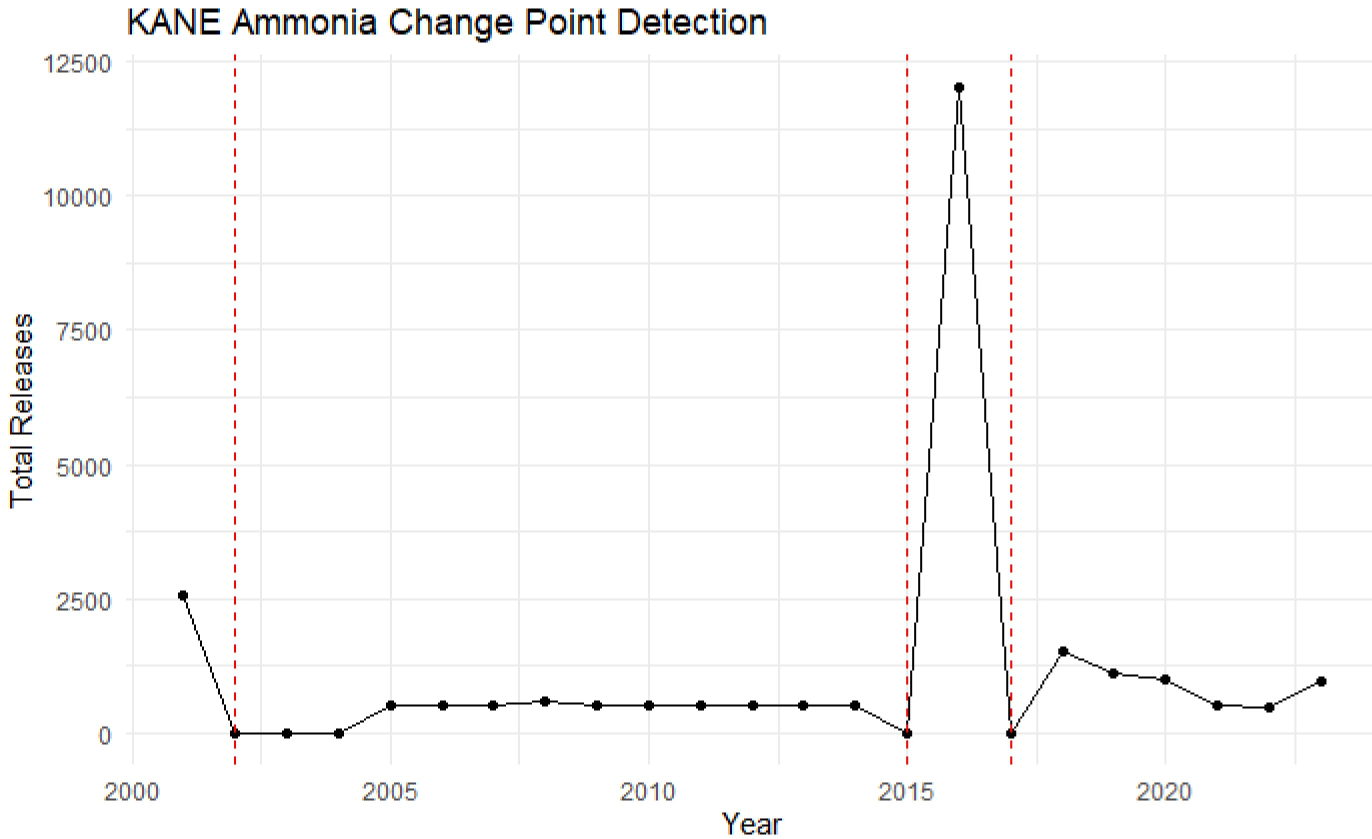

HENRY Ammonia Change Point Detection
Total Releases
12000
8000
4000
0
2008
2012
2016
2020
Year

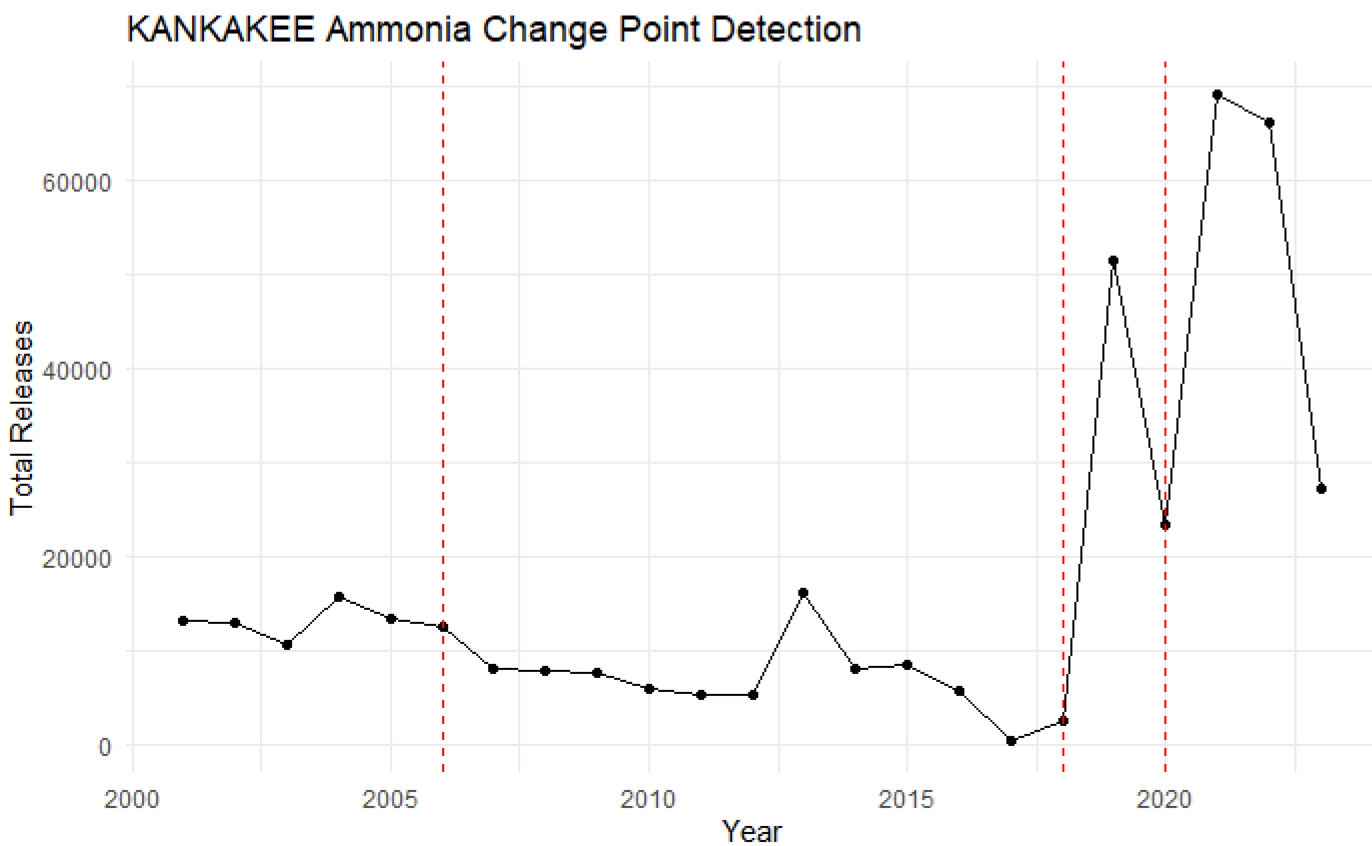


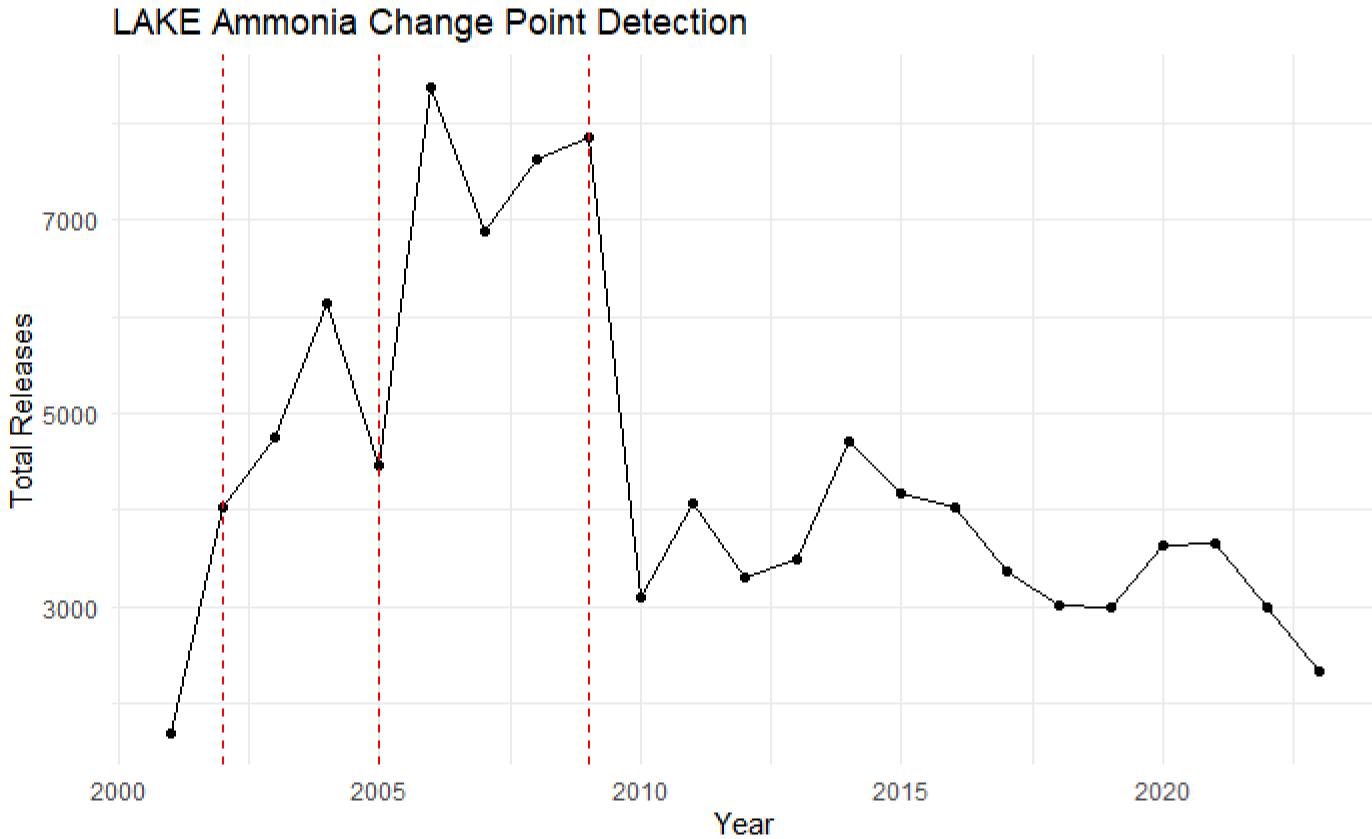
JASPER Ammonia Change Point Detection
Total Releases
Year
4000
3000
2000
1000
0
2000
2005
2010
2015
2020

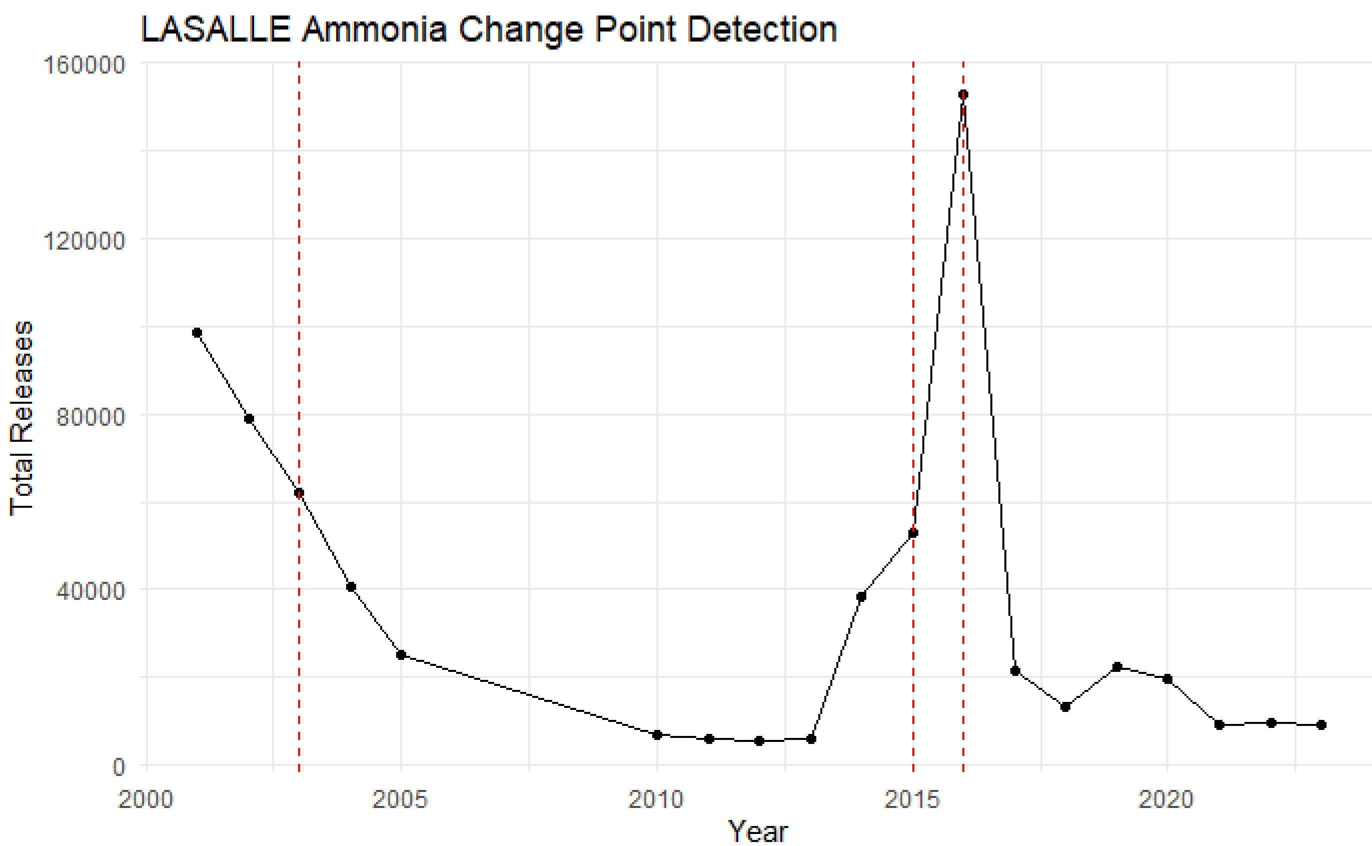


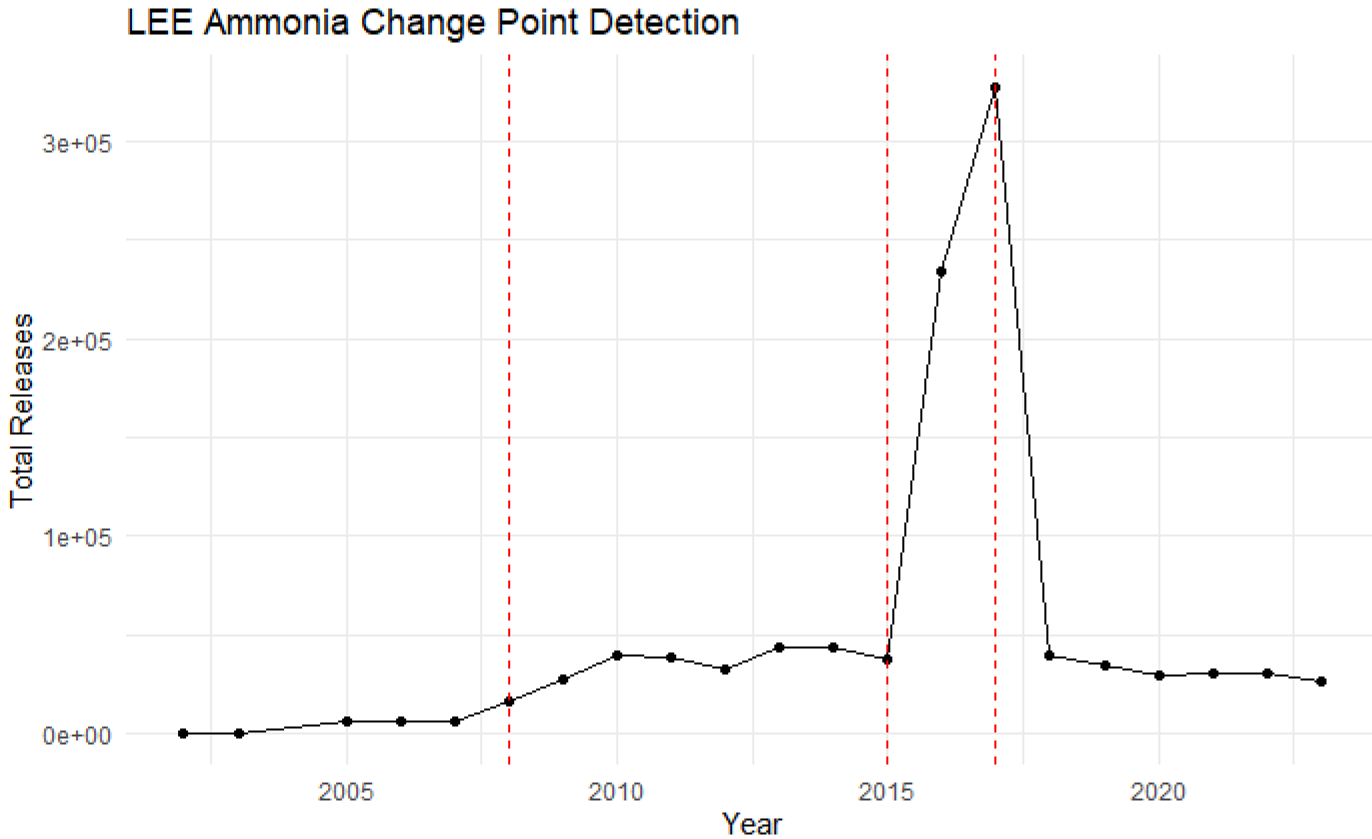
JO DAVIESS Ammonia Change Point Detection
Total Releases
Year
1400000
1200000
1000000
800000
600000
2000
2005
2010
2015
2020

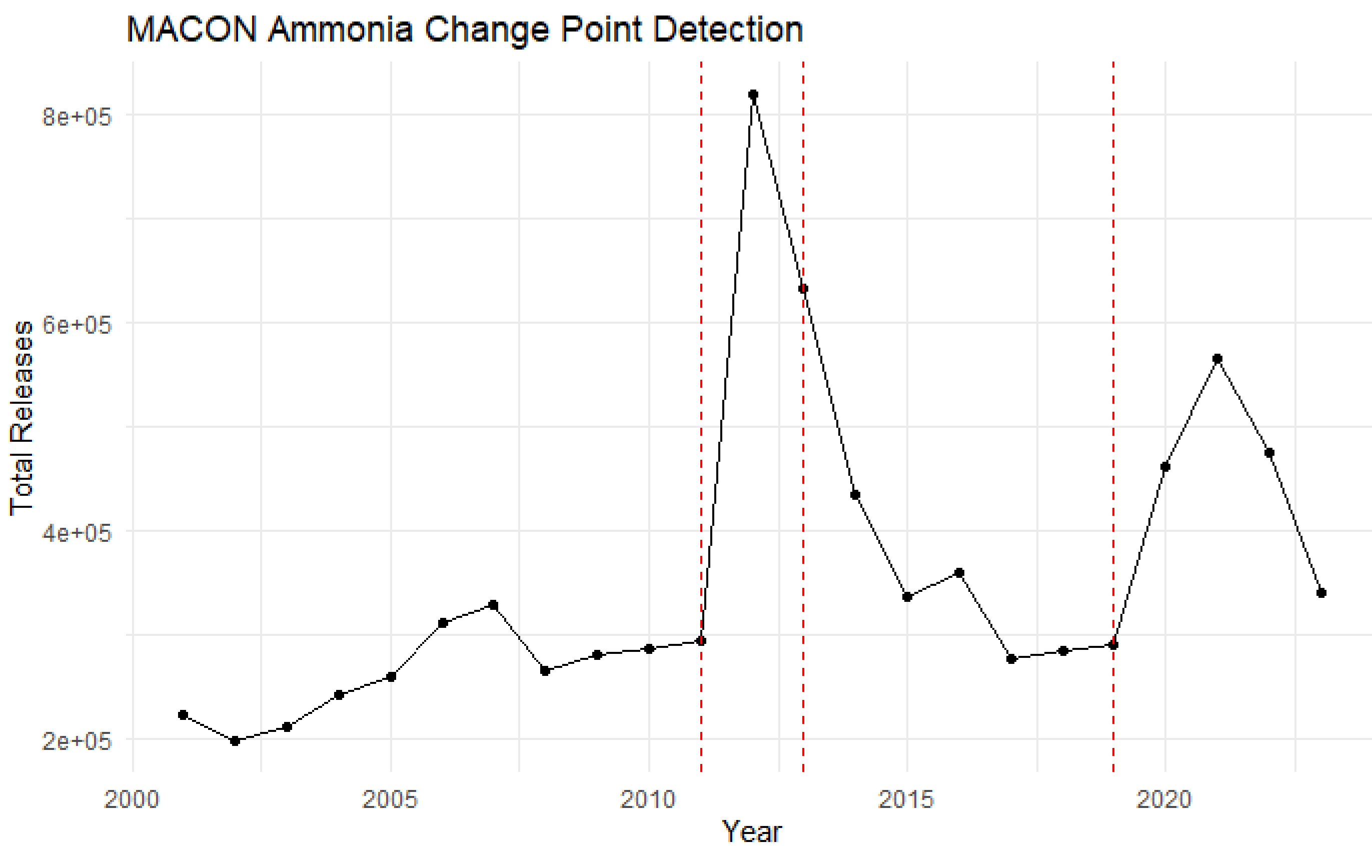

KANE Ammonia Change Point Detection
12500
10000
7500
5000
2500
0
Total Releases
2000
2005
2010
2015
2020
Year

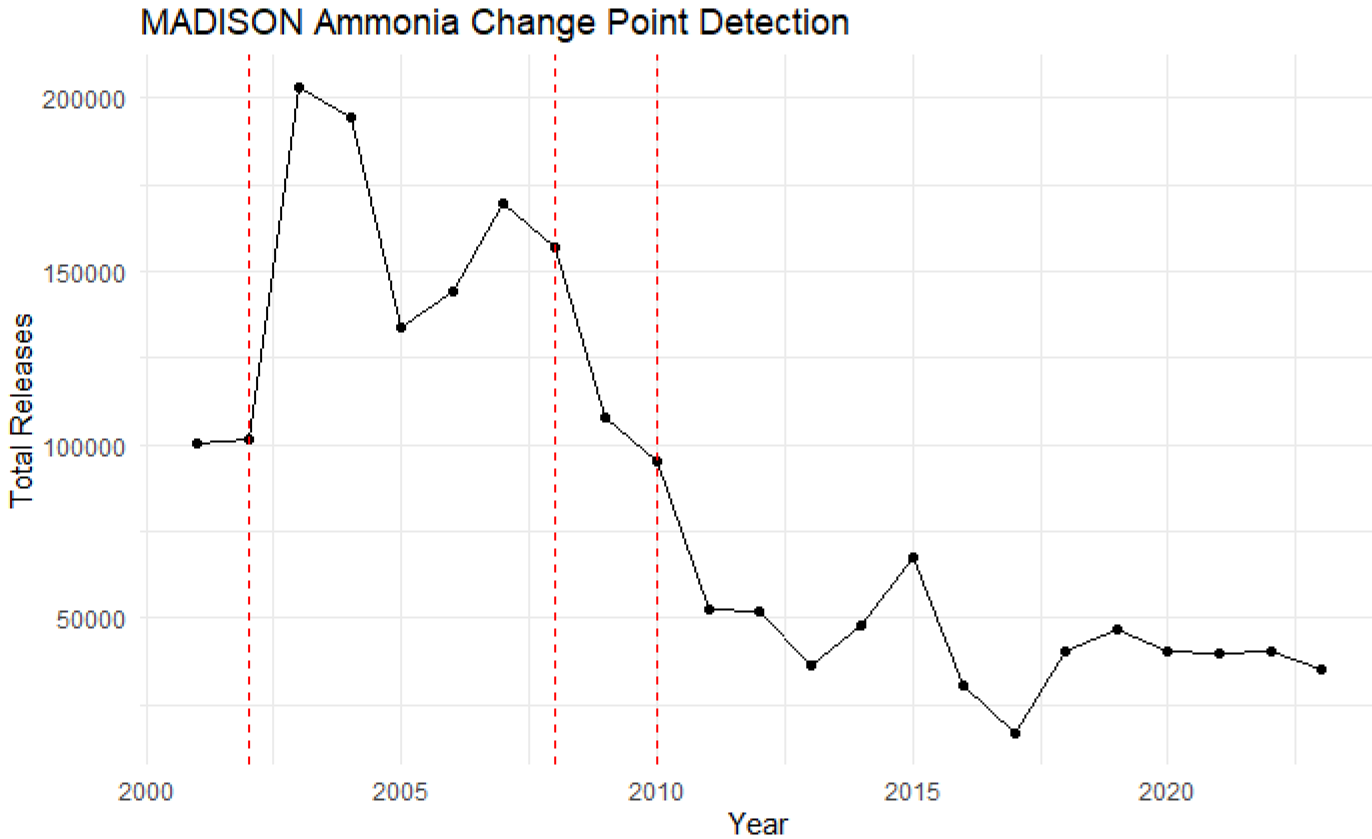

KANKAKEE Ammonia Change Point Detection
60000
40000
20000
0
Total Releases
2000
2005
2010
2015
2020
Year

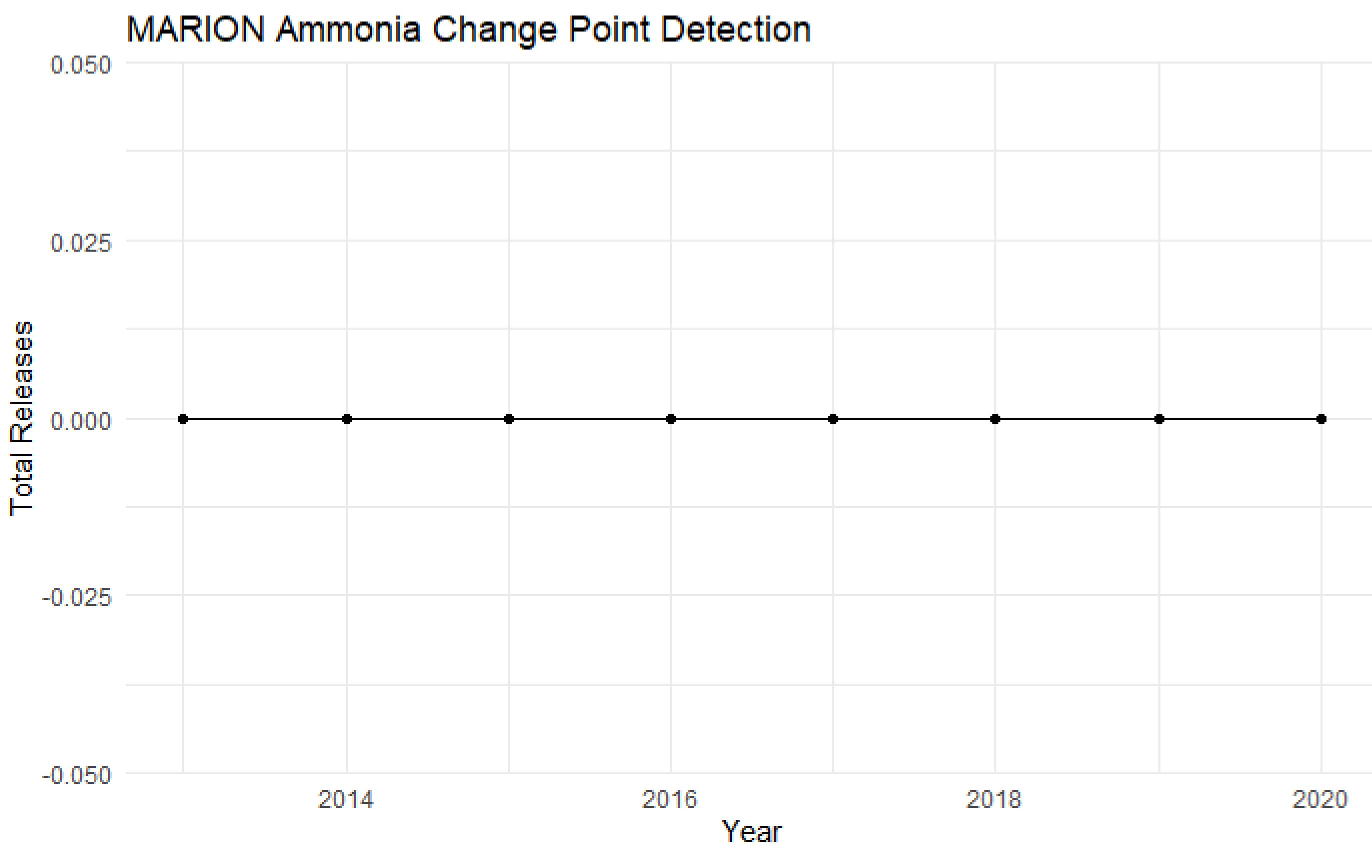

LAKE Ammonia Change Point Detection
Total Releases
7000
5000
3000
2000
2005
2010
2015
2020
Year

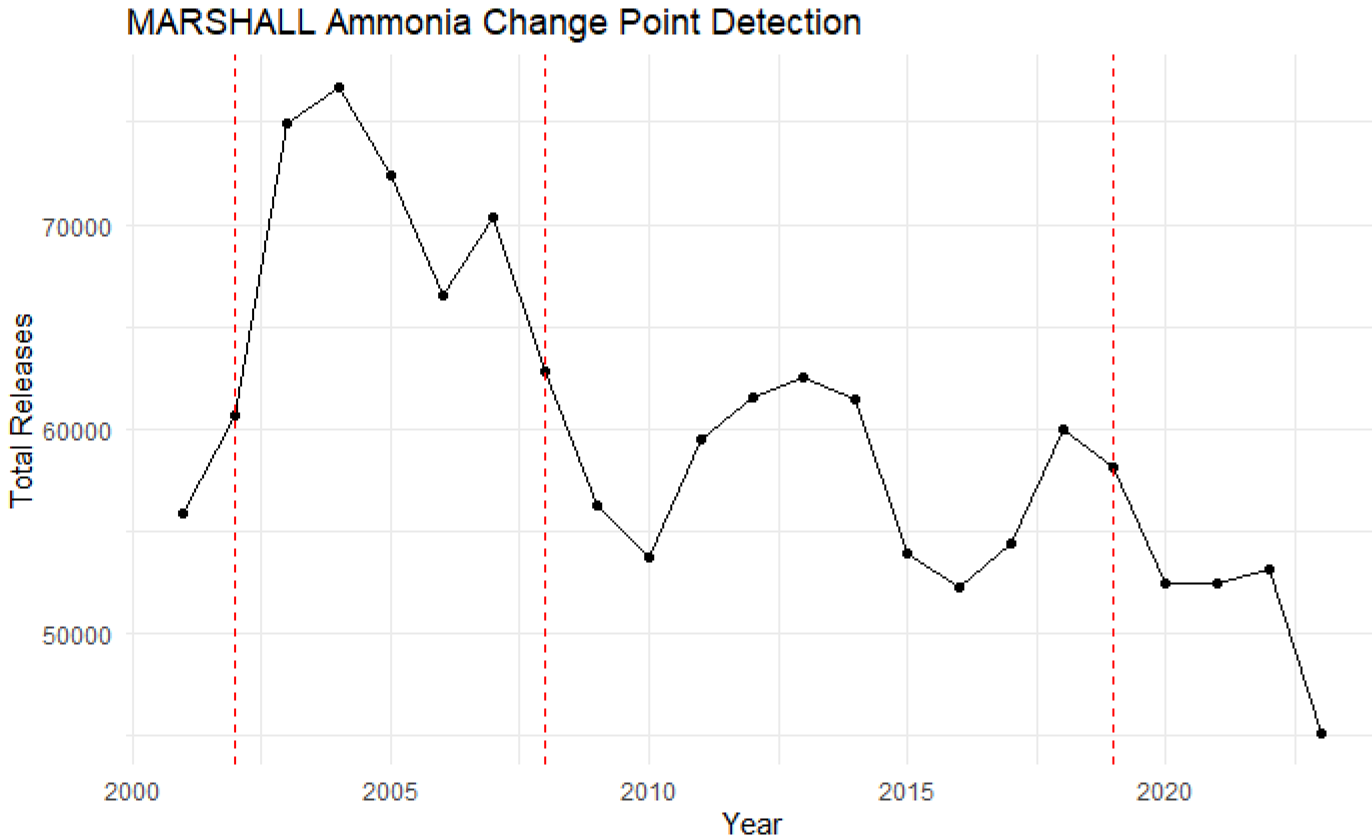

LASALLE Ammonia Change Point Detection
Total Releases
160000
120000
80000
40000
0
2000
2005
2010
2015
2020
Year

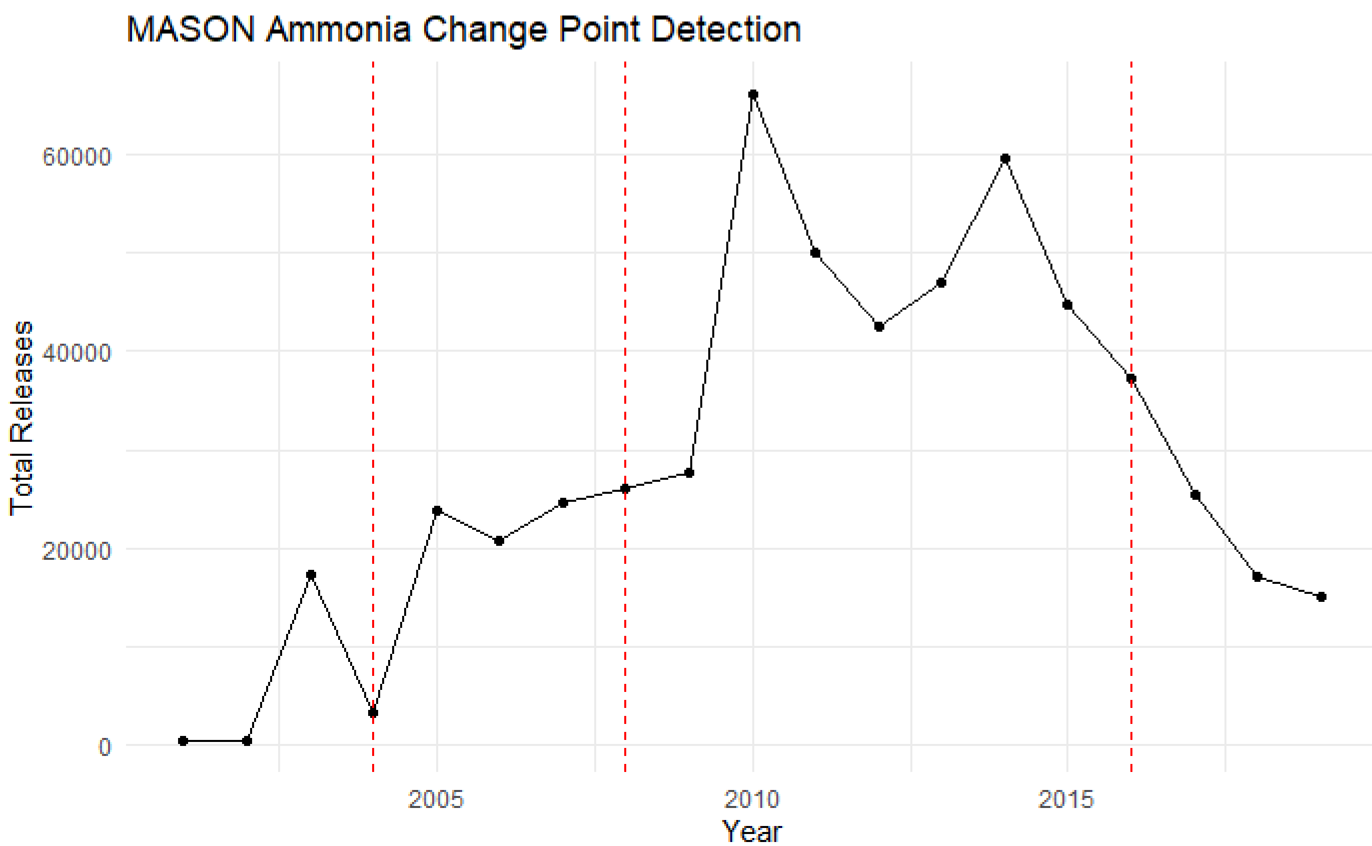

LEE Ammonia Change Point Detection
Total Releases
3e+05
2e+05
1e+05
0e+00
2005
2010
2015
2020
Year


MACON Ammonia Change Point Detection
Total Releases
8e+05
6e+05
4e+05
2e+05
2000
2005
2010
2015
2020
Year

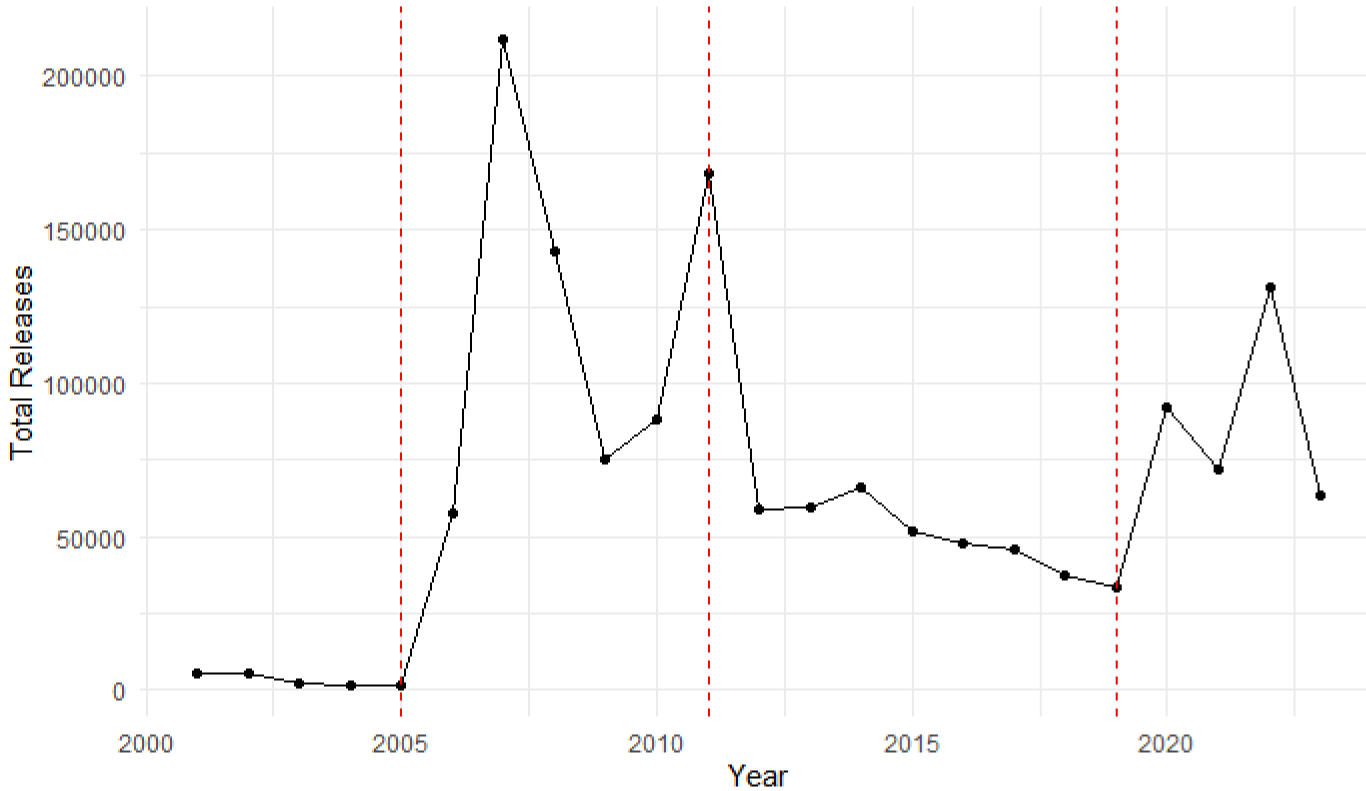

MADISON Ammonia Change Point Detection
200000
150000
100000
50000
Total Releases
2000
2005
2010
2015
2020
Year


MARION Ammonia Change Point Detection
0.050
0.025
0.000
-0.025
-0.050
Total Releases
2014
2016
2018
2020
Year

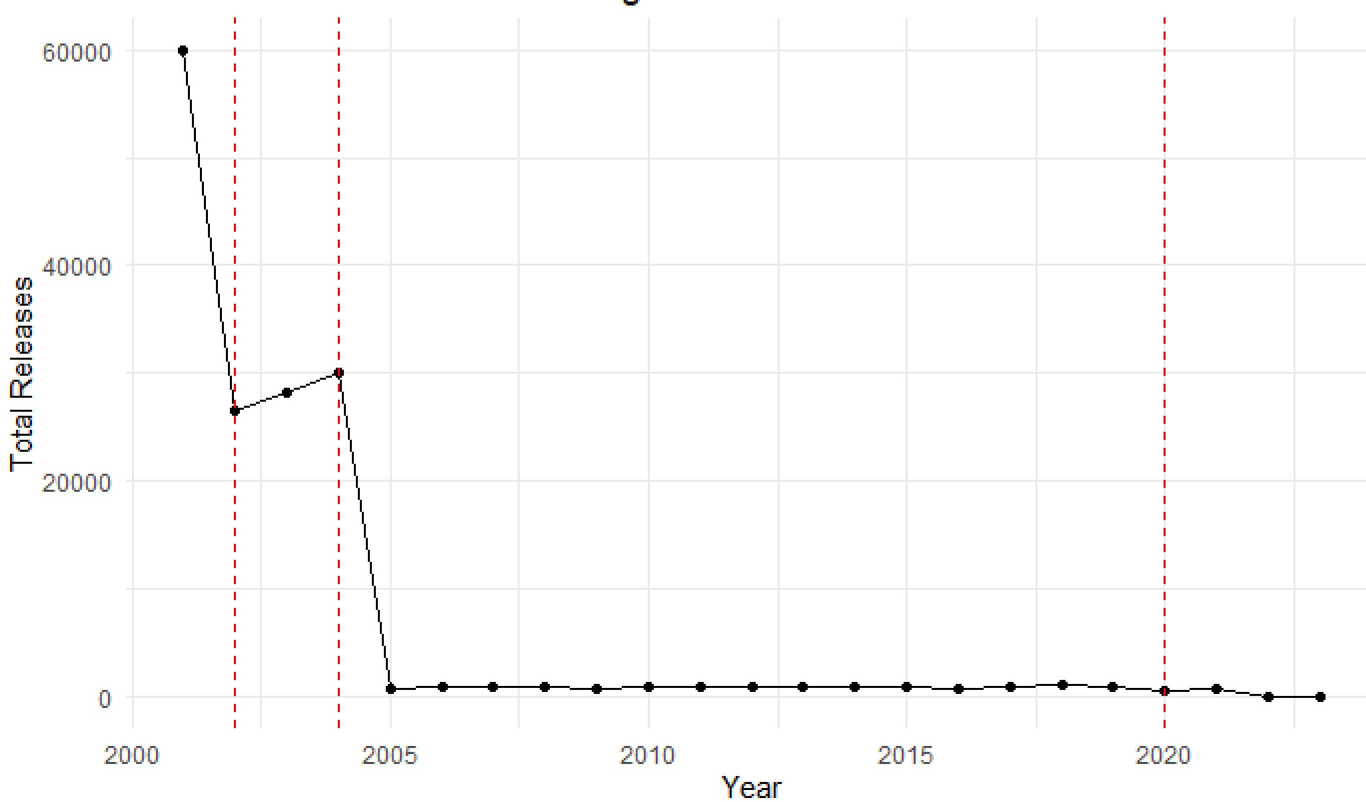

MARSHALL Ammonia Change Point Detection
Total Releases
50000
60000
70000
2000
2005
2010
2015
2020
Year

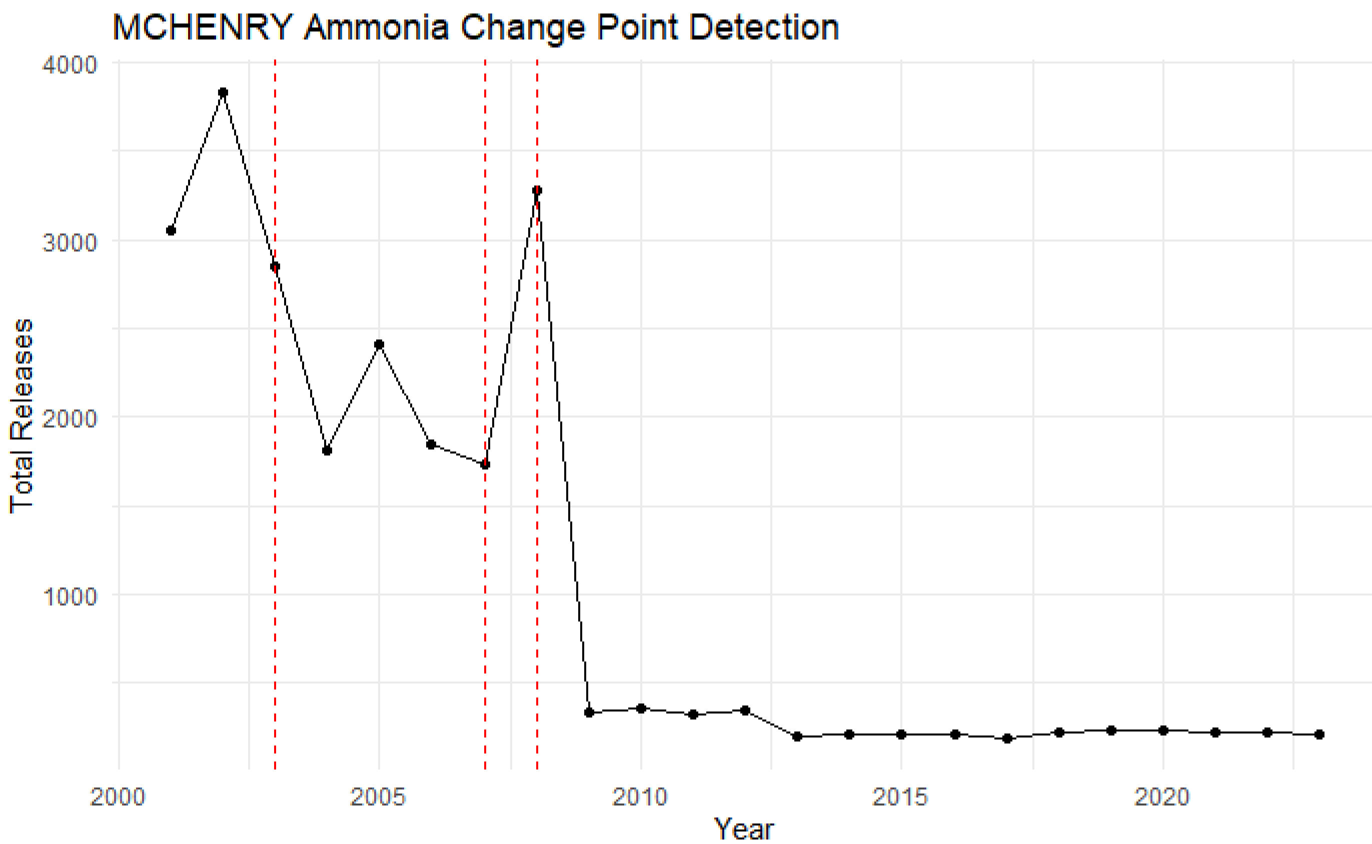

MASON Ammonia Change Point Detection
Total Releases
0
20000
40000
60000
2005
2010
2015
Year

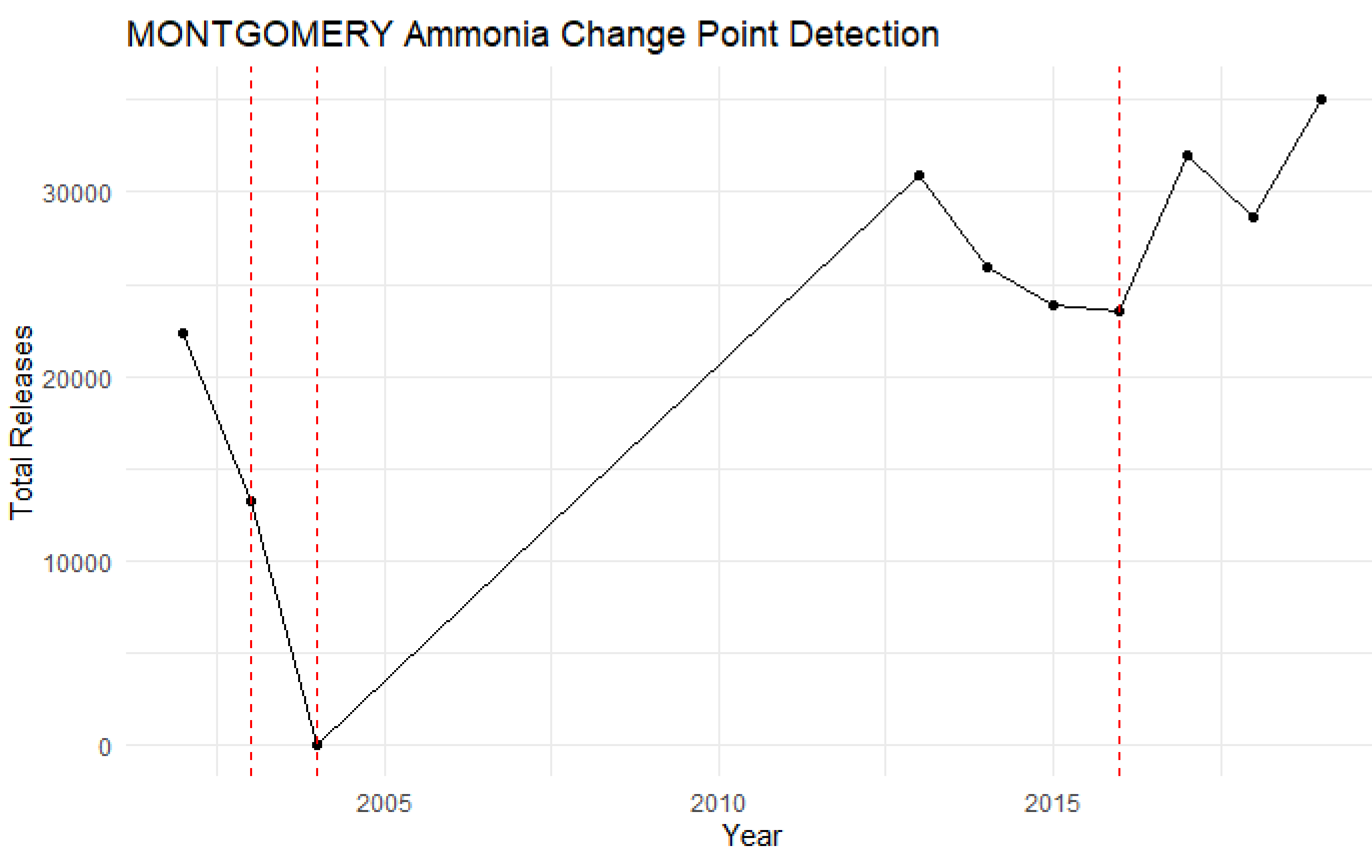
MASSAC Ammonia Change Point Detection

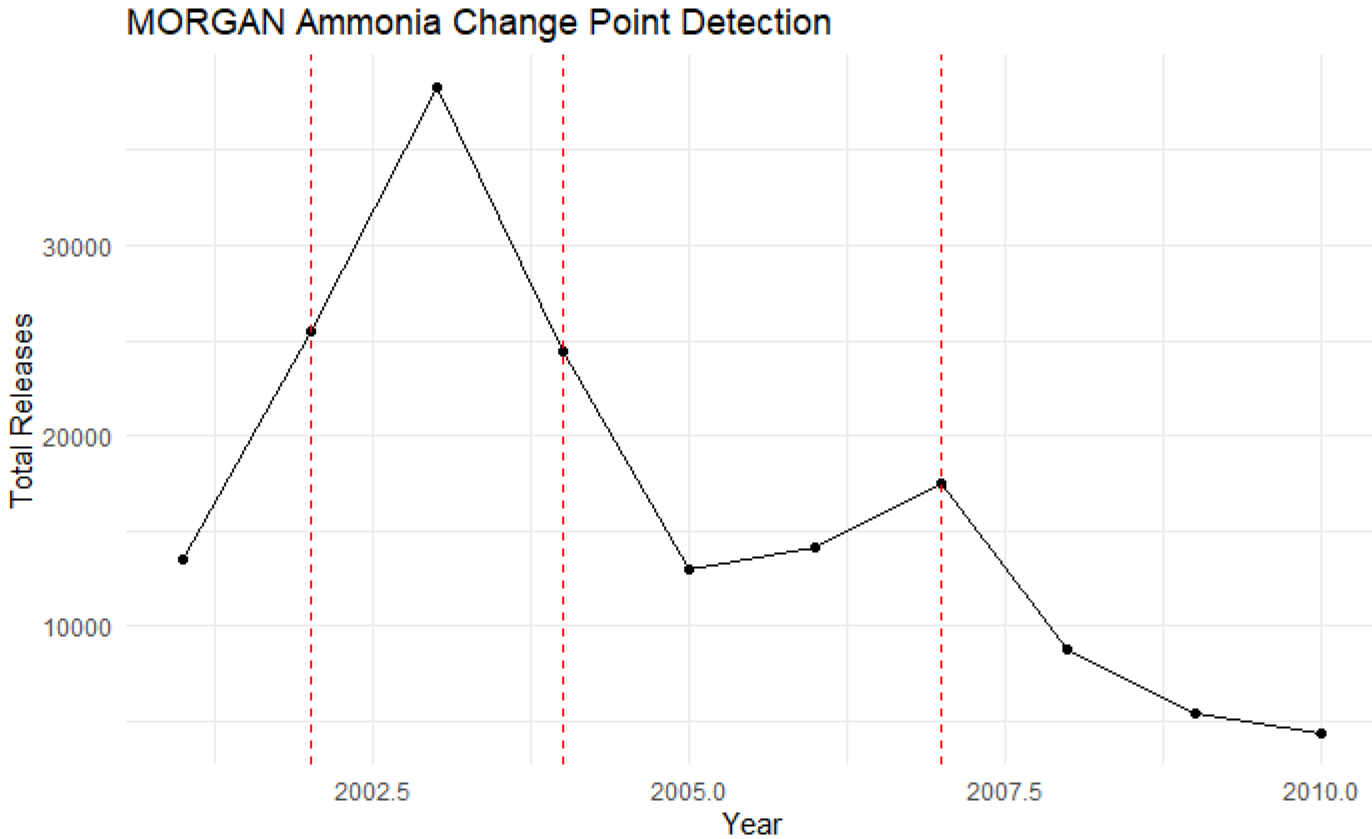


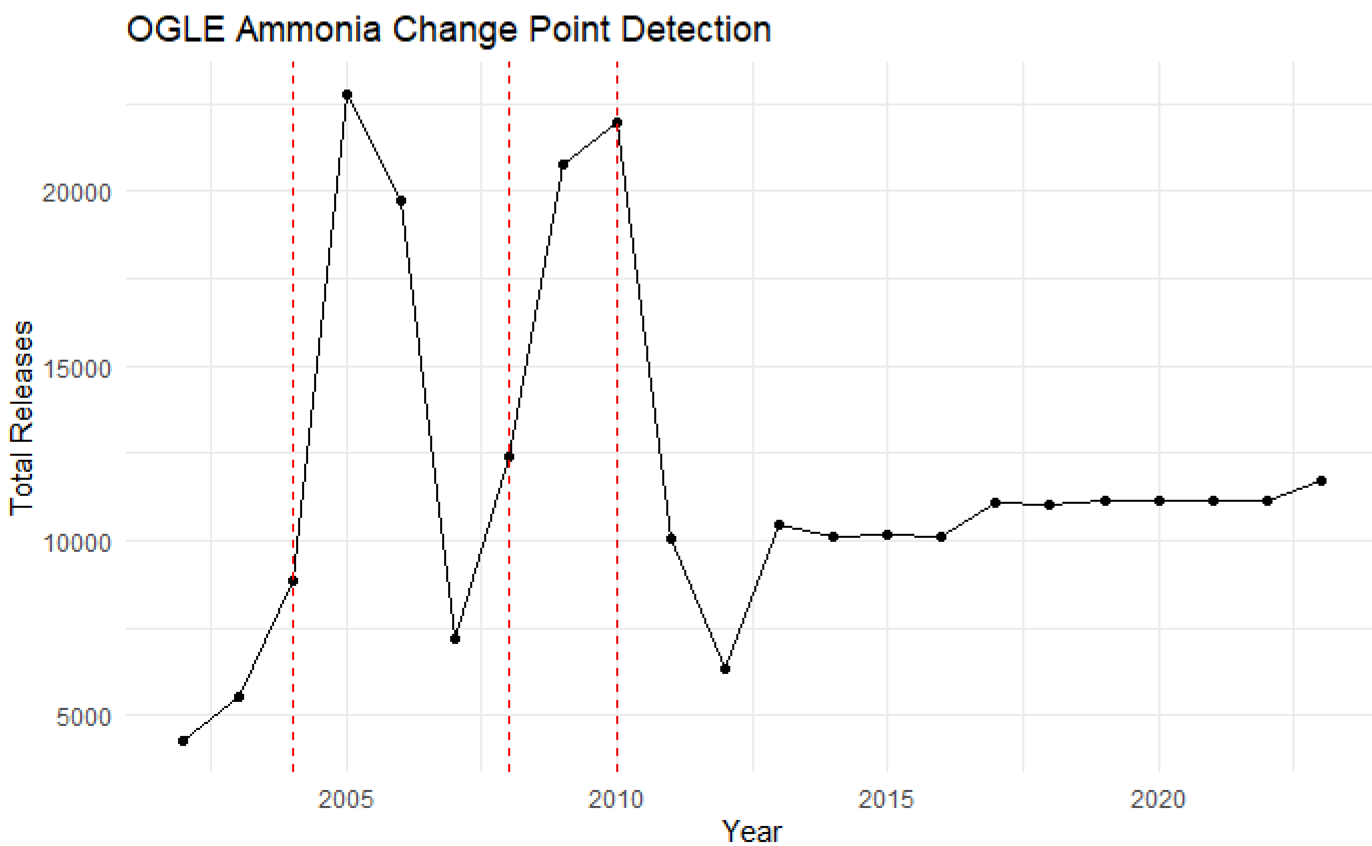
MCDONOUGH Ammonia Change Point Detection

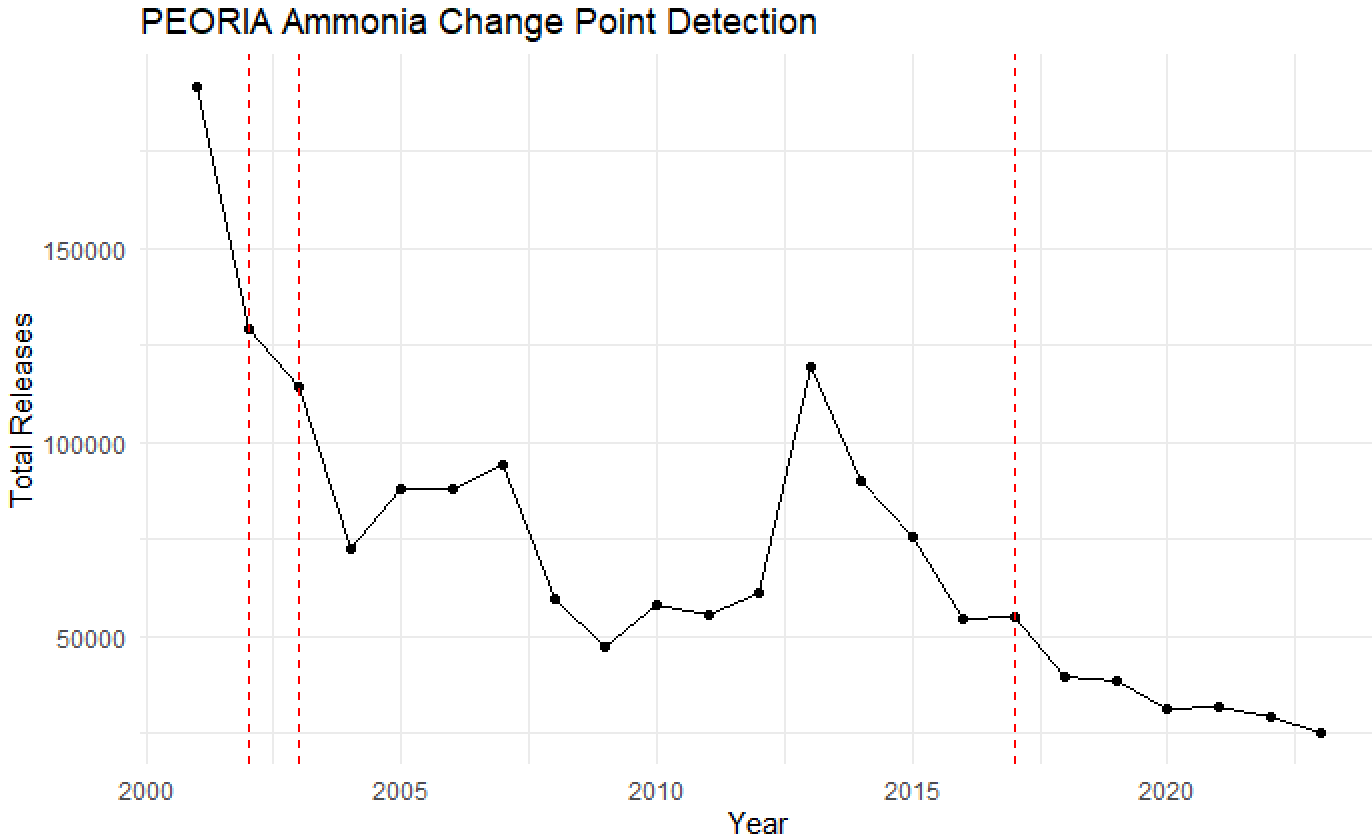


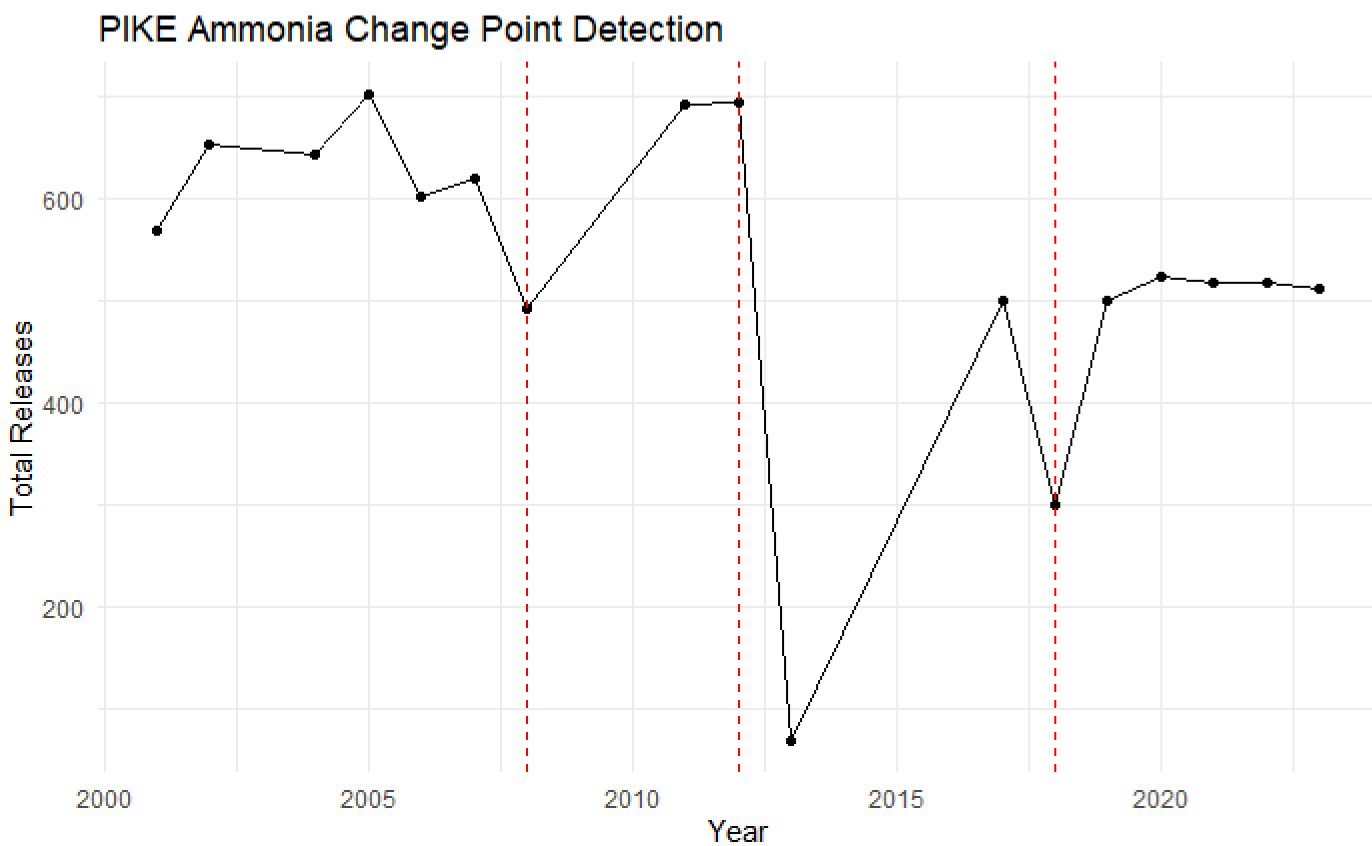

MCHENRY Ammonia Change Point Detection
Total Releases
4000
3000
2000
1000
2000
2005
2010
2015
2020
Year

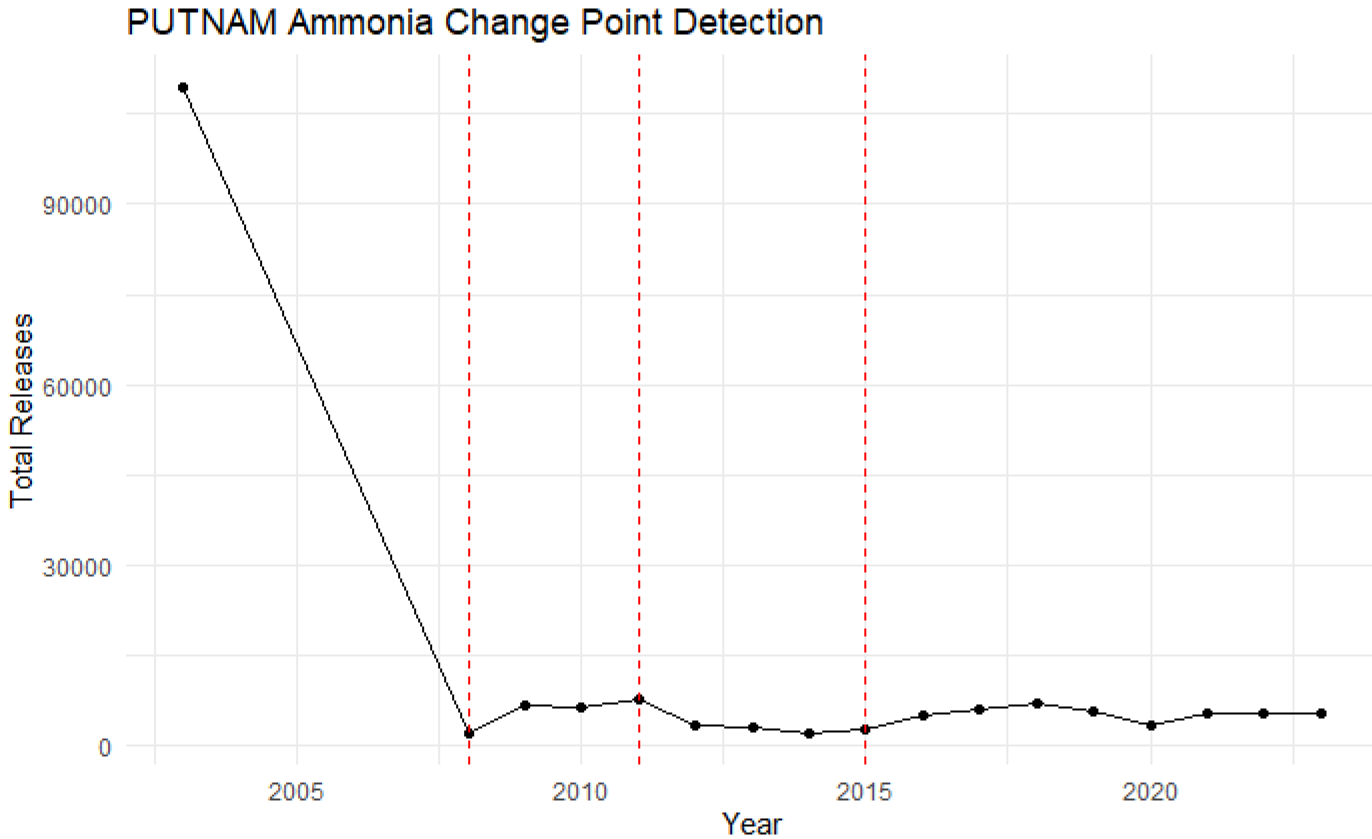

MONTGOMERY Ammonia Change Point Detection
Total Releases
30000
20000
10000
0
2005
2010
2015
Year

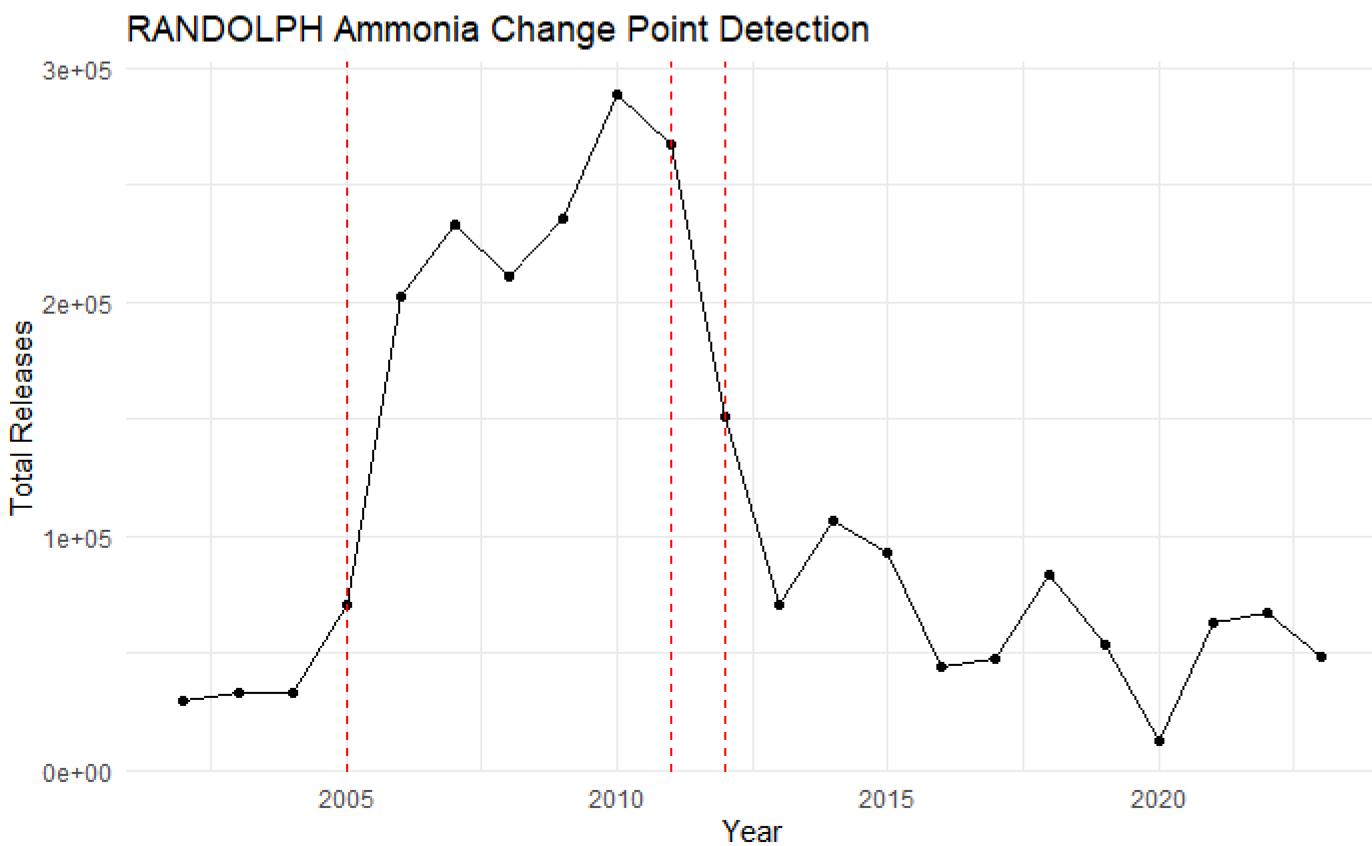

MORGAN Ammonia Change Point Detection
30000
20000
10000
Total Releases
2002.5
2005.0
2007.5
2010.0
Year

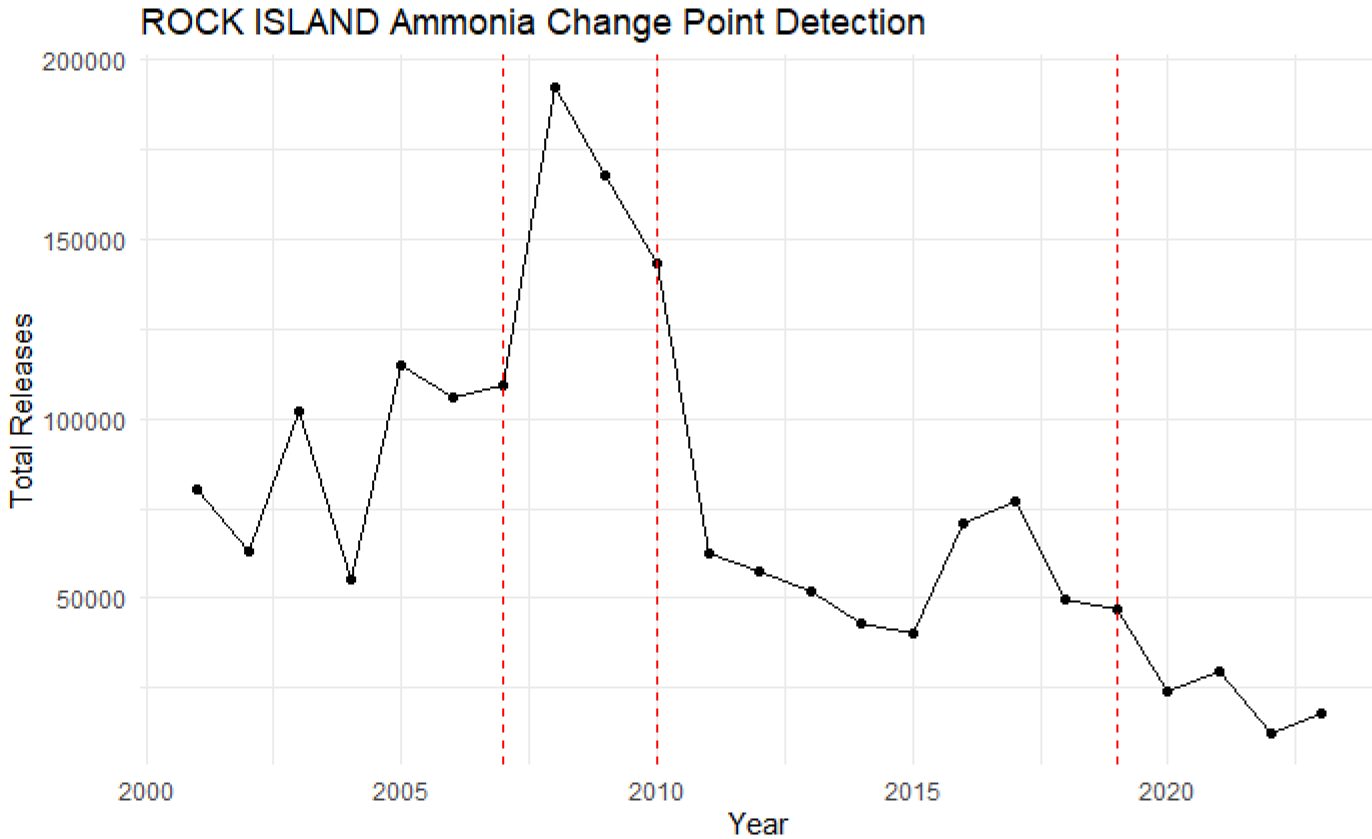

OGLE Ammonia Change Point Detection
20000
15000
10000
5000
Total Releases
2005
2010
2015
2020
Year

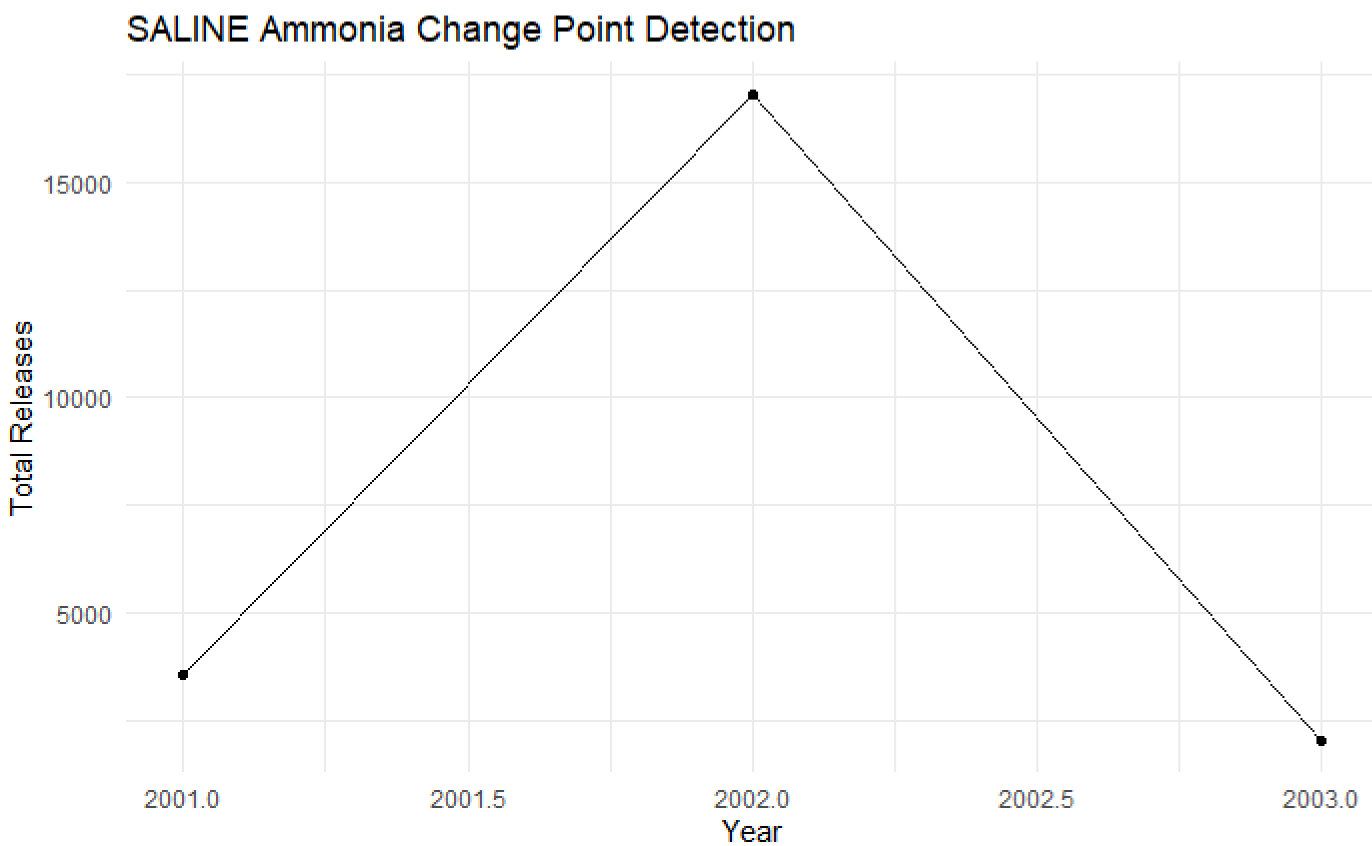

PEORIA Ammonia Change Point Detection
Total Releases
150000
100000
50000
2000
2005
2010
2015
2020
Year

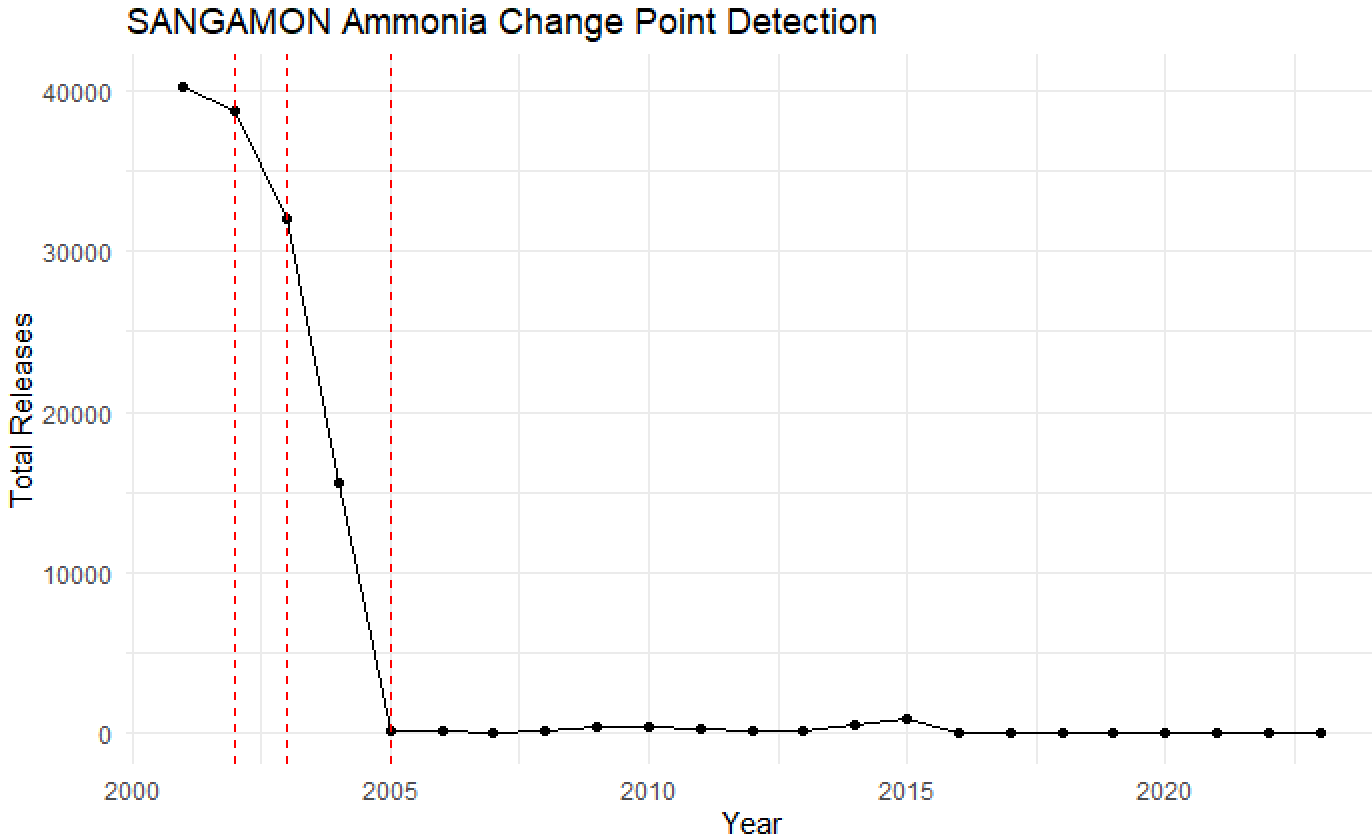

PIKE Ammonia Change Point Detection
Total Releases
600
400
200
2000
2005
2010
2015
2020
Year

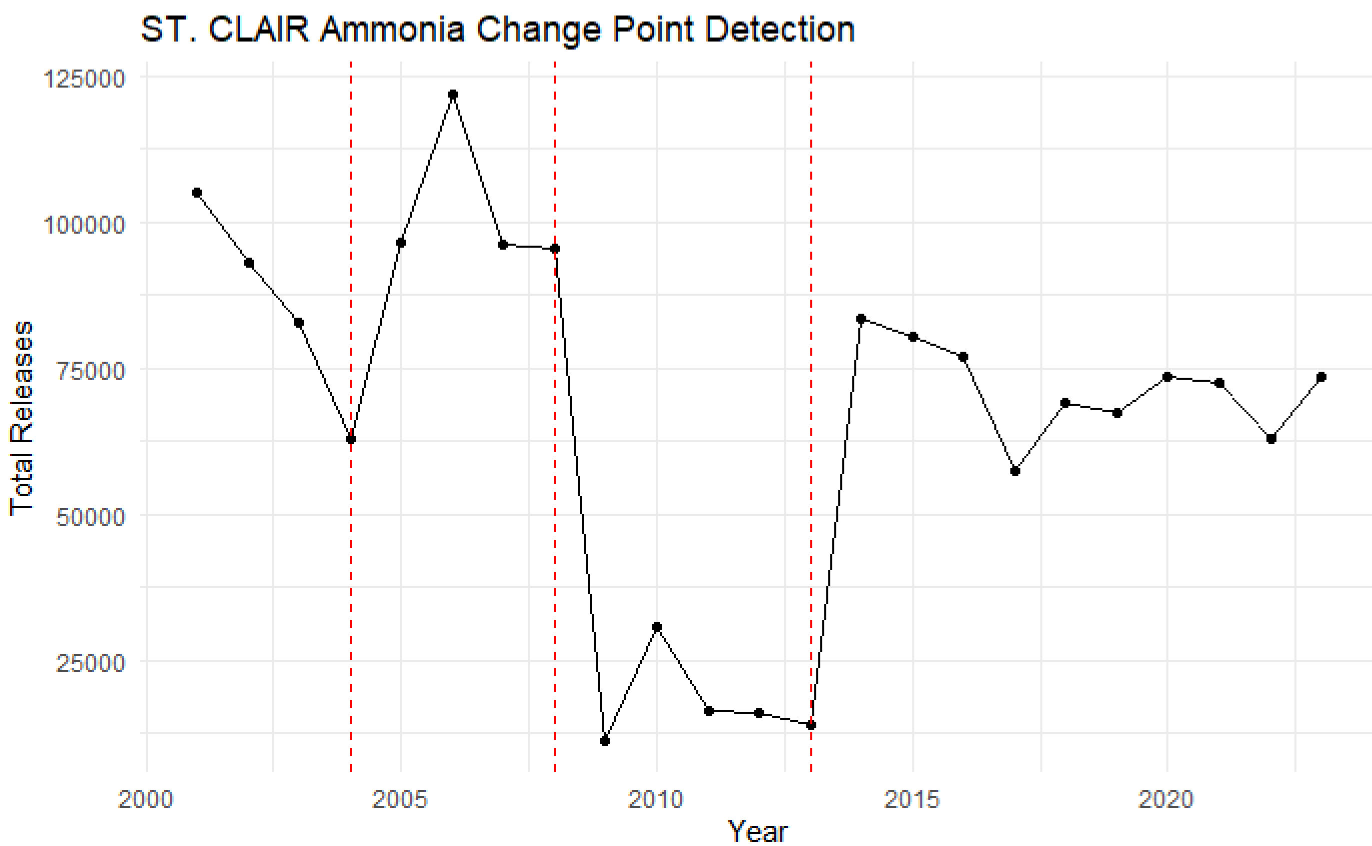

PUTNAM Ammonia Change Point Detection
Total Releases
90000
60000
30000
0
2005
2010
2015
2020
Year

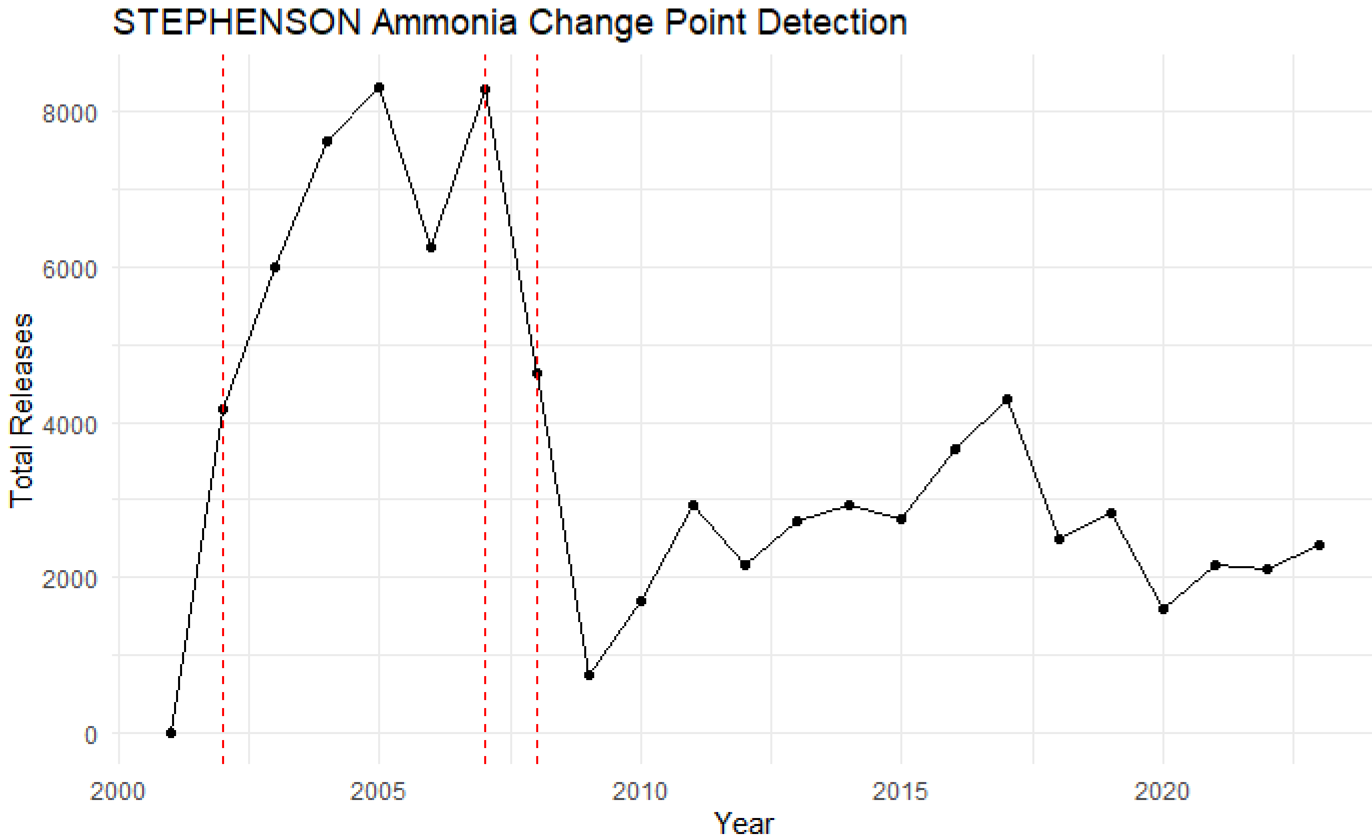

RANDOLPH Ammonia Change Point Detection
Total Releases
3e+05
2e+05
1e+05
0e+00
2005
2010
2015
2020
Year

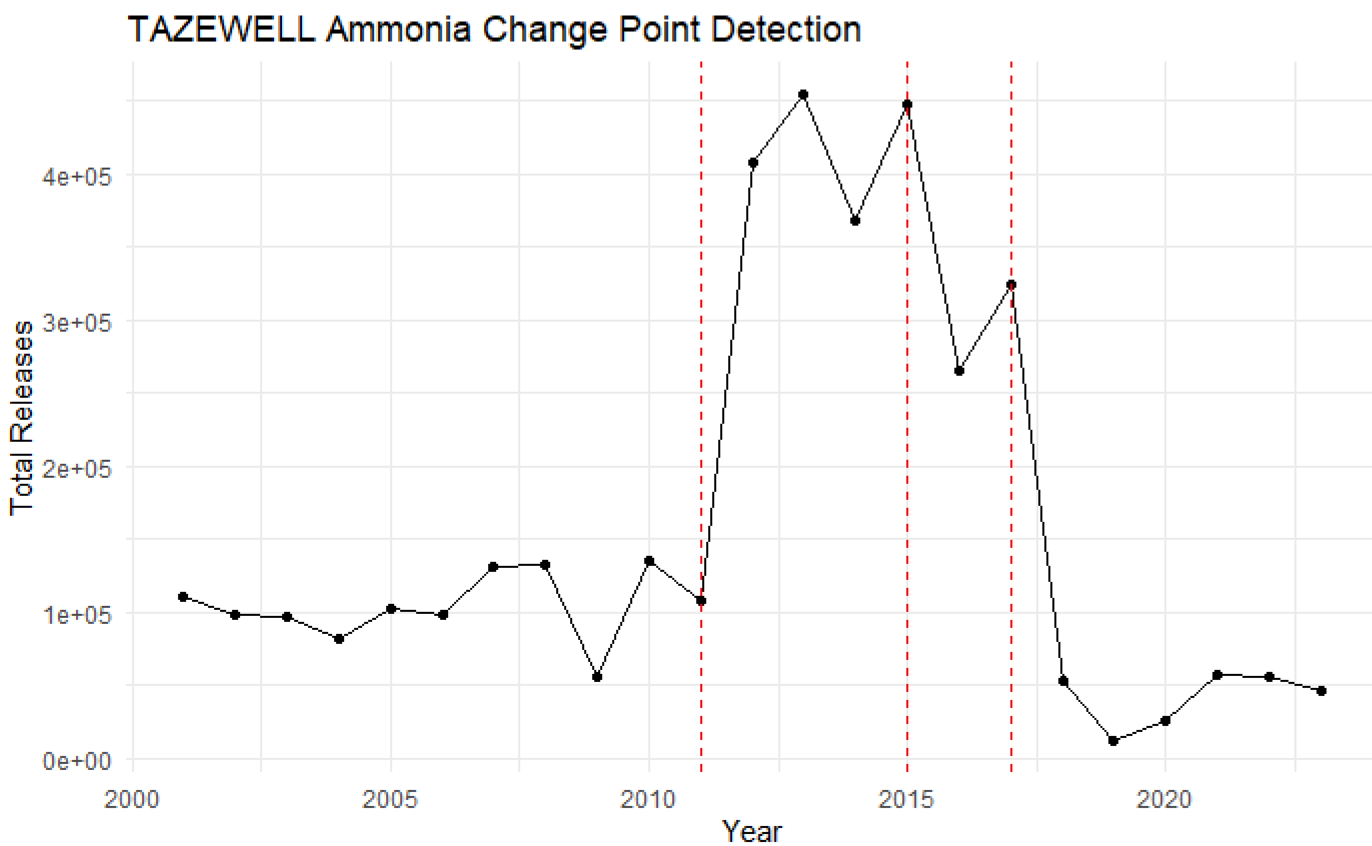

ROCK ISLAND Ammonia Change Point Detection
Total Releases
200000
150000
100000
50000
2000
2005
2010
2015
2020
Year

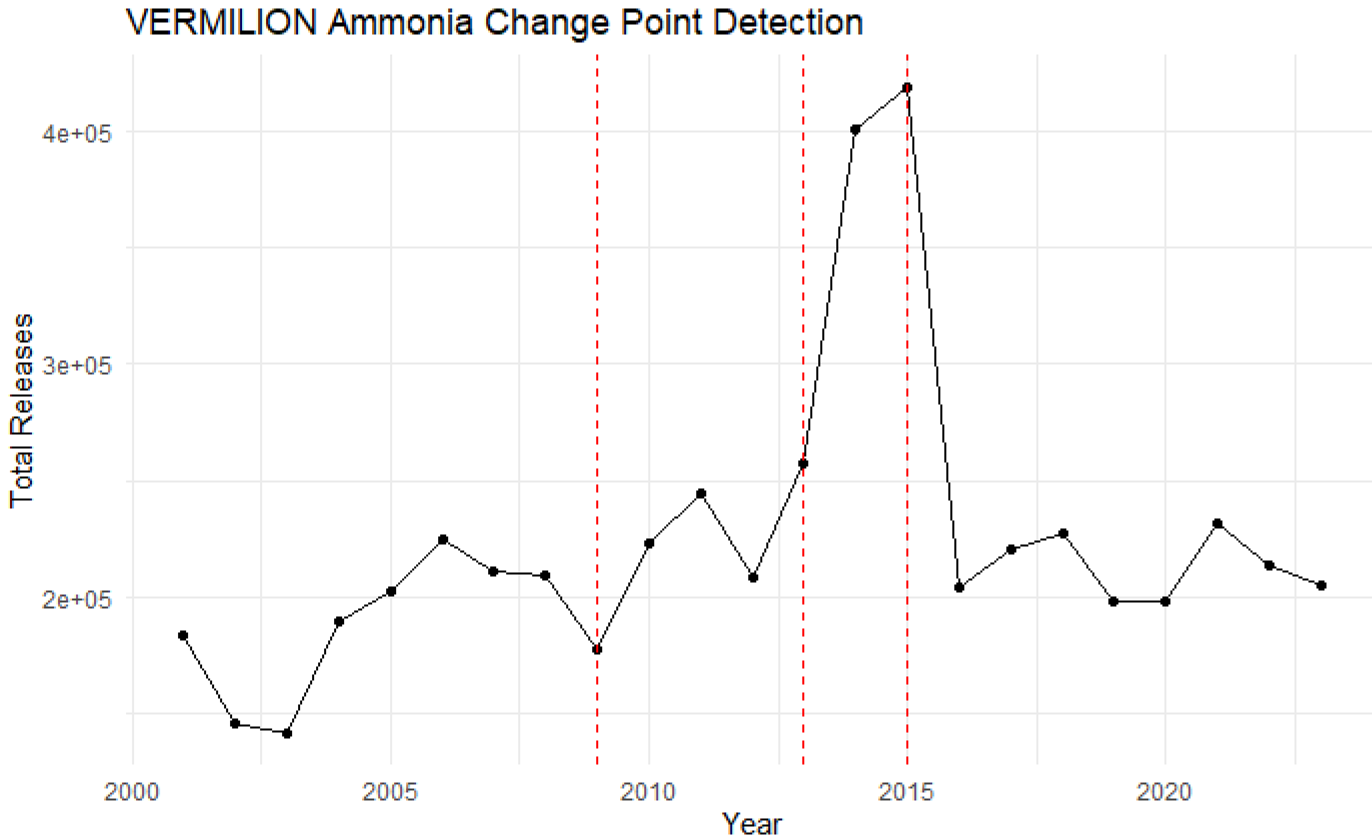

SALINE Ammonia Change Point Detection
Total Releases
15000
10000
5000
2001.0
2001.5
2002.0
2002.5
2003.0
Year


SANGAMON Ammonia Change Point Detection
Total Releases
40000
30000
20000
10000
0
2000
2005
2010
2015
2020
Year

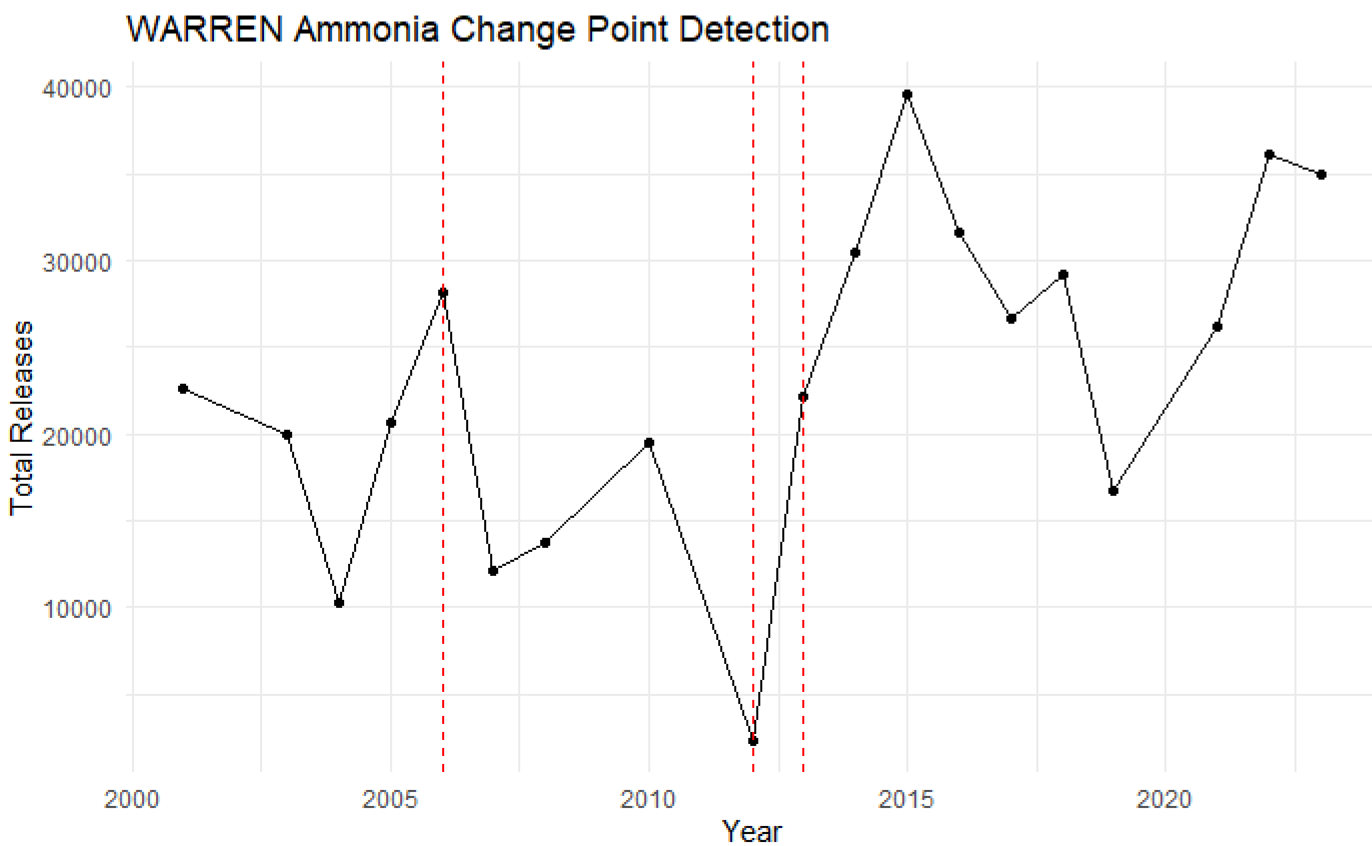

ST. CLAIR Ammonia Change Point Detection
Total Releases
125000
100000
75000
50000
25000
2000
2005
2010
2015
2020
Year


STEPHENSON Ammonia Change Point Detection
Total Releases
0
2000
4000
6000
8000
2000
2005
2010
2015
2020
Year

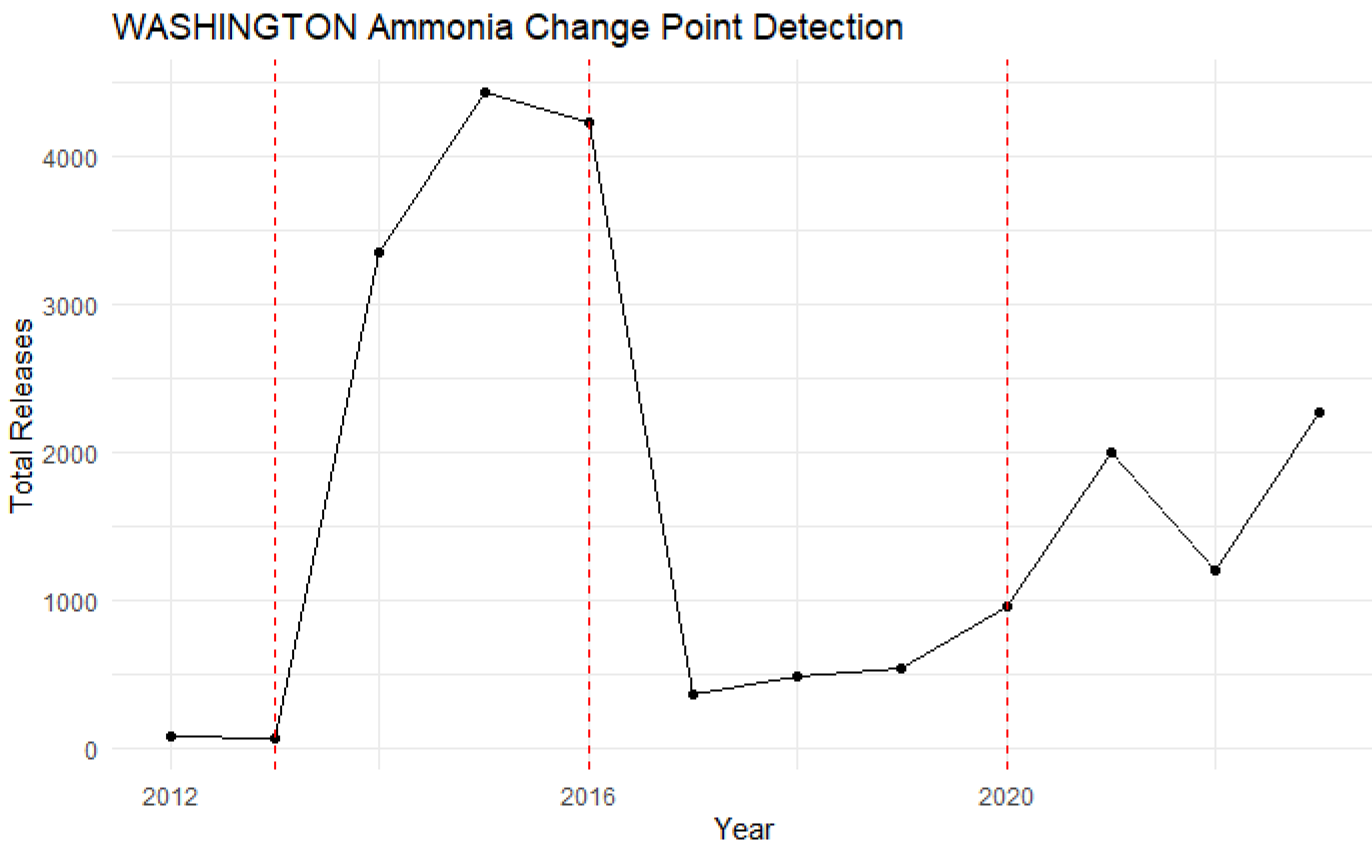

TAZEWELL Ammonia Change Point Detection
Total Releases
0e+00
1e+05
2e+05
3e+05
4e+05
2000
2005
2010
2015
2020
Year

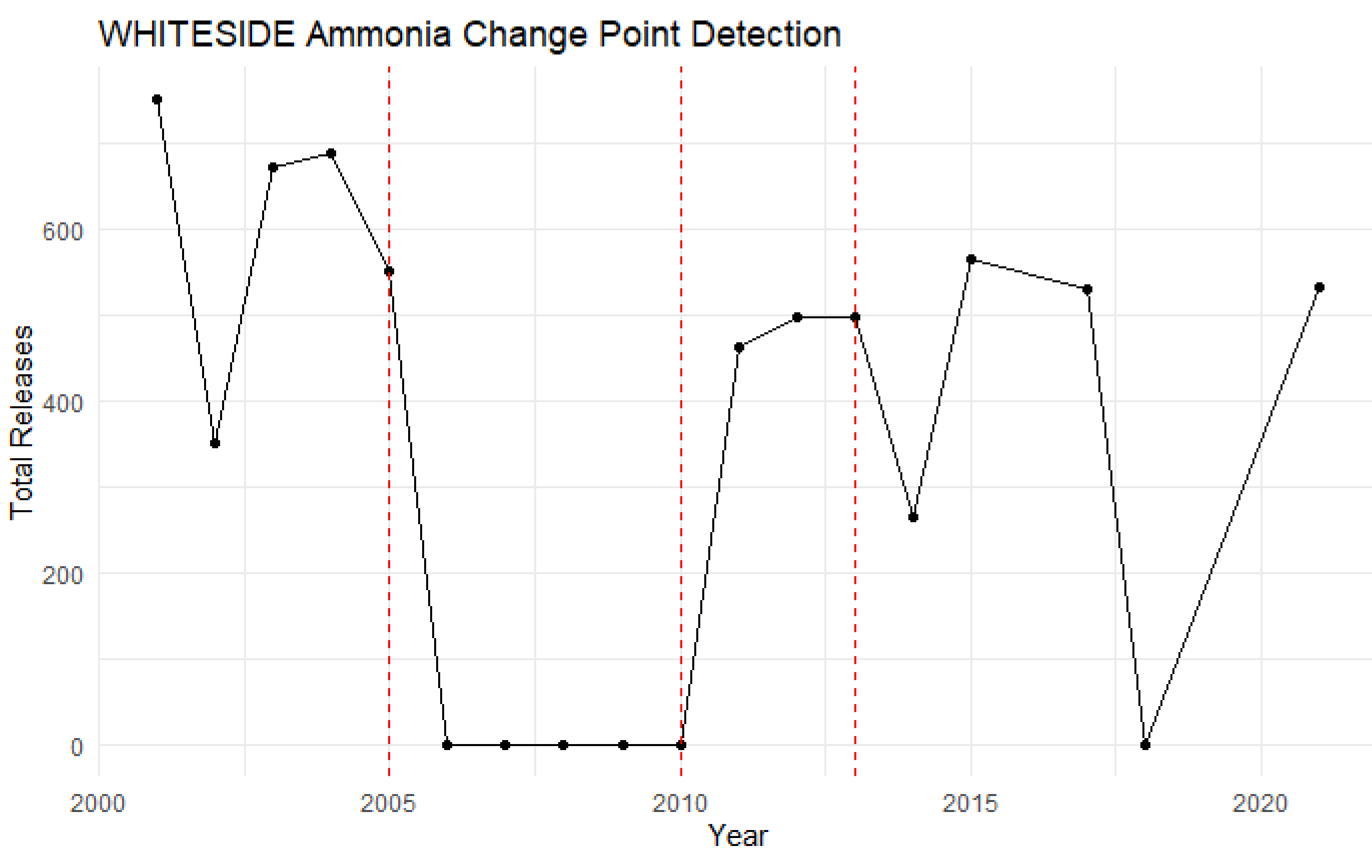


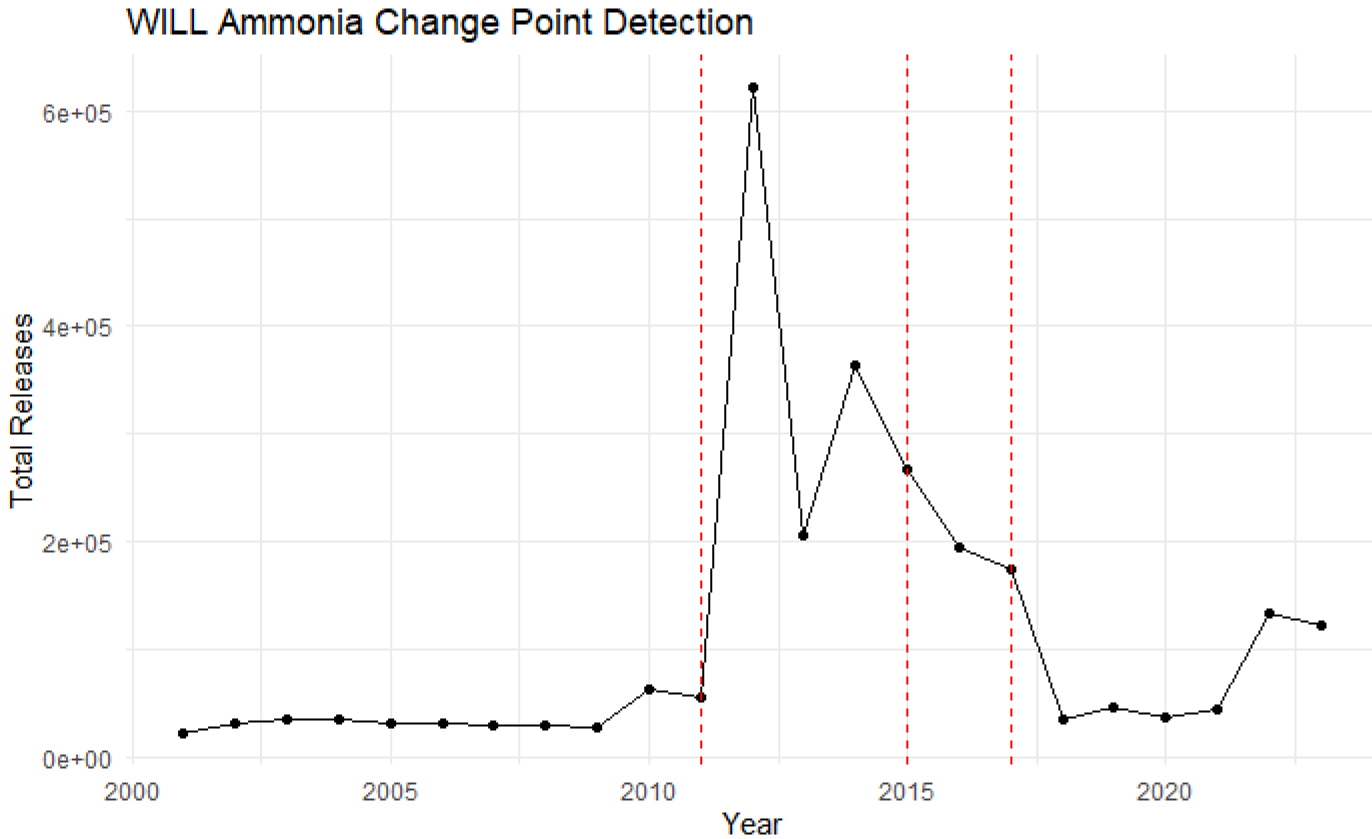
VERMILION Ammonia Change Point Detection
4e+05
3e+05
2e+05
Total Releases
2000
2005
2010
2015
2020
Year

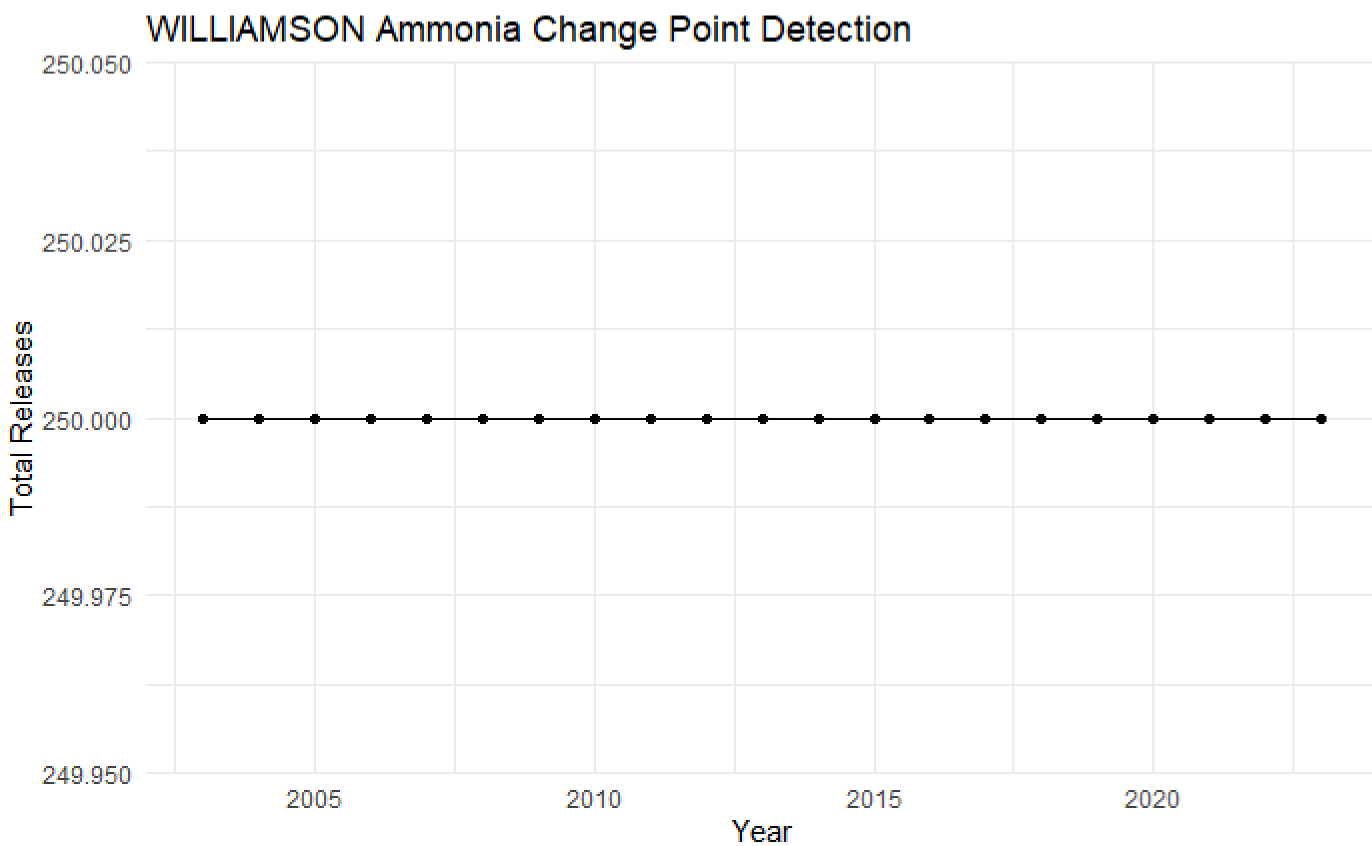


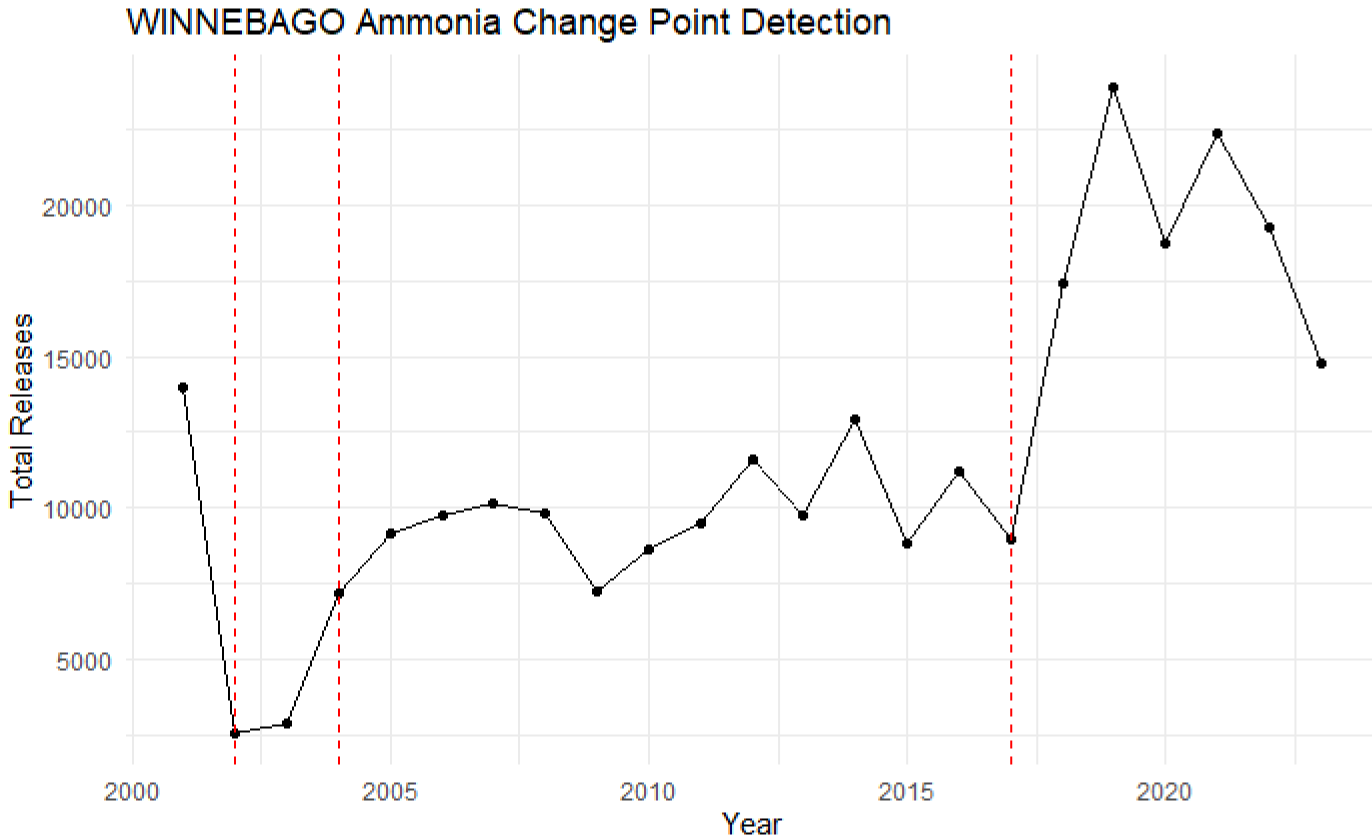
WARREN Ammonia Change Point Detection
40000
30000
20000
10000
Total Releases
2000
2005
2010
2015
2020
Year

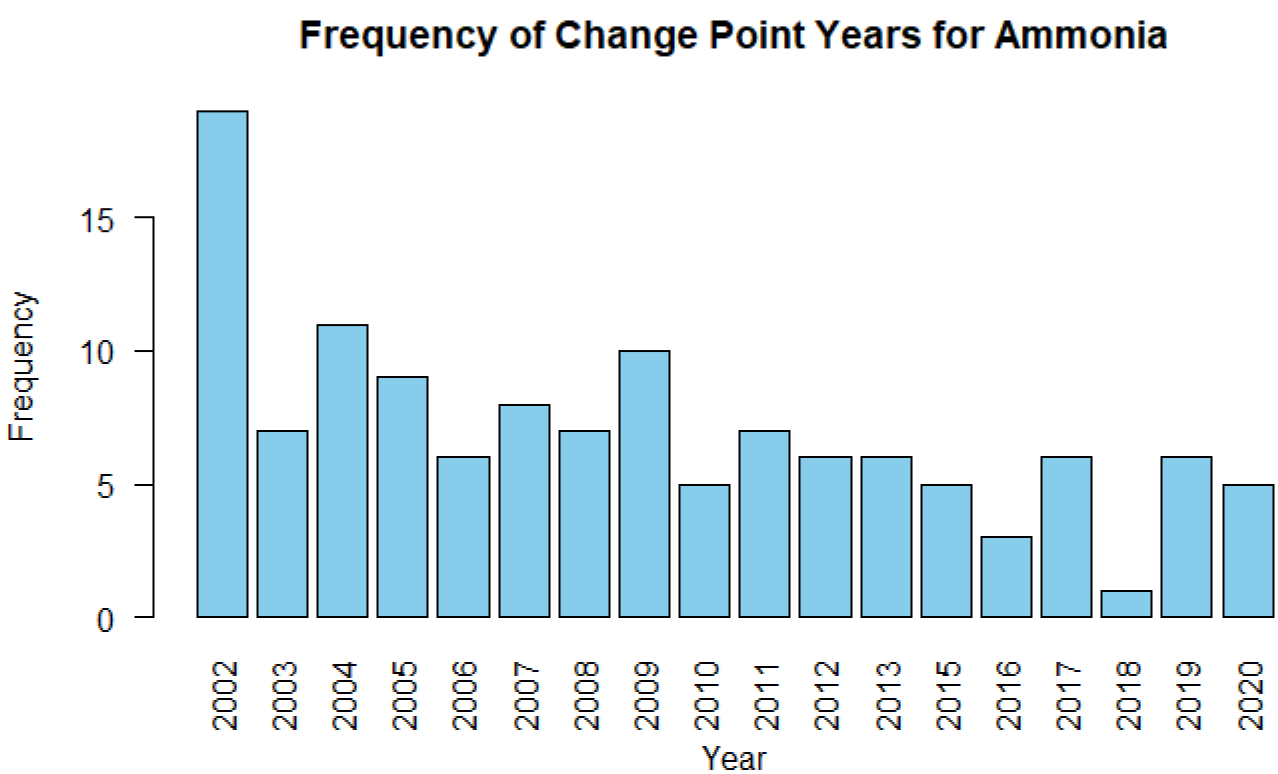

WASHINGTON Ammonia Change Point Detection
Total Releases
0
1000
2000
3000
4000
2012
2016
2020
Year


WHITESIDE Ammonia Change Point Detection
Total Releases
0
200
400
600
2000
2005
2010
2015
2020
Year


WILL Ammonia Change Point Detection
Total Releases
6e+05
4e+05
2e+05
0e+00
2000
2005
2010
2015
2020
Year


WILLIAMSON Ammonia Change Point Detection
Total Releases
250.050
250.025
250.000
249.975
249.950
2005
2010
2015
2020
Year

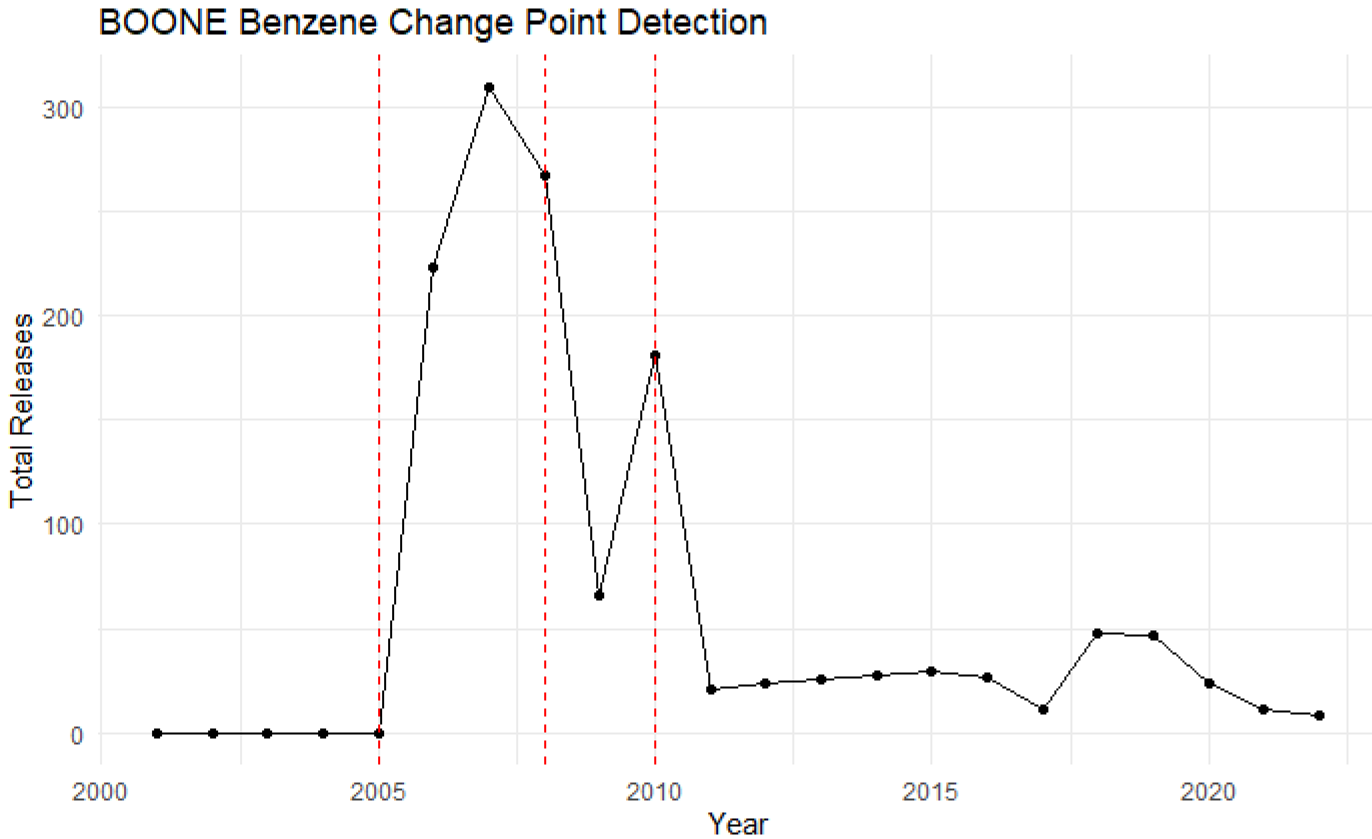


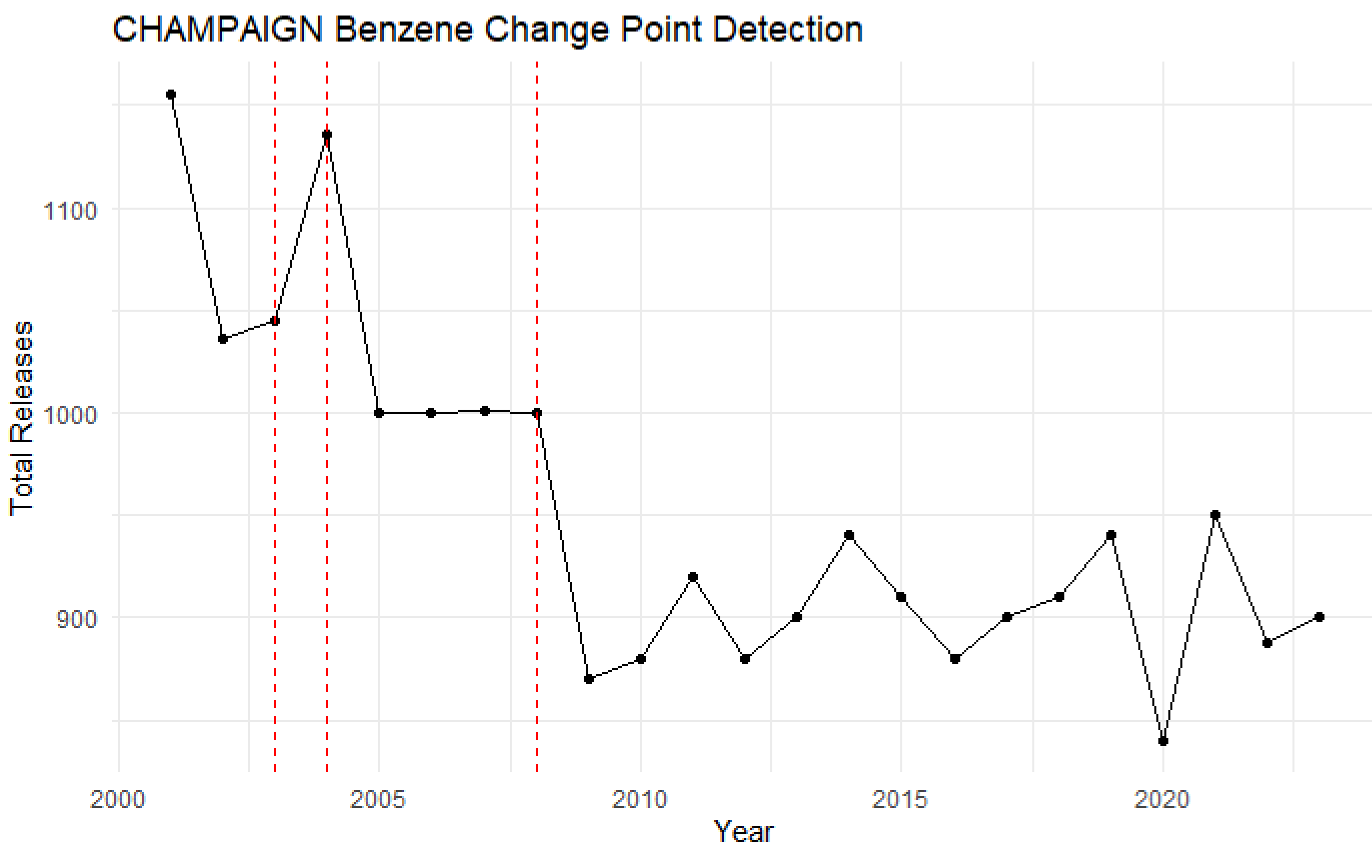


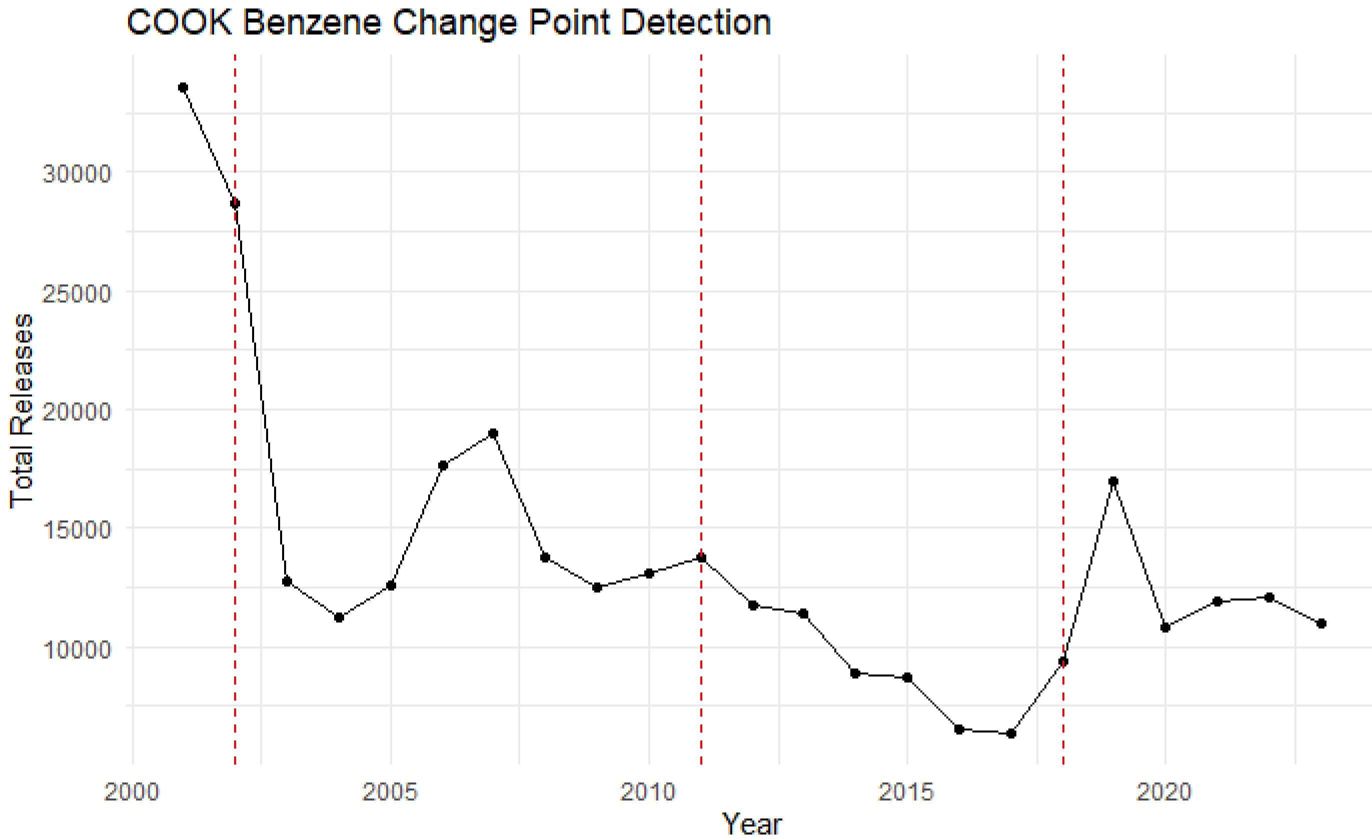
Benzene example

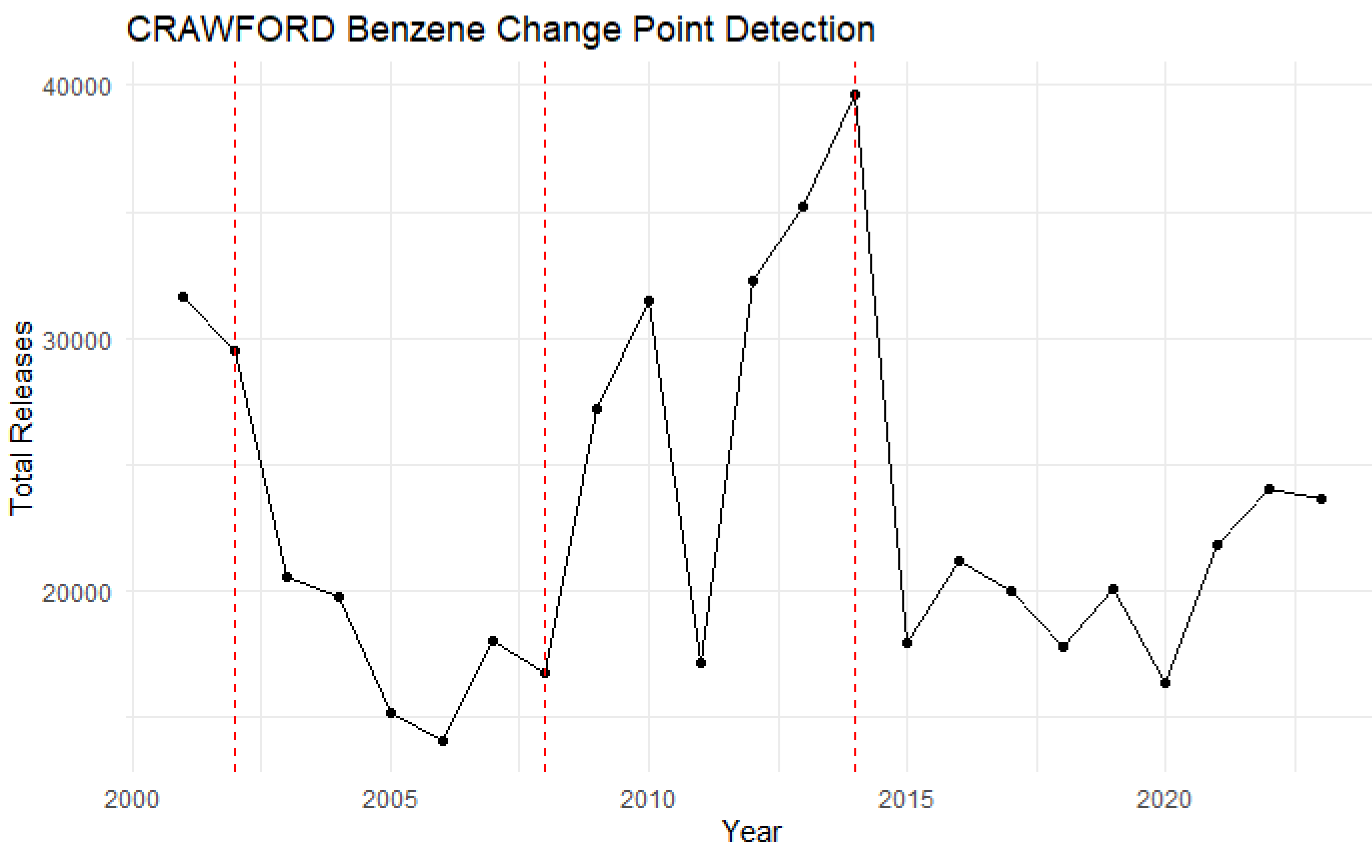
Hide

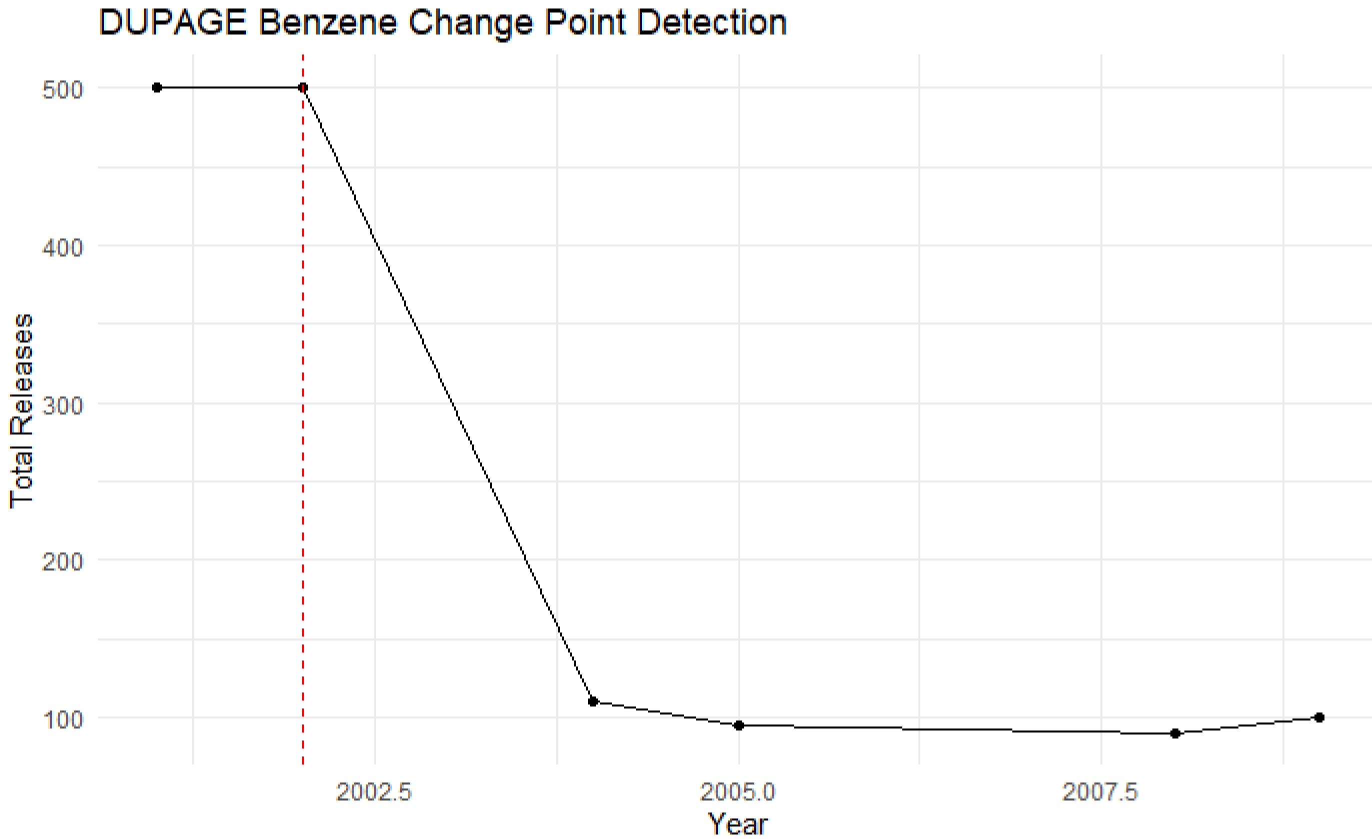
```
analyze_change_points("Benzene", chemical_by_county, change_points)
```

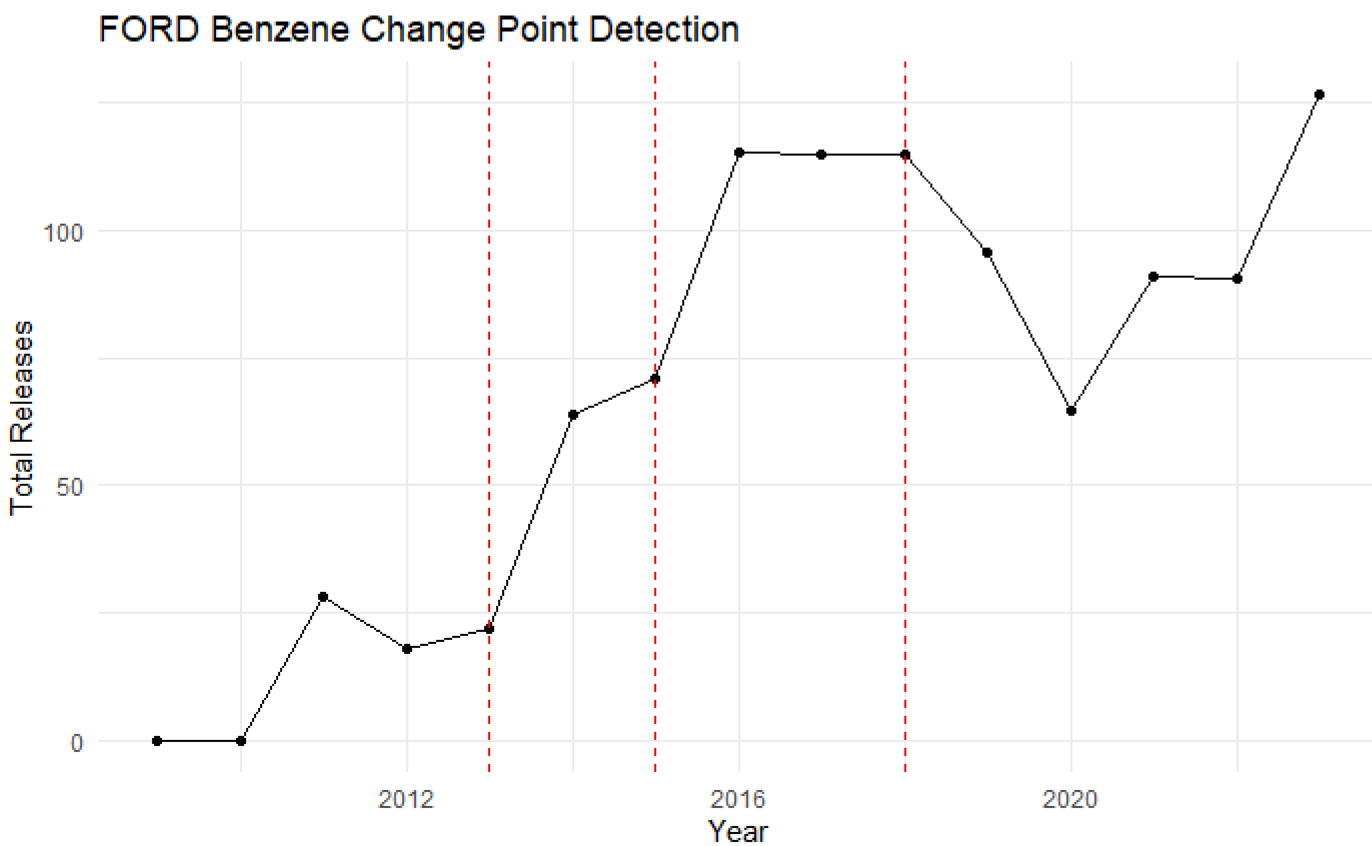

BOONE Benzene Change Point Detection
Total Releases
0
100
200
300
2000
2005
2010
2015
2020
Year

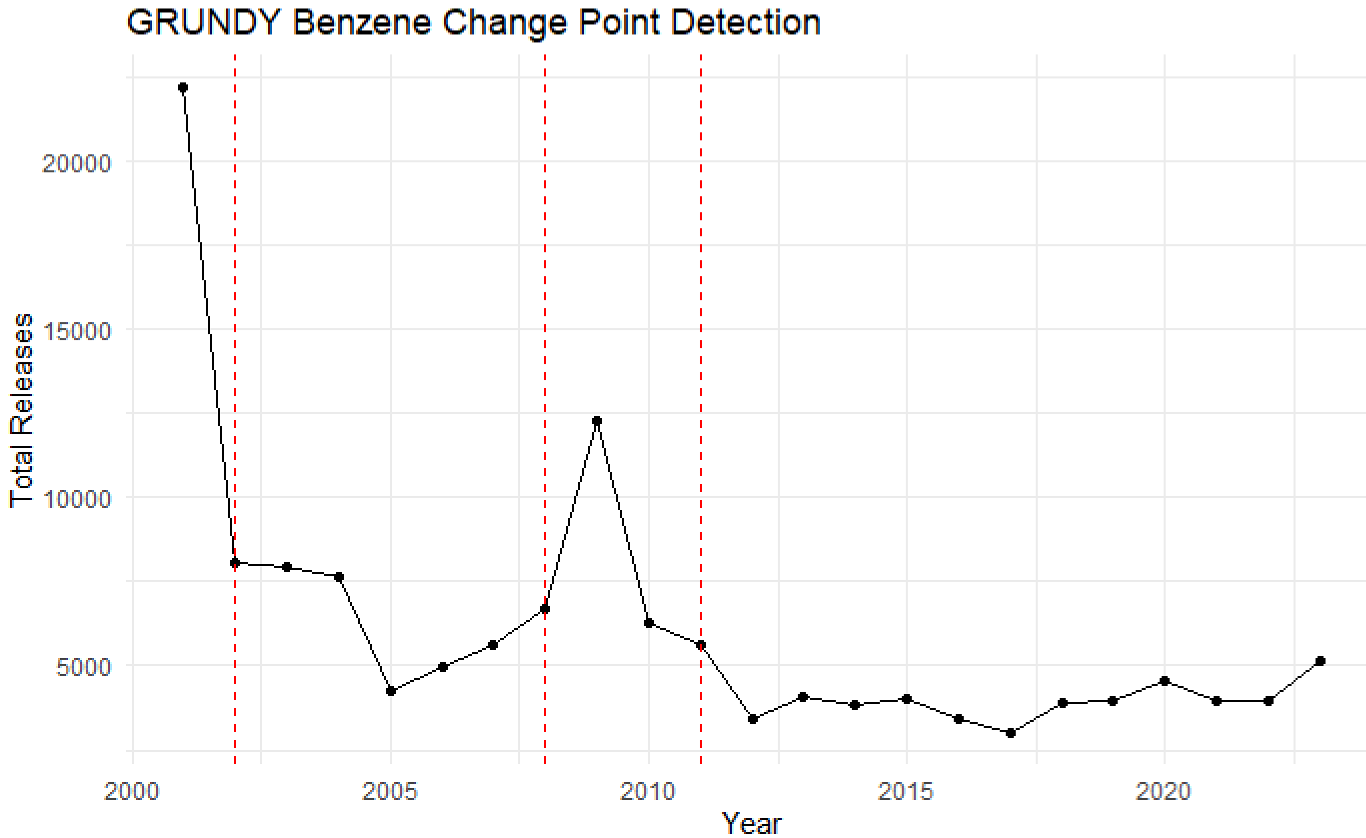

CHAMPAIGN Benzene Change Point Detection
Total Releases
900
1000
1100
2000
2005
2010
2015
2020
Year

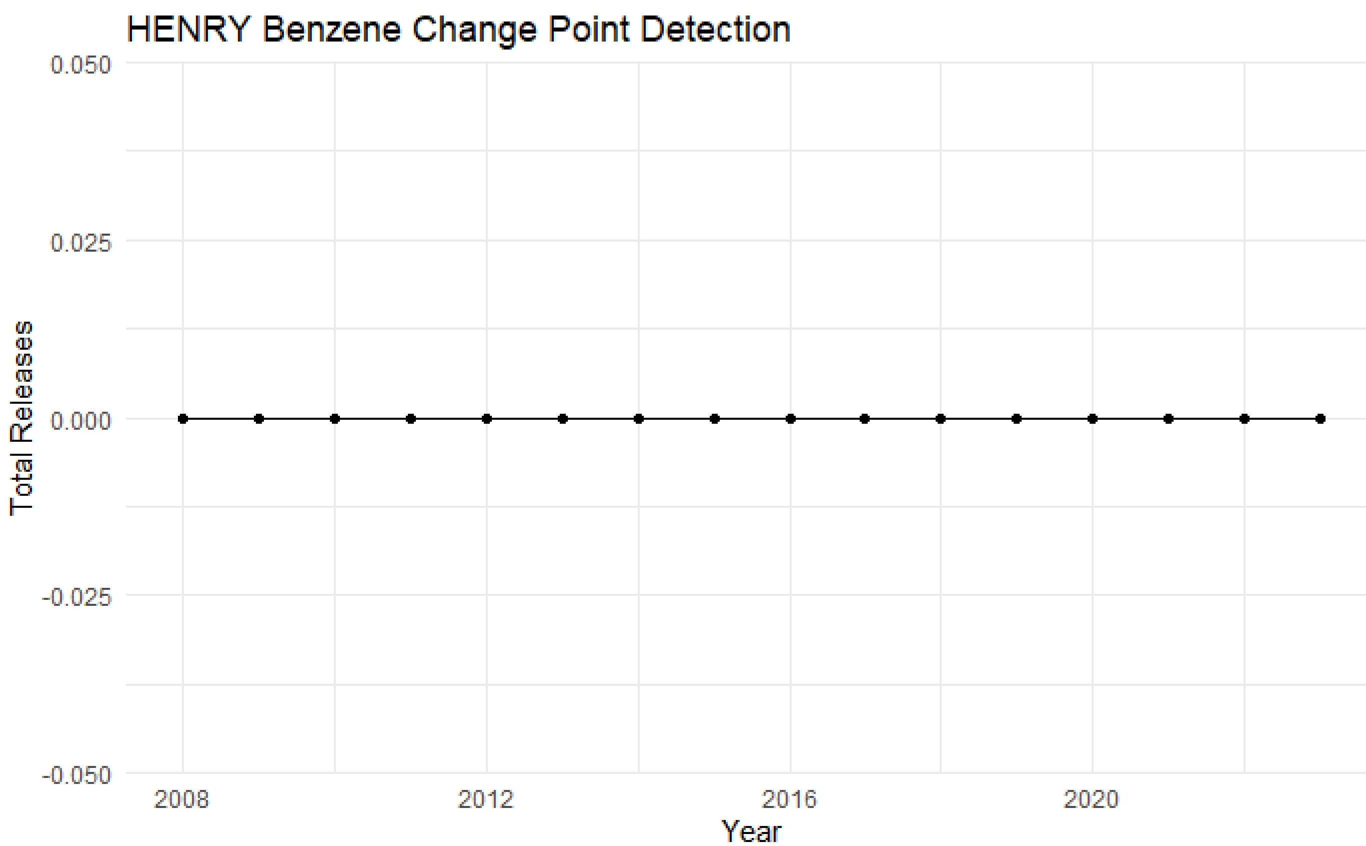

COOK Benzene Change Point Detection
Total Releases
30000
25000
20000
15000
10000
2000
2005
2010
2015
2020
Year

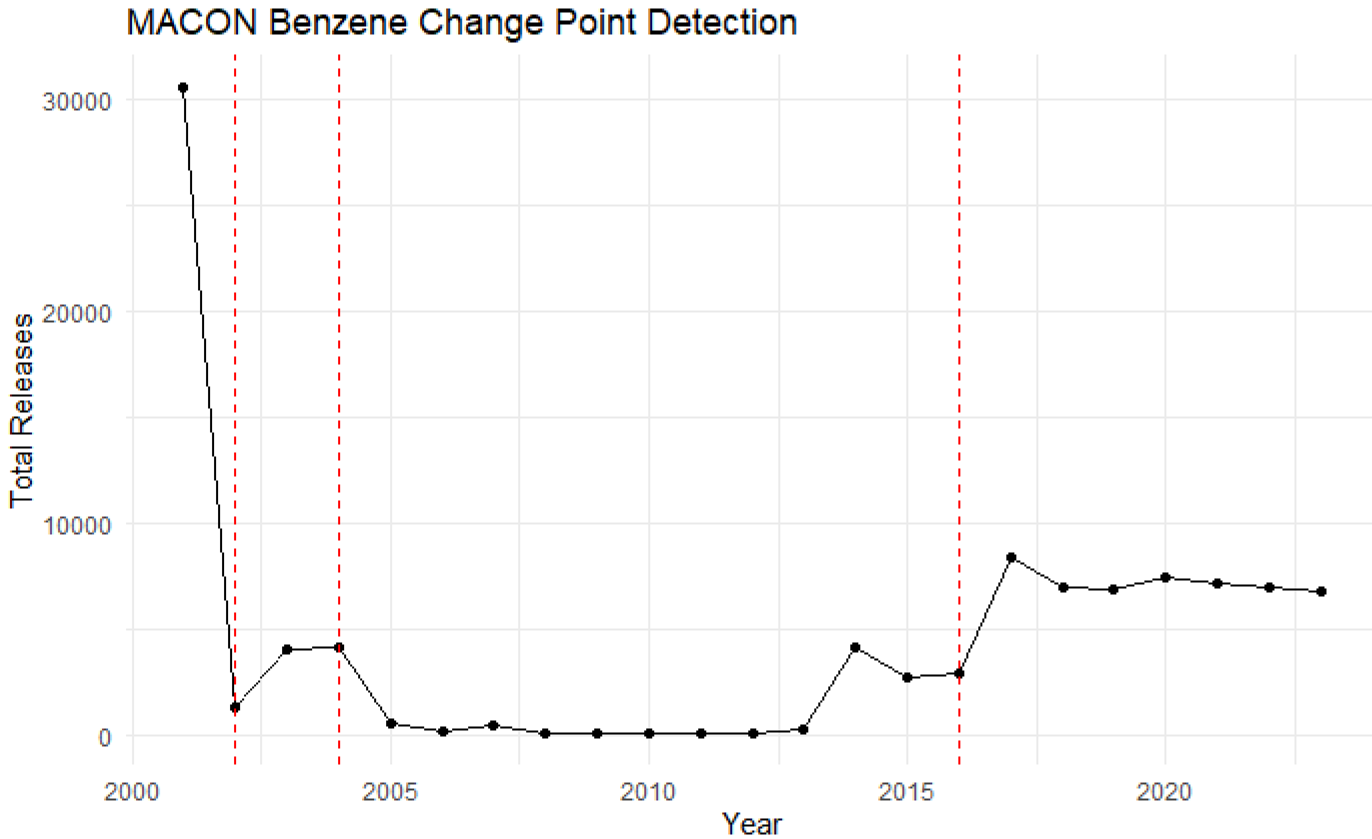

CRAWFORD Benzene Change Point Detection
Total Releases
40000
30000
20000
2000
2005
2010
2015
2020
Year

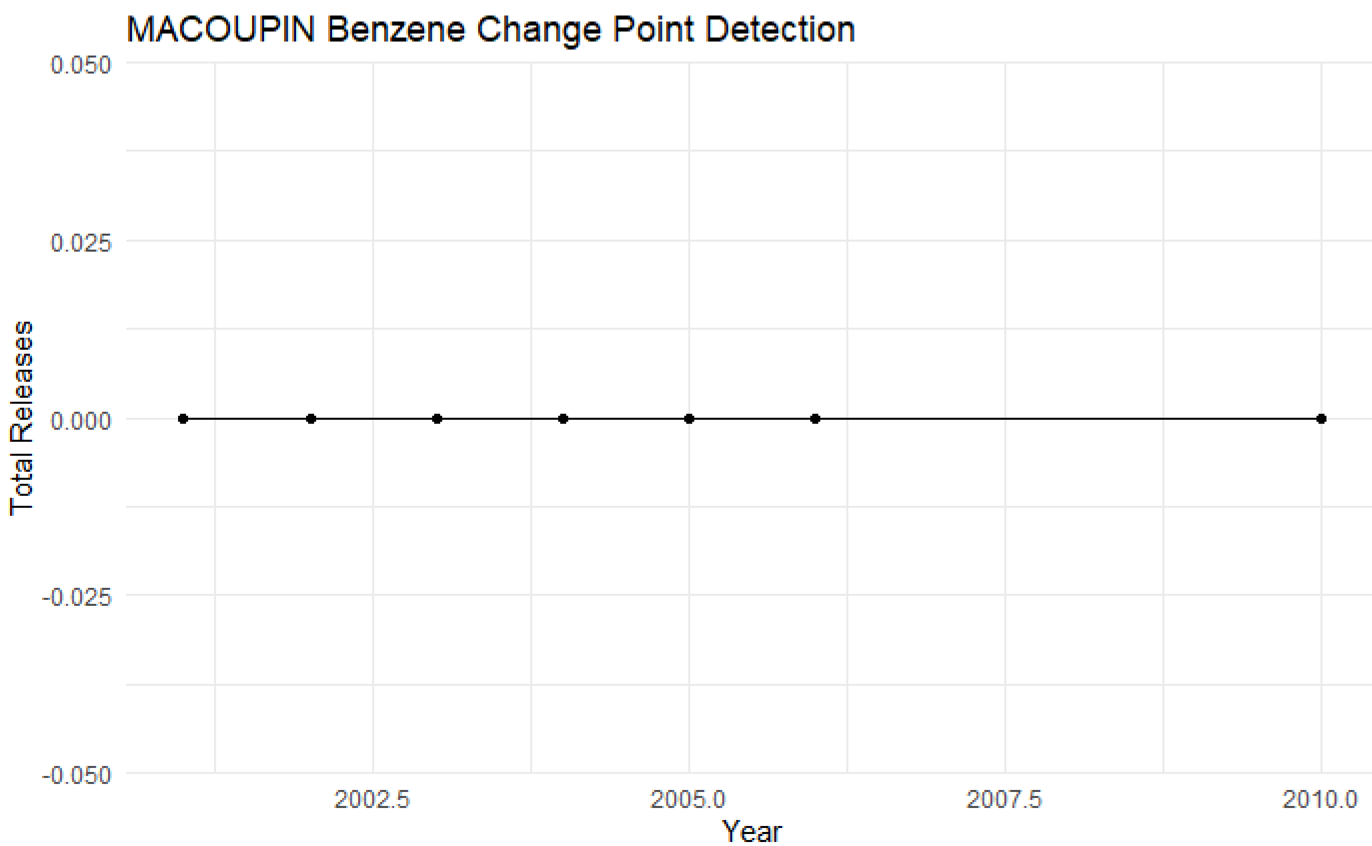

DUPAGE Benzene Change Point Detection
Total Releases
500
400
300
200
100
2002.5
2005.0
2007.5
Year

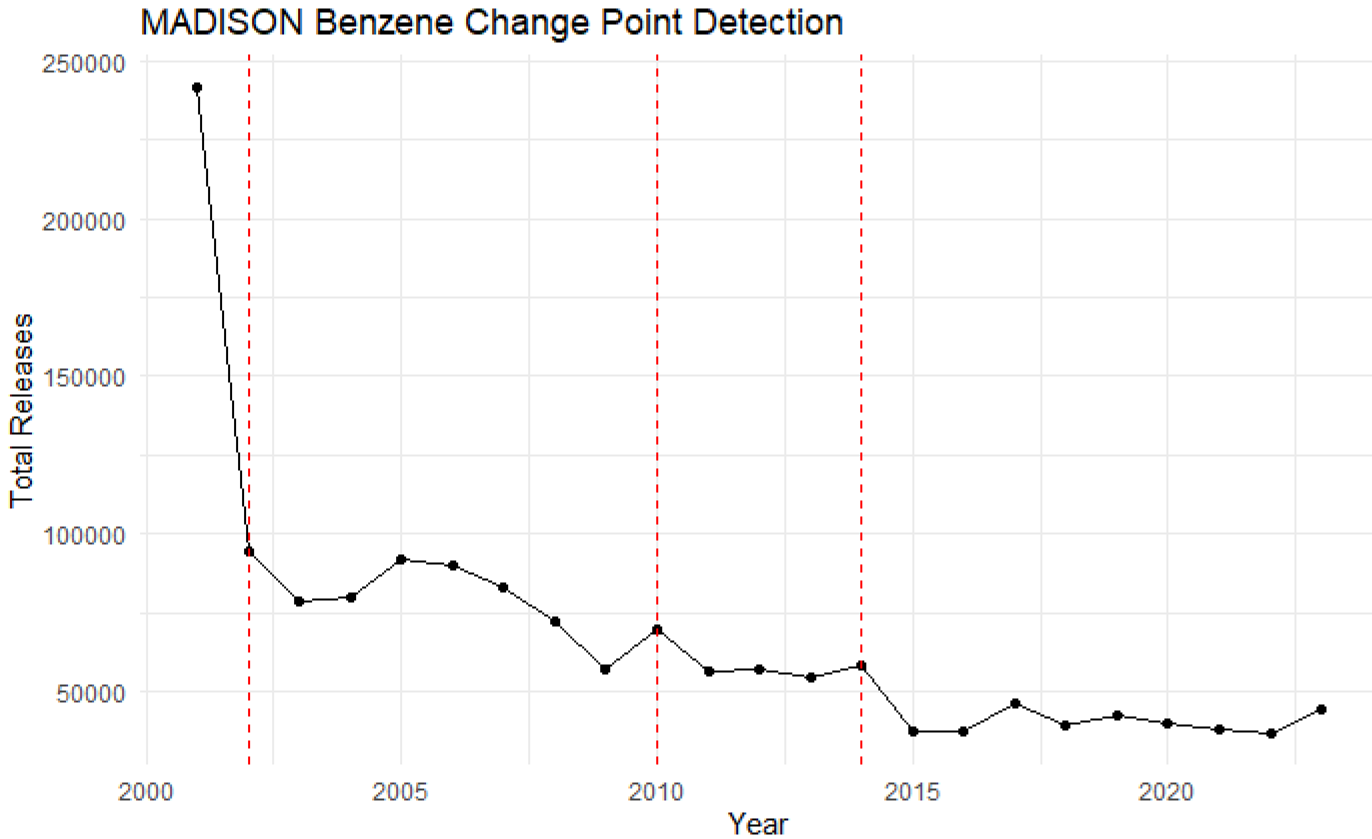

FORD Benzene Change Point Detection
Total Releases
100
50
0
2012
2016
2020
Year

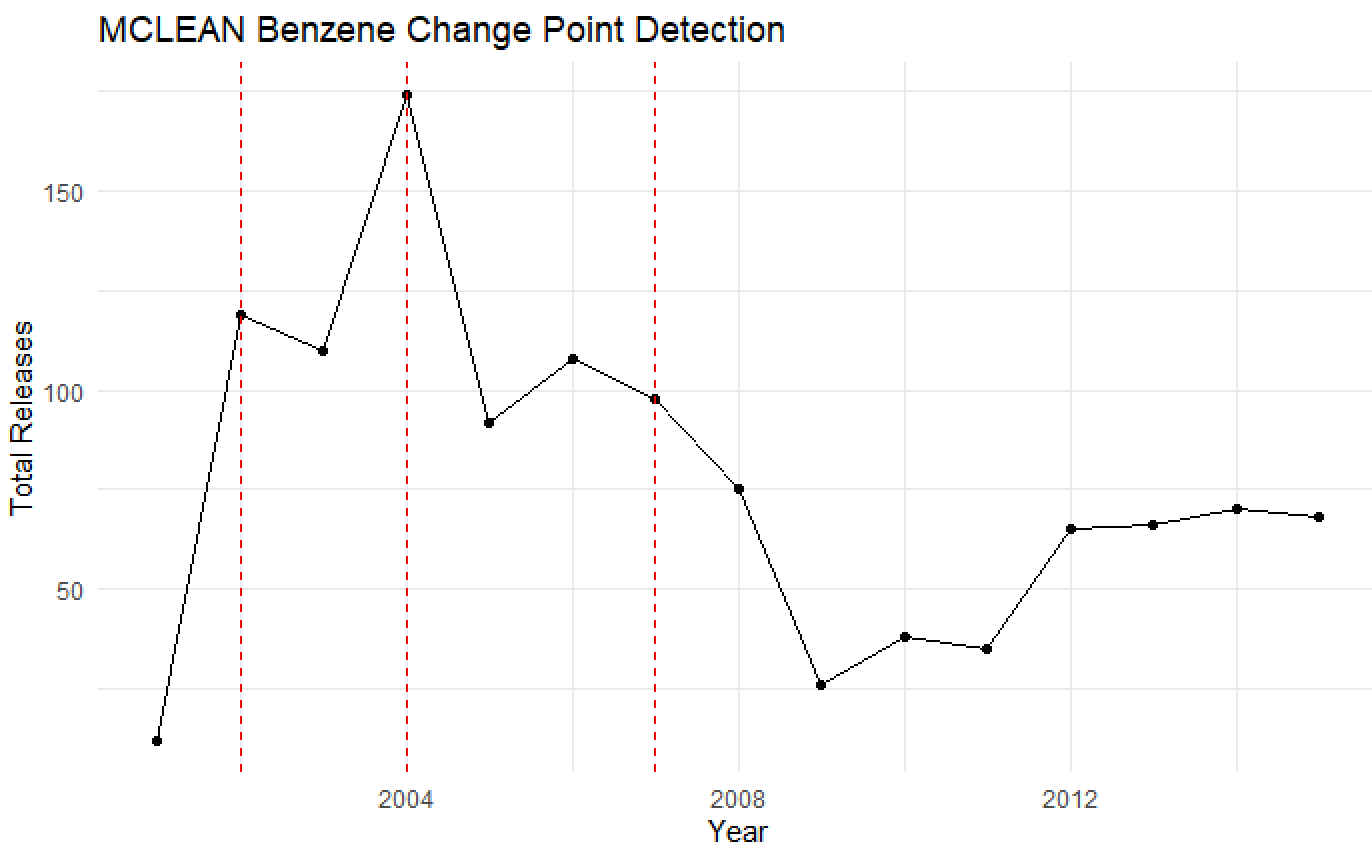

GRUNDY Benzene Change Point Detection
20000
15000
10000
5000
Total Releases
2000
2005
2010
2015
2020
Year

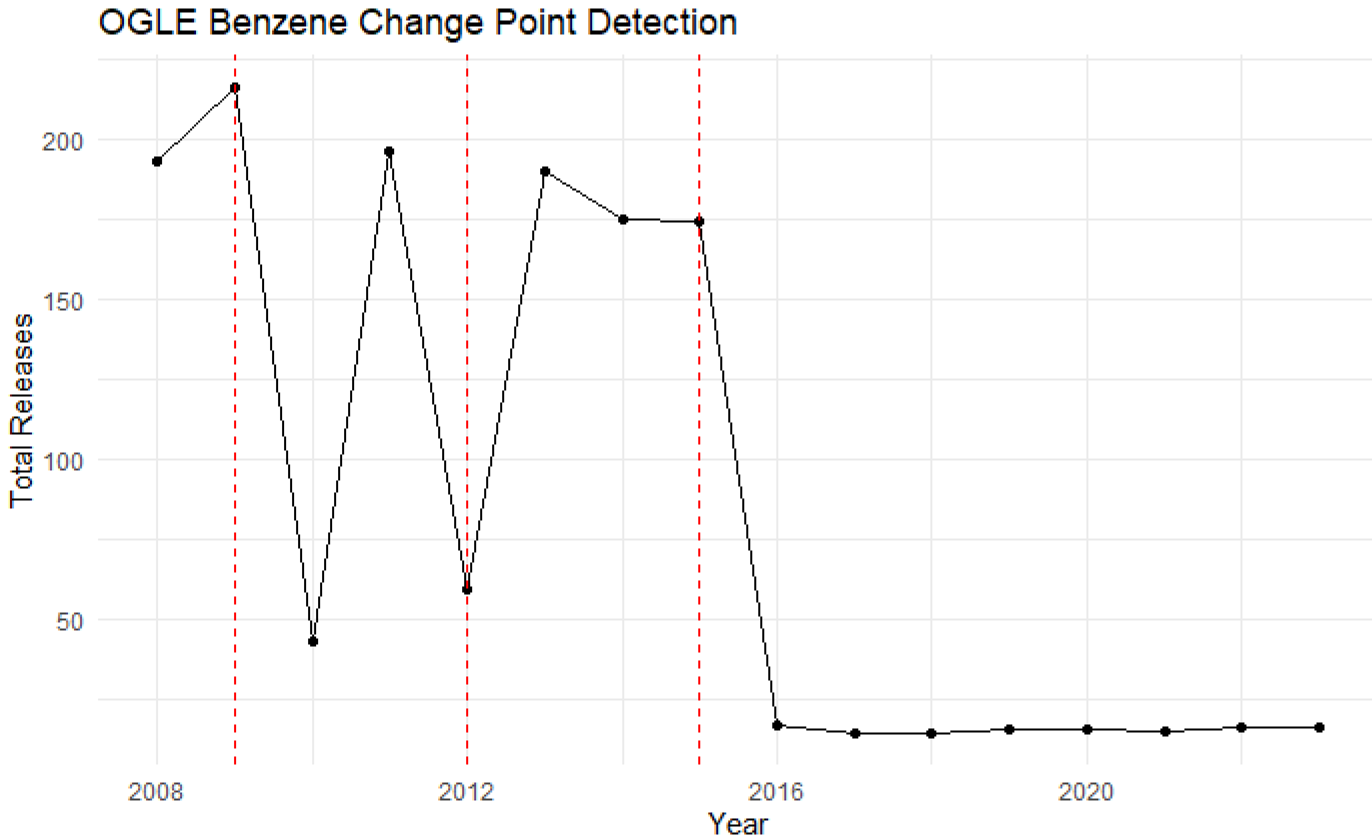

HENRY Benzene Change Point Detection
0.050
0.025
0.000
-0.025
-0.050
Total Releases
2008
2012
2016
2020
Year

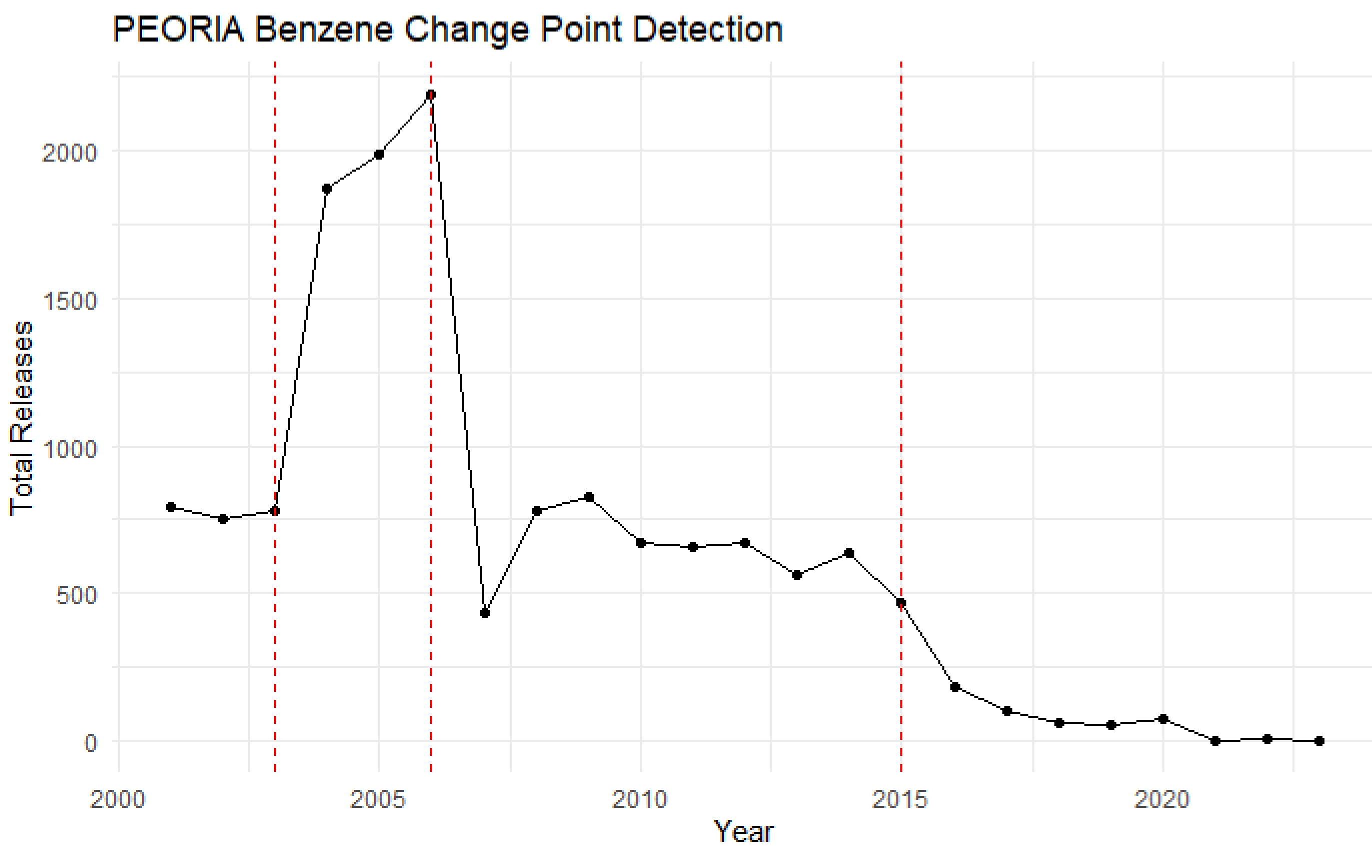

MACON Benzene Change Point Detection
Total Releases
30000
20000
10000
0
2000
2005
2010
2015
2020
Year

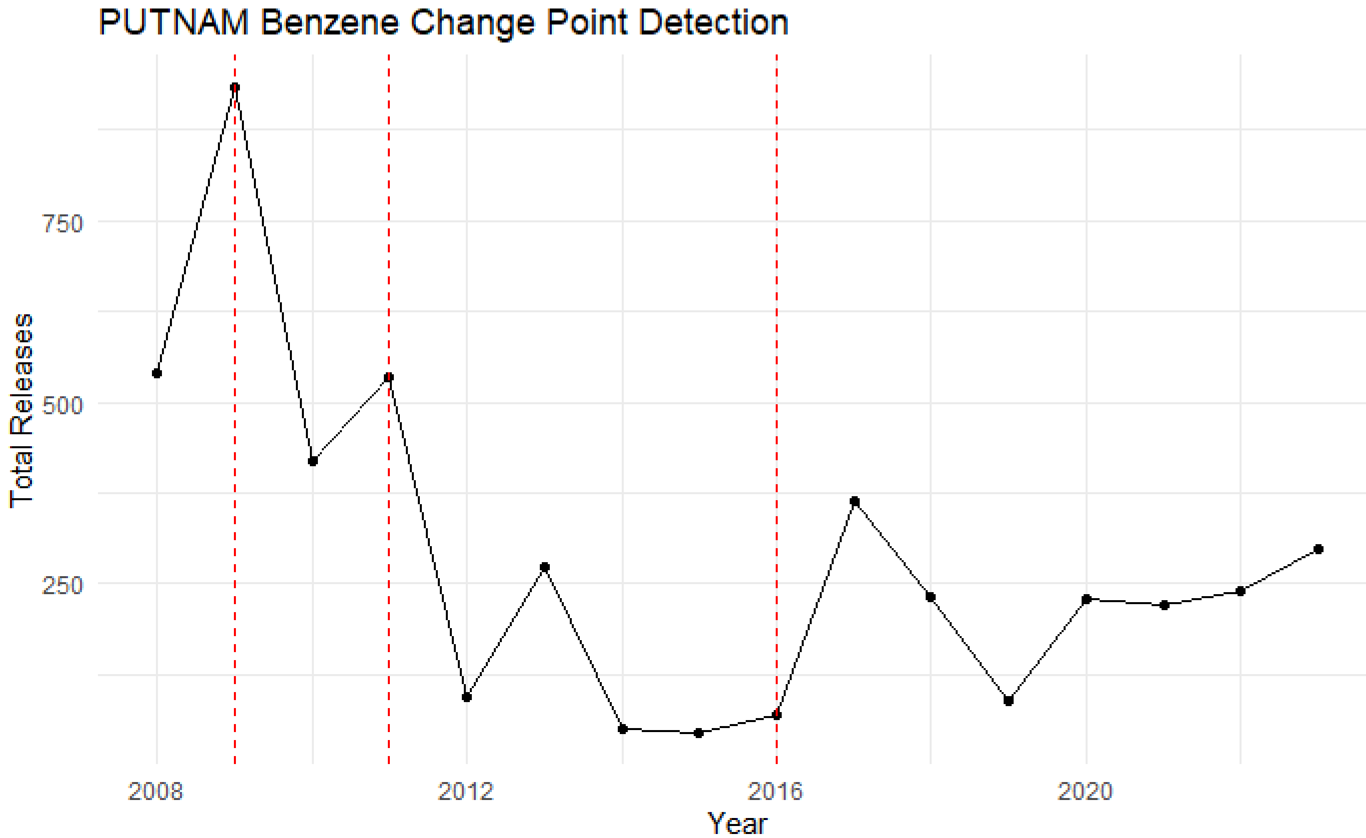

MACOUPIN Benzene Change Point Detection
Total Releases
0.050
0.025
0.000
-0.025
-0.050
2002.5
2005.0
2007.5
2010.0
Year

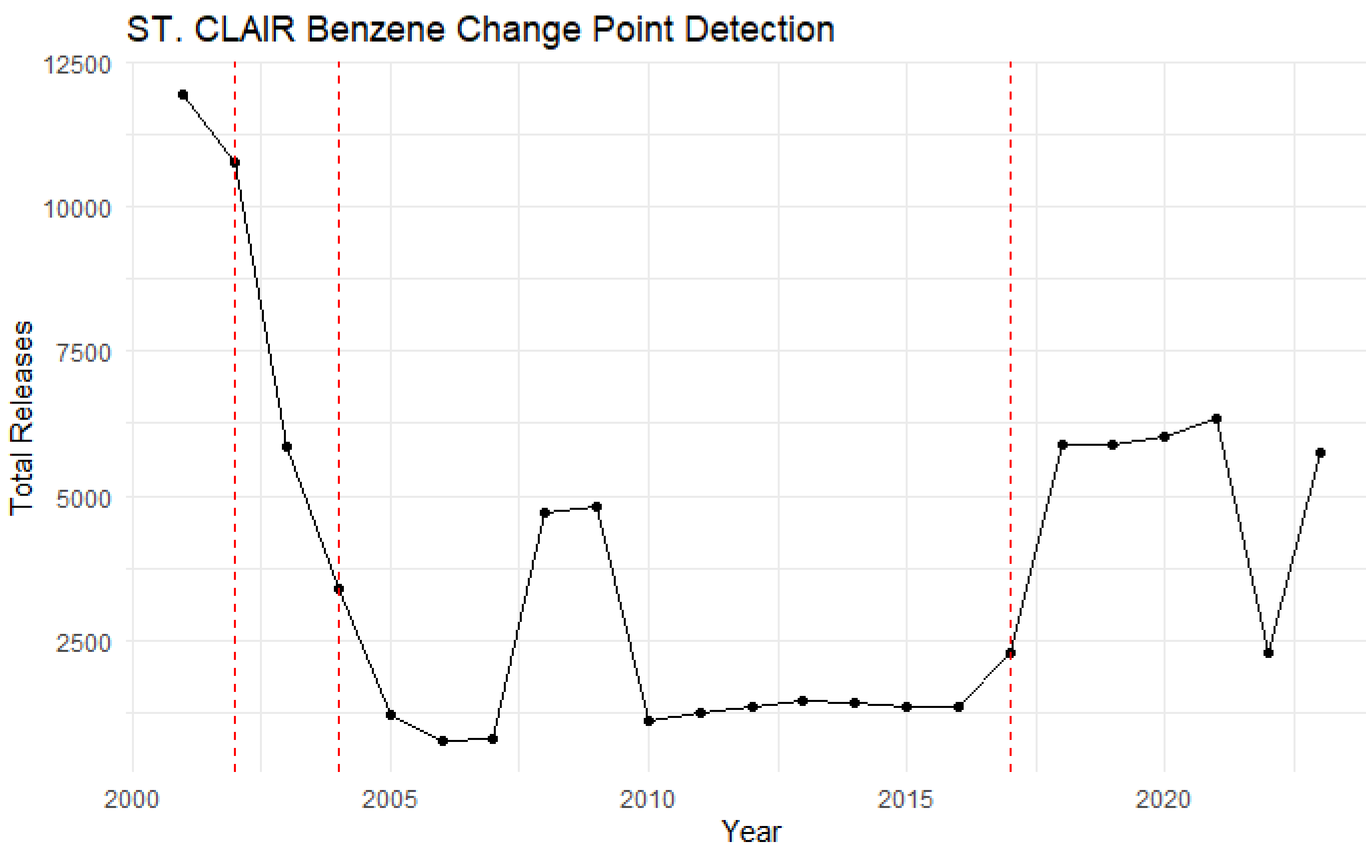

MADISON Benzene Change Point Detection
Total Releases
250000
200000
150000
100000
50000
2000
2005
2010
2015
2020
Year

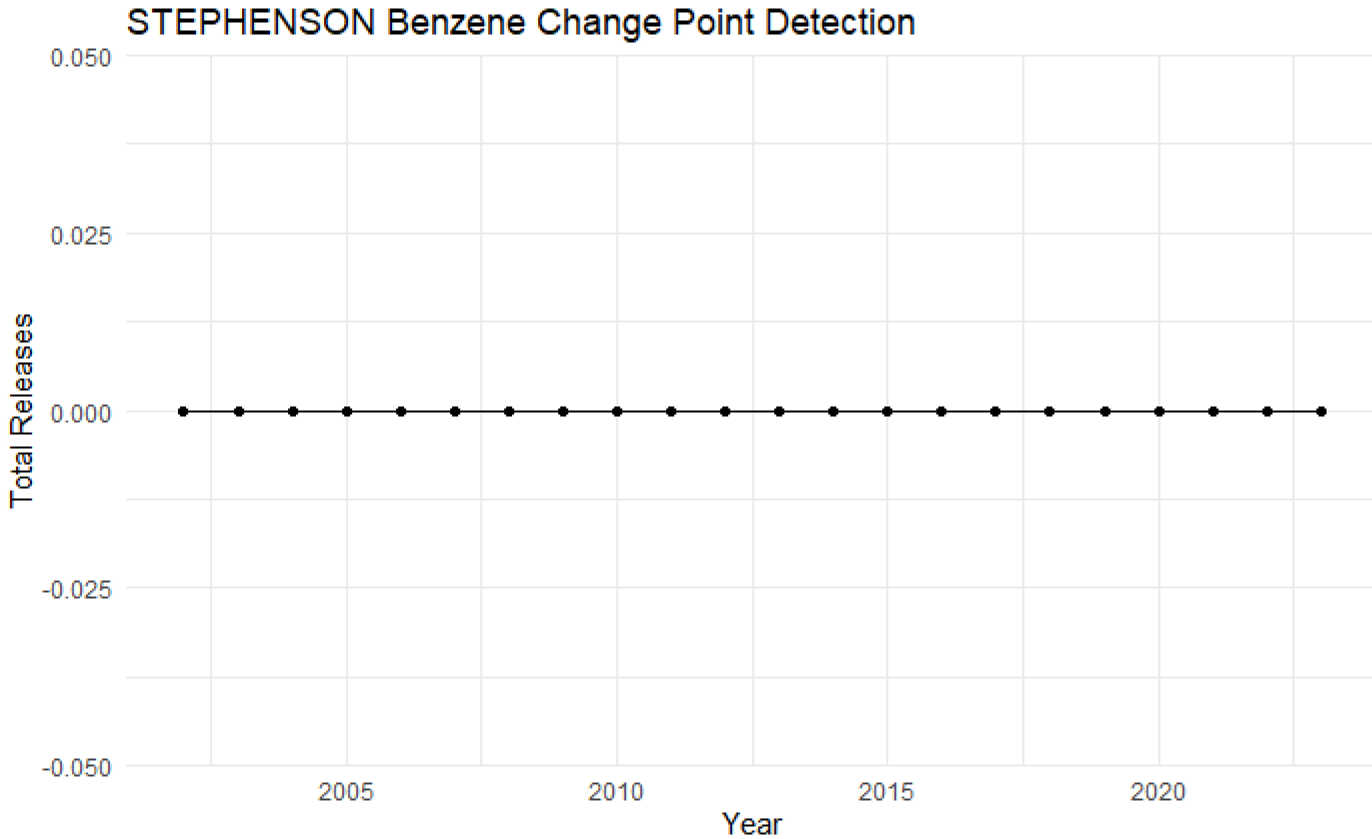

MCLEAN Benzene Change Point Detection
Total Releases
150
100
50
2004
2008
2012
Year

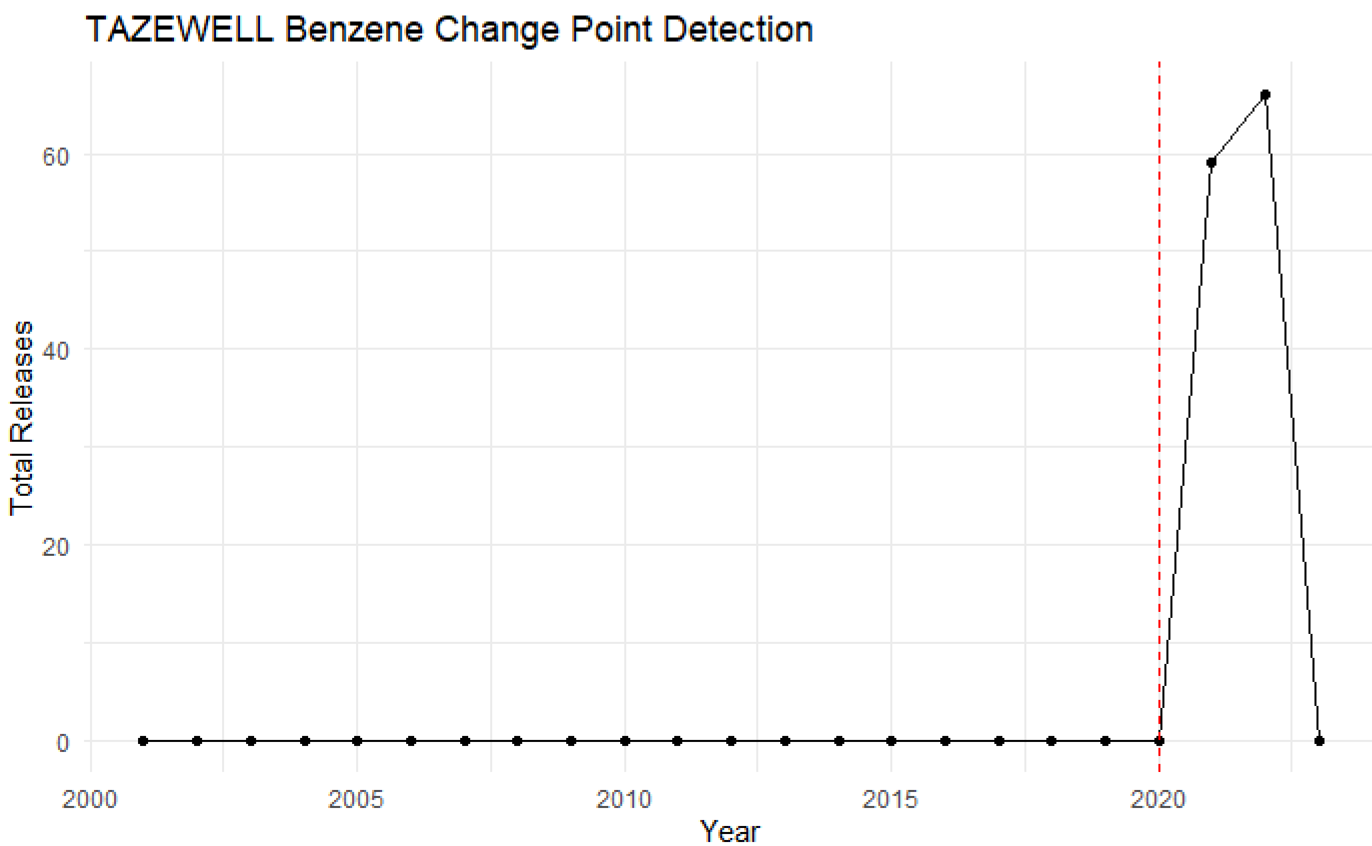

OGLE Benzene Change Point Detection
Total Releases
Year
200
150
100
50
2008
2012
2016
2020

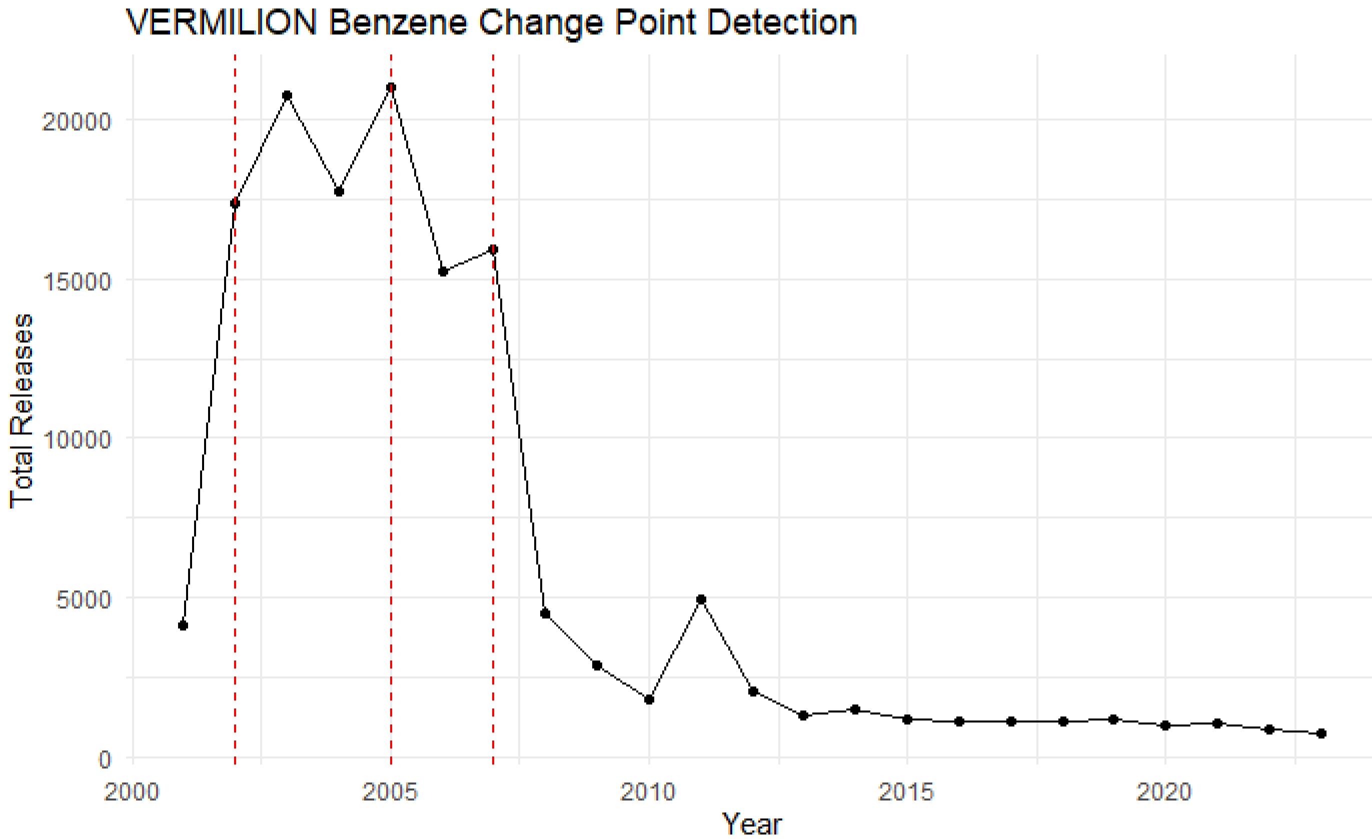

PEORIA Benzene Change Point Detection
Total Releases
Year
2000
1500
1000
500
0
2000
2005
2010
2015
2020

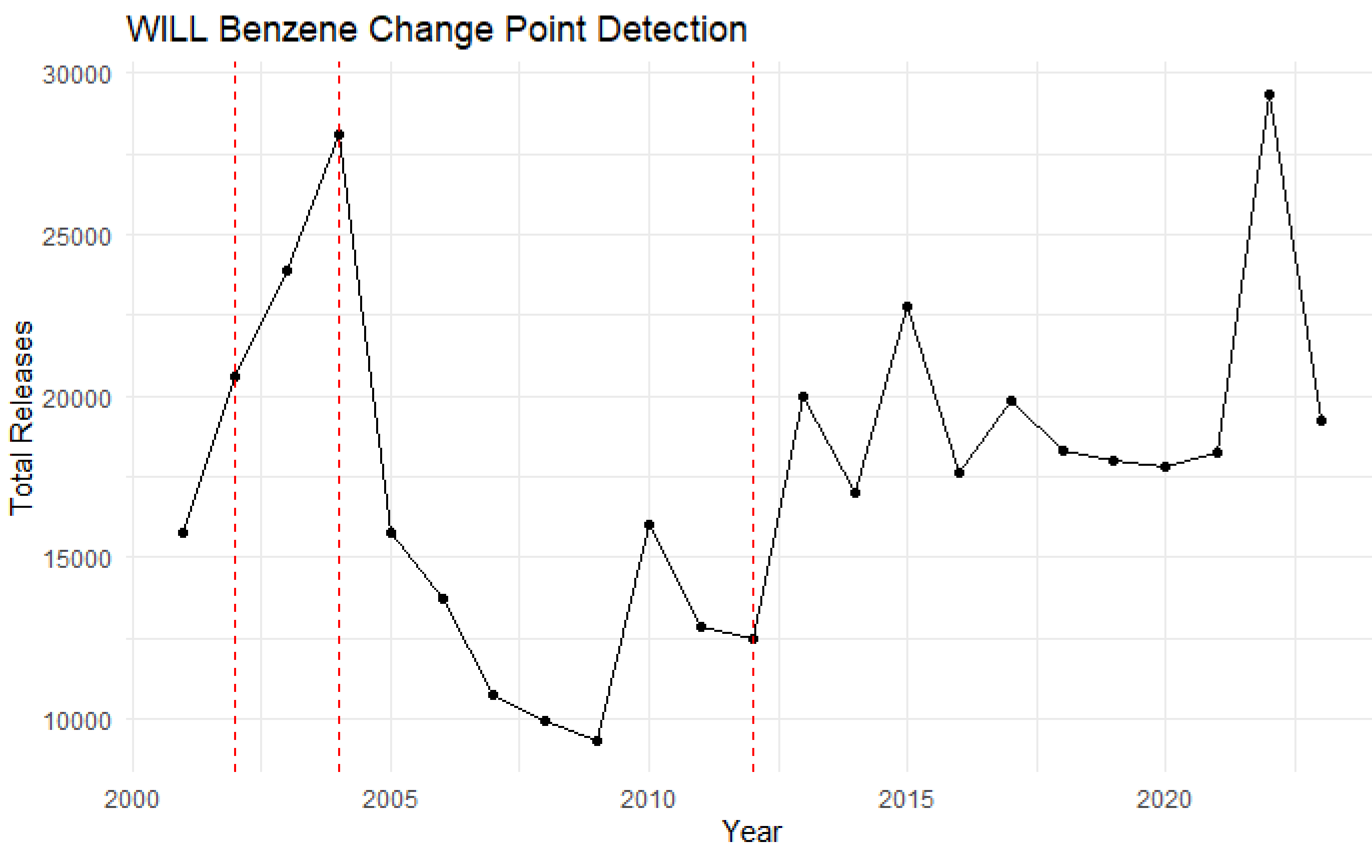

PUTNAM Benzene Change Point Detection
Total Releases
Year
250
500
750
2008
2012
2016
2020

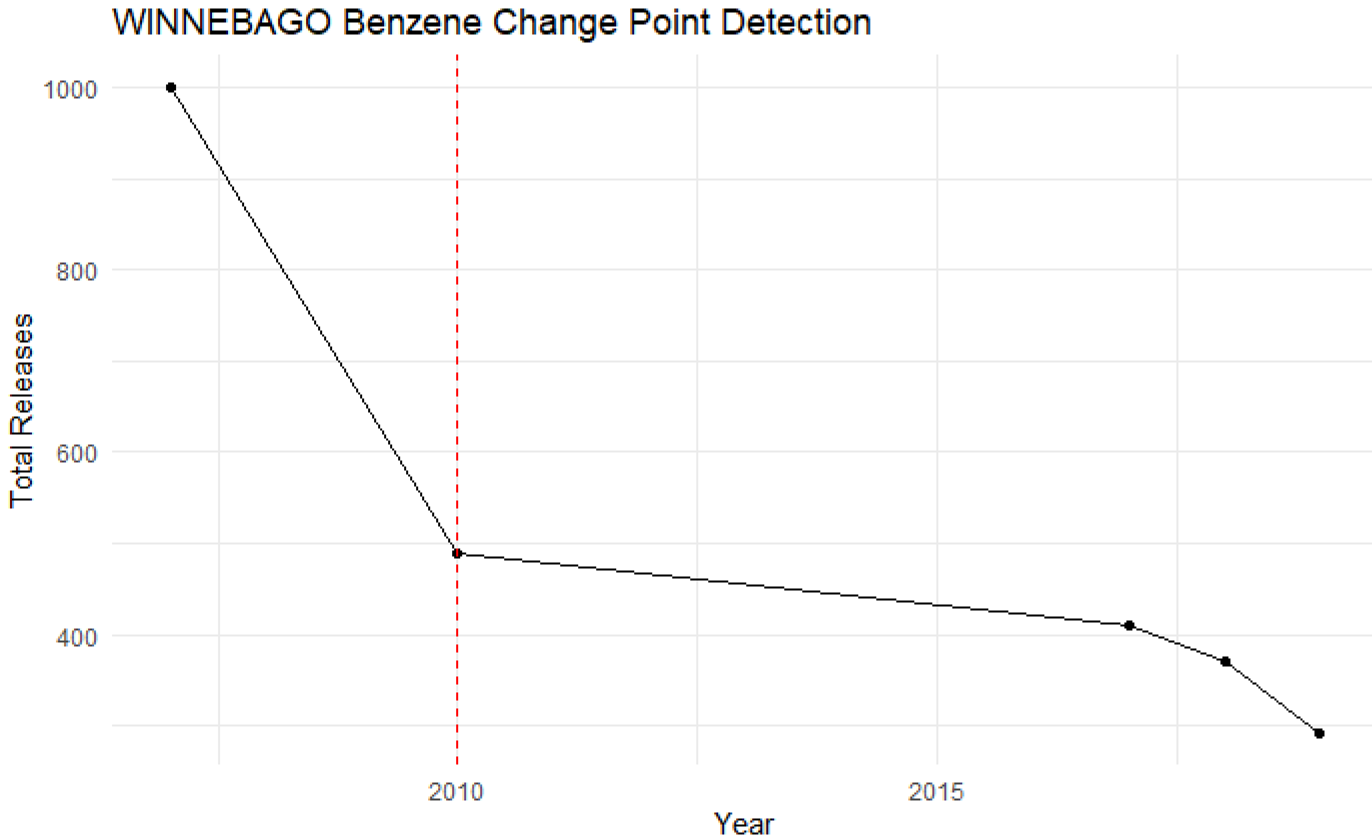

ST. CLAIR Benzene Change Point Detection
Total Releases
Year
2500
5000
7500
10000
12500
2000
2005
2010
2015
2020

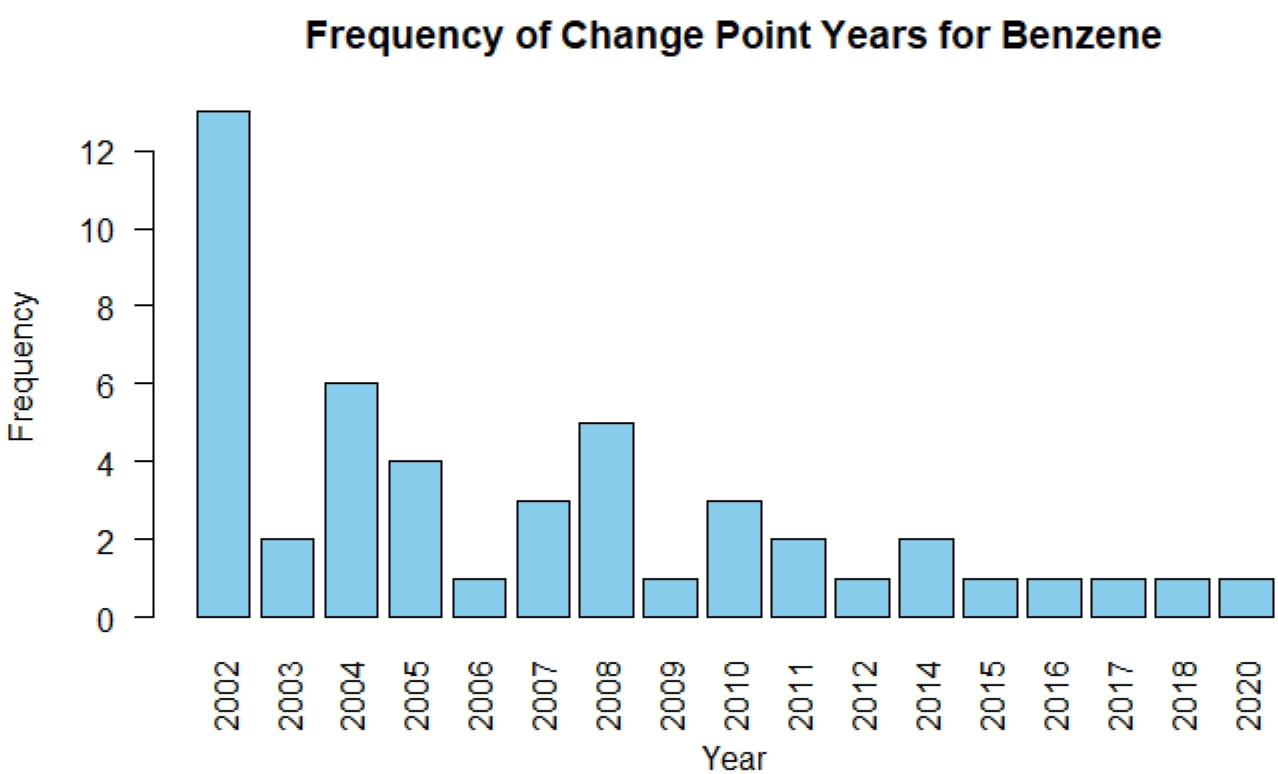

STEPHENSON Benzene Change Point Detection
0.050
0.025
0.000
-0.025
-0.050
Total Releases
2005
2010
2015
2020
Year


TAZEWELL Benzene Change Point Detection
60
40
20
0
Total Releases
2000
2005
2010
2015
2020
Year


VERMILION Benzene Change Point Detection
Total Releases
Year
0
5000
10000
15000
20000
2000
2005
2010
2015
2020


WILL Benzene Change Point Detection
Total Releases
Year
10000
15000
20000
25000
30000
2000
2005
2010
2015
2020





Change point detection by year

Hide

```
change_points <- list()

# Group by year and chemical this time
valid_data <- chemical_by_county %>%
  group_by(year, chemical) %>%
  summarise(total_releases = sum(total_releases, na.rm = TRUE)) %>%
  filter(n() >= 3) %>%
  ungroup()
```

```
`summarise()` has grouped output by 'year'. You can override using the `.groups` argument.
```

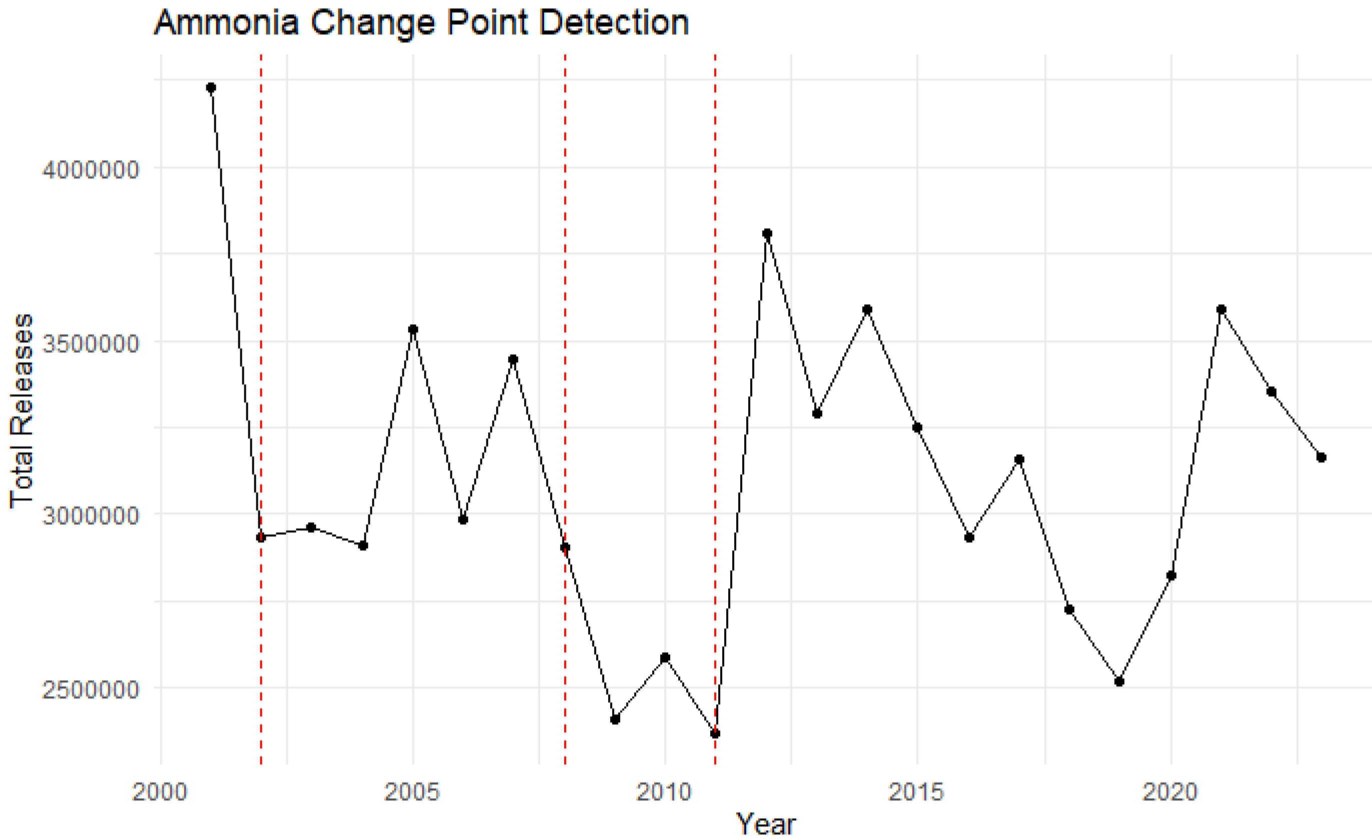
Hide

```
# Only unique chemicals since no counties
unique_chemicals <- unique(valid_data$chemical)

# For each unique chemical, take the subset of data and run get change points function
for (i in 1:length(unique_chemicals)) {
  chemical1 <- unique_chemicals[i]
  subset_data <- subset(valid_data, chemical == chemical1)

  result <- get_change_points(subset_data)

  if (!is.null(result)) {
    change_points[[chemical1]] <- result
  } else {
    change_points[[chemical1]] <- "No valid data or change points"
  }
}

plot_change_points <- function(df, change_points) {
  df$year <- as.numeric(df$year)

  # Order by year
  df <- df[order(df$year), ]

  # Create change point plot of total releases
  change_point_plot <- ggplot(df, aes(x = year, y = total_releases)) +
    geom_line() +
    geom_point() +
    labs(title = paste(df$chemical[1], "Change Point Detection"),
         x = "Year", y = "Total Releases") +
    theme_minimal()

  # Plotting change points using a vertical red line
  if (!is.null(change_points) && length(change_points) > 0) {
    change_point_plot <- change_point_plot + geom_vline(xintercept = df$year[change_points], color = "red",
linetype = "dashed")
  }
  print(change_point_plot)
}

analyze_change_points <- function(chemical_name, chemical_by_county, change_points) {
  subset_data <- subset(chemical_by_county, chemical == chemical_name)  # Only chemical for this subset
  change_point_result <- change_points[[chemical_name]]

  if (!is.null(change_point_result)) {
    plot_change_points(subset_data, change_point_result)
    change_point_years <- change_point_result  # Collect change point years
  } else {
    change_point_years <- c()
  }
}
```

Ammonia change points statewide

Hide

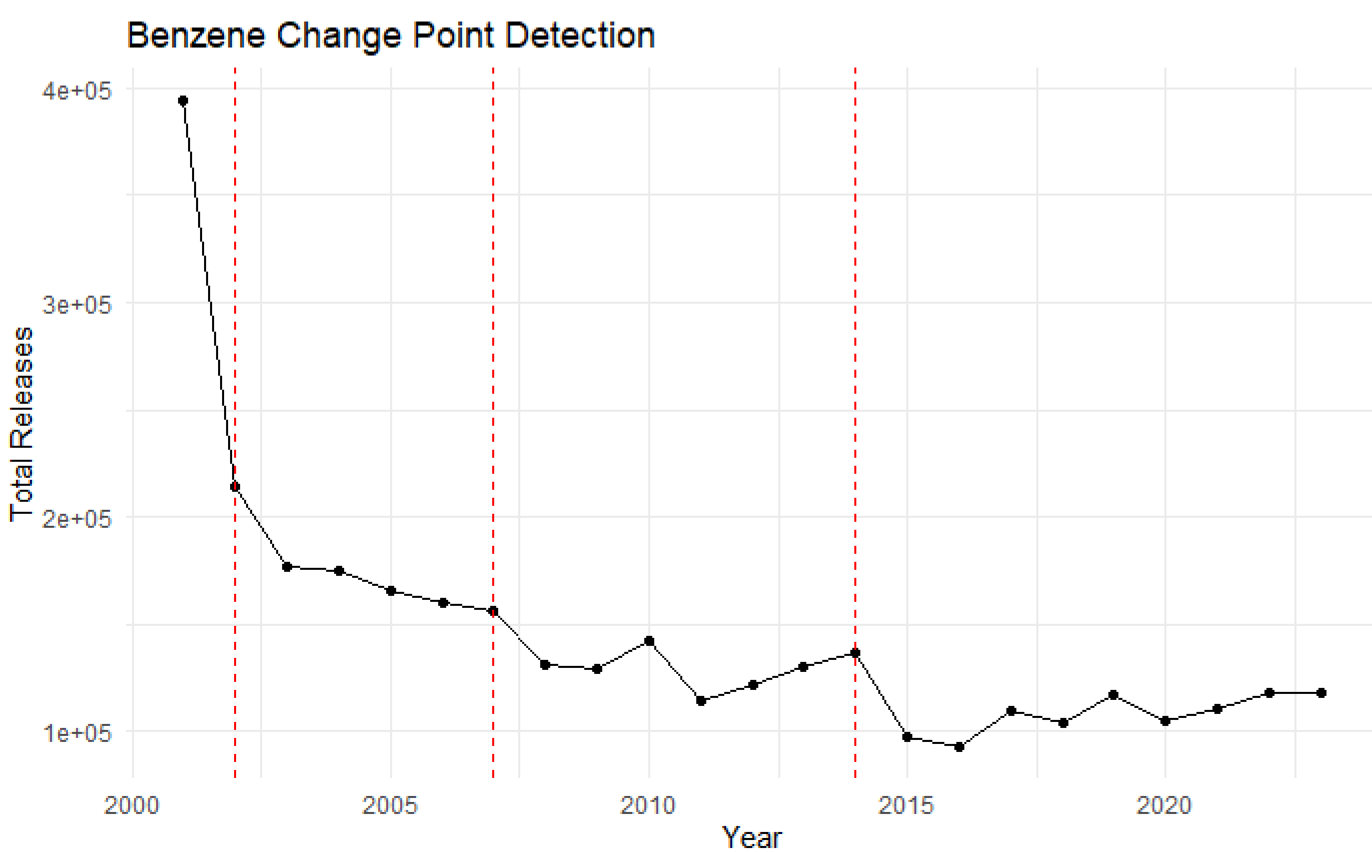
```
analyze_change_points("Ammonia", valid_data, change_points)
```



Benzene change points statewide

Hide

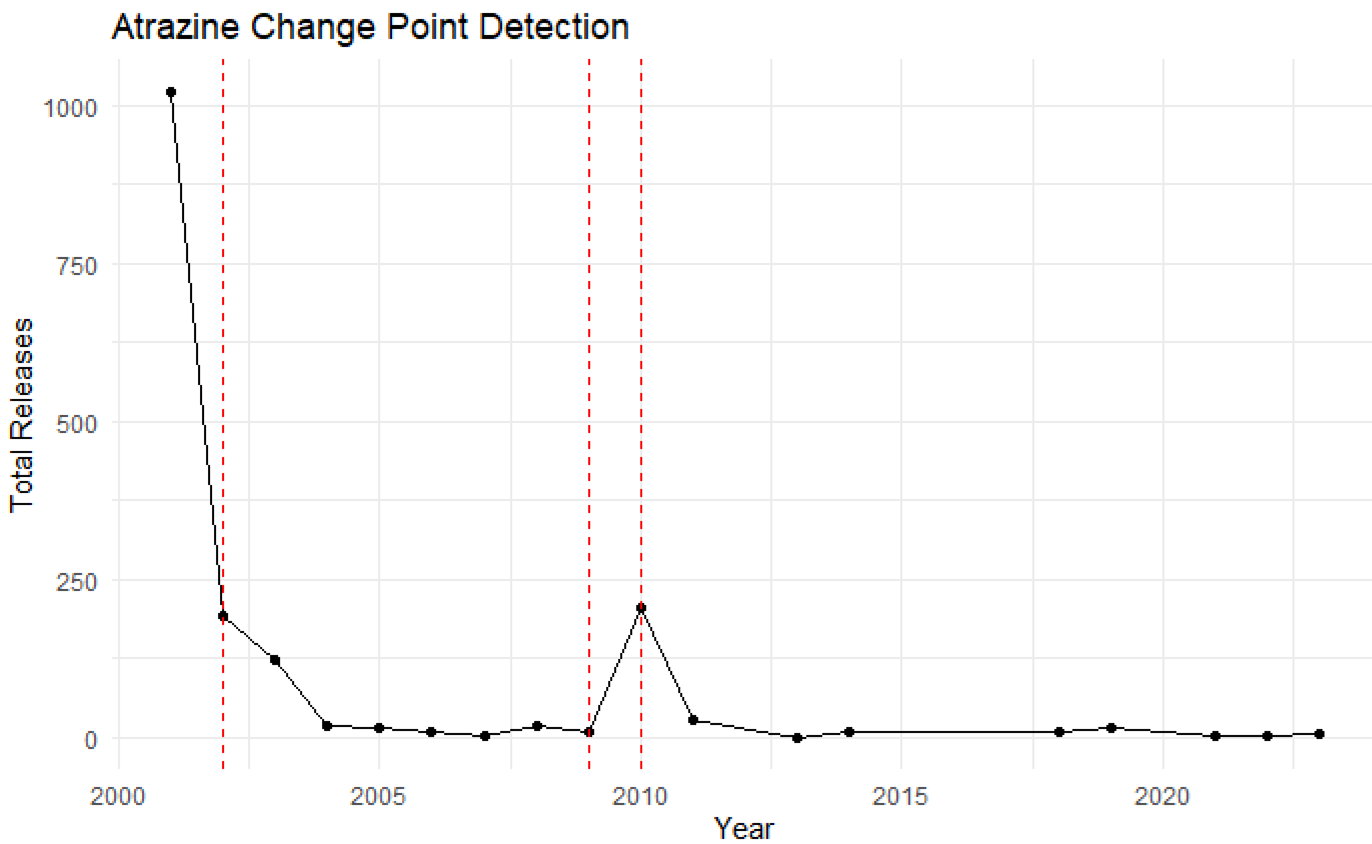
```
analyze_change_points("Benzene", valid_data, change_points)
```



Atrazine change points statewide

Hide

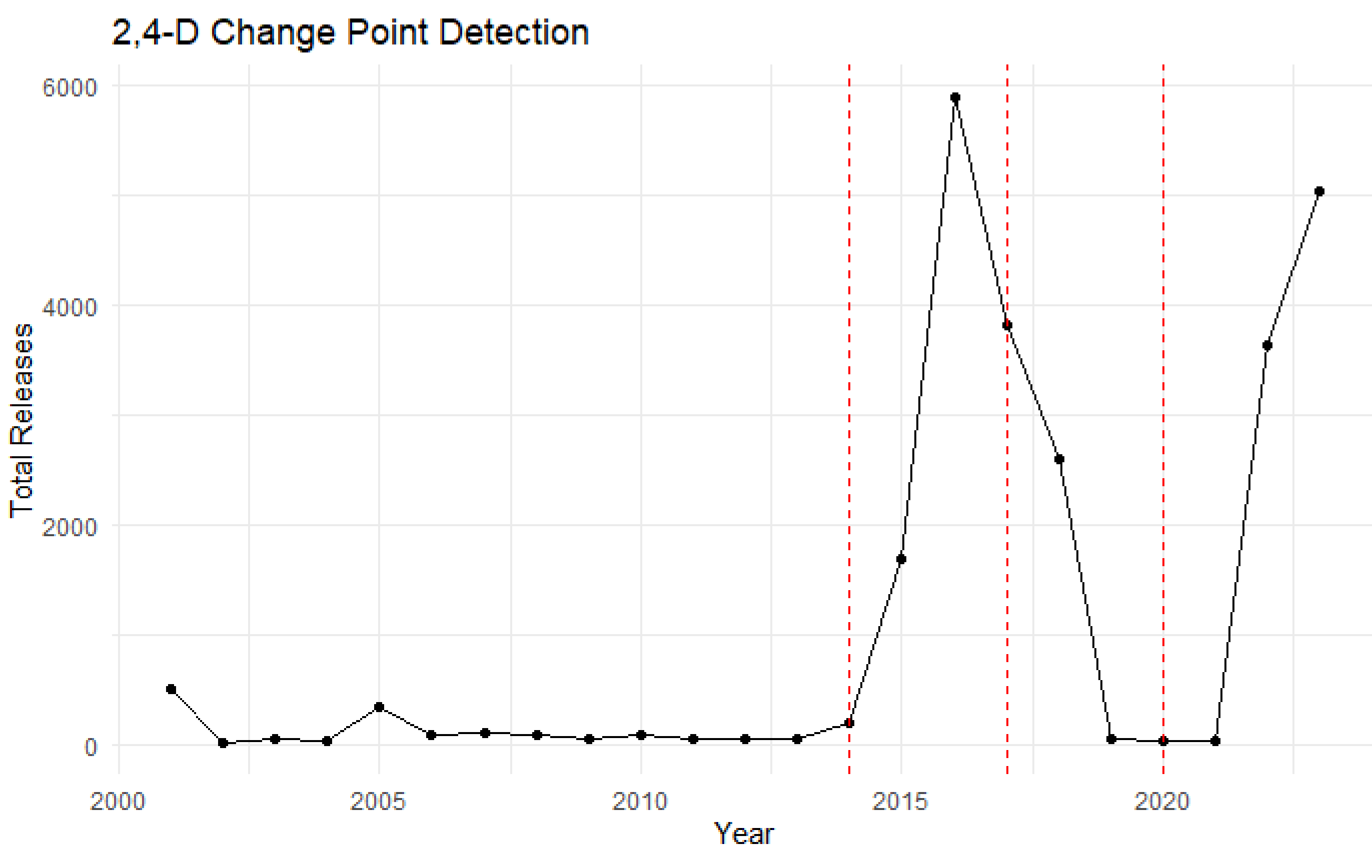
```
analyze_change_points("Atrazine", valid_data, change_points)
```



2,4-D change points statewide

Hide

```
analyze_change_points("2,4-D", valid_data, change_points)
```



7. Geospatial analysis for TRI program data

Average yearly increase function

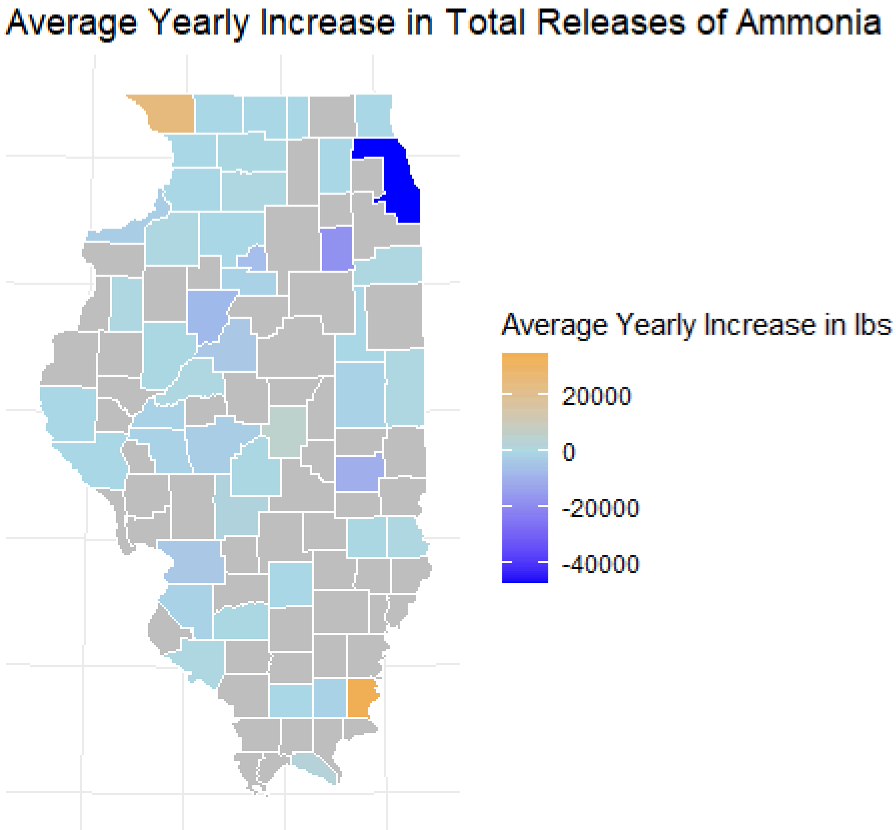
Hide

```
plot_avg_yearly_increase <- function(chemical_name, chemical_by_county, illinois_counties) {

  # Filter by chemical of interest, group by county, arrange by year, calculate yearly change, take the mea
n and drop groups
  avg_increase_data <- chemical_by_county %>%
    filter(chemical == chemical_name) %>%
    group_by(county) %>%
    arrange(year) %>%
    mutate(yearly_change = total_releases - lag(total_releases)) %>%
    summarize(avg_yearly_increase = mean(yearly_change, na.rm = TRUE), .groups = "drop") %>%
    rename(county_name = county)

  # Convert county names to regular format ex "Jo Daviess County"
  avg_increase_data$county_name <- paste(tools::toTitleCase(tolower(avg_increase_data$county_name)), "Count
y")

  # Set up counties data with average increase data by county name
  spatial_data <- left_join(illinois_counties, avg_increase_data, by = "county_name")

  # Plot using Albers projection to minimize warping, remove axis labels, and apply nice color scheme
  plot <- ggplot() +
    geom_polygon(data = spatial_data,
                 mapping = aes(x = long, y = lat, group = group, fill = avg_yearly_increase),
                 color = "#ffffff", size = 0.25) +
    coord_map(projection = "albers", lat0 = 39, lat1 = 45) +
    scale_fill_gradient2(midpoint = 0, low = "blue", mid = "lightblue", high = "orange",
                         na.value = "grey", name = "Average Yearly Increase in lbs") +
    labs(title = paste("Average Yearly Increase in Total Releases of", chemical_name)) +
    theme_minimal() +
    theme(axis.title = element_blank(),
          axis.text = element_blank(),
          axis.ticks = element_blank())

  print(plot)
}
```

Ammonia average yearly increase (Jo Daviess county of interest)

Hide

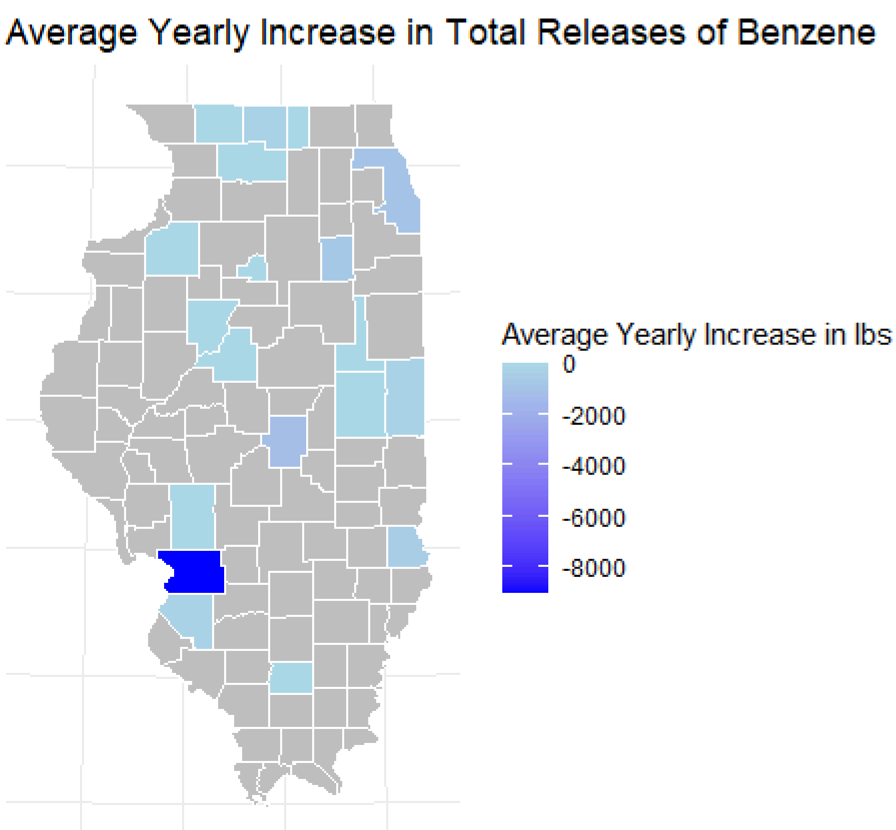
```
plot_avg_yearly_increase("Ammonia", chemical_by_county, illinois_counties)
```



Benzene average yearly increase (Benzene generally decreasing)

Hide

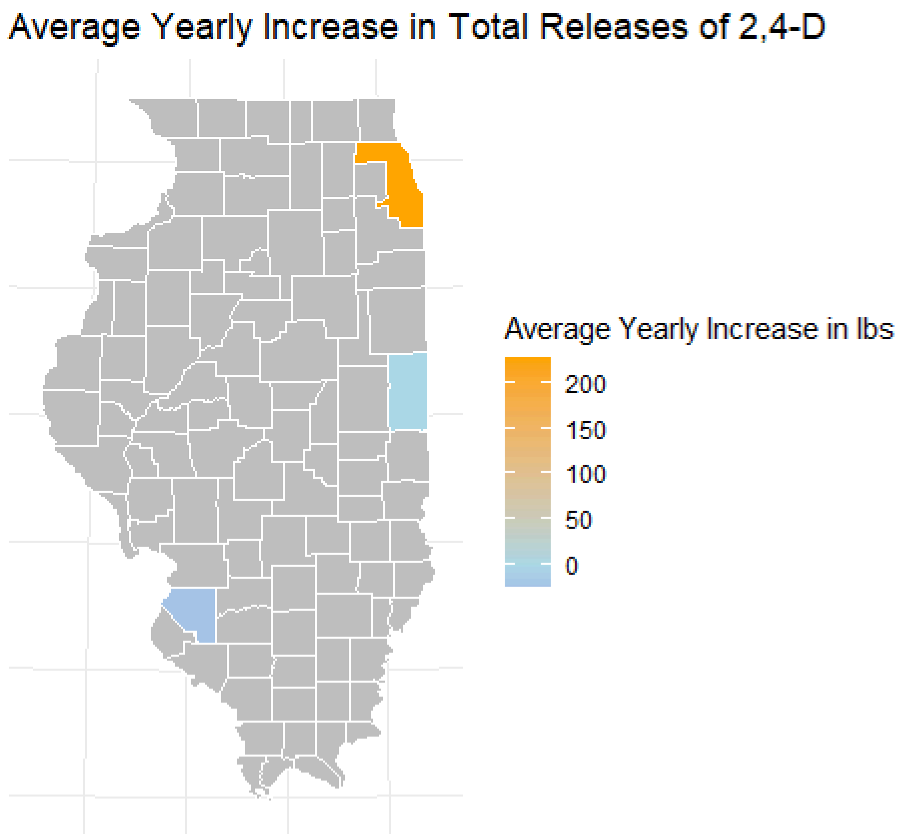
```
plot_avg_yearly_increase("Benzene", chemical_by_county, illinois_counties)
```



2,4-D average yearly increase (Cook county of interest)

Hide

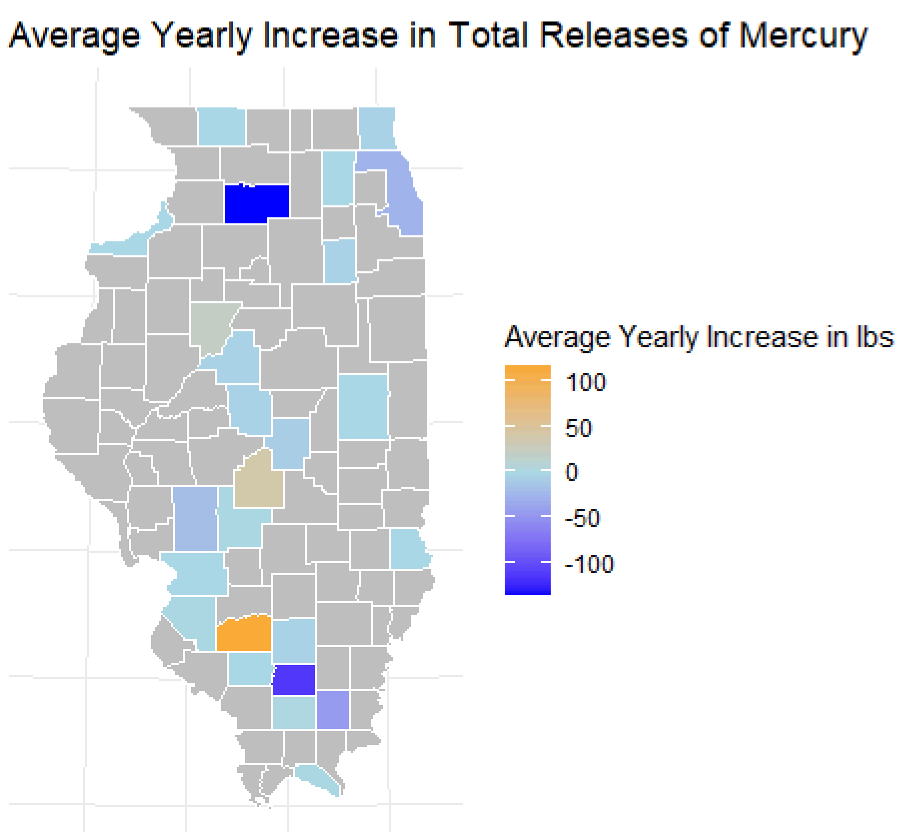
```
plot_avg_yearly_increase("2,4-D", chemical_by_county, illinois_counties)
```



Mercury average yearly increase

Hide

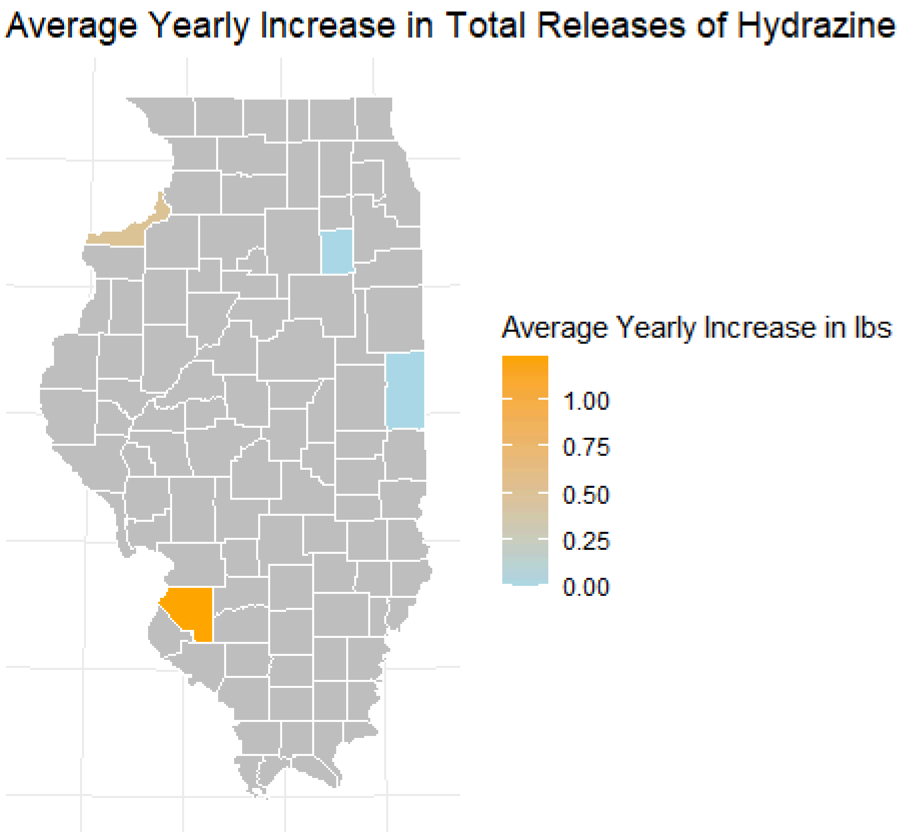
```
plot_avg_yearly_increase("Mercury", chemical_by_county, illinois_counties)
```



Hydrazine average yearly increase

Hide

```
plot_avg_yearly_increase("Hydrazine", chemical_by_county, illinois_counties)
```

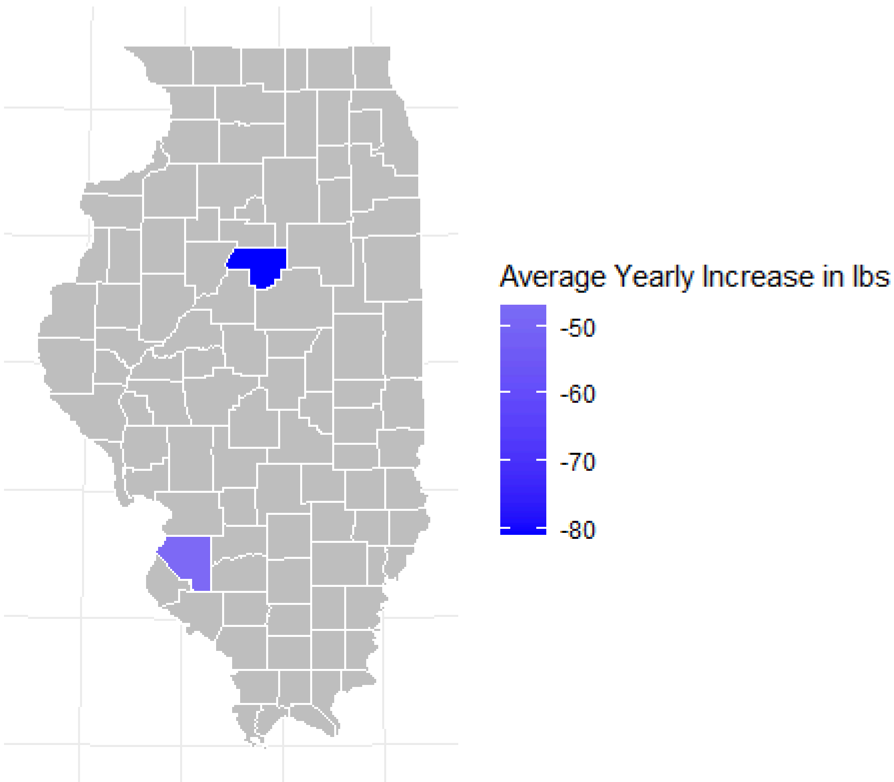


Atrazine average yearly increase

Hide

```
plot_avg_yearly_increase("Atrazine", chemical_by_county, illinois_counties)
```

Average Yearly Increase in Total Releases of Atrazine

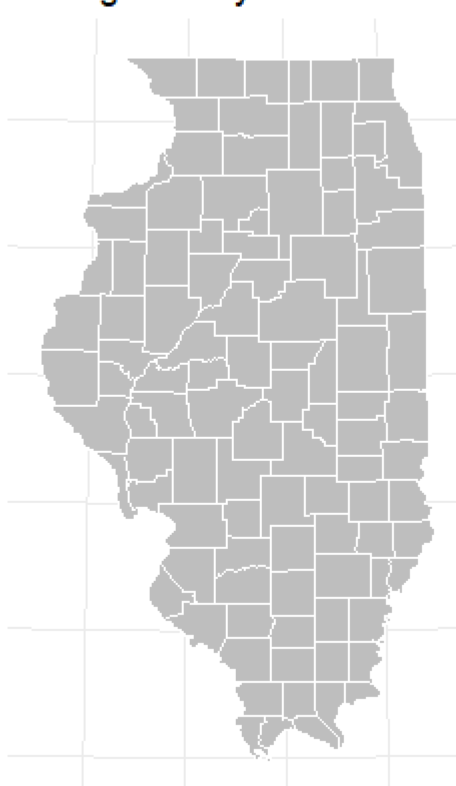


Acetochlor shows no reported releases

Hide

```
plot_avg_yearly_increase("Acetochlor", chemical_by_county, illinois_counties)
```

Average Yearly Increase in Total Releases of Acetochlor



```
USGS WATER QUALITY DATASET ANALYSIS SECTION
```

1. Load in and format the USGS data

Hide

```
chemical_data <- fread("C:/Users/carso/Downloads/waterdata/waterdata.csv", sep = "\t", fill = TRUE, quote =
"")
```

```
|--------------------------------------------------|
|==================================================|
```

Hide

```
# Format date field
chemical_data$sample_dt <- as.Date(chemical_data$sample_dt, format="%Y-%m-%d")
chemical_data <- chemical_data %>% drop_na(result_va)
```

Examine data

Hide

```
head(chemical_data)
```

agency_cd <chr>	site_no <S3: integer64>	sample_dt <date>	sample_tm <chr>	sample_end_dt <S3: IDate>	sample_end_tm <chr>
USGS	3384500	1955-10-01		1955-10-10	
USGS	3384500	1955-10-01		1955-10-10	
USGS	3384500	1955-10-01		1955-10-10	
USGS	3384500	1955-10-01		1955-10-10	
USGS	3384500	1955-10-01		1955-10-10	
USGS	3384500	1955-10-01		1955-10-10	

6 rows | 1-6 of 34 columns

Hide

```
summary(chemical_data)
```

```
  agency_cd           site_no                  sample_dt           sample_tm         sample_end_dt      
sample_end_tm      sample_start_time_datum_cd
 Length:3742875     Min.   :        3336280   Min.   :1901-01-02   Length:3742875     Min.   :1951-05-31  
Length:3742875     Length:3742875            
 Class :character   1st Qu.:        5527500   1st Qu.:1984-11-05   Class :character   1st Qu.:1957-06-30  
Class :character   Class :character          
 Mode  :character   Median :        5572000   Median :1989-10-31   Mode  :character   Median :1981-06-16  
Mode  :character   Mode  :character          
                    Mean   : 63559090058734   Mean   :1992-11-10                      Mean   :1979-03-01  
                    3rd Qu.:        5595200   3rd Qu.:1997-01-07                      3rd Qu.:1990-06-21  
                    Max.   :423784088133401   Max.   :2024-03-08                      Max.   :2020-02-20  
                                              NA's   :3534                            NA's   :3725526     
 tm_datum_rlbty_cd  coll_ent_cd         medium_cd          project_cd          aqfr_cd              tu_id      
body_part_id     hyd_cond_cd        samp_type_cd      
 Length:3742875     Length:3742875     Length:3742875     Length:3742875     Length:3742875     Min.   :      
0    Min.   : 6        Length:3742875     Length:3742875    
 Class :character   Class :character   Class :character   Class :character   Class :character   1st Qu.:      
0    1st Qu.:59        Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :      
0    Median :94        Mode  :character   Mode  :character  
                                                                                                 Mean   : 73
193    Mean   :69                                                  
                                                                                                 3rd Qu.:163
344    3rd Qu.:94                                                  
                                                                                                 Max.   :163
998    Max.   :94                                                  
                                                                                                 NA's   :373
7504   NA's   :3737504                                             
 hyd_event_cd       sample_lab_cm_tx      parm_cd       remark_cd           result_va          val_qual_tx       
meth_cd             dqi_cd             rpt_lev_va     
 Length:3742875     Length:3742875     Min.   :    1   Length:3742875     Min.   :-3.830e+02   Length:37428
75     Length:3742875     Length:3742875     Min.   :    0  
 Class :character   Class :character   1st Qu.:  610   Class :character   1st Qu.: 1.000e+00   Class :chara
cter   Class :character   Class :character   1st Qu.:    0  
 Mode  :character   Mode  :character   Median : 1027   Mode  :character   Median : 8.000e+00   Mode  :chara
cter   Mode  :character   Mode  :character   Median :    1  
                                       Mean   :18787                      Mean   : 9.134e+03                     
Mean   :   28  
                                       3rd Qu.:32730                      3rd Qu.: 6.300e+01                     
3rd Qu.:    8  
                                       Max.   :99995                      Max.   : 1.000e+10                     
Max.   :18950  
                                                                                                                 
NA's   :3308987  
  rpt_lev_cd          lab_std_va      prep_set_no          prep_dt          anl_set_no             anl_dt         
result_lab_cm_tx    anl_ent_cd          V34          
 Length:3742875     Min.   :      0   Length:3742875     Min.   :20010530   Length:3742875     Min.   :2001
0606   Length:3742875     Length:3742875     Mode:logical  
 Class :character   1st Qu.:      0   Class :character   1st Qu.:20110624   Class :character   1st Qu.:2009
1114   Class :character   Class :character   NA's:3742875  
 Mode  :character   Median :      0   Mode  :character   Median :20130916   Mode  :character   Median :2013
0825   Mode  :character   Mode  :character                 
                    Mean   :    955                      Mean   :20137288                      Mean   :2013
3588                                                         
                    3rd Qu.:      1                      3rd Qu.:20170814                      3rd Qu.:2017
0731                                                         
```

```
                    Max.   :1301508                    Max.   :20240221                     Max.   :2024
0307
                    NA's   :3741510                    NA's   :3421702                      NA's   :3282
264
```

2. Load in and format parameter data

Hide

```
param_data <- read.delim("C:/Users/carso/Downloads/waterdata/param_dictionary.txt", sep = "\t", header = FA
LSE, stringsAsFactors = FALSE, col.names = "field")

# Format param data so we can match it. Don't need field
param_data <- param_data %>%
  mutate(
    parm_cd = as.numeric(sub(" - .*", "", field)),
    description = sub("^[0-9]+ - ", "", field)
  ) %>%
  select(-field)
```

3. Function for filtering by chemicals in the parameter codes and using a SARIMA model to forecast

Hide

```
forecast_chemical <- function(chemical_name, start_date, param_data, chemical_data, forecast_horizon = 36)
{

  # Get data for chemical of interest using grepl
  data <- param_data %>%
    filter(grepl(chemical_name, description, ignore.case = TRUE))

  # Get parameter codes that match that chemical and select data after start date
  filtered_chemical_data_after_start <- chemical_data %>%
    filter(parm_cd %in% data$parm_cd, sample_dt > as.Date(start_date)) %>%
    select(site_no, sample_dt, sample_end_dt, result_va, result_lab_cm_tx) %>%
    mutate(year_month = format(sample_dt, "%Y-%m"))

  # Get outlier threshold
  outlier_threshold <- quantile(filtered_chemical_data_after_start$result_va, 0.99, na.rm = TRUE)

  # Remove outliers above the threshold
  filtered_chemical_data_no_outliers <- filtered_chemical_data_after_start %>%
    filter(result_va <= outlier_threshold)

  # Get data by year month and get mean
  data_by_month <- filtered_chemical_data_no_outliers %>%
    group_by(year_month) %>%
    summarise(result_va = mean(result_va, na.rm = TRUE)) %>%
    ungroup() %>%
    arrange(year_month)

  # Convert year month to date
  data_by_month <- data_by_month %>%
    mutate(year_month = as.Date(paste0(year_month, "-01")))

  # Create time series using the data
  result_va_ts_monthly <- ts(
    data_by_month$result_va,
    start = c(
      as.numeric(format(min(data_by_month$year_month), "%Y")),
      as.numeric(format(min(data_by_month$year_month), "%m"))
    ),
    frequency = 12
  )

  # Fit SARIMA model
  sarima_model <- auto.arima(result_va_ts_monthly, seasonal = TRUE)

  # Get forecast
  forecast_monthly <- forecast(sarima_model, h = forecast_horizon)

  # Plot forecast
  plot(
    forecast_monthly,
    main = paste("Forecast of Monthly Average result_va for", chemical_name),
    xlab = "Year",
    ylab = "Average Amount in Water Supply"
  )

  # Extract forecasted values
```

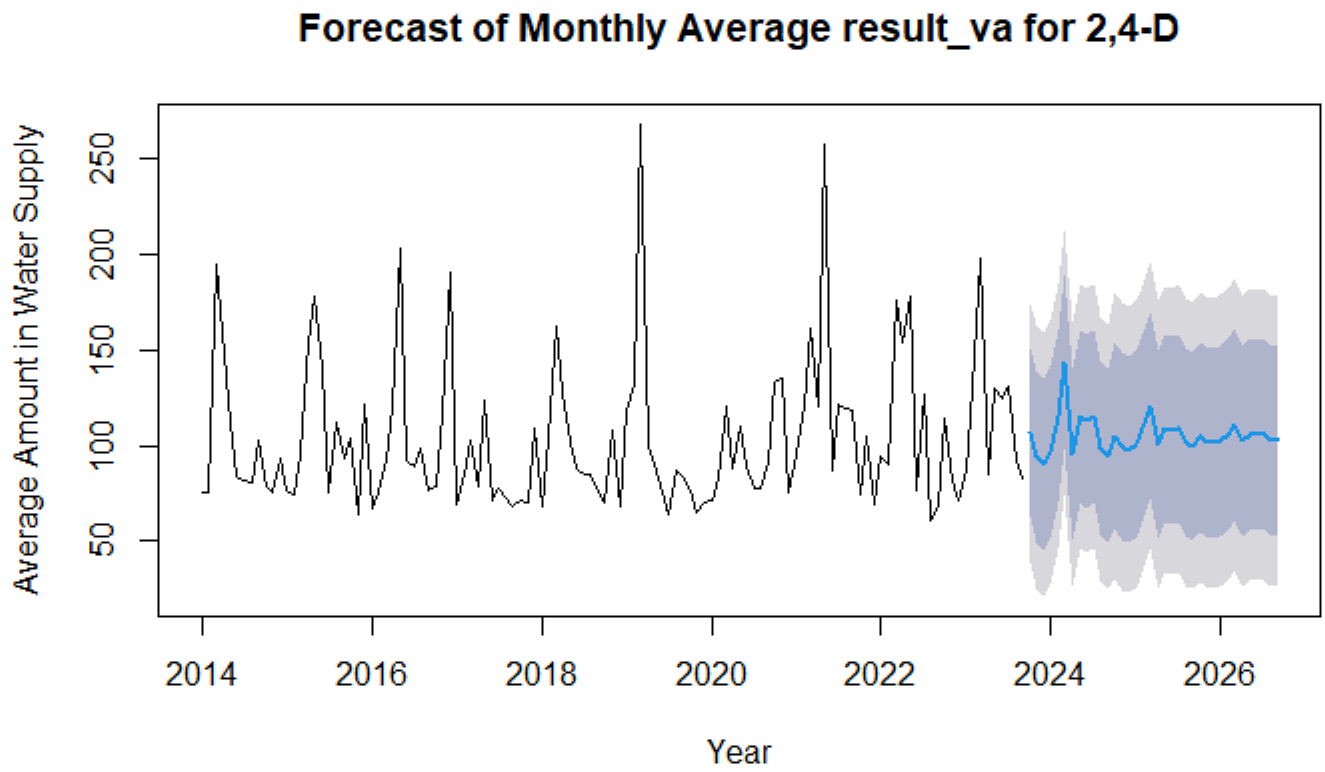
```
  forecast_values <- as.data.frame(forecast_monthly)$`Point Forecast`

  # Calculate the difference between the first and last forecasted values
  forecast_increase <- forecast_values[length(forecast_values)] - forecast_values[1]

  # Print increase or decrease
  if (forecast_increase > 0) {
    cat("The forecast indicates an increase of", forecast_increase, "over the next", forecast_horizon, "mon
ths.\n")
  } else if (forecast_increase < 0) {
    cat("The forecast indicates a decrease of", abs(forecast_increase), "over the next", forecast_horizon,
"months.\n")
  } else {
    cat("The forecast indicates no significant change over the forecast period.")
  }

  # Print summary
  print(summary(sarima_model))

  # Return the forecast and model summary
  return(list(
    model = sarima_model,
    forecast = forecast_monthly
  ))

}
```

2,4-D SARIMA forecast post 2014 since the EPA approved use in 2014. It picks up a decrease but when reading the chart you can see its just the top end of the season its starting at so note this as possible increase

Hide

```
result <- forecast_chemical(
  chemical_name = "2,4-D",
  start_date = "2014-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates a decrease of 4.559721 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
         ar1    sar1      mean
      0.2180  0.4260  103.8303
s.e.  0.0899  0.0846    6.4850

sigma^2 = 1177:  log likelihood = -579.34
AIC=1166.69   AICc=1167.05   BIC=1177.74

Training set error measures:
                    ME    RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 0.1651061 33.8604 25.16816 -7.974954 24.06877 0.8373728 -0.006531367
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean

Coefficients:
         ar1    sar1      mean
      0.2180  0.4260  103.8303
s.e.  0.0899  0.0846    6.4850

sigma^2 = 1177:  log likelihood = -579.34
AIC=1166.69   AICc=1167.05   BIC=1177.74

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	106.73410	62.77299	150.6952	39.50138	173.9668
Nov 2023	93.95748	48.96375	138.9512	25.14550	162.7695
Dec 2023	90.04299	45.00077	135.0852	21.15685	158.9291
Jan 2024	96.62030	51.57577	141.6648	27.73064	165.5100
Feb 2024	115.11246	70.06782	160.1571	46.22263	184.0023
Mar 2024	143.82212	98.77747	188.8668	74.93227	212.7120

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	95.64435	50.59971	140.6890	26.75451	164.5342
May 2024	114.85187	69.80723	159.8965	45.96203	183.7417
Jun 2024	112.65474	67.61009	157.6994	43.76490	181.5446
Jul 2024	115.21937	70.17472	160.2640	46.32952	184.1092

1-10 of 36 rows Previous **1** 2 3 4 Next

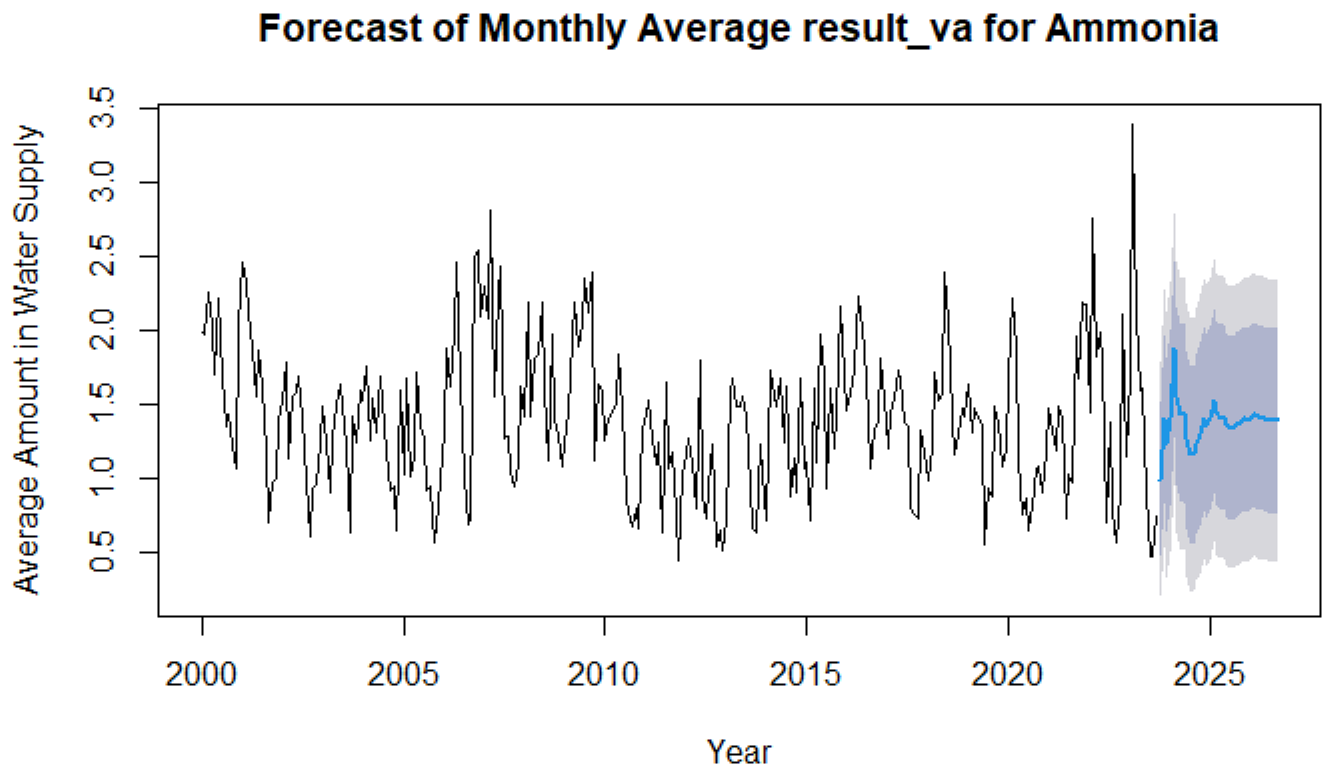
```
NA
```

Ammonia SARIMA forecast. Slight increase detected though it starts at the bottom of the season. Peaks are higher for years post 2020.

Hide

```
result <- forecast_chemical(
  chemical_name = "Ammonia",
  start_date = "2000-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates an increase of 0.3990033 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(1,0,1)(1,0,0)[12] with non-zero mean

Coefficients:
         ar1      ma1    sar1    mean
      0.6366  -0.1770  0.2611  1.4056
s.e.  0.0826   0.1033  0.0624  0.0713

sigma^2 = 0.1626:  log likelihood = -144.12
AIC=298.24   AICc=298.45   BIC=316.5

Training set error measures:
                       ME      RMSE       MAE       MPE     MAPE      MASE         ACF1
Training set -0.003353968 0.4003976 0.3046954 -10.10617 25.75158 0.7259315 -0.006283936
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,1)(1,0,0)[12] with non-zero mean

Coefficients:
         ar1      ma1    sar1    mean
      0.6366  -0.1770  0.2611  1.4056
s.e.  0.0826   0.1033  0.0624  0.0713

sigma^2 = 0.1626:  log likelihood = -144.12
AIC=298.24   AICc=298.45   BIC=316.5

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	0.994671	0.4779016	1.511440	0.2043404	1.785002
Nov 2023	1.409832	0.8411104	1.978553	0.5400476	2.279615
Dec 2023	1.225905	0.6374376	1.814373	0.3259215	2.125889
Jan 2024	1.408235	0.8119518	2.004518	0.4962985	2.320171
Feb 2024	1.878981	1.2795596	2.478401	0.9622454	2.795716
Mar 2024	1.557139	0.9564509	2.157826	0.6384660	2.475811

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	1.435825	0.8346245	2.037025	0.5163684	2.355281
May 2024	1.447479	0.8460711	2.048887	0.5277050	2.367253
Jun 2024	1.265575	0.6640833	1.867067	0.3456727	2.185478
Jul 2024	1.158952	0.5574259	1.760478	0.2389973	2.078906

1-10 of 36 rows Previous **1** 2 3 4 Next

```
NA
```

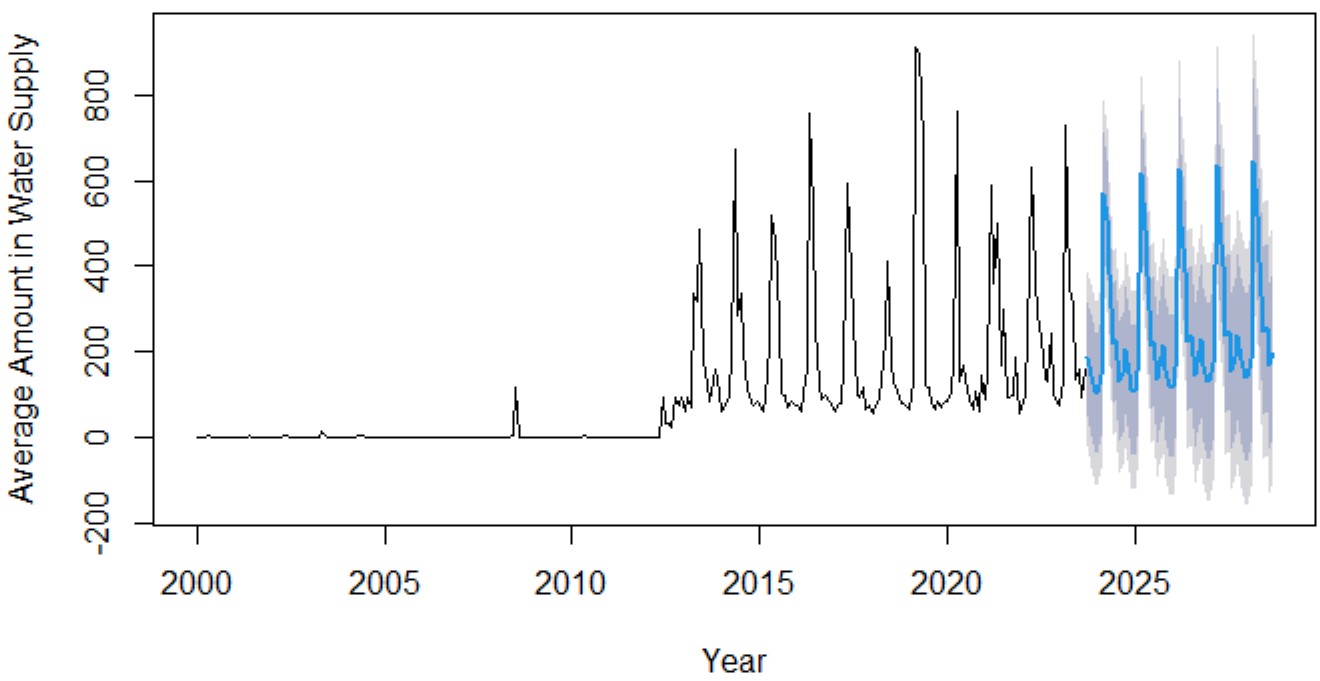
Atrazine SARIMA forecast. It picks up a decrease but when reading the chart you can see its just the top end of the season its starting at. Note this as possible increase because the bottom line of the prediction looks like it is increasing.

Hide

```
result <- forecast_chemical(
  chemical_name = "Atrazine",
  start_date = "2000-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 60
)
```

```
The forecast indicates an increase of 9.810902 over the next 60 months.
Series: result_va_ts_monthly
ARIMA(1,0,0)(0,1,2)[12] with drift

Coefficients:
         ar1     sma1    sma2   drift
      0.3471  -0.6308  0.1374  0.8493
s.e.  0.0586   0.0626  0.0596  0.4157

sigma^2 = 10762:  log likelihood = -1655.09
AIC=3320.18   AICc=3320.4   BIC=3338.23

Training set error measures:
                    ME     RMSE     MAE       MPE     MAPE      MASE        ACF1
Training set -0.261197 100.7862 43.3391 -5201.369 5237.927 0.8554528 -0.02752251
```

Forecast of Monthly Average result_va for Atrazine



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,0)(0,1,2)[12] with drift

Coefficients:
         ar1     sma1    sma2   drift
      0.3471  -0.6308  0.1374  0.8493
s.e.  0.0586   0.0626  0.0596  0.4157

sigma^2 = 10762:  log likelihood = -1655.09
AIC=3320.18   AICc=3320.4   BIC=3338.23

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	185.3289	52.380483	318.2774	-17.998174	388.6560
Nov 2023	146.7128	5.983487	287.4421	-68.514116	361.9397
Dec 2023	105.3472	-36.290642	246.9851	-111.269209	321.9637
Jan 2024	102.6465	-39.100410	244.3935	-114.136714	319.4298
Feb 2024	143.6000	1.839930	285.3601	-73.203327	360.4034
Mar 2024	571.4437	429.682035	713.2054	354.637941	788.2495

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	509.5392	367.777377	651.3011	292.733182	726.3453
May 2024	382.9636	241.201671	524.7254	166.157463	599.7696
Jun 2024	219.1350	77.373074	360.8968	2.328865	435.9411
Jul 2024	228.8925	87.130655	370.6544	12.086446	445.6986

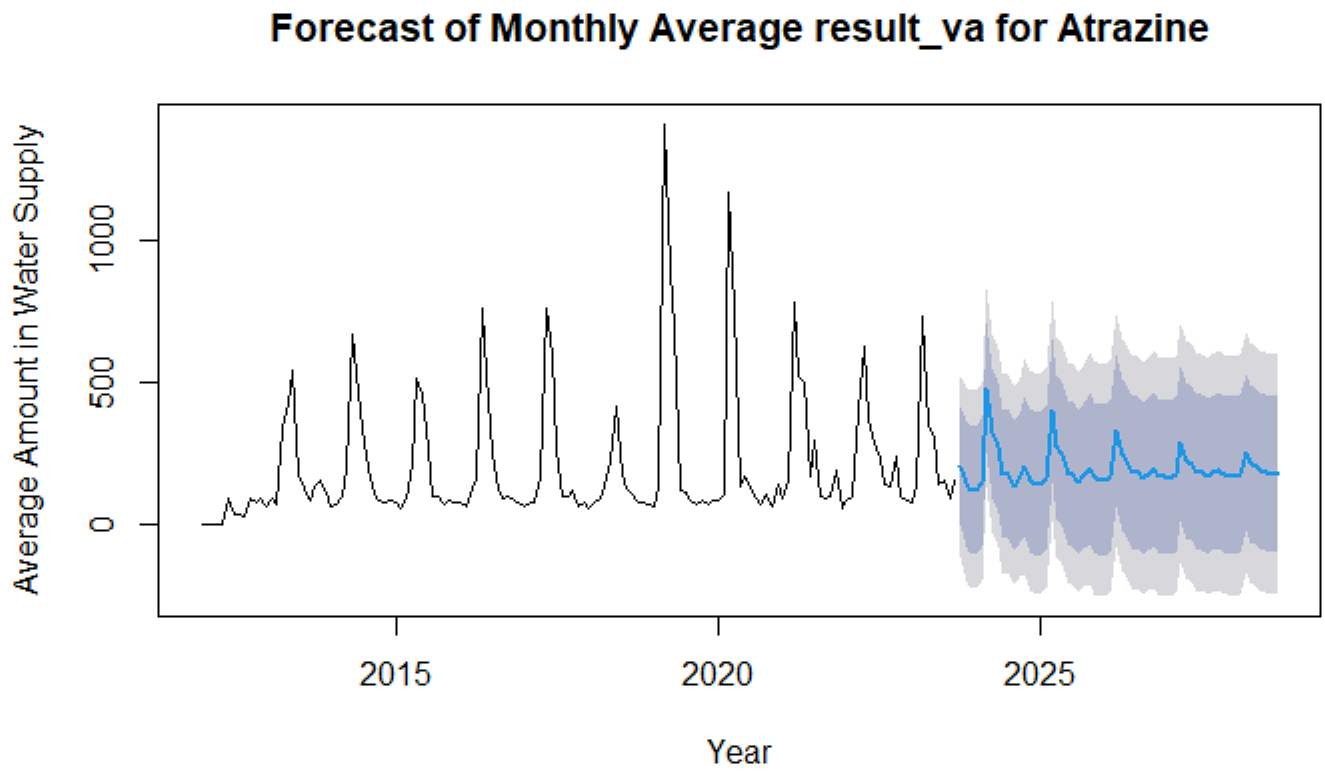
1-10 of 60 rows   Previous **1** 2 3 4 5 6 Next

```
NA
```

Hide

```
result <- forecast_chemical(
  chemical_name = "Atrazine",
  start_date = "2012-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 60
)
```

```
The forecast indicates a decrease of 24.82873 over the next 60 months.
Series: result_va_ts_monthly
ARIMA(1,0,0)(2,0,0)[12] with non-zero mean

Coefficients:
         ar1    sar1    sar2      mean
      0.4224  0.4888  0.1316  186.4247
s.e.  0.0769  0.0848  0.0850   52.9737

sigma^2 = 26271:  log likelihood = -918.06
AIC=1846.12   AICc=1846.56   BIC=1860.86

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 7.044405 159.7668 80.60892 -3622.318 3639.897 0.8120273 0.01773624
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,0)(2,0,0)[12] with non-zero mean

Coefficients:
         ar1    sar1    sar2      mean
      0.4224  0.4888  0.1316  186.4247
s.e.  0.0769  0.0848  0.0850   52.9737

sigma^2 = 26271:  log likelihood = -918.06
AIC=1846.12   AICc=1846.56   BIC=1860.86

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	206.7542	-0.9627693	414.4711	-110.92147	524.4298
Nov 2023	144.3263	-81.1631498	369.8158	-200.53006	489.1827
Dec 2023	119.8762	-108.6393367	348.3918	-229.60817	469.3606
Jan 2024	119.2531	-109.7982370	348.3045	-231.05070	469.5569
Feb 2024	150.2334	-78.9134265	379.3802	-200.21643	500.6832
Mar 2024	481.0344	251.8705034	710.1982	130.55849	831.5102

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	321.8427	92.6757852	551.0096	-28.63784	672.3232
May 2024	272.9874	43.8199581	502.1548	-77.49396	623.4687
Jun 2024	175.2758	-53.8917552	404.4433	-175.20572	525.7573
Jul 2024	178.1494	-51.0181315	407.3170	-172.33211	528.6309

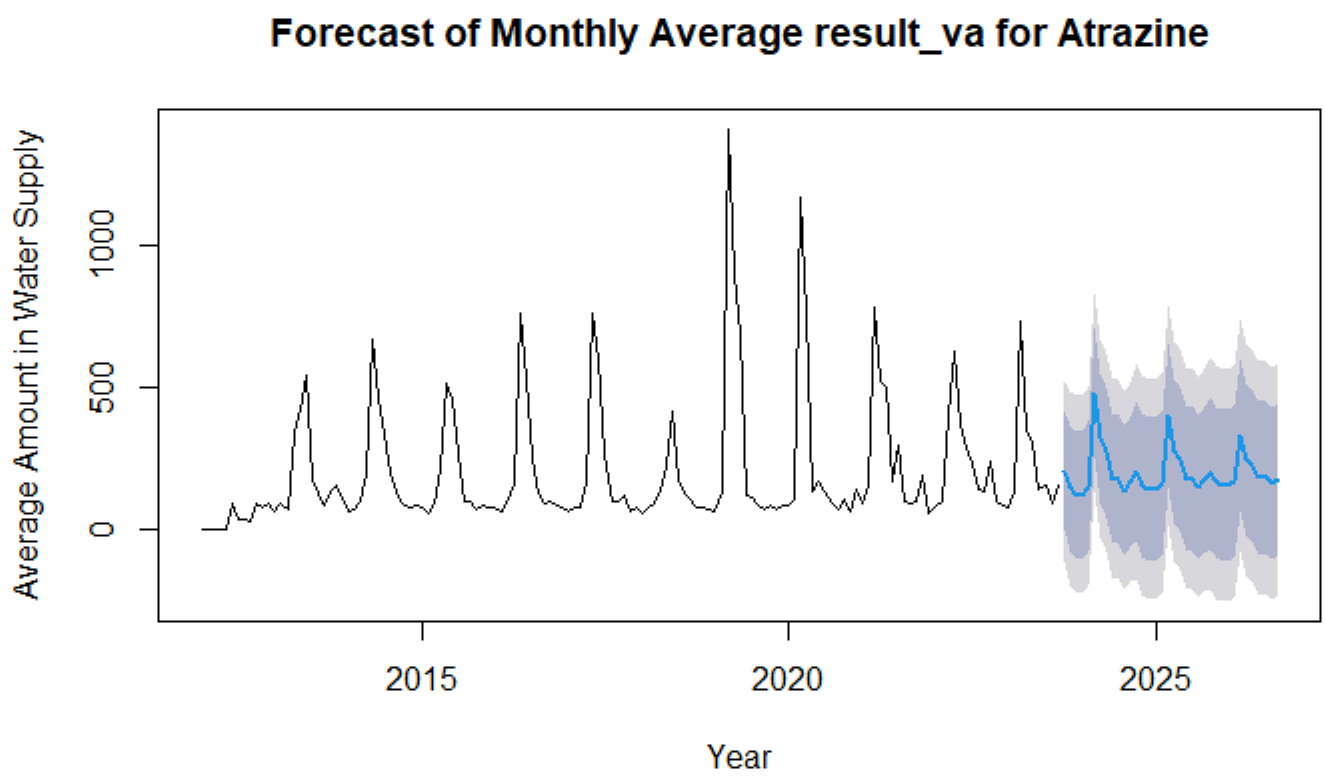
1-10 of 60 rows Previous **1** 2 3 4 5 6 Next

```
NA
```

Hide

```
result <- forecast_chemical(
  chemical_name = "Atrazine",
  start_date = "2012-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates a decrease of 30.01999 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(1,0,0)(2,0,0)[12] with non-zero mean

Coefficients:
         ar1    sar1    sar2      mean
      0.4224  0.4888  0.1316  186.4247
s.e.  0.0769  0.0848  0.0850   52.9737

sigma^2 = 26271:  log likelihood = -918.06
AIC=1846.12   AICc=1846.56   BIC=1860.86

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 7.044405 159.7668 80.60892 -3622.318 3639.897 0.8120273 0.01773624
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,0)(2,0,0)[12] with non-zero mean

Coefficients:
         ar1     sar1     sar2      mean
      0.4224   0.4888   0.1316  186.4247
s.e.  0.0769   0.0848   0.0850   52.9737

sigma^2 = 26271:  log likelihood = -918.06
AIC=1846.12   AICc=1846.56   BIC=1860.86

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	206.7542	-0.9627693	414.4711	-110.92147	524.4298
Nov 2023	144.3263	-81.1631498	369.8158	-200.53006	489.1827
Dec 2023	119.8762	-108.6393367	348.3918	-229.60817	469.3606
Jan 2024	119.2531	-109.7982370	348.3045	-231.05070	469.5569
Feb 2024	150.2334	-78.9134265	379.3802	-200.21643	500.6832
Mar 2024	481.0344	251.8705034	710.1982	130.55849	831.5102

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	321.8427	92.6757852	551.0096	-28.63784	672.3232
May 2024	272.9874	43.8199581	502.1548	-77.49396	623.4687
Jun 2024	175.2758	-53.8917552	404.4433	-175.20572	525.7573
Jul 2024	178.1494	-51.0181315	407.3170	-172.33211	528.6309

1-10 of 36 rows Previous **1** 2 3 4 Next

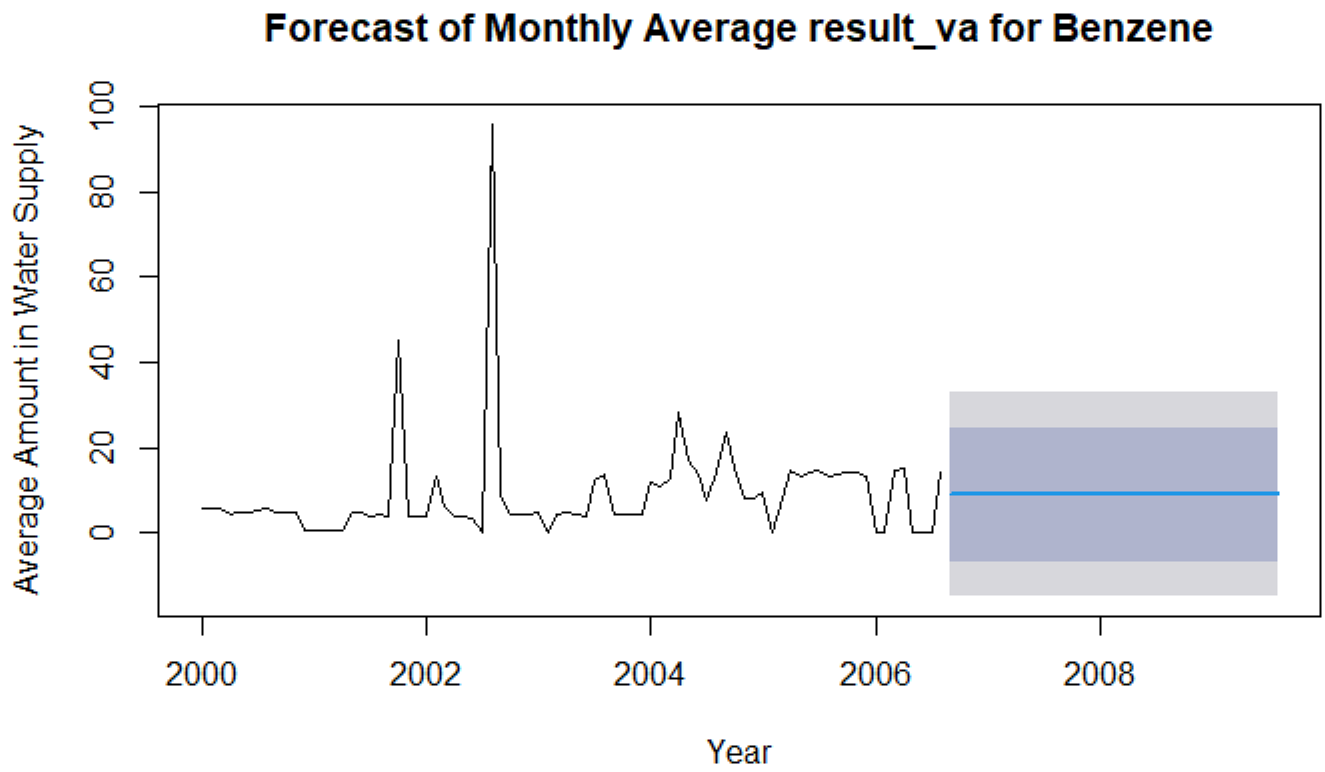
```
NA
```

Benzene SARIMA forecast. Not enough data to predict anything with the model.

Hide

```
result <- forecast_chemical(
  chemical_name = "Benzene",
  start_date = "2000-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates no significant change over the forecast period.Series: result_va_ts_monthly
ARIMA(0,0,0) with non-zero mean

Coefficients:
        mean
      8.8200
s.e.  1.3542

sigma^2 = 148.6:  log likelihood = -313.05
AIC=630.1   AICc=630.26   BIC=634.87

Training set error measures:
                        ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set -3.098129e-15 12.11215 6.678332 -1437.831 1465.066 0.7630948 0.04636652
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(0,0,0) with non-zero mean

Coefficients:
        mean
      8.8200
s.e.  1.3542

sigma^2 = 148.6:  log likelihood = -313.05
AIC=630.1   AICc=630.26   BIC=634.87

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Sep 2006	8.819998	-6.80028	24.44028	-15.06916	32.70915
Oct 2006	8.819998	-6.80028	24.44028	-15.06916	32.70915
Nov 2006	8.819998	-6.80028	24.44028	-15.06916	32.70915
Dec 2006	8.819998	-6.80028	24.44028	-15.06916	32.70915
Jan 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915
Feb 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Mar 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915
Apr 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915
May 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915
Jun 2007	8.819998	-6.80028	24.44028	-15.06916	32.70915

1-10 of 36 rows Previous **1** 2 3 4 Next

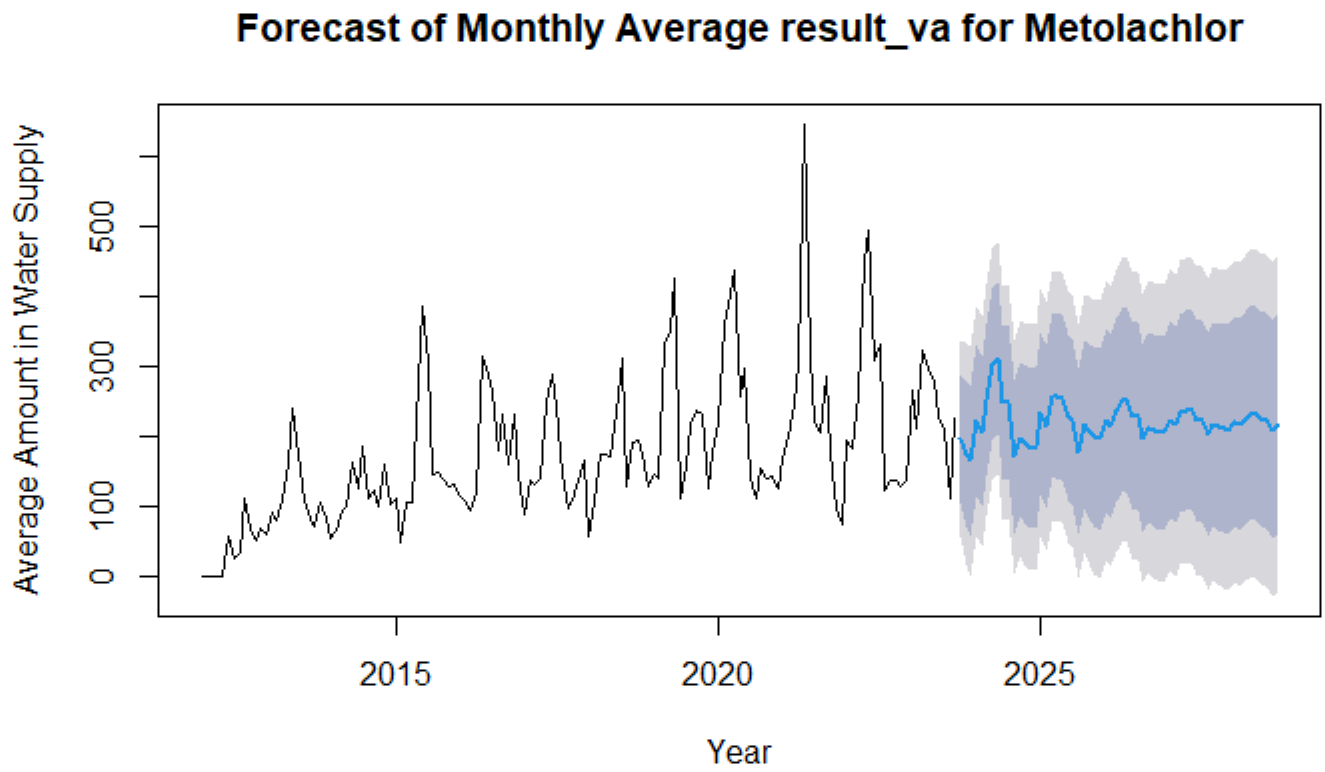
```
NA
```

Metolachlor SARIMA forecast. Found post 2012 was where most measurements happen. Increase detected

Hide

```
result <- forecast_chemical(
  chemical_name = "Metolachlor",
  start_date = "2012-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 60
)
```

```
The forecast indicates an increase of 20.30976 over the next 60 months.
Series: result_va_ts_monthly
ARIMA(3,1,1)(2,0,0)[12]

Coefficients:
         ar1      ar2      ar3      ma1    sar1    sar2
      0.4670  -0.0092  -0.0830  -0.9430  0.2009  0.2915
s.e.  0.0938   0.0947   0.0904   0.0306  0.0843  0.0856

sigma^2 = 5145:  log likelihood = -795.94
AIC=1605.88   AICc=1606.73   BIC=1626.48

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 8.180157 69.92323 48.94842 -2.731818 30.29614 0.6899594 -0.01712536
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(3,1,1)(2,0,0)[12]

Coefficients:
         ar1      ar2      ar3      ma1    sar1    sar2
      0.4670  -0.0092  -0.0830  -0.9430  0.2009  0.2915
s.e.  0.0938   0.0947   0.0904   0.0306  0.0843  0.0856

sigma^2 = 5145:  log likelihood = -795.94
AIC=1605.88   AICc=1606.73   BIC=1626.48

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	197.0880	105.16698	289.0090	56.5069437	337.6690
Nov 2023	175.6214	71.84622	279.3967	16.9109318	334.3320
Dec 2023	165.2023	58.00066	272.4039	1.2515446	329.1530
Jan 2024	223.1507	115.50947	330.7920	58.5276290	387.7738
Feb 2024	207.5748	99.79162	315.3581	42.7346376	372.4150
Mar 2024	252.7217	144.79833	360.6450	87.6671505	417.7762

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	302.7518	194.60507	410.8986	137.3556384	468.1480
May 2024	312.4816	204.04138	420.9218	146.6366019	478.3265
Jun 2024	247.9121	139.14397	356.6803	81.5655813	414.2587
Jul 2024	250.5685	141.46640	359.6705	83.7112589	417.4257

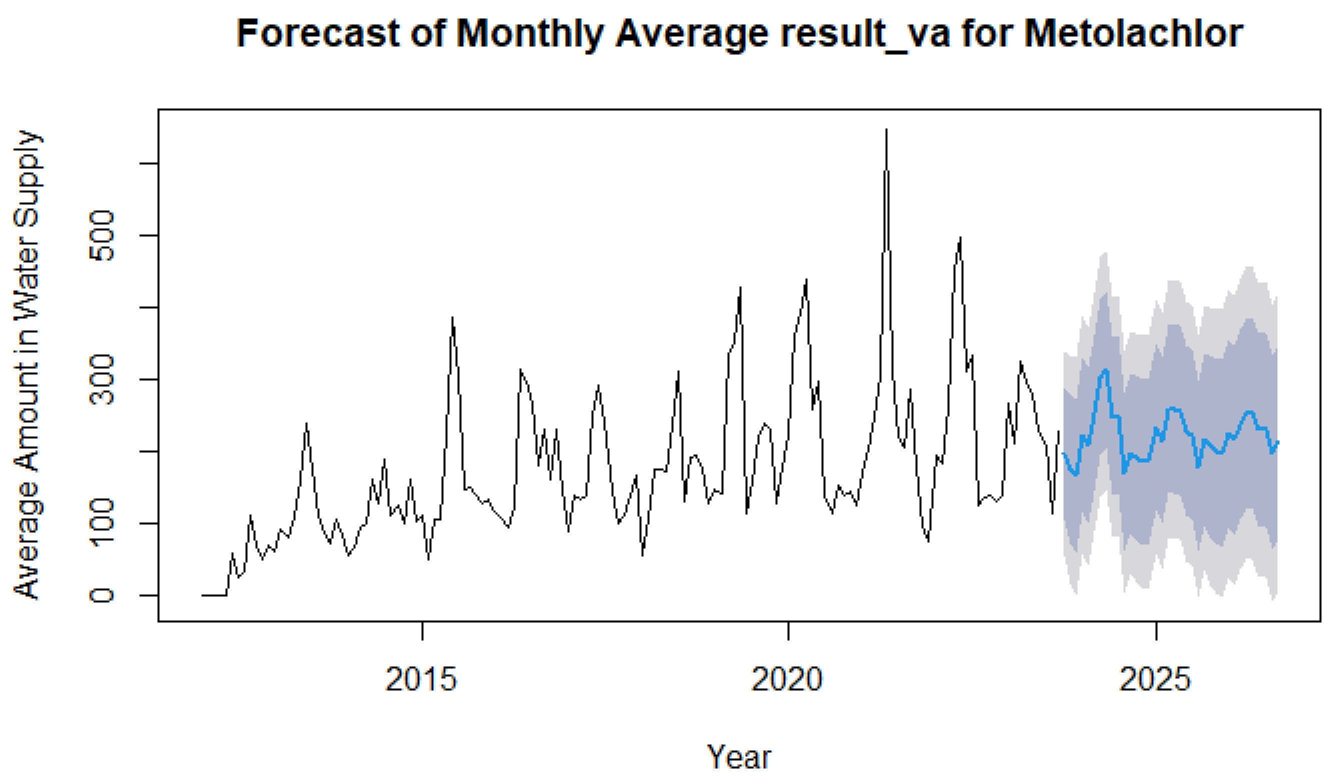
1-10 of 60 rows Previous **1** 2 3 4 5 6 Next

```
NA
```

Hide

```
result <- forecast_chemical(
  chemical_name = "Metolachlor",
  start_date = "2012-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates an increase of 15.71984 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(3,1,1)(2,0,0)[12]

Coefficients:
         ar1      ar2      ar3      ma1    sar1    sar2
      0.4670  -0.0092  -0.0830  -0.9430  0.2009  0.2915
s.e.  0.0938   0.0947   0.0904   0.0306  0.0843  0.0856

sigma^2 = 5145:  log likelihood = -795.94
AIC=1605.88   AICc=1606.73   BIC=1626.48

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 8.180157 69.92323 48.94842 -2.731818 30.29614 0.6899594 -0.01712536
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(3,1,1)(2,0,0)[12]

Coefficients:
         ar1      ar2      ar3      ma1    sar1    sar2
      0.4670  -0.0092  -0.0830  -0.9430  0.2009  0.2915
s.e.  0.0938   0.0947   0.0904   0.0306  0.0843  0.0856

sigma^2 = 5145:  log likelihood = -795.94
AIC=1605.88   AICc=1606.73   BIC=1626.48

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	197.0880	105.16698	289.0090	56.506944	337.6690
Nov 2023	175.6214	71.84622	279.3967	16.910932	334.3320
Dec 2023	165.2023	58.00066	272.4039	1.251545	329.1530
Jan 2024	223.1507	115.50947	330.7920	58.527629	387.7738
Feb 2024	207.5748	99.79162	315.3581	42.734638	372.4150
Mar 2024	252.7217	144.79833	360.6450	87.667151	417.7762

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	302.7518	194.60507	410.8986	137.355638	468.1480
May 2024	312.4816	204.04138	420.9218	146.636602	478.3265
Jun 2024	247.9121	139.14397	356.6803	81.565581	414.2587
Jul 2024	250.5685	141.46640	359.6705	83.711259	417.4257

1-10 of 36 rows Previous **1** 2 3 4 Next

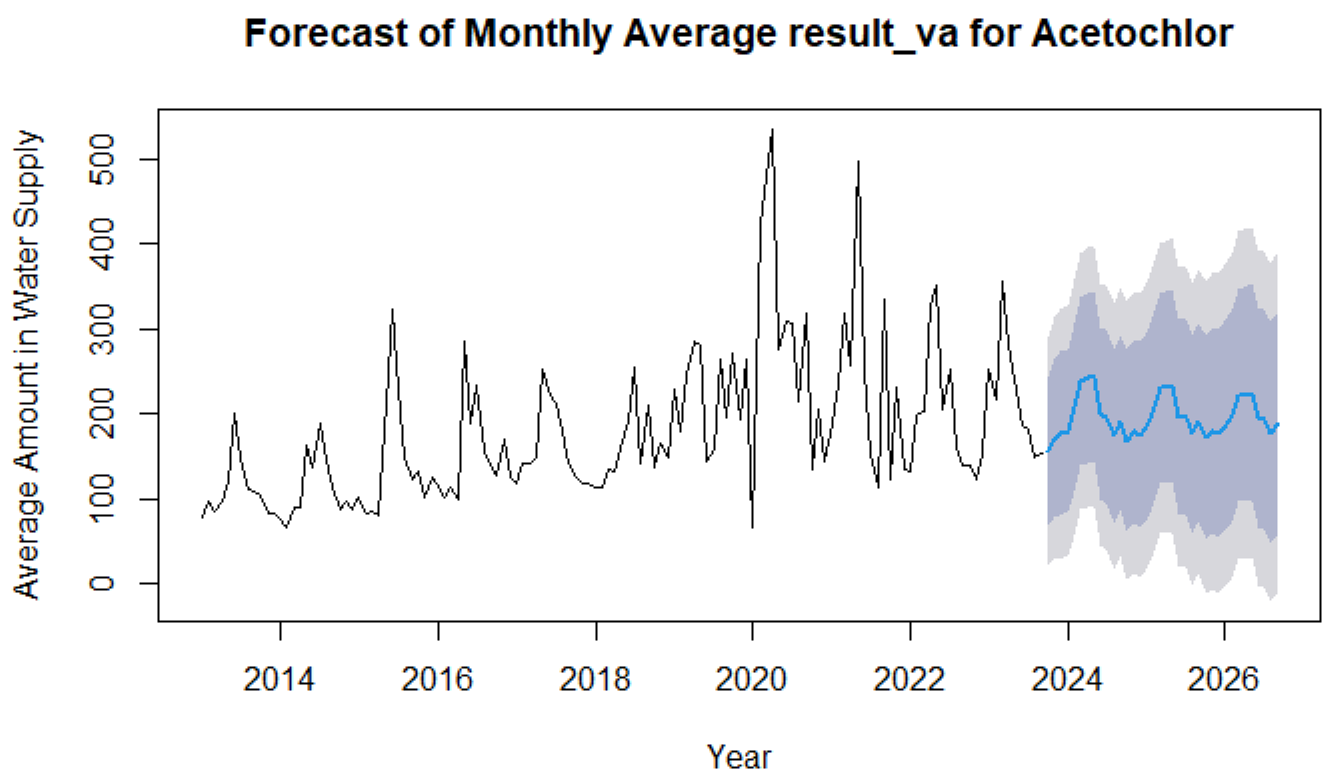
```
NA
```

Acetochlor SARIMA forecast. Increase detected

Hide

```
result <- forecast_chemical(
  chemical_name = "Acetochlor",
  start_date = "2013-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates an increase of 33.41419 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(1,1,3)(1,0,1)[12]

Coefficients:
          ar1     ma1      ma2      ma3    sar1      sma1
      -0.7580  0.1155  -0.5537  -0.3269  0.8118   -0.6299
s.e.   0.1298  0.1511   0.1120   0.0859  0.1512    0.1961

sigma^2 = 4671:  log likelihood = -720.85
AIC=1455.7   AICc=1456.63   BIC=1475.66

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE         ACF1
Training set 2.732303 66.46761 45.09393 -7.075041 24.36949 0.7112018 -0.001749886
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,1,3)(1,0,1)[12]

Coefficients:
          ar1     ma1      ma2      ma3    sar1     sma1
      -0.7580  0.1155  -0.5537  -0.3269  0.8118  -0.6299
s.e.   0.1298  0.1511   0.1120   0.0859  0.1512   0.1961

sigma^2 = 4671:  log likelihood = -720.85
AIC=1455.7   AICc=1456.63   BIC=1475.66

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	155.9643	68.37263	243.5561	22.004373	289.9243
Nov 2023	170.8929	77.87231	263.9136	28.630168	313.1557
Dec 2023	176.1579	79.71257	272.6032	28.657487	323.6583
Jan 2024	180.2952	83.84165	276.7487	32.782230	327.8081
Feb 2024	207.0990	108.67142	305.5265	56.567032	357.6309
Mar 2024	238.2643	139.67166	336.8569	87.479886	389.0486

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	242.3073	142.38497	342.2296	89.489283	395.1253
May 2024	243.8287	143.56596	344.0914	90.490089	397.1673
Jun 2024	199.4642	98.18505	300.7434	44.571090	354.3574
Jul 2024	196.7184	94.97382	298.4631	41.113471	352.3234

1-10 of 36 rows Previous **1** 2 3 4 Next

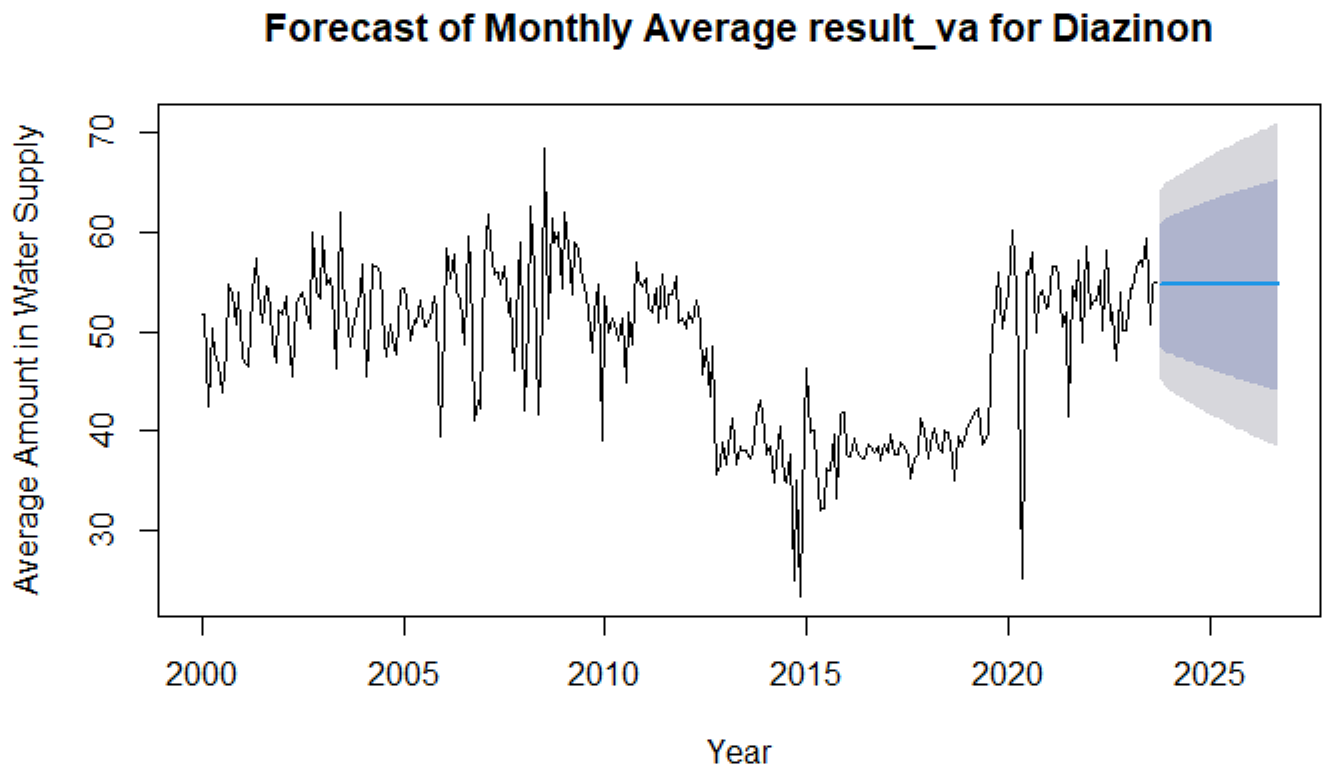
```
NA
```

Diazinon SARIMA forecast. Data has inconsistent drops and peaks, no forecast

Hide

```
result <- forecast_chemical(
  chemical_name = "Diazinon",
  start_date = "2000-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates a decrease of 0.02370768 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(0,1,2)

Coefficients:
          ma1      ma2
      -0.6385  -0.1300
s.e.   0.0579   0.0557

sigma^2 = 23.09:  log likelihood = -848.16
AIC=1702.31   AICc=1702.4   BIC=1713.26

Training set error measures:
                     ME     RMSE      MAE        MPE     MAPE      MASE         ACF1
Training set 0.06639633 4.779556 3.344419 -0.8824121 7.284437 0.6176786 -0.001053708
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(0,1,2)

Coefficients:
          ma1      ma2
      -0.6385  -0.1300
s.e.   0.0579   0.0557

sigma^2 = 23.09:  log likelihood = -848.16
AIC=1702.31   AICc=1702.4   BIC=1713.26

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	54.67059	48.51285	60.82833	45.25313	64.08804
Nov 2023	54.64688	48.09910	61.19466	44.63292	64.66084
Dec 2023	54.64688	47.94575	61.34801	44.39839	64.89538
Jan 2024	54.64688	47.79583	61.49794	44.16910	65.12466
Feb 2024	54.64688	47.64911	61.64465	43.94472	65.34904
Mar 2024	54.64688	47.50542	61.78835	43.72495	65.56881

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	54.64688	47.36455	61.92921	43.50952	65.78424
May 2024	54.64688	47.22636	62.06740	43.29818	65.99559
Jun 2024	54.64688	47.09070	62.20306	43.09070	66.20306
Jul 2024	54.64688	46.95743	62.33633	42.88688	66.40688

1-10 of 36 rows Previous **1** 2 3 4 Next

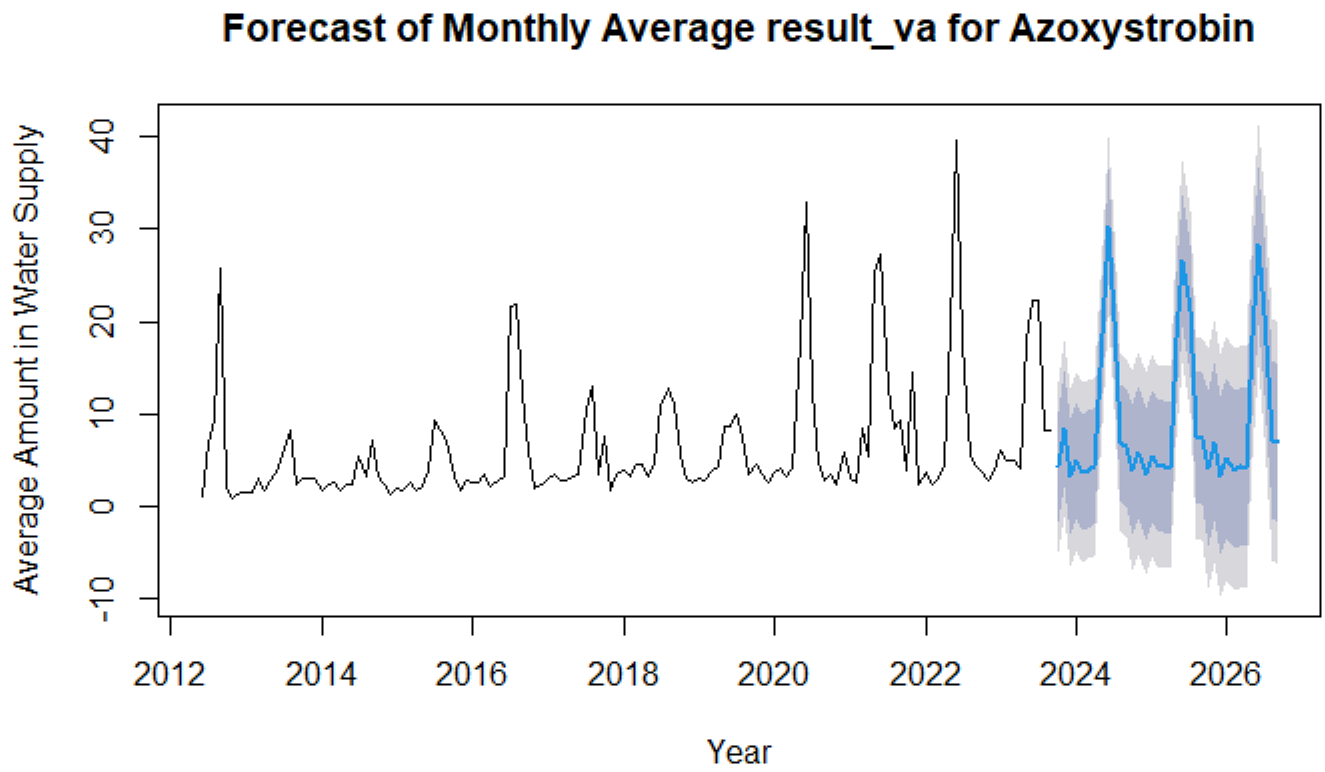
```
NA
```

Azoxystrobin SARIMA forecast. Increase detected

Hide

```
result <- forecast_chemical(
  chemical_name = "Azoxystrobin",
  start_date = "2000-01-01",
  param_data = param_data,
  chemical_data = chemical_data,
  forecast_horizon = 36
)
```

```
The forecast indicates an increase of 2.598618 over the next 36 months.
Series: result_va_ts_monthly
ARIMA(1,0,0)(1,1,0)[12]

Coefficients:
         ar1     sar1
      0.3123  -0.4643
s.e.  0.0886   0.0955

sigma^2 = 21.82:  log likelihood = -367.59
AIC=741.17   AICc=741.37   BIC=749.63

Training set error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 0.4181157 4.424291 2.443488 -8.475068 41.71326 0.8250618 0.01634823
```



Hide

```
result
```

```
$model
Series: result_va_ts_monthly
ARIMA(1,0,0)(1,1,0)[12]

Coefficients:
         ar1     sar1
      0.3123  -0.4643
s.e.  0.0886   0.0955

sigma^2 = 21.82:  log likelihood = -367.59
AIC=741.17   AICc=741.37   BIC=749.63

$forecast
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Oct 2023	4.261333	-1.72511658	10.247783	-4.894152	13.41682
Nov 2023	8.402956	2.13130043	14.674613	-1.188714	17.99463
Dec 2023	3.184986	-3.11380208	9.483773	-6.448179	12.81815
Jan 2024	4.907953	-1.39347543	11.209381	-4.729250	14.54516
Feb 2024	3.680759	-2.62092656	9.982445	-5.956837	13.31836
Mar 2024	3.993377	-2.30833415	10.295088	-5.644258	13.63101

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Apr 2024	4.263749	-2.03796386	10.565463	-5.373889	13.90139
May 2024	18.522842	12.22112863	24.824556	8.885203	28.16048
Jun 2024	30.294207	23.99249304	36.595920	20.656567	39.93185
Jul 2024	20.157130	13.85541644	26.458844	10.519491	29.79477

1-10 of 36 rows Previous **1** 2 3 4 Next

```
NA
```