Counterfactual Adjusted Causal Model

Brain K. Masinde

```r
# clear the working space
rm(list = ls())

library(here)
library(stats) # need this to calculate Mahalanobis Distance
library(parallel) # parallelize
library(dplyr)
library(FNN)
library(cluster)
library(ggplot2)
library(rpart)
library(caret)
```

Counterfactual Data Input

```r
# we need the renaming function for cleaning
melor_2015 <- read.csv(here("data", "clustered_M15_CF_data2.csv"))
```

Counterfactual predictions

Importing trained models

```r
# Import trained BASE models
#   From folder: adjusted SCM/new base models

base_models_list <- list()

# base models file path
base_file_path <- here("adjusted SCM/new base models")
base_wind_model <- readRDS(file.path(base_file_path,
                                     "dec_base_wind_model_tuned.rds"))


base_class_full_model <- readRDS(file.path(base_file_path,
                                     "damage_fit_class_full.rds"))
base_reg_model <- readRDS(file.path(base_file_path,
                                     "base_reg_model.rds"))
```

```r
base_models_list <- list("base_wind_model" = base_wind_model,
                         "base_class_full_model" = base_class_full_model,
                         "base_reg_model" = base_reg_model)
```

```r
# Import trained Truncated models
#   From folder: adjusted SCM/new trunc models

# empty list
trunc_models_list <- list()

trunc_file_path <- here("adjusted SCM/new trunc models")

trunc_wind_model <- readRDS(file.path(trunc_file_path,
                                      "trunc_wind_model_tuned.rds"))


trunc_reg_model <- readRDS(file.path(trunc_file_path,
                                     "trunc_reg_model.rds"))

trunc_models_list <- list("trunc_wind_model" = trunc_wind_model,
                          "trunc_reg_model" = trunc_reg_model)
```

```r
# calling hurdle function
source(here("R", "adj_hurdle_function.R"))


adj_counterfactual_hurdle_preds  <- adj_hurdle_function(df = melor_2015,
                                               scm_models_base = base_models_list,
                                               scm_models_high = trunc_models_list,
                                               threshold = 0.3 # threshold in train/test models is 0.35
                                               )
```

```r
# append the results to the counterfactual dataset
melor_2015  <- melor_2015 %>%
    mutate(damage_preds = adj_counterfactual_hurdle_preds)
```

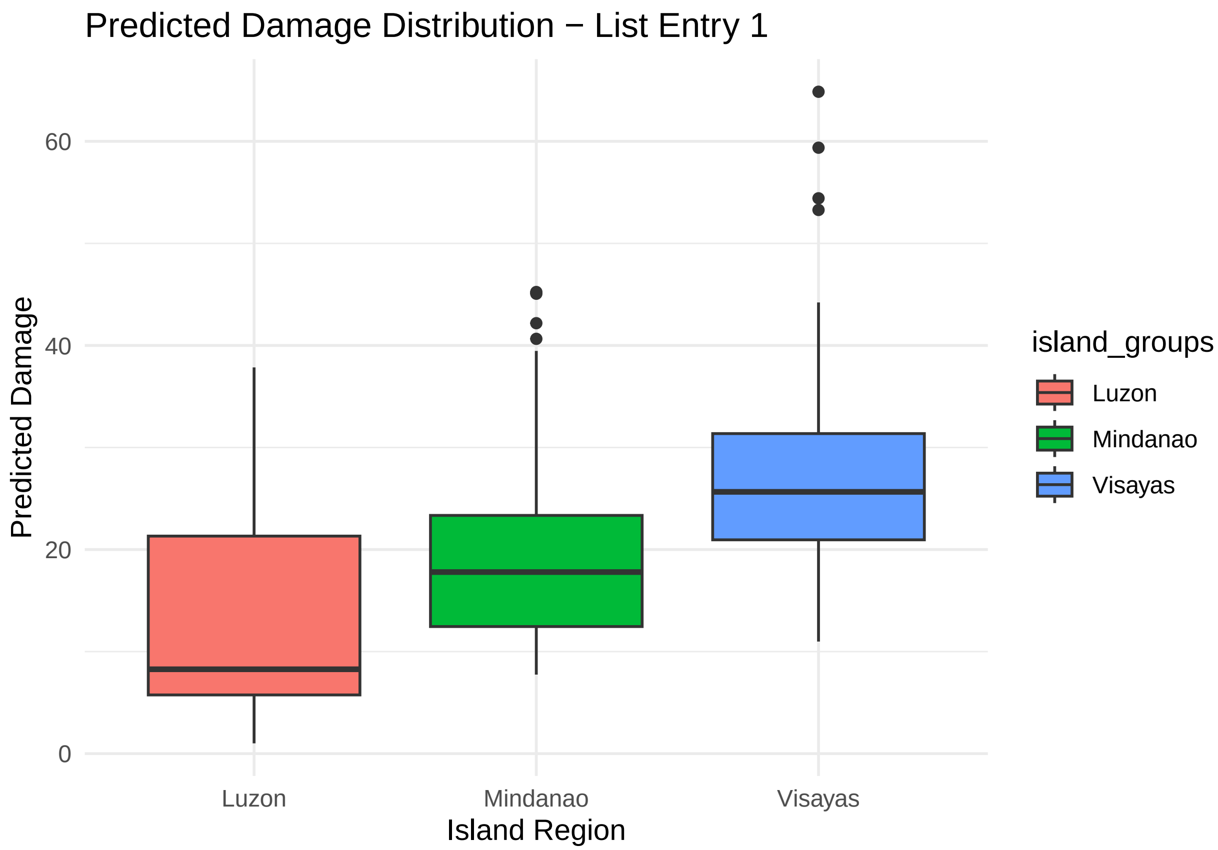
Counteractual Results

```r
# convert the Cluster column to factor
melor_2015$Cluster <- as.factor(melor_2015$Cluster)

# extract cluster_levels
cluster_levels <- levels(melor_2015$Cluster)

# Source function: counterfactual_results
source(here("R", "counterfactual_results.R"))


cf_results <- counterfactual_results(cf_data = melor_2015,
                                     cluster_levels = cluster_levels)
```

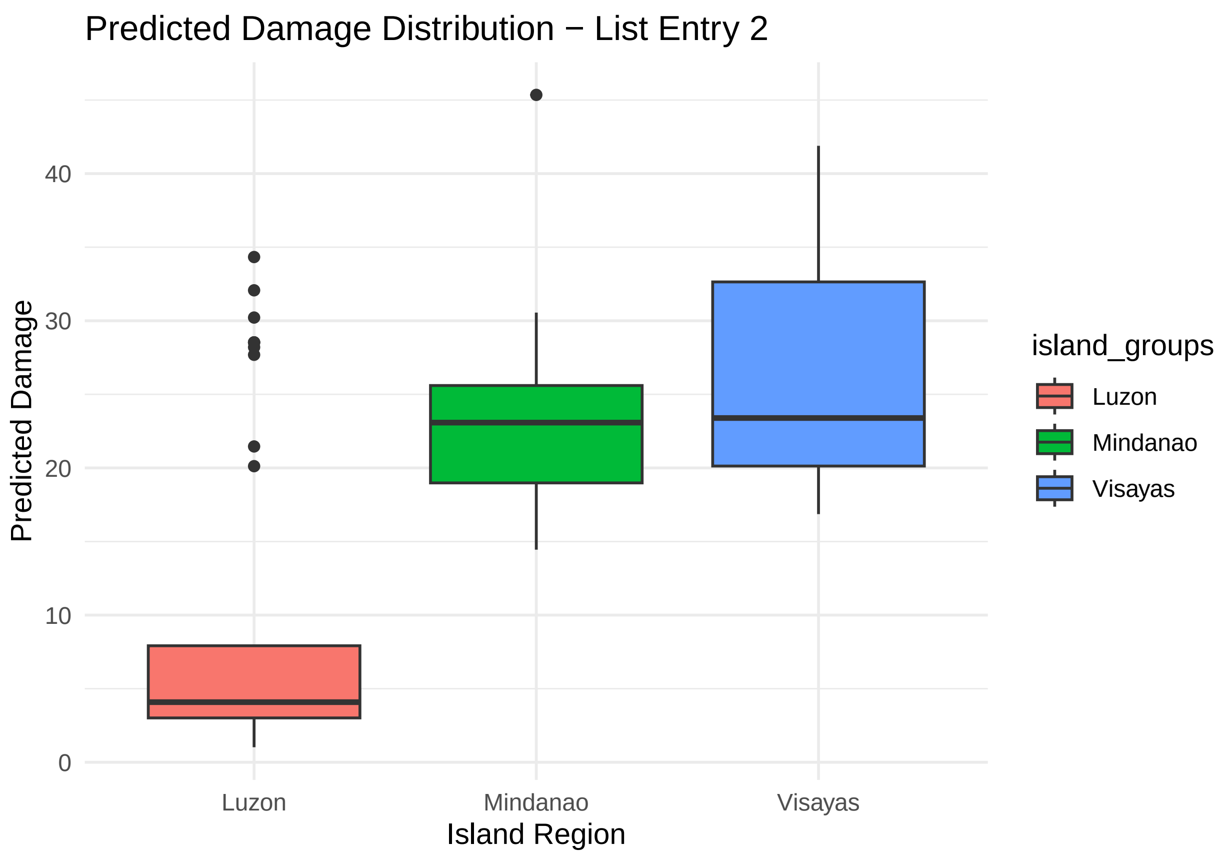
```
## currently evaluating cluster: 1currently evaluating cluster: 2currently evaluating cluster: 3currentl
```

```
# Check the list to confirm plots are stored
print(cf_results$plots)
```

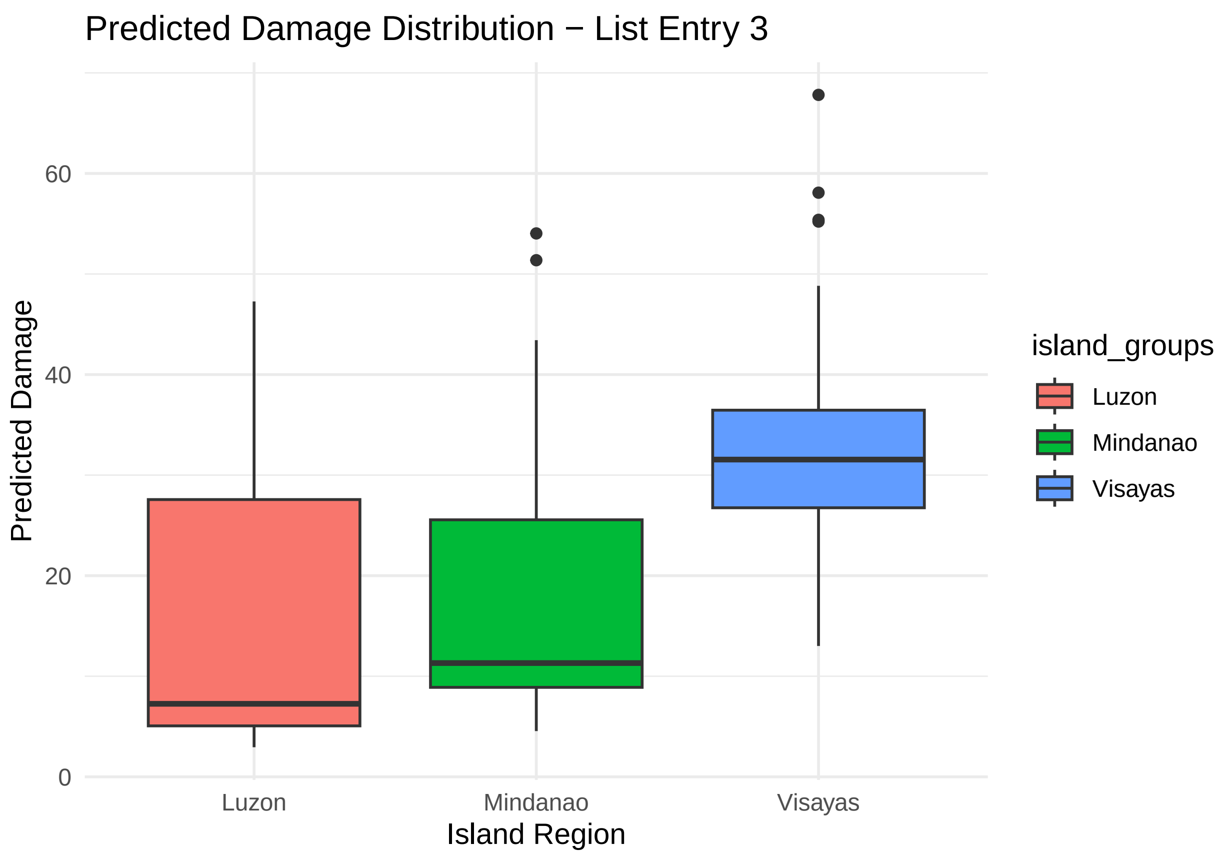
```
## [[1]]
```



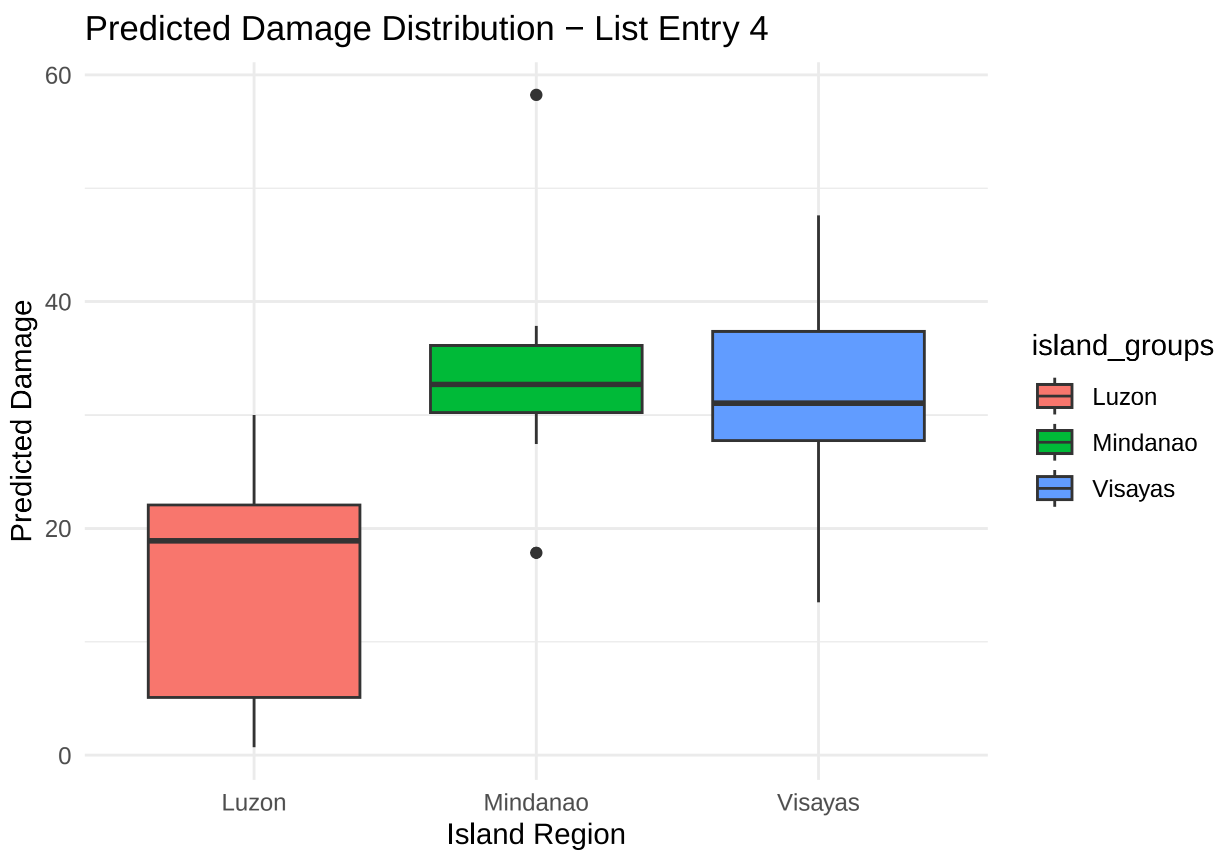

```
##
## [[2]]
```


Predicted Damage Distribution − List Entry 2
Predicted Damage
0
10
20
30
40
Luzon
Mindanao
Visayas
Island Region
island_groups
Luzon
Mindanao
Visayas

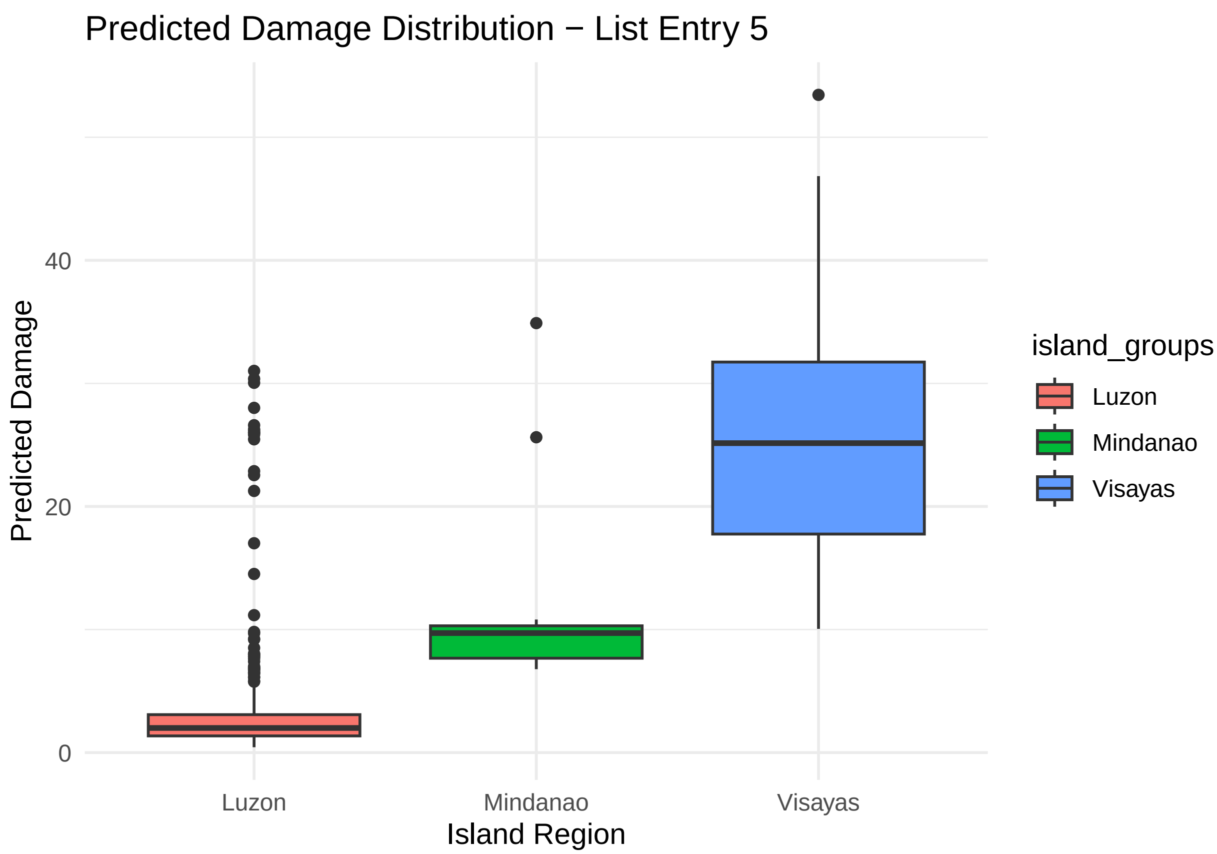
```
##
## [[3]]
```


Predicted Damage Distribution − List Entry 3
Predicted Damage
60
40
20
0
Luzon
Mindanao
Visayas
Island Region
island_groups
Luzon
Mindanao
Visayas

```
##
## [[4]]
```


Predicted Damage Distribution − List Entry 4
60
40
20
0
Predicted Damage
Luzon
Mindanao
Visayas
Island Region
island_groups
Luzon
Mindanao
Visayas

```
##
## [[5]]
```




```r
print(cf_results$median)
```

```
## [[1]]
## # A tibble: 3 x 2
##   island_groups median_damage
##   <chr>                 <dbl>
## 1 Luzon                  8.27
## 2 Mindanao              17.8
## 3 Visayas               25.7
##
## [[2]]
## # A tibble: 3 x 2
##   island_groups median_damage
##   <chr>                 <dbl>
## 1 Luzon                  4.08
## 2 Mindanao              23.1
## 3 Visayas               23.4
##
## [[3]]
## # A tibble: 3 x 2
##   island_groups median_damage
##   <chr>                 <dbl>
## 1 Luzon                  7.26
## 2 Mindanao              11.3
## 3 Visayas               31.5
```

```
##
## [[4]]
## # A tibble: 3 x 2
##   island_groups median_damage
##   <chr>                 <dbl>
## 1 Luzon                  18.9
## 2 Mindanao               32.7
## 3 Visayas                31.0
##
## [[5]]
## # A tibble: 3 x 2
##   island_groups median_damage
##   <chr>                 <dbl>
## 1 Luzon                  2.00
## 2 Mindanao               9.71
## 3 Visayas               25.1
```

```
print(cf_results$averages)
```

```
## [[1]]
## # A tibble: 3 x 2
##   island_groups mean_damage
##   <chr>               <dbl>
## 1 Luzon                13.1
## 2 Mindanao             19.5
## 3 Visayas              28.7
##
## [[2]]
## # A tibble: 3 x 2
##   island_groups mean_damage
##   <chr>               <dbl>
## 1 Luzon                9.46
## 2 Mindanao            24.4
## 3 Visayas             27.4
##
## [[3]]
## # A tibble: 3 x 2
##   island_groups mean_damage
##   <chr>               <dbl>
## 1 Luzon                15.0
## 2 Mindanao             16.7
## 3 Visayas              32.1
##
## [[4]]
## # A tibble: 3 x 2
##   island_groups mean_damage
##   <chr>               <dbl>
## 1 Luzon                15.4
## 2 Mindanao             34.2
## 3 Visayas              32.5
##
## [[5]]
## # A tibble: 3 x 2
##   island_groups mean_damage
```

```
##   <chr>                <dbl>
## 1 Luzon                 3.27
## 2 Mindanao             11.2
## 3 Visayas              25.0
```

Output the counterfactual predictions

Saving the counterfactual predictions for mapping differences between this adjusted causal model and the associational XGBOOST model in QGIS.

```
CF_output <- melor_2015 %>%
  select(Mun_Code, Municipality, Cluster, damage_preds) %>%
  rename(adj_scm_CF_M15 = damage_preds)

write.csv2(CF_output, file = here("adjusted SCM/outputs", "scm_CF_M15.csv"))
```