

DATA589_Project

Justin Chan, CM Tong (Kenny)

2023-04-14

Dataset

URL: <https://www.gbif.org/occurrence/download/0165170-230224095556074> Species: Ascomycota

Introduction:

Provide a brief description of the study system from which the data come and an outline of what questions you intend on exploring with the data, citing any relevant literature. Length: ca. 1 page.

Methods:

Briefly describe the data and what variables are included. Provide a detailed description of the analytical workflow that was applied to the data, citing any relevant literature and statistical packages employed. There should be enough information that anyone can reproduce the workflow if they had access to the data. Length: As long as necessary.

```
# install.packages("rgbif")
library(rgbif)
?rgbif

species <- c("Ascomycota")

# Get number of occurrence records from rgbif
#occ_count(scientificName = species) #16069587 in the dataset

#Filter Ascomycota data to only in BC
asc_count <- occ_count(scientificName = species,
                       hasCoordinate = TRUE,
                       country = "CA",
                       stateProvince = "British Columbia")

asc_bc_data <- occ_data(scientificName = species,
                        hasCoordinate = TRUE,
                        country = "CA",
                        stateProvince = "British Columbia") ## 102 illegal points stored in attr(",r

class(asc_bc_data) #gbif_data
```

```

## [1] "gbif_data"

asc_bc_data <- asc_bc_data$data #gbif_data to data.frame
head(asc_bc_data) #contain "Ascomycota" data only in BC

## # A tibble: 6 x 75
##   key      scien~1 decim~2 decim~3 issues datas~4 publi~5 insta~6 publi~7 proto~8
##   <chr>    <chr>     <dbl>   <dbl>   <chr>   <chr>   <chr>   <chr>   <chr>
## 1 407561~ Lobari~    48.5    -123. cdc,c~ 50c950~ 28eb1a~ 997448~ CA      DWC_AR~
## 2 407558~ Sphaer~    48.5    -123. cdc,c~ 50c950~ 28eb1a~ 997448~ CA      DWC_AR~
## 3 401160~ Physci~    48.4    -123. cdc,c~ 50c950~ 28eb1a~ 997448~ CA      DWC_AR~
## 4 401160~ Everni~    48.6    -123. cdc,c~ 50c950~ 28eb1a~ 997448~ US      DWC_AR~
## 5 401171~ Cladon~    48.8    -124. cdc,c~ 50c950~ 28eb1a~ 997448~ CA      DWC_AR~
## 6 401152~ Nectri~    49.1    -122. cdc,c~ 50c950~ 28eb1a~ 997448~ US      DWC_AR~
## # ... with 65 more variables: lastCrawled <chr>, lastParsed <chr>,
## #   crawlId <int>, hostingOrganizationKey <chr>, basisOfRecord <chr>,
## #   occurrenceStatus <chr>, taxonKey <int>, kingdomKey <int>, phylumKey <int>,
## #   classKey <int>, orderKey <int>, familyKey <int>, genusKey <int>,
## #   speciesKey <int>, acceptedTaxonKey <int>, acceptedScientificName <chr>,
## #   kingdom <chr>, phylum <chr>, order <chr>, family <chr>, genus <chr>,
## #   species <chr>, genericName <chr>, specificEpithet <chr>, ...

```

Data Records

Firstly, screen through some sample records of the dataset to see what information it contains and which attributes would and would not be useful and valuable in our analysis.

```

asc_bc_data <- apply(asc_bc_data, 2, as.character)

# check the working directory and ensure that you have write access to that directory
getwd()

## [1] "C:/Users/cmton/Documents/MDS/Block6/589_SpecialTopic/project/github/DATA589_Project"

#add write acces
# setwd("C:/Users/champ/Desktop/MDS/Class/Block6/DATA589/Project")

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
# save the data as CSV
write.csv(asc_bc_data, "data/asc_bc_data.csv", row.names = FALSE)
asc_bc_data <- read.csv("data/asc_bc_data.csv")

head(asc_bc_data)

##          key scientificName decimalLatitude decimalLongitude
## 1 4075614861   Lobaria pulmonaria (L.) Hoffm.        48.48630       -123.4483
## 2 4075584864 Sphaerophorus tuckermanii Räsänen        48.48630       -123.4483
## 3 4011601239   Physcia tenella (Scop.) DC.        48.41779       -123.3424
## 4 4011605204   Evernia prunastri (L.) Ach.        48.62953       -123.4649
## 5 4011712210   Cladonia squamosa                  48.78214       -123.5684
## 6 4011520235   Nectria cinnabarina (Tode) Fr.       49.12191       -122.1845

```

```

##      issues          datasetKey
## 1 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 2 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 3 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 4 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 5 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
## 6 cdc,cdround 50c9509d-22c7-4a22-a47d-8c48425ef4a7
##          publishingOrgKey           installationKey
## 1 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
## 2 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
## 3 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
## 4 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
## 5 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
## 6 28eb1a3f-1c15-4a95-931a-4af90ecb574d 997448a8-f762-11e1-a439-00145eb45e9a
##      publishingCountry     protocol          lastCrawled
## 1             CA DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
## 2             CA DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
## 3             CA DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
## 4             US DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
## 5             CA DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
## 6             US DWC_ARCHIVE 2023-04-18T14:51:01.646+00:00
##          lastParsed crawlId          hostingOrganizationKey
## 1 2023-04-19T13:19:58.881+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 2 2023-04-19T13:22:36.456+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 3 2023-04-19T13:25:38.604+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 4 2023-04-19T13:26:08.205+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 5 2023-04-19T13:21:04.336+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
## 6 2023-04-19T13:27:02.546+00:00      358 28eb1a3f-1c15-4a95-931a-4af90ecb574d
##      basisOfRecord occurrenceStatus taxonKey kingdomKey phylumKey classKey
## 1 HUMAN_OBSERVATION        PRESENT  5260693       5    95   180
## 2 HUMAN_OBSERVATION        PRESENT  5471125       5    95   180
## 3 HUMAN_OBSERVATION        PRESENT  2608944       5    95   180
## 4 HUMAN_OBSERVATION        PRESENT  2605261       5    95   180
## 5 HUMAN_OBSERVATION        PRESENT  7411883       5    95   180
## 6 HUMAN_OBSERVATION        PRESENT  5251836       5    95   320
##      orderKey familyKey genusKey speciesKey acceptedTaxonKey
## 1     1055     8375  2600355  5260693      5260693
## 2     1048     4810  2608411  5261338      5261338
## 3 10861608     8369  2600367  2608944      2608944
## 4     1048     8305  2605254  2605261      2605261
## 5     1048     8328  2607519  7411883      7411883
## 6     1290     4152  2560952  5251836      5251836
##          acceptedScientificName kingdom      phylum      order
## 1      Lobaria pulmonaria (L.) Hoffm. Fungi Ascomycota Peltigerales
## 2 Sphaerophorus globosus (Huds.) Vain. Fungi Ascomycota Lecanorales
## 3      Physcia tenella (Scop.) DC. Fungi Ascomycota Caliciales
## 4      Evernia prunastri (L.) Ach. Fungi Ascomycota Lecanorales
## 5      Cladonia squamosa Fungi Ascomycota Lecanorales
## 6 Nectria cinnabarina (Tode) Fr. Fungi Ascomycota Hypocreales
##      family      genus      species genericName
## 1  Lobariaceae  Lobaria  Lobaria pulmonaria  Lobaria
## 2 Sphaerophoraceae Sphaerophorus  Sphaerophorus globosus  Sphaerophorus
## 3  Physciaceae  Physcia  Physcia tenella  Physcia
## 4  Parmeliaceae  Evernia  Evernia prunastri  Evernia

```

```

## 5     Cladoniaceae      Cladonia      Cladonia squamosa      Cladonia
## 6     Nectriaceae       Nectria       Nectria cinnabarina      Nectria
##   specificEpithet taxonRank taxonomicStatus iucnRedListCategory
## 1     pulmonaria      SPECIES      ACCEPTED                  NE
## 2     tuckermanii     SPECIES      SYNONYM                  NE
## 3     tenella         SPECIES      ACCEPTED                  NE
## 4     prunastri       SPECIES      ACCEPTED                  NE
## 5     squamosa        SPECIES      ACCEPTED                  NE
## 6     cinnabarina     SPECIES      ACCEPTED                  NE
##   dateIdentified coordinateUncertaintyInMeters      continent
## 1 2023-01-01T22:23:29                      200 NORTH_AMERICA
## 2 2023-01-01T22:47:09                      200 NORTH_AMERICA
## 3 2023-01-01T23:24:42                      2 NORTH_AMERICA
## 4 2023-01-02T01:00:32                      355 NORTH_AMERICA
## 5 2023-01-02T02:31:54                      36 NORTH_AMERICA
## 6 2023-01-02T03:56:46                      305 NORTH_AMERICA
##   stateProvince year month day      eventDate
## 1 British Columbia 2023    1    1 2023-01-01T11:26:00
## 2 British Columbia 2023    1    1 2023-01-01T12:01:00
## 3 British Columbia 2023    1    1 2023-01-01T14:15:00
## 4 British Columbia 2023    1    1 2023-01-01T14:48:15
## 5 British Columbia 2023    1    1 2023-01-01T13:08:28
## 6 British Columbia 2023    1    1 2023-01-01T15:55:00
##   modified      lastInterpreted
## 1 2023-03-24T21:03:25.000+00:00 2023-04-19T13:19:58.881+00:00
## 2 2023-03-24T21:03:25.000+00:00 2023-04-19T13:22:36.456+00:00
## 3 2023-01-01T23:55:15.000+00:00 2023-04-19T13:25:38.604+00:00
## 4 2023-01-02T16:32:20.000+00:00 2023-04-19T13:26:08.205+00:00
## 5 2023-01-02T02:33:24.000+00:00 2023-04-19T13:21:04.336+00:00
## 6 2023-01-02T06:01:33.000+00:00 2023-04-19T13:27:02.546+00:00
##   references
## 1 https://www.inaturalist.org/observations/145614607
## 2 https://www.inaturalist.org/observations/145614670
## 3 https://www.inaturalist.org/observations/145619638
## 4 https://www.inaturalist.org/observations/145626873
## 5 https://www.inaturalist.org/observations/145627018
## 6 https://www.inaturalist.org/observations/145637879
##   license isInCluster
## 1 http://creativecommons.org/licenses/by/4.0/legalcode FALSE
## 2 http://creativecommons.org/licenses/by/4.0/legalcode FALSE
## 3 http://creativecommons.org/publicdomain/zero/1.0/legalcode FALSE
## 4 http://creativecommons.org/licenses/by-nc/4.0/legalcode FALSE
## 5 http://creativecommons.org/licenses/by-nc/4.0/legalcode FALSE
## 6 http://creativecommons.org/licenses/by-nc/4.0/legalcode FALSE
##   datasetName      recordedBy
## 1 iNaturalist research-grade observations      Noah How
## 2 iNaturalist research-grade observations      Noah How
## 3 iNaturalist research-grade observations      Brian Starzomski
## 4 iNaturalist research-grade observations      Dan Kells
## 5 iNaturalist research-grade observations      klbarry
## 6 iNaturalist research-grade observations      Darcy Kehler
##   identifiedBy geodeticDatum      class countryCode country
## 1 Noah How          WGS84 Lecanoromycetes      CA Canada
## 2 Myung Jin (John) Kang      WGS84 Lecanoromycetes      CA Canada

```

```

## 3      Brian Starzomski      WGS84 Lecanoromycetes      CA  Canada
## 4      Dan Kells            WGS84 Lecanoromycetes      CA  Canada
## 5      Stewart Wechsler     WGS84 Lecanoromycetes      CA  Canada
## 6      Darcy Kehler        WGS84 Sordariomycetes      CA  Canada
## rightsHolder identifier http://www.inaturalist.org.nick      verbatimEventDate
## 1      Noah How             145614607                 nhow          2023/01/01 11:26 AM
## 2      Noah How             145614670                 nhow          2023/01/01 12:01 PM
## 3 Brian Starzomski        145619638                 bstarzomski    2023/01/01 2:15 PM
## 4      Dan Kells            145626873                 dougiefir      2023-01-01 14:48:15
## 5      klbarry              145627018                 klbarry        2023-01-01 13:08:28-08:00
## 6      Darcy Kehler        145637879                 lophopanopeus  2023/01/01 3:55 PM
## collectionCode gbifID      verbatimLocality
## 1      Observations         4075614861               Capital, BC, Canada
## 2      Observations         4075584864               Capital, BC, Canada
## 3      Observations         4011601239               Rockland, Victoria, BC, Canada
## 4      Observations         4011605204               Capital, BC V8L, Canada
## 5      Observations         4011712210               North Cowichan, BC, CA
## 6      Observations         4011520235               Fraser Valley, BC, Canada
## occurrenceID taxonID catalogNumber
## 1 https://www.inaturalist.org/observations/145614607  48711   145614607
## 2 https://www.inaturalist.org/observations/145614670  123231  145614670
## 3 https://www.inaturalist.org/observations/145619638  228059  145619638
## 4 https://www.inaturalist.org/observations/145626873  123175  145626873
## 5 https://www.inaturalist.org/observations/145627018  117963  145627018
## 6 https://www.inaturalist.org/observations/145637879  118012  145637879
## institutionCode eventTime http://www.inaturalist.org.captive identificationID
## 1      iNaturalist          11:26:00-08:00           wild       324236558
## 2      iNaturalist          12:01:00-08:00           wild       324241852
## 3      iNaturalist          14:15:00-08:00           wild       324250052
## 4      iNaturalist          14:48:15-08:00           wild       324268496
## 5      iNaturalist          13:08:28-08:00           wild       324287531
## 6      iNaturalist          15:55:00-08:00           wild       324301946
##
## 1
## 2
## 3
## 4
## 5 [C. furcata] (https://www.inaturalist.org/observations/107152719) doesn't have such numerous, large
## 6
## occurrenceRemarks informationWithheld      name
## 1      <NA>                <NA>    Lobaria pulmonaria (L.) Hoffm.
## 2      <NA>                <NA>    Sphaerophorus tuckermanii Räsänen
## 3      <NA>                <NA>    Physcia tenella (Scop.) DC.
## 4      <NA>                <NA>    Evernia prunastri (L.) Ach.
## 5      <NA>                <NA>    Cladonia squamosa
## 6      <NA>                <NA>    Nectria cinnabarina (Tode) Fr.

```

Broadly speaking, from the above records screening, the dataset information can be categorised into the following types according to the dataset attribuets :

- Species Taxonomy Information : There are many information about the species/family/genus/kingdom/class etc which correspond to the structural classification system of the fungi family.
- Collected Record Information : Contains how observation is made (by human?), date, time and location (coordinate, continents, etc) it is collected, the sample collector/institution

Data Cleaning

As the dataset contains too many information, including many detailed timestamps, key information and dataset/records identifiers which should not be valuable in our analysis and complicate our subsequent analysis, we have performed preliminary cleaning procedure and perform attributes selection, in order to extract those potentially useful attributes and make our analysis more focused. The cleaned list if shown below.

```
# View(data.frame(names(asc_bc_data)))
cleaned_asc_bc <- asc_bc_data[, c("decimalLongitude", "decimalLatitude", "order", "family", "genus", "year", "month", "day", "eventDate", "occurrenceStatus", "class", "verbatimEventDate", "collectionCode", "gbifID", "verbatimLocality")]
names(cleaned_asc_bc)

## [1] "decimalLongitude"           "decimalLatitude"
## [3] "order"                     "family"
## [5] "genus"                     "species"
## [7] "genericName"               "specificEpithet"
## [9] "coordinateUncertaintyInMeters" "stateProvince"
## [11] "year"                      "month"
## [13] "day"                       "eventDate"
## [15] "occurrenceStatus"          "class"
## [17] "countryCode"                "country"
## [19] "verbatimLocality"           "taxonID"
## [21] "catalogNumber"              "institutionCode"
## [23] "eventTime"                  "verbatimEventDate"
## [25] "collectionCode"             "gbifID"
## [27] "verbatimLocality.1"
```

Initial Coordinates Plotting

```
# convert cleaned_asc_bc to SpatialPointsDataFrame-class
#reference: https://www.nceas.ucsb.edu/sites/default/files/2020-04/OverviewCoordinateReferenceSystems.pdf
#           https://scisus.org/2014/05/

names(cleaned_asc_bc)

## [1] "decimalLongitude"           "decimalLatitude"
## [3] "order"                     "family"
## [5] "genus"                     "species"
## [7] "genericName"               "specificEpithet"
## [9] "coordinateUncertaintyInMeters" "stateProvince"
## [11] "year"                      "month"
## [13] "day"                       "eventDate"
## [15] "occurrenceStatus"          "class"
## [17] "countryCode"                "country"
## [19] "verbatimLocality"           "taxonID"
## [21] "catalogNumber"              "institutionCode"
## [23] "eventTime"                  "verbatimEventDate"
```

```

## [25] "collectionCode"                  "gbifID"
## [27] "verbatimLocality.1"

sort(unique(cleaned_asc_bc$individualCount)) # notice if some points correspond to zero abundance

## NULL

sort(unique(cleaned_asc_bc$occurrenceStatus)) # check for different indications of "absent", which cou

## [1] "PRESENT"

absence_rows <- which(cleaned_asc_bc$individualCount == 0 | cleaned_asc_bc$occurrenceStatus %in% c("abs
length(absence_rows)

## [1] 0

if (length(absence_rows) > 0) {
  cleaned_asc_bc <- cleaned_asc_bc[-absence_rows, ]
}

library(sp)
sp_asc_bc <- SpatialPointsDataFrame(coords = cleaned_asc_bc[, c("decimalLongitude", "decimalLatitude")]
                                      data = cleaned_asc_bc)

proj4string(sp_asc_bc) <- CRS("+proj=longlat +datum=WGS84")

# install.packages("rgdal")
library(rgdal)

#project basemap to match spu shapefile
tran_sp_asc_bc <- spTransform(sp_asc_bc, CRS("+proj=aea +lat_0=45 +lon_0=-126 +lat_1=50 +lat_2=58.5 +x_0=0 +y_0=0 +datum=NAD83 +units=m +no_defs"))
summary(tran_sp_asc_bc)

## Object of class SpatialPointsDataFrame
## Coordinates:
##               min     max
## decimalLongitude 606101.5 1798367
## decimalLatitude  370250.1 1007523
## Is projected: TRUE
## proj4string :
## [+proj=aea +lat_0=45 +lon_0=-126 +lat_1=50 +lat_2=58.5 +x_0=1000000
## +y_0=0 +datum=NAD83 +units=m +no_defs]
## Number of points: 500
## Data attributes:
##   decimalLongitude decimalLatitude   order      family
##   Min.    :-131.9   Min.    :48.32  Length:500      Length:500
##   1st Qu.:-123.5   1st Qu.:48.43  Class :character  Class :character
##   Median :-123.5   Median :48.46  Mode   :character  Mode   :character
##   Mean   :-123.4   Mean   :48.72
##   3rd Qu.:-123.4   3rd Qu.:48.80

```

```

##  Max.    :-114.9   Max.    :54.04
##
##      genus           species       genericName       specificEpithet
##  Length:500        Length:500       Length:500        Length:500
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##      coordinateUncertaintyInMeters stateProvince          year         month
##  Min.    : 1.0            Length:500       Min.    :2023  Min.    :1
##  1st Qu.:  5.0            Class :character  1st Qu.:2023  1st Qu.:1
##  Median :  8.0            Mode  :character  Median :2023  Median :1
##  Mean   : 554.4           Mode  :character  Mean   :2023  Mean   :1
##  3rd Qu.: 26.0            Mode  :character  3rd Qu.:2023  3rd Qu.:1
##  Max.   :111003.0          Mode  :character  Max.   :2023  Max.   :1
##  NA's   :98
##
##      day       eventDate       occurrenceStatus       class
##  Min.    : 1.00  Length:500       Length:500        Length:500
##  1st Qu.: 9.75  Class :character  Class :character  Class :character
##  Median :23.00  Mode  :character  Mode  :character  Mode  :character
##  Mean   :18.74
##  3rd Qu.:27.00
##  Max.   :29.00
##
##      countryCode       country       verbatimLocality       taxonID
##  Length:500        Length:500       Length:500        Min.    : 48711
##  Class :character  Class :character  Class :character  1st Qu.: 54734
##  Mode  :character  Mode  :character  Mode  :character  Median : 117948
##                                         Mean   : 166623
##                                         3rd Qu.: 194116
##                                         Max.   :1398622
##
##      catalogNumber     institutionCode       eventTime       verbatimEventDate
##  Min.    :145614607  Length:500       Length:500        Length:500
##  1st Qu.:146308188  Class :character  Class :character  Class :character
##  Median :147241211  Mode  :character  Mode  :character  Mode  :character
##  Mean   :146966899
##  3rd Qu.:147567074
##  Max.   :147668836
##
##      collectionCode      gbifID       verbatimLocality.1
##  Length:500        Min.    :4.011e+09  Length:500
##  Class :character  1st Qu.:4.018e+09  Class :character
##  Mode  :character  Median :4.029e+09  Mode  :character
##                                         Mean   :4.026e+09
##                                         3rd Qu.:4.029e+09
##                                         Max.   :4.080e+09
##

```

#update to spTransformed Longitude & Latitude

```

cleaned_asc_bc$decimalLongitude <- tran_sp_asc_bc@coords[, 1]
cleaned_asc_bc$decimalLatitude <- tran_sp_asc_bc@coords[, 2]

```

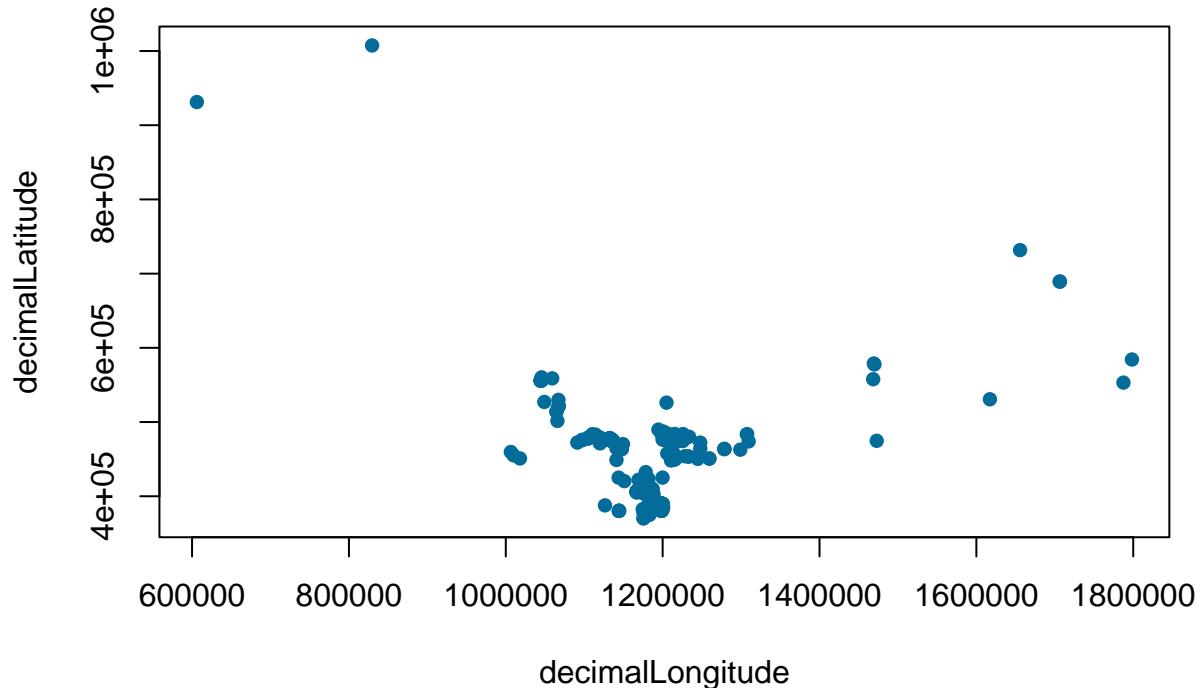
```

# save the data as CSV
write.csv(cleaned_asc_bc, "data/clean_sp_asc_bc", row.names = FALSE)
cleaned_asc_bc <- read.csv("data/clean_sp_asc_bc")

#plot(decimalLatitude ~ decimalLongitude,
#      pch = 16,
#      col = "#046C9A",
#      data = cleaned_asc_bc)

library(spatstat)
# install.packages("maptools")
library(maptools)
#Visualise the data
plot(decimalLatitude ~ decimalLongitude,
      pch = 16,
      col = "#046C9A",
      data = cleaned_asc_bc)

```



From the initial coordinate plot, we have identified the following very preliminary observation :

- Points clustering is highly likely as a large group of data points is spotted at the decimal Longitude between 1000K and 1300K and Latitude between 4e+05 and 5e+05.
- Therefore is only a single major and significant large clustering only. Although there are some other scattered points observed near the Longitude range 1500K-1800K, they are simply incomparable to the large clustering.

BC Windows Data Walkthrough

Let's see what the BC_Covariates.Rda file provide us which may help us to identify and choose appropriate fields to be used in the covariates analysis in the later part of this analysis. It should contain the information on the BC province :

- Windows
- Elevation
- Forest
- Dist Water

```
load("data/BC_Covariates.Rda")

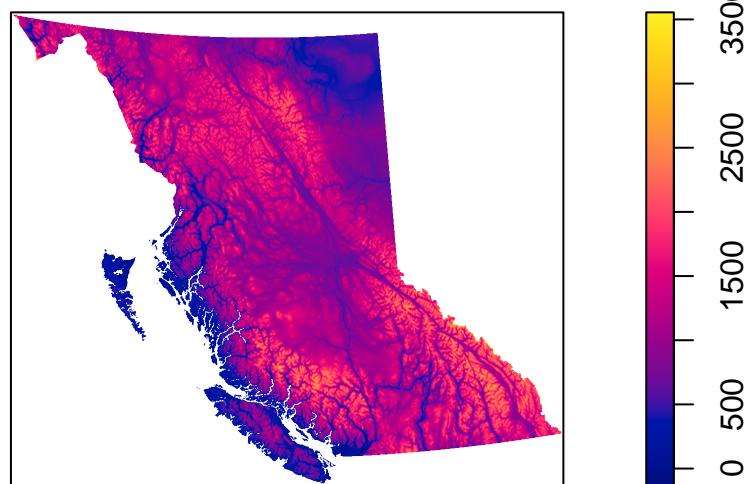
#observe windows
BC<-DATA
BC_win <- DATA$Window
BC_win <- as.owin(DATA$Window)
plot(BC_win,
      pch = 16,
      cols = "red",
      main = "BC windows data")
```

BC windows data



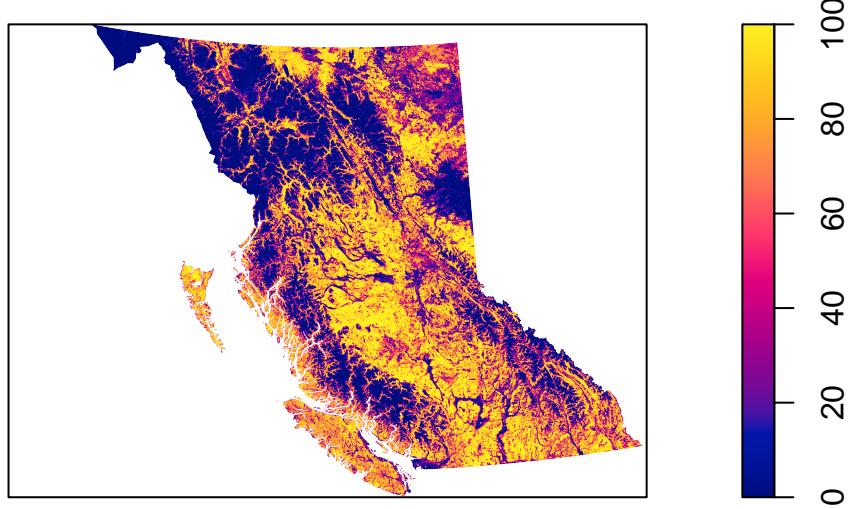
```
#Elevation
plot(BC$Elevation, main = "Elevation")
```

Elevation



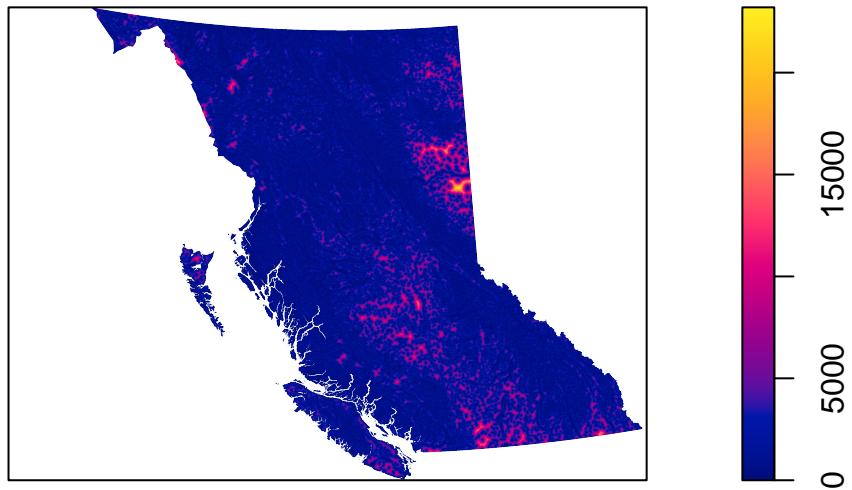
```
#Forest  
plot(BC$Forest, main = "Forest")
```

Forest



```
#Dist_Water  
plot(BC$Dist_Water, main = "Dist_Water")
```

Dist_Water



Spatial Dataset Conversion and Preparation

As the current format of the fungi data is not in a PPP class object, to facilitate subsequent analysis and plots, we would firstly convert it to the PPP, which would make subsequent plotting and analysis library much more accessible.

```
# names(cleaned_asc_bc$decimalLongitude)
# class(cleaned_asc_bc$decimalLongitude)

# convert into ppp object
asc_data_ppp <- ppp(x = cleaned_asc_bc$decimalLongitude,
                      y = cleaned_asc_bc$decimalLatitude,
                      window = BC_win)

length(asc_data_ppp)

## [1] 5

anyDuplicated(asc_data_ppp)

## [1] TRUE
```

```

asc_data_ppp

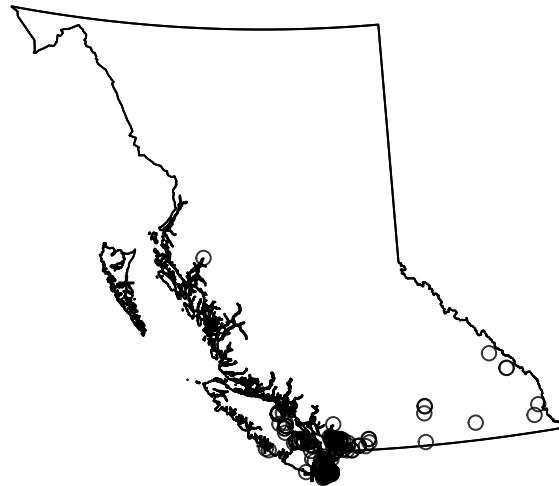
## Planar point pattern: 400 points
## window: polygonal boundary
## enclosing rectangle: [273874.9, 1870573.4] x [369042.8, 1735666.4] units
## *** 100 illegal points stored in attr(),"rejects" ***

# length(cleaned_asc_bc$decimalLongitude)
# length(cleaned_asc_bc$decimalLatitude)
# length(cleaned_asc_bc$species)
# length(marks(asc_data_ppp))
#
# marks(asc_data_ppp) <- factor(cleaned_asc_bc$species)
# length(cleaned_asc_bc$species)

plot(asc_data_ppp)

```

asc_data_ppp



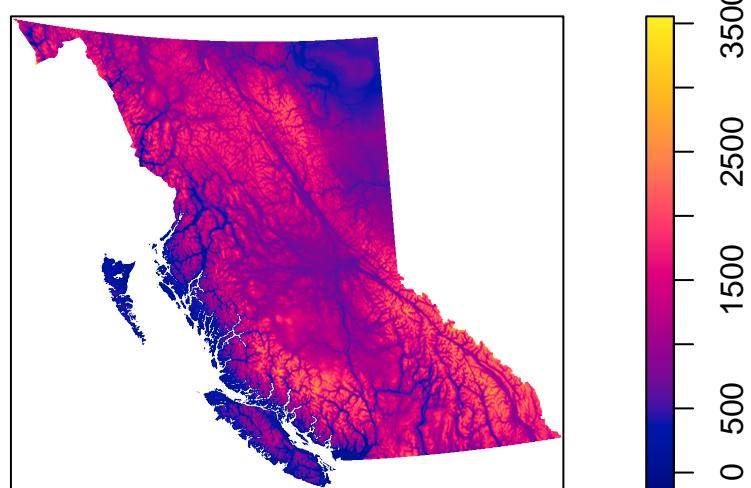
```

#
# marks(asc_data_ppp) <- BC[[1]]$Region

plot(BC$Elevation, main = "elevation data")

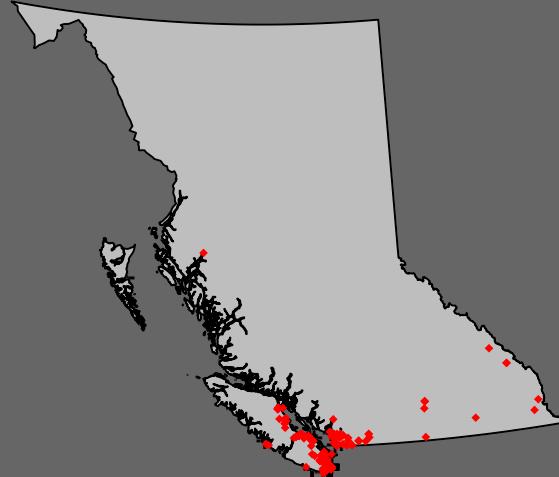
```

elevation data



```
plot(asc_data_ppp,
  which.marks = "species", # Which mark to use
  col = "grey", #The colour of the window
  cols = 'red', #The colours of the points
  cex = 0.6,
  pch = 18, # The plotting symbol
  main = "Ascomycota in BC", # The title
  par(bg="grey40", cex.main = 2),
  cex = 0.6,
  legend = T) # Turn of the legend depending on needs`
```

Ascomycota in BC

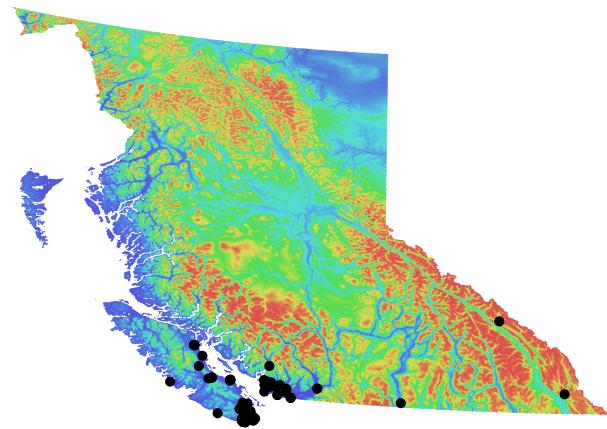


```
col_pal <- rev(rainbow(100,end = 4/6))

fig <- persp(BC$Elevation, # source data
             theta = 10,
             phi = 40, # rotation
             expand = 0, # z-axis expansion
             border = NA, #remove grid borders
             apron = FALSE, #apron around edge
             shade = 0.3, # shading
             box = FALSE, # axes on/off
             main = "BC Parks Elevation", # title
             visible = TRUE, #Supporting calculations
             colmap = col_pal) # colour pallet

perspPoints(asc_data_ppp, Z = BC$Elevation, M = fig, pch = 16, cex = 0.7, col="black")
```

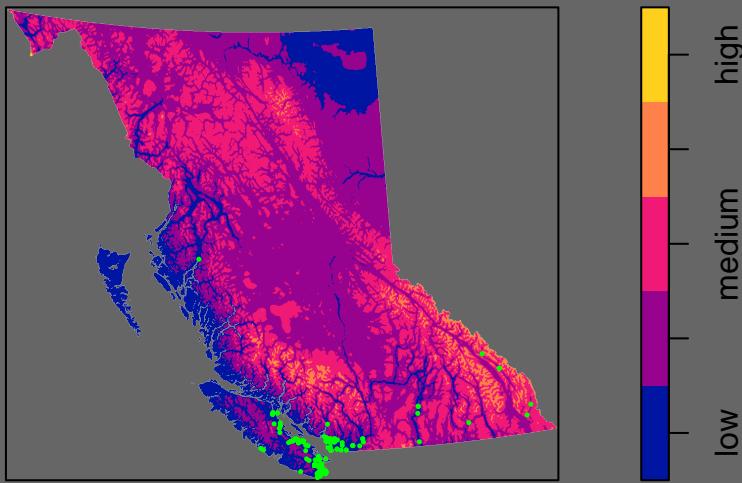
BC Parks Elevation



```
plot(cut(BC$Elevation,5,
labels = c("low","low-medium","medium","medium-high","high")),
par(bg="grey40", cex.main = 2),
main = "Elevation classes")

points(asc_data_ppp, pch = 16, cex = 0.35, col="green")
```

Elevation classes



```
cut <- cut(BC$Elevation, 5,
labels = c("low", "low-medium", "medium", "medium-high", "high"))
table(cut[asc_data_ppp]) #most in low elevation

##
##          low   low-medium       medium medium-high        high
##          315           8            1            0            0

# nn_dist <- nndist(asc_data_ppp)
# marks(asc_data_ppp) <- nn_dist
# plot(asc_data_ppp, main = "Ascomycota Distance", which.marks = "Dist", pch = 16)
BC$Elevation[asc_data_ppp]

## [1] 91.271680 91.271680 23.397900 62.692406 51.428791 261.980282
## [7] 261.980282 261.980282 74.498003 74.498003 510.926444 32.118248
## [13] 62.814557 81.548294 81.548294 76.084493 706.409676 65.194619
## [19] 8.000000 -35.179669 -35.179669 -35.179669 -35.179669 76.084493
## [25] 76.084493 76.084493 32.118248 32.118248 195.284587 61.665185
## [31] 206.588250 66.000000 65.619479 288.070603 288.070603 288.070603
## [37] 288.070603 288.070603 355.967892 355.967892 355.967892 355.967892
## [43] 355.967892 355.967892 355.967892 288.070603 288.070603 288.070603
## [49] 288.070603 288.070603 288.070603 288.070603 288.070603 288.070603
## [55] 288.070603 288.070603 430.115273 430.115273 288.070603 288.070603
## [61] 355.967892 430.115273 430.115273 288.070603 288.070603 288.070603
```

```

## [67] 288.070603 288.070603 288.070603 225.481611 225.481611 198.143926
## [73] 81.126810 -35.053897 -35.053897 39.369451 39.369451 294.384024
## [79] -35.053897 -35.053897 -35.053897 -35.053897 -35.053897 -35.053897
## [85] 71.801786 -35.053897 187.957696 187.957696 171.936613 167.107333
## [91] 40.396331 40.396331 40.396331 40.396331 40.396331 40.396331
## [97] 1038.434881 777.098704 1670.964740 69.289789 94.422608 69.289789
## [103] 39.092076 82.583477 14.698512 55.883115 36.396924 49.126893
## [109] 45.652949 50.701483 50.701483 50.784868 49.126893 44.246494
## [115] 25.428184 25.428184 25.428184 339.591607 21.000000 21.000000
## [121] 21.000000 25.141153 1119.185852 108.960033 45.652949 32.118248
## [127] 23.899560 92.143706 42.880855 54.069165 50.784868 130.735146
## [133] 147.005958 50.784868 1318.635190 131.236704 114.939521 131.236704
## [139] 160.090881 44.246494 1130.775108 46.432839 -3.734969 20.714393
## [145] 44.000000 122.489656 53.985568 177.952547 46.432839 117.749693
## [151] 50.784868 46.432839 72.717695 433.194915 33.995941 72.717695
## [157] 72.717695 48.896700 25.547370 55.243721 55.243721 2.895689
## [163] 32.118248 197.677291 46.440137 46.440137 71.801786 198.143926
## [169] 23.397900 32.118248 80.825577 102.071038 433.786763 224.330101
## [175] 224.330101 189.297238 214.439717 189.990055 203.326184 203.326184
## [181] 203.326184 203.326184 203.326184 206.588250 22.034083 18.242945
## [187] -35.053897 45.657363 510.926444 510.926444 197.677291 197.677291
## [193] -57.458214 60.833698 50.784868 206.588250 -35.053897 -35.053897
## [199] 32.118248 181.306638 27.681348 27.681348 176.437454 176.437454
## [205] 37.000000 32.118248 46.432839 26.520313 62.814557 4.412054
## [211] 315.807853 315.807853 40.609456 135.806165 50.784868 66.741829
## [217] 31.212766 53.507950 267.141364 233.374089 129.407323 197.677291
## [223] 1119.185852 67.901429 145.876827 235.703330 334.238754 1108.387686
## [229] 206.488760 206.488760 206.488760 193.400186 115.019727 32.118248
## [235] 27.407411 32.118248 32.118248 32.118248 32.118248 44.246494
## [241] 44.246494 26.520313 26.520313 26.520313 26.520313 26.520313
## [247] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [253] 26.520313 26.520313 26.520313 26.520313 26.520313 404.858785
## [259] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [265] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [271] 26.520313 26.520313 32.118248 32.118248 43.852612 43.852612
## [277] 43.852612 26.520313 26.520313 26.520313 26.520313 26.520313
## [283] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [289] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [295] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [301] 26.520313 26.520313 26.520313 26.520313 26.520313 26.520313
## [307] 26.520313 26.520313 26.520313 26.520313 62.814557 62.814557
## [313] 62.814557 62.814557 62.814557 64.156157 23.804111 45.652949
## [319] 26.520313 36.396924 -6.244630 123.164317 205.363512 205.363512

```

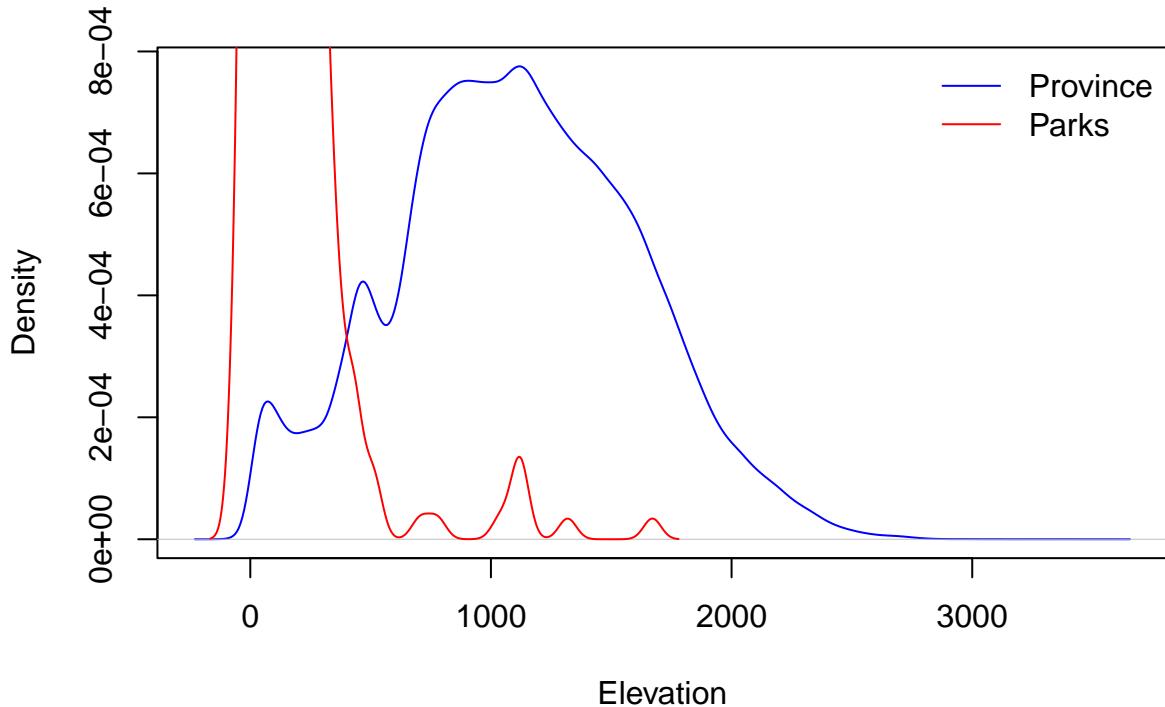
```

library("kdensity")
asc_density <- density(BC$Elevation[asc_data_ppp])
province_density <- density(BC$Elevation$v, na.rm=T)

plot(province_density, main = "KDE of elevation values within province and park locations", xlab = "Ele-
lines(asc_density, col = "red")
legend("topright", legend = c("Province", "Parks"), lty = 1, col = c("blue", "red"), bty = "n")

```

KDE of elevation values within province and park locations



```
#  
# plot(province_density, main = "KDE of elevation values within province and park locations", xlab = "Elevation", ylab = "Density")  
# lines(park_density, col = "red")  
# legend("topright", legend = c("Province", "Parks"), lty = 1, col = c("blue", "red"), bty = "n")
```

First Moment Descriptive Statistics

```
intensity(asc_data_ppp)
```

```
## [1] 4.218242e-10
```

Homogeneity Studies, Quadrat Test and Hotspot Analysis

Note that the intensity is a very small number. This is consistent with our plots above. There are not many points in the whole BC windows. Overwhelming majority of the points are sparsely located in the BC. Let's verify this with the Quadrat Count Plot.

```
Q <- quadratcount(asc_data_ppp,  
                    nx = 4,  
                    ny = 4)
```

```
#Plot the output
```

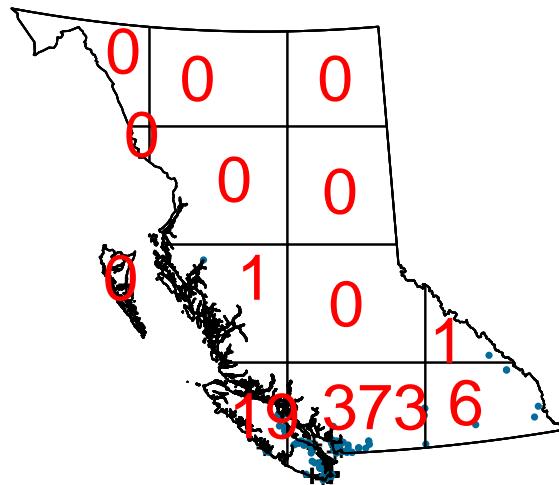
```

plot(asc_data_ppp,
  pch = 16,
  cex = 0.5,
  cols = "#046C9A",
  main = " Ascomycota Locations - Quadrat Count")

plot(Q, cex = 2, col = "red", add = T)

```

Ascomycota Locations – Quadrat Count



```

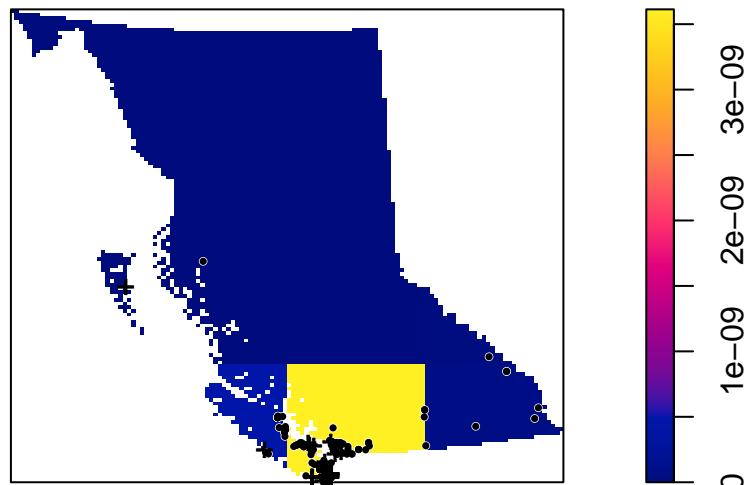
plot(intensity(Q, image = T),
  main = "Ascomycota intensity")

plot(asc_data_ppp,
  pch = 16,
  cex = 0.6,
  cols = "white",
  add = T)

plot(asc_data_ppp,
  pch = 16,
  cex = 0.5,
  cols = "black",
  add = T)

```

Ascomycota intensity



```
#Quadrat test of homogeneity
quadrat.test(Q)
```

```
##
## Chi-squared test of CSR using quadrat counts
##
## data:
## X2 = 2818.1, df = 12, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 13 tiles (irregular windows)
```

```
R <- bw.ppl(asc_data_ppp)
LR <- scanLRTS(asc_data_ppp,r=R)
plot(LR, "Likelihood Ratio Test")
plot(asc_data_ppp>window, add=T)
```

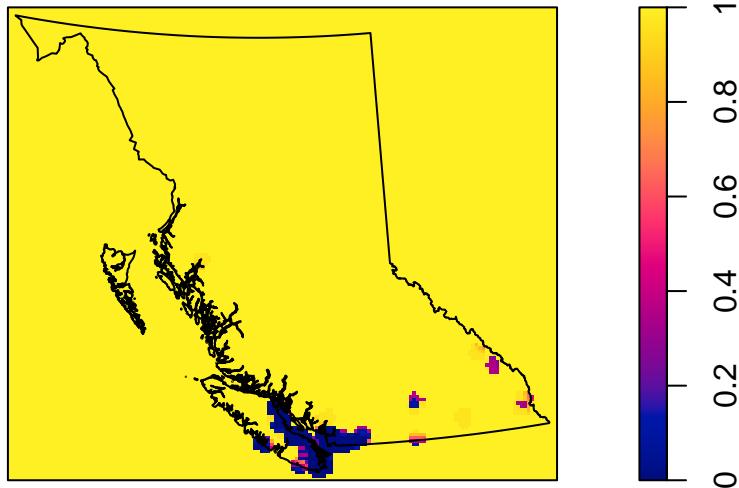
LR



```
pvals <- eval.im(pchisq(LR,
                         df = 1,
                         lower.tail = FALSE))

#Plot the output
plot(pvals, main = "Local p-values")
plot(asc_data_ppp>window, add=T)
```

Local p-values



Once again, this has strongly verified the spatial inhomogeneity nature of the fungi distribution :

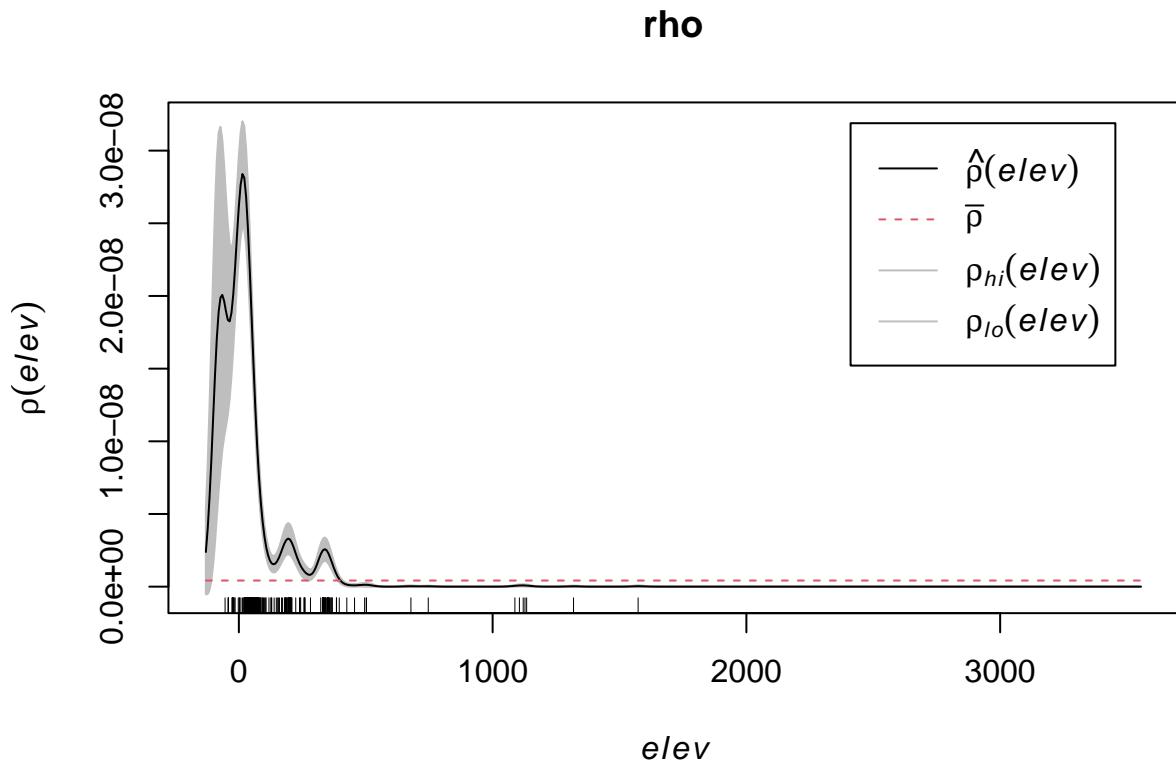
- About 90% of the regions have zero point.
- Less than 5% of the regions contain more than 85% of all the points
- Quadrat Test also indicates a significant deviation from homogeneity.
- Hot spot analysis shows only a single prominent hot spot

Covariate Study

Covariate Variable : Elevation

```
elev <- BC$Elevation
b <- quantile(elev,probs=(0:4)/4,type=2)
Zcut <- cut(elev,breaks=b)
V <- tess(image=Zcut)
quadratcount(asc_data_ppp,tess=V)
```

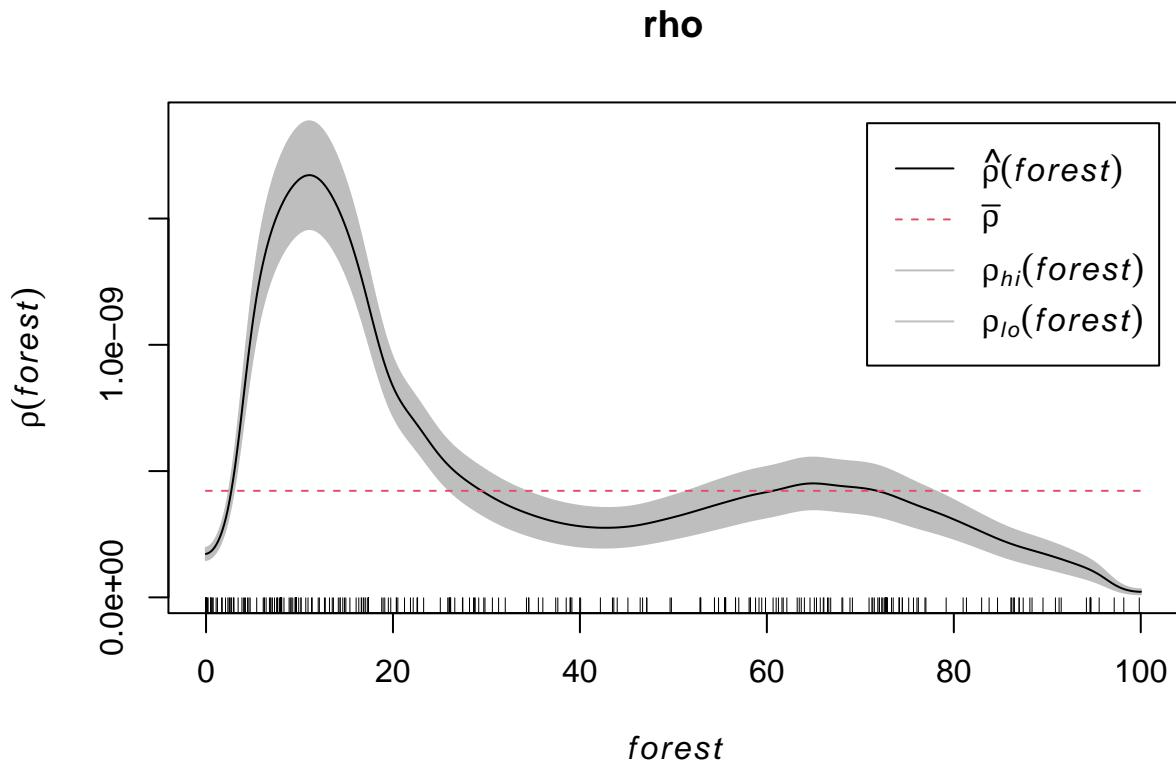
```
## tile
##          (-130,761]      (761,1.1e+03]  (1.1e+03,1.46e+03] (1.46e+03,3.56e+03]
##          392                  2                 5                  1
```



Covariate Variable : Forest

```
forest <- BC$Forest
b <- quantile(forest,probs=(0:4)/4,type=2)
Zcut <- cut(forest,breaks=b)
V <- tess(image=Zcut)
quadratcount(asc_data_ppp,tess=V)
```

```
## tile
##      (0,11.6] (11.6,50.2] (50.2,86.9] (86.9,100]
##          197         93         81         29
```



Covariate Variable : Distance to Water

```

dist <- BC$Dist_Water
b <- quantile(dist,probs=(0:4)/4,type=2)
Zcut <- cut(dist,breaks=b)
V <- tess(image=Zcut)
quadratcount(asc_data_ppp,tess=V)

```

```

## tile
##          (0,483]      (483,1.1e+03]  (1.1e+03,2.18e+03] (2.18e+03,2.32e+04]
##          213                 47            38                  96

```

Covariate Variable Observation

Observation :

- For Elevation Level :
 - It is obvious that the fungi is overwhelmingly correlated to the low elevation level.
 - The proportion of occurrence that appears in the lowest elevation sector accounts for more than 97% of the identified points.
- For Forest and Distance to Water :

- It is observed that the fungi is correlated to both Forest and distance .
- The proportion of occurrence that appears in the respective smallest sectors account for about 50% of the identified points.

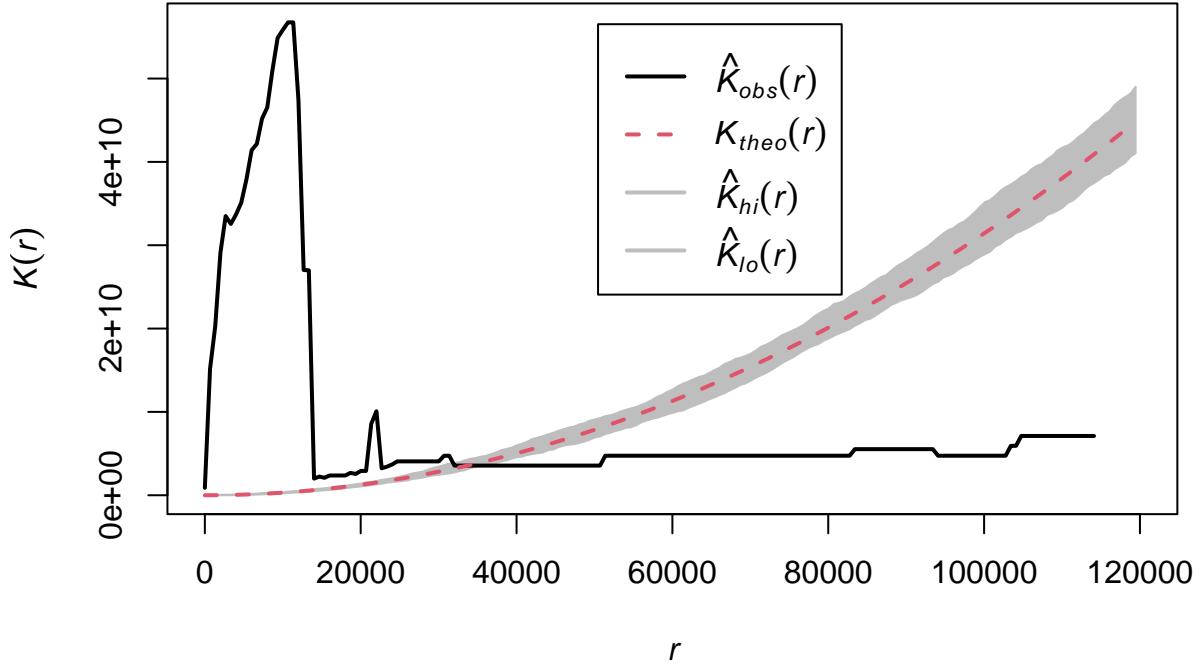
Second Moment Descriptives

Ripley's K-function

```
# Bootstrapped CIs
# rank = 1 means the max and min
# Border correction is to correct for edges around the window
# values will be used for CI
E_asc <- envelope(asc_data_ppp,
                    Kest,
                    correction="border",
                    rank = 1,
                    nsim = 19,
                    fix.n = T)

## Generating 19 simulations of CSR with fixed number of points ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19.
##
## Done.

# visualise the results
plot(E_asc,
      main = "",
      lwd = 2,
      xlim=c(0,120000))
```



Now, we are going to correct for inhomogeneity...

```

lambda_asc <- density(asc_data_ppp, bw.ppl)
Kinhom_asc <- Kinhom(asc_data_ppp, lambda_asc)

#Estimate a strictly positive density
lambda_asc_pos <- density(asc_data_ppp,
                           sigma=bw.ppl,
                           positive=TRUE)

#Simulation envelope (with points drawn from the estimated intensity)
E_asc_inhom <- envelope(asc_data_ppp,
                         Kinhom,
                         simulate = expression(rpoispp(lambda_asc_pos)),
                         correction="border",
                         rank = 1,
                         nsim = 19,
                         fix.n = TRUE)

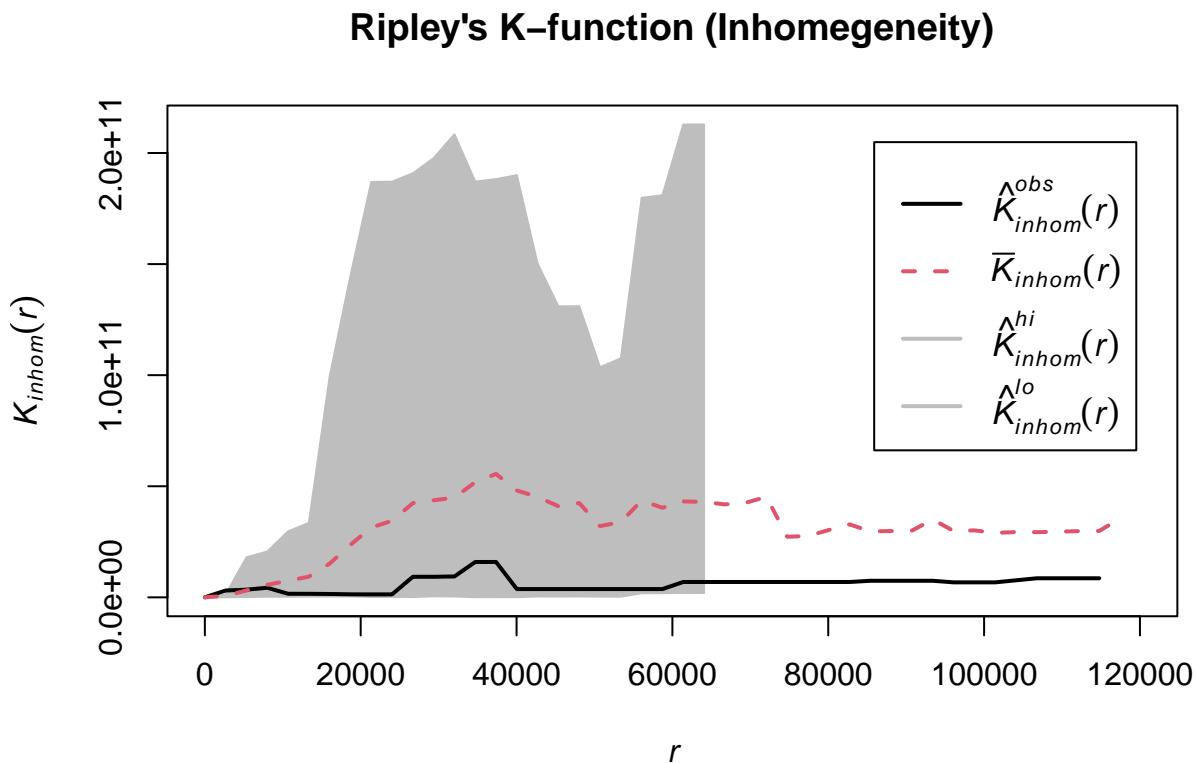
## Generating 19 simulations by evaluating expression ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19.
##
## Done.

```

```

# visualise the results
# par(mfrow = c(1,2))
plot(E_asc_inhom,
     main = "Ripley's K-function (Inhomegeneity)",
     lwd = 2,
     xlim=c(0,120000))

```



For homogeneous case, the observed data show a strong clustering at small r value range. When corrected for inhomogeneity, such clustering pattern is no longer observed.

Pair Correlation Function

```

# Estimate the g function
pcf_asc <- pcf(asc_data_ppp)

pcf_asc

## Function value object (class 'fv')
## for the function r -> g(r)
##
##          Math.label      Description
##          r              distance argument r
## theo  g[Pois](r)    theoretical Poisson g(r)
## trans hat(g)[Trans](r) translation-corrected estimate of g(r)

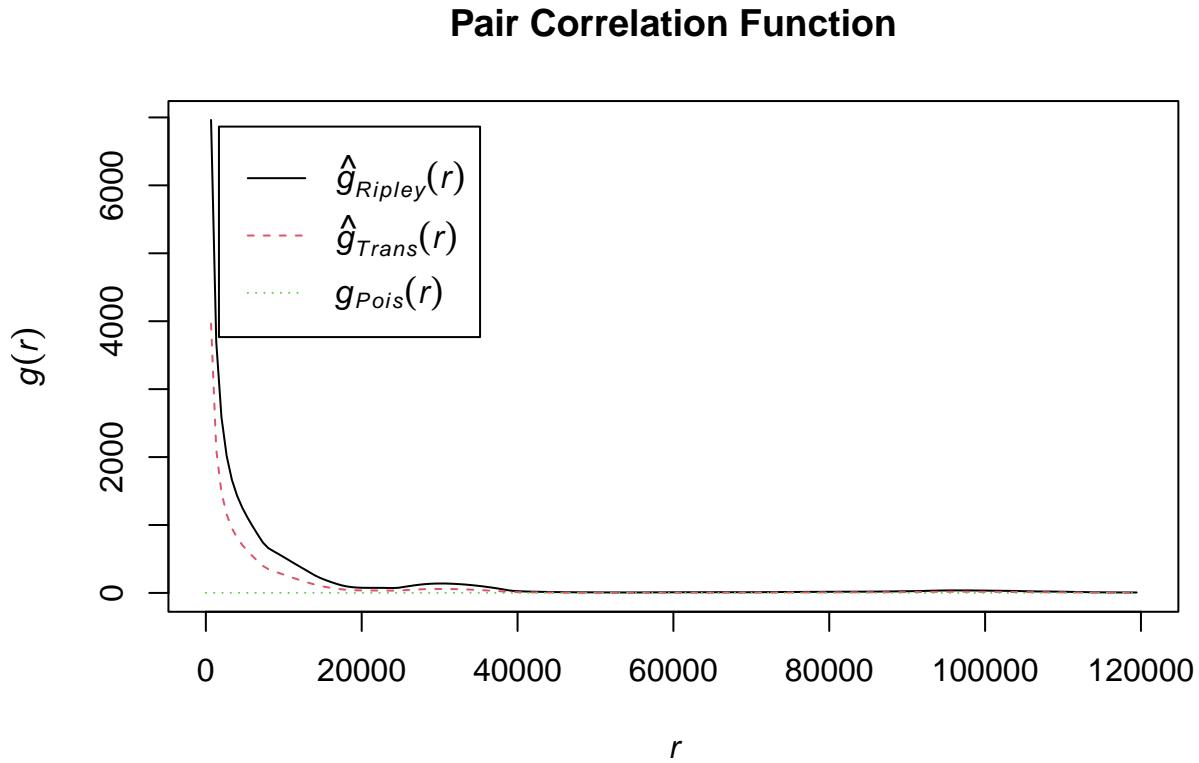
```

```

## iso  hat(g)[Ripley](r) isotropic-corrected estimate of g(r)
## .....
## Default plot formula: .~r
## where "." stands for 'iso', 'trans', 'theo'
## Recommended range of argument r: [0, 341660]
## Available range of argument r: [0, 341660]

plot(pcf_asc, xlim=c(0,120000), main="Pair Correlation Function")

```



The above estimator also assumes homogeneity. Let's relax this assumption.

```

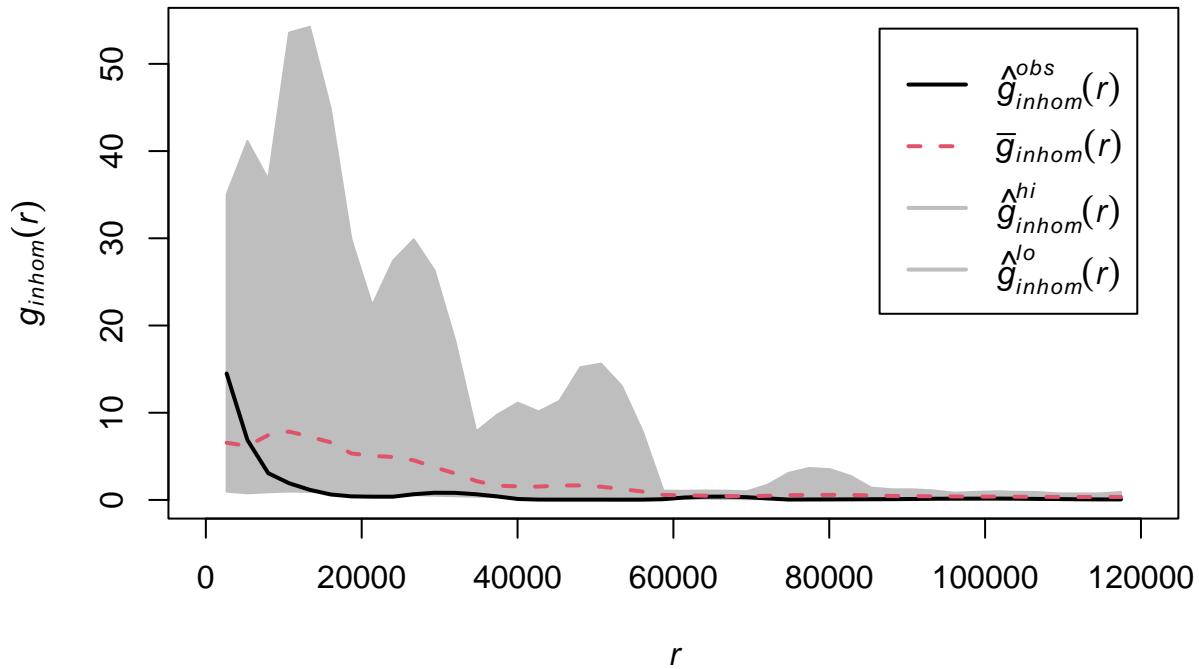
#Simulation envelope (with points drawn from the estimated intensity)
pcf_asc_inhom <- envelope(asc_data_ppp,
                           pcfinhom,
                           simulate = expression(rpoispp(lambda_asc_pos)),
                           rank = 1,
                           nsim = 19)

## Generating 19 simulations by evaluating expression ...
## 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19.
##
## Done.

plot(pcf_asc_inhom,
      xlim = c(0,120000),

```

```
main = "",  
lwd = 2)
```



When corrected for homogeneity, the locations of the fungi appear not to have any significant correlations.

Model Fitting and Selection with AIC

```
fit <- ppm(asc_data_ppp ~ Elevation + Dist_Water + Forest, data = BC)
fit

## Nonstationary Poisson process
## Fitted to point pattern dataset 'asc_data_ppp'
##
## Log intensity: ~Elevation + Dist_Water + Forest
##
## Fitted trend coefficients:
##   (Intercept)    Elevation    Dist_Water      Forest
## -1.649783e+01 -9.746289e-03  5.074702e-05 -2.772079e-02
##
##             Estimate       S.E.     CI95.lo     CI95.hi Ztest
## (Intercept) -1.649783e+01 8.719049e-02 -1.666872e+01 -1.632694e+01 *** 
## Elevation   -9.746289e-03 3.872771e-04 -1.050534e-02 -8.987240e-03 *** 
## Dist_Water   5.074702e-05 2.503222e-05  1.684771e-06  9.980928e-05 *  
## Forest      -2.772079e-02 1.803605e-03 -3.125579e-02 -2.418579e-02 ***
```

```

##                  Zval
## (Intercept) -189.215943
## Elevation    -25.166190
## Dist_Water    2.027268
## Forest       -15.369657
## *** Fitting algorithm for 'glm' did not converge ***

```

Elevation and Forest are significant but Dist_Water isn't. OK, then try higher order.

```

fit <- ppm(asc_data_ppp ~ Elevation + I(Elevation^2) + Dist_Water + I(Dist_Water^2), data = BC)
fit

```

```

## Nonstationary Poisson process
## Fitted to point pattern dataset 'asc_data_ppp'
##
## Log intensity: ~Elevation + I(Elevation^2) + Dist_Water + I(Dist_Water^2)
##
## Fitted trend coefficients:
##      (Intercept)      Elevation  I(Elevation^2)      Dist_Water  I(Dist_Water^2)
## -1.751444e+01 -1.298691e-02  3.757783e-06   2.313973e-04 -2.416955e-08
##
##             Estimate        S.E.      CI95.lo      CI95.hi Ztest
## (Intercept) -1.751444e+01 8.075694e-02 -1.767272e+01 -1.735616e+01 ***
## Elevation   -1.298691e-02 5.148358e-04 -1.399597e-02 -1.197785e-02 ***
## I(Elevation^2) 3.757783e-06 1.849431e-07  3.395301e-06  4.120264e-06 ***
## Dist_Water   2.313973e-04 7.552735e-05  8.336647e-05  3.794282e-04 **
## I(Dist_Water^2) -2.416955e-08 1.138818e-08 -4.648997e-08 -1.849130e-09 *
##
##                  Zval
## (Intercept) -216.878412
## Elevation    -25.225346
## I(Elevation^2) 20.318594
## Dist_Water     3.063756
## I(Dist_Water^2) -2.122337
## *** Fitting algorithm for 'glm' did not converge ***

```

OK, after trying several rounds, the following seems to be a improved fit :

```

fit_simple <- ppm(asc_data_ppp ~ Elevation + Forest, data = BC)
fit_simple

```

```

## Nonstationary Poisson process
## Fitted to point pattern dataset 'asc_data_ppp'
##
## Log intensity: ~Elevation + Forest
##
## Fitted trend coefficients:
##      (Intercept)      Elevation      Forest
## -16.449678447 -0.009601581 -0.027922868
##
##             Estimate        S.E.      CI95.lo      CI95.hi Ztest
## (Intercept) -16.449678447 0.0831309100 -16.61261204 -16.286744858 ***
## Elevation   -0.009601581 0.0003758118 -0.01033816 -0.008865003 ***
## Forest      -0.027922868 0.0018099253 -0.03147026 -0.024375480 ***

```

```

##          Zval
## (Intercept) -197.87680
## Elevation    -25.54891
## Forest       -15.42764
## *** Fitting algorithm for 'glm' did not converge ***

fit <- ppm(asc_data_ppp ~ Elevation + I(Elevation^2) + Forest + I(Forest^2), data = BC)
fit

## Nonstationary Poisson process
## Fitted to point pattern dataset 'asc_data_ppp'
##
## Log intensity: ~Elevation + I(Elevation^2) + Forest + I(Forest^2)
##
## Fitted trend coefficients:
## (Intercept)   Elevation I(Elevation^2)      Forest   I(Forest^2)
## -1.621033e+01 -1.173001e-02  3.259440e-06 -4.196879e-02  1.826759e-04
##
##           Estimate      S.E.      CI95.lo      CI95.hi Ztest
## (Intercept) -1.621033e+01 9.367906e-02 -1.639394e+01 -1.602673e+01 ***
## Elevation   -1.173001e-02 4.272309e-04 -1.256736e-02 -1.089265e-02 ***
## I(Elevation^2) 3.259440e-06 1.877594e-07  2.891438e-06  3.627441e-06 ***
## Forest      -4.196879e-02 5.890907e-03 -5.351476e-02 -3.042283e-02 ***
## I(Forest^2)  1.826759e-04 6.812603e-05  4.915130e-05  3.162004e-04 **
##
##          Zval
## (Intercept) -173.041168
## Elevation    -27.455891
## I(Elevation^2) 17.359658
## Forest       -7.124334
## I(Forest^2)   2.681440
## *** Fitting algorithm for 'glm' did not converge ***

```

Now, let's see if the complicated model is worth it.

```
#AIC values
AIC(fit); AIC(fit_simple)
```

```
## [1] 15464.31
```

```
## [1] 15508.45
```

```
#Delta AIC
AIC(fit_simple) - AIC(fit)
```

```
## [1] 44.14666
```

Also, conduct a anova LRT test to compare the two models :

```
anova(fit_simple, fit, test = "LRT")
```

```

## Analysis of Deviance Table
##
## Model 1: ~Elevation + Forest      Poisson
## Model 2: ~Elevation + I(Elevation^2) + Forest + I(Forest^2)  Poisson
##   Npar Df Deviance  Pr(>Chi)
## 1     3
## 2     5  2  48.147 3.508e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The following conclusion is drawn : - The model with quadratic terms provides a better fit to the data - With the delta AIC of 44, the extra complexity is well supported. - So, a good model should be

$$\lambda_{ASC}(u) = e^{-16.2 - 0.017 \text{elevation}(u) - 0.000003 \text{elevation}(u)^2 - 0.042 \text{forest}(u) + 0.00018 \text{forest}(u)^2}$$

Model Validation

```

#Run the quadrat test
quadrat.test(fit, nx = 4, ny = 4)

##
## Chi-squared test of fitted Poisson model 'fit' using quadrat counts
##
## data: data from fit
## X2 = 585.05, df = 8, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 13 tiles (irregular windows)

```

This has small p value, suggesting significant deviation from our model's prediction. Room for further improvement is therefore expected, but it does not provide hint for how to achieve any improvement.

PPP Residuals

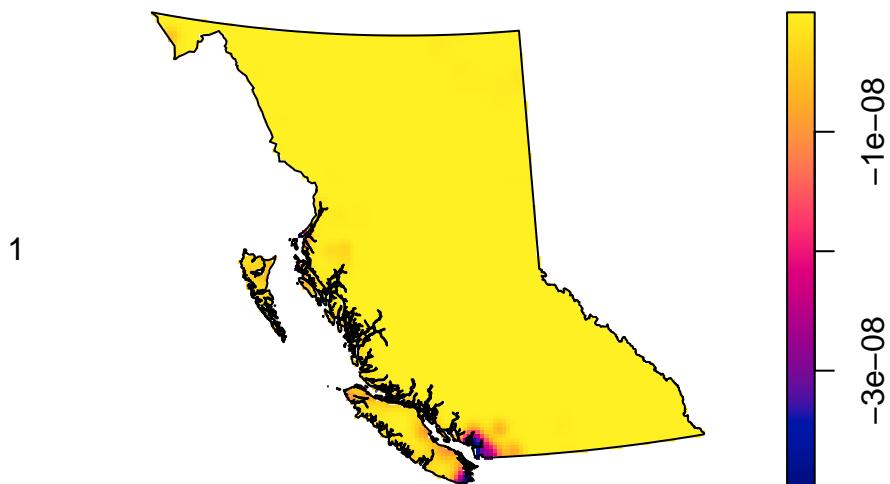
```

#Calculate the residuals
res <- residuals(fit)

#Visualise
plot(res,
      cols = "transparent",
      main="Residuals Plots")

```

Residuals Plots



From the plot, it's observed that :

- A clear pattern is detected : Majority of the residual plots are in the “high residuals region”
- The residuals are small in magnitude (e^{-8}).
- The negative residual values suggest over-prediction.
- there is room for improvement for the model

Higher Order Polynomial Fitting with Spline and Validation

As the above analysis indicated a room for improvement, we will finally try to add higher-order polynomials with the spline packages.

```
library(splines)

#Fit the PPP model
fit_smooth <- ppm(asc_data_ppp ~ bs(Elevation, 7) + bs(Forest, 3), data = BC, use.gam = TRUE)

fit_smooth

## Nonstationary Poisson process
## Fitted to point pattern dataset 'asc_data_ppp'
##
## Log intensity: ~bs(Elevation, 7) + bs(Forest, 3)
##
```

```

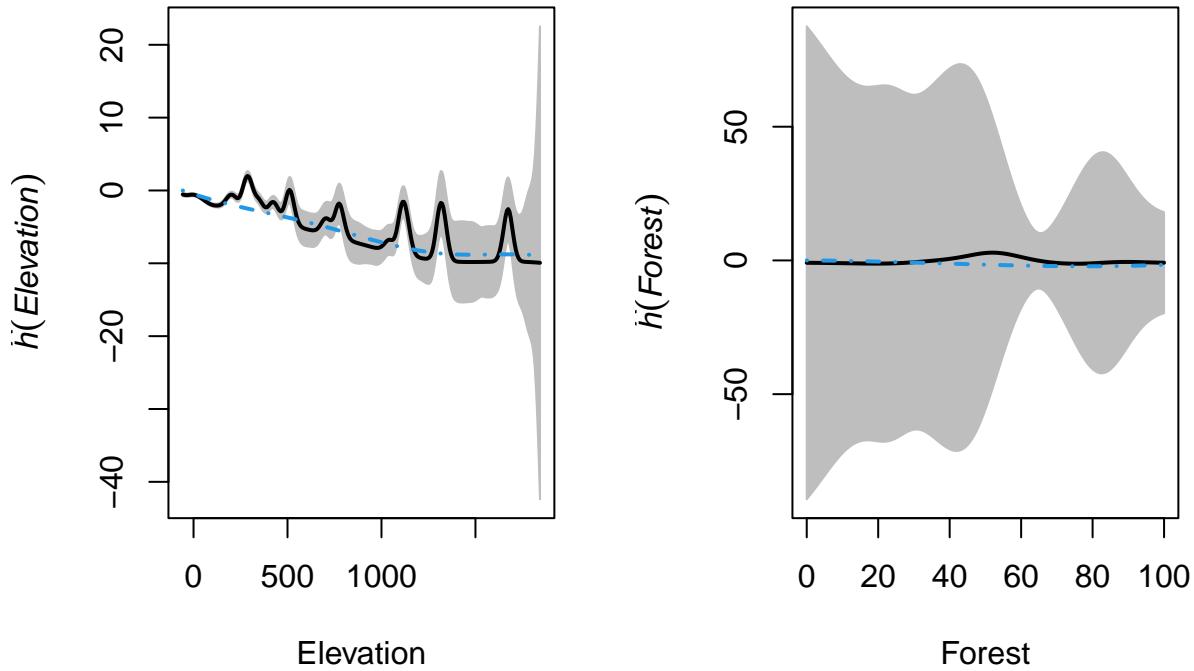
## Fitted trend coefficients:
##      (Intercept) bs(Elevation, 7)1 bs(Elevation, 7)2 bs(Elevation, 7)3
## -15.1837101      -2.1215999      -3.9107238     -10.2916364
## bs(Elevation, 7)4 bs(Elevation, 7)5 bs(Elevation, 7)6 bs(Elevation, 7)7
## -6.9600818      -20.1516125      49.1493840    -1129.0583122
##   bs(Forest, 3)1   bs(Forest, 3)2   bs(Forest, 3)3
## -0.2110034      -3.4557922      -1.6215046
##
## For standard errors, type coef(summary(x))

#Calculate the partial residuals as a function of elevation
par_res_elev <- parres(fit_smooth, "Elevation")

#Calculate the relative intensity as a function of gradient
par_res_forest <- parres(fit_smooth, "Forest")

#Side by side plotting
par(mfrow = c(1,2))
plot(par_res_elev,
      legend = FALSE,
      lwd = 2,
      main = "",
      xlab = "Elevation")
plot(par_res_forest,
      legend = FALSE,
      lwd = 2,
      main = "",
      xlab = "Forest")

```



```

#AIC values
AIC(fit); AIC(fit_smooth)

## [1] 15464.31

## [1] 15440.68

#Delta AIC
AIC(fit) - AIC(fit_smooth)

## [1] 23.62747

#Likelihood ratio test
anova(fit, fit_smooth, test = "LRT")

## Analysis of Deviance Table
##
## Model 1: ~Elevation + I(Elevation^2) + Forest + I(Forest^2)    Poisson
## Model 2: ~bs(Elevation, 7) + bs(Forest, 3)    Poisson
##   Npar Df Deviance  Pr(>Chi)
## 1     5
## 2    11  6  35.622 3.264e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

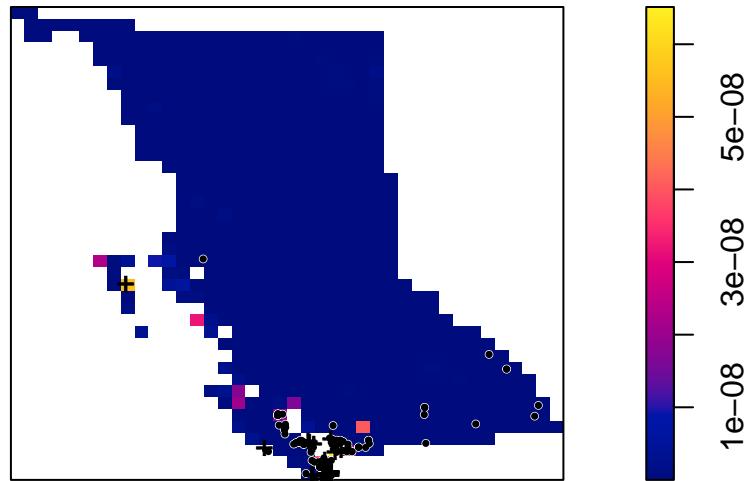
```

All these suggest that this complex models provides a better fit to the data. Let's finally visualize the predictions as before.

```
#Plot the model predictions
plot(fit_smooth,
      se = FALSE,
      superimpose = FALSE,
      main = "Estimated Fungi intensity")

#Overlay the locations
plot(asc_data_ppp,
      pch = 16,
      cex = 0.6,
      cols = "white",
      add = TRUE)
plot(asc_data_ppp,
      pch = 16,
      cex = 0.5,
      cols = "black",
      add = TRUE)
```

Estimated Fungi intensity



From this visualisation, the following is observed :

- Although the model is not yet perfect, it is progressively having improvement after rounds of variables selection process.

- Considering the fact that we are predicting the locations of one species of fungi in a biodiverse continent based only on Elevation and Forest, and have no information on all of the many other factors that would significantly influence fungi growth (e.g. humidity, moisture level, temperature, pH value, oxygen content, etc.)

Results :

Discussion:

Provide a brief summary of your findings. Length: ca. 1 page.

References:

Include references to all necessary literature.