Supplementary Material

Quantifying Novel Ecosystems

August, 2025

3.1 BNI Calculation

3.1.1 Obtaining and Cleaning the Data

After cleaning the data obtained from the British Trust for Ornithology (Gillings et al., 2019) and AVONET (Tobias et al., 2022), separate data frames can be formed with a list of species present in each Atlas period, with accompanying data frames containing the trait measurements for each of these unique species, separated by Atlas period. Examples of which are shown below:

Example Distribution Data Frame Structure

Grid	Scientific_name
HP40	Gavia stellata
IL83	Cepphus grylle
NX64	Alca torda
IX99	Uria aalge
IG53	Sterna hirundo

Example Traits Data Frame Structure

scientific_name	Beak.Length_Culmen	Beak.Length_Nares	Beak.Width	Beak.Depth	Tarsus.Length	Wing.Length	Kipps.Distance
Gavia stellata	63.1	37.4	7.5	12.2	66.1	275.4	130.7
Gavia arctica	71.0	41.3	8.5	14.2	72.8	285.8	145.3
Gavia immer	97.0	60.1	10.5	23.0	86.1	365.2	178.0
Tachybaptus ruficollis	23.9	11.5	4.4	6.7	35.4	100.0	31.6
Podiceps cristatus	57.5	35.4	6.6	10.6	60.9	183.6	79.0
Podiceps nigricollis	26.2	15.4	5.5	5.9	39.6	127.2	55.1
Fulmarus glacialis	41.3	23.6	11.9	14.4	48.8	321.1	162.5
Puffinus puffinus	34.6	27.3	6.4	7.9	41.7	232.5	143.2
Hydrobates pelagicus	13.7	6.1	2.3	3.5	22.2	119.0	69.8
Hydrobates leucorhous	17.8	10.6	3.1	4.5	23.1	150.4	88.2
Morus bassanus	98.7	98.0	24.8	32.0	60.0	485.8	300.5
Phalacrocorax carbo	82.7	57.0	13.2	18.3	64.2	340.2	124.2
Gulosus aristotelis	66.0	50.2	10.6	13.1	61.6	262.6	87.0

3.1.2 Distance Matrix (Dist.mat)

For each Atlas period, a trait matrix needs to be curated by removing the “Scientific_name” column and calculating the Euclidean distance between trait values. These distances were then scaled by dividing each distance pair by its maximum distance to develop distance matrices, examples shown below, that are suitable for substitution into appropriate areas of the BNI function provided by Schittko et al. (2020).

Example Distance Matrix Structure

row_name	Acanthis.flammea	Accipiter.gentilis	Accipiter.nisus	Acrocephalus.schoenobaenus	Acrocephalus.scirpaceus
Acanthis flammea	0.0000000	0.0871359	0.0265659	0.0017036	0.0018137

row_name	Acanthis.flammea	Accipiter.gentilis	Accipiter.nisus	Acrocephalus.schoenobaenus	Acrocephalus.scirpaceus
Accipiter gentilis	0.0871359	0.0000000	0.0626882	0.0878111	0.0877126
Accipiter nisus	0.0265659	0.0626882	0.0000000	0.0275994	0.0274756
Acrocephalus schoenobaenus	0.0017036	0.0878111	0.0275994	0.0000000	0.0005397
Acrocephalus scirpaceus	0.0018137	0.0877126	0.0274756	0.0005397	0.0000000
Actitis hypoleucos	0.0055425	0.0829637	0.0223200	0.0066343	0.0067793
Aegithalos caudatus	0.0033709	0.0875166	0.0269278	0.0037794	0.0035284
Aix galericulata	0.0546310	0.0345380	0.0332320	0.0551890	0.0551543
Aix sponsa	0.0627306	0.0276714	0.0413734	0.0632457	0.0632037
Alauda arvensis	0.0044023	0.0836452	0.0225092	0.0056313	0.0055784
Alca torda	0.0686021	0.0256346	0.0480275	0.0690593	0.0690336
Alcedo atthis	0.0039106	0.0862443	0.0266496	0.0036033	0.0036456

3.1.3 Years Since Introduction (YSI)

To calculate years since introduction (YSI), the ‘mutate()’ command from the R package ‘dplyr’ (v1.1.0; Wickham et al., 2023) was used to subtract the introduction date from the end year of that particular Atlas period. For example, the calculation of the YSI for species in the first Atlas period would be:

```
T1YSI <- T1_intro_dates %>% mutate(YSI = 1972 – time)
```

The dates to subtract from for the second and third Atlas periods will be 1991 and 2011, respectively. Leaving you with three data frames for each Atlas period displaying each species name and its calculated years since introduction:

Example YSI Data Frame Structure

Scientific_name	time	YSI
Acanthis flammea	1492	480
Accipiter gentilis	1492	480
Accipiter nisus	1492	480
Acrocephalus schoenobaenus	1492	480
Acrocephalus scirpaceus	1492	480
Actitis hypoleucos	1492	480
Aegithalos caudatus	1492	480
Aix galericulata	1886	86
Aix sponsa	1873	99
Alauda arvensis	1492	480

3.1.4 Community Matrix (Com)

To make a presence-absence matrix for the BNI calculation, the original distribution data frames need to be manipulated further. For this they are converted to a table using ‘table()’ , following with ‘as.data.frame.matrix()’ from the base R package (v4.2.3; R Core Team, 2023). With these three data frames completed, they can be put into the R function to calculate BNI per grid cell. An example of the data frame matrix structure is shown here:

Example Community Matrix Structure

Grid	Acanthis.flammea	Accipiter.gentilis	Accipiter.nisus	Acrocephalus.schoenobaenus	Acrocephalus.scirpaceus	Actitis.hypoleucos
IB61	0	0	0	0	0	0
IB70	0	0	0	1	1	0
IB71	1	0	1	1	1	0
IB72	0	0	0	1	1	0

Grid	Acanthis.flammea	Accipiter.gentilis	Accipiter.nisus	Acrocephalus.schoenobaenus	Acrocephalus.scirpaceus	Actitis.hypoleucos
IB73	0	0	0	0	0	0
IB80	1	0	0	1	1	0
IB81	1	0	0	1	1	0
IB82	1	0	0	1	0	0
IB83	1	0	0	1	0	0
IB84	0	0	0	0	0	0
IB90	1	0	0	1	0	0
IB91	0	0	1	1	0	0

3.1.5 Calculating BNI

Using ‘com’, ‘dist.mat’ and ‘YSI’, the BNI function was ran three times to accommodate for each of the BTO breeding bird survey Atlas periods in order to get a measure of novelty across space and time. Each output was saved and exported as a .csv file using ‘write_csv()’ from the ‘readr’ package (Wickham et al., 2023). The outputs of BNI, Rao’s Q and BNIs are saved as data frames and should appear as such:

Example BNIs Output

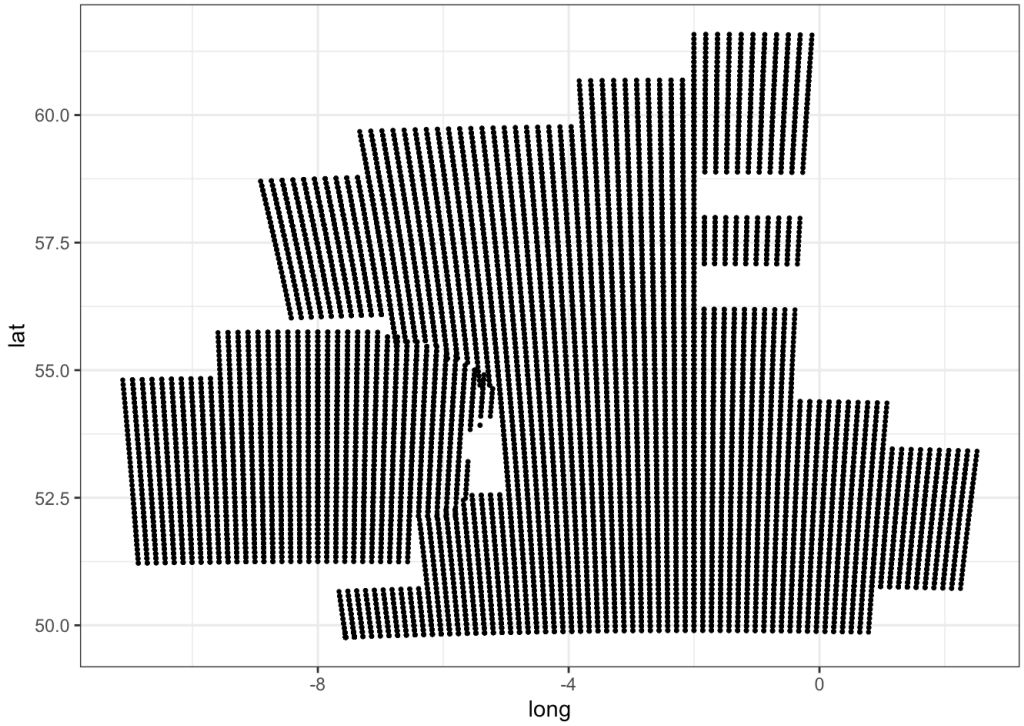
Grid	BNIs
IS76	0.0417057
IS77	0.0430961
IS78	0.0000000
IS79	0.0000000
IS80	0.0000000
IS81	0.0000000
IS82	0.0000000
IS83	0.0206145
IS84	0.0000000
IS85	0.0098637
IS86	0.0000000
IS87	0.0000000

3.1.6 Index Outputs

For this section, the packages required were mapview, sf, and RColorBrewer (v.11.2; Appelhans T et al., 2023; v1.0-15; Pebesma, E., and Bivand, R., 2023; v1.1-3; Neuwirth E, 2022, respectively). To manipulate the results data into a suitable format for mapping, the row names needed to be moved into the first column, resulting in a column for grid cells and another for BNIs values repeating for all three data frames with the BNIs results. BNIs was mapped over BNI as it is standardised over Rao’s Q, bounding the results between 0 and 1, thus making results more interpretable in a map format, following the approach of Schittko et al. (2020) to show the spatial relationship of biotic novelty.

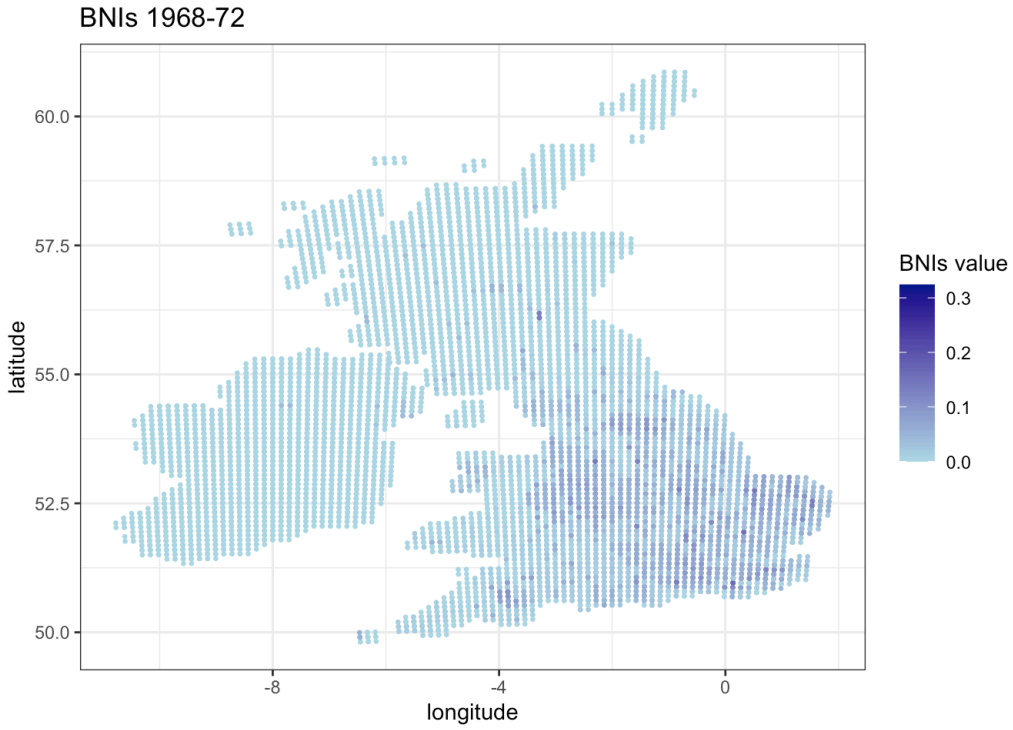
To get coordinates for mapping, the file ‘grid_square_coordinates_lookup.csv’ from the downloaded BTO breeding bird survey data was read into R. Initially, ‘ggplot’ (Wickham,2016), was used to produce a quick map of grid cells to ensure they produced a map of the UK and Ireland with:

```
ggplot(grids, aes(x = long, y = lat)) +
  geom_point(size = 0.5)+
  theme_bw()
```



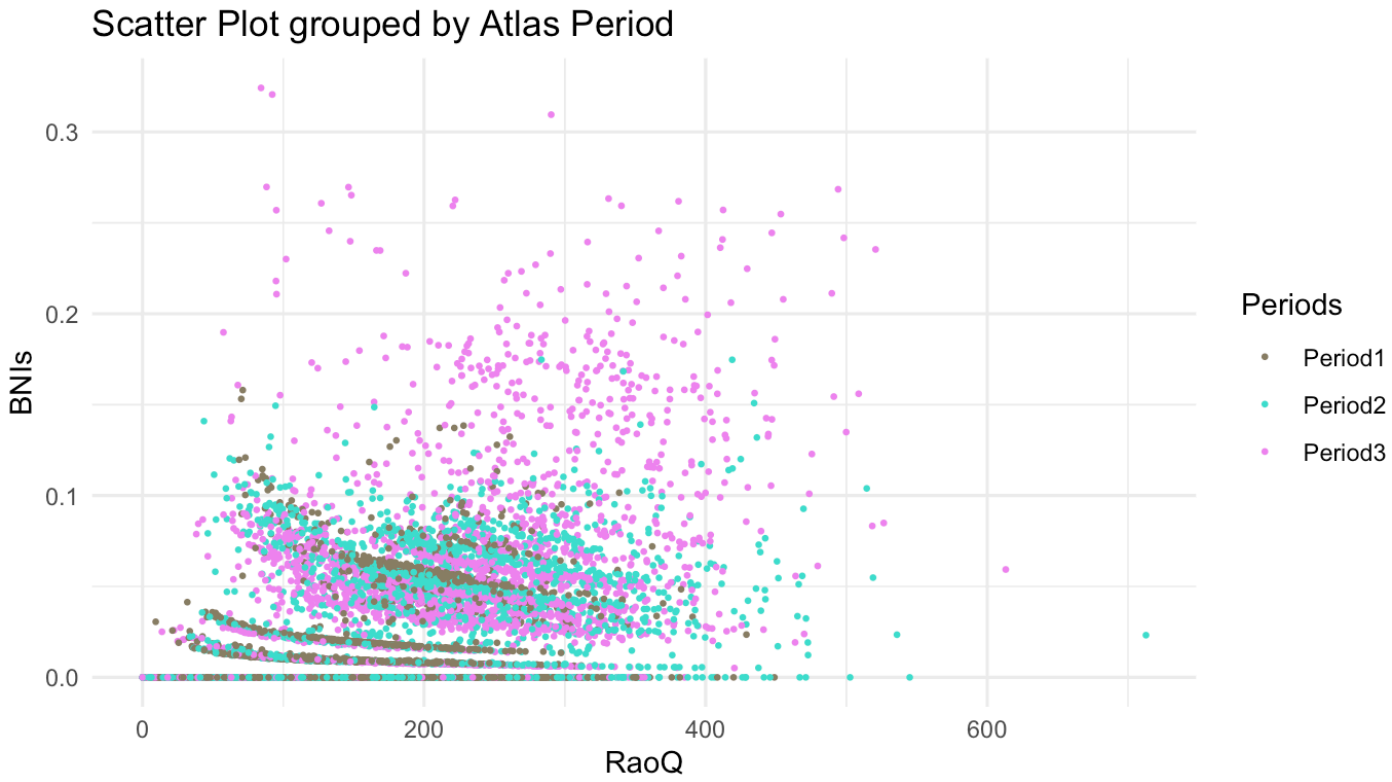
Then 'left_join' was used to join this data frame with the BNI function results 'by = "grid"', removing multiples with "multiple = all" and subsequently removing NAs. Next, the data frames were converted to spatial data using 'st_as_sf', setting the coordinates reference system to crs = 27700, which identifies the OSGB coordinates reference system. Here, the minimum and maximum BNIs values were calculated across all Atlas periods, using 'min()' and 'max()'. These were saved as new data frames, which will be incorporated into the map code to provide values for a consistent scale bar across maps. The code used to produce the maps is shown below.

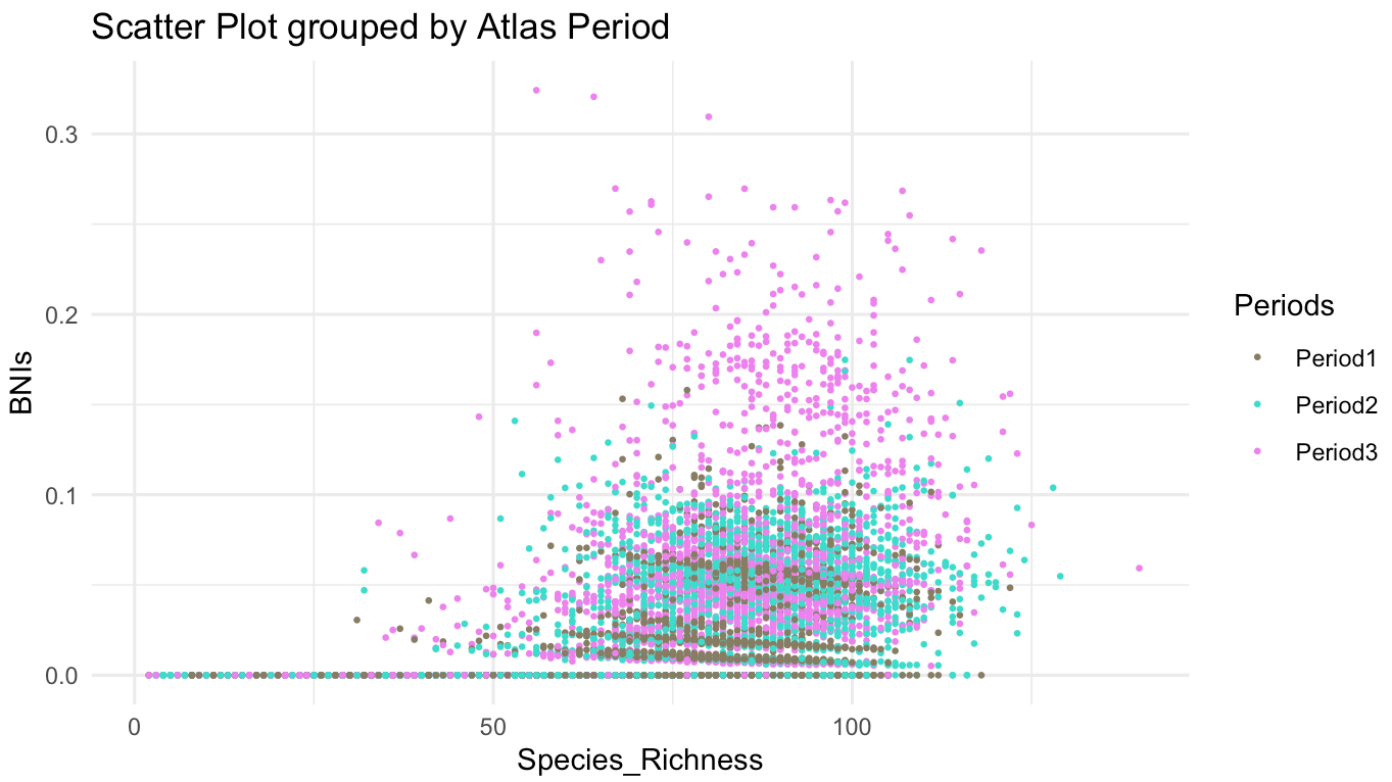
```
ggplot(grid_BNIs1, aes(x = long, y = lat, color = grid_BNIs1$BNIs)) +
  geom_point(size = 0.5) +
  scale_color_gradient(low = "lightblue", high = "darkblue",
                       limits = c(min_value_BNIs, max_value_BNIs),
                       name = "BNIs value") +
  labs(title = "BNIs 1968-72",
       x = "longitude",
       y = "latitude") +
  theme_bw()
```



To further investigate the drivers of BNIs values, BNIs was plotted against Rao's Q and species richness to produce scatter plots. There is a rough clustering of points, especially at the lower end of Rao's Q values, suggesting that at lower RaoQ values, there is a wider range of BNIs values, while at higher RaoQ values, BNIs values are more consistently low. Moreover, there is significant overlap between points from different

Atlas periods, inferring that the relationship between Rao's Q, species richness and BNIs is not strongly influenced by the Atlas period. In the scatter depicting the relationship between Rao's Q and BNIs, there is a slight inverse relationship observed, showing that as Rao's Q values increase, BNIs values decrease. This suggests that communities with higher diversity tend to have lower novelty. This is counterintuitive, as you might expect more diverse communities to increase the chances of including invasive species, but the results in the next scatter support this inference, as increased species richness does not necessarily mean higher BNIs values.





However, what is evidenced in these plots is that even communities with fewer species can exhibit either high or low levels of novelty. This is expected when using a highly mobile taxonomic group such as birds, as they frequent different grid cells at different periods, thus constantly changing the functional measures of that cell, causing functional diversity to grow non-linearly throughout time.

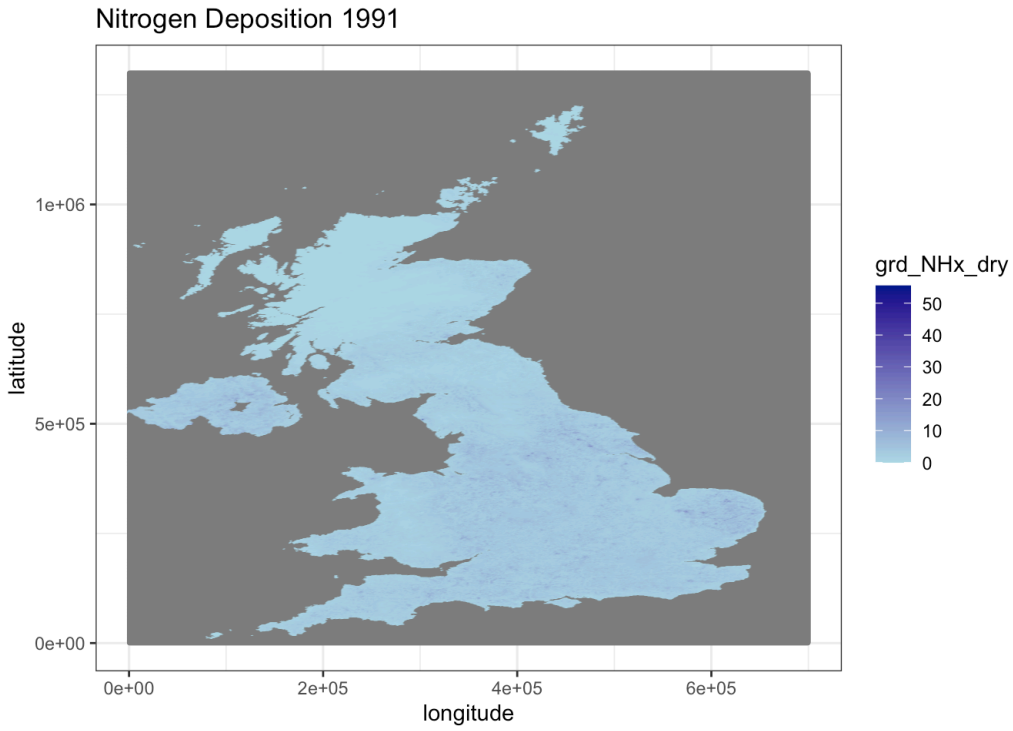
A stronger positive correlation between BNIs and Rao's Q was expected as they both contain a measure of functional distance between pairs of species. However, it is most likely the inclusion of the coexistence matrix in the BNI formula that creates this disparity between measures.

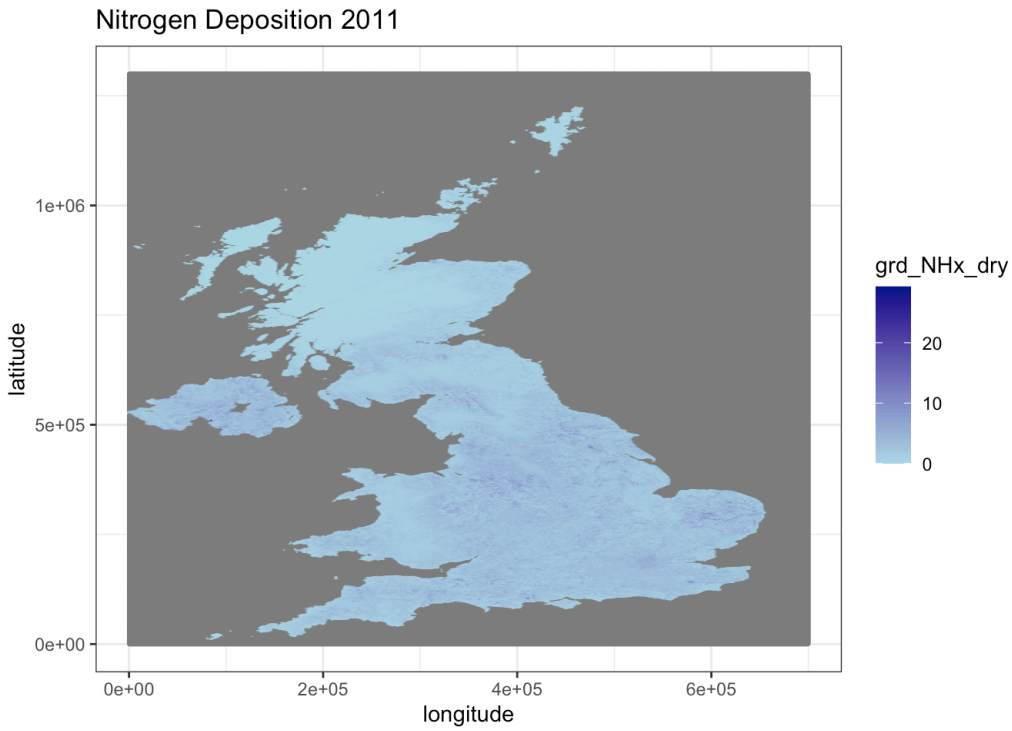
3.2 SED Calculation

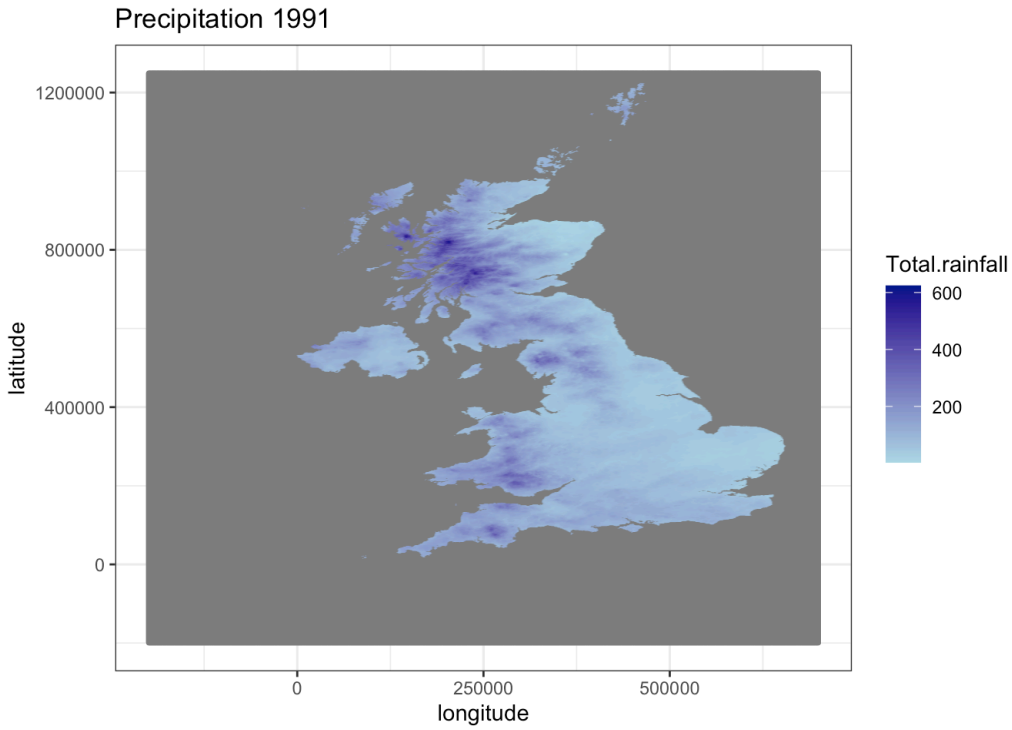
3.2.1 Obtaining and Cleaning the Data

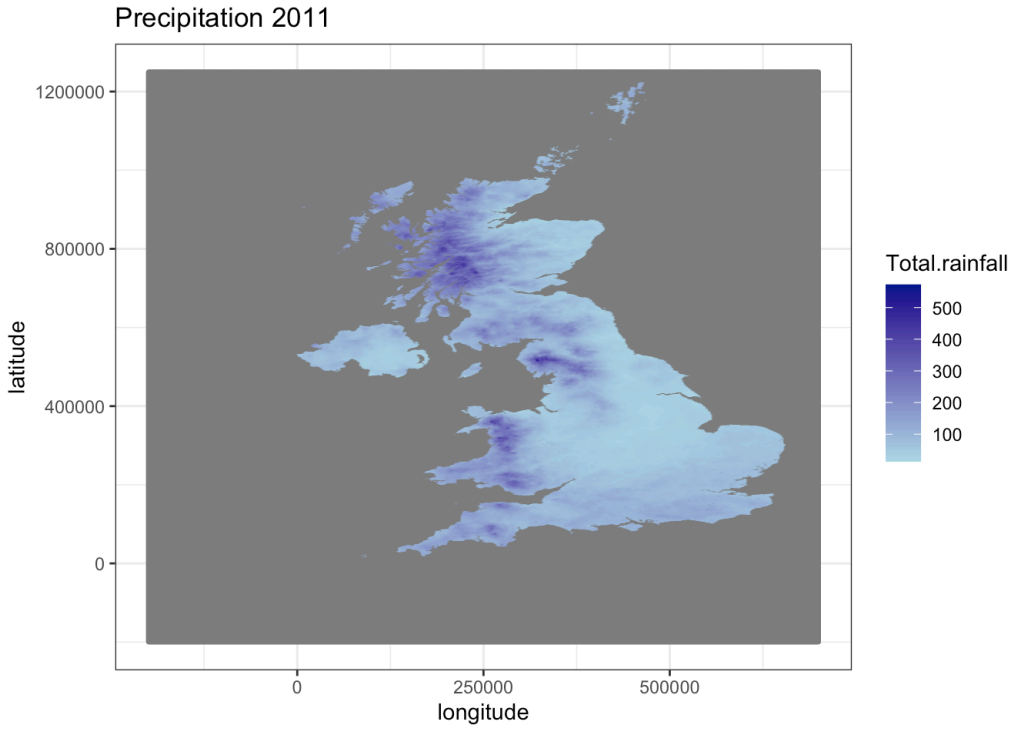
Maps and boxplots as the initial exploration of the downloaded data for nitrogen deposition, precipitation and temperature were produced.

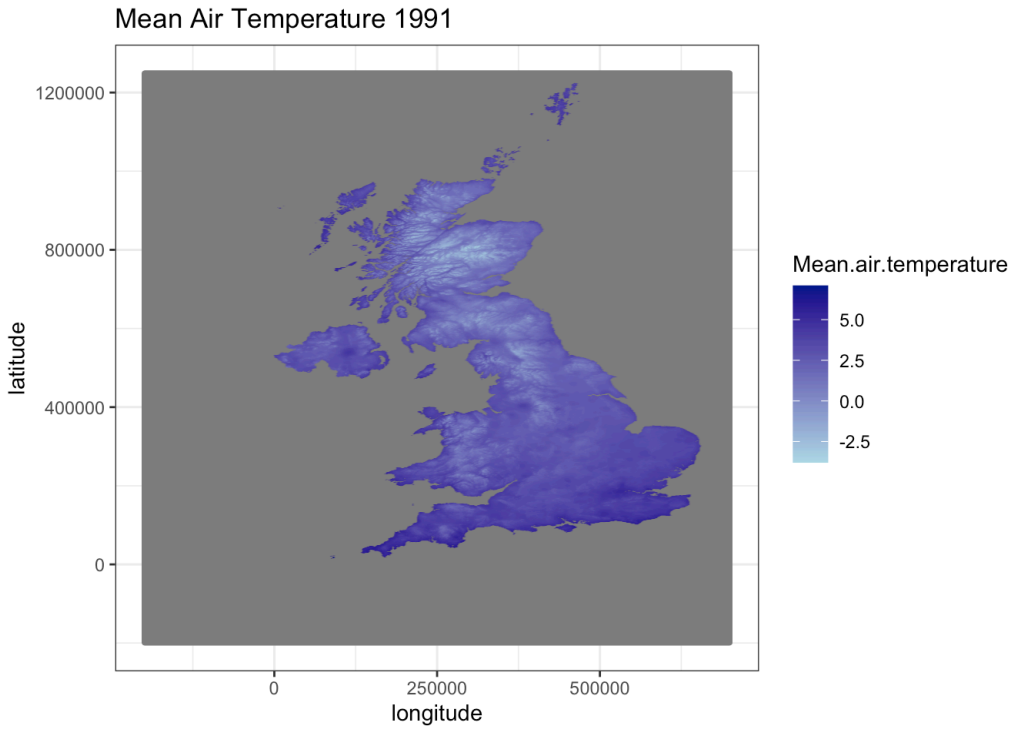
```
ggplot(N1991, aes(x = x, y = y, color = grd_NHx_dry)) +
  geom_point(size = 0.5) +
  scale_color_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "Nitrogen Deposition 1991",
       x = "longitude",
       y = "latitude") +
  theme_bw()
```

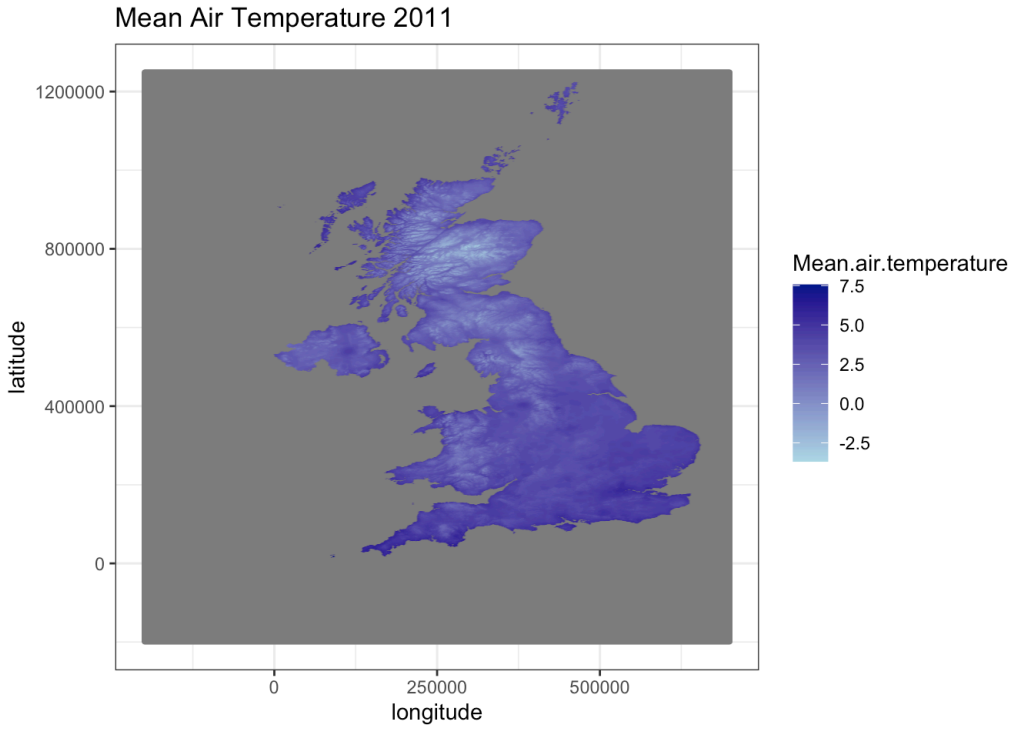


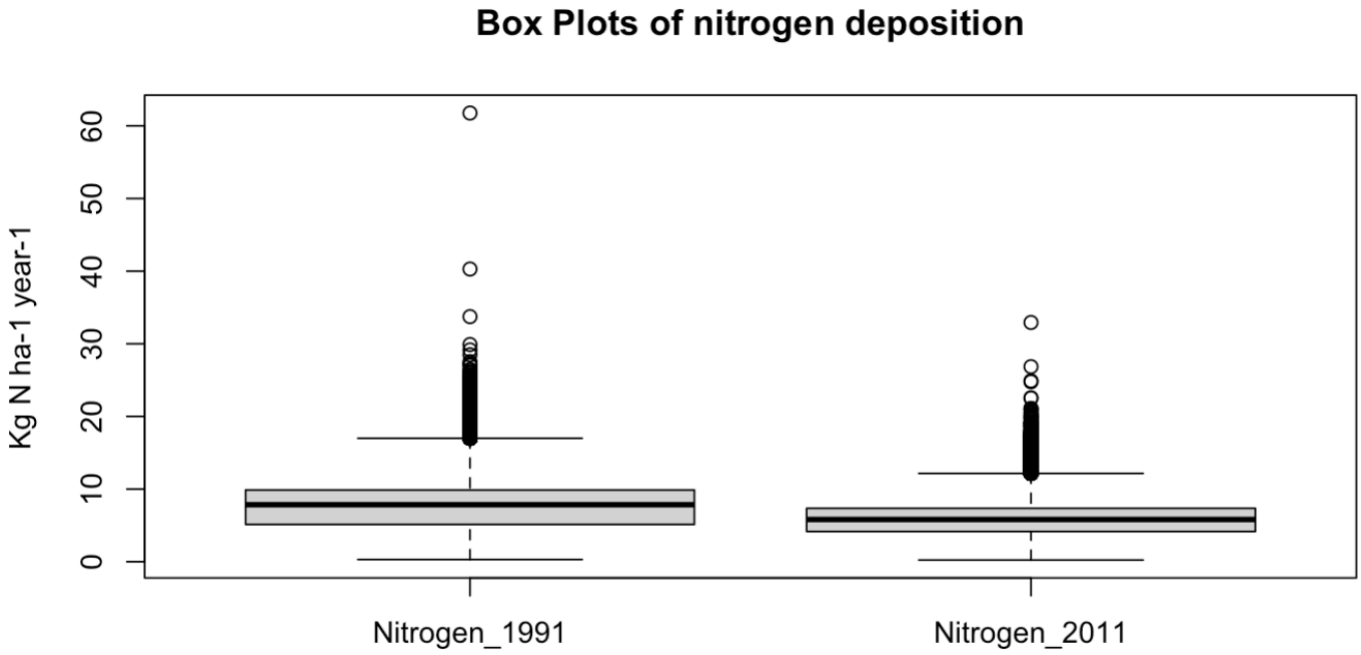

Nitrogen Deposition 2011
1e+06
5e+05
0e+00
latitude
0e+00
2e+05
4e+05
6e+05
longitude
grd_NHx_dry
20
10
0

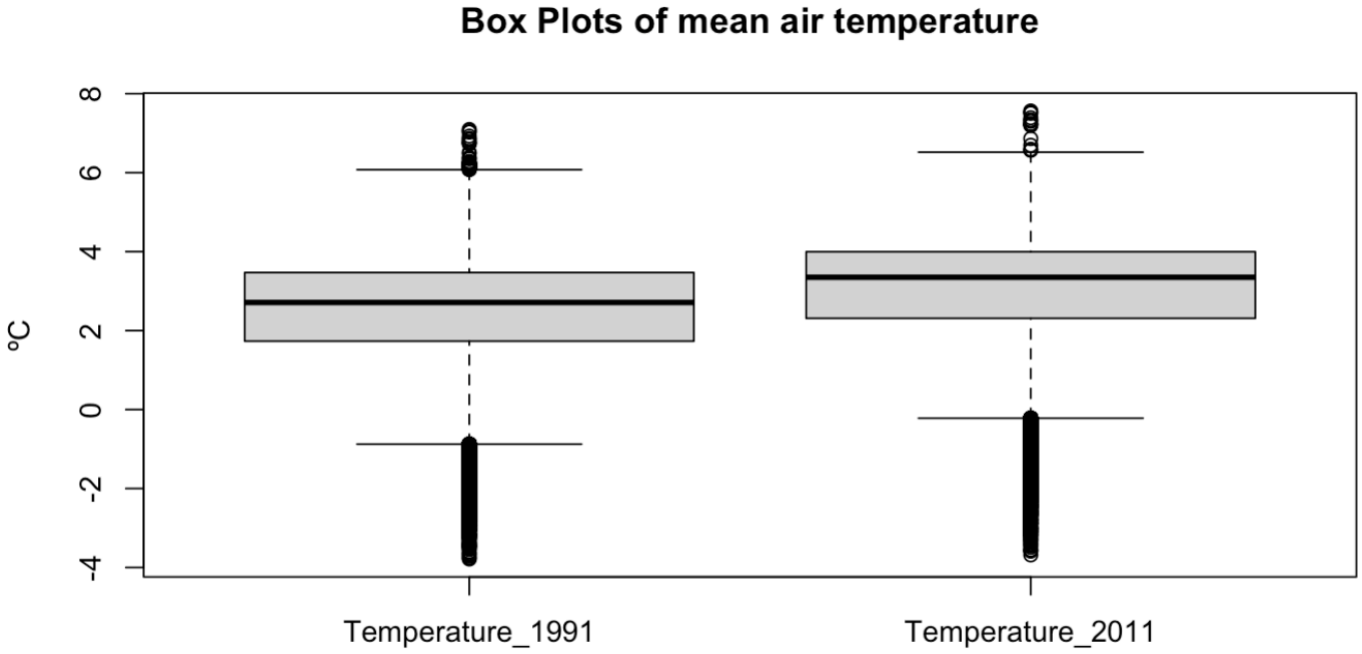

Precipitation 1991
1200000
800000
400000
0
latitude
0
250000
500000
longitude
Total.rainfall
600
400
200

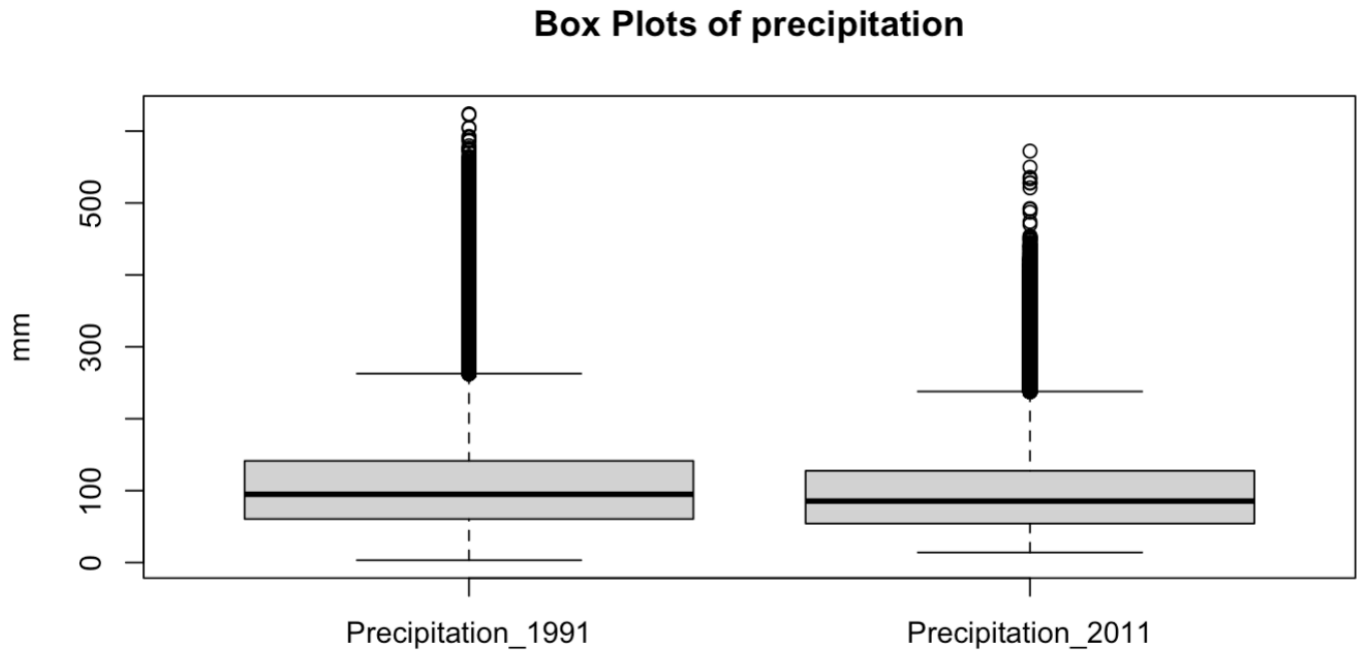

Precipitation 2011
1200000
800000
400000
0
latitude
0
250000
500000
longitude
Total.rainfall
500
400
300
200
100


Mean Air Temperature 1991
1200000
800000
400000
0
latitude
0
250000
500000
longitude
Mean.air.temperature
5.0
2.5
0.0
-2.5


Mean Air Temperature 2011
1200000
800000
400000
0
latitude
0
250000
500000
longitude
Mean.air.temperature
7.5
5.0
2.5
0.0
-2.5


Box Plots of nitrogen deposition
Kg N ha-1 year-1
0
10
20
30
40
50
60
Nitrogen_1991
Nitrogen_2011





To ensure results from the SED calculation were comparable to those from the BNI, the variables needed to be converted to OSGB grid cells and aggregated into 10km cells. The method for doing so, with an example using the data of mean air temperature in 1991, is shown below:

```
#remove NAs from cleaned variable (e.g. temperature 1991)
temp_1991_rm <- na.omit(temp_1991)

#isolate lon and lat from dfs
#rename columns to lon and lat
coords <- temp_1991_rm[, -c(3)]
colnames(coords) <- c("lon", "lat")

#create an sf object
coordinates_sf <- st_as_sf(coords, coords = c("lon", "lat"), crs = 27700)

#convert to OSGB coordinate system (EPSG:27700)
coordinates_osgb <- st_transform(coordinates_sf, 27700)

#extract Easting and Northing
easting <- st_coordinates(coordinates_osgb)[, 1]
northing <- st_coordinates(coordinates_osgb)[, 2]

#calculate grid cell indices (adjust grid size as needed)
grid_size <- 1000  # 1km grid size (in m)
grid_cell_x <- floor(easting / grid_size)
grid_cell_y <- floor(northing / grid_size)

#create grid cell IDs
grid_cell_ids <- paste0("Grid_", grid_cell_x, "_", grid_cell_y)

#add grid cell IDs to the data frame
coords$grid_cell <- grid_cell_ids

#create grid cell boundaries (optional)
grid_cell_boundaries <- st_make_grid(coordinates_osgb, cellsize = grid_size)

#create a data frame for grid cell boundaries
grid_df <- data.frame(
  xmin = st_bbox(grid_cell_boundaries)[1],
  xmax = st_bbox(grid_cell_boundaries)[3],
  ymin = st_bbox(grid_cell_boundaries)[2],
  ymax = st_bbox(grid_cell_boundaries)[4]
)

#align values from original df to new grid cells
temp_1991_grid <- cbind(temp_1991_rm, coordinates_osgb)
```

3.2.2 Creating a Function to Calculate SED

Here is an R Script which contains a function to calculate the standardised Euclidean distance according the formula from Radeloff et al. (2015); accompanied with dummy data sets which can be used to run the function as an example.

```
# SED equation from Radeloff et al., (2015) ----
# paper accessible via: https://onlinelibrary.wiley.com/doi/abs/10.1890/14-1781.1

#       SED = ∑√((bki-aki)^2/skt)

#INPUTS (format of grid cell x value, all need to be same length)
#       dtr - temperature
#       ndep - nitrogen deposition
#       precip - precipitation
#OUTPUTS (x = name of variable)
#       sd_x - standard deviation of variable x at time point a (one value)
#       x_calc - results from (bki-aki)^2
#       x_div - results from x_calc/sd_x
#       sum_all - sum of all variables per grid cell (need to combine all results from x_div into one df)
#       SED - square root of sum_all to give SED of each grid cell


dtr <- data.frame(
  grid = c("a", "b", "c", "d", "e", "f"),
  bki = c(12,11,14,10, 11,13),
  aki = c(8,9,10,6,5,7))

precip <- data.frame(
  grid = c("a", "b", "c", "d", "e", "f"),
  bki = c(1750,1500,4000,3000,2500,1500),
  aki = c(1000,1500,1250,1500,800,700))

ndep <- data.frame(
  grid = c("a", "b", "c", "d", "e", "f"),
  bki = c(20,19,12,17,11,10),
  aki = c(10,15,11,16,9,13))
```

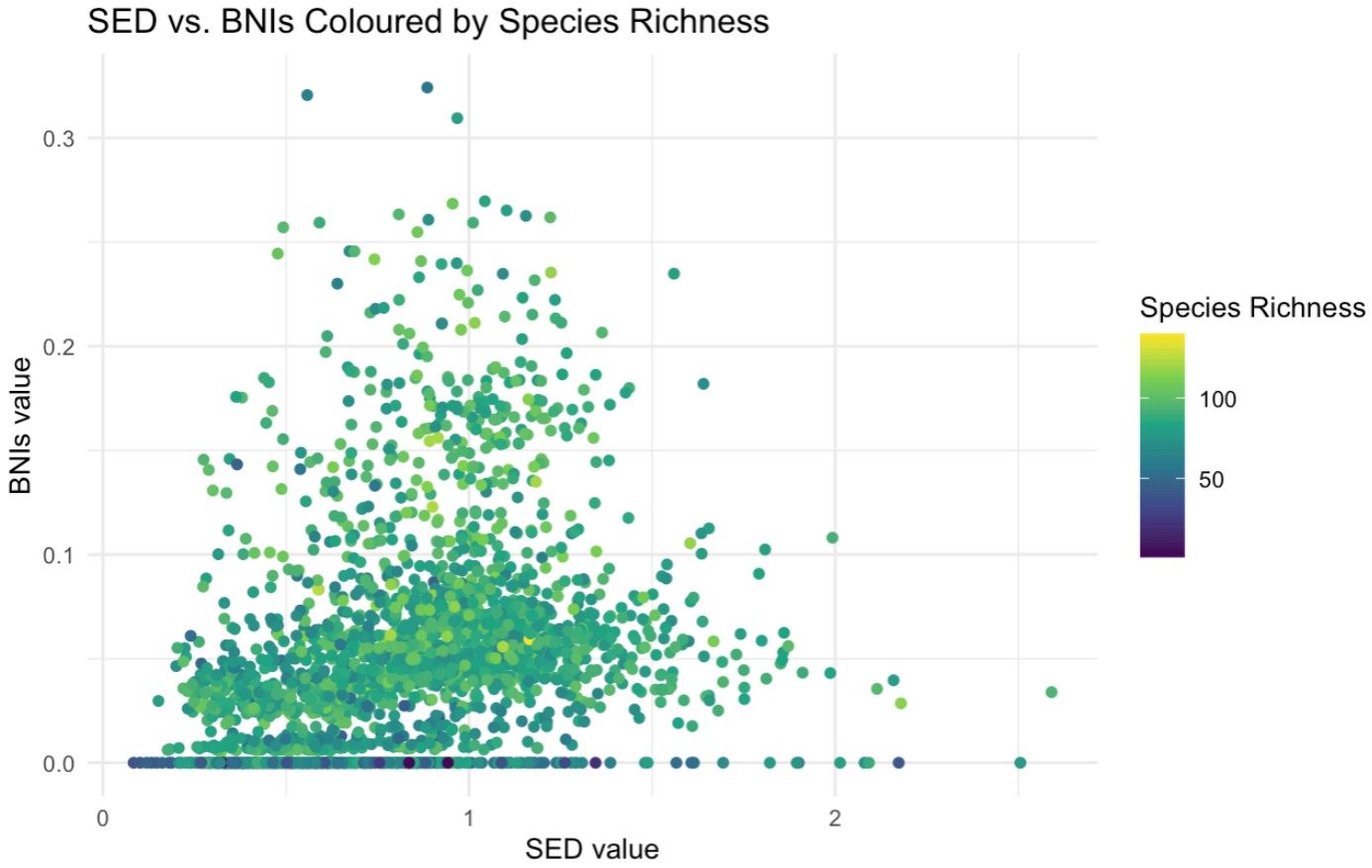
Example Data Frame Structure (Nitrogen Deposition)

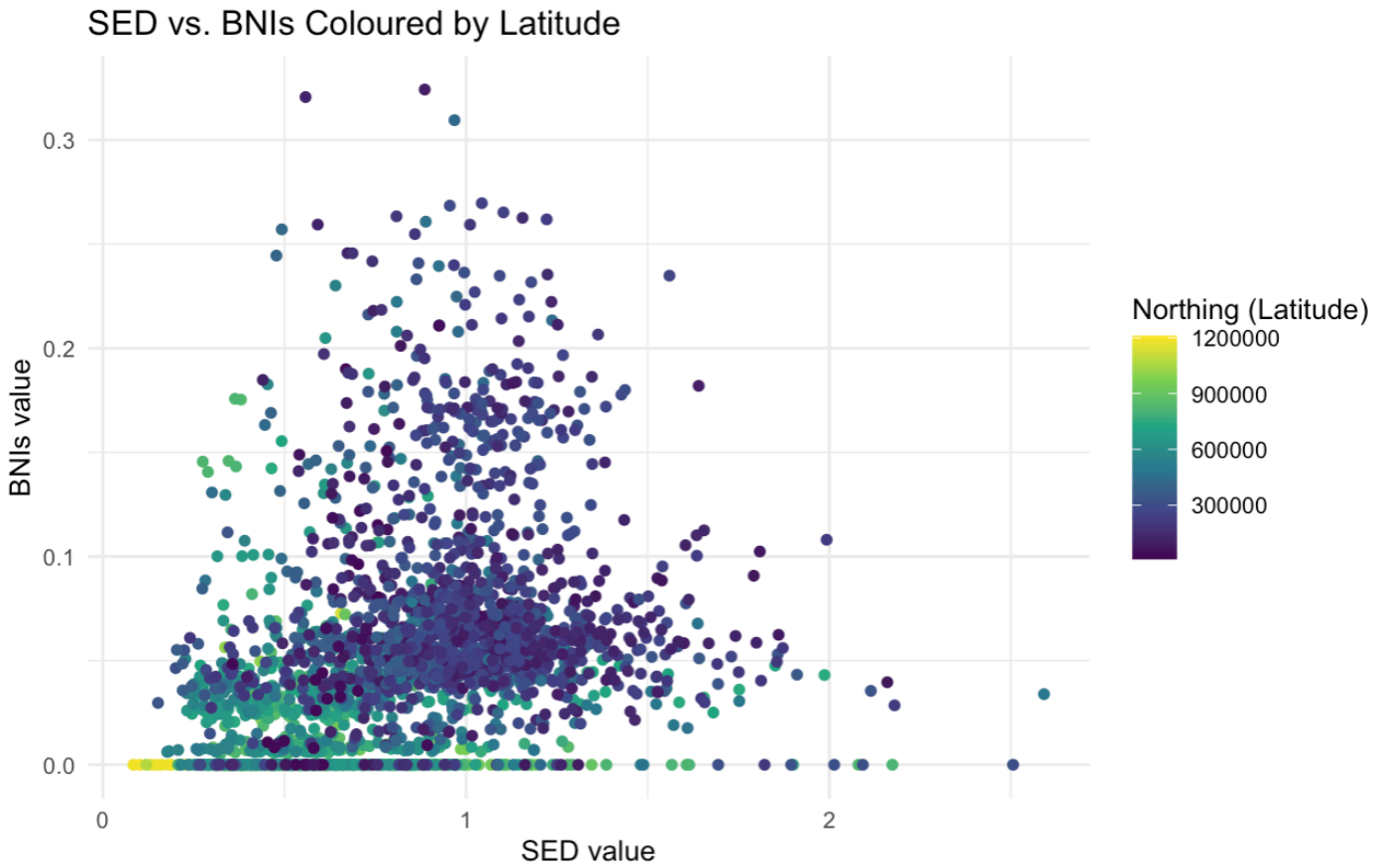
grid	bki	aki
a	20	10
b	19	15
c	12	11
d	17	16
e	11	9
f	10	13

```
SED_calc <- function(SED){

#calculate standard deviation for each variable at the baseline time point
sd_dtr <- sd(dtr$aki)
sd_ndep <- sd(ndep$aki)
sd_precip <- sd(precip$aki)

#calculate (bki-aki)^2 for all
dtr_calc <- (dtr$bki - dtr$aki)^2
precip_calc <- (precip$bki - precip$aki)^2
ndep_calc <- (ndep$bki - ndep$aki)^2

#divide previous result over sd
dtr_div <- dtr_calc/sd_dtr
precip_div <- precip_calc/sd_precip
ndep_div <- ndep_calc/sd_ndep

#combine datasets
combined_calcs <- cbind(dtr_div,precip_div,ndep_div)

#sum of results by grid cell
sum_all <- rowSums(combined_calcs)

#square root to get SED
SED <- sqrt(sum_all)

#join to grid cell

SED_grid <- cbind(dtr, SED)
}

results<-SED_calc(SED)
```

3.3 Comparing the Results of the BNI and SED Methdologies

Following the scatter plot showing the correlation of BNIs values against SED, an attempt to discover any underlying patterns in the distribution of BNIs and SED values was made using the same scatter graph but coloured according to the species richness and latitude of each grid cell. Species richness was taken from the calculation of the BNI identified in the community matrix (com), and latitude was extracted from the northing values provided from the spatial data used to map the SED results.





4.1 Dissimilarity Metric

4.1.1 Methods

The mean pairwise trait distance for each grid cell was calculated by applying a function to the matrix of presences and absences created in the previous analysis. The function is shown below, applied to the community and distance matrix.

```
#calculate mean pairwise functional distance for a grid cell
calculate_mean_fdist <- function(pa_row, dist_matrix) {
  present_species <- names(pa_row)[pa_row == 1]

  #if only one or zero species are present, the mean pairwise distance is undefined (or 0)
  if (length(present_species) < 2) {
    return(NA) # Or 0, depending on your ecological interpretation
  }

  #subset the distance matrix for the present species
  sub_matrix <- dist_matrix[present_species, present_species]

  #extract the lower triangle of the sub-matrix (excluding the diagonal)
  lower_triangle_values <- sub_matrix[lower.tri(sub_matrix)]

  #calculate the mean of these values
  mean_distance <- mean(lower_triangle_values)
  return(mean_distance)
}
```

Values for mean pairwise trait distance and species richness for both Atlas periods were merged, creating a data frame ready for the calculations to be carried out:

Example Data Frame Structure

grid	sp_2	sp_3	fun_2	fun_3
HP40	36	33	0.0560791	0.0652170
HP50	49	48	0.0712676	0.0961043
HP51	43	35	0.0665660	0.0780528
HP60	44	50	0.0651198	0.0615158
HP61	54	56	0.0964794	0.0662533
HT93	48	54	0.0631889	0.0695797

The calculations consisted of subtracting values from the second period from those in the third, followed by a scale and centre to bring all features to a similar range, so that each variable contributes more equally to the output. Example code is shown below:

```
difference_calc <- diss_example %>%
  mutate(difference_sp = sp_3 - sp_2)
```

4.2 PCA Metric

4.2.1 Methods

For a PCA to be carried out, the values for each variable must be converted into long format with a column for grid cells at atlas period 2 and 3 and their corresponding variable values. This was performed with ‘pivot_longer()’ from the tidyr package (v1.3.0; Wickham et al., 2023), for data to appear as in the example below. These data were then scaled and centered before performing the PCA. This resulting data frame will be used for the PCA, with a separate PCA being carried out for the biotic and abiotic variables. In the table below “sp” represents species richness, “frich” functional richness and “fun” is mean pairwise functional trait distance.

Example Data Frame Structure

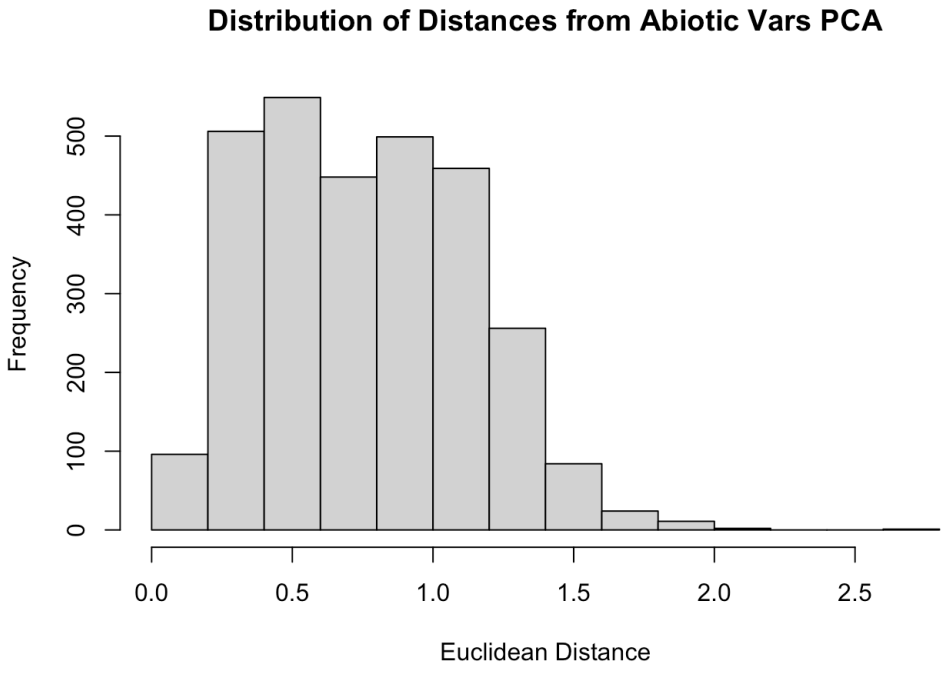
grid	sp	frich	fun
HP40_2	36	0.2113078	0.0560791
HP40_3	33	0.2151196	0.0652170
HP50_2	49	0.2865402	0.0712676
HP50_3	48	0.2697235	0.0961043
HP51_2	43	0.2764588	0.0665660
HP51_3	35	0.2236087	0.0780528
HP60_2	44	0.2224890	0.0651198
HP60_3	50	0.2446262	0.0615158

grid	sp	frich	fun
HP61_2	54	0.3376343	0.0964794
HP61_3	56	0.2850461	0.0662533

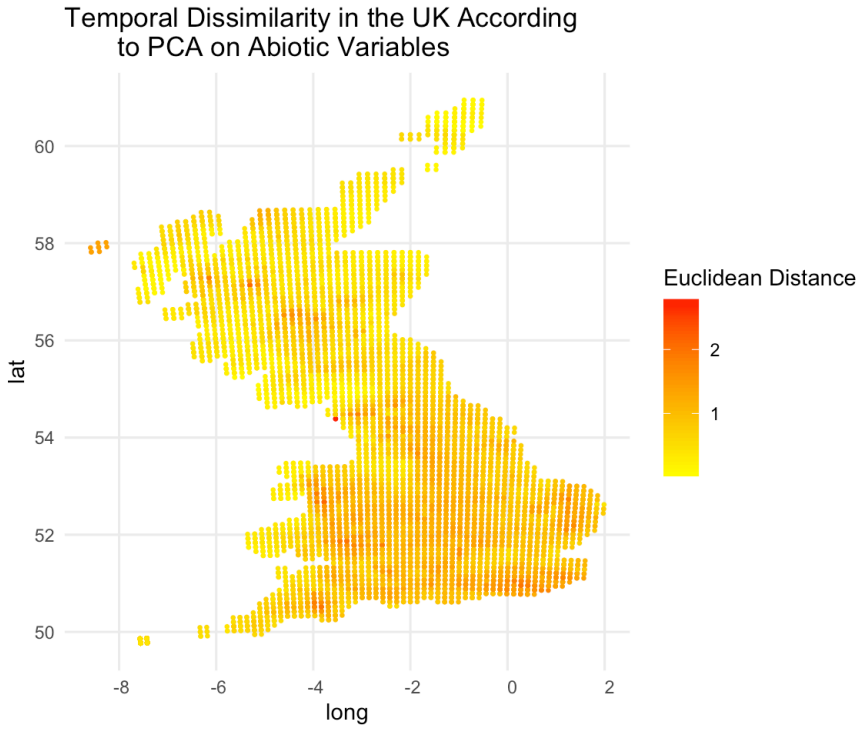
4.2.2 Results

With the PCA results, Euclidean distance can then be calculated between the grid cell's position in PCA space at period 2 and its position at period 3 to get a measure of temporal dissimilarity. This is completed by separating the data by time point to get the position of each grid cell within the first principal component at periods 2 and 3 from the PCA data frame, with an example workflow shown below:

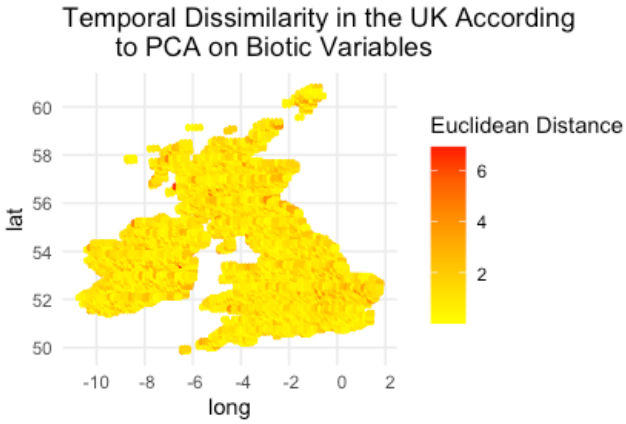
```
#separate data by time point
pca_2_abio <- pca_df_abio %>%
  filter(grepl("_2$", Grid)) %>%
  rename(PC1_2 = PC1, PC2_2 = PC2) %>%
  mutate(base_grid = gsub("_2$", "", Grid)) %>%
  select(base_grid, PC1_2, PC2_2)

pca_3_abio <- pca_df_abio %>%
  filter(grepl("_3$", Grid)) %>%
  rename(PC1_3 = PC1, PC2_3 = PC2) %>%
  mutate(base_grid = gsub("_3$", "", Grid)) %>%
  select(base_grid, PC1_3, PC2_3)

#join PCA results from period 2 and 3
merged_pca_abio <- inner_join(pca_2_abio, pca_3_abio, by = "base_grid")

#calculate Euclidean distance
distance_df_abio <- merged_pca_abio %>%
  mutate(distance = sqrt((PC1_2 - PC1_3)^2 + (PC2_2 - PC2_3)^2))


#histogram of differences (temporal distance measured between the same grid cell at each time point)
hist(distance_df_abio$distance,
     main = "Distribution of Distances from Abiotic Vars PCA",
     xlab = "Euclidean Distance")
```

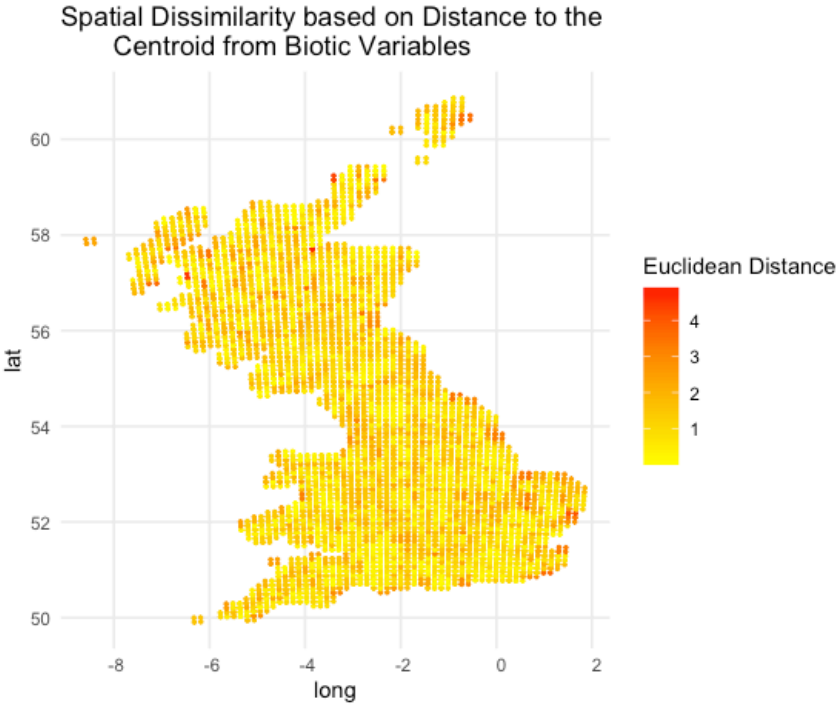


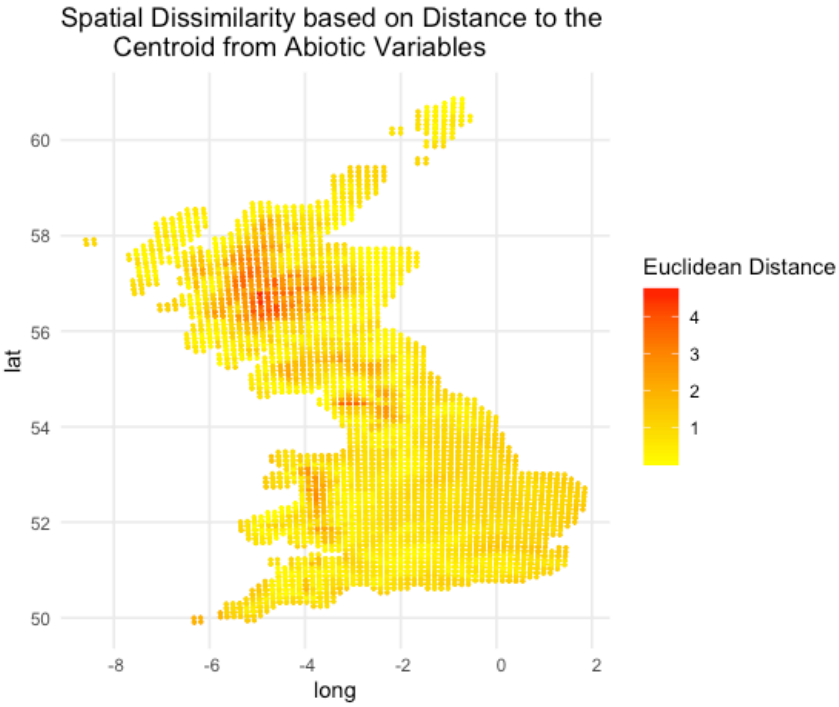
```
#join to 'grids' file downloaded from the BTO to map
names(distance_df_abio)[1] <- "grid"
grid_PCA_dist_temp <- merge(distance_df_abio, grids, by = "grid", miltiple = "all")

#map
ggplot(grid_PCA_dist_temp, aes(x = long, y = lat, colour = distance)) +
  geom_point(size = 0.5) + # Adjust point size as needed
  scale_color_continuous(low = "yellow", high = "red") +
  labs(title = "Temporal Dissimilarity in the UK According
       to PCA on Abiotic Variables",
       colour = "Euclidean Distance") +
  coord_sf() +
  theme_minimal()
```



This process was repeated for the biotic PCA and then using the biotic and abiotic PCA to visualise spatial novelty from the centroid, shown in the figures below. To calculate distance to the centroid, first the average value for all grid cells needs to be calculated for the abiotic and biotic PCA separately. Secondly, the distance of all grid cells to that average value is then calculated to get a novelty measure. It is these values that are then mapped using ‘ggplot2’ (Wickham, 2016).


Temporal Dissimilarity in the UK According to PCA on Biotic Variables
Euclidean Distance
6
4
2
lat
long
60
58
56
54
52
50
-10
-8
-6
-4
-2
0
2


Spatial Dissimilarity based on Distance to the Centroid from Biotic Variables
Euclidean Distance
4
3
2
1
lat
long
60
58
56
54
52
50
-8
-6
-4
-2
0
2


Spatial Dissimilarity based on Distance to the Centroid from Abiotic Variables
Euclidean Distance
4
3
2
1
lat
long
60
58
56
54
52
50
-8
-6
-4
-2
0
2