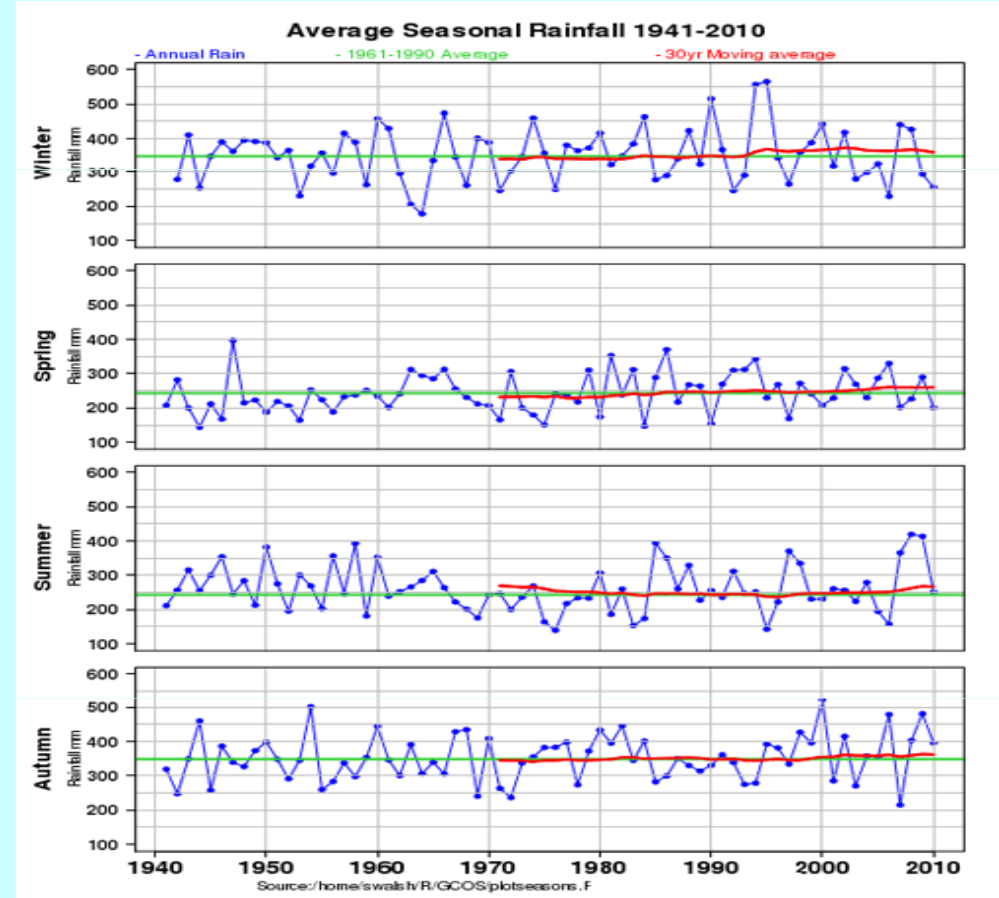


# Homogenising Ireland's monthly precipitation records



Source: Met Éireann



John Coll



# Structure of presentation

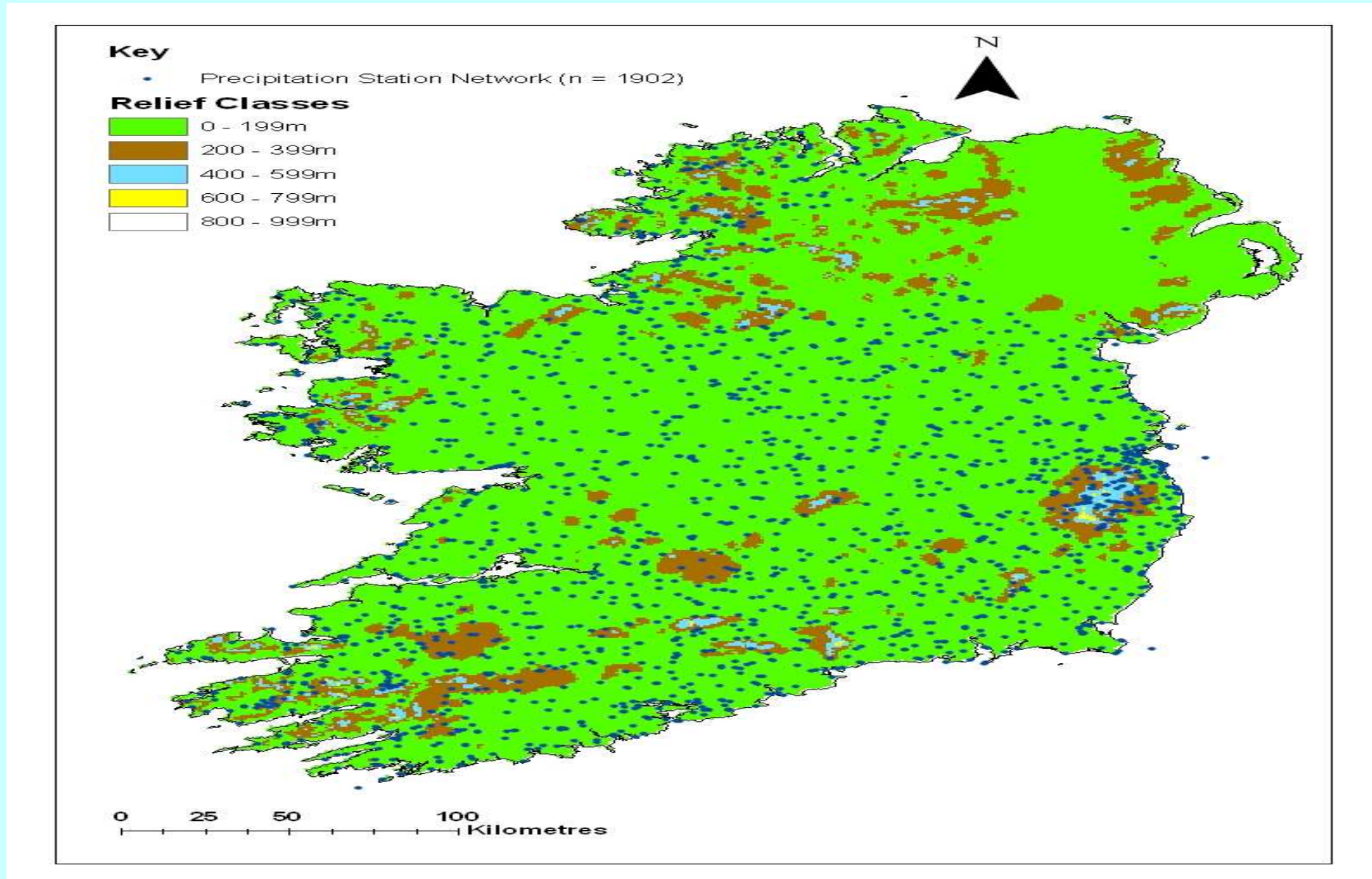
- Motivation
- The station network & auditing the series
- Statistical and spatial exploration of series data
- Application of HOME-R – initial set of stations (n = 88)

# EPA funding motivation

- Extends the work carried out by Met Éireann staff as part of their contribution to the COST Action HOME. The output (HOME-R) used to process Irish data & identify break points & secular trends due to non climatic causes
- Removing non climate influences permits more authoritative statements to be made on the extent to which Irish climate is changing in accordance with global trends
- Assists Met Éireann's contribution to the European effort in climate data homogenisation by enhancing capacity
- Enables improved collaboration between Met Éireann & university-based research to be achieved at a time of reduced resources for both

# The precipitation station network database

=> Variable electronic record length, 1941 – 2014 (but some from 1800s)



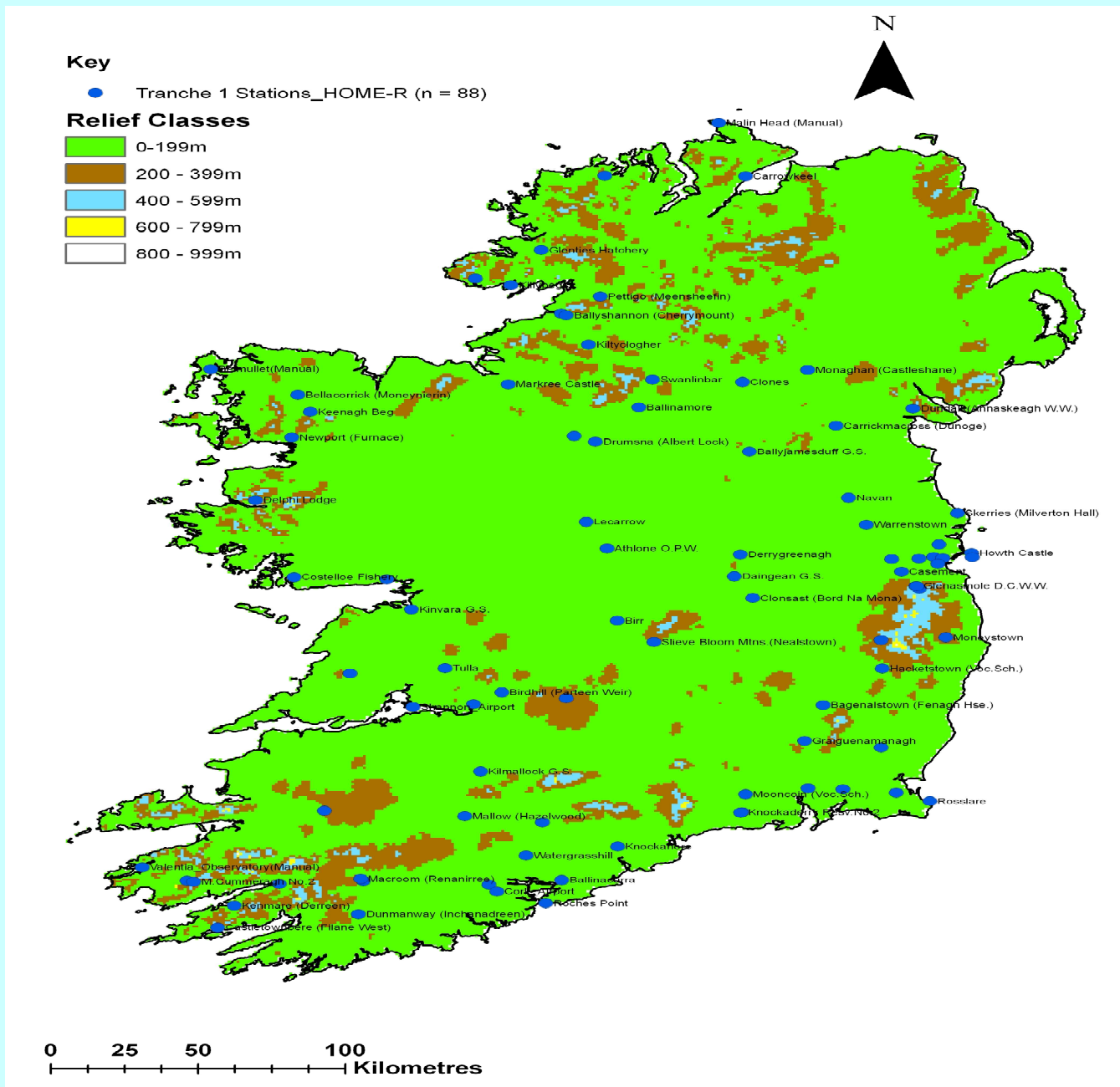
# Preliminary work

1. Assembling the R code libraries & capacity – data management & handling
2. Defining the ~‘scale of the task’. Audit in RStudio, R & spreadsheets
3. Exploring the statistical properties of the series

## Current phase

1. Assembling metadata - longer intact record stations (based on audit)
- 2. Initialising HOME-R and assessing results**
3. Repeating down through other station record categories  
⇒ Iterate & modify – ‘learning by doing’

# HOME-R: Test application I



- Station coverage - sparse in places

- Record length - variably intact 1941-2014

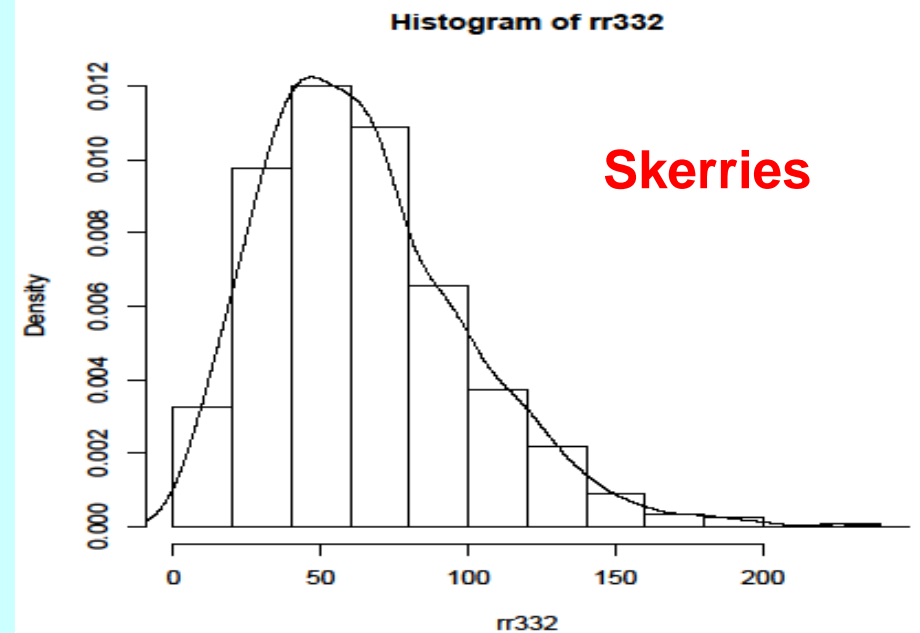
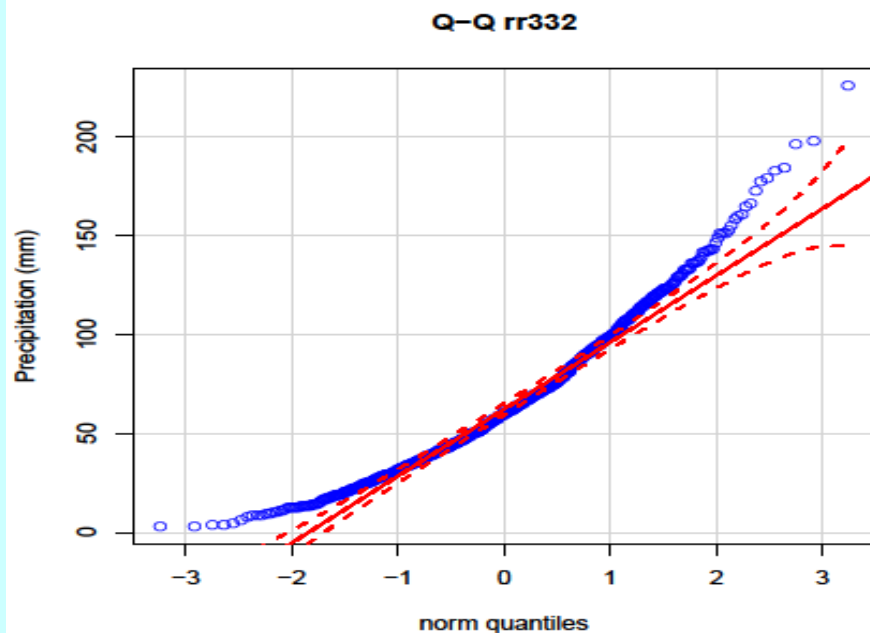
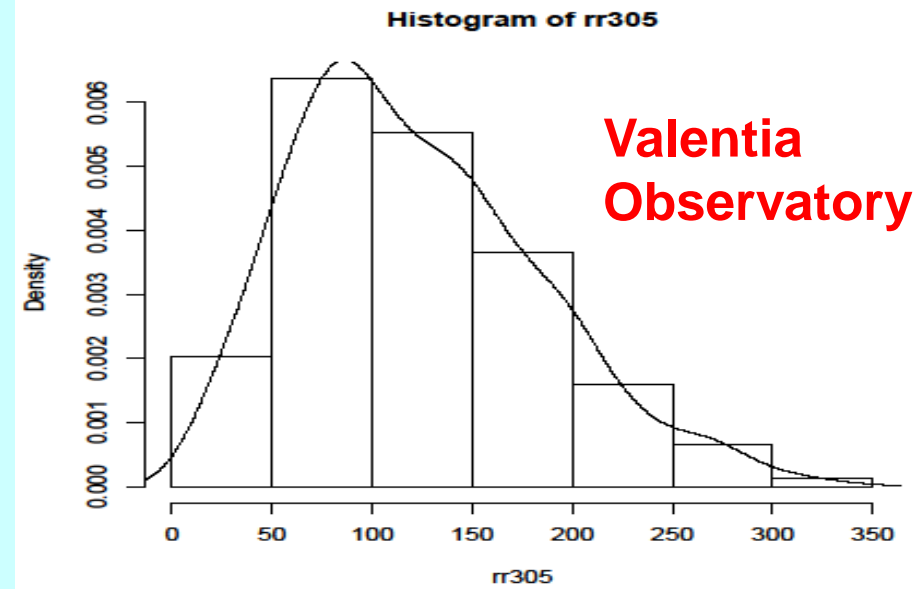
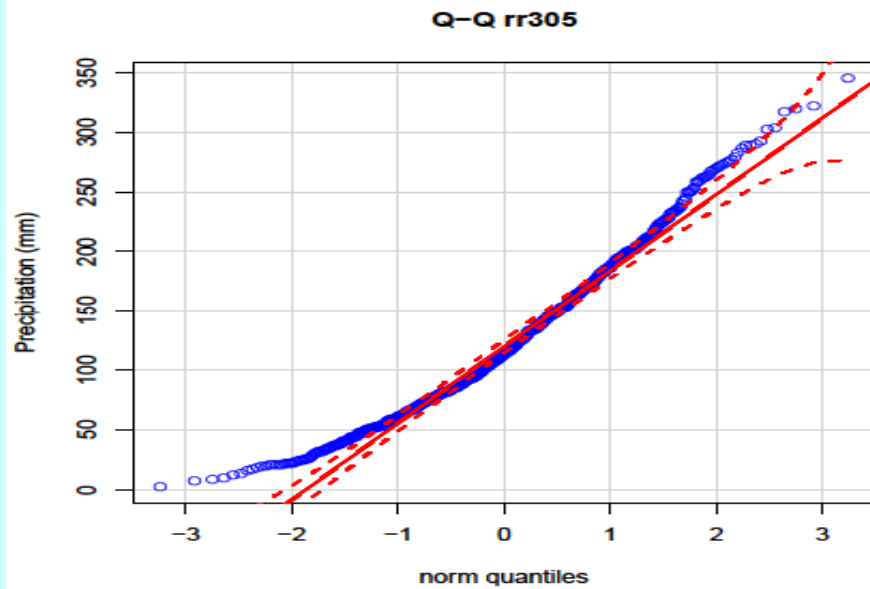
- Lot of missing entries ~1940s – mid-1960s & ~mid-1980s – 2000s

# Modified Exploratory Data Analysis (EDA)

- Clarify the objectives of the investigation
- Investigate the structure & quality of the data
- Carry out EDA (descriptive statistics)
- Compare findings with previous results
- Look for troublesome series & any that may warrant special analysis
- Search for outliers, missing data etc. that may affect the analysis
- See if transformations or robust statistics are necessary

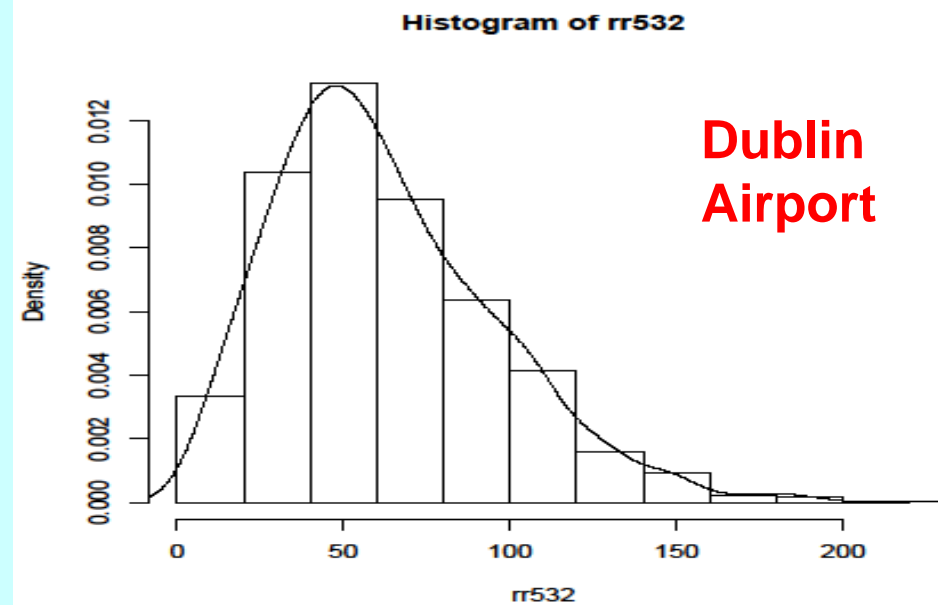
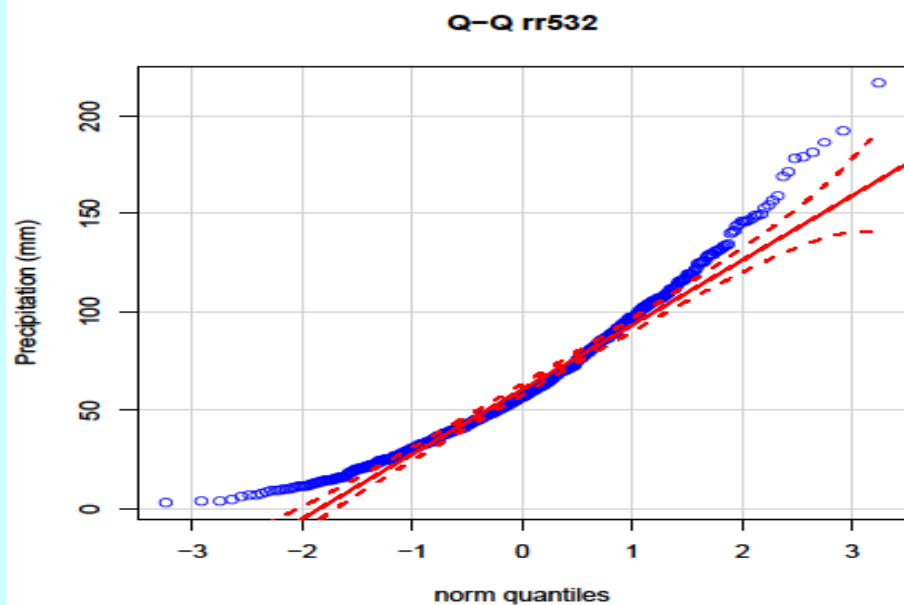
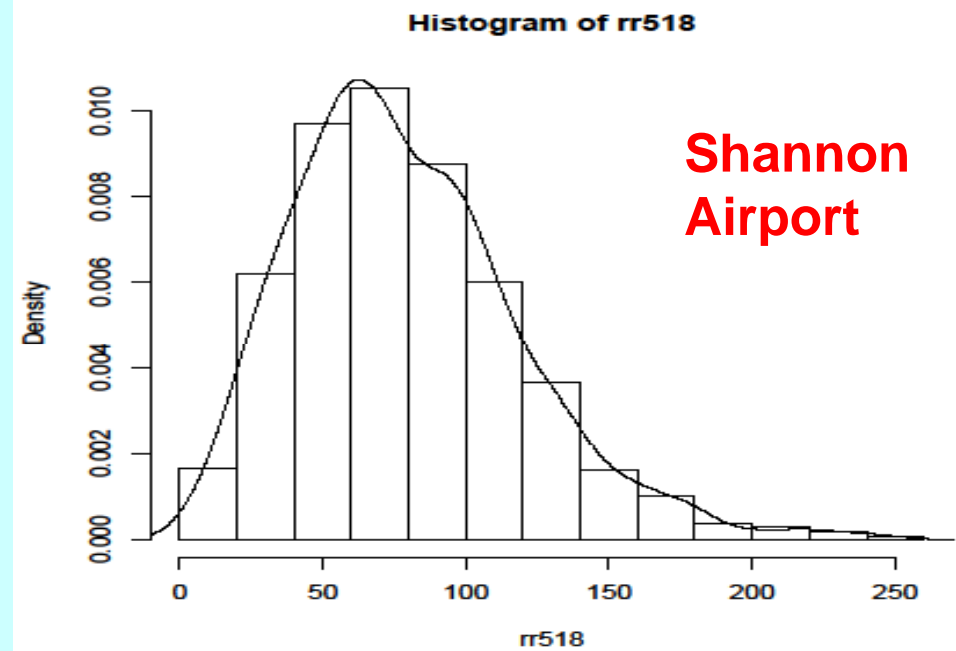
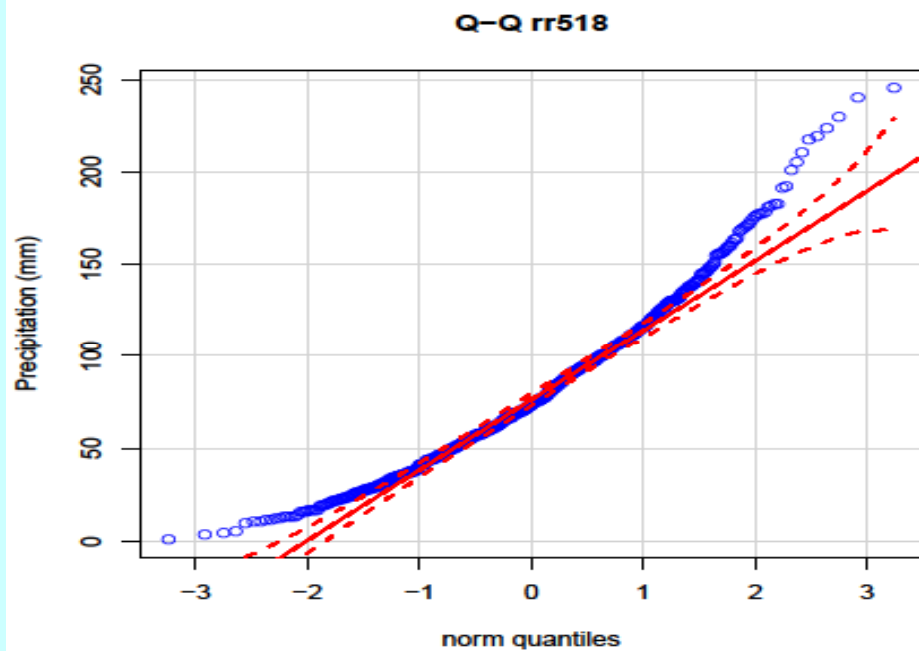
**=> Flexible in approach - treating each series & their inter-relationships uniquely**

# Individual station data structure

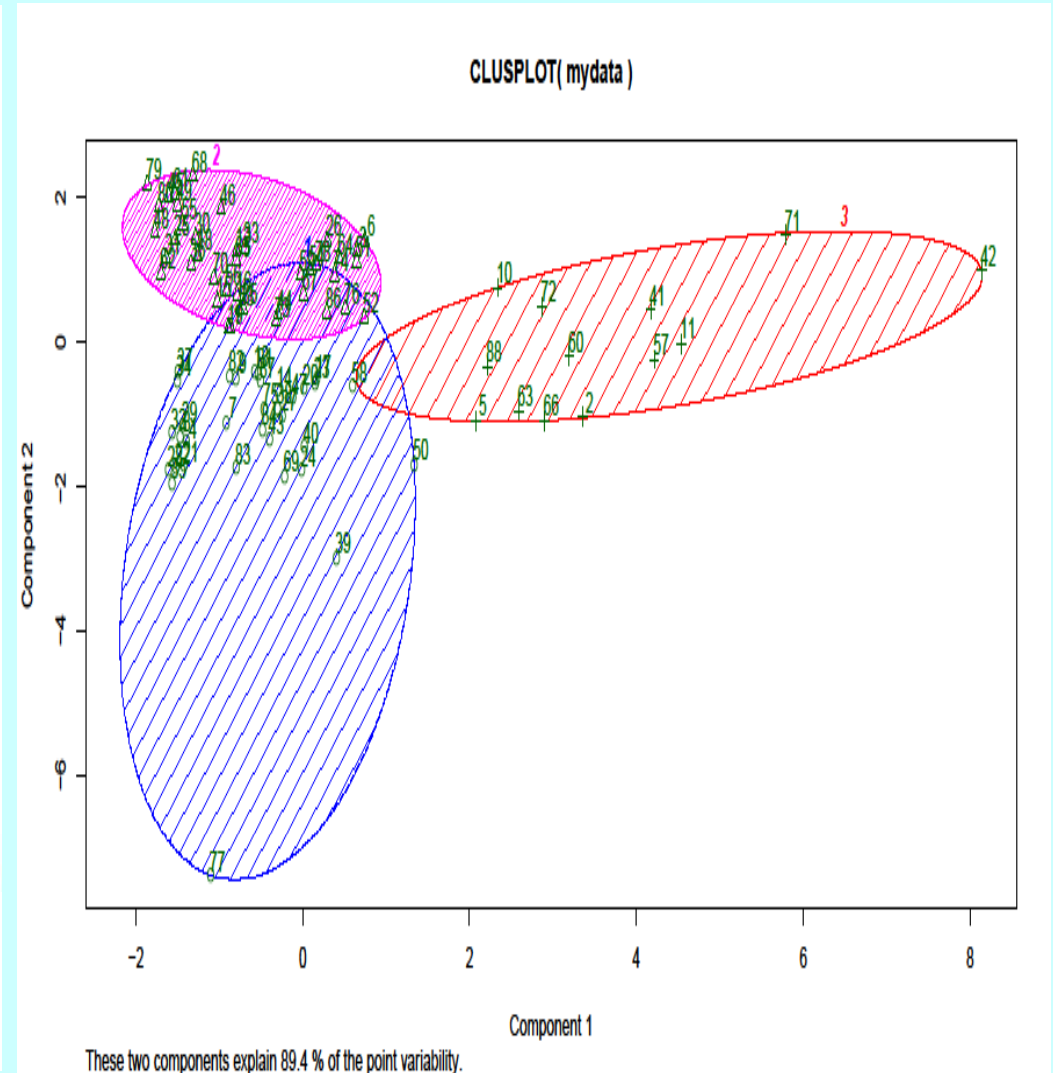
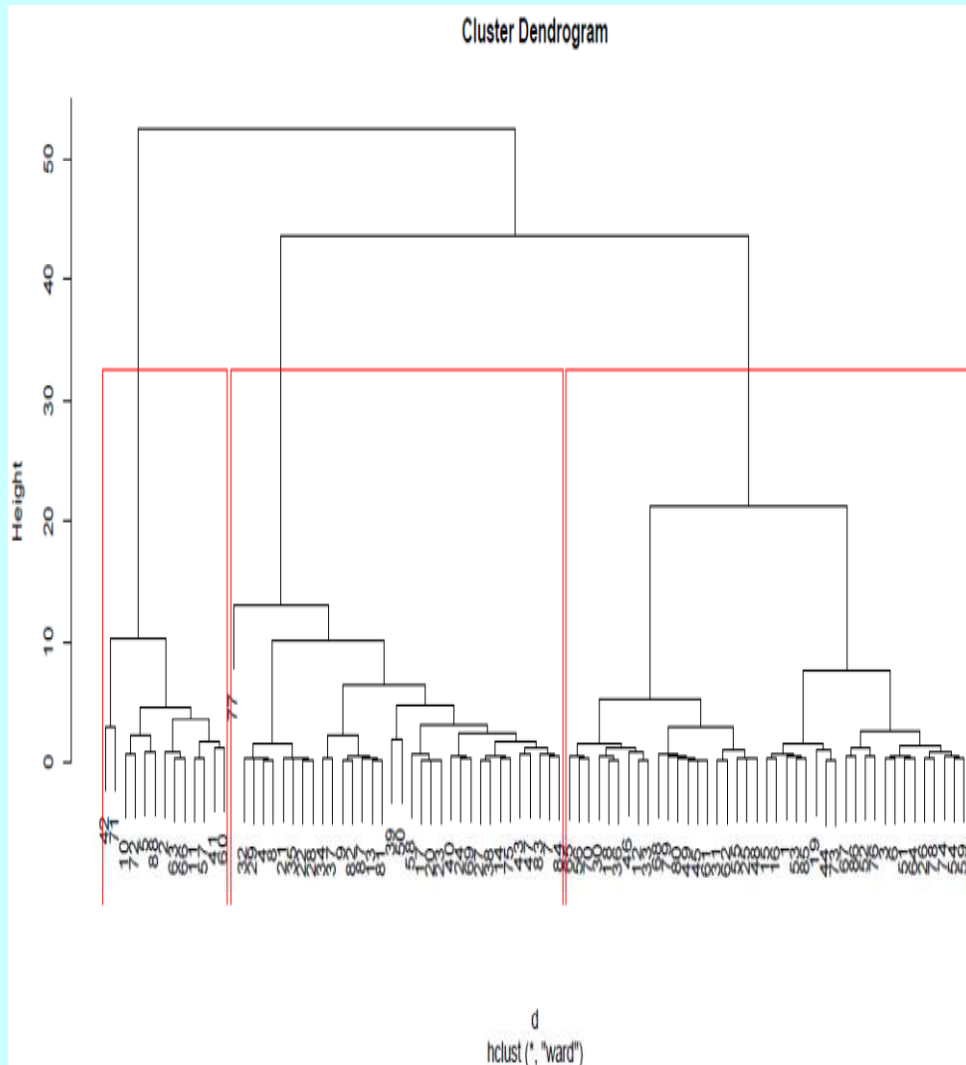




# Individual station data structure



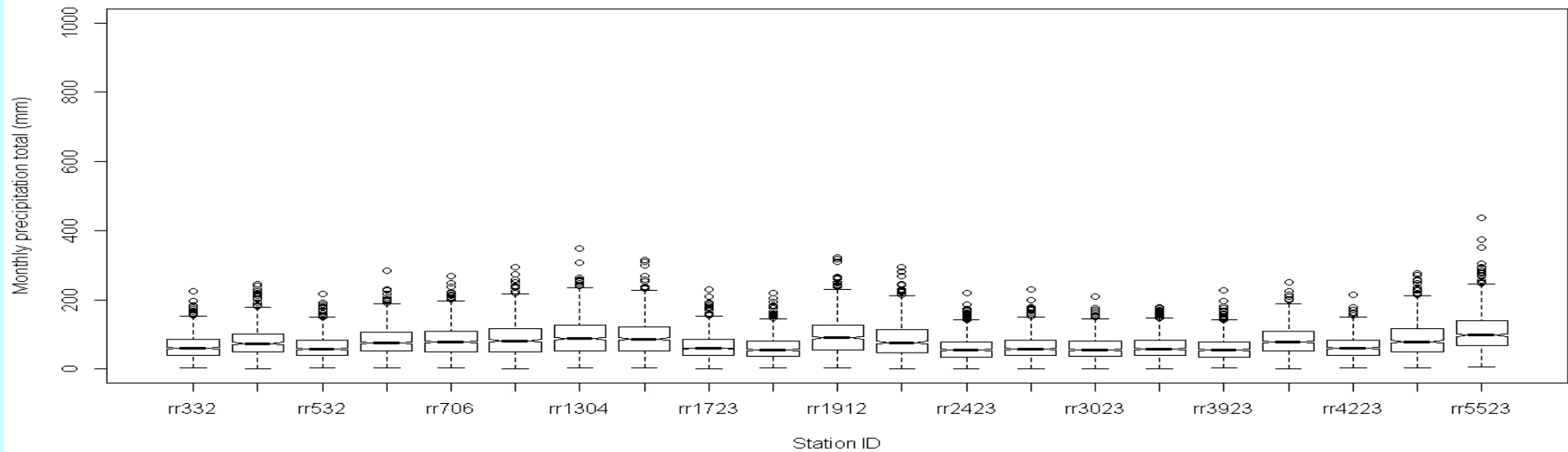
# K-means cluster visual (station n = 88)



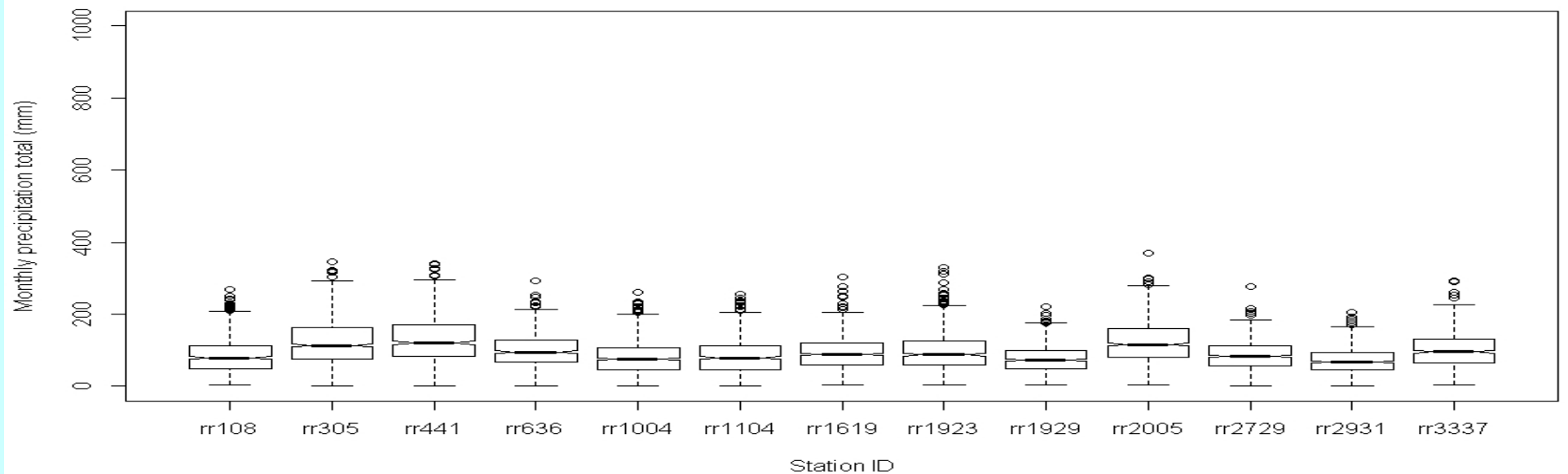
**Based on:** Mean, Variance, Kurtosis, Skewness, Range, Min, Max (from Descriptive Statistics)

# K-means cluster boxplots (1941-2010 subset)

K-means cluster 1 Boxplots (1941-2010 sub-set)

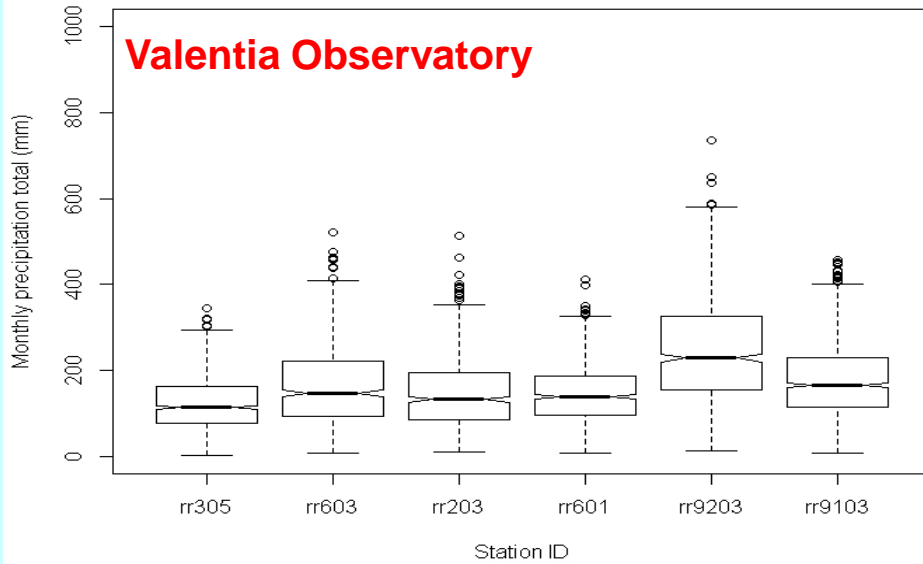


K-means cluster 3 Boxplots (1941-2010 sub-set)

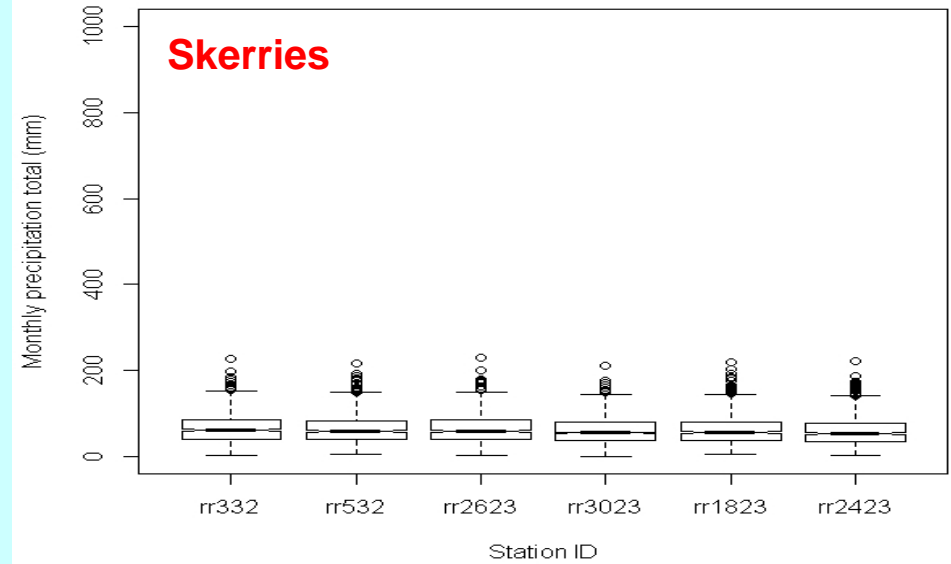


# Correlation group boxplots

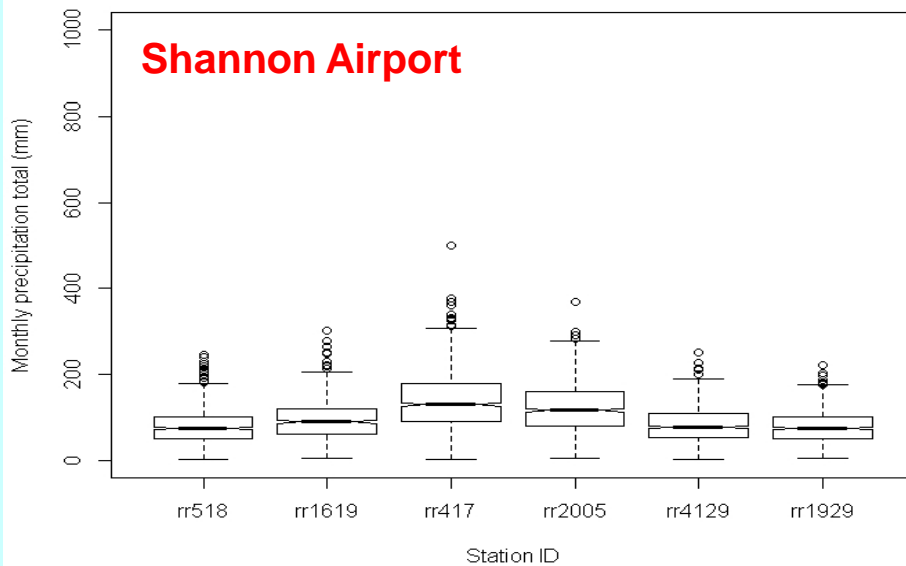
1941-2010: Full Series Boxplot + 5 Most Correlated Stations



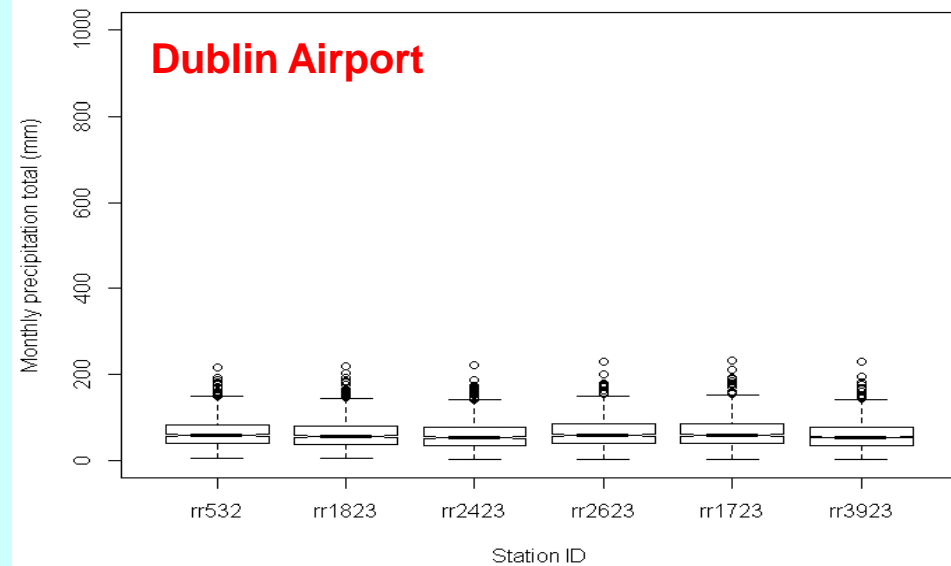
1941-2010: Full Series Boxplot + 5 Most Correlated Stations



1941-2010: Full Series Boxplot + 5 Most Correlated Stations

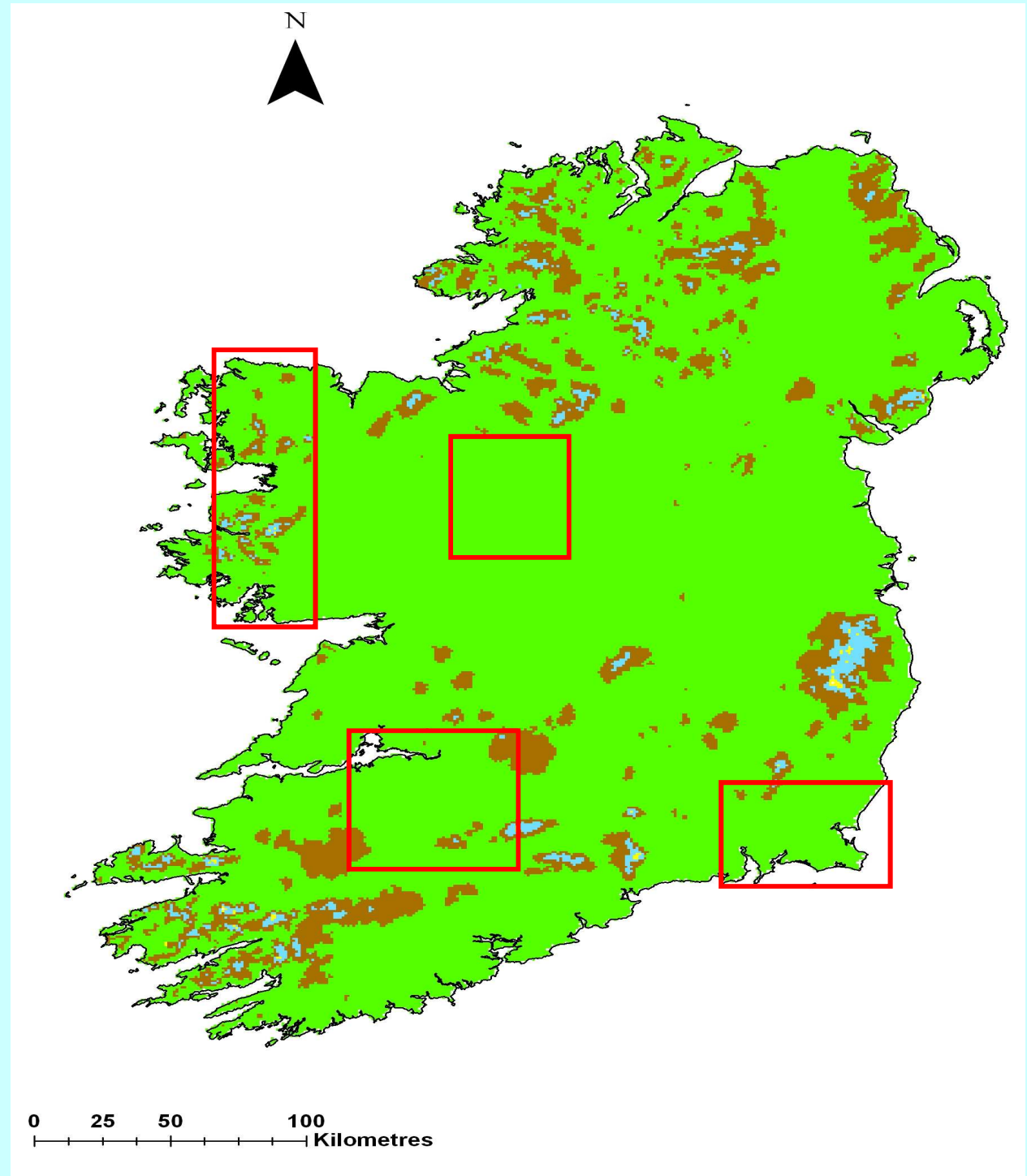


1941-2010: Full Series Boxplot + 5 Most Correlated Stations

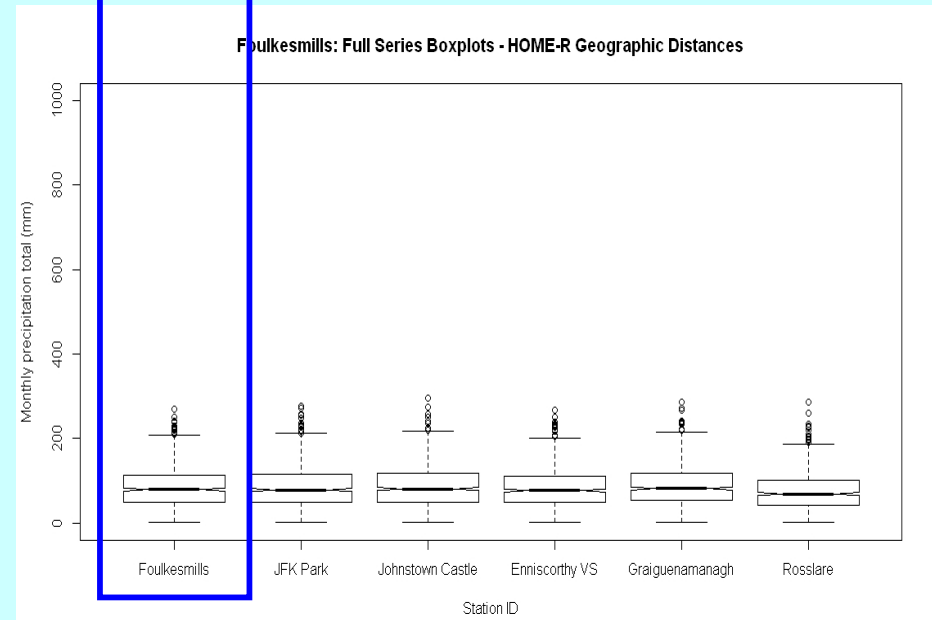
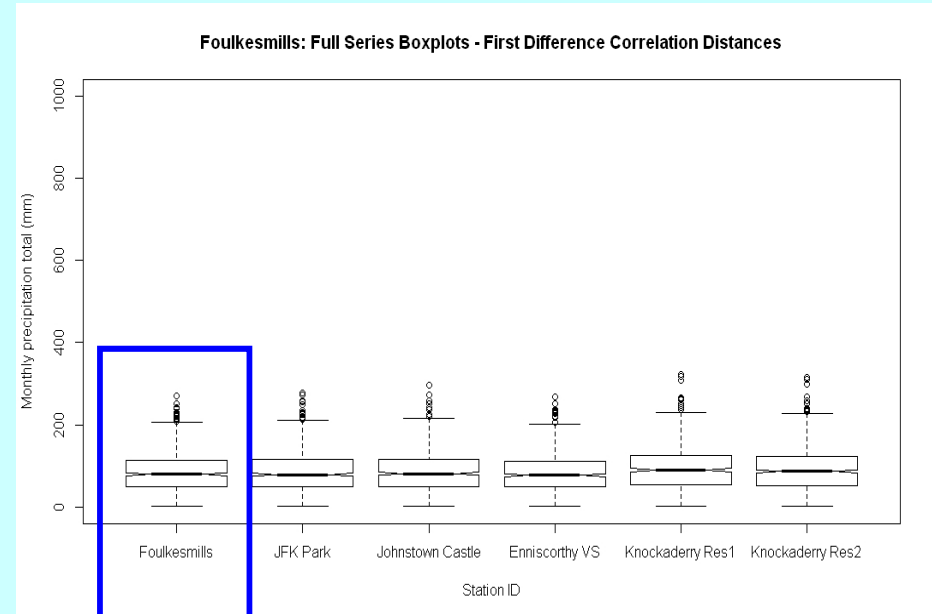
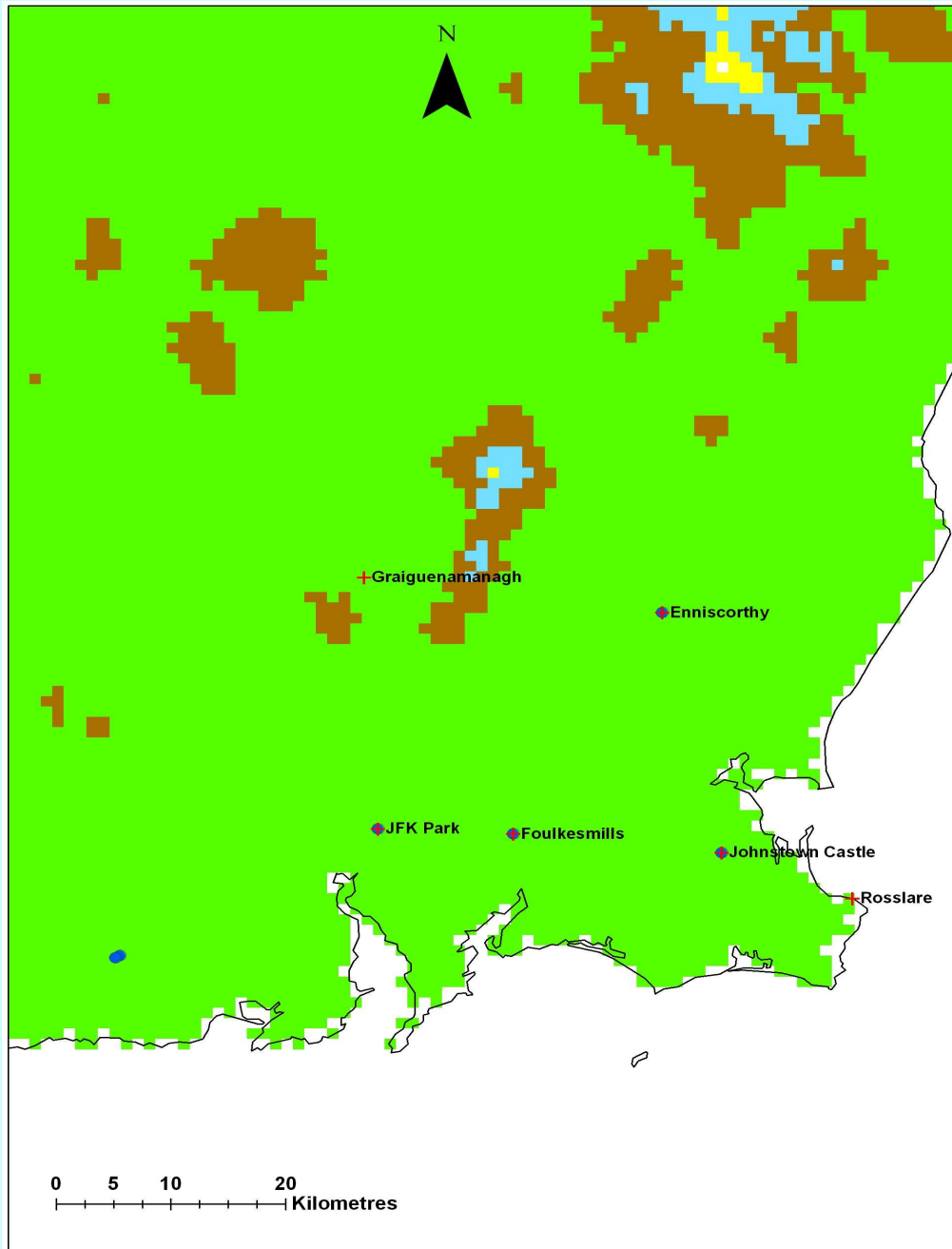


# Results (HOME-R)

Initialisation – 4 case studies from the first analysis (1941 – 2010) for x 88 station series

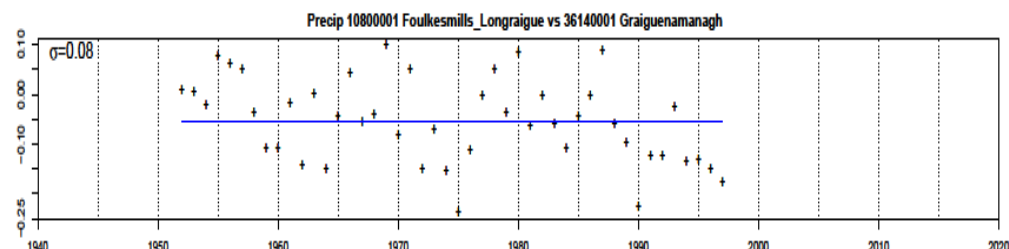
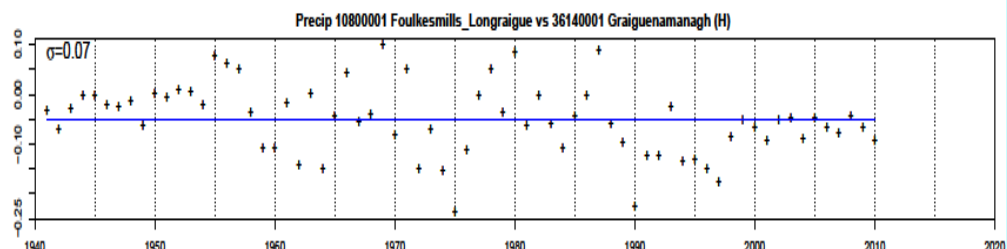
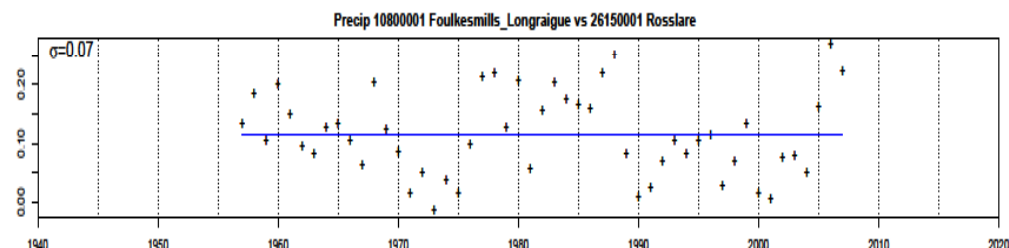
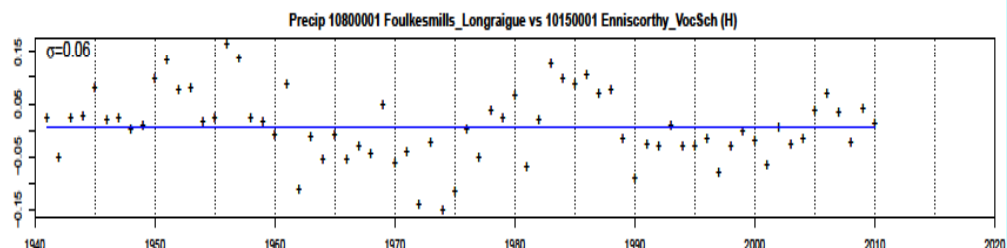
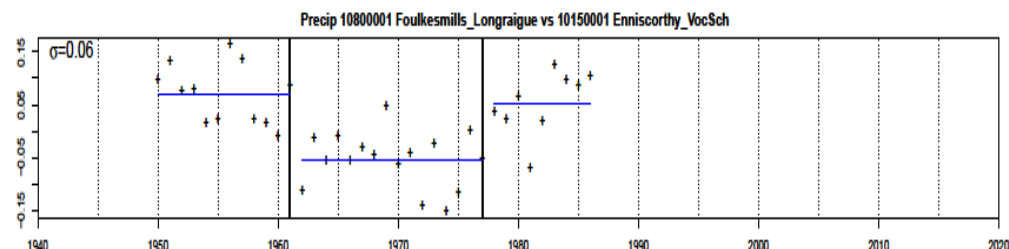
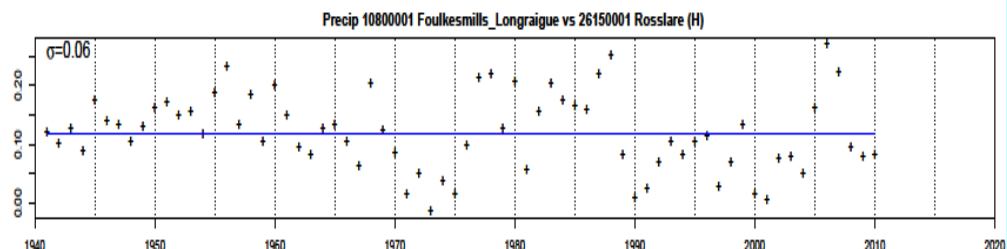
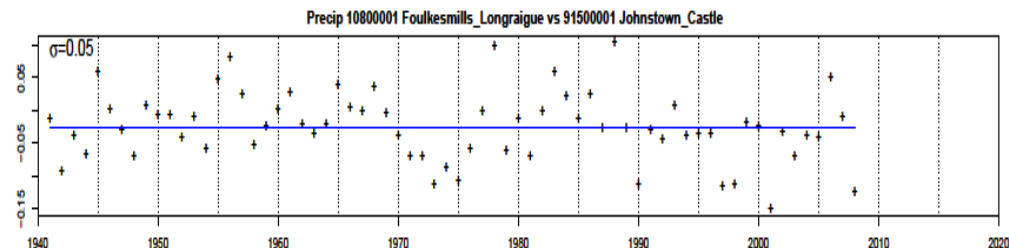
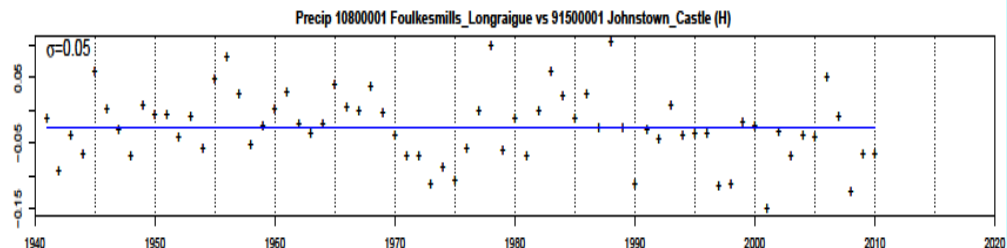
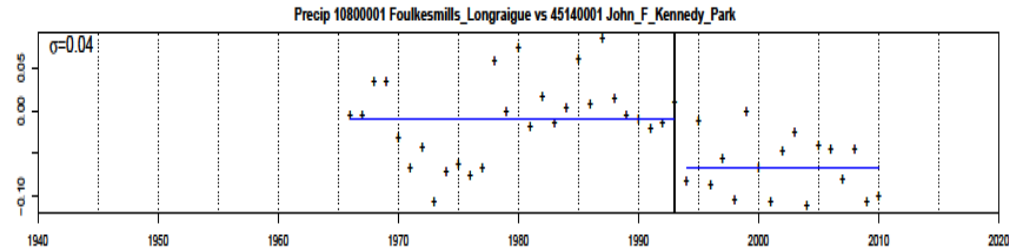
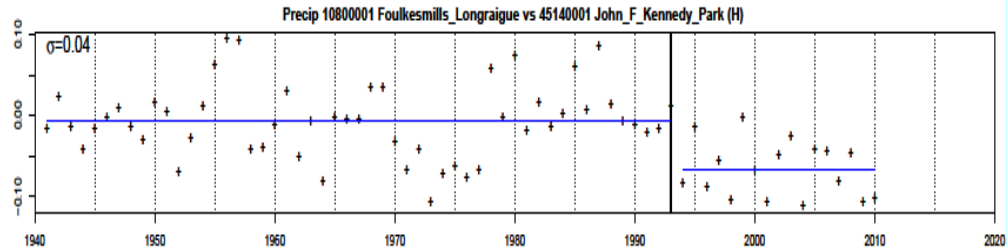


# (1) Foulkesmills & reference series

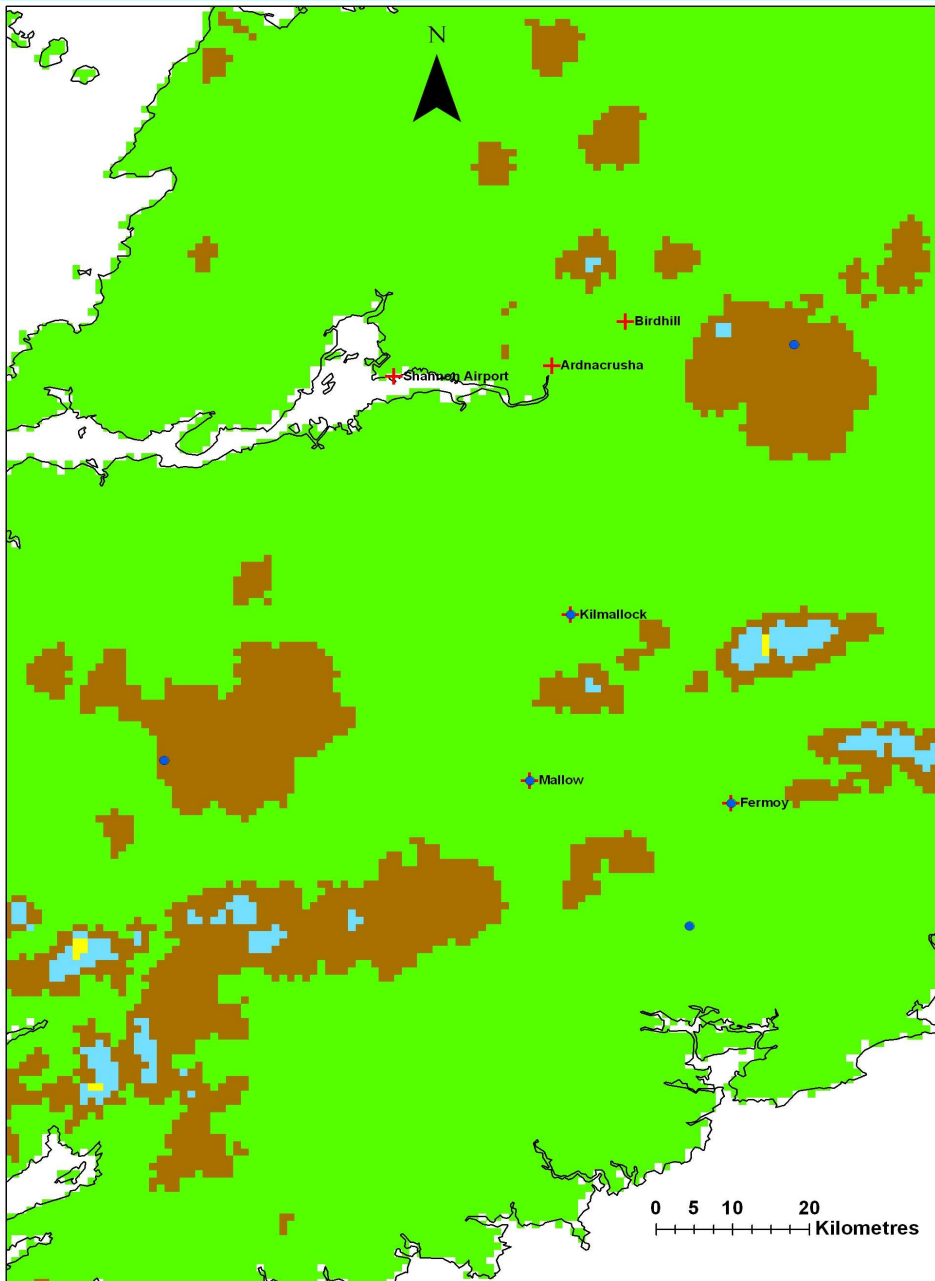


# 'detect\_horr....a'

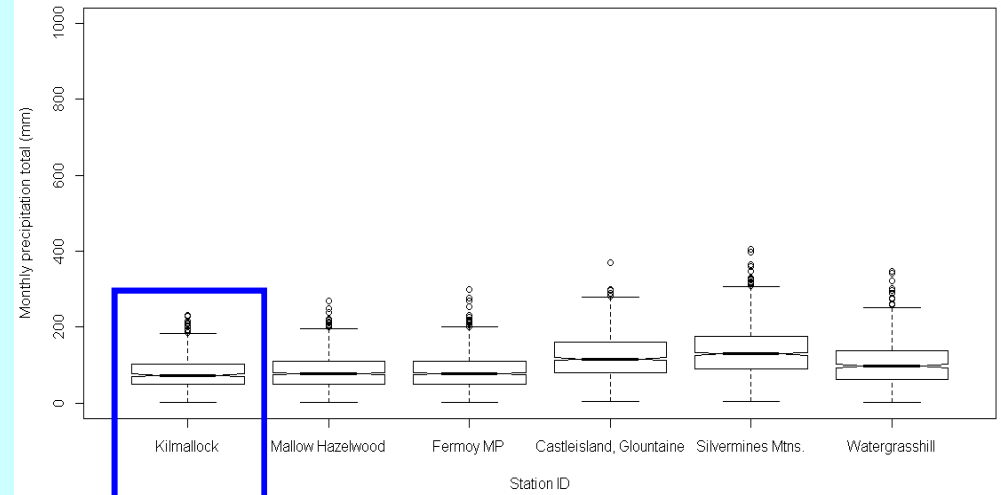
# 'detect\_qcrr....a'



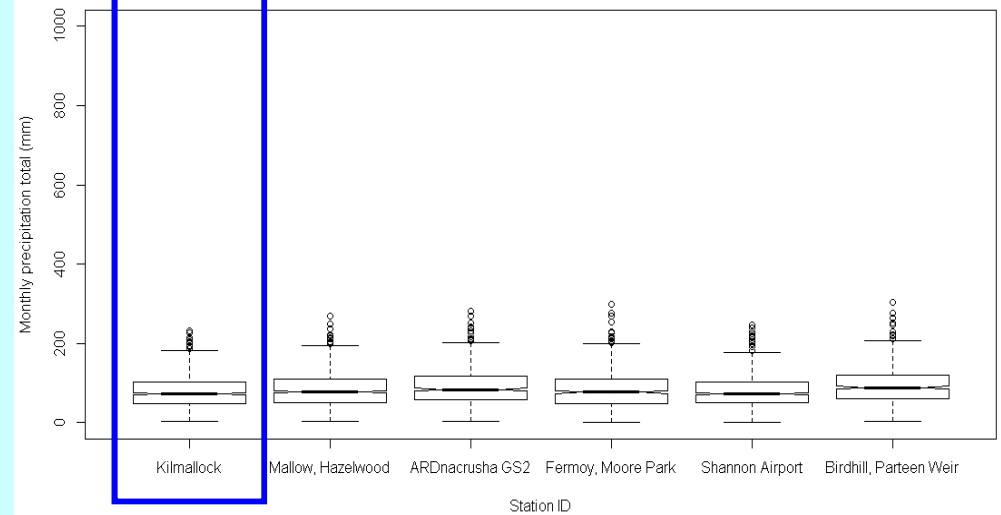
## (2) Kilmallock & reference series



Kilmallock: Full Series Boxplots - First Difference Correlation Distances

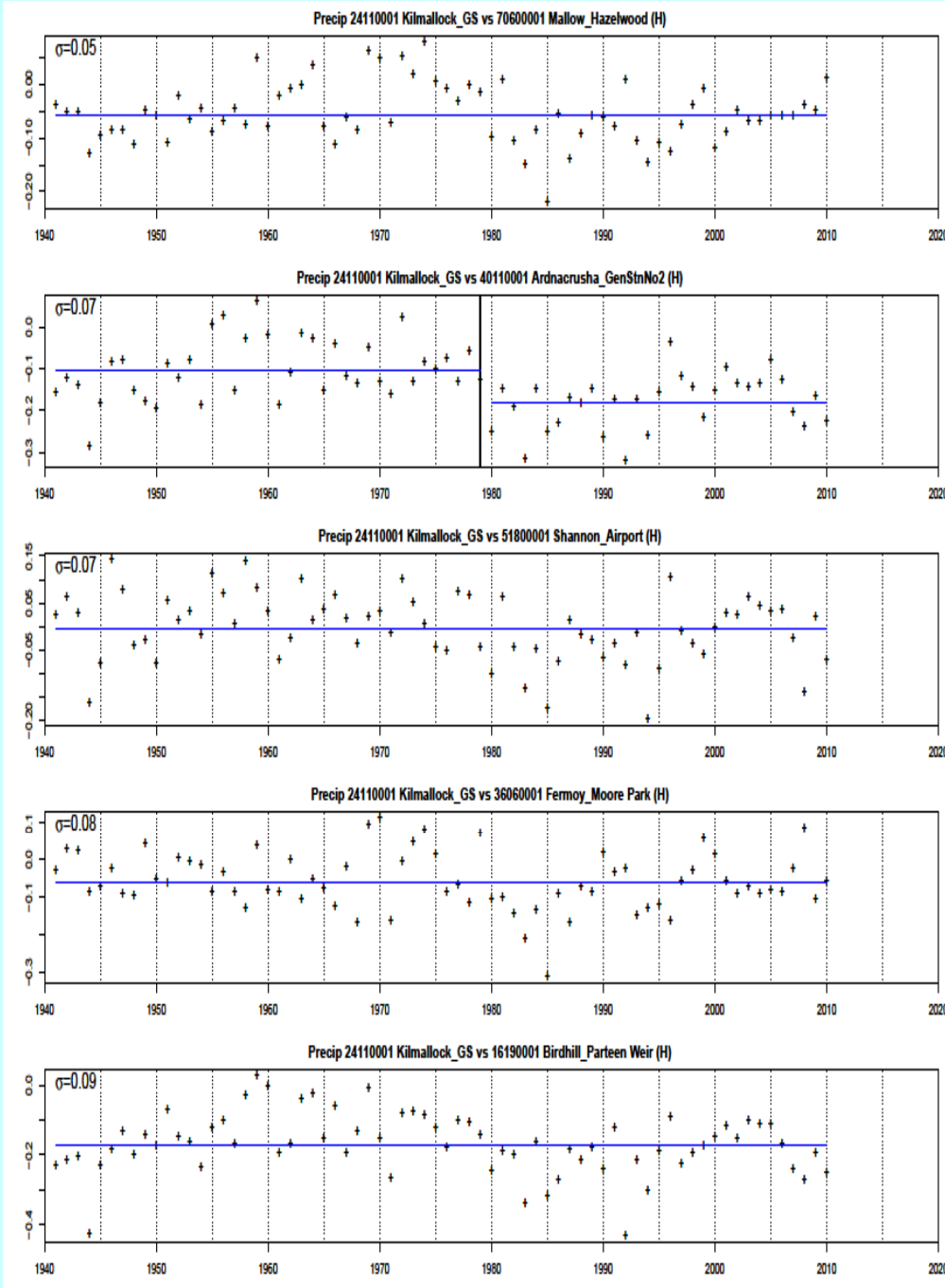


Kilmallock: Full Series Boxplots - HOME-R Geographic Distances

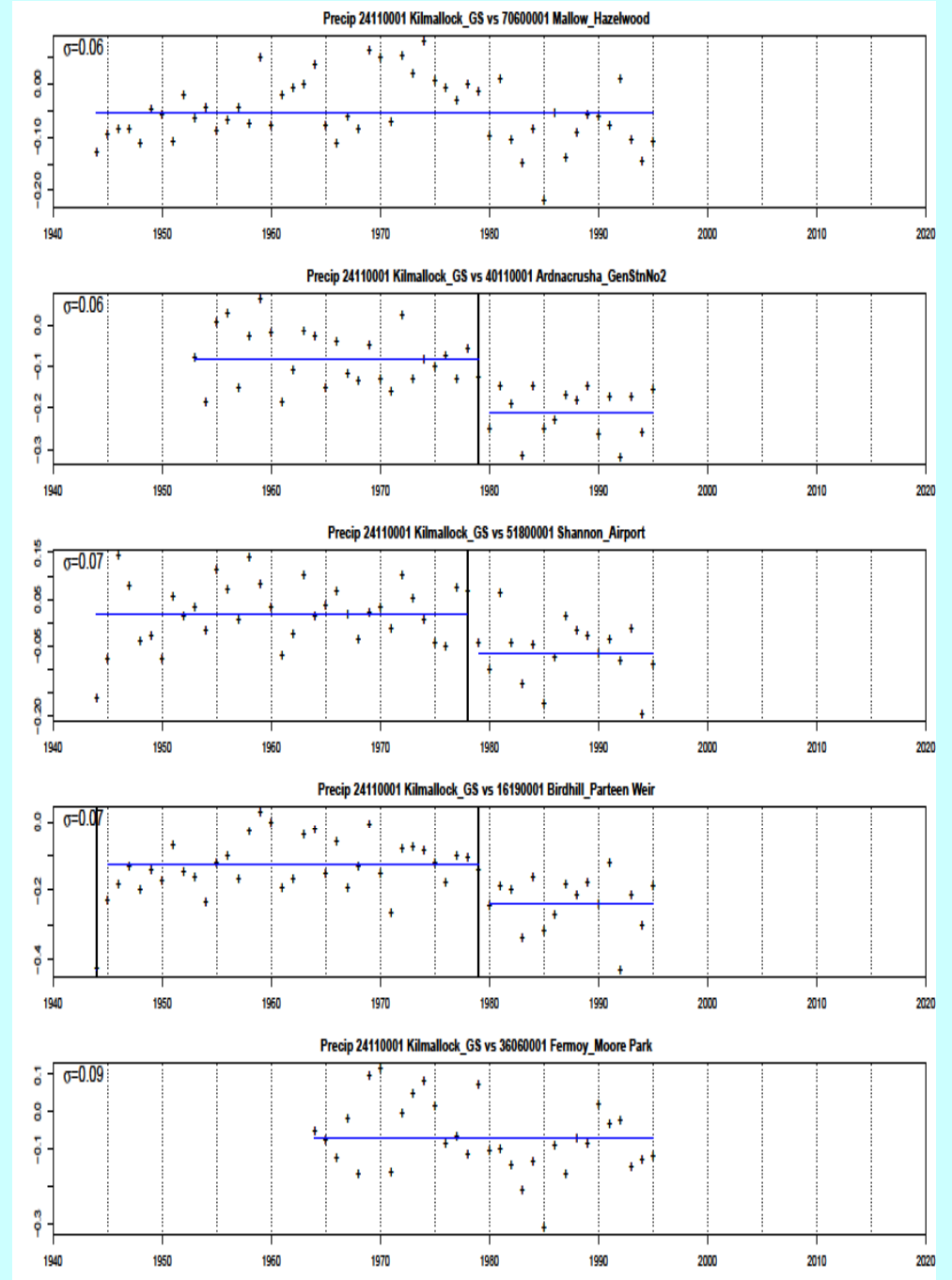




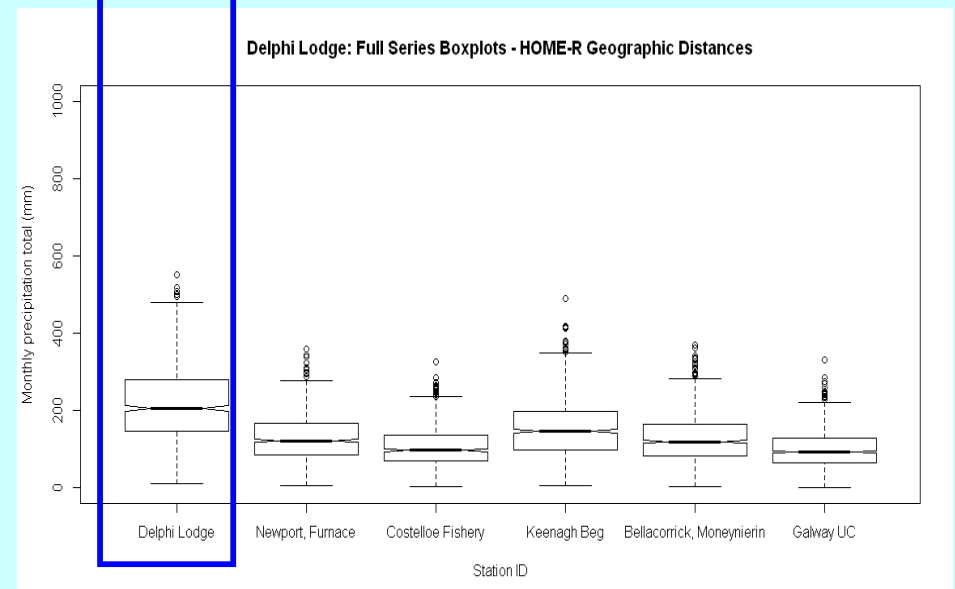
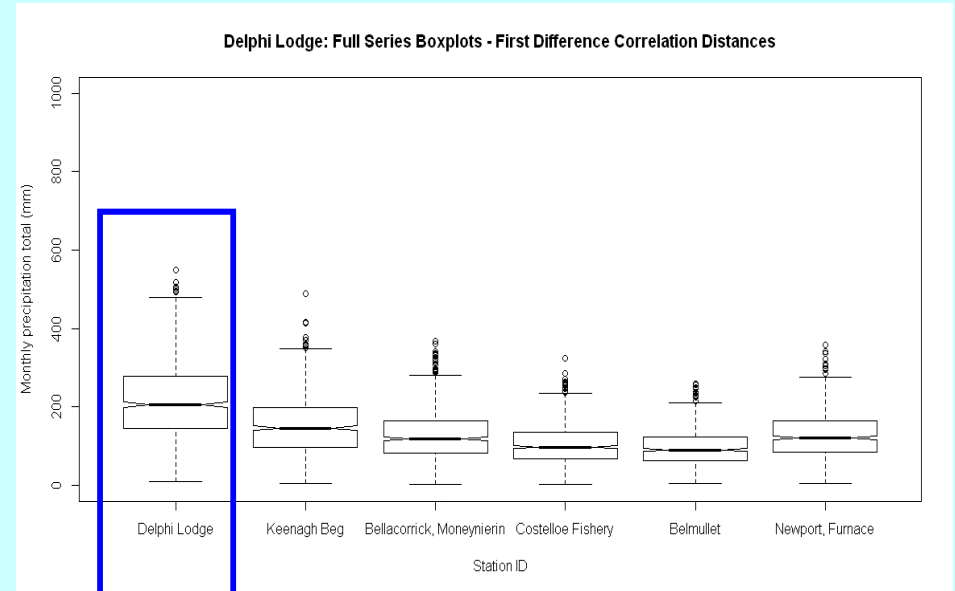
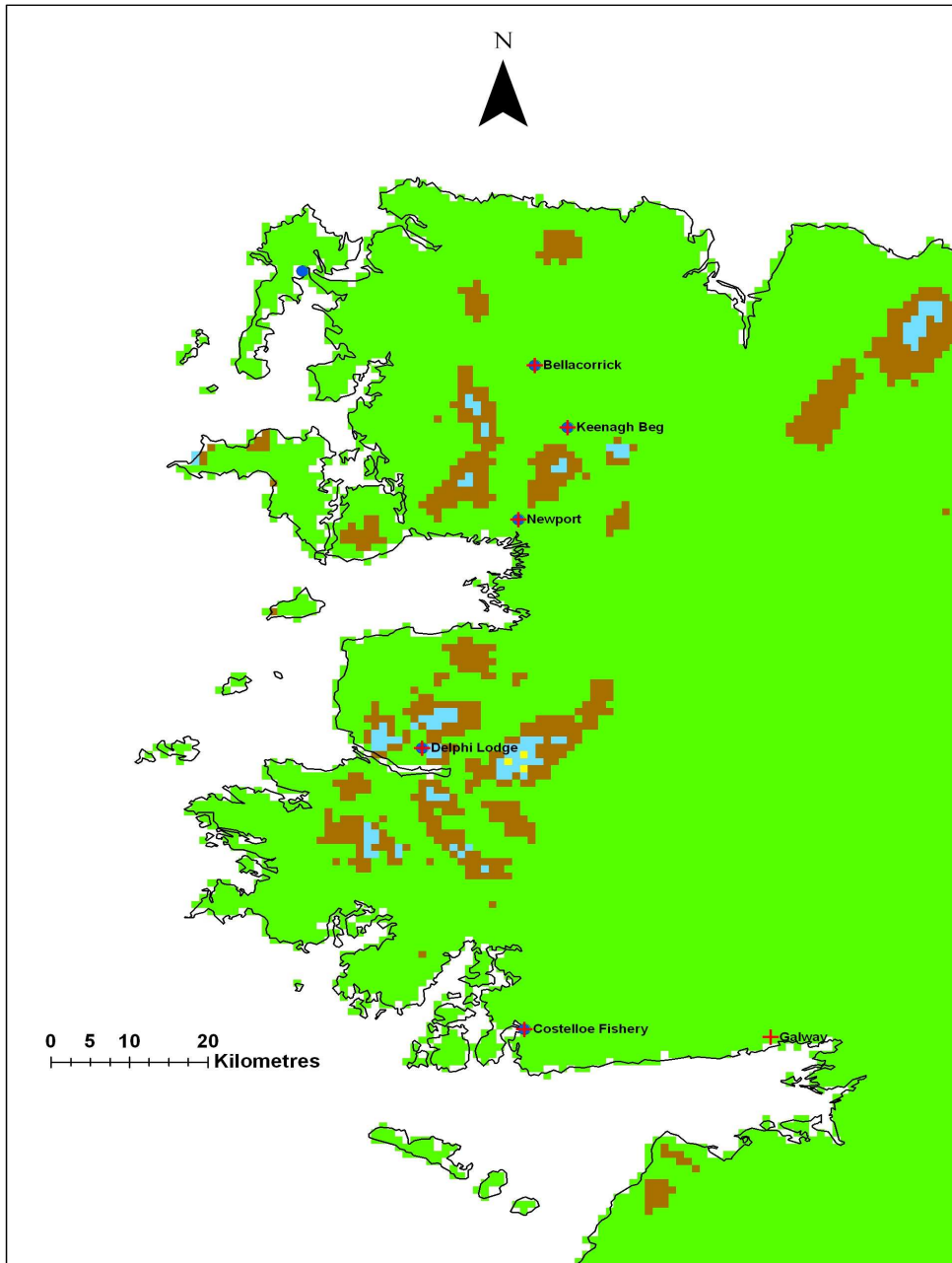
# 'detect\_horr....a'



# 'detect\_qcrr....a'

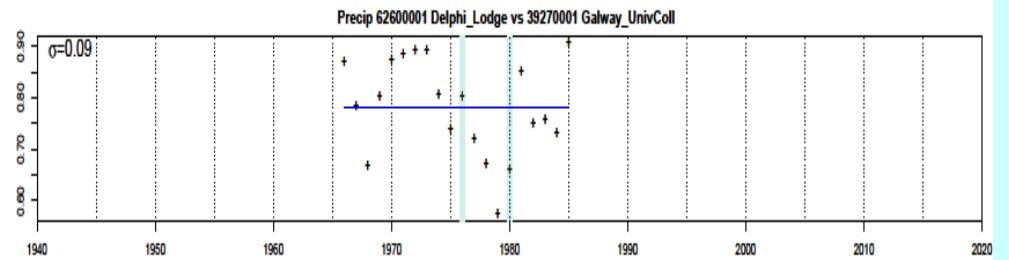
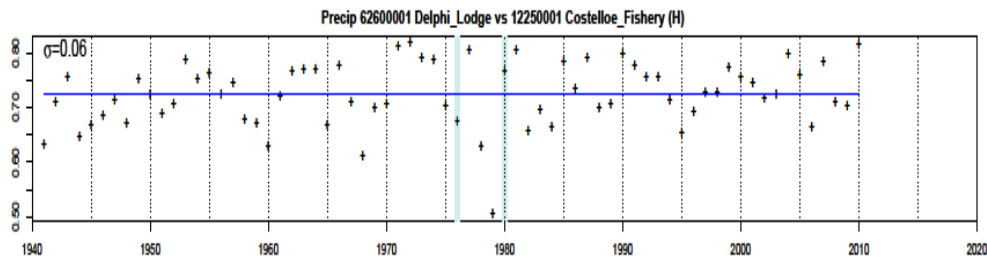
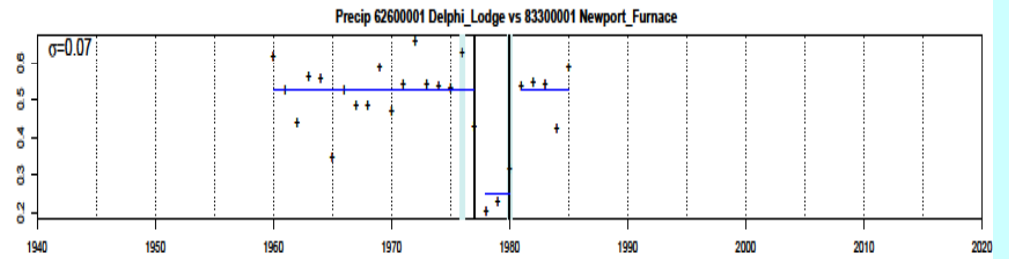
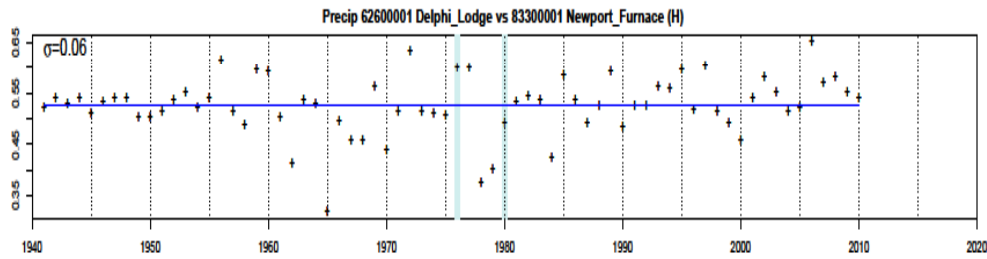
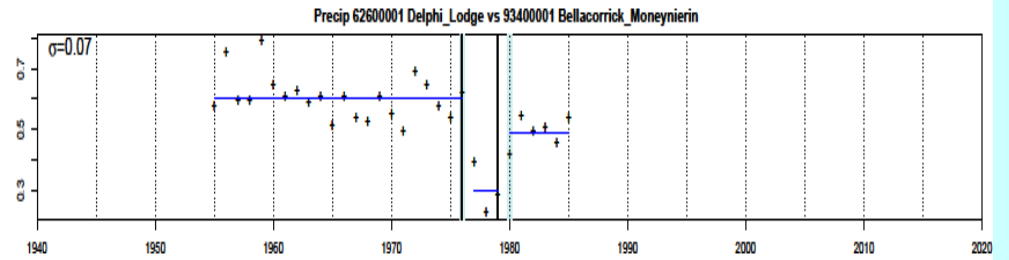
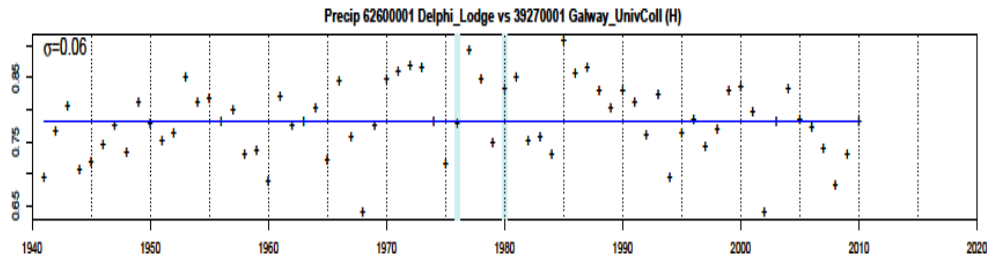
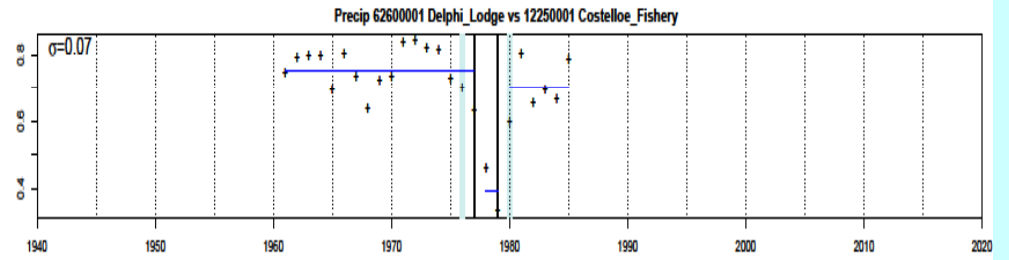
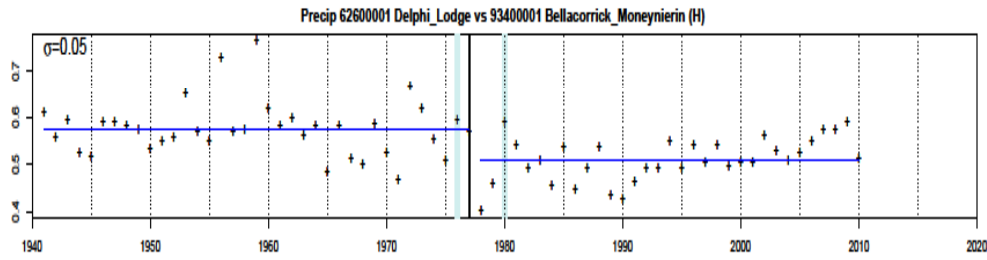
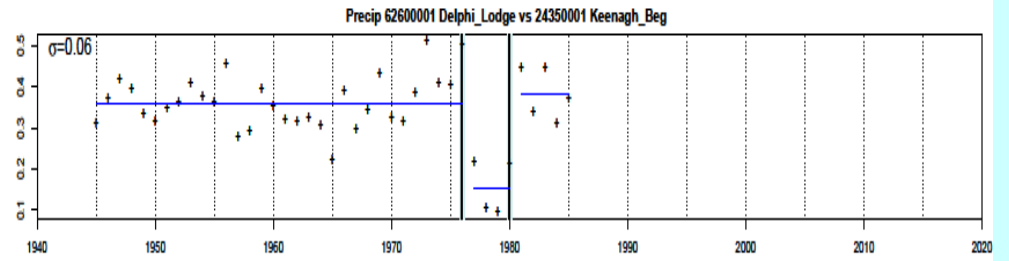
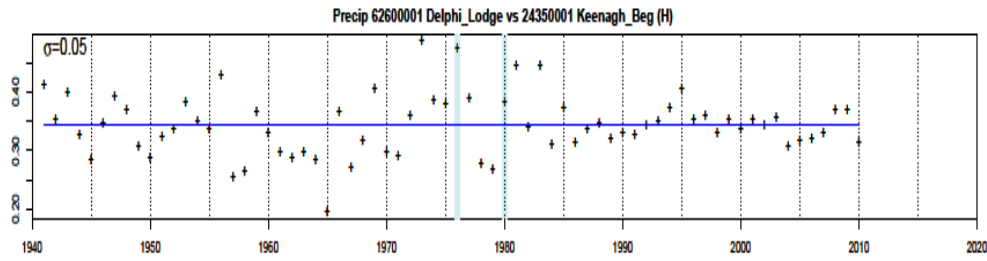


# (3) Delphi Lodge & reference series

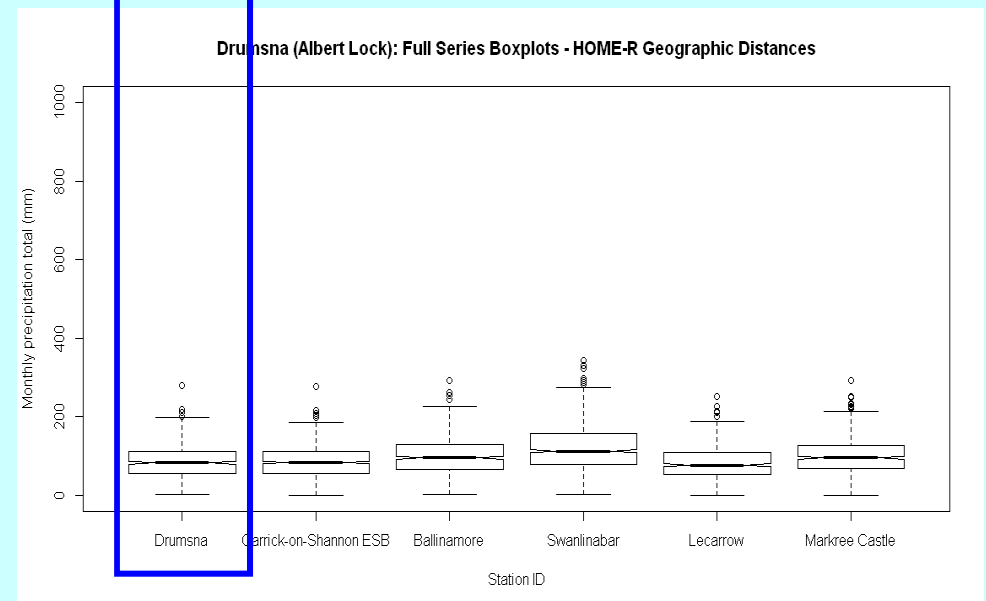
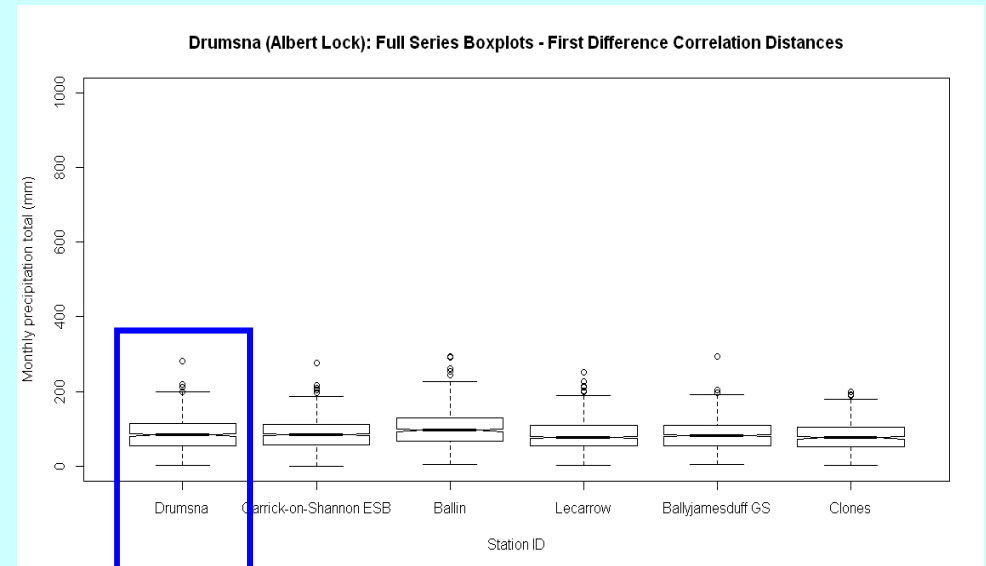
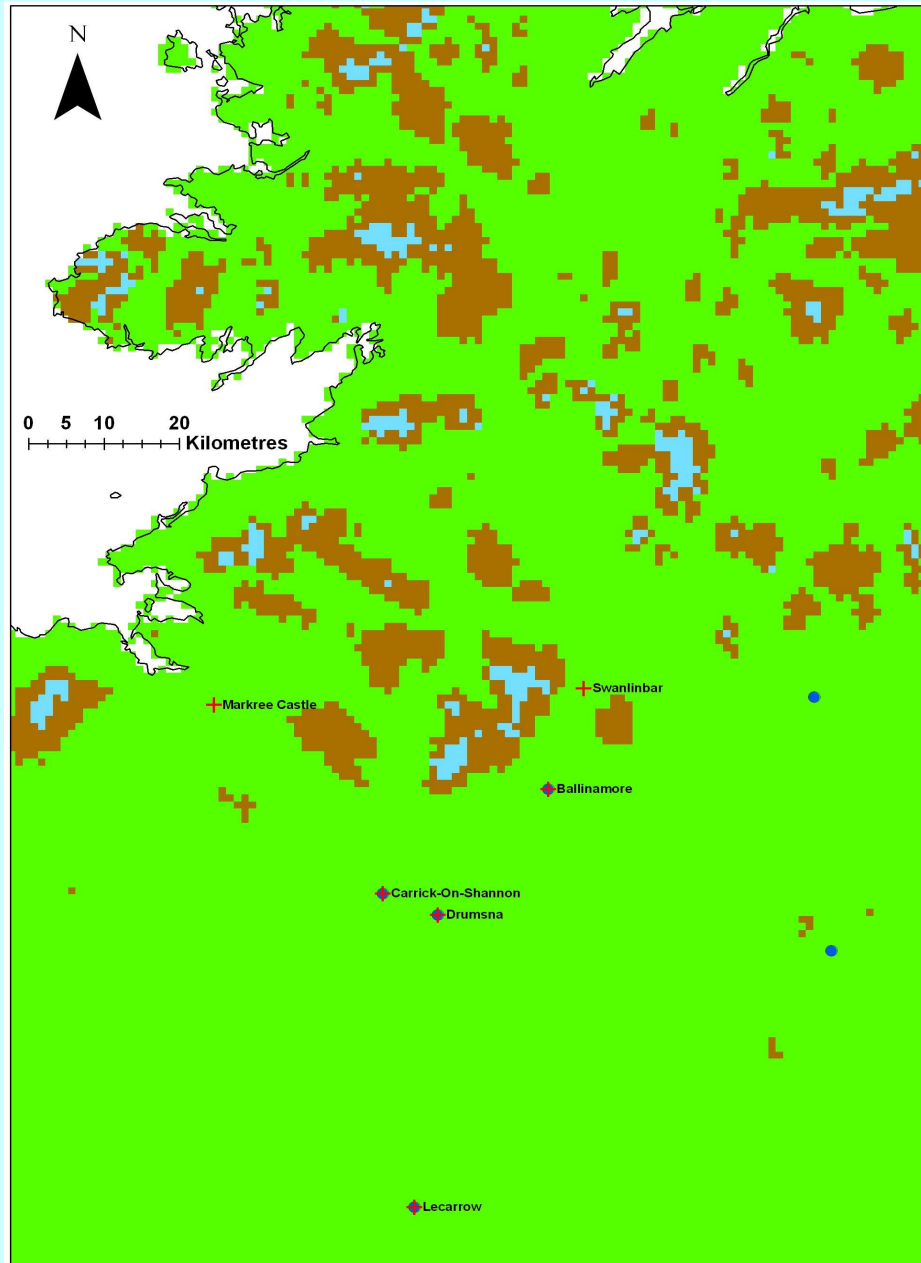


# 'detect\_horr....a'

# 'detect\_qcrr....a'

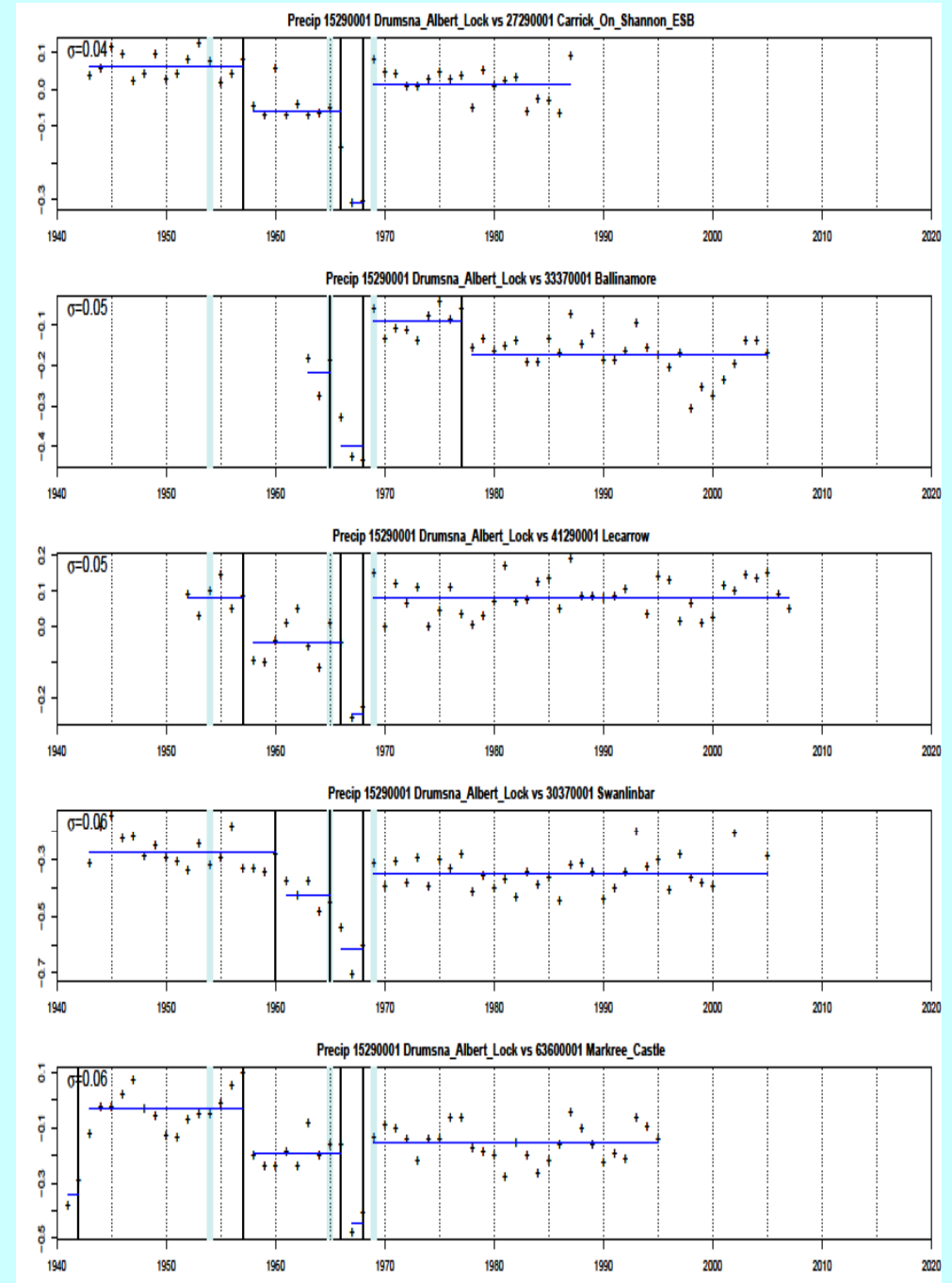
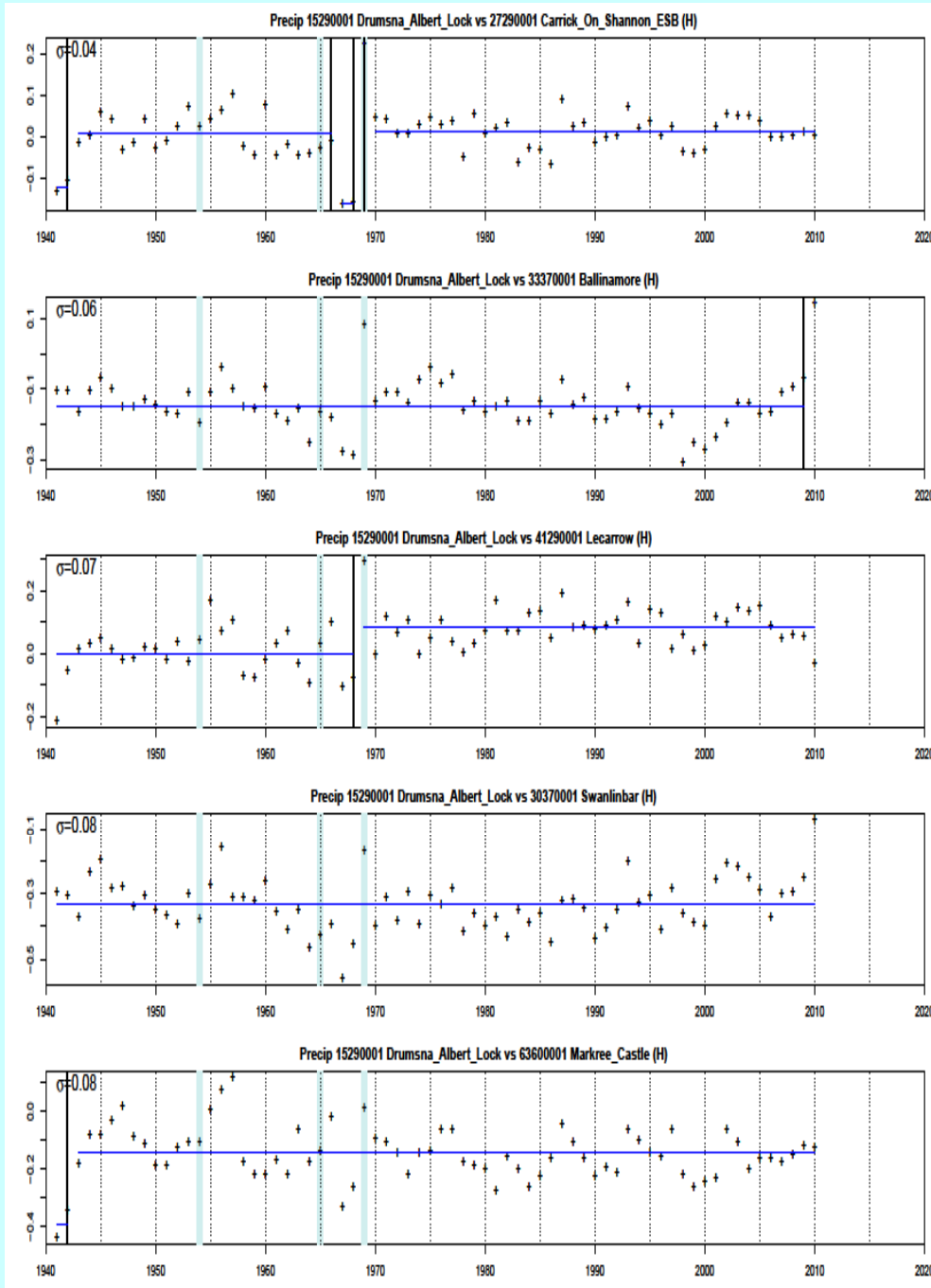


# (4) Drumsna (Albert Lock) & reference series



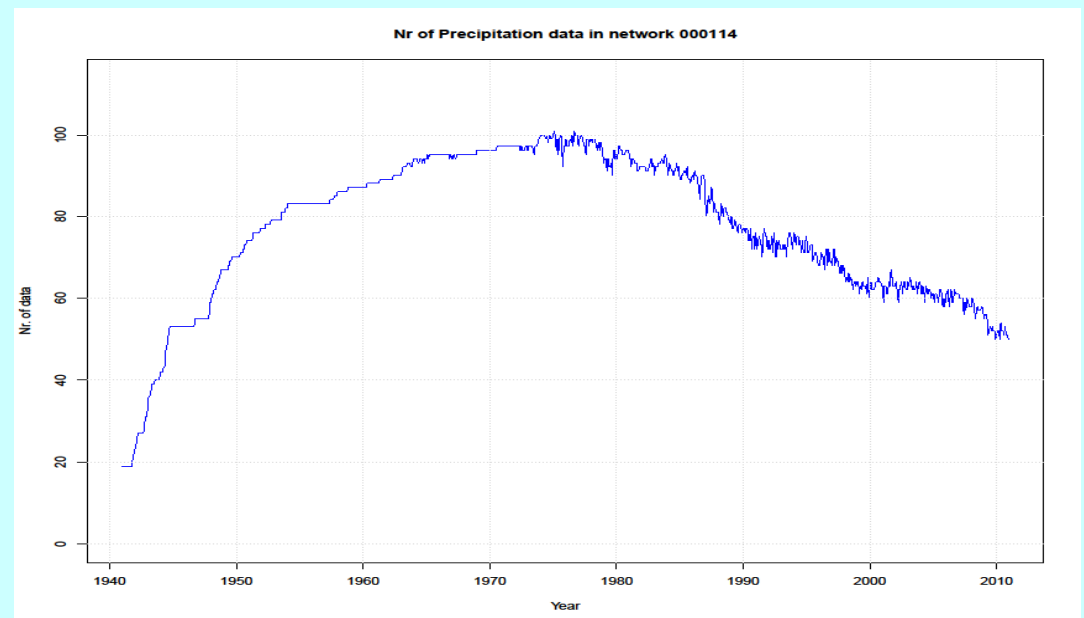
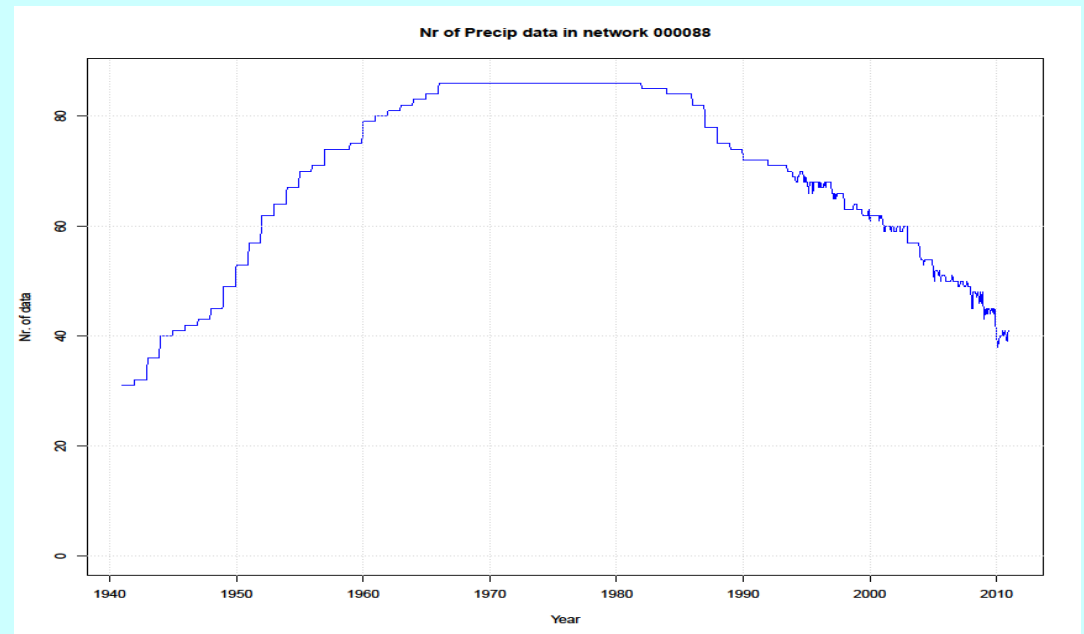
# 'detect\_horr....a'

# 'detect\_qcrr....a'



# Results summary

- Pairwise detection has so far only found breaks in 11 of the records based on the reference networks used
- Therefore in 77 stations no break detected for the networks used (so far!)
- Initial homogenised records are for 1941-2010, so series need extended to 2013 & exercise repeated
- Lot of missing data in the records earlier & later in series - other records need to be introduced (new Reference networks defined)
- But see Diagnostics left!



# Straight on?



Thanks!