10TH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES AND 5TH CONFERENCE ON SPATIAL INTERPOLATION TECHNIQUES IN CLIMATOLOGY AND METEOROLOGY

BUDAPEST, HUNGARY
12-14 October 2020

Development and analysis of long-term quality assured daily precipitation series for Ireland

Ciara Ryan[1,2], Mary Curley[2], Conor Murphy[1] and Seamus Walsh[2]

[1]Irish Climate Analysis and Research Units, Maynooth University, Kildare, Ireland.

[2]Climate, Research & Applications Division, Met Éireann, Glasnevin, Dublin 9, Ireland.

ICARUS
Irish Climate Analysis and Research Units

Maynooth University
National University of Ireland Maynooth

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

MET éireann

An Roinn Tithíochta,
Rialtais Áitiúil agus Oidhreachta
Department of Housing,
Local Government and Heritage

Overview

1) Integrating data rescue into the classroom: ***The Bulletin of the American Meteorological Society (BAMS). Published***

Novel integration of data transcription and quality assurance with teaching and learning at Maynooth University through development of approaches to citizen science in the classroom.

2) Publication of rescued data: ***Geoscience Data Journal. Published***

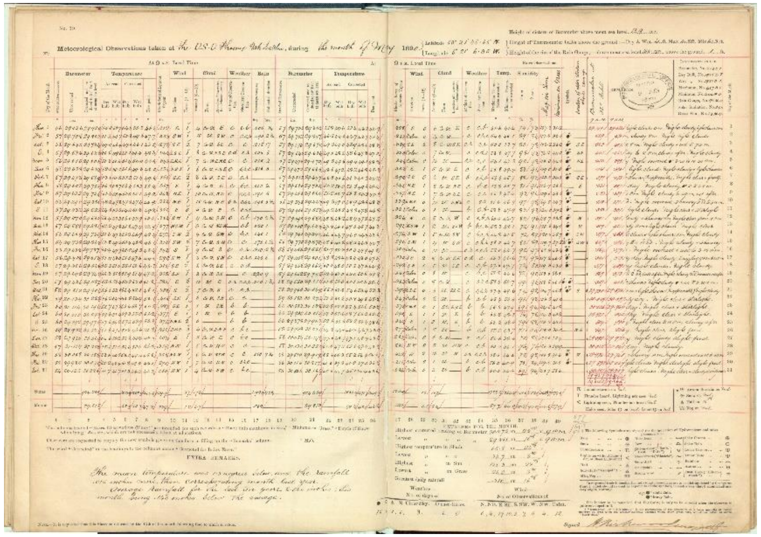
History of meteorological observations in Ireland. Data description paper with link to the dataset and metadata.

3) Development and analysis of long-term daily rainfall series 1900-2019: ***International Journal of Climatology. Submission date: November 2020.***

Description of quality assurance and homogeneity methods. Analysis of ETCCDI indices for derived rainfall series**.** Assessment of the long-term, quality assured datasets to assess changes in the characteristics of extreme events.

Data imaging

Station	Start year	End year
Birr	1873	1951
Blacksod	1884	1956
Fitzwilliam Square	1869	1935
Malin Head	1885	1955
Markree	1869	1968
Phoenix Park	1866	1959
Roches Point	1873	1956
Valentia	1873	1950

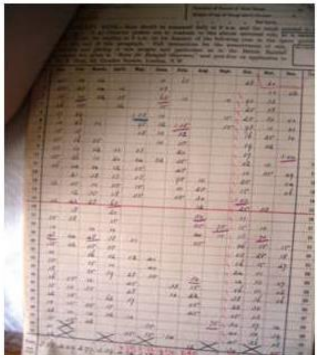


Metis EDS Gamma professional digital scanner used to image historical meteorological registers held in the archives. Digital images stored in Met Éireann’s database.

Integrating data rescue into the classroom

Transcription from digital image format to digital numerical format was largely undertaken by undergraduate students at Maynooth University as part of a novel crowdsourcing initiative to integrate data rescue activities into the classroom.

An innovative approach to data rescue by developing a research-led project to engage students in data rescue tasks.



The study explored i) the potential for integrating data rescue activities into the classroom, ii) the ability of students to produce reliable transcriptions, and iii) the achieved learning outcomes for students.

The work was facilitated by the provision of student aids including written guidelines, transcription templates with an automated quality-assurance check, a video tutorial, in-class workshops and an online discussion forum.

Data Rescue project - summary

- Following the success of the initial project, a further two iterations were executed across three cohorts of undergraduate students at Maynooth University.

- In total, 3616 station years of rainfall data (~1.32 million daily values) were transcribed by students.

- Transcriptions were double-keyed.

- Option for students to continue working with the data in Research Methods module (semester 2)

- Methodology and resources published in the Bulletin of the American Meteorological Society (BAMS). *Bull. Amer. Meteor. Soc.* (2018) **99** (9): 1757–1764. Available at: https://doi.org/10.1175/BAMS-D-17-0147.1

Quality Control: part 1

At each stage of the transcription process, quality assurance measures were employed to preserve the integrity of the data being rescued. Keying guidelines were developed ensuring conformity to World Meteorological Organisation (WMO) standards (WMO, 2016).

Monthly totals were examined against the derived sum of the daily entries to identify potentially incorrect data entries. The data were double keyed and the entries from different transcribers compared.

Where the entries agreed, the value was provisionally accepted as the raw data value. If the values disagreed, the original record was manually examined to ascertain the true observed value. An examination of errors across all transcriptions revealed a percentage error of <1%.

Multiday accumulations were identified and flagged using the original records as a reference. A description of numerical flag values is included in the metadata files. These indicator flags will facilitate the re-distribution of multiday accumulations to the respective days on which no observation was recorded.

As a final check for transcription errors, the upper and lower 1% of observations (non-zero precipitation) were examined for each individual station record. Values identified as outliers were cross-checked against the original record.

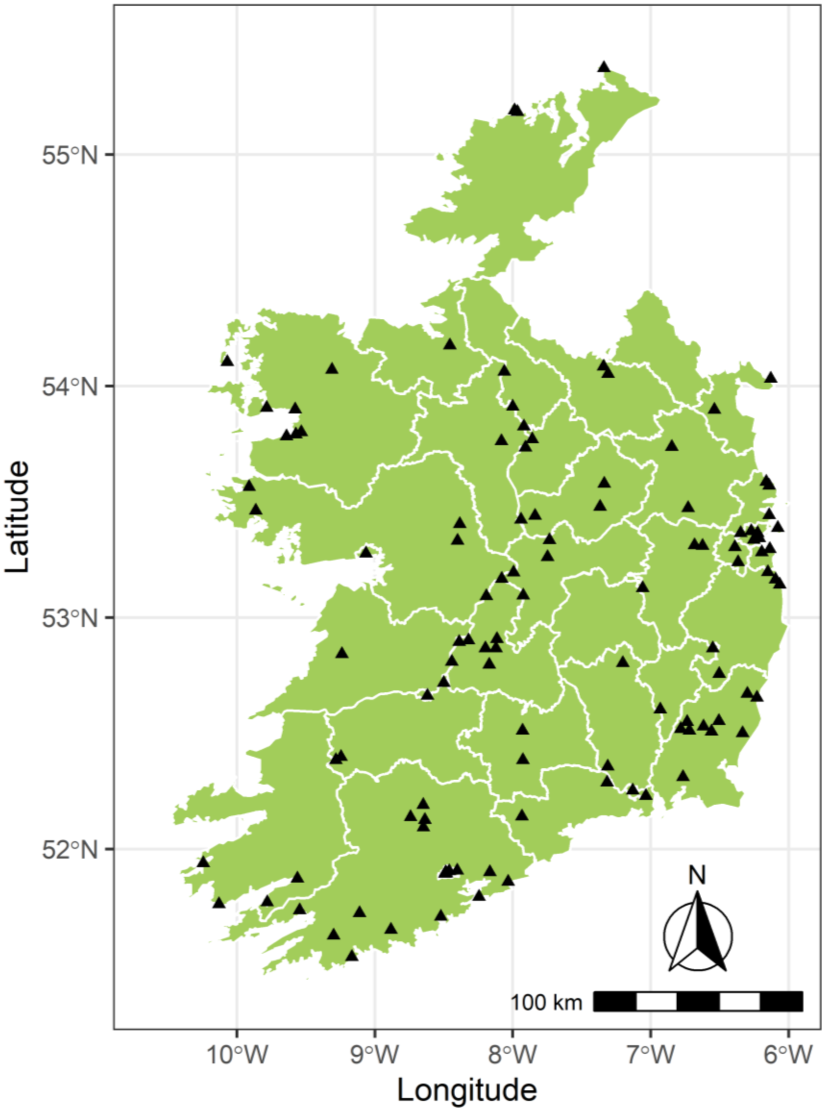
Access to the rescued data

The paper published in Geoscience Data Journal presents the raw data and associated metadata. It is envisaged that by presenting the data in its original state it can be easily integrated into current international data rescue initiatives, e.g. Copernicus Climate Change Service Global Land and Marine Observations Database, and that future research will have recourse to the raw data.

The data are freely available from the edepositIreland data centre (http://hdl.handle.net/2262/91347). The dataset comprises daily rainfall data for 114 stations at various locations throughout Ireland for varying time periods.

Individual station folders contain two files: a data file in ASCII format and a corresponding metadata text file.

Rainfall values run continuously from start date to end date of the data recovery period, with missing values denoted using a −999 indicator.



Quality Control: part 2

Generally, quality assurance tests are divided into 5 categories: (i) Basic integrity checks, (ii) Tolerance checks (iii) Internal consistency checks (iv) Temporal consistency checks and (v) Spatial consistency checks.

Note: Internal consistency checks identify inconsistencies between parameters (e.g. precipitation and snow-depth) and so are not applied here.

Suspect values: Adjust, accept or reject?

Does the value agree with the original record?
Does the metadata provide any information?
Was the event noted in neighbouring stations?
Is the value physically reasonable for this station/region/season?

QC1: Check for non-numeric values – check structure of each file (e.g. Year, Month, Day, Ind = INTEGER; Rain = NUMERIC). Also, check that all years are between 1864-2019; months between 1-12; days between 1-31.

QC2: Check for negative precipitation values (set -999 to NA).

QC3: Check for potential monthly accumulated values by identifying months with only one value recorded which is in excess of 2 times the mean daily rainfall intensity for that month.

QC4: Anomalous sequence of zero precipitation – generally result from zeros being used in place of missing value code. Flag if zeros persist for ≥ 1 month duration.

Quality Control cont… Climdex_extraQC

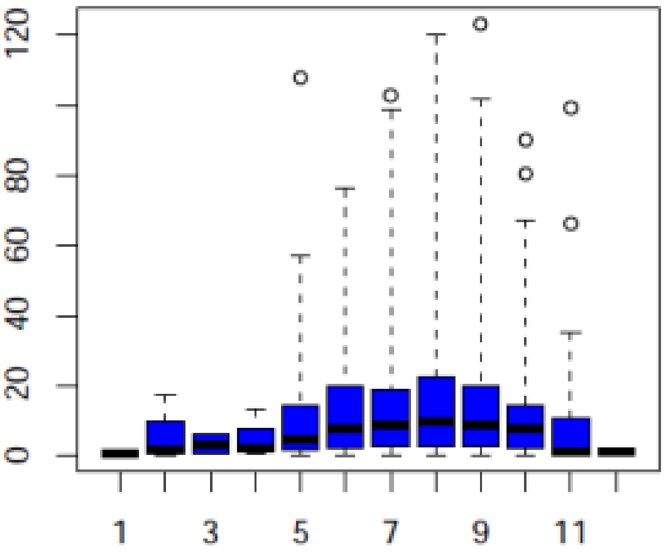
QC5: Duplicate dates control – Climdex_extraQC

QC6: Out of range values, based on fixed threshold values – Climdex_extraQC (here the maximum daily rainfall for Ireland is used) 243.5MM = highest daily total 18th Sep 1993 Cloone, Kerry.

QC7: Rounding problems evaluation – Climdex_extraQC Visual assessment to identify rounding issues by plotting the values after the decimal point. It shows how frequently each of the 10 possible values (.0 to .9) appears.

QC8: Climatological outlier test: Outliers, based on Interquartile range exceedance – Climdex_extraQC – this is a percentile based approach and therefore suitable for asymmetric distribution. Use non zero rain days. Also, exclude flagged multiday accumulations.

As can be seen in the example below, the mystation_boxes.pdf file, produces boxplots of precipitation data flagging as outliers all those values falling outside a range with p25 – 5 interquartile ranges (lower bound) and p75 + 5 interquartile ranges (upper bound).



Quality Control cont…

Temporal consistency checks:

QC9: Temporal check to determine whether or not the month in question is consistent with the sample population of other such months for that station.

The temporal check for outliers for a particular station is based on the premise that an individual monthly value should be similar (in statistical sampling sense) to values for the same month for other years. Outliers are identified utilising the sample distribution of each calendar month separately for each station.

Extreme values are flagged based on limits determined from a multiple of the interquartile range (IQR) calculated for each station/month.

An outlier is flagged when $X_i - {}_q50 > f(\mathrm{IR})$ where X_i is the monthly mean of the year i, ${}_\mathrm{q}50$ is the median, and f is the multiplication factor.

Quality control cont..

Spatial Consistency checks:

Check data for consistency with nearest neighbours according to the following criteria:

1. **Number of dry days** in month:
   - 2 days more/less than max/min of nearest neighbour
2. **Number of wet days** rr >= 0.2 mm in month:
   - 3 days more/less than max/min of nearest neighbour
3. **Days >= 5mm** in month
   - 3 days more/less than max/min of nearest neighbour
4. **Days>=10mm** in month
   - 2 days more or less than min/max nearest neighbour
5. **Total Monthly Fall** (Normalised, i.e. as fraction of long-term average)
   - +/- 25% of the max/min normalised rainfall of nearest neighbours

     i.e., > 1.25 max of NN or < 0.75 min of NN

Infilling/joining station series to produce long-term series

Potential long-term stations were identified based on:
record length; continuity of record; amount of missing data; availability of nearby stations.

Joining/Infilling was carried out using a spatial interpolation method – kriging (R' package geoR). This method performed well in a previous assessment carried out by Walsh (2013) using Irish station data.

Stations have been grouped based on the quality of the original record:

Group A stations:
Continuous record from start year to end year.
Monthly values available where no daily value exists.

Group B stations:
(i) Continuous record with small amount of missing data infilled using monthly values where available and neighboring station data
Or
(ii) Record extended by joining to nearby station.

Group C stations:
Joined/infilled using nearby station data but less confidence in infilling due to low station density during missing period.

st_id	station_name	lat	lon	elevation	start	end	group
1530	ARMAGH	54.352	-6.65	62	1838	2019	A
1929	ATHLONE O.P.W.	53.422	-7.942	37	1902	2019	A
2375	BELMULLET	54.228	-10.01	9	1884	2019	A
2012	CASHEL (Ballinamona)	52.511	-7.929	80	1911	2019	A
1529	DRUMSNA (Albert Lock)	53.911	-8	45	1903	2019	A
108	FOULKESMILL (Longraigue)	52.311	-6.766	71	1874	2019	A
417	INAGH (Mt.Callan)	52.842	-9.238	122	1908	2019	A
1575	MALIN HEAD	55.372	-7.339	20	1885	2019	A
1275	MARKREE	54.175	-8.456	34	1874	2019	A
1519	MEELICK (Victoria Lock)	53.167	-8.081	39	1902	2019	A
175	PHOENIX PARK	53.364	-6.35	48	1881	2019	A
1819	PORTUMNA O.P.W.	53.092	-8.192	35	1929	2019	A
1075	ROCHES POINT	51.793	-8.244	40	1873	2019	A
2275	VALENTIA OBSERVATORY	51.938	-10.24	24	1875	2019	A
1812	WATERFORD (Tycor)	52.253	-7.131	49	1890	2019	A
3310	ABBEYFEALE (Caherlane)	52.352	-9.284	155	1925	2019	B
2528	BALLYFORAN (BORD NA MONA)	53.44	-8.303	47	1925	2019	B
675	BALLYHAISE	54.051	-7.31	78	1900	2019	B
944	CREESLOUGH (Carrownamaddy)	55.133	-7.95	88	1908	2019	B
1375	DUNSANY	53.516	-6.66	83	1900	2019	B
4015	ENNISCORTHY (Brownswood)	52.463	-6.561	18	1900	2019	B
1923	GLENASMOLE D.C.W.W.	53.239	-6.367	158	1900	2019	B
201	GLENGARRIFF (Ilnacullin)	51.735	-9.546	7	1914	2019	B
1475	GURTEEN	53.053	-8.009	75	1900	2019	B
2115	HACKETSTOWN (Voc.Sch.)	52.861	-6.553	189	1918	2019	B
603	KENMARE (DERREEN)	51.769	-9.781	24	1912	2019	B
4513	KILKENNY (Lavistown House) II	52.636	-7.197	58	1900	2019	B
6019	KILLALOE DOCKS	52.81	-8.449	40	1902	2019	B
5131	KILSKYRE (Robinstown)	53.693	-6.963	87	1900	2019	B
706	MALLOW (Hazelwood)	52.19	-8.65	94	1900	2019	B
875	MULLINGAR	53.537	-7.362	101	1900	2019	B
1338	OMEATH	54.087	-6.256	12	1900	2019	B
8212	PORTLAW-MAYFIELD II	52.291	-7.301	8	1900	2019	B
6329	STROKESTOWN (Carrowclogher)	53.753	-8.108	52	1908	2019	B
2227	CARNDOLLA	53.403	-9.016	24	1900	2019	C
1433	WESTPORT (Carrabawn)	53.792	-9.527	56	1909	2019	C

Homogeneity testing:

Detection and adjustment of breaks carried out using RHtests software (Wang et al. 2010)

RHtests_dlyPrcp software package is specifically designed for homogenization of daily precipitation data. In this study breaks are detected at monthly scale and adjustments applied to daily.

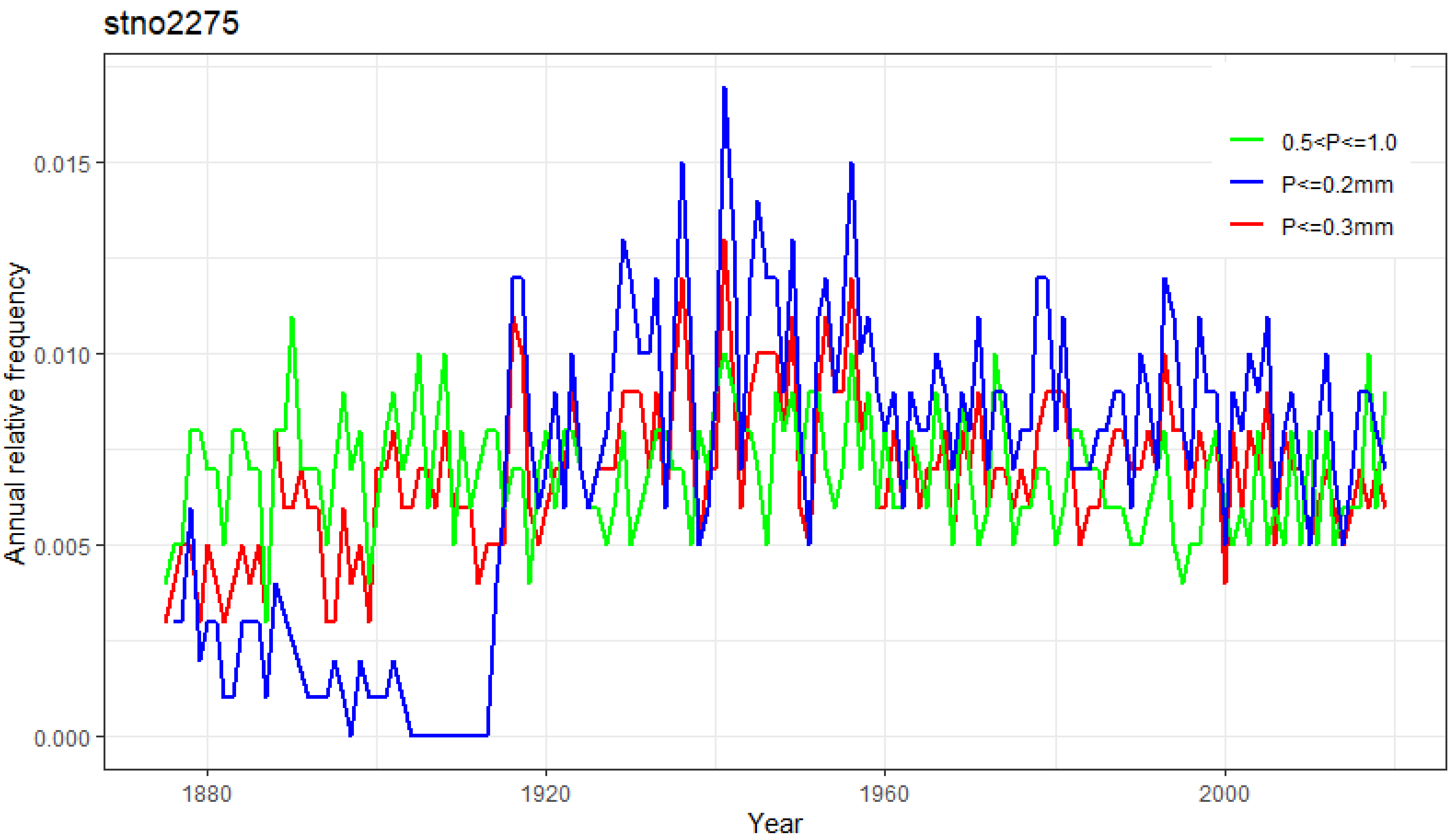
Software detects both Type 1 and Type 0 changepoints and allows for testing of known changepoints.

Monthly break detection good at detecting documented breakpoints.
Daily break detection also good at detecting breakpoints BUT also detects many other undocumented breakpoints.

Outputs both the mean-adjusted and QM-adjusted data series, however, inhomogeneities remain after the application of mean adjustment.

Note that the QM adjustments could be problematic if a discontinuity is present in the frequency of precipitation measured (see Wang et al. 2010 for details).

Frequency discontinuities



Provisional results:

18 stations found to be homogenous

21 breaks detected across 17 stations

st_id	station_name	breaks	reason
1575	Malin Head	1921	Readings were made at the telegraphic reporting station (Lloyds tower at a height of 230 ft). In 1921 station moved to the coast guard station and at a height of approximately 20 ft above msl.
1529	Drumsna	1917; 1942	No documented reason for break detected in 1917 but this break also detected by HOMER previous work by Noone et al. (2015). New gauge installed in 1942. Comparison of old site over 13-month period showed new gauge recording 154% of old site.
2275	Valentia	1993	No documented reason for detected breakpoint. Beofre accepting breakpoint further investigation is required to determine if this is a natural associated with a change in the NAO index in 1994.
1929	Athlone	1926	Reports of gauge leaking resulting in low readings. Recommendation for new gauge to be installed, however, no documentation to say that this was carried out. The height of the gauge above ground was changed from 2ft to 1ft in Jan 1928.
2375	Belmullet	1914; 1956	Reports of change in station elevation in 1914. Sept 1956 new station established

Next steps:

Assessment of the long-term, quality assured series to assess changes in the characteristics of extreme events. For this purpose, indices derived by the Expert Team on Climate Change Detection and Indices (ETCCDI) will be extracted and investigated for evidence of trend and variability in long-term records across Ireland.

- RX1day: Monthly maximum 1-day precipitation
- Rx5day: Monthly maximum consecutive 5-day precipitation
- SDII: Simple daily intensity index
- PRCPTOT: Annual total PRCP in wet days (RR>=1mm)
- CDD: Maximum number of consecutive days with RR<1mm Days
- CWD: Maximum number of consecutive days with RR>=1mm Day
- R95p: Annual total PRCP when RR>95th percentile
- R99p: Annual total PRCP when RR>99th percentile

Thank you

Correspondence to: Ciara Ryan (ciara.ryan@met.ie)

ICARUS
Irish Climate Analysis and Research Units

Maynooth University
National University
of Ireland Maynooth

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

MET
éireann

An Roinn Tithíochta,
Rialtais Áitiúil agus Oidhreachta
Department of Housing,
Local Government and Heritage