Met Éireann

Climatological Note No. 15

Long-term rainfall averages for Ireland, 1981-2010

Séamus Walsh

Glasnevin Hill, Dublin 9
2016

Title: Long-term rainfall averages for Ireland, 1981-2010.

Author: Séamus Walsh

Table of Contents

1. ABSTRACT **1**

2. INTRODUCTION **1**

3. DATA **3**

3.1 THE RAINFALL NETWORK 3
3.2 QUALITY CONTROL 3
3.3 REDISTRIBUTION OF CUMULATIVE TOTALS 5
3.4 ESTIMATION OF MISSING DATA 5

4. CALCULATION OF STATION LONG-TERM AVERAGES **6**

5. CREATION OF GRIDDED DATASETS **6**

5.1 ANNUAL AVERAGE RAINFALL 8
5.2 MONTHLY RAIN LTAS 8
5.3 VERIFICATION OF GRIDDING 10
5.4 DAYS OF RAIN 10

6. COMPARISON WITH 1961-1990 LTAS **11**

7. SUMMARY AND CONCLUSIONS **15**

8. ACKNOWLEDGEMENTS **15**

9. REFERENCES **16**

1. Abstract

Long-Term Averages (LTA) or Climate Normals are 30-year averages of weather elements. They are used to describe the current climate and to place current weather in context. Met Éireann has produced a suite of LTAs covering the period 1981-2010, which have replaced the 1961-1990 LTAs for day-to-day comparison purposes. LTAs of monthly rainfall and days of rain greater than or equal to 0.2mm, 1mm and 10mm have been compiled for over 750 locations. Using these data and data for Northern Ireland, provided by the United Kingdom Meteorological Office, gridded values have been produced at resolution of 1km^2.

This paper details the process involved in generating the new LTAs, from data collection through the network of rainfall observers, to quality control and estimation of missing values, and finally, the geostatistical methods used to produce the gridded datasets. A comparison is also made between the LTAs for 1961-1990 and 1981-2010, which shows an increase of approximately 6% in Average Annual Rainfall, and other seasonal and monthly changes.

2. Introduction

It is usual to place current weather events in context by comparing them to long-term averages (LTAs) or 'normals'. LTAs are also used as a benchmark for climate models and across a wide range of environmental disciplines.

Climate normals or LTAs are 30-year averages of a particular parameter, usually produced in 10 year cycles; 1951-1980, 1961-1990 etc. The previously available LTAs for the Irish rainfall station network cover the 1961-1990 period (Fitzgerald, 1996). In 2002/2003 LTAs of rainfall were produced on a 5km^2 grid as part of a British Irish Council (BIC) project using UKCIPS software. A gridded 1961-1990 Annual Average Rainfall (AAR), at 2km^2 resolution was produced as part of the Flood Studies Update (Fitzgerald, 2007). There is a need for up-to-date sets of LTAs to reflect the most recent 30-year period, and because GIS systems are now widely in use, there is also a need for quality gridded datasets at high spatial resolution.

This paper describes the production of new sets of monthly and annual LTAs for rainfall observing sites in the Republic of Ireland (ROI) covering the period 1981-2010, and the production of gridded data sets for the Island of Ireland including data for Northern Ireland provided by the United Kingdom Meteorological Office (UKMO). Figure 1 shows the locations for which LTAs were calculated. Station values for the 1961-1990 period are also produced for ROI stations using a consistent analysis method to allow for comparison with the new LTA period.

In the case of rainfall, averages have been produced for precipitation depth and days of rain greater than 0.2mm, 1mm and 10mm. These averages are used in the production of monthly, seasonal and annual weather summaries. The compilation of the new averages involves data collection and quality control.It relies on statistical techniques for data analysis and infilling, and geostatistical techniques for the generation of gridded data sets. See Figure 2.

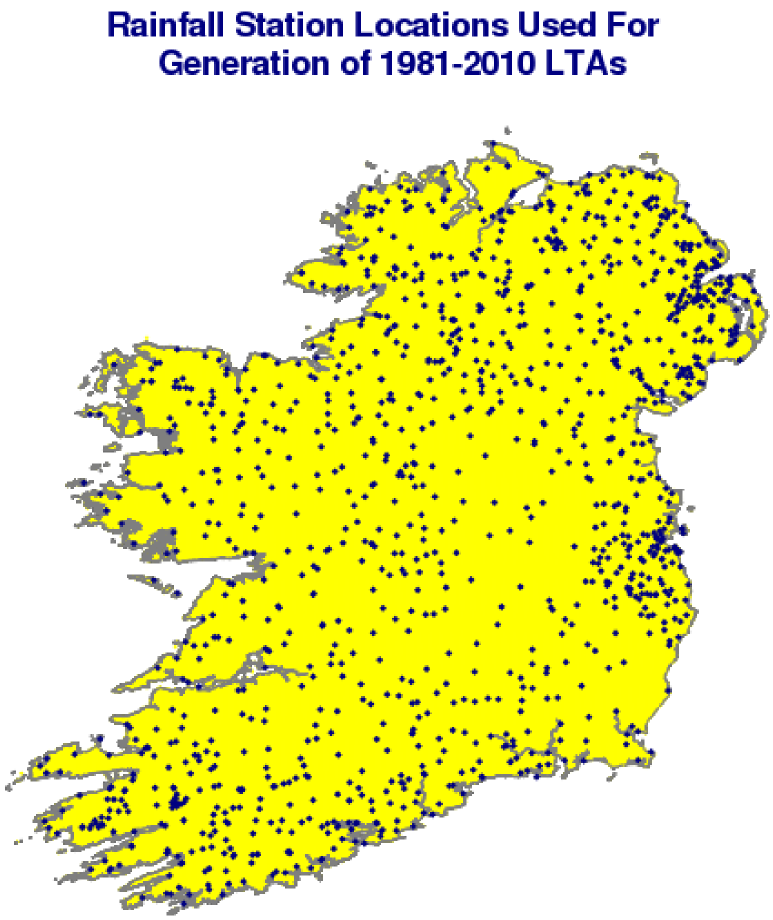

Figure 1: Rainfall Station Locations used for generation of 1981-2010 LTAs.

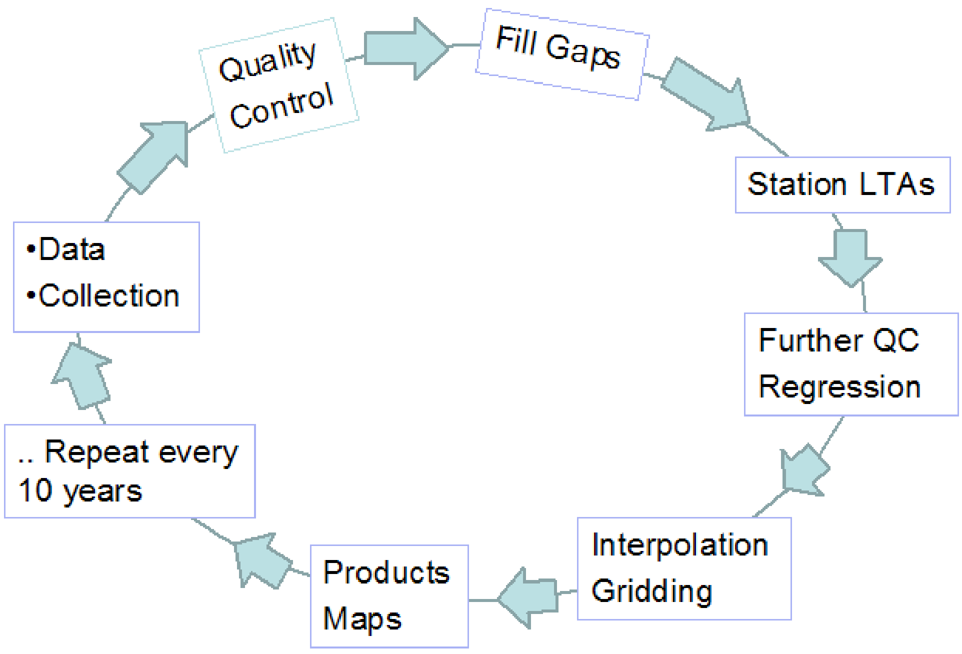

Figure 2: Step-by-step process to produce LTAs

3. Data

3.1 The rainfall network

A daily rainfall total is recorded at rainfall stations every day at 0900UTC and assigned to the previous day. If a reading is missed a cumulative value is taken the next day and flagged as such. There are also a number of rain gauges in remote locations which are read once a month. Readings are sent by post to Met Éireann after the end of each month. The data are then quality controlled and entered into the Met Éireann database. The rainfall network is, for the most part, operated by voluntary observers. The number of locations varies from year to year. Figure 3 shows the approximate number of stations reporting each year since 1941. The fluctuation in the number of recording stations is one of the issues that must be addressed to ensure a consistent dataset.

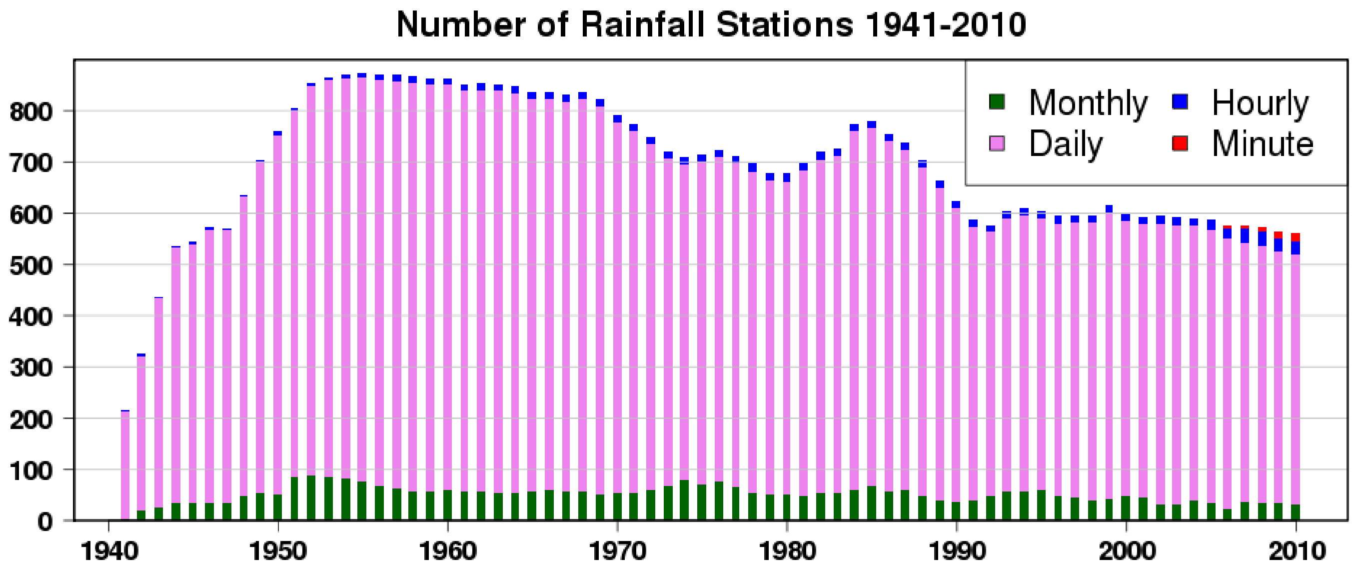

Figure 3: Number and type of rainfall stations reporting each year (1941-2010)

3.2 Quality Control

Data used were extracted from the Met Éireann (MÉ) database, which contains daily and monthly precipitation records from 1941. All data in the database have undergone basic quality control (QC). However, as QC programs and techniques were developed and improved over a number of years, a new QC system was introduced within Met Éireann to include the use of spatial techniques. See Figure 4. All data were run through the new QC (approximately 14 million observations).

A summary of rainfall QC checks for daily and monthly rainfall is shown in Table 1. Each reading is assigned a flag if it fails a QC test. If a reading fails more than three tests it is excluded from the analysis and treated as if it were missing. It is possible for a recording station to fail for daily tests but not to fail monthly. In such a case the monthly total will be used but the individual daily values will not.

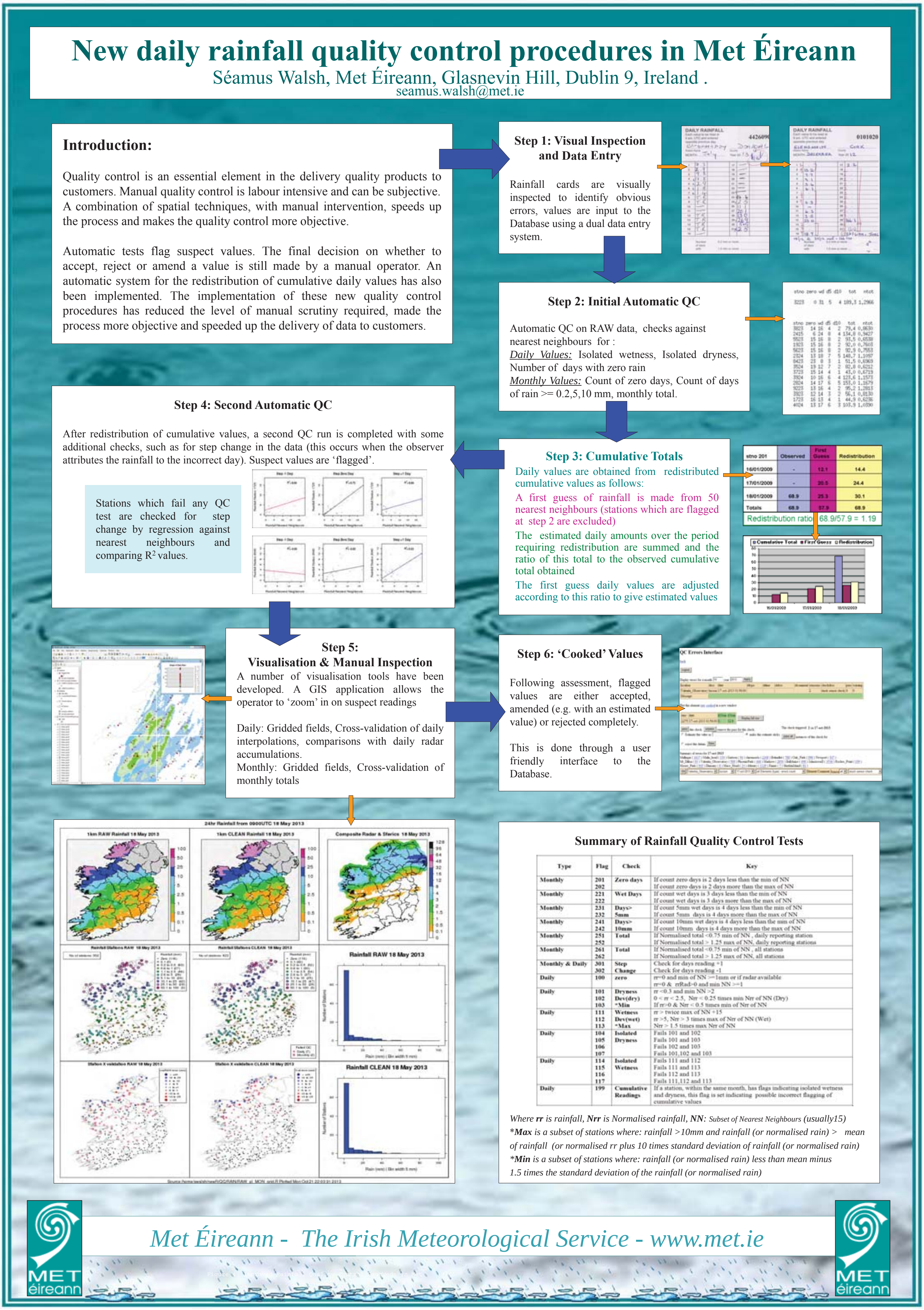

Figure 4: Improved Rainfall Quality Control Procedures in Met Éireann

Period	Check for	No of tests	Summary of tests
Daily	Zeros, Isolated Dryness, Isolated Wetness	7	Various checks, using totals, normalised totals, magnitude of deviation from standard deviations of rainfall of Nearest neighbours
Monthly	Zero days, Wet Days Days>5mm, 10mm Isolated Dryness, Isolated Wetness, Total	14	Various checks, using counts of days, totals, normalised totals, magnitude of deviation from standard deviations of totals of nearest neighbours

Table 1: Summary of Rainfall Quality Control.

3.3 Redistribution of cumulative totals

'Days of rain' parameters require daily values so all cumulative rainfall totals were redistributed to daily values as follows: a first guess of rainfall is made for each run of days for each station requiring cumulative redistribution by inverse distance weighted interpolation from 50 nearest neighbours; the estimated daily amounts for the run of days requiring redistribution are summed, the ratio of this total to the observed cumulative total is calculated; the final estimated daily values are then obtained by adjusting the first guess values according to this ratio. See Table 2.

Gauge 201	Observed Cumulative reading	IDW daily first guess	Daily Redistribution
Day 1	-	14.4	12.1
Day 2	-	24.4	20.5
Day 3	68.9	30.1	25.3
Totals	**68.9**	**57.9**	**68.9**

Table 2: Redistribution of Cumulative Daily Totals.

3.4 Estimation of missing data

Rainfall observations are made on a daily basis so missing daily values are always followed by a cumulative total. As these cumulative values are redistributed as above, missing rainfall values occur on a monthly basis only. Three methods to estimate missing monthly data were investigated. The first was to use weighted ratios of nearby stations, the second was to use weighted spatial regression (Hubbard, 2005) and the third was to do a spatial interpolation of monthly data. The first and second methods used relationships derived from the period (at least 5 years) for which overlaps existed for the station missing data and at least five nearest neighbours. The third method used the output of a regression-kriging interpolation model.

Each method was assessed by removing a random selection of months from observing stations with a complete dataset for the 1981-2010 period, and estimating the removed values using each method. Interpolation and weighted spatial regression performed equally well with weighted means also showing some skill. A combination of the three methods was used with weights assigned to each method according the root-mean-square-error (RMSE) of the cross-validation for each method.

The combination infilling method was verified by omitting data for a number of stations with a complete record and using the methods described above to fill in the gaps. A summary of the verification statistics for the infilling of missing monthly data is shown in Table 3.

Mean Error (mm)	Mean Abs Error (mm)	RMSE (mm)	Mean Monthly Rainfall (mm)	Number of Stations	Number of months
-0.5	7.3	10.6	108	45	1350

Table 3: Summary cross-validation of monthly total infilling methods.

To produce 'days of rain' parameters, a complete daily record for the period is necessary. For stations that had gaps in data, the total for missing months was estimated, then the procedure outlined above for redistribution of cumulative totals was used to generate daily values, with the estimated monthly total used as the cumulative total.

4. Calculation of station long-term averages

Following the estimation of missing daily and/or monthly data, it was possible to compile monthly LTAs for stations that now had a complete set of monthly values for the period. Further tests were carried out on the data at this stage; ratios of monthly totals to annual totals were examined and regression analysis was carried out. A number of outlier stations were removed.

The monthly totals were found to be unreliable for a number of rainfall stations in mountain locations, which are read on a monthly basis. At this stage only annual average rainfall was computed for these stations. The unreliability of these totals is likely due to monthly readings being taken on an irregular basis. LTAs were compiled for stations with at least three complete years during the 30 year period, giving a total of approximately 750 stations.

5. Creation of gridded datasets

Data were gridded at $1km^2$ resolution on the Irish Grid (OSI, 1996) for the rainfall totals and days of rain greater than or equal to 0.2mm, 1mm and 10mm. Seasonal values are derived from the monthly values.

There are various methods for interpolation of climatological parameters; a good overview is given by Hengel (2007). Almost all methods involve detrending the data; that is, removing the dependency on geographical effects by regression analysis, and customising the regression analysis model to the parameter and month. This method produces a trend that is dependent on regression variables only (and can be calculated at each grid point), and also a residual for each data point. The regression residuals are then interpolated onto the grid and the regression model trend is evaluated at each grid point and added to the interpolated residual field to produce the final LTA product.

Table 4 shows the independent variables used in the regression analysis. The interpolation was carried out using the R packages gstat (IDW) and geoR (kriging). Kriging uses a weighted average of neighbouring samples to estimate the 'unknown' value at a given location; the weights are optimised using the semi-variogram model. The technique also provides a "standard error", which may be used to quantify confidence levels. It is used when data points show spatial correlation.

The gridding techniques were applied to the whole Island of Ireland. This was made possible by the provision of LTA data for stations in Northern Ireland by the UKMO. These data were produced using similar techniques to those described in this paper (Perry, 2005; Hollis D., private communication to S. Walsh, 2012).

Rainfall data were available for over 1100 locations. The regression was carried against the variables as shown in Table 4 using stepwise regression of a linear model. Only one of the parameters of Coast and Distance to Sea was used, chosen as follows: stepwise regression was carried out using each value of these parameters in turn in the model, the parameter that yielded the highest R^2 value was used in the final regression model.

Independent Variable	**Use**	**Purpose**
Easting and Northing	1st, 2nd order and cross product	To capture spatial trends
Elevation	Elevation above mean sea level	To model elevation
Elevation in station vicinity	Mean altitude in a 5km radius offset by 10km to N, NE, E, SE, S, SW, W, NW	To model terrain effects
Coast	Percentage of land within a 5,10,15,20 or 25km radius	To model coastal effects (best fitting value)
Distance to Sea	Cartesian distance to nearest coast	To model coastal effects

Table 4: Summary variables used in the rainfall regression model.

5.1 Annual Average Rainfall

The Annual Average Rainfall (AAR) grid was produced first. In deriving the variogram model any residuals greater than 2.5 times the standard deviation of the residuals were omitted. However, they were included in the interpolation process and a small number of inconsistent outliers were removed during this process. The regression model trend was evaluated at each grid point, and this was then added to the residuals interpolated to each grid point to produce the final rainfall field (Figure 5).

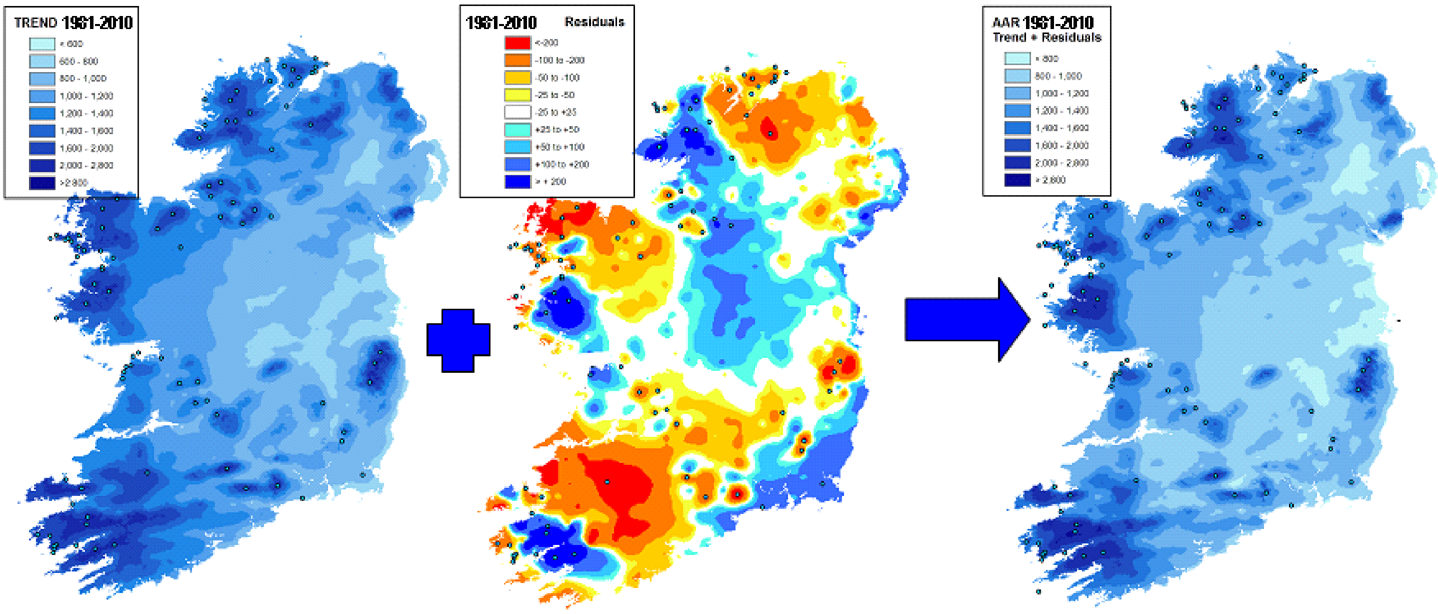

Figure 5: The AAR regression trend is added to interpolated residuals to give the final AAR field.

5.2 Monthly Rain LTAs

When the AAR grid had been produced, the monthly values (Figure 6) were obtained as follows: monthly values were normalised by dividing by the AAR, the regression model was applied to the resulting normalised monthly rain, the residuals interpolated by kriging and added to the regression model trend evaluated at each point and this was then de-normalised by multiplying by the AAR.

Finally, the sum of all monthly totals was constrained to equal the AAR at each grid point. This approach ensured consistency between the monthly and annual totals and enabled the generation of monthly LTAs for stations with annual totals only (i.e. the mountain gauges).

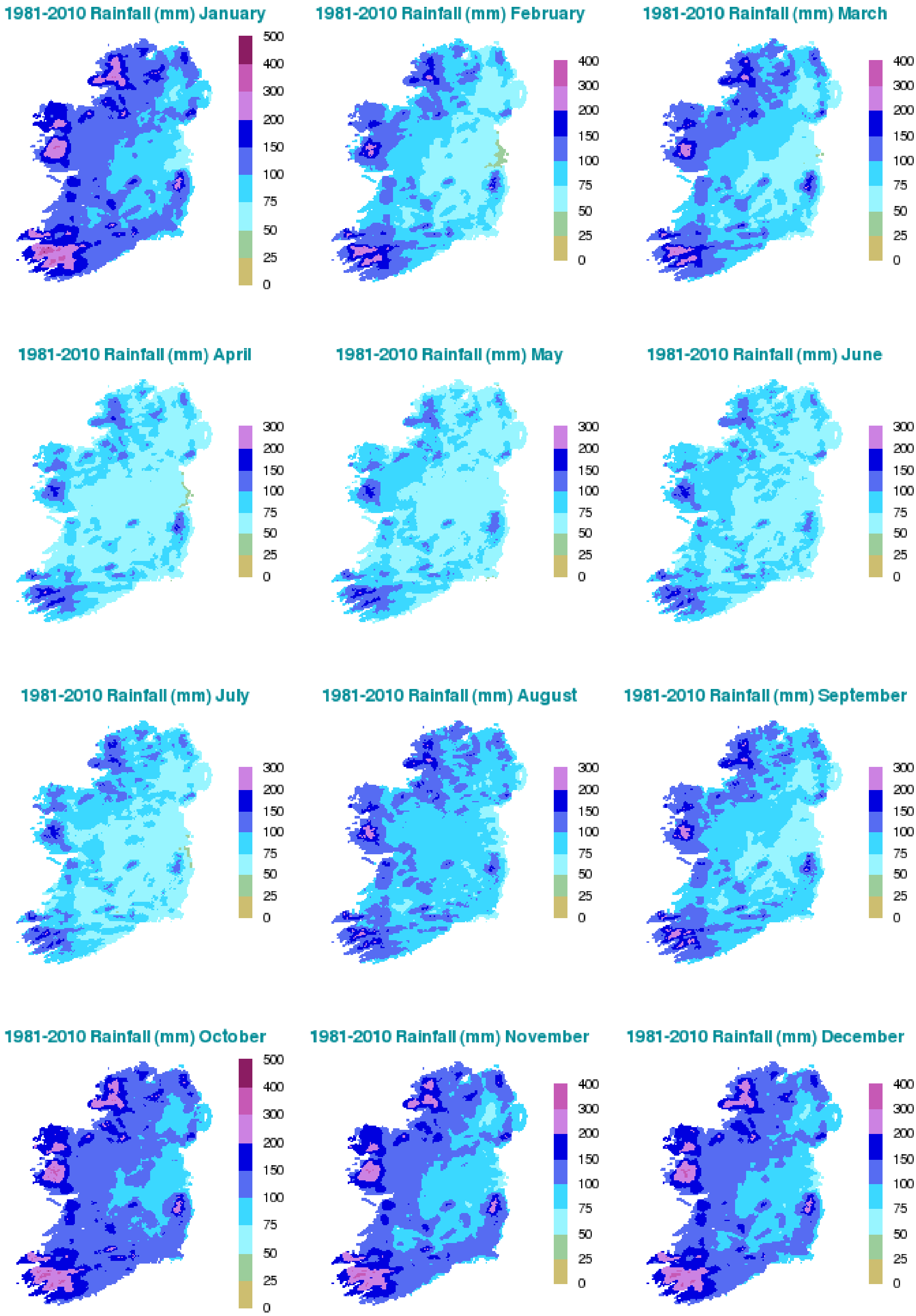

Figure 6: Monthly Rainfall Averages 1981-2010.

5.3 Verification of Gridding

Interpolation of monthly and annual rainfall was verified by the leave-one-out cross-validation method (LOOCV). Each observation is left out and its value predicted using the remaining members of the dataset. The verification statistics for monthly and annual rainfall are shown in Table 5. RMSEr is the normalised RMSE, the RMSE of the prediction errors divided by the total variation. As a rule of thumb a value of RMSEr of 0.4 or less means 85% or more of the variation is accounted for by the regression-kriging model (Hengl, 2007).

	ERR	**MAE**	**RMSE**	**RMSEr**
JAN	-0.21	6.55	9.41	0.23
FEB	-0.15	4.60	6.59	0.21
MAR	-0.13	5.10	7.17	0.24
APR	-0.09	3.47	4.86	0.26
MAY	-0.08	3.55	4.90	0.30
JUN	-0.12	3.74	5.15	0.35
JUL	-0.09	4.21	5.82	0.32
AUG	-0.16	4.86	6.64	0.31
SEP	-0.12	4.52	6.42	0.26
OCT	-0.23	5.83	8.19	0.23
NOV	-0.28	5.78	8.13	0.22
DEC	-0.25	6.11	8.55	0.22
ANN	-1.30	59.31	92.74	0.22

Table 5: Verification statistics for rainfall **ERR**: Mean Error, **MAE**: Mean Absolute Error, **RMSE**: Root Mean Squared Error, **RMSEr**: Normalised RMSE.

5.4 Days of Rain

The production of ‘days of rain’ parameters requires a complete daily record. Where stations were missing complete months of data, the estimated monthly total was treated as a cumulative total. The procedure previously outlined for redistribution of cumulative totals was applied to generate daily values. The data were fitted to a regression model. Due to the absence of spatial correlation in the residuals, the residuals were interpolated by Inverse Distance Weighted Interpolation (IDW). The trend was evaluated at each grid point and the interpolated residuals added to give the final gridded field. Annual ‘days of rain’ totals are shown in Figure 7.

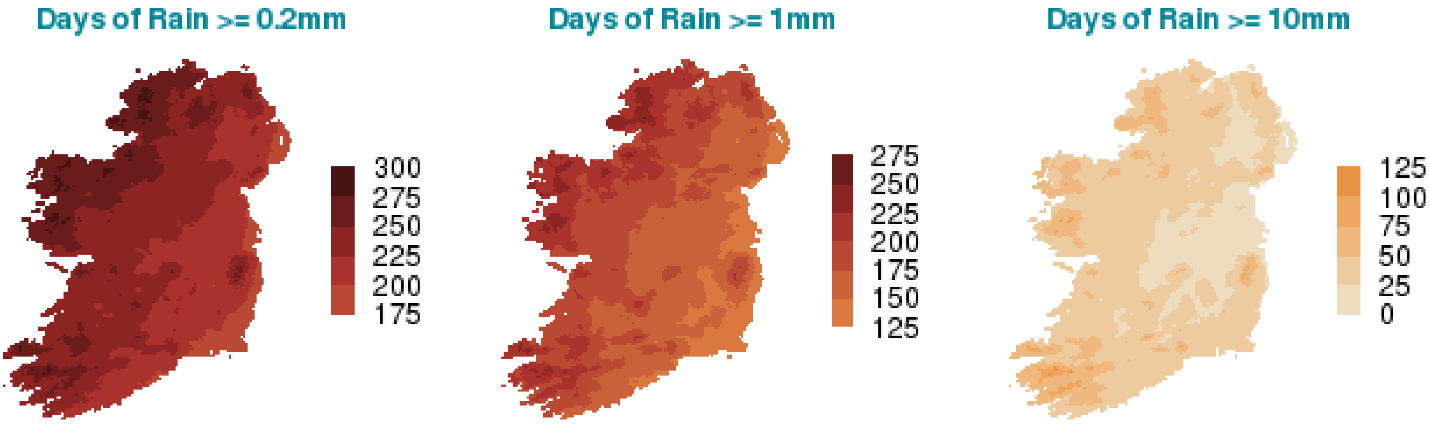

Figure 7: Annual 'Days of Rain' Totals 1981-2010 for >= 0.2mm, >= 1mm and >= 10mm.

6. Comparison with 1961-1990 LTAs

The 1981-2010 LTAs were compared with the 1961-1990 LTAs that had previously been in use. Comparisons are now made with the new LTA period. To enable such a comparison, the datasets covering both periods need to be consistent. Station LTAs were generated for the 1961-1990 period using the same methods as for the 1981-2010 period, including quality control and infilling methods. Comparisons were made between stations that had over 50% coverage in both LTA periods for monthly seasonal and annual rainfall totals (ROI stations only; approximately 250 stations). Average Annual Rainfall showed an average increase of approximately 6%, generally ranging from 2% or 3% in the East to 8% or 9% in the West. The changes in annual, monthly and seasonal mean rainfall are shown in Table 6.

	Annual	**Winter**	**Spring**	**Summer**	**Autumn**
%	105.9	102.2	106.6	110.3	104.5

Table 6: 1981-2010 Annual and Seasonal rainfall as a % of 1961-1990

Smoothed maps derived from the station data are shown in Figures 8, 9 and 10. All months, with the exception of September, show an average increase in rainfall. The month with the largest increase is July (17%), there are also significant regional differences in some months, e.g. February. Seasonal differences reflect the monthly with all seasons on average recording an increase, but the South and East record decreases in winter. See Table 7.

	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
%	100.4	103.2	105.3	112.5	102.1	112.4	116.6	102.9	96.0	109.5	108.9	101.3

Table 7: 1981-2010 Monthly rainfall as a % of 1961-1990

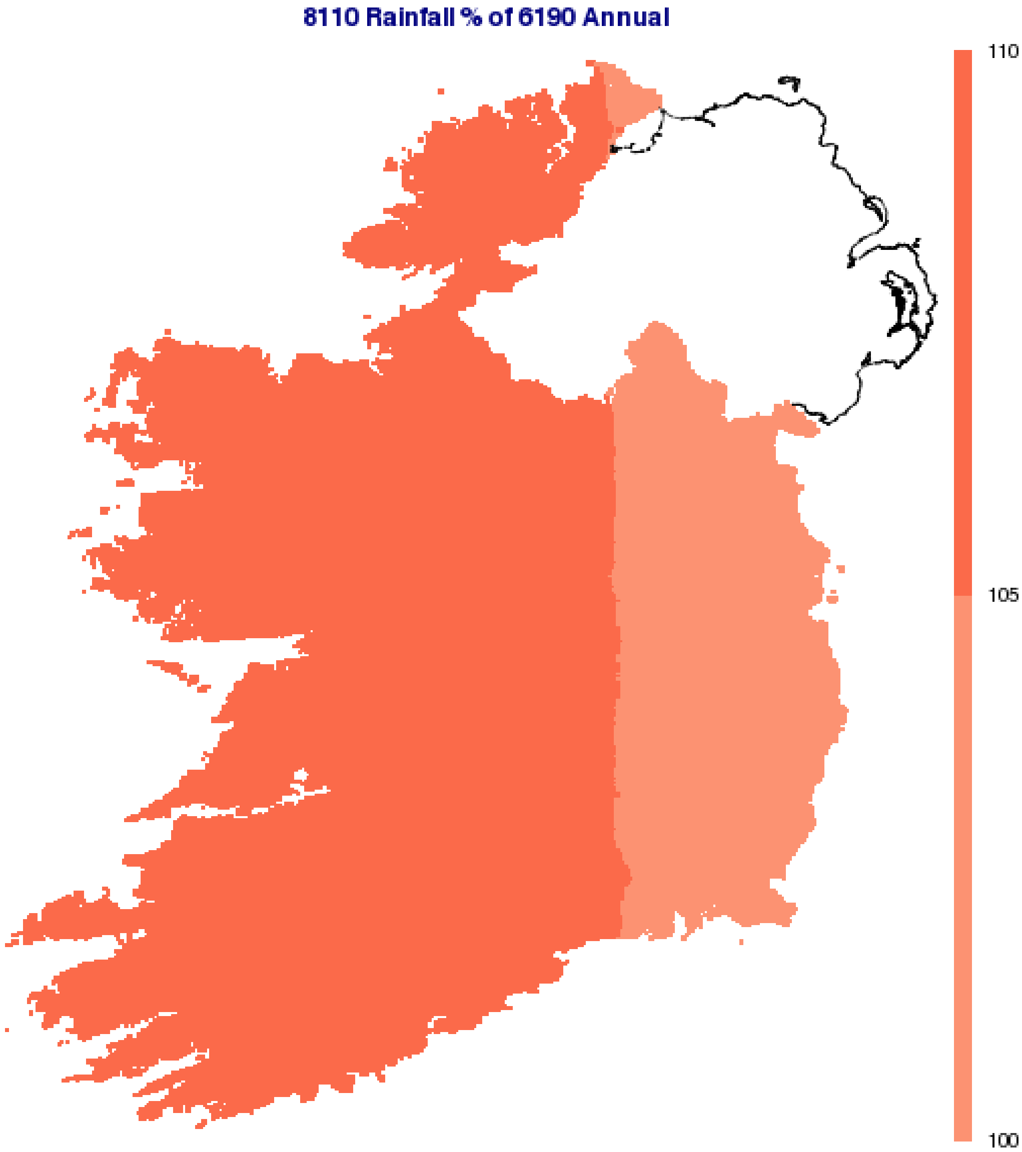

Figure 8: Annual Rainfall 1981-2010 as a % of 1961-1990

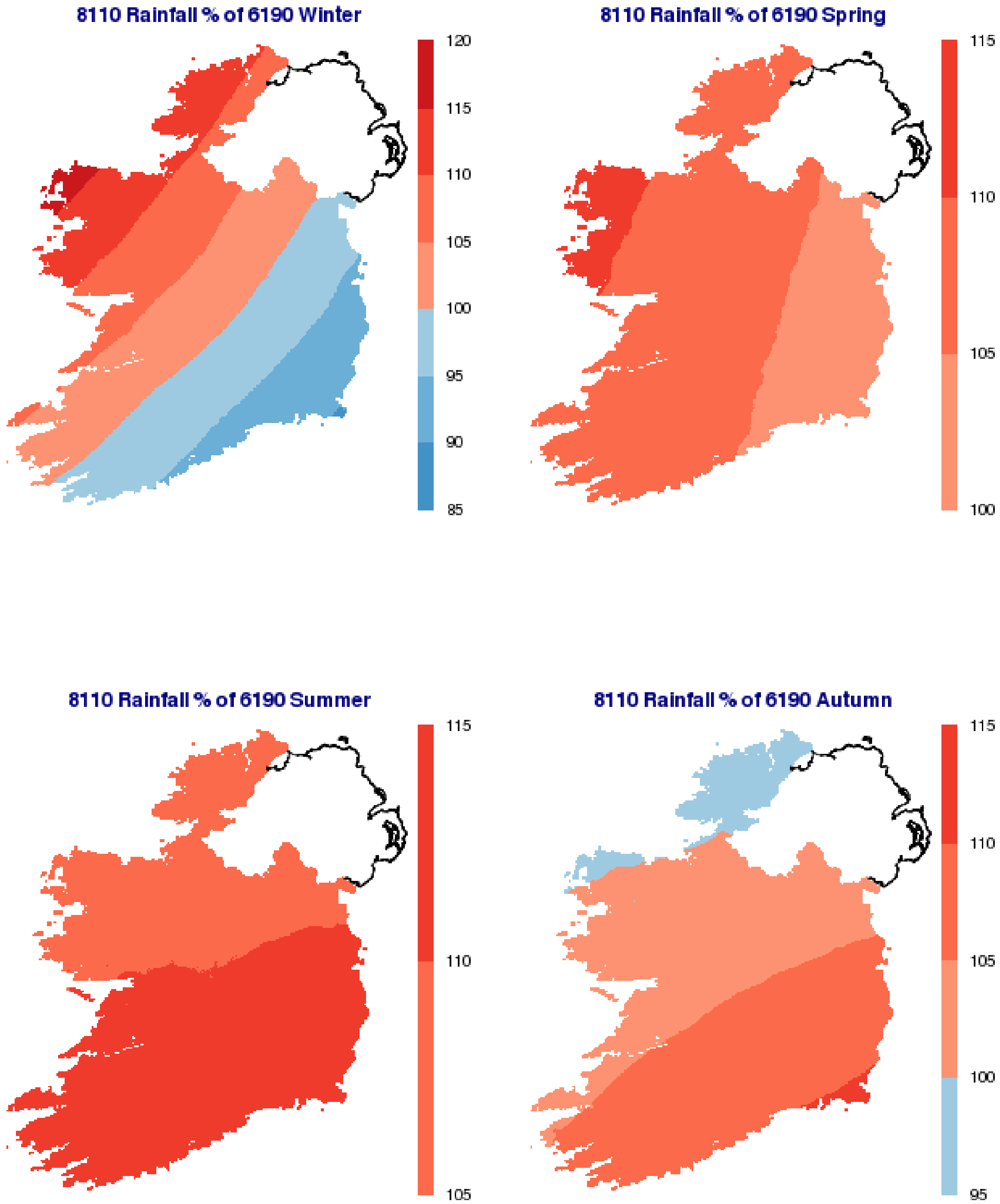

Figure 9: Seasonal Rainfall 1981-2010 as % of 1961-1990.

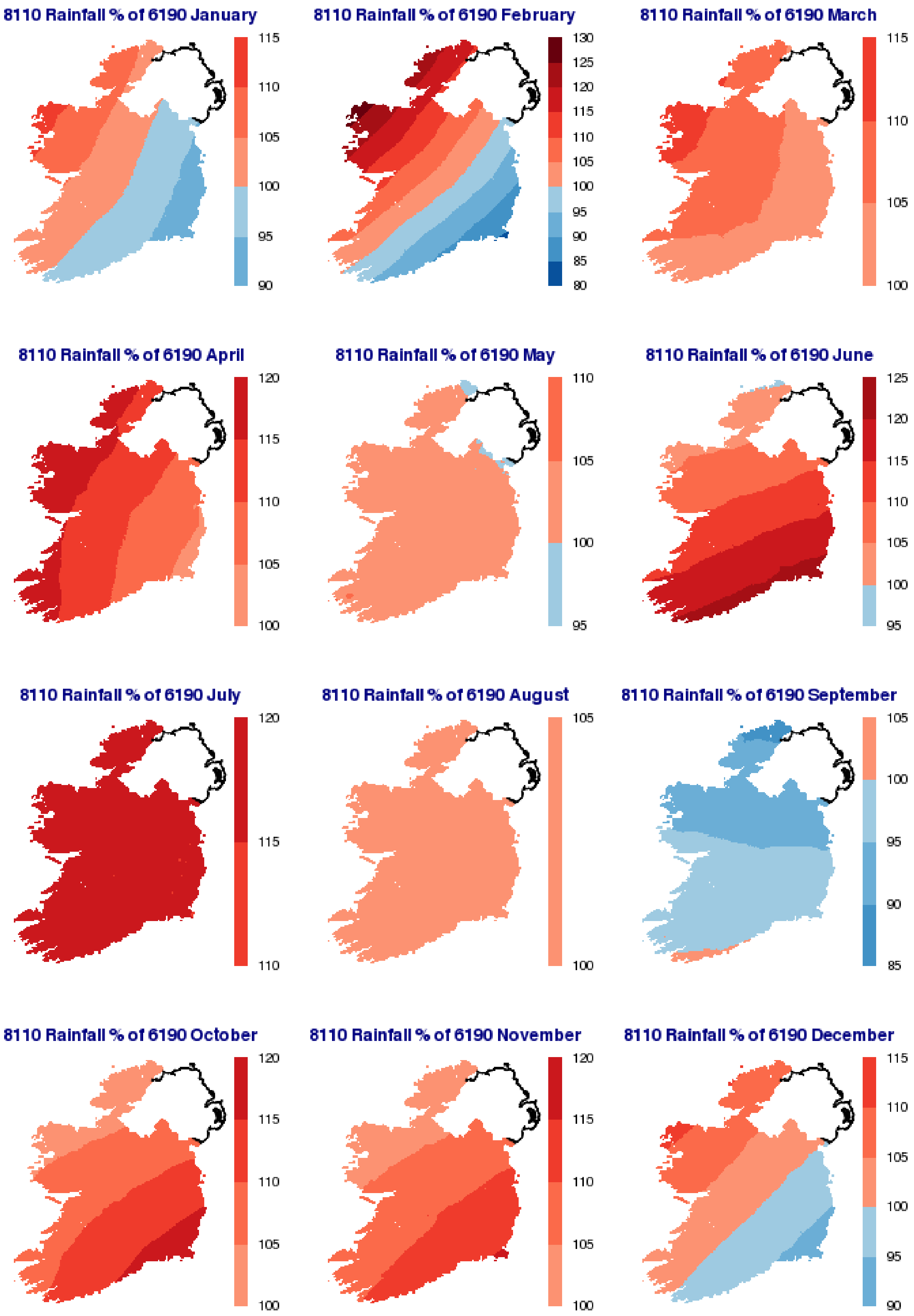

Figure 10: 1981-2010 Monthly Rainfall as a % of 1961-1990.

7. Summary and Conclusions

LTAs of monthly, seasonal and annual rainfall totals, and days of rain greater than 0.2mm, 1mm and 10mm have been generated for approximately 750 rainfall station in the Republic of Ireland for the 30-year period 1981-2010. These data have been combined with the corresponding data for Northern Ireland to produce gridded datasets for the Island of Ireland at 1km^2 resolution using geostatistical methods. Comparisons with the 1961-1990 station LTA point to an approximately 6% increase in average annual rainfall between the two averaging periods with corresponding increases in most months and seasons, although there are some regional differences. Verifications statistics for the infilling and gridding techniques indicate that the datasets are of high quality and can be used with confidence.

8. Acknowledgements

Particular thanks to the Met Éireann climate and rainfall voluntary observers without whom this work would not have been possible. Thanks to my colleagues in Met Éireann for valuable help, assistance and advice. Thanks also to John Prior and Dan Hollis of the United Kingdom Meteorological Office for provision of data for Northern Ireland.

9. References

Fitzgerald D. and Forrestal, F. (1996) *Monthly and annual averages of rainfall for Ireland 1961-1990*. Climatological Note No. 10, Dublin: Meteorological Service. Available at: http://hdl.handle.net/2262/70486

Fitzgerald D.L., (2007). *Estimation of point rainfall frequencies*. Met Éireann Technical Note 61, Dublin: Met Éireann.: Dublin. Available at: http://hdl.handle.net/2262/70546

Hengl, T., (2007) A *Practical guide to Geostatistical Mapping of Environmental Variables*. EUR 22904 EN Scientific and Technical Research Series, Luxembourg: Office for Official Publications of the European Communities.. Available at: http://eusoils.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/other/EUR22904en.pdf

Hubbard, K. G., S. Goddard, W. D. Sorensen, N. Wells, and T. T. Osugi (2005) Performance of Quality Assurance Procedures for an Applied Climate Information System, *Journal of Atmospheric and Oceanic Technology,* 22,105-112. Available at: http://cse.unl.edu/~goddard/Papers/Journals/JAOT-Jan05.pdf

Ordnance Survey Ireland, 1996. *The Irish Grid:* a description of the co-ordinate reference system used in Ireland. Dublin: OS*i*. Available at: http://www.osi.ie/wp-content/uploads/2015/04/The-Irish-Grid-A-Description-of-the-Coordinate-Reference-System-Used-in-Ireland.pdf .

Perry, M. and Hollis, D. (2005). The development of a new set of long-term climate averages for the UK, *International Journal of Climatology*, 25, 1023–1039. Available at: 10.1002/joc.1160

R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.. Available at: http://www.R-project.org.