Earth Syst. Sci. Data Discuss., 8, 83–112, 2015
www.earth-syst-sci-data-discuss.net/8/83/2015/
doi:10.5194/essdd-8-83-2015

Open Access
Earth System
Science
Data
Discussions

CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological use

V. D. J. Keller, M. Tanguy, I. Prosdocimi, J. A. Terry, O. Hitt, S. J. Cole, M. Fry, D. G. Morris, and H. Dixon

NERC Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxon, OX10 8BB, UK

Received: 11 December 2014 – Accepted: 2 January 2015 – Published: 27 January 2015

Correspondence to: M. Tanguy (malngu@ceh.ac.uk)

Published by Copernicus Publications.

Abstract

The Centre for Ecology & Hydrology – **G**ridded **E**stimates of **A**real **R**ainfall (CEH-GEAR) dataset was developed to provide reliable 1 km gridded estimates of daily and monthly rainfall for Great Britain (GB) and Northern Ireland (NI) (together with approximately 3500 km^2 of catchment in the Republic of Ireland) from 1890 onwards. The dataset was primarily required to support hydrological modelling.

The rainfall estimates are derived from the Met Office collated historical weather observations for the UK which include a national database of raingauge observations. The natural neighbour interpolation methodology, including a normalisation step based on average annual rainfall, was used to generate the daily and monthly rainfall grids. To derive the monthly estimates, rainfall totals from monthly and daily (when complete month available) read raingauges were used in order to obtain maximum information from the raingauge network. The daily grids were adjusted so that the monthly grids are fully consistent with the daily grids. The CEH-GEAR dataset was developed according to the guidance provided by the British Standards Institution.

The CEH-GEAR dataset contains 1 km grids of daily and monthly rainfall estimates for GB and NI for the period 1890–2012. For each day and month, CEH-GEAR includes a secondary grid of distance to the nearest operational raingauge. This may be used as an indicator of the quality of the estimates. When this distance is greater than 100 km, the estimates are not calculated due to high uncertainty.

CEH-GEAR is available free of charge for commercial and non-commercial use subject to licensing terms and conditions. doi:10.5285/5dc179dc-f692-49ba-9326-a6893a503f6e.

1 Introduction

Estimates of areal daily or monthly rainfall over extended periods are often required for hydrological purposes such as catchment management of water resources (e.g.

Young et al., 2003), catchment modelling (e.g. Bell et al., 2013; Young et al., 2006), peak flow estimation (e.g. Prosdocimi et al., 2014), and groundwater recharge (e.g. Sorensen et al., 2014). More widely, they are required by a variety of disciplines, for example to model or explain processes such as the atmospheric deposition of nitrogen in geosciences (Dore et al., 2012) and the relationship between rainfall and cholera in epidemiology (Eisenberg et al., 2013).

In the UK, point measurements of daily and monthly rainfall data have been collected using standardised storage raingauges since the late 19th century (Burt, 2010; Eden, 2009). Here rainfall is defined as total precipitation which is the sum of liquid precipitation plus the liquid equivalent of any solid precipitation (Met Office, 2012a) and is in accordance with the British Standards Institution (BS 7843-4:2012), UK convention for areal rainfall calculations. The UK network of raingauges grew from around 450 in 1860 to approximately 3500 by 1900 and peaked at around 6250 in 1974 (Eden, 2009). By 2009, data were recorded at 3285 sites (Burt, 2010). While the current national raingauge network is dense in global terms, the resulting rainfall information is limited to a set of discrete points in space and time. Practical considerations, such as those relating to the suitability of sites and the cost of maintaining the network, mean that there is considerable spatial variation in the density of the network. Nevertheless, interpolation techniques can then be used to provide rainfall estimates across a continuous area based on the rainfall data collected.

The Met Office has developed a method for generating 5 km grids of daily, monthly and annual estimates of rainfall for the UK from 1961 onwards (Perry and Hollis, 2005; Perry et al., 2009). However for hydrological purposes there are often requirements for finer spatial resolutions to model river flows accurately at a catchment level (Bell et al., 2013; Cole and Moore, 2008; Young et al., 2006), as well as longer time series to allow assessment of hydrological change (in particular daily data prior to 1961 when computer-held raingauge data are less prevalent). Spatial rainfall fields, represented as daily 1 km grids are required for the estimation of catchment average rainfall time series for input into generalised rainfall–runoff models. As the optimisation of parameters of any model will tend to compensate for measurement error within both the input data and the calibration flow data, it is essential that the methods used for estimating rainfall data are accurate and consistent in approach for both calibration and subsequent application.

The aim of this paper is to outline the development of the Centre for Ecology & Hydrology – **G**ridded **E**stimates of **A**real **R**ainfall (CEH-GEAR) dataset, a 1 km daily and monthly rainfall dataset for Great Britain (GB) and Northern Ireland (NI) (together with approximately 3500 km^2 of catchment in the Republic of Ireland) for the period 1890–2012. A description of the data (Sect. 2) used to generate this dataset is presented followed by the rainfall interpolation method (Sect. 3). Quality control of the daily rainfall data is described (Sect. 4) and validation results of the gridded rainfall estimates presented using an independent raingauge network over Scotland (Sect.5). Finally, some recommendations are provided regarding the use and limitations of this dataset.

2 Data

2.1 Raingauge rainfall observations

The aim of the CEH-GEAR dataset is to produce temporally consistent areal rainfall data for as long a period as possible. This dataset makes use of the Met Office collated historical weather observations for the UK, specifically the daily and monthly rainfall accumulations (liquid precipitation plus the liquid equivalent of any solid precipitation (Met Office, 2012b)) from a national network of raingauges (Met Office, 2012a). These rainfall data are collected by a range of organizations from an irregularly spaced and constantly evolving network of manual and automated raingauges (Eden, 2009). For the period 1961–2000, there is an average of one rainfall station per 49 km^2 (4400 stations) (Perry and Hollis, 2005), with the peak density occurring in 1974. While the UK raingauge network expanded rapidly during the late 19th and early 20th century, only a limited proportion of the pre-1961 data is currently available in digital form. The national database contains records of rainfall accumulations over a range of durations, however this paper will focus on daily and monthly accumulations from both manual and automated raingauges. Maps of all daily and monthly raingauges used to generate the CEH-GEAR dataset are presented in Fig. 1.

When developing and using spatially aggregated rainfall data based on raingauge observations, it is important to consider the uncertainties in the source measurements. Extensive international trials have shown that the main sources of error in raingauge measurement include those due to adhesion of water to the gauge surface, in- and out-splash, wetting and evaporation. However, the largest source of error is caused by the wind around the raingauge, leading to a systematic underestimation of the rainfall amount (Rodda and Dixon, 2012). Indeed long-term trials have shown that, for the UK standard Met Office Mk2 raingauge (British Standards Institute, 2011a), these errors lead to significant systematic undercatches of around 5 % in the estimation of average annual rainfall, a figure that can rise to 16 % in highly exposed areas (Rodda and Smith, 1986). While alterations to the siting of gauges, for example by locating rims at ground level, can reduce undercatch, this is not systematically done within the UK. The high spatial and temporal variation in the degree of underestimation means that data held in the national archive cannot be routinely corrected for under-catch.

The magnitude of errors in rainfall fields derived from point measurements is mainly a function of the local density of the raingauge network. The meteorological forcing is also important: errors are likely to be smaller for frontal rainfall than for thunderstorms or localised showers associated with warm sector weather.

2.2 Standard period Average Annual Rainfall (SAAR)

The distribution of raingauges across the UK is not uniform. Many stations are situated in locations of easy access, and often near population centres which tend to be lower in altitude and therefore dryer (British Standards Institute, 2011b). Thus, to avoid a downward bias in the gridded rainfall estimates, there is a need to normalise the raingauge

Title Page
Abstract | Instruments
Data Provenance & Structure
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

rainfall totals before interpolation, and the most suitable available variable for this is average annual rainfall (AAR).

The version used for Great Britain (GB) was the Met Office 1 km grid for the 1961–1990 standard period (SAAR (61–90)). This was developed by Spackman (Spackman, 1993) by deriving grid point values of average annual rainfall values for a 10 km grid using monthly data from approximately 13 100 raingauges. These values were then contoured and gridded at a 1 km resolution using a bi-cubic spline interpolation procedure.

For Northern Ireland (NI), the Met Eireann 1 km grid of 1961–1990 long term average rainfall was used (Walsh, 2012a). This dataset, which covers the whole of Ireland, has been derived from raingauge observations, using a regression analysis (Walsh, 2012b).

2.3 Weather radar rainfall estimates

Over recent decades, weather radars have played an increasingly important role in areal rainfall estimation, particularly in real-time applications. Weather radars can give good qualitative estimates of rainfall across extensive areas at fine spatial and temporal resolutions (e.g. 1 km and 5 min resolution for the UK) and data are usually available within minutes of the observation time. As a consequence, a major use is for flood forecasting where radar can detect the location, extent and evolution of convective storms that raingauge networks rarely sample well, if at all. The UK weather radar network has only been operational since 1985, when it was launched with just four radars (Kitchen and Illingworth, 2011). Since its inception there have been many changes to the radar processing that have improved the quality of the rainfall estimates and the UK network coverage has now expanded to 15 radars (Kitchen and Illingworth, 2011).

However, raingauges still provide more accurate quantitative rainfall estimates at a point and are the only option for generating long-term timeseries of areal rainfall. Whilst merging radar and raingauge information to form historical daily or monthly totals has the potential to provide improved areal rainfall estimates, radar data have not been used in the production of the current version of CEH-GEAR. This is in part due to the comparatively short duration available for the radar rainfall estimates (~ 30 years) compared to the raingauge based observations. It was therefore considered that CEH-GEAR would have greater temporal consistency if it was based solely on raingauge data.

3 Areal rainfall estimation procedure

3.1 Introduction

Areal rainfall methods seek to represent the spatial distribution of rainfall over a catchment, a region, or even a country. Within CEH-GEAR, a grid interval of 1 km was chosen as this aligns to the resolution of the available SAAR grids used for normalisation and because there are few locations in the UK where the raingauge density is sufficient to justify a finer resolution.

There are many spatial interpolation methods available, however they all have specific features and therefore are not suitable to all environmental datasets (Li and Heap, 2011, 2008). There are four principal categories of procedures for estimating the rainfall at each grid point. Although all of the procedures may be applied directly to the gauged values, it is generally recommended that they are applied to values that have been normalised by SAAR (British Standards Institute, 2011b), as discussed in Sect. 2.2.

The first category is termed the domain method, where each operational raingauge is considered to represent a contiguous area of the surrounding surface (referred to as domain), and each grid point in that domain is allocated the rainfall recorded at the raingauge. Domains are most commonly defined on the basis of proximity, and this type of estimation of point values is known as nearest neighbour interpolation. This is the basis of the well established Thiessen procedure for areal rainfall estimation (Thiessen, 1911). A serious drawback within this type of approach is the presence of discontinuities at the edges of domains: this is of particular concern when using the grid

Title Page
Abstract | Instruments
Data Provenance & Structure
Tables | Figures
I◀ | ▶I
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

to estimate areal rainfall in small catchments with an area of a similar spatial resolution as the raingauge domains.

The second category involves the fitting of a mathematical surface to the observations from a selection of local gauges. The two main drawbacks of this approach are the risk of unjustifiable or unrealistic extrapolation, and sensitivity to the selection procedure: discontinuities can arise where a gauge with a particularly low or high observation drops in or out of the local selection.

The third category involves the fitting of a mathematical surface to the observations from all gauges, and computing the value at every grid point from this surface. This is computationally impractical for the large area and number of raingauges applicable to CEH-GEAR.

Within the fourth category, grid point or areal rainfall at time t ($\overline{R_t}$), is estimated as a weighted average of the rainfall observations from a selection of local gauges:

$$\overline{R_t} = \sum_{i=1}^{n} w_i r_{i,t} \qquad (1)$$

where n is the number of gauges, w_i is the weight applied to raingauge i ($w_i \in [0,1]$) and $r_{i,t}$ is the observed rainfall depth from raingauge i at time t.

The British Standards Institute "Guide to the acquisition and management of meteorological precipitation data" (British Standards Institute, 2011b) recommends a set of such interpolation techniques, including the triangular planes method (Jones, 1983), the natural neighbour interpolation also called Voronoi interpolation (Gold, 1989; Ledoux and Gold, 2005; Sibson, 1981), and kriging (Delhomme, 1978; Webster and Oliver, 2001). The natural neighbour method was selected for CEH-GEAR as it produces smooth rainfall surfaces without the boundary discontinuities that occur between adjacent polygons in the Thiessen polygon method, and, it is relatively simple to implement.

3.2 CEH-GEAR interpolation method

A schematic of the interpolation methodology used to derive daily and monthly 1 km grids for the UK is presented in Fig. 2. The grids are generated using the natural neighbour interpolation methodology, including a normalisation step based on average annual rainfall. Note that the derivation of the daily grids involves two stages: an initial estimate from daily gauges alone, followed by multiplication by a correction grid to give consistency with monthly grids that have been derived from all available gauged data – daily and monthly. This is discussed further in Sect. 3.3.

The natural neighbour interpolation method is a development of the Thiessen approach (Gold, 1989; Ledoux and Gold, 2005; Sibson, 1981). First, for each operational raingauge i at time step t, its Thiessen polygon $T_{i,t}$ is defined: this is the polygon within which no other operational gauge is closer. Traditionally this was derived manually by connecting the perpendicular bisectors of the lines connecting neighbouring gauges. In the automated grid-based implementation used here, it is approximated by the set of grid points for which no other gauge is closer.

Then, for each grid point α, the Thiessen polygons are reconstructed ($\hat{T}$) treating the grid point as an additional gauge. The grid point then possesses its own Thiessen polygon $\hat{T}_{\alpha,t}$ at time step t, which overlaps part of the original Thiessen polygons ($T_{i,t}$) for the neighbouring raingauges (only one in the case a raingauge being coincident with the grid point). Each raingauge i at time t that has part of its original Thiessen polygon $T_{i,t}$ overlapped by the Thiessen polygon for the grid point ($\hat{T}_{\alpha,t}$) is included in the rainfall interpolation at the grid point α and the weight associated with raingauge i is proportional to the area of overlap: Area($T_{i,t} \cap \hat{T}_{\alpha,t}$). The natural neighbour weight ($w_{i,t}(\alpha)$) of a neighbouring raingauge i, when interpolating at point α at time t is:

$$w_{i,t}(\alpha) = \frac{\mathrm{Area}(T_{i,t} \cap \hat{T}_{\alpha,t})}{\mathrm{Area}(\hat{T}_{\alpha,t})} \tag{2}$$

A schematic illustrating the natural neighbour method is provided in Fig. 3. In automated grid-based implementation, the areas are approximated by the number of grid points contained in the polygon. Whilst estimating the monthly grids, all monthly rainfall observations and daily data from raingauges with a full month recorded, are used to construct the Thiessen polygons.

The estimated rainfall for a grid point α, at time t ($\mathrm{rc}(\alpha,t)$), is then derived using the natural neighbour interpolation and SAAR (61–90) normalised rainfall:

$$\mathrm{rc}(\alpha,t) = \mathrm{SAAR}(\alpha)\sum_{i=1}^{n} w_{i,t}(\alpha)\frac{r_{i,t}}{\mathrm{SAAR}_i} \tag{3}$$

where SAAR_i and SAAR_α are the SAAR values at raingauge i and grid point α respectively. It is worth remarking that the weights sum to one and are independent of the actual raingauge rainfall totals so only need to be recalculated when a raingauge drops in or out of the network.

As the selected grid point α moves away from a particular raingauge (but within the domain of the raingauge network), the weight for the gauge diminishes gradually to zero until it is no longer a natural neighbour. Therefore the natural neighbour interpolation method provides a gradually varying surface, unlike the Thiessen approach which consists of a series of plateaux with sharp edges between them. Nevertheless, it should be noted that the method can give rise to discontinuities in gradient at gauge locations, although these are of minor concern for areal rainfall applications.

The natural neighbour interpolation method, although more computationally demanding than the triangular planes method, makes greater use of the locally available data as it uses all neighbouring recording gauges instead of only three. Importantly, this method is less computationally demanding than kriging methods whilst providing comparable interpolation results; the main difference is that kriging provides a map of the standard error statistic of the gridded rainfall estimates.

3.3 Monthly correction procedure

The same interpolation methodology is applied to derive daily grids and monthly grids. However the raingauge network, and therefore data, used may be different: the daily grids are derived based on daily raingauges only whereas the monthly grids make use of both the monthly raingauges and the daily raingauges with complete record for the month. Although the monthly grids may be more reliable, due to the higher number of raingauges used, the consequence is that the gridded monthly estimates and the monthly totals based on daily grid estimates may differ. Thus a correction step was added, after the creation of the monthly grids and the provisional daily grids, to ensure that the monthly sum of daily rainfall depth matches the estimated monthly depth (Fig. 2). For a given month, when all daily grids are estimated from interpolating the daily raingauge data (provisional daily grids), these estimates are summed up to provide a monthly estimate from daily data ($\mathrm{MR_d}$). For each grid point, this estimate ($\mathrm{MR_d}$) is compared with the monthly gridded value ($\mathrm{MR_m}$). To ensure that the daily grids and the monthly grids are in agreement, a correction factor $\frac{\mathrm{MR_m}}{\mathrm{MR_d}}$ is applied to each daily point estimate for the month.

3.4 Calculation thresholds

The accuracy of the rainfall estimates is affected by the density of the raingauge network and the distance to the closest raingauges. For the pre-1961 grids there was a concern that the lower density of digitised raingauge data would give rise to unrepresentative estimates in some locations that were a long way from any raingauge. It was therefore decided not to compute a rainfall estimate when a grid point was more than 100 km from the nearest operational raingauge. The effect of this threshold varies according to the availability of digitised raingauge data: for example, out of a total of 244 343 UK grid points, the number of points excluded on 1 January 1890, 1910 and 1960 were respectively 46 394, 20 604 and 34 (Fig. 4a–c). From 1961 onwards, the 100 km threshold has virtually no effect, with only some remote Scottish islands affected on isolated days (Fig. 5). In order to provide users – especially modellers – with the spatial and temporal extend of gaps, two sets of three ancillary grids were produced (one set for monthly data and one for daily data):

- year of the first missing data for each grid point,
- year of the last missing data for each grid point,
- total number of days with missing data for each grid point for the whole period.

The dataset also contains, for every day and month, a grid of the distance to the closest operational raingauge.

4 Quality control of the input rainfall data

Causes of error in raingauge data include hydrometric and meteorological factors (Sect. 2.1), and human factors such as misreading and typing errors. Rainfall observations held in the national database are subject to extensive quality control by both the raingauge operators and by the Met Office at the point of submission to the archive. A further quality control procedure is applied during the production of the CEH-GEAR dataset to identify erroneous raingauge observations in the daily rainfall input dataset. The procedure was designed to further scrutinise exceptionally high rainfall values by comparing the daily measured rainfall with an estimate of the 200 year return period 1 day rainfall at the gauge location. This estimate was made using the latest Flood Estimation Handbook rainfall depth-duration-frequency model, which is a development of the model documented in Stewart et al. (2010). For the period 1961–2012, there were 687 observations in Great Britain and 34 in Northern Ireland that exceeded the 200 year return period rainfall.

For those high rainfall events exceeding the 200 year return period rainfall, a manual investigation was undertaken to identify whether the extreme rainfall recorded was genuine. The identified high rainfall events were cross-referenced with a historical database of extreme events for the UK for the period 1886–2005 published by Svensson et al. (2009). Those events present in the historical extreme events database were considered to be genuine. Then for each of the remaining events, the raingauge data was investigated using a time series plotter in order to identify likely multiday rainfall accumulations which had not been flagged as such in the historical records. Any high rainfall identified as the result of a multiday accumulation was rejected.

For the remaining events, each selected event was compared with the 3 nearest raingauges stations within a radius of 10 km. In instances where the 3 raingauges were recording 20 % or more of the investigated rainfall event, the event was classified as genuine. Where significantly lower rainfall depth ($< 20\,\%$) was recorded at these neighbouring gauges then the selected rainfall event was considered erroneous and was therefore rejected from the input dataset. Where no decision could be made, then a manual investigation was required and the number of neighbouring raingauges investigated increased (up to 10 within a 10 km radius). Where uncertainty remained the event was classified as genuine, as the recording may be the result of localised rainfall.

5 Validation of the method

The suitability of the natural neighbour method as a daily rainfall interpolation procedure for the UK was assessed using measured rainfall data for the period 2007–2010 from the tipping bucket raingauge network operated by SEPA (Scottish Environment Protection Agency). Scotland was chosen because rainfall interpolation is generally more demanding there, because of the higher spatial variability of rainfall (due to the terrain) and the relatively sparse raingauge network.

The SEPA tipping-bucket network has around 200 raingauges with a resolution (bucket size) of 0.2 mm and provides 15 min rainfall totals for use in real-time flood forecasting (Cranston et al., 2012). An automated quality-control procedure (Howard et al., 2012) has been applied to the data with the aim of removing any major errors that may exist. Simple tests are first performed on each individual raingauge record, before more involved comparisons to neighbours are made. Robust statistics (median and median absolute deviation) form the basis for identifying and removing outliers. To ensure the quality-controlled tipping-bucket records provided an independent source of validation data, a validation subset of 121 tipping-bucket raingauges was used (Fig. 6): from this subset only the days when the tipping bucket was at least 5 km away from any of the daily gauges used to derive the rainfall grids, were retained.

The accuracy of the daily rainfall estimates was assessed by means of:

- Absolute errors (ϵ): absolute value of the difference between the estimated rainfall (rc_p) and the observed values at the gauges (rc_o):

$$\epsilon = |\mathrm{rc}_p - \mathrm{rc}_\mathrm{o}| \tag{4}$$

- Absolute relative errors (δ): ratio of the absolute error and the observed value, absolute relative errors are only computed where $\mathrm{rc}_\mathrm{o} > 0$:

$$\delta = \left| \frac{\mathrm{rc}_p - \mathrm{rc}_\mathrm{o}}{\mathrm{rc}_\mathrm{o}} \right| \tag{5}$$

The repartition of the absolute error ϵ (Eq. 4) across several ranges of observed events (rc_o) was analysed (Table 1). Overall ϵ is equal to 0 in about 25 % of the cases, and smaller than 0.5 mm in approximately 57 % of the cases: an encouraging result. For smaller events (i.e. $\mathrm{rc}_\mathrm{o} < 2$ mm), about 78 % of the absolute errors are ≥ 0.5 mm. For increasing levels of observed rainfall, high values of ϵ are more frequent, although where $\mathrm{rc}_\mathrm{o} \geq 20$ mm (48 % of studied events) ϵ is equal or lower than 5 mm: a relatively small error when compared to the observed rainfall. Indeed, results for the relative absolute error (δ) (Table 2) indicate that although for events of higher intensities ϵ can be quite high, these are still relatively small compared to the actual observed values (low values of δ).

An important influence on the quality of the estimate in the natural neighbour method is the representativeness of the nearby gauges and the density of the raingauge network in the vicinity of the interpolated point. The effect of the network density on ϵ values is significant (Fig. 7). The x axis represents the distance (km) to the closest gauge used in the interpolation procedure (Fig. 7). This information is provided as a simple indicator of the network density, although the number of gauges used in the estimation, the average distance and other network characteristics are also likely to have an effect. This specific analysis also included 18 automatic raingauges, which had been excluded from the error analysis because they were within 5 km of a daily or monthly gauge. The red line (Fig. 7) represents a smoothed estimate of the median function of the absolute relative error obtained by quantile regression: generally, the median increases as the distance to the closest gauge used in the interpolation increases.

The monthly correction procedure (Sect. 3.3) is necessary to ensure the monthly sum of daily estimated rainfall depths and the estimated monthly grids match. Nevertheless, it is preferable that such adjustments have a minimal impact on the daily estimates. For the same Scottish validation gauges used in Table 1, the absolute difference (φ) between the final estimates (est_{mc}), including the monthly correction, and the provisional estimates (est_{pr}) (Fig. 2), obtained from the interpolation of the observed daily measurements, is calculated:

$$\varphi = |\text{est}_{\text{mc}} - \text{est}_{\text{pr}}| \tag{6}$$

Overall for more than 90 % of the cases, φ is less than or equal to 0.5 mm (Table 3): the largest proportion of large differences occurs for higher rainfall events, where a difference larger than 5 mm remains relatively small.

6 Limitations

The CEH-GEAR dataset is derived from daily and monthly raingauge data using the natural neighbour interpolation method combined with a normalisation step based on average annual rainfall. As such, the quality of the rainfall estimates are highly dependent on the accuracy of the raingauge data, hence the need for quality control of the input data. The quality control procedure focussed on high daily rainfall events and identified a set of recorded events that resulted from a multiday accumulation and therefore were discarded from the input dataset. Although measures are in place to flag erroneous raingauge data, some erroneous data may still remain in the underlying data. However, the Met Office national database of raingauge observations (Met Office, 2012a) remains the most appropriate and abundant source of rainfall observation from which to derive gridded time series of daily and monthly rainfall in the UK.

The density of the raingauge network in the vicinity of a grid point is also an important factor when assessing the quality of the rainfall estimates (Sect. 5). Users should be aware that the uncertainties in these gridded estimates will generally increase with increasing distance to the closest gauge used in the interpolation and, depending on their application of the data, should make appropriate use of the associated grids of closest gauge distance. Only a fraction of the pre-1961 raingauge data is available in digital form (Sect. 2.1), digitising the rest of the data would considerably improve the CEH-GEAR rainfall estimates for the period 1890–1960.

7 Data Access and Terms of use

The Centre for Ecology & Hydrology – **G**ridded **E**stimates of **A**real **R**ainfall (CEH-GEAR) dataset is available from http://doi.org/10.5285/5dc179dc-f692-49ba-9326-a6893a503f6e. The data will be hosted on a THREDDS server managed by CEH-Lancaster. The following citation should be used for every use of the data:

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D. G., Keller, V. D. J. (2014). Gridded estimates of daily and monthly areal rainfall for the UK (1890–2012) [CEH-GEAR]. NERC-Environmental Information Data Centre http://dx.doi.org/10.5285/5dc179dc-f692-49ba-9326-a6893a503f6e.

The dataset is available for download free of charge from the CEH Information Gateway. License terms apply.

Acknowledgements. The CEH-GEAR dataset is based on UK raingauge data provided under license by the Met Office, and those organisations contributing to this national dataset (including the Met Office, Environment Agency, Scottish Environment Protection Agency (SEPA) and Natural Resources Wales) are gratefully acknowledged. Richard Maxey, Michael Cranston and colleagues from SEPA are thanked for arranging access to the Scottish tipping bucket raingauge data for validation purposes. Last but not least, Nuria Bachiller-Jareno from CEH-Wallingford is specially thanked for her help in producing the metadata and acquiring a doi and Oliver Robertson from CEH-Wallingford is gratefully acknowledged for setting up the licensing terms.

References

Bell, V. A., Davies, H. N., Kay, A. L., Marsh, T. J., and Brookshaw, A.: Developing a large-scale water-balance approach to seasonal forecasting: application to the 2012 drought in Britain, Hydrol. Process., 27, 3003–3012, 2013.

British Standards Institute: BS7843-1:2012, Guide for the Design, Development and Review of a Raingauge Network, BSI, London, 2011a.

British Standards Institute: BS7843-4:2012, Guide for the Estimation of Areal Rainfall, BSI, London, 2011b.

Burt, S.: British Rainfall 1860–1993, Weather, 65, 121–128, 2010.

Cole, S. J. and Moore, R. J.: Hydrological modelling using raingauge- and radar-based estimators of areal rainfall, J. Hydrol., 358, 159–181, 2008.

Cranston, M., Maxey, R., Tavendale, A., Buchanan, P., Motion, A., Cole, S., Robson, A., Moore, R. J., and Minett, A.: Countrywide flood forecasting in Scotland: challenges for hydrometeorological model uncertainty and prediction, in: Weather Radar and Hydrology, Proc. Exeter Symp., April 2011, edited by: Moore, R. J., Cole, S. J., and Illingworth, A. J., IAHS Publ., 351, International Association of Hydrological Sciences, 2012.

Delhomme, J. P.: Kriging in the hydrosciences, Adv. Water Resour., 1, 251–266, 1978.

Dore, A. J., Kryza, M., Hall, J. R., Hallsworth, S., Keller, V. J. D., Vieno, M., and Sutton, M. A.: The influence of model grid resolution on estimation of national scale nitrogen deposition and exceedance of critical loads, Biogeosciences, 9, 1597–1609, doi:10.5194/bg-9-1597-2012, 2012.

Eden, P.: Traditional weather observing in the UK: an historical overview, Weather, 64, 239–245, 2009.

Eisenberg, M. C., Kujbida, G., Tuite, A. R., Fisman, D. N., and Tien, J. H.: Examining rainfall and cholera dynamics in Haiti using statistical and dynamic modeling approaches, Epidemics, 5, 197–207, 2013.

Gold, C. M.: Surface interpolation, spatial adjacency and GIS, in: Three Dimensional Applications in Geographical Information Systems, edited by: Raper, J., Taylor and Francis, London, 21–35, 1989.

Howard, P. J., Cole, S. J., Robson, A. J., and Moore, R. J.: Raingauge quality-control algorithms and the potential benefits for radar-based hydrological modelling, in: Weather Radar and Hydrology, Proc. Exeter Symp., April 2011, edited by: Moore, R. J., Cole, S. J., and Illingworth, A. J., IAHS Publ., 351, International Association of Hydrological Sciences, 2012.

Jones, S. B.: The Estimation of Catchment Average Point Rainfall Profiles, Institute of Hydrology, UK, Report No. 87., 1983.

Kitchen, M. and Illingworth, A.: From observations to forecasts – Part 13: The UK weather radar network – past, present and future, Weather, 66, 291–297, 2011.

Ledoux, H. and Gold, C.: An efficient natural neighbour interpolation algorithm for geoscientific modelling, in: Developments in Spatial Data Handling, Springer, Berlin, Heidelberg, 97–108, 2005.

Met Office: Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853–current). NCAS British Atmospheric Data Centre, available at: http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0 (last access: 23 January 2015), 2012a.

Met Office: Met Office Surface Data Users Guide, Section 5.6 Precipitation, available at: http://badc.nerc.ac.uk/data/ukmo-midas/ukmo_guide.html#5.6 (last access: 23 January 2015), 2012b.

Li, J. and Heap, A. D.: A Review of Spatial Interpolation Methods for Environmental Scientists – Geoscience Australia Record 2008/23, Geoscience Australia, Canberra, Australia, 2008.

Li, J. and Heap, A. D.: A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors, Ecol. Inform., 6, 228–241, 2011.

Perry, M. and Hollis, D.: The generation of monthly gridded datasets for a range of climatic variables over the UK, Int. J. Climatol., 25, 1041–1054, 2005.
Perry, M., Hollis, D., and Elms, M.: The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK, Climate Memorandum No. 24, National Climate Information Centre, Met Office, Exeter, UK, 2009.
Prosdocimi, I., Kjeldsen, T. R., and Svensson, C.: Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK, Nat. Hazards Earth Syst. Sci., 14, 1125–1144, doi:10.5194/nhess-14-1125-2014, 2014.
Rodda, J. C. and Dixon, H.: Rainfall measurement revisited, Weather, 67, 131–136, 2012.
Rodda, J. C. and Smith, S. W.: The significance of the systematic error in rainfall measurement for assessing wet deposition, Atmos. Environ., 20, 1059–1064, 1986.
Sibson, R.: A brief description of natural neighbour interpolation, in: Interpreting Multivariate data, edited by: Barnett, V., John Wiley and Sons, Chichester, 21–36, 1981.
Sorensen, J. P. R., Finch, J. W., Ireson, A. M., and Jackson, C. R.: Comparison of varied complexity models simulating recharge at the field scale, Hydrol. Process., 28, 2091–2102, 2014.
Spackman, E.: Calculation and Mapping of Rainfall Averages for 1961–90, University of Salford, Manchester, 1993.
Stewart, E. J., Jones, D. A., Svensson, C., Morris, D. G., Dempsey, P., Dent, J. E., Collier, C. G., and Anderson, C. A.: Reservoir Safety – Long Return Period Rainfall, Draft final report, Wallingford, 2010.
Svensson, C., Folwell, S., Dempsey, P., Dent, J., and Fung, C. F.: Reservoir safety – long return period rainfall, Report on data consolidation – revised version, Project Number WS194/2/39, Milestone 2 Defra, Unpublished, 77 pp., 2009.
Thiessen, A. H.: Precipitation averages for large areas, Mon. Weather Rev., 39, 1082–1084, 1911.
Walsh, S.: A Summary of Climate Averages 1981–2010 for Ireland, Climatological Note, 14, Dublin, 2012a.
Walsh, S.: New Long-Term Rainfall Averages for Ireland, Tullamore, Ireland, 3–12, 2012b.
Webster, R. and Oliver, M.: Geostatistics for Environmental Scientists, John Wiley & Sons, Ltd, Chichester, 2001.
Young, A. R., Grew, R., and Holmes, M. G. R.: Low flows 2000: a national water resources assessment and decision support tool, Water Sci. Technol., 48, 119–126, 2003.

Young, A. R., Keller, V., and Griffiths, J.: Predicting low flows in ungauged basins: a hydrological response unit approach to continuous simulation, in: Climate Variability and Change – Hydrological Impacts, IAHS Publication, 134, 2006.

Table 1. Repartition (%) of the absolute errors (ϵ (mm), Eq. 4) across different ranges of observed rainfall (rc_o) events for the observed data of the Scottish validation gauges.

Range of rc_o	Number rc_o	$\epsilon = 0$	$0 < \epsilon \leq 0.5$	$0.5 < \epsilon \leq 2$	$2 < \epsilon \leq 5$	$5 < \epsilon \leq 10$	$\epsilon > 10$
$rc_o \geq 0$	75 796	25.4	31.5	23.8	12.6	4.6	2.0
$rc_o > 0$	47 944	4.1	35.3	33.9	18.2	6.2	2.3
$0 \leq rc_o < 2$	48 593	38.7	39.0	15.5	4.0	1.6	1.3
$2 \leq rc_o < 5$	10 662	2.4	26.2	48.7	19.0	2.9	0.8
$5 \leq rc_o < 10$	8367	1.8	16.1	39.7	33.6	8.0	0.8
$10 \leq rc_o < 20$	6019	1.1	11.4	28.2	36.2	18.4	4.6
$rc_o \geq 20$	2155	0.5	5.1	14.1	28.4	29.0	22.8

Table 2. Repartition (%) of the absolute relative errors (δ, Eq. 5) across different ranges of observed rainfall (rc_o) events for the observed data of the Scottish validation gauges.

Range of rc_o	Number rc_o	$\delta = 0$	$0 < \delta \leq 0.3$	$0.3 < \delta \leq 0.6$	$0.6 < \delta \leq 1$	$1 < \delta \leq 2$	$2 < \delta$
$rc_o > 0$	47 944	4.1	38.4	25.3	19.5	6.1	6.6
$0 < rc_o < 2$	20 741	7.2	16.5	23.6	29.0	9.8	13.9
$2 \leq rc_o < 5$	10 662	2.4	43.8	27.9	17.3	6.2	2.3
$5 \leq rc_o < 10$	8367	1.8	57.1	27.8	11.1	2.0	0.3
$10 \leq rc_o < 20$	6019	1.1	67.1	23.8	7.2	0.7	0.0
$rc_o \geq 20$	2155	0.5	69.8	24.4	5.2	0.1	0.0

Table 3. Repartition (%) of the difference in absolute relative errors (Eq. 7) between the monthly corrected estimates and the standard estimates across different range of observed rainfall (rc_o) events for the observed data of the Scottish validation data.

Range of rc_o	Number of events	$\varphi = 0$	$0 < \varphi \leq 0.2$	$0.2 < \varphi \leq 0.5$	$0.5 < \varphi \leq 1$	$1 < \varphi \leq 5$	$\varphi > 5$
$rc_o \geq 0$	75 796	56.6	29.4	6.5	3.9	3.4	0.2
$rc_o > 0$	47 944	43.0	37.0	9.4	5.6	4.7	0.3
$0 \leq rc_o < 2$	48 593	69.8	25.8	2.4	1.0	0.8	0.1
$2 \leq rc_o < 5$	10 662	40.1	41.9	11.2	4.7	2.0	0.1
$5 \leq rc_o < 10$	8367	31.5	36.5	15.7	9.9	6.3	0.1
$10 \leq rc_o < 20$	6019	26.9	29.1	16.6	13.0	13.8	0.6
$rc_o \geq 20$	2155	20.8	21.5	12.4	14.6	26.7	4.0

Title Page
Abstract
Instruments
Data Provenance & Structure
Tables
Figures
◀I
▶I
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

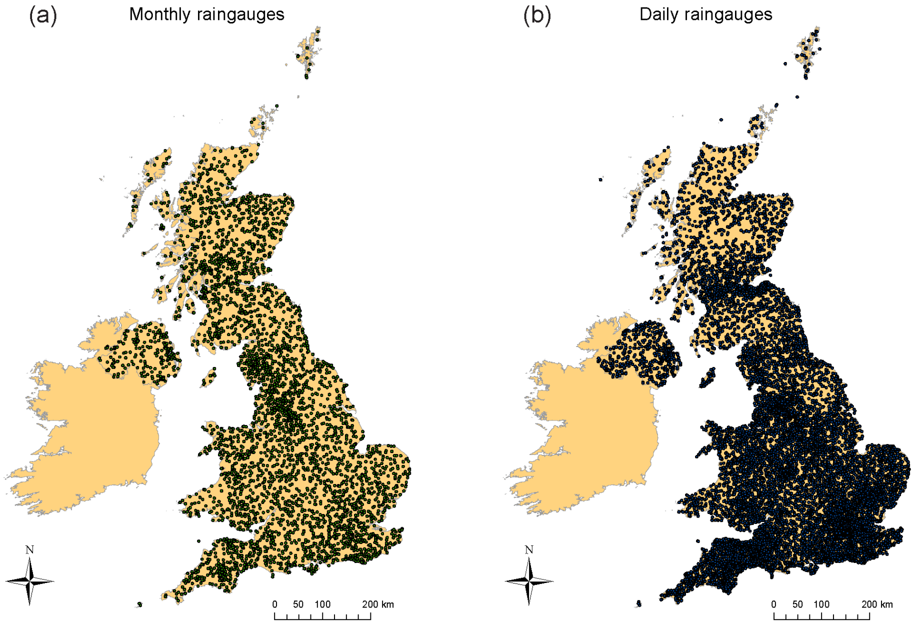

Figure 1. Maps of all raingauges used to derive the CEH-GEAR dataset: **(a)** monthly raingauges and **(b)** daily raingauges.

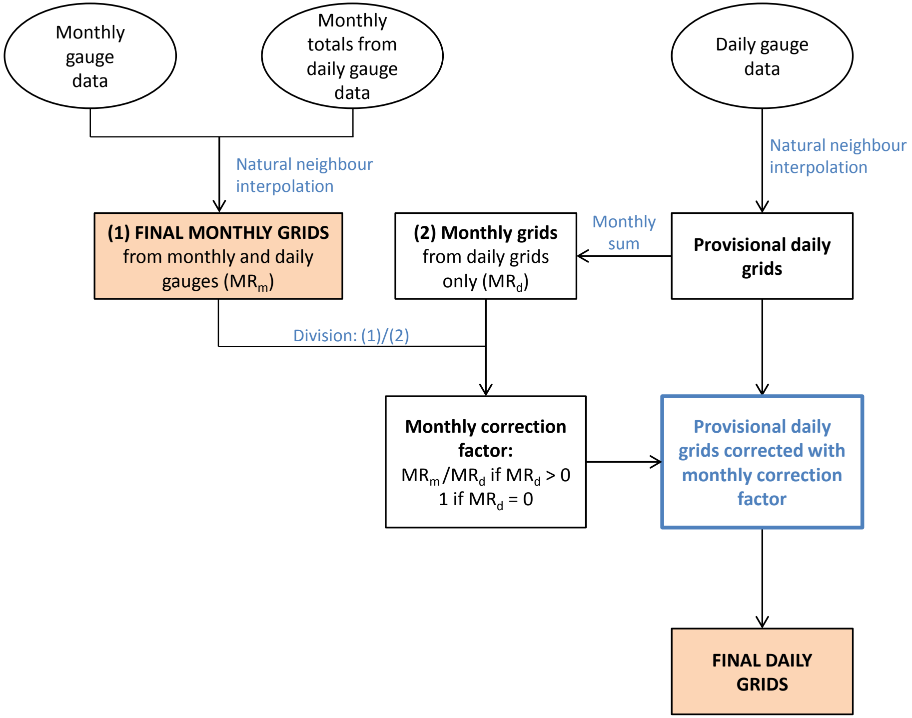

Figure 2. Schematic overview of the generation of daily and monthly 1 km gridded rainfall estimates for the UK.

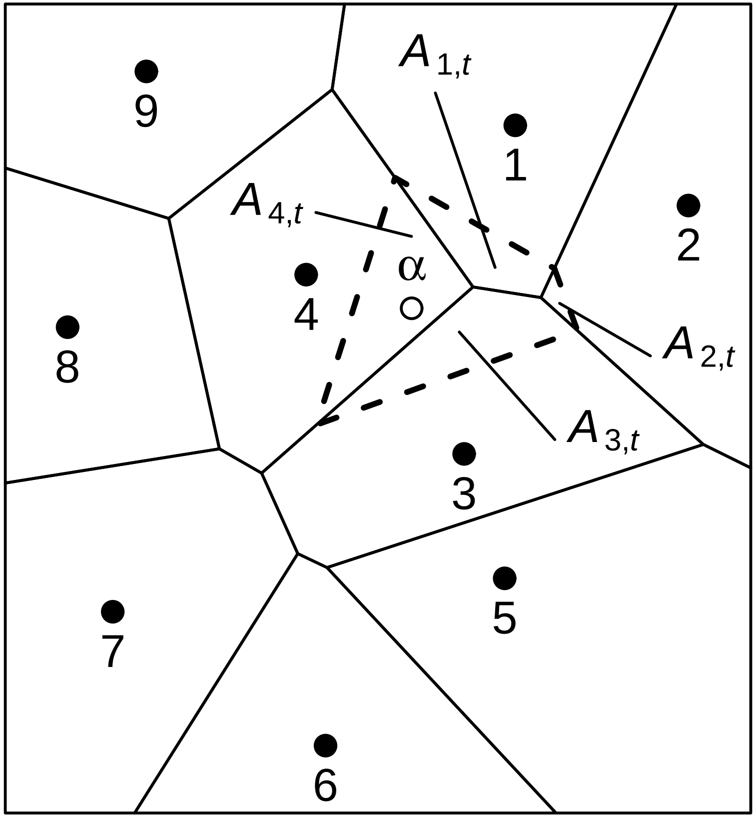

Figure 3. Illustration of the natural neighbour method for time step t. The solid lines represent the original Thiessen polygons $T_{i,t}$ for each of the i raingauges (solid circles). The dashed line represents the Thiessen polygon $\hat{T}_{\alpha,t}$ for the grid point α (open circle). The overlap between $T_{i,t}$ and $\hat{T}_{\alpha,t}$ is labelled $A_{i,t}$ ($A_{i,t} = \text{Area}(T_{i,t} \cap \hat{T}_{\alpha,t})$, Eq. 2).

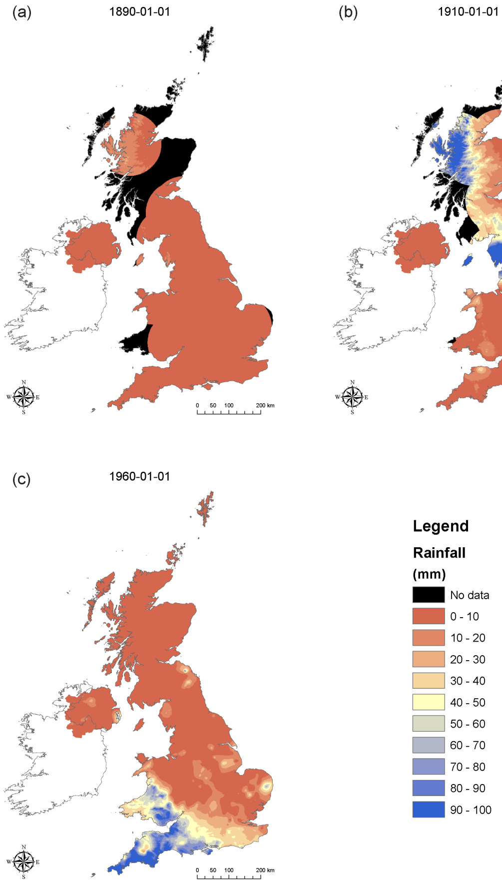

Figure 4. Rainfall map showing missing data (in black) **(a)** on the 1 January 1890 **(b)** on the 1 January 1910 and **(c)** on the 1 January 1960.

Title Page
Abstract
Instruments
Data Provenance & Structure
Tables
Figures
I◀
▶I
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

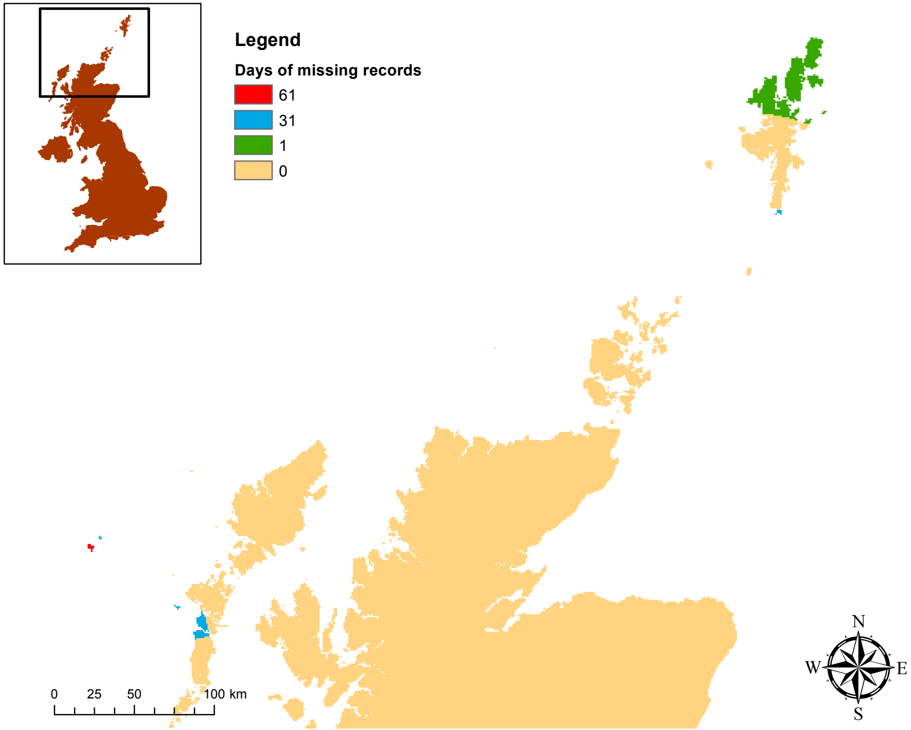

Figure 5. Map of total days of missing records (for daily grids) for the period post-1961 for northern Scotland (no missing records in rest of UK).

Title Page
Abstract
Instruments
Data Provenance & Structure
Tables
Figures
|◀
▶|
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

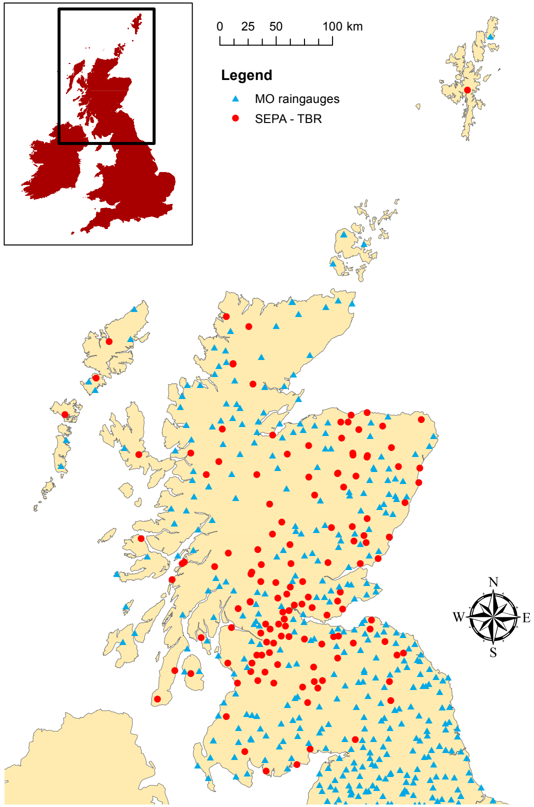

Figure 6. Map of SEPA tipping bucket raingauges (TBR) used for validation purposes (red circles) and Met Office (MO) raingauges network (blue triangles). The MO raingauges are represented on the map only if they have data available for at least 50 % of the days between 2007–2010 (validation period).

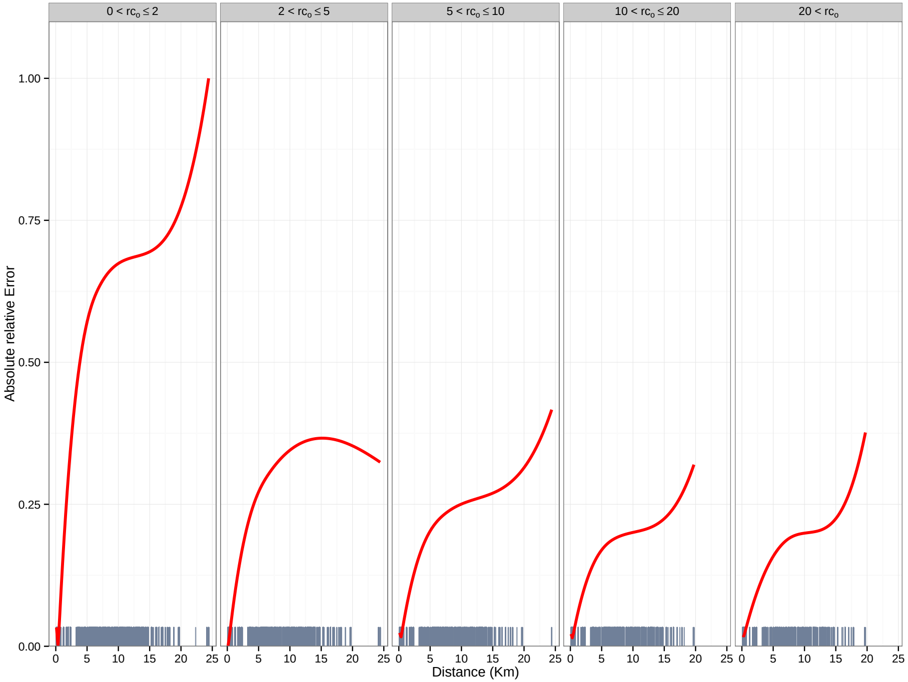

Figure 7. Median absolute relative error represented as a function of the distance to the closest raingauge for different observed rainfall event ranges. The grey lines along the *x* axis indicate the distance between the tipping bucket and the closest rain gauge used in the estimation procedure.