Exploration of Rain Gauge Quality Issues in Northern England

S Fadhel* and M S Saleh
Water Resources Engineering Department, Mustansiriyah University, Baghdad, Iraq
E-mail: alrubaiye@uomustansiriyah.edu.iq

Abstract. The rain gauge, which is the preferred device for rain rate quantification, is susceptible to various types of errors. The problems of inconsistency and incompleteness of rainfall time-series, in addition to data quality issues obstruct hydrometeorological analysis. Daily rainfall data from 30 gauges in Northern England were quality controlled by a double quality check procedure for the period 2006–2014 to solve any specific data quality issues. It emerged that fault readings happened only over a few days compared with the high percentage of missing data which could reach years. Thus, factors that impact gauges’ missing data were inspected. It was found that snowfalls, land cover-land use and bird distribution are the dominant factors that malfunctioning the rain gauge for long time periods, often producing successive missing days that could be extended months in total. However, when the missing data happened on separate days, then it seemed that high winds related with light rainfall resulted in drifting away the rainfall from the rain gauge and thus caused missing data. Moreover, the rainfall itself was not one of the major factors to malfunction the gauge, because the values of missed rainfall were not quite extreme to do this.

1. Introduction

Hydrological and environmental models as well as water-quality models are reliant on accurate rainfall information. Rain gauge is one of the traditional devices that are used to measure rainfall intensity at a point. However, gauges are prone to different types of errors and require regular maintenance and calibration to provide accurate rainfall measurements [1]. Point rainfall measurements are affected by both random and systematic errors [2, 3]. While random errors can be put down to micro-climatic alterations in the vicinity of the gauge, systematic errors can be traced to factors such as blockages, extreme rainfall, wind, wetting, and losses due to evaporation [2-4]. The above mentioned factors could also malfunction the gauges producing missing data. However, in case the gauges are located in inaccessible areas, then the malfunction may last for a long period (e.g. month) before the problem is solved [2, 4].

Several instruments including weather radars and computer models occasionally rely on data from rain gauges for their calibration [5, 6]. As such, it is essential that such data should be reliable, complete, and free of considerable biases. Before the observations derived from rain gauges and other sensors are integrated, it is crucial that issues related to the detected biases and incomplete data are resolved.

The removal of these types of errors that impede the performance of rain gauges can be realized through the design and management of an effective automated monitoring network [6-8]. However, outliers within gauges can be detected by quality control procedure either by comparison with historical observations or with neighbouring gauges [4, 9]. While, many authors adopt a sequence type
of quality control procedures where the data are checked in each step and outliers detected and removed. A threshold is set, and outliers exceeding the specified threshold are distinguished and discarded. Subsequently, data with other forms of homogeneity flaws in the dataset are identified. These include (a) data occur too frequently; (b) a series of too low or zero rainfall data; and (c) the impractical occurrence of dry months during a rainy season [8, 10, 11]. Thus, different proposed procedures to detect outliers vary depending on the adopted methodology for such detection.

Most of the previous studies addressed errors within gauges and propose solutions for such problems separately (e.g. studies addressed either factors affect gauge systematic error; gauge’s quality control; or infill gauge’s missing data). However, only few studies that give a thorough analysis of all these problems; address each problem in detail and solve it in one study. In addition, regarding gauge’s quality control problem, almost all but few literatures adopt neighbouring gauges for the quality control of single gauge. Radar data can play an important role to check the quality of gauges especially during convective storms where a gauge that receives rainfall does not need to be spatially consistent with its neighbours. While, only few studies used radar data for such check [3, 12]. This is the case especially for networks with sparse gauges, where neighbouring gauges located far away from the gauge under test. For such cases neighbouring gauges may not be useful for single gauge quality check.

Thus, the aims of this study are (1) to produce a quality controlled precipitation database via a two-step quality control of daily rain gauge data using the nearest gauges and radar data; (2) to address the factors that cause systematic errors i.e. faults and absences in the gauge measurements. They’re supposed to be a third objective in this study which is to fill in the identified missing days in the incomplete 9-years’ time series and correct the outliers resulting from this quality check for the gauges which have these problems. However, due to space constraints we will perform the third objective in a separate study. In this way a corrected and complete rainfall time series should be achieved for hydrologic modelling and water-resource studies.

2. The study area and meteorological data
The study area is located in West Yorkshire, Northern England and comprises an area of approximately $60 \times 70$ km (Figure 1). The catchment has medium-size topographic gradients and the altitude ranges from 30 to 350m above sea level with a mean elevation of 200m. The mean variance in elevation between the stations is 180m and the greatest elevation variation between the highest and the lowest stations is 340m. Variances in altitudes are much less relevant across shorter station distribution, as the greatest difference in elevation between nearby stations in the catchment is 80m. This enables greater precipitation gradients locally associated with variations in altitude to be looked on as insignificant. Since the distances between the nearby gauges which are used to assess the gauges in the test are short and the elevation variance of those nearby gauges is small.

The daily rain gauge data are provided by the British Atmospheric Data Centre (BADC) for 30 gauges located in the study area for the period (1st of January 2006 – 31st of December 2014); these daily gauges run by the Met Office are either manual or automatic gauges. For our case study the manual gauges are the standard Met Office MK2 splay-based -type gauges of a 127mm diameter, while the automatic gauges are tipping bucket gauges that have a bucket capacity of 0.2 mm of rain. Although the BADC dataset has been subjected to a quality check by the Met Office [13], such data still have some sources of error and need further quality check (as can be seen in the next section). Figure 1 shows the location and the type of gauges for our study area and it is clear that the tipping bucket gauges comprise half of the gauges; the other half are the manual MK2 splay-based gauges.

To address the systematic error that can cause faults or absences in the gauge readings, we need another complete rainfall dataset to have a clear idea about the annual and monthly rainfall rate for the study area. Thus, a 1 km gridded daily rainfall dataset is adopted in this study, provided by CEH – Gridded Estimates of Areal Rainfall (CEH – GEAR) for the period 2006-2014. The CEH data are usually used to study the systematic error effect on the missing data and are not used for the quality
check since this has already been created by interpolating gauges with different densities for different years [14].

Figure 1. The study area and rain gauge locations.

The composite radar data covering the study area are provided by the UK Met Office radar network through the BADC with spatial and temporal resolutions of 1 km and 5 minutes respectively. The catchment area is within the coverage of three single-polarisation C-band weather radars (Hameldon Hill, High Moorsley and Ingham) which are located 30 km, 95km and 90km away from the study area respectively [15]. Quality control and corrections of some sources of errors related to the radar rainfall data were implemented by the UK Met Office Nimrod System [16]. Usually the radar data are checked and corrected using the gauge data. However, in this study, the radar data was adopted purely as an assistance tool for further quality checking of the gauge data. This means it is used as a qualitative check instead of a quantitative check. However, since the radar has missing data, it was not useful to adopt it for the gauge’s systematic error study since we need a monthly and annual rainfall values for such study. Thus, the radar data have been aggregated to a daily time scale before its use in the quality check.

A 1 km gridded daily temperature and wind speed for the study area for the period (2006-2014) was provided by the Climate Hydrology and Ecology research Support System meteorology dataset (CHESS-met). These data were used to address their effects on the gauge’s systematic error.

Two snow gauges provided by the Met Office for the period 2006-2014 were used to check the snowfalls events within the study area and to address the impact of snow on gauge’s systematic error (Figure 1).

3. Methodology

3.1. Data quality control
A two steps quality check procedure which derived by Fadhel et al. [12] (for the same study area) was adopted in this study. The procedure includes: (i) the spatial consistency check, (ii) zero values check. For both quality checks, each gauge is compared with the nearest gauges. The gauges that are flagged up as a result of any of the two checks are subjected to another test using radar data (i.e. radar pixels where the gauge under test and neighbouring gauges are located). This is because rainfall is highly variable in space and time, so the performance of a gauge during a convective storm is not necessarily consistent spatially with that of neighbouring gauges. The neighbouring stations were selected by: the closest distance to the gauge under the test with 15 km as a maximum distance; the higher the Pearson coefficient; and with average elevation differences of 200m (see [12] for the details of the quality check procedure). Figure 2 shows the number of flagged days from the above double quality check procedure for gauges which report a quality problem during the whole time series. It is worth mentioning that gauges without a bar plot in Figure 2 are free from quality issues.

Figure 2. Outliers resulted from the double quality check procedure. The gauges ID on X axis are the same in Figure 1 with sequence 1-30. Numbers placed at the top of the bar represent the total percentage of missing data for each gauge.

3.2. Systematic Errors in precipitation measurement

Usually errors within gauges are divided in two types: random and systematic errors. In this study we addressed four factors that could cause systematic errors (i.e. faults and missing data) within the gauge, those factors are: a) land cover and land use; b) rainfall and wind (separately and interactively); c) temperature and snowfall (conjointly); and d) bird distribution (blockage due to bird droppings).

The temporal availability of daily rainfall time series for each of the 30 gauges are plotted as a chronogram and presented in Figure 3. It is clear from Figure 3 that almost all the gauges except two have missing data. The chronogram plot also shows that the period length of missing data are either long with successive days of missing data, or short periods with only one or two missing days. Long periods of successive missing data refer to gauge malfunctioning. Factors affecting missing data for both short and long periods are investigated to see what factors cause gauge malfunctioning and whether it is different from those causing lost of data for short periods.

Since 93% of the gauges in this study have missing data, the percentages of missing data (Pmd hereafter) for each gauge for the whole period (2006-2014) were researched and gauges with high, medium and low percentages of missing data were identified. The first group of gauges having the highest Pmd where Pmd ranging between 10-35%. The second group of gauges with medium Pmd where Pmd is in the range 0.5-10%. Gauges with Pmd less than 0.5% compose the third group with
lowest Pmd. It is clear from Figure 2 that the gauges in the case study are divided equally between the three groups of missing data with one third of total gauges for each group. By looking at the gauge type and period length of missing data for each group it is clear that all the gauges with the highest Pmd are manual gauges, and the period length of missing data for those gauges between 1.1-3 years. While the period length of those gauges with a medium Pmd is between 22 days and 6 months; all but three of these gauges are the tipping bucket type. Finally, the last group of the lowest Pmd comprises 8 tipping bucket gauges and only two manual gauges with the length of missing data between one and two days.

It is clear from the chronogram plot (Figure 3) that some gauges share the same period of missing data. By examining the gauge type, and the group that they belong to, it was discovered that six manual gauges out of ten from group 1 share thirteen periods of missing data (Table 1), and the length of each period varies from 1-3 months. Those six gauges (1, 2, 4, 7, 10, 15 in Figure 1) are mainly located in one area that has similar topographical conditions with a maximum distance of 18 km from the centre of the study area. However, gauge 56024 (i.e., 28 in Figure 1) which has 31.86% missing data is located to the south of the six gauges at 31.85 km from the centre of the study area. That may explain why this gauge does not have similar periods of missing data as the other six gauges even though it has similar Pmd. Thus, it is clear that if manual gauges are located in one area, close to each other, and at a comparable terrain level, then there is a high possibility that such gauges will share the same period of missing data due to the same effect of the factors that cause gauge systematic error or gauges malfunctioning.

It should also be noted that sometimes the gauge’s rainfall values are already available in the metadata; however such gauges data did not undergo a quality check by the Met Office. In addition, it was noticed that the rainfall value for these dates is the sum for few days and not only for one day. Thus, those days are considered as missing data for the purposes of this study.

Although the reasons for the gauge’s systematic errors are known to hydrologists, this study tried to answer these questions: which one of the above mentioned factors has the major effect on gauge’s systematic error? Do these factors have the same effect for different gauge types? The next section will answer these questions by addressing the effect of four factors on gauge fault and absence measurements.

To study the effect of rainfall rate and temperature on systematic error within gauges, we analysed the rain gauge’s missing data in three ways. Firstly, by calculating the yearly Pmd for each gauge and the total number of gauges (NG hereafter) that report missing data, because the number of gauges which report missing data varies during the period of the study. Secondly, by finding the monthly Pmd and NG over the whole period (2006-2014) it should be possible to discover which months or seasons have the highest readings. The final analysis is performed by calculating the monthly Pmd and NG for each year to see whether the missing data are consistent with the monthly change in rainfall and temperature for each year. For each of the three cases of analysis the corresponding effects of rainfall and temperature will be examined to see how these factors affect the gauge systematic error.
Table 1. Long period of successive missing data starting from these dates for the six manual gauges with the highest Pmd. The gauges are (1, 2, 4, 7, 10, 15) in Figure 1.

| Dates        | No. of successive missing dates |
|--------------|---------------------------------|
| 02/09/2006   | 30                              |
| 03/10/2006   | 91                              |
| 01/03/2007   | 32                              |
| 01/11/2007   | 62                              |
| 02/05/2008   | 31                              |
| 02/10/2008   | 31                              |
| 03/02/2009   | 27                              |
| 02/10/2009   | 31                              |
| 02/12/2009   | 31                              |
| 01/02/2010   | 29                              |
| 01/04/2010   | 31                              |
| 01/07/2010   | 33                              |
| 01/12/2011   | 32                              |

Figure 3. Chronogram showing the temporal availability of daily rainfall data before quality control.

4. Results and Discussion

4.1. Land cover-land use

The Land Cover Map 2007 from the Centre of Ecology and Hydrology [17] has been adopted in this study to address how the land cover type could affect the systematic error within the gauges and produce faulty and missing data. The dominant types of land cover for the study area are a mix of
urban, improved grassland and heather (Figure 4). Also there are small different agricultural areas. The colour key of Figure 4 can be read from [17]; Page 14.

![Land Cover Map 2007 from the Centre of Ecology and Hydrology (CEH) for the study area.](image)

By looking at the gauge locations it could be seen that almost all of the gauges in group 1 and most of the first half gauges from group 2 are located either in improved grassland or in heather grassland, and both areas are effective places for herding sheep. Thus, it can be speculated that those gauges could either be damaged by animals or blocked by grass. In addition, in the case of the manual gauges, such gauges should be checked regularly by volunteers who record the daily gauge reading and help to report any problems observed within the gauges. However, gauges installed in such areas may be far from the volunteer’s place, thus data could be lost because the volunteer may have other duties and so not record the gauge reading accurately or at all.

In contrast, most of the gauges from group 3 (with the lowest Pmd) and the second half gauges from group 2 are located either in or close to urban areas, Figure 4. Thus, here the risk that the gauges will be damaged by animals or blocked by grass should be low. In addition, those manual gauges located in such areas are checked regularly by volunteers since the gauge locations are close from the volunteer’s place. Gauges installed in or close to urban areas become easily accessible to fix anything wrong noticed within the gauge report in the case of the tipping bucket type; or anomalies are reported by the volunteer in the case of the manual gauges.

But it is also evident from Figure 4 that other factors can also affect gauge recording negatively more than any damage caused by the land cover-land use factor. Since there are a few manual gauges within or close to urban areas that still experience the highest percentage of missing data (e.g. gauges 16597, 2526 and 2545 (i.e. 2, 3, 6 in Figure 1)).
4.2. Rainfall

Figure 5 shows the annual Pmd, NG, and the annual rainfall. It is clear from Figure 5b that most of the gauges in group 1 are located in an area which experiences high annual rainfall, since they are located at a high terrain level. However, the gauges from groups 2 and 3 are located in both areas with both low and high terrain levels in the study area; consequently they receive a varied amount of rainfall depending on the location. Thus, gauges could malfunctioning in areas of high rainfall intensity especially in the case of an extreme rainfall. In tipping bucket gauges, the inflow of water from the rain gauge funnel can interrupt, or even terminate the functioning of the tipping mechanism in case of heavy rainfall. In such a situation, the bucket tip rate is either slows down, or completely stops [18].

From the initial analysis, it is clear from Figure 5a that both the Pmd and the NG have decreased gradually from 2006 to 2014 with an exception for the year 2010 which shows high levels for both readings. The years 2006 and 2011 recorded the highest and lowest values of Pmd respectively. However, the annual rainfall analysis shows surprising results. For example, 2012 shows the highest annual rainfall but this year shows a low Pmd value (Figure 5b and 5a respectively). Also, 2010 shows the lowest annual rainfall but high values in both the Pmd and the NG. The annual rainfall for 2006 which shows the highest NG and Pmd is between the maximum and minimum annual rainfall for the whole period.

Since the annual rainfall could not be established as a reason for the prevalence of missing data, the next step of the analyses was undertaken to address the monthly rainfall and the corresponding monthly gauge’s Pmd and NG over the whole period. Figure 6a shows the monthly Pmd and NG over the whole period (2006-2014) and it shows that the highest Pmd and NG are for the months January, February, and October-December. It is clear that the highest Pmd and NG are concentrated within the autumn and winter seasons, since the monthly rainfall values for these seasons is mainly higher than the corresponding values for the other months, as shown clearly in Figure 6b. Thus, it can be seen that the total monthly Pmd and NG are consistent with corresponding total monthly rainfall.

Finally the monthly rainfall was analysed in relation to the corresponding monthly gauge’s Pmd and NG for each year of the period (2006-2014), and here the result for the year 2007 is given as an example. It was noticed that in most cases - but not all - the monthly Pmd and NG are consistent with increasing rainfall values for that month. For example, it is clear from Figure 7a that the Pmd increased gradually from October until its peak in December; such an increase is consistent with the corresponding monthly rainfall increase (Figure 7b). Interestingly, in other cases we noticed that both records increased in some months while the corresponding monthly rainfall was low, e.g. March and May. In addition, we also noticed a case where a peak monthly rainfall, January (Figure 7b) shows the lowest corresponding monthly Pmd (Figure 7a). From this, it can be suggested that it is not always the rainfall that causes the gauge data errors, since there are several months with high rainfall amounts but low Pmd.
Figure 5. The annual Pmd and rainfall over the whole period. Black stars are gauges from group1, blue stars are gauges from group2, and cyan stars are gauges from group3.
So what rainfall values cause gauge systematic error and gauge malfunctioning and loss of data for long periods of successive days? For this analysis five manual and four tipping bucket gauges were selected with the highest Pmd for both types of gauges. As mentioned earlier, most of the gauges with the highest Pmd have experienced missing data over successive days, often reaching months. For such cases the rainfall values for only the first day were extracted from the successive days because it is assumed that the gauge showed errors on that day. Thus making it unreasonable to analyse the rest of the days since the gauge had already stopped working properly. To find the rainfall value for the missing dates, the corresponding dates from the CEH dataset for the pixels where those gauges are located were extracted. It was found that all manual gauges except one show missing rainfall values between low and medium rainfall but the gauge could not have been malfunctioning by such rainfall values. It therefore seems reasonable to conclude that rainfall itself is not the main cause for the gauge’s systematic error, even though the monthly rainfall in most cases was consistent with the corresponding monthly Pmd. Thus, there has to be another factor associated with rainfall that produces gauge absence readings.

So, wind data was analysed as well, and the wind speed was extracted for these dates for the pixels where those gauges are located. Wind maps were also plotted for the study area in order to confirm that those dates which reported missing data were windy or not (results are not shown).

Since the rainfall catch of the gauge can be affected by high winds, and this is greatest for light rain [19], missing light rainfall data can be explained as being due to a corresponding high wind speed (more than 10 knots) on these dates. However, the most extreme rainfall value that appears in the four manual gauges is 36 mm; the wind speed on this date is not high enough to have caused data to be lost. This rainfall value is additionally not large enough to malfunction the manual gauge since the capacity of this type of gauge is 75 mm for the inner collector and 140 mm for an extra overflow storage [20]. Hence, it seems there is another factor that causes gauge malfunctioning in such cases. On the contrary, the missed rainfall values for the last manual gauge show both the most extreme rainfall values between those gauges, and the windiest values that have most certainly caused the rainfall to not be caught by the gauge.

The same reason for the missing of light rainfall values due to wind speed was found in the case of tipping bucket gauges. However the maximum rainfall value that was lost from the tipping bucket gauge was 22.5mm, which is really not enough to malfunction the tipping bucket, since this type of gauge performs well within a rainfall rate of 0.6-30mm [18].

It is worth mentioning that the study area showed extreme rainfall events, which can cause flooding to the area, such as that experienced on 24-25 Jun 2007. However, apparently the gauges were already damaged before these events, and that is why it can be concluded that the rainfall was not the main reason for data errors in the case study.
Figure 6. The monthly Pmd and rainfall over the whole period. Black stars are gauges from group1, blue stars are gauges from group2, and cyan stars are gauges from group3.
Figure 7. The monthly Pmd and rainfall for the year 2007. Black stars are gauges from group1, blue stars are gauges from group2, and cyan stars are gauges from group3.
4.3. Temperature
The previous rainfall analysis is repeated here for temperature, and we found that most of the gauges with a high \( P_{\text{md}} \) from group 1 are located within areas of cold temperature since those gauges are located at a higher terrain level. However, those gauges with medium and low \( P_{\text{md}} \)s from group 2 and 3 are distributed equally in areas with low and high temperature due to the corresponding high and low terrain level in these areas.

Likewise with the rainfall’s effects on gauge malfunctioning and by logical thinking, the temperature itself cannot deflect the gauge directly but the consequences from having a low and especially a freezing temperature can affect the gauge. Since in cold weather there is a high possibility of experiencing snowfall within the UK, snow can block the gauge’s funnel causing damage to gauges and also take a longer time to melt \([18]\). Accordingly, snowfall events were checked using two gauges provided by the Met Office for the period 2006-2014. Most of snowfall events were consistent with the corresponding missing data from the gauges for these dates, and especially for the dates which showed a long period of successive missing data. Table 2 shows an example of the long period of successive missing data for gauge 2513 (i.e. 15 in Figure 1) which is consistent with snowfall events for these years. It is worth mentioning that in some cases there were snowfalls events for two or three successive months (case 2, 3 and 8 in Table 2). That explains the long period of successive missing days since snow took long time to melt. Moreover, in some cases the gauge already malfunctioning before snowfalls events by few days and because of snowfalls it last damage for longer period (case 1, 5, 8 in Table 2). In addition, the year 2010 experienced the highest snowfalls events, that explain the high annual readings for both \( P_{\text{md}} \) and NG for that year (Figure 5a), although the annual rainfall was the lowest for such year (Figure 5b). Thus, snowfall associated with freezing temperature is an important factor that cause gauge systematic errors and gauge malfunctioning.

4.4. Bird distribution
As is clear from the above analysis, it is not always the land cover-land use, rainfall, wind or the snowfall that affects the gauges and causes data to be lost. However, birds could also be considered as a major contributor to gauge malfunctioning through their droppings. Indeed “small birds seem to find a rain gauge rim an excellent toilet perch” \([18]\). The Bird Atlas Mapstore was used to check the bird distribution for our study are. In addition, the Royal Society for the Protection of Birds (RSPB) was also referred to, both check the information and the availability of birds. The bird distribution in our study area showed increasing percentages for some kinds of birds for the period 2007-2011. Some of those birds are dominant year round, while other kinds of birds appear only during the summer season.

The bird distribution was found to be dependent on the location, i.e. areas planted with heather grassland are preferable for some kinds of birds. While for improved grassland areas other kinds of birds available there, but they cannot be available in heather grassland. Thus, gauges with missing data during light or zero rainfall (cross-referenced by checking the CEH data) associated with light wind speed and when no snowfall records shown on these dates then bird droppings may have been the cause of the damage for those gauges. Table 3 shows an example of gauge malfunctioning for long periods due to either bird droppings or the land cover-land use factor. It is clear from Table 3 that most of these missing dates were either in summer or spring. Thus, in the case of a long period of missing data in these seasons, then there is a high probability that the gauge was malfunctioning due to either bird droppings or the land cover - land used factor. Also, and as mentioned earlier, in cases where the gauges were located in not an easily accessible areas then the malfunction lasted for a long period before it repair.
| Dates         | No. of successive missing dates | Dates         | Rainfall (mm) | Wind speed (knot) |
|--------------|-------------------------------|---------------|---------------|------------------|
| *1/3/2007    | 32                            | 4/9/2006      | 27            | 4.508            |
| 3/1/2008     | 90                            | 3/10/2006     | 90            | 0.948            |
| 2/12/2008    | 61                            | 2/5/2007      | 60            | 0.000            |
| 3/2/2009     | 27                            | 3/7/2007      | 11            | 9.880            |
| *2/12/2009   | 31                            | 15/7/2007     | 17            | 6.422            |
| 1/2/2010     | 29                            | 2/9/2007      | 29            | 4.044            |
| 1/4/2010     | 18                            | 1/11/2007     | 29            | 0.084            |
| *1/11/2010   | 93                            | 2/5/2008      | 30            | 0.000            |
| 1/12/2011    | 32                            | 2/10/2008     | 30            | 0.206            |
| 3/1/2012     | 30                            | 5/7/2009      | 10            | 3.891            |
|              |                               | 20/7/2009     | 26            | 0.200            |
|              |                               | 19/8/2009     | 13            | 0.000            |
|              |                               | 2/10/2009     | 30            | 3.681            |
|              |                               | 1/7/2010      | 32            | 6.337            |
|              |                               | 1/8/2012      | 31            | 1.100            |

*Mean gauges already damaged before snowfalls and last damaged for longer period due to snowfalls.

5. Conclusions
A two-step quality control procedure, including a spatial consistency check and a zero values check, were performed on thirty daily precipitation gauges in Northern England over the period 2006–2014. The proposed quality control of the gauge stations compares precipitation time series from individual rain gauges with neighbouring stations and radar data. The data quality check shows only few days with quality issues. However, the missing data within the gauges was a serious problem. Thus, the effect of land cover -land use, rainfall and wind, temperature and snowfall, and bird distribution on gauges’ systematic error was addressed.

For our case study it was found that the highest percentage of missing data is from 10 out of 15 manual gauges which comprise half of the gauges in our study. Most of those gauges are placed at a high terrain level where there are cold temperatures and high annual rainfall. However, tipping bucket gauges, which comprise the other half for our case study, are placed in both high and low terrain level, which means some of those gauges have a similar conditions to the above mentioned manual gauges with the highest Pmd. Thus, regarding the gauge type, manual gauges are affected more than automatic gauges by the factors that produce gauge systematic error.

When addressing the factors that cause gauge systematic error, we found that in case the period length of missing data lasts for nearly a month or more of successive days within summer or spring, then mostly land cover land use or bird distribution are the main reason for this, since most of the manual gauges with high Pmd are placed within improved grassland or heather grassland areas. Thus the gauges could be blocked by the grass; by bird droppings or damaged by animals. However, in the case of long successive days of missing data within winter or autumn, then snowfalls are the main reason for gauge malfunctioning, since snowfalls can block the gauge and take a long time to melt. In addition, it seems like the gauges are placed in not easily accessible areas, because when they became damaged it took long time until they were repaired and back in use again.
For our case study it was surprising to find that the rainfall was not a reason for the gauges malfunctioning and indeed to stop working for a few successive days. This is because the rainfall values on these days were not extreme enough to malfunction the gauges, whether those gauges are manual or automatic. Despite this, there were occurrences of extreme rainfall where the study area became flooded on these dates, but often the gauges were already damaged before these events. It is worth mentioning that the missed rainfall values were extracted from the CEH data from the pixels where those gauges are located. Moreover, in case the period length of missing data occurred on separate days, then wind speed associated with light rainfall found to be the main driver for these missing data. Thus for the case study it was found that the rainfall itself or temperature itself were not the main reason for the gauge to have missing data. However, the wind speed and snowfalls associated with the former factors respectively produced the gauge’s missing data.

It is worth mentioning that the gauge’s annual percentage of missing data was found to decrease gradually over the period 2006-2014 with the exception of 2010, because this year shows the highest snowfall events. That’s mean measurements were taken by the Met Office that run these gauges to reduce the effects of the above mentioned factors as much as possible on gauge systematic error. In addition, the highest Pmd were found to be concentrated over autumn and winter seasons since these seasons found to be windier and having lots of snowfall events than the other seasons.

Although the study is focused on the gauges in Northern England, its methodology and procedure are relevant to many other parts of the world. We hope this paper will encourage more similar studies around the world in different climate and geographic situations so that a comprehensive pattern on the influential factors (likely variable to the local conditions) to gauge quality issues could be obtained. Such studies will be very valuable not only to data users, but also to data providers (for better operations and maintenance of rain gauge networks).

Acknowledgements
The authors thank the British Atmospheric Data Centre, and the UK Met Office for providing the rain gauges and the radar dataset. We also thank the Centre of Ecology and Hydrology for providing the gridded temperature data through (CHESS-met) and the Gridded Estimates of Areal Rainfall (CEH_GEAR) dataset.

References
[1] Sevruk B 1996 Adjustment of tipping-bucket precipitation gauge measurements Atmos. Res. 42 237–46
[2] Sciuto G, Bonaccorso B, Cancelliere A and Rossi G 2009 Quality control of daily rainfall data with neural networks J. Hydrol. 364 13-22
[3] Yeung H.Y, Man C, Chan S. T and Seed A 2014 Development of an operational rainfall data quality-control scheme based on radar-raingauge co-kriging analysis Hydrol Sci J. 59 7 1293-307
[4] Upton G J G and Rahimi A R 2003 On-line detection of errors in tipping-bucket raingauges J. Hydrol. 278 197-212
[5] Pedersen L, Jensen N E and Madsen H 2010 Calibration of local area weather radar—Identifying significant factors affecting the calibration Atmos. Res. 97 129–43
[6] Liu S, Li Y, Pauwels V R N and Walker J P 2018 Impact of rain gauge quality control and interpolation on streamflow simulation: An application to the Warwick Catchment Australia Front. Earth Sci. 5 114
[7] Qi Y, Martinaitis S, Zhang J and Cocks S 2016 A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS System J. Hydrometeorol. 17 1675-91
[8] Blenkinsop S, Lewis E, Chan S C and Fowler H J 2017 An hourly precipitation dataset and climatology of extremes for the UK Int. J. Climatol. 37 722–40
[9] Delahaye F, Kirstetter P E, Dubreuil V, Machado L A T, Vila D A and Clark R 2015 A consistent gauge database for daily rainfall analysis over the Legal Brazilian Amazon J. Hydrol. 527 292-304

[10] Westerberg I, Walther A, Guerrero J. L, Coello Z, Halldin S, Xu C Y and Lundin L C 2010 Precipitation data in a mountainous catchment in Honduras: quality assessment and spatiotemporal characteristics Theor. Appl. Climatol. 101 381-96

[11] Lewis E, Quinn N, Blenkinsop S, Fowler H J, Freer J, Tanguy M, Hitt O, Coxon G, Bates P and Woods R 2018 A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr J. Hydrol. 564 930–43

[12] Fadhel S, Rico-Ramirez M A and Han D 2017 Exploration of an adaptive merging scheme for optimal J. Hydroinformatics. 192 225-237

[13] Met Office 2012 Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current) NCAS British Atmospheric Data Centre

[14] Keller V D J, Tanguy M, Prosdocimi I, Terry J A, Hitt O, Cole S J, Fry M, Morris D G and Dixon H 2015 CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications Earth Syst. Sci. Data 7 143–55

[15] UK Met Office 2009 Fact Sheet No. 15 — Weather Radar.

[16] Harrison D L, Scovell R W and Kitchen M 2009 High-resolution precipitation estimates for hydrological uses Proceedings of the Institution of Civil Engineers-Water Management 162 125-35

[17] Morton D, Rowland C, Wood C, Meek L, Marston C, Smith G, Wadsworth R and Simpson I C 2011 Final Report for LCM2007 - the new UK Land Cover Map. Countryside Survey Technical Report No. 11/07 NERC/Centre for Ecology & Hydrology 112pp. (CEH Project Number: C03259)

[18] Burt S 2012 The Weather Observer’s Handbook New York NY: Cambridge University Press

[19] Shaw E.M 1994 Hydrology in practice 3rd ed. Chapman & Hall 2-6 boundary row London

[20] Strangeways I 2010 A history of rain gauges Weather 65