Restless climate: Lessons using data skills

Royal Geographical Society with IBG

Advancing geography and geographical learning

Lesson 4: Reconstructing rainfall patterns of the distant past

Lesson objectives

- To evaluate the consistency of two very long reconstructed rainfall records for the English Lake District using the Mann Whitney *U*-test
- To appreciate the value of very long rainfall records for assessing the severity of extreme weather events – such as Storm *Desmond* – when set in a much longer context

Setting the scene

November 2015 was unusually wet and, by the end of the month, many northern and western UK river catchments were saturated. Then, on 5/6 December, Storm *Desmond* brought exceptionally heavy and persistent rainfall across the same areas. A new record was set in Honister Pass, Cumbria where 341 mm of rainfall fell in 24 hours. In Thirlmere, 405 mm fell in 48 hours. Records were also broken for peak river flows across a swathe of northern England. Storms *Eva* (23/24 December) and *Frank* (29/30 December) added to the misery of flooded communities. By Boxing Day there were 500 Flood Warnings in place across northern England and Wales. December 2015 was subsequently ranked as the wettest month in the Lake District since 1788.

It is estimated that Storm Desmond caused £500 million of damage to homes, businesses and infrastructure across Cumbria[1]. Additional, indirect costs due to disruption of commerce, water/energy services, transport, and communication systems are harder to quantify. Inevitably, the flooding attracted considerable attention by the media, public and politicians. Questions were asked about the effectiveness of flood management strategies and whether the winter was a one-off, or symptomatic of climate change.

Very long rainfall records are an important resource for placing such extreme weather events in their historic context. Unfortunately, weather records longer than 50 years are relatively rare even in the UK. This is because meteorological stations are opened then closed (to save money), or re-sited to make way for other land uses. Instruments and observer practices have changed too. Pre-digital era paper records may be mislaid or accidentally destroyed. Hence, there is growing interest in ‘bridging’ overlapping and neighbouring weather records to create longer, homogeneous series that are broadly representative of a wider region. This lesson looks at one such rainfall index for the English Lake District and shows how different sources of information can be used to test the reliability of record-breaking values.

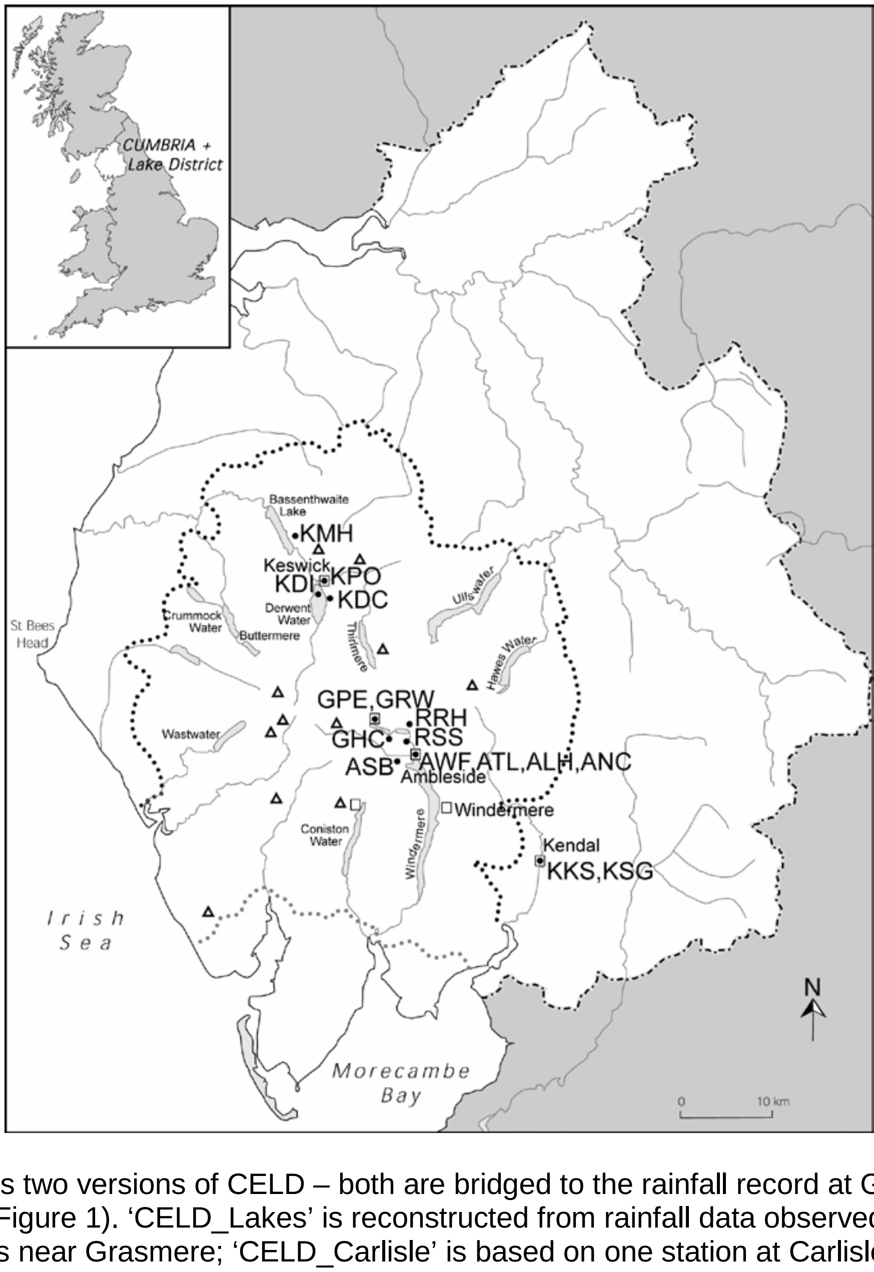
The data

The original Central English Lake District (CELD) rainfall index covered the years 1788 to 2000. The series was constructed by statistically overlapping monthly precipitation records from Kendal (1788–1899), Keswick (1790–1879), Grasmere (1870–1950) and Rydal (1901–2000) (Figure 1).

[1] https://www.theguardian.com/uk-news/2015/dec/08/storm-desmond-damage-cumbria-estimated-500m

The Grasmere record provides the ‘bridge’ between the very early and modern records. The CELD series was last updated to the end of 2016 using rainfall data collected at Rydal Hall. But publication of a long, homogenous rainfall record for Carlisle (1757–2012) (Todd et al., 2015) also raises the possibility of extending the CELD series back to 1757 (Figure 2).

Figure 1. Rain gauges in the Lake District: Ambleside, Lesketh How (ALH), Nook Cottage (ANC), Skelwith Bridge (ASB), The Lakes (ATL), Wansfell (AWF); Grasmere High Close (GHC), Pavement End (GPE); Kendal, Kirbie School (KKS), **St Georges Church (KSG)**; Keswick, Mirehouse (KMH), Deer Close (KDC), Derwent Island (KDI), **Post Office (KPO)**; Rydal, **Stepping Stones (RSS)**, **Rydal Hall (RRH)**. Sites in **bold** were used in the CELD series. Source: Barker et al. (2004)



This lesson uses two versions of CELD – both are bridged to the rainfall record at Grasmere High Close (GHC in Figure 1). ‘CELD_Lakes’ is reconstructed from rainfall data observed at four weather stations near Grasmere; ‘CELD_Carlisle’ is based on one station at Carlisle. The underpinning data have been made available by the authors of the following studies:

(a) Barker, P.A., Wilby, R.L. and Borrows, J. 2004. A 200-year precipitation index for the central English Lake District. *Hydrological Sciences Journal*, **49**, 769-785.

(b) Todd, B., Macdonald, N. and Chiverrell, R.C. 2015. Revision and extension of the composite Carlisle rainfall record, northwest England: 1757-2012. *International Journal of Climatology*, **35**, 3593-3607.

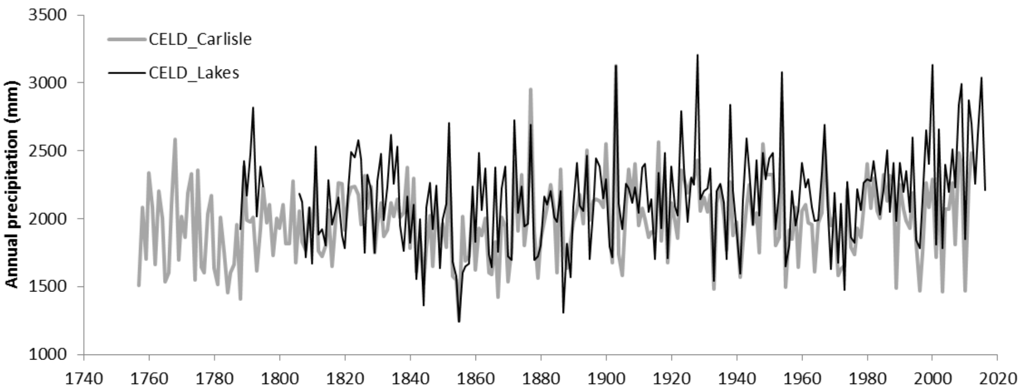
Tasks

1. **How comparable are the reconstructed rainfall series for the Central English Lake District (Grasmere) to those at Carlisle?**
   Download the Microsoft Excel file ‘L4_Data_LakeDistrict.xlsx’, the accompanying ‘Datasheet 4’. [Note that the full data set was used to plot Figure 2]. The CELD estimates for the last 20 years of overlapping data at Carlisle and Grasmere are provided below:

   Table 1. CELD rainfall estimates and rank order from Carlisle and Grasmere, 1993-2012.

Year	Carlisle (mm)	Grasmere (mm)	Carlisle (rank)	Grasmere (rank)
1993	1934.0	2049.2	28	27
1994	2189.9	2599.0	22	8
1995	1816.4	1845.5	31	30
1996	1472.7	1784.4		35
1997	1753.7	2137.4		23
1998	2263.5	2653.8	19	7
1999	2083.5	2406.0	24	12
2000	2292.2	3130.8	16	
2001	1718.2	1809.8		33
2002	2349.0	2656.1	15	6
2003	1466.0	1784.9		34
2004	2070.7	2396.2	26	14
2005	2071.4	2198.5	25	21
2006	2267.6	2512.0	17	9
2007	1813.5	2233.7	32	20
2008	2483.2	2837.0	11	
2009	2403.9	2994.7	13	
2010	1469.4	1848.0		29
2011	2265.0	2871.1	18	
2012	2484.6	2697.5	10	
Mean				
Maximum				
Minimum				
IQR				

Figure 2. Annual rainfall series for the Lake District 1757-2016 reconstructed using records for Carlisle (CELD_Carlisle) and near Grasmere (CELD_Lakes).



Use the data in the first two columns of Table 1 to derive the mean, maximum, minimum and inter-quartile range (IQR) for both series. How similar are the two blocks of data? How representative are these 20 years compared to earlier periods of data (see Figure 2)?

How unusual was the **annual** total for Grasmere in 2015 (3041.4 mm) when compared with the data in Table 1?

2. **How similar are the CELD estimates based on Carlisle and Grasmere rainfall data?** Produce a scatterplot to compare the two CELD estimates using data for the period 1993-2012. Carefully inspect the raw data for outliers. Suggest reasons for any large differences between the estimates for individual years.

   Tip: Add Carlisle to the map of rain gauges (Figure 1) and note the distance scale. Look up the elevations for Carlisle and Grasmere.

   Take it further: Use the Excel CORREL function to derive the Pearson correlation between the CELD_Carlisle and CELD_Lakes indices. Obtain the correlation coefficient for the years 1993-2012 then for all data in the period 1788-1870. How and why does the correlation between the indices depend on the period of record used?

3. **How similar are the rainfall distributions of CELD_Carlisle and CELD_Lakes?** Use the Mann-Whitney U-test to compare the annual CELD_Carlisle and CELD_Lakes data for the 20 years 1993-2012. State your null (H_0) and alternative (H_1) hypotheses. Then manually calculate the U-test in the following steps:
   a) Complete the missing cells in the last two columns of Table 1. These are ranked values based on THE COMBINED SAMPLE. The top 5 and lowest 5 ranks are missing;
   b) Sum the ranks in each column;
   c) Obtain the sample size (m) for the LARGER of the two sums of the ranks (R_m);
   d) Obtain the sample size (n) for the SMALLER of the two sums of the ranks (R_n);
   e) Using R_m, m and n, and the equation in Box 1, calculate the U-statistic.

Box 1. The Mann-Whitney U-test
$$U = mn + \frac{m(m+1)}{2} - \sum R_m$$ Where m and n are the sample sizes of the two groups of data, and R_m is the smaller of the two sums of the ranks.

   f) Refer to a table of critical values for U. If U is less than or equal to the critical value given for sample sizes m and n, then the null hypothesis (H_0) must be rejected. Alternatively, if U is greater than the critical value accept the null hypothesis.

   Based on the result obtained in step f), do the two indices have the same distribution?

   Take it further: Repeat steps a) to d) using the FULL data set (i.e. years 1757-2016) to calculate the U-statistic. Then use the equation in Box 2 to calculate the z-statistic for your value of U. Again determine whether H_0 is accepted or rejected. For example, if the significance level is set at p=0.05 then z must be greater than 1.96 to reject H_0.

   Tip: The Excel RANK() and SUM() functions can help with steps a) and b) respectively.

Box 2. Calculating the significance of the Mann-Whitney *U*-test for large samples (one or both groups with *N*>20)

The following equation is used when one or both groups have more than 20 data values. The equation rests on the assumption that *U* approaches a normal distribution for large samples, so significance can be established using a *z*-test:

$$z = \frac{U - (\frac{1}{2}mn)}{\sqrt{\frac{1}{12}mn(m+n+1)}}$$

The larger the **absolute** value of *z*, the less likely the value of *U* could have occurred by chance and more likely H_0 is rejected. Critical values of *z* are given below for commonly used significance levels (p):

Critical value of test statistic $\lvert z \rvert$	Significance level (p)
1.645	0.10
1.960	0.05
2.576	0.01
3.291	0.001

4. **How unusual was the annual rainfall during 2000-2016 compared with 1900-1916?** Use the data in the first two columns of Table 2 to derive the mean, maximum, minimum and inter-quartile range (IQR) for the two periods. What are the most noteworthy differences between the samples? Is 17 years enough data to make a reliable comparison?

Table 2. Annual rainfall and rank order for CELD_Lakes in 1900-1916 and 2000-2016.

1900-1916 (mm)	2000-2016 (mm)	1900-1916 (rank)	2000-2016 (rank)
2290.4	3130.8	15	
1794.9	1809.8		
1716.7	2656.1		8
3127.0	1784.9		
2109.3	2396.2	25	12
1923.6	2198.5	28	22
2258.5	2512.0	17	10
2226.1	2233.7	20	18
2120.8	2837.0	24	6
2232.6	2994.7	19	
2043.9	1848.0	27	29
2377.3	2871.1	13	
2403.4	2697.5	11	7
2050.6	2259.2	26	16
2146.9	2628.9	23	9
1702.4	3041.4		
		Mean	
		Maximum	
		Minimum	
		IQR	

Take it further: Repeat steps a) to e) in Task 3 using data from CELD_Lakes, this time comparing samples of annual rainfall totals for the period 2000-2016 with 1900-1916. Begin by entering the top 5 and lowest 5 ranks in Table 2 (last two columns).

Based on the *U*-statistic, was the distribution of annual rainfall the same in the two periods?

Plenary

Return to the main lesson questions: (1) how consistent are reconstructions of rainfall for the Lake District (especially for the eighteenth and nineteenth centuries); (2) has rainfall in the region been unusual in recent years?

Ask the students to vote on whether the Carlisle record could be used to reliably estimate rainfall over the Lake District during the years 1757-1787. Compile a table of evidence supporting ‘yes’, ‘no’ and ‘not sure’ positions. What other sources of information could be used to check the reliability of the reconstructed rainfall extremes?

Hint: See Table 1 in http://www.mangeogsoc.org.uk/pdfs/watkins_whyte.pdf

Discuss the various ways in which annual rainfall in the Lake District has changed over the last century. How are interpretations of rainfall trends limited by use of *annual* data?

Further reading

For an accessible account of the exceptional rainfall totals and river flows experienced in December 2015, see: Barker, L., Hannaford, J., Muchan, K., Turner, S. and Parry, S. 2016. The winter 2015/2016 floods in the UK: a hydrological appraisal. *Weather*, **71**, 324-333.

For a note on the record breaking rainfall total in the Lake District, see: Wilby, R.L. and Barker, P.A. 2016. Wettest December in the Lake District for over 200 years. *Weather*, **71**, 76.

For a clear explanation of the Mann Whitney U-test with several worked examples, see: http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric4.html [accessed 2 February 2017]

Link to Pickering flooding resource