New FEH Catchment Descriptors 2025

Updates to *FARL*, *SAAR*, *URBEXT* and *BFIHOST*

Gianni Vesuviano, John Wallbank, Adam Griffin, James Miller

Issue number 3.2
2025-07-08

UK Centre for Ecology & Hydrology

Contents

1. Motivation

The Flood Estimation Handbook's flood frequency methods are underpinned by the *QMED* catchment descriptor equation – a way of estimating the flood with an annual exceedance probability of 50% (2-year return level) using a catchment's quantitative hydrological properties. The *QMED* equation (Kjeldsen *et al.*, 2008) gives an "as-rural" estimate of flow, which can be adjusted using knowledge of urbanization in the catchment. This process makes use of the following five catchment descriptors:

- *AREA* – Catchment Area: the area draining to the measurement point (gauging station or otherwise).
- *SAAR* – Standard-period Average Annual Rainfall: a measurement of the mean annual rainfall total falling on a catchment during the period 1961-1990 (Spackman, 1993).
- *FARL* – Flood Attenuation due to Reservoirs and Lakes: a measurement of the size and location of water bodies, as a proxy for how water storage impacts river flows (Bayliss, 1999).
- $BFIHOST_{19}$ – Baseflow index from HOST soil classes: An estimate of the proportion of baseflow to total flow based on a regression model against the gridded HOST soil classification dataset (Boorman *et al.*, 1995; Griffin *et al.*, 2019).
- $URBEXT_{2000}$ – Urban Extent in the year 2000: A weighted calculation of the amount of urban and suburban land cover within a catchment in the year 2000 (Bayliss *et al.*, 2006).

This report presents a summary of updates to four of these:

- $SAAR_{9120}$ – *SAAR* averaged over the period 1991-2020.
- $FARL_{2015}$ – *FARL* as determined using 2015 spatial data.
- $URBEXT_{2015}$ – Urban extent as determined using 2015 spatial data.
- $BFIHOST_{19SCALED}$ – an update to $BFIHOST_{19}$ to better account for correlations with *FARL* and $FARL_{2015}$.

Each of the catchment descriptors has been updated for best use in flood frequency estimates in risk assessments, and is designed to best describe the current state of a catchment, rather than the historical state. This is because flood frequency estimation is usually performed on catchments with very little, very poor quality, or no flow data, so there is no historical data to reference. The updated catchment descriptors should be better suited for future research and guidance on future estimates of flow under the influence of climate change and anthropogenic change. It is noted that 2015 spatial data is used in preference to newer data as the UKCEH Land Cover Map 2015 (LCM2015: Rowland *et al.*, 2017a; 2017b; 2017c; 2017d) has received both automatic and manual quality control, whereas more recent versions (LCM2017-LCM2023) have received, and are only intended to receive, automatic quality control.

The following sections give, for each new descriptor, an overview of the processes undertaken and how the new dataset differs from older versions. Full descriptions of the processes will be published in individual catchment descriptor reports.

Differences focus on the changes at NRFA stations, but are also shown for a selection of approximately 1500 ungauged locations chosen to be representative the GB river network (covering wide ranges and combinations of *AREA*, *BFIHOST*, *FARL*, *SAAR*, *URBEXT* and *FPEXT*). It is difficult to give precise reasons for large differences between catchment descriptor versions at any ungauged locations, but large-scale patterns in version differences are discussed. Note that *SAAR* is the only fully spatial dataset, all other values are only calculated on a catchment scale at locations on the river network.

Due to the timings of this work, current and new descriptors are compared at the gauged stations included in version 12.1 of the NRFA peak flow dataset, not version 13. However, the underlying datasets of new and current catchment descriptors are identical for both versions, so all the conclusions are the same for the V12.1 and V13 datasets.

2. New catchment descriptors

2.1 $SAAR_{9120}$

Overview

The new $SAAR_{9120}$ descriptor contains the mean annual rainfall over the years 1991 to 2020 inclusive. It is therefore conceptually similar to the existing *SAAR* and $SAAR_{4170}$ descriptors, which contain the mean annual rainfall over the years 1961-90 and 1941-70 inclusive, respectively. Like *SAAR* and $SAAR_{4170}$, $SAAR_{9120}$ exists in both catchment-average values and point values on a 1-km grid. To avoid ambiguity, this report refers to *SAAR* as $SAAR_{6190}$.

Input data

For all parts of the UK and Isle of Man (IoM) except Northern Ireland (NI), $SAAR_{9120}$ is based on the Met Office's HadUK-Grid 30-year mean annual rainfall for the period 1991-2020 (Met Office *et al.*, 2023). This Met Office dataset is natively at the same 1-km resolution as $SAAR_{6190}$ and $SAAR_{4170}$, but is offset from them by 500 metres in both the *x*- and y-directions. However, the Met Office dataset does not cover the whole of the area represented on the FEH Web Service; there are no data around many coastal areas. To cover the whole of the UK (excluding Northern Ireland), the Met Office dataset was buffered outwards by 1 km following five stages.

1. A regression relationship was produced to model the natural logarithm of 1991-2020 mean annual rainfall as a function of mean altitude for each 1-km square, and easting, northing and *DLILLE*[1] of the centre point of each square.
2. The regression value was calculated for the whole FEH Web Service domain.
3. Error value (i.e. natural logarithm of Met Office value minus regression value) was mapped.
4. All locations adjacent to cells with an error value were infilled with the mean error value of all adjacent cells (sharing either an edge or a corner). This buffered the grid of error values by 1 kilometre.
5. Modelled natural logarithm of 1991-2020 mean annual rainfall was added to the buffered error map and the result transformed exponentially. This, rounded to the nearest integer, is $SAAR_{9120}$.

Values in the resulting $SAAR_{9120}$ map are identical to the Met Office data (to the nearest integer) in all places where they overlap – the buffering stages above were only performed in order to generate values for locations where the FEH Web Service extends beyond the boundaries of the Met Office dataset.

[1] "Distance from Lille" – straight-line distance from the point 750000, 80000 on the British National Grid (OSGB1936) – first defined in the Flood Estimation Handbook, volume 2 (Faulkner, 1999).

For Northern Ireland, $SAAR_{9120}$ is based on Met Éireann's 30-year mean annual rainfall for the period 1991-2020 (Curley *et al.*, 2023). This dataset covers the whole island of Ireland, including all parts available on the FEH Web Service. This Met Éireann dataset is natively at the same 1-km resolution as $SAAR_{6190}$ and $SAAR_{4170}$ and is not offset relative to them (meaning that it is offset relative to the $SAAR_{9120}$ grid for the UK and IoM except NI). To match this with the other grid, bilinear interpolation was used to shift the Met Éireann data by 500 metres in both the *x*- and y-directions. As the Met Éireann data were already in integer format, this shifted data constitutes the $SAAR_{9120}$ grid for Northern Ireland.

$SAAR_{9120}$ vs $SAAR_{6190}$ vs $SAAR_{4170}$

Figure 1 and Figure 2 show the value of $SAAR_{9120}$ divided by $SAAR_{6190}$ for the UK and IoM except Northern Ireland, and Northern Ireland respectively. The $SAAR_{6190}$ grids were shifted by 500 metres in the *x*- and *y*-directions to align them with the $SAAR_{9120}$ so that division could take place. Therefore, mean annual rainfall values are not being compared at exactly the same locations.
In Figure 1, $SAAR_{9120}$ values are shown to be similar to $SAAR_{6190}$ in the English Midlands and generally slightly higher everywhere else. The greatest increases are in Scotland southeast of the Caledonian Canal and the far north of England. There are small, widely-distributed areas throughout the map where $SAAR_{9120}$ is less than $SAAR_{6190}$ as well as slightly larger regions, such as parts of Scotland northwest of the Caledonian Canal. The larger-scale proportional reductions in this region relate to mean annual rainfall estimates falling from extremely large (for the UK) values to smaller extremely large values. Given that the dark red regions are intermixed with (smaller) deep blue regions, the "reduction" from $SAAR_{6190}$ to $SAAR_{9120}$ is partly due to a spatial re-distribution of estimated mean annual rainfall, which may have also been slightly affected by the 500-metre shift in $SAAR_{6190}$ necessary to allow the differences to be mapped. It is important to note that the 500-metre shift exists only in Figure 1, not in the data. Figure 3 shows millimetre values of $SAAR_{9120}$ and $SAAR_{6190}$ in northwest Scotland, with 50-km gridlines overlaid to simplify comparison of identical locations. This makes it clear that the only major differences are in the sizes, locations and shapes of the areas experiencing over 3000 mm per year. As the vast majority of this area, particularly the wetter, higher-altitude parts, has never been gauged by a ground-level gauge suitable for estimating average annual rainfall (Perry *et al.*, 2009), it is not necessarily the case that the higher extreme values in the 1961-90 average dataset are more accurate than the slightly less extreme values in the 1991-2020 dataset.

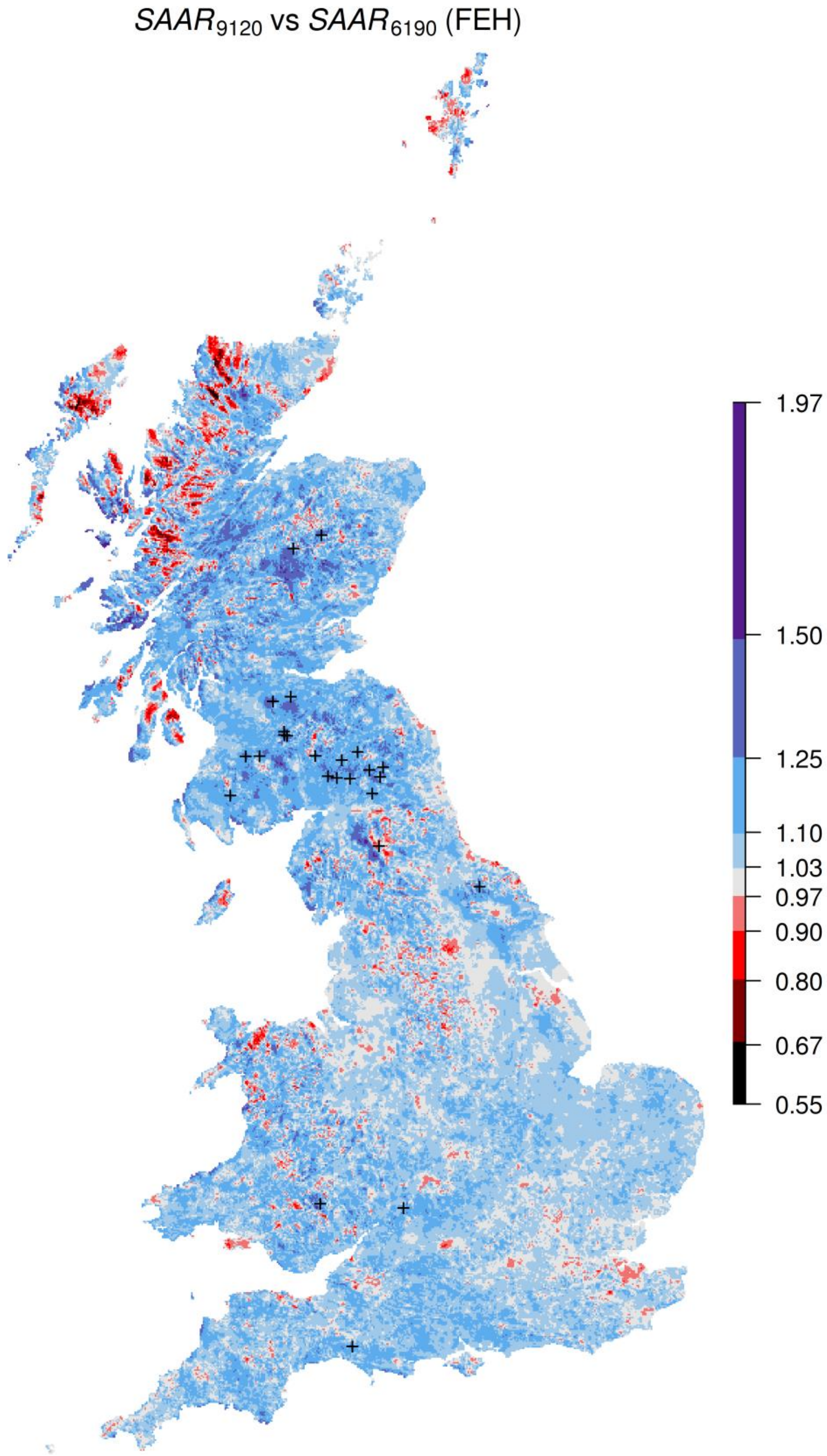

Figure 1 – *SAAR*$_{9120}$ divided by *SAAR*$_{6190}$ for the UK and Isle of Man except Northern Ireland. Black crosses show centroids of 25 catchments where *SAAR*$_{9120}$ has the greatest percentage increase over *SAAR*$_{6190}$.

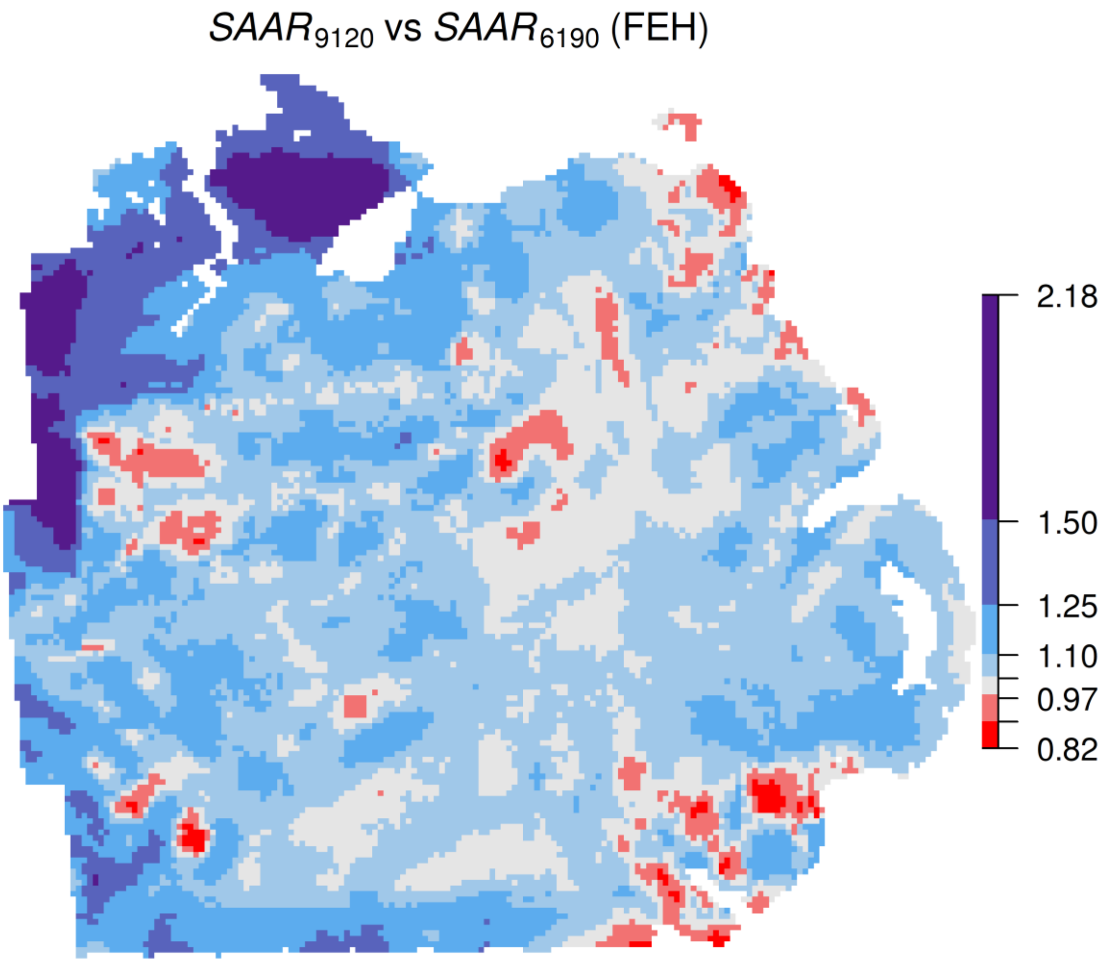

Figure 2 – $SAAR_{9120}$ divided by $SAAR_{6190}$ for Northern Ireland.

In Figure 2, there is little difference between $SAAR_{9120}$ and $SAAR_{6190}$ values across Northern Ireland, with most differences in the −3% to +10% range. There are more significant increases in Donegal, but these can be ignored as the FEH Web Service domain does not extend into these areas, so they will not be available publicly. Here, $SAAR_{9120}$ is considered more trustworthy than $SAAR_{6190}$ as the Met Éireann dataset from which it was derived is intended for accuracy across the whole of Ireland, whereas $SAAR_{6190}$ wasn't. In general, the reductions of 10% or more are confined to the wettest few square kilometres of Northern Ireland and again relate more to changes in the size and shape of the very wettest areas.

For comparison, Figure 4 shows the value of $SAAR_{6190}$ divided by $SAAR_{4170}$ for the UK and IoM except NI. This clarifies that the differences between $SAAR_{9120}$ and $SAAR_{6190}$ are less structured and systematic than the differences between $SAAR_{6190}$ and $SAAR_{4170}$, even though $SAAR_{6190}$ and $SAAR_{4170}$ have 10 years in common, while $SAAR_{9120}$ and $SAAR_{6190}$ have no years in common. The overall reduction in average annual rainfall from $SAAR_{4170}$ to $SAAR_{6190}$ coincides with the transition to a "flood-poor" period. This report is not intended to compare $SAAR_{6190}$

and $SAAR_{4170}$, but to compare the scale of these changes to the mean annual rainfall descriptor against previous changes.

Table 1 lists the 25 catchments where $SAAR_{9120}$ is most different from $SAAR_{6190}$. These are all in GB and marked by plus symbols on Figure 1. As these are catchment-average values, $SAAR_{6190}$ values are derived from a grid that has not been shifted spatially from where it is in the FEH Web Service. All of these 25 largest differences are positive and more than 14%. In contrast, the largest negative difference is only 7.1%. Nineteen of the top 25 catchments with the largest differences are concentrated around south Scotland, an area that has been shown to be getting wetter at a relatively fast rate (Otto *et al.*, 2018), and where gridded $SAAR_{9120}$ is high relative to $SAAR_{6190}$ over a large spatial extent. The remaining six catchments are located in other areas where gridded $SAAR_{9120}$ is greater than $SAAR_{6190}$ over a reasonable spatial extent.

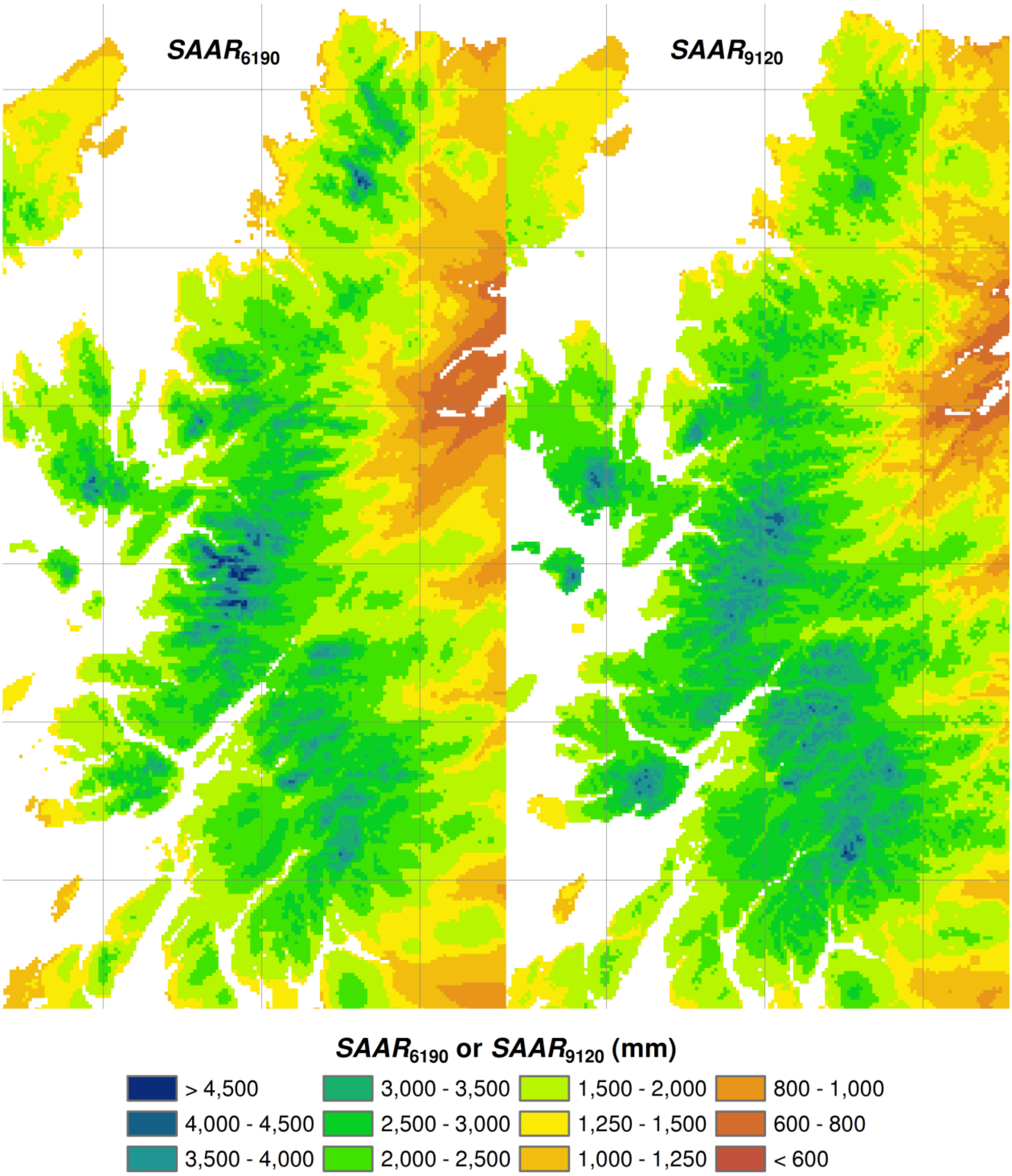

Figure 3 – $SAAR_{6190}$ (left) and $SAAR_{9120}$ (right) in northwest Scotland.

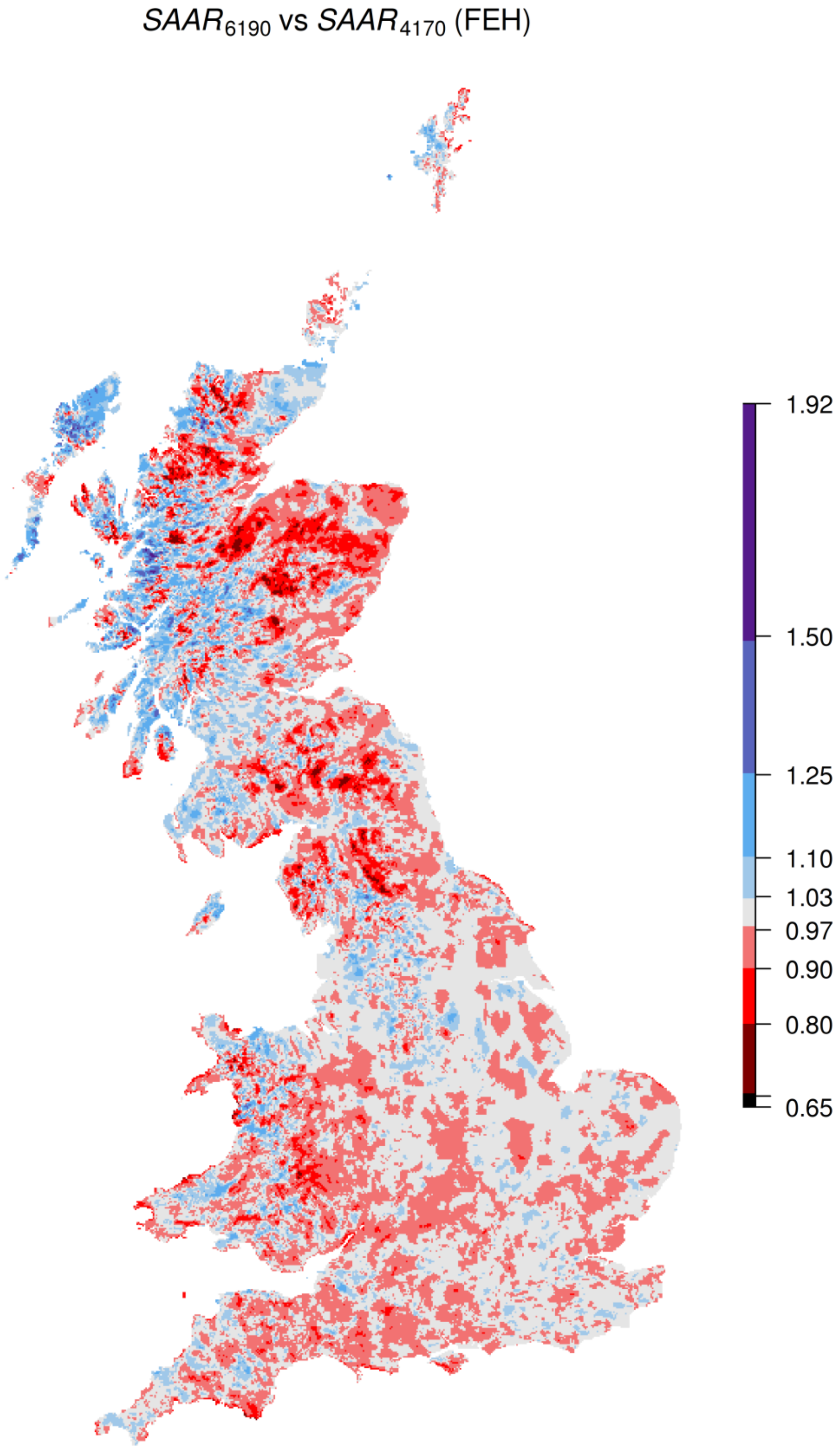

Figure 4 – $SAAR_{6190}$ divided by $SAAR_{4170}$ for the UK and Isle of Man except Northern Ireland.

Table 1 – 25 stations with largest percentage difference between $SAAR_{9120}$ and $SAAR_{6190}$.

Station	$SAAR_{6190}$ (mm)	$SAAR_{9120}$ (mm)	Difference (%)
25808	1499	1972	+24.0
76011	1096	1411	+22.3
84007	928	1143	+18.8
23011	1199	1459	+17.8
79003	1512	1815	+16.7
21012	1149	1367	+15.9
79006	1485	1766	+15.9
56012	1241	1472	+15.7
27010	987	1167	+15.4
77003	1291	1517	+14.9
77002	1423	1671	+14.8
23010	993	1166	+14.8
21017	1740	2043	+14.8
12007	1334	1566	+14.8
77001	1358	1592	+14.7
19005	964	1130	+14.7
84003	1165	1364	+14.6
80003	2468	2883	+14.4
84004	1223	1428	+14.4
12006	1048	1223	+14.3
54026	726	847	+14.3
23033	1024	1194	+14.2
44015	1033	1203	+14.1
21008	936	1090	+14.1
84018	1205	1402	+14.1

Reasons for differences

For a full comparison of the underlying methods and gauging networks see Appendix 1.
The key differences between the two datasets:

- FEH $SAAR_{6190}$ used (in 1993 when developed) more extensive statistical regression to smooth the dataset, beyond what the Met Office considers necessary for monitoring the UK climate.
- FEH $SAAR_{9120}$ is offset by 500m from HadUK-Grid, so some differences observed in the figures are due to misalignment of the datasets.
- There is a large change in the number and location of raingauges over the 60 years of interest, with a shift to fewer, more uniformly distributed

raingauges across the UK (Figure 5). For example, 4822 stations were used in 1961 in annual calculations, and 2706 stations used in 2021 in annual calculations.

- The Met Office’s methods have been transferred through different programming languages and platforms, which may introduce very small differences, although efforts were made at the time to ensure these were minimized.
- There is also the actual difference in rainfall due to inter-decadal variability and climate change, which is the difference the product aims to capture.

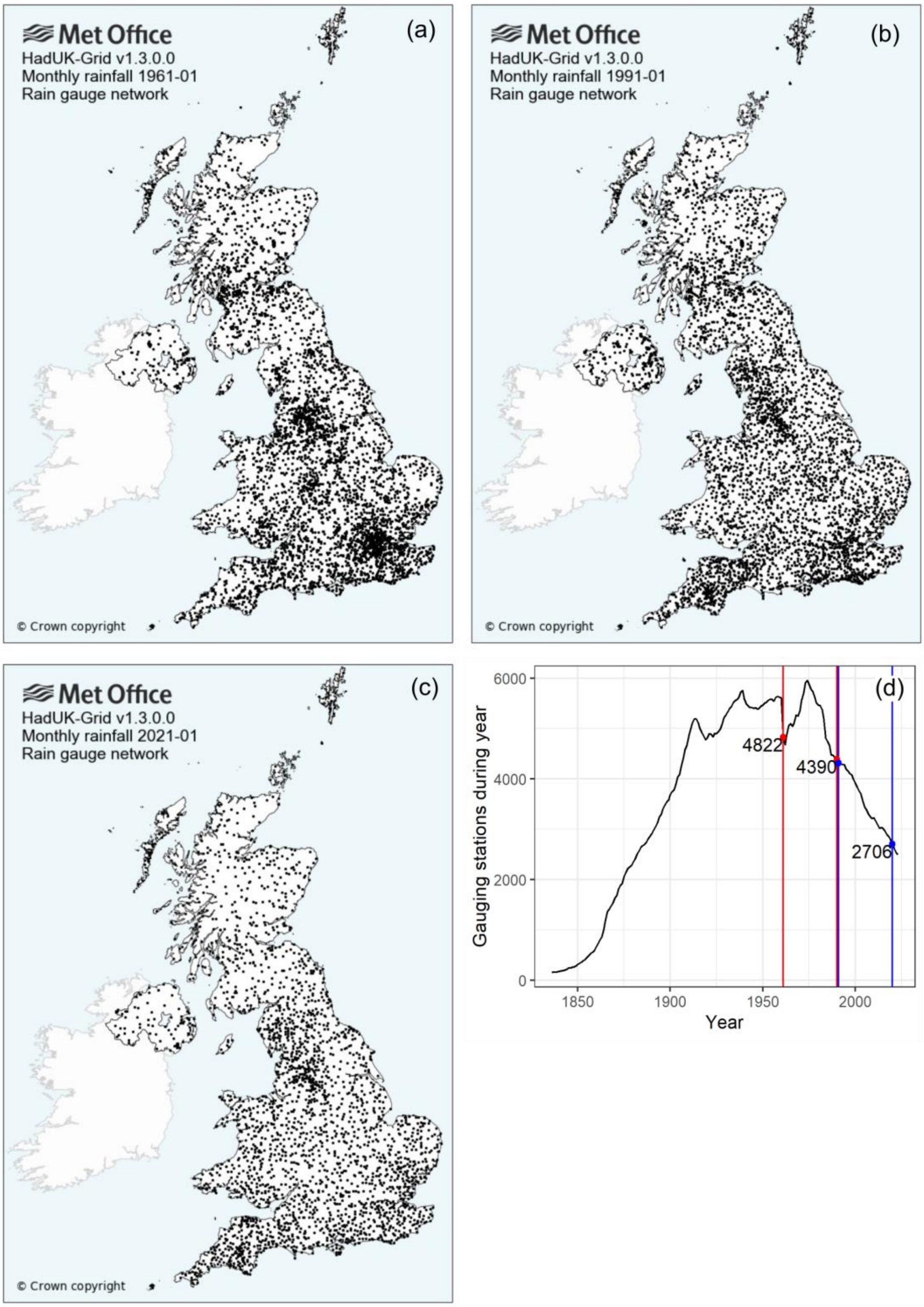

Figure 5 – Coverage of stations used in monthly rainfall calculations for HadUK in (a) January 1961, (b) January 1991, (c) January 2021. Annual station counts given in (d), with standard periods 1961-90 and 1991-2020 highlighted in red and blue respectively.

2.2 $URBEXT_{2015}$

Overview

The new $URBEXT_{2015}$ descriptor measures where urban and suburban areas are, using data collected in the years 2014 and 2015, and aggregates them at catchment-scale. It is therefore conceptually similar to $URBEXT_{2000}$ and $URBEXT_{1990}$. Like these earlier versions, only catchment-average values of $URBEXT_{2015}$ exist.

Input data

$URBEXT_{2015}$ is derived from the UKCEH Land Cover Map 2015 (LCM2015) 25-metre gridded data product (Rowland *et al.*, 2017a; 2017b), which was created by classifying data remotely sensed by Landsat-8 in 2014-2015 into 21 broad habitat classes, two of which are “urban” and “suburban”. At a catchment-average level, $URBEXT_{2015}$ is calculated identically to $URBEXT_{1990}$ as the fraction of the catchment classified as urban plus half of the fraction of the catchment classified as suburban. This is slightly different from how $URBEXT_{2000}$ was calculated from LCM2000, which included the “inland bare ground” category in the calculation (Bayliss *et al.*, 2006). “Inland rock”, the closest category to “inland bare ground” in LCM2015, is almost always found in rural areas only, so was ignored in the calculation of $URBEXT_{2015}$. Gravel car parks, railway sidings and derelict industrial land, which were often classed as “inland bare ground” in LCM2000, were classed as “urban” in LCM2015, with only one identified exception, in southwest Birmingham.

To calculate $URBEXT_{2015}$, a three-class version of LCM2015 was produced with all except the urban and suburban classes set to a general “rural” category. This was aggregated to 50-metre resolution by combining every 2 × 2 group of 25-metre cells into one, resulting in two 50-metre grids aligned with the 50-metre river network underlying the FEH Web Service: one for the UK and IoM excluding NI and another for Northern Ireland. The value in each 50-metre grid cell encoded the number of 25-metre urban and suburban cells contributing to it, where an urban cell was weighted twice as heavily as a suburban cell (a 50-metre cell aggregating four suburban cells would have the same value as one aggregating two urban and two rural cells: one half). Catchment-average values were then calculated by summing the total value of all the 50-metre cells within a catchment and dividing by the quantity of 50-metre cells in that catchment. These values were rounded to four decimal places to give $URBEXT_{2015}$.

LCM2015 follows the Ireland-NI border, while the FEH Web Service domain extends slightly into Ireland, as there are locations in Northern Ireland with headwaters partly in Ireland. These headwater areas are small and do not contain significant settlements, so they were assumed fully rural for calculation of cross-border catchment-average $URBEXT_{2015}$ values.

$URBEXT_{2015}$ vs $URBEXT_{2000}$ vs $URBEXT_{1990}$

Figure 5 compares $URBEXT_{1990}$, $URBEXT_{2000}$ and $URBEXT_{2015}$ values for all 917 NRFA Peak Flow v12.1 catchments. As expected, there is a general increase in urbanization from 1990 to 2000 and 2000 to 2015, with the number of catchments in the lowest urbanization category ("essentially rural") falling and the number of catchments in all other categories slightly rising. However, the majority of catchments remain essentially rural. Categories defined by Bayliss *et al.* (2006) for $URBEXT_{2000}$ are used here for all eras of *URBEXT*. These categories are presented in Table 2.

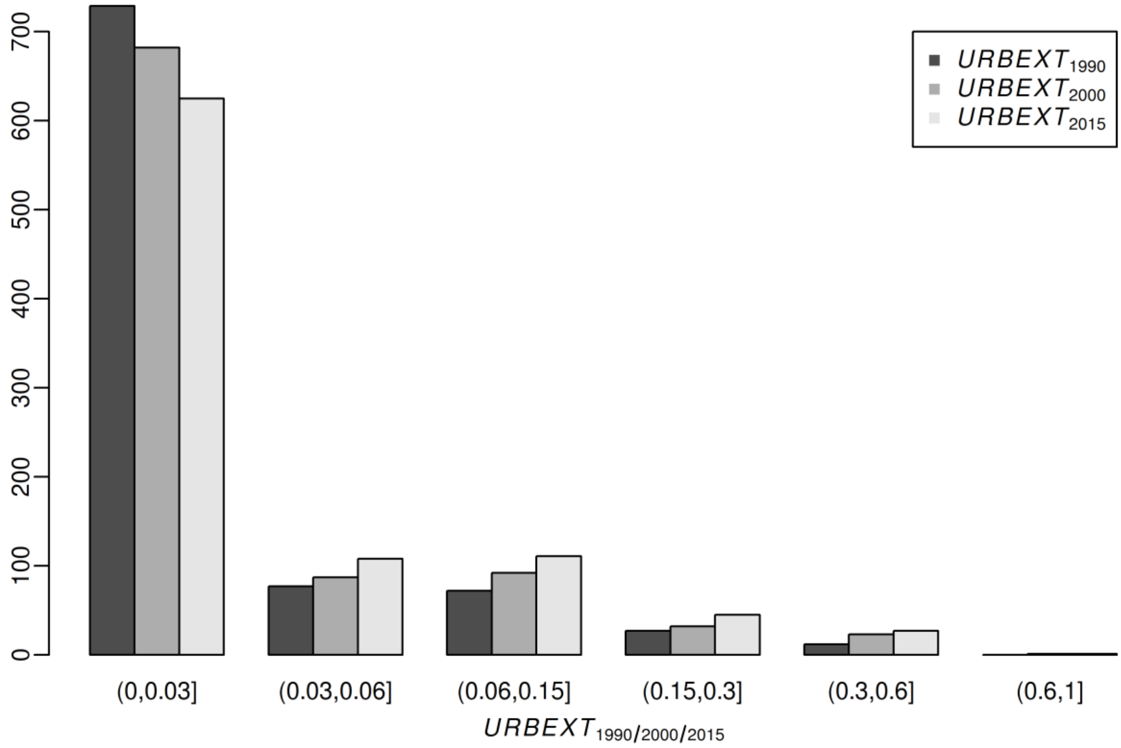

Figure 6 – $URBEXT_{1990}$, $URBEXT_{2000}$ and $URBEXT_{2015}$ value histograms for 917 catchments in NRFA Peak Flow dataset v12.1.

Table 2 – Urbanization categories used by Bayliss et al. (2006) for $URBEXT_{2000}$, and here for $URBEXT_{1990}$ and $URBEXT_{2015}$ in Figure 6.

Category	$URBEXT_{2000}$
Essentially rural	$0.00 \le URBEXT_{2000} < 0.03$
Slightly urbanized	$0.03 \le URBEXT_{2000} < 0.06$
Moderately urbanized	$0.06 \le URBEXT_{2000} < 0.15$
Heavily urbanized	$0.15 \le URBEXT_{2000} < 0.30$
Very heavily urbanized	$0.30 \le URBEXT_{2000} < 0.60$
Extremely heavily urbanized	$0.60 \le URBEXT_{2000} < 1.00$

Figure 6 shows the change in value from $URBEXT_{2000}$ to $URBEXT_{2015}$, which is an estimate of the change in urbanization from 2000 to 2015. As expected, most catchments urbanized slightly. Also as expected, some catchments urbanized considerably, particularly around the edges of settlements that were already large in 2000 (e.g. Liverpool/Manchester, Glasgow/Edinburgh, Vale of Glamorgan, and along the Great Eastern Main Line). A handful of catchments at the edges of London appear to have de-urbanized between 2000 and 2015, with five having an $URBEXT_{2015}$ value that is more than 0.02 less than the $URBEXT_{2000}$ value. All of these have $URBEXT_{2015}$ values above 0.43 and one is nested within another. Figure 7 shows the LCM2000 and LCM2015 data underlying three of these catchments, and it is clear that the lower LCM2015 value is due to the reclassification of land at the edge of London from urban in LCM2000 to suburban in LCM2015. There are five other gauged stations where $URBEXT_{2000}$ exceeds $URBEXT_{2015}$ by more than 0.01, all near London, and all for the same reason (Figure 8 and Table 3). All other 907 gauged catchments either have essentially the same $URBEXT_{2000}$ and $URBEXT_{2015}$ values (difference of 0.01 or less) or a higher $URBEXT_{2015}$ value. It is noted that $URBEXT_{2000}$ very likely overestimates urban area, as the total urban area in LCM2000 significantly exceeds that in LCM2007 (not shown here) – 6.7% vs 5.9% of the UK.

Table 3 – Catchments in the NRFA Peak Flow dataset v12.1 where $URBEXT_{2015}$ is more than 0.01 less than $URBEXT_{2000}$.

Station	$URBEXT_{2000}$	$URBEXT_{2015}$	Difference
37019	0.3391	0.3239	−0.0152
38022	0.5774	0.5048	−0.0726
39005	0.4992	0.4860	−0.0132
39049	0.4020	0.3885	−0.0135
39055	0.5347	0.4515	−0.0832
39056	0.3429	0.3285	−0.0144
39093	0.4781	0.4377	−0.0404
39096	0.5115	0.4845	−0.0270
39134	0.4841	0.4332	−0.0509
40016	0.2607	0.2490	−0.0117

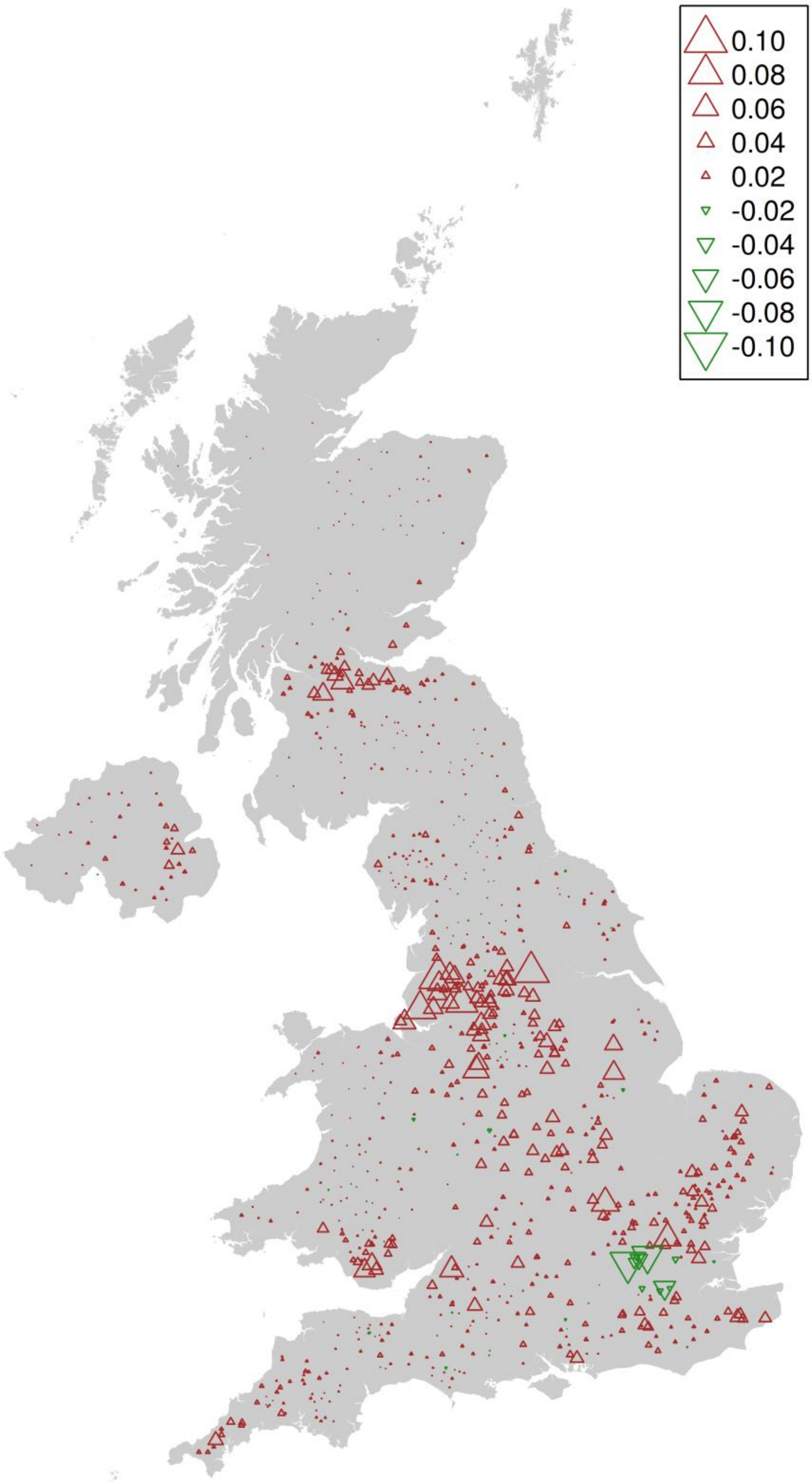

Figure 7 – *URBEXT*$_{2015}$ value minus *URBEXT*$_{2000}$ value for 917 NRFA Peak Flow v12.1 catchments.

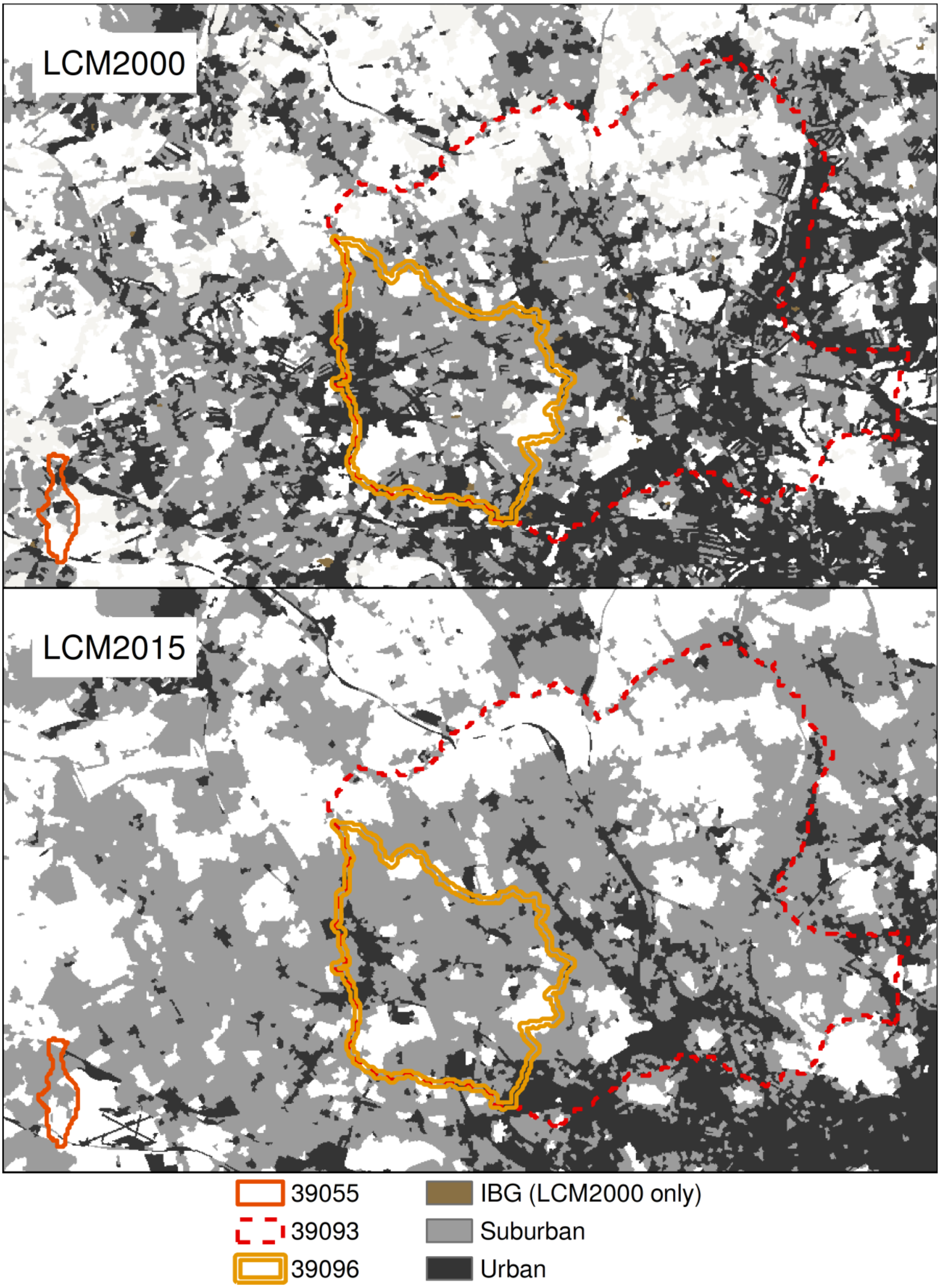

Figure 8 – LCM2000 and LCM2015 representations of northwest London.

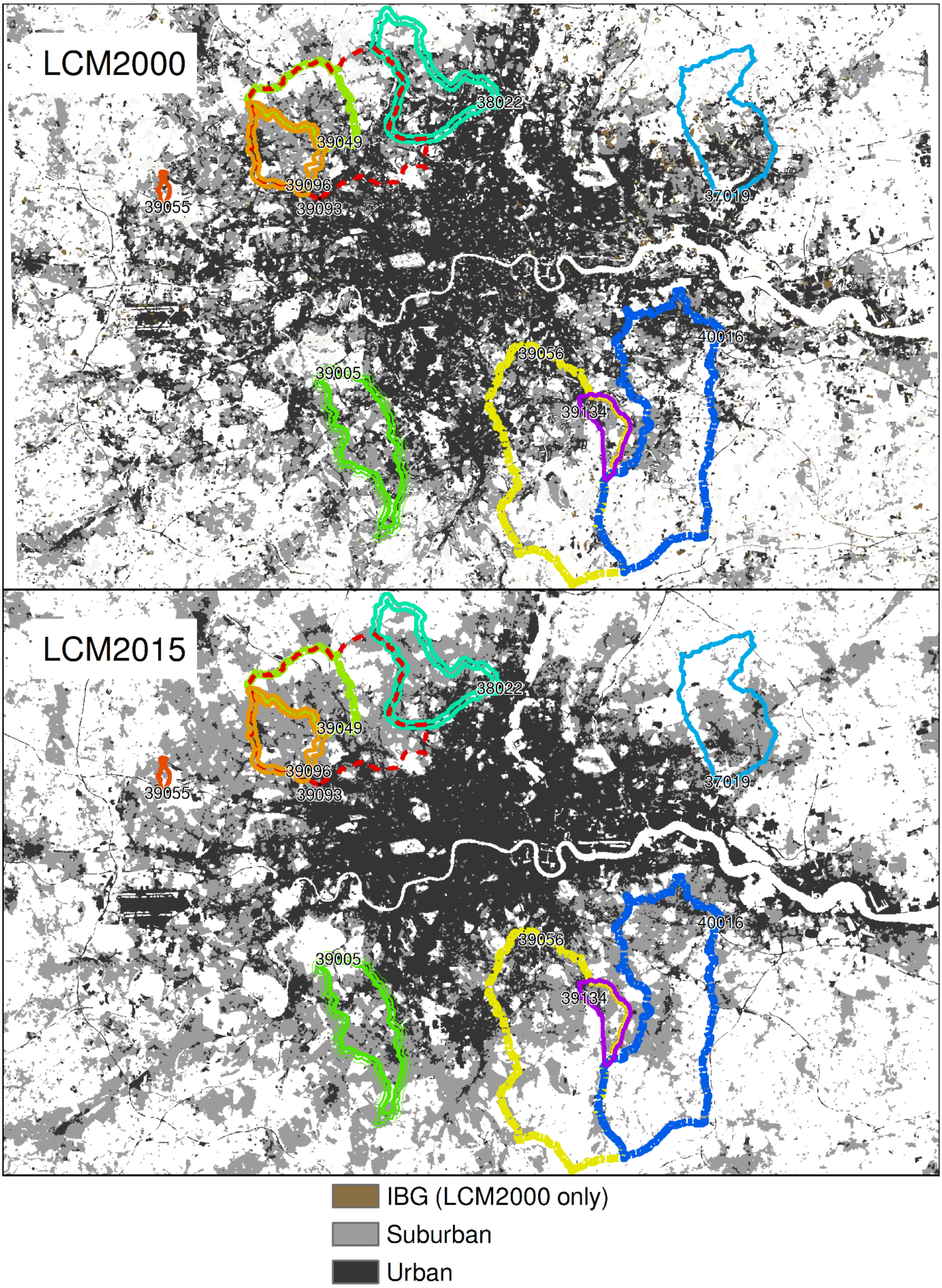

Figure 9 – LCM2000 and LCM2015 representations of Greater London.

Figure 9 plots the difference between $URBEXT_{2015}$ and $URBEXT_{2000}$ for 932 of the 1568 ungauged catchments, the remaining 636 being excluded for being and remaining fully rural (i.e. $URBEXT_{2015} = URBEXT_{2000} = 1$). Much like Figure 6, most catchments show a slight increase in urbanization, while a few in generally urban areas show larger increases. Only 25 catchments show a decrease of 0.02 or more, and 24 of these are smaller than 6 km^2, so very sensitive to the classification of land as “urban” or “suburban”. The other catchment, of 100.3 km^2 is in north London, near 38022.

Some catchments near Birmingham show larger apparent decreases in urbanization of 0.1 or more, which appear to be linked to the reclassification of areas at the edge of Birmingham from urban to suburban, similarly to Greater London (Figure 11).

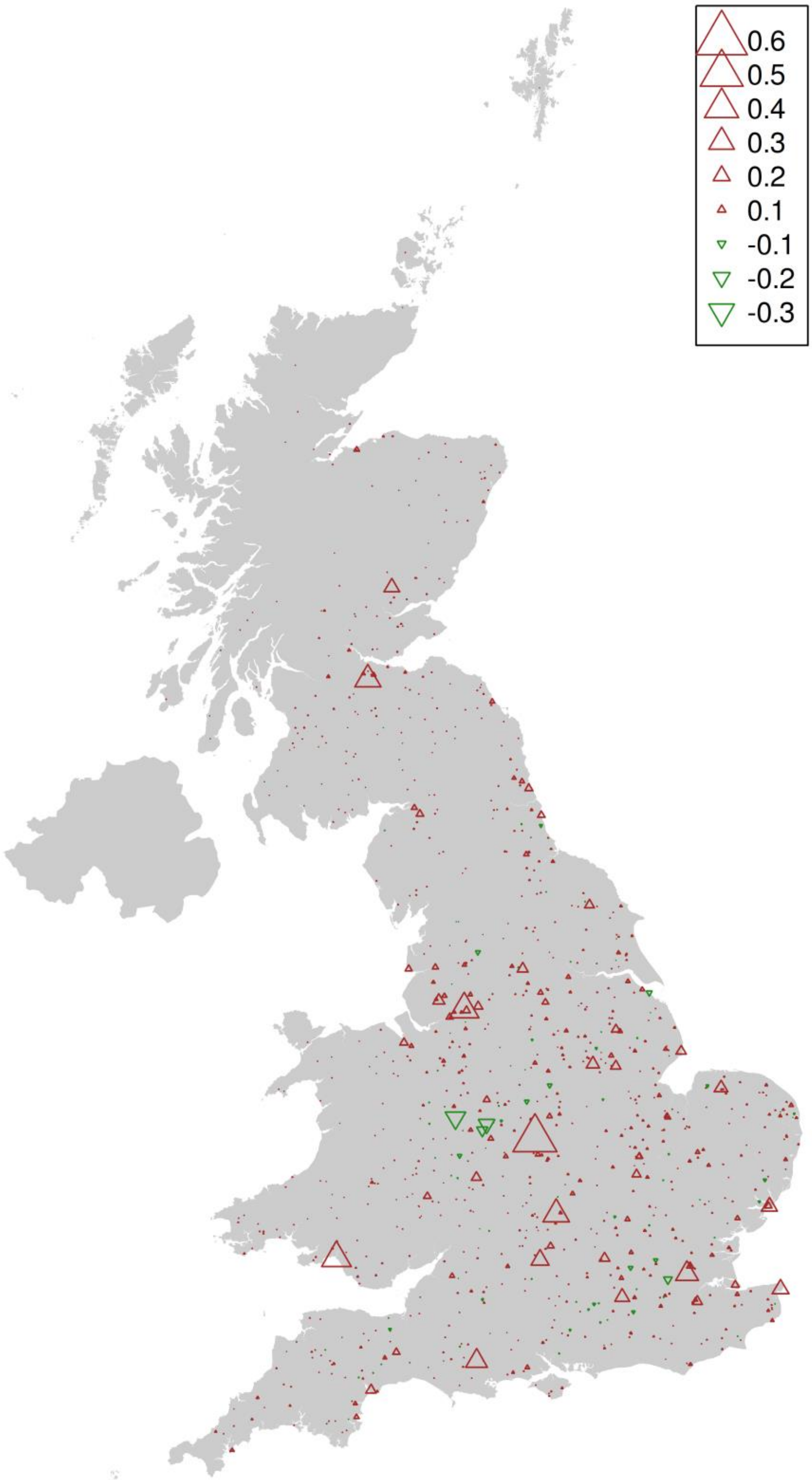

Figure 10 – *URBEXT*$_{2015}$ value minus *URBEXT*$_{2000}$ value for 932 ungauged catchments.

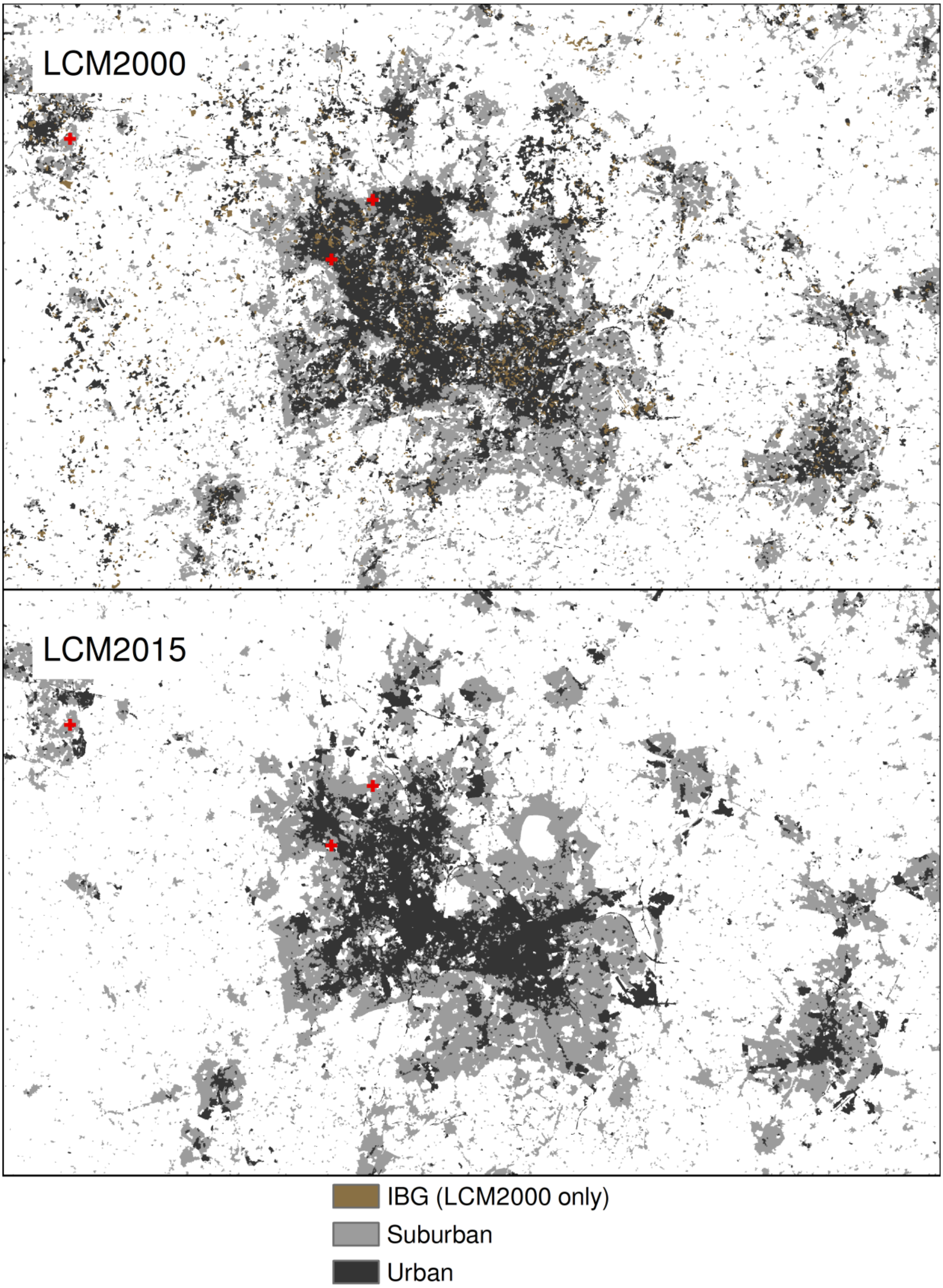

Figure 11 – LCM2000 and LCM2015 representations of Birmingham.

2.3 *FARL*$_{2015}$

Overview

The new *FARL*$_{2015}$ descriptor estimates in-river flow attenuation resulting from on-line upstream reservoirs and lakes, mapped using data collected in the years 2014-2015. It is therefore conceptually similar to *FARL*. As both *FARL* and *FARL*$_{2015}$ values depend on the size and location of reservoirs and lakes relative to the entire drainage area, it is impossible to produce point values of either descriptor, so only catchment-average values exist.

Input data

The main datasets used in the creation of *FARL*$_{2015}$ were the UKCEH Land Cover Map 2015 (LCM2015) vector dataset (Rowland *et al.*, 2017c; 2017d), the UKCEH 1:50k digital river network (Moore *et al.*, 1994), and the IHDTM digital terrain model underpinning the FEH Web Service (Morris & Flavin, 1990; 1994). CORINE Land Cover 2018 (CLC2018: Copernicus Land Monitoring Service, 2020) and the Ordnance Survey (OS) Terrain 50 DTM (Ordnance Survey, 2021) were also used in minor roles. LCM2015 is mainly based on remotely sensed data from Landsat-8, while CLC2018 is based on Sentinel-2 data, with Landsat-8 used to fill gaps. The UKCEH digital river network was originally based on Ordnance Survey 1:50k mapping.

FARL$_{2015}$ was calculated in the same way as *FARL*. However, the use of more recent land cover data meant that more lakes and reservoirs were identified, while the use of vector data meant that the surface areas of water bodies were more accurate and water bodies could more accurately be classed as on- or off-line.

The IHDTM was used to define catchments, as it was for all other descriptors. The OS DTM was used only to assess whether groups of water body polygons that touched each other should be combined and treated as single water bodies, rather than multiple separate ones. Water body polygons that touched were combined if their OS DTM elevations were within 0.1 metres. Due to the unavailability of the OS DTM in Northern Ireland, the IHDTM was used to determine whether different polygons represented the same water body, however elevation differences of up to 1 metre were permitted. Since LCM2015 follows the Ireland-Northern Ireland border, but some catchments cross the border, CLC2018 was used as a source for water bodies in Ireland that affect *FARL*$_{2015}$ calculations for places in Northern Ireland.

Cotswold Water Park (Figure 11) is an example of an area where vector representation of water bodies can be used both to correct the on-line/off-line status of several lakes and to separate lakes that are incorrectly combined in the 50-metre raster land cover data used to create *FARL*.

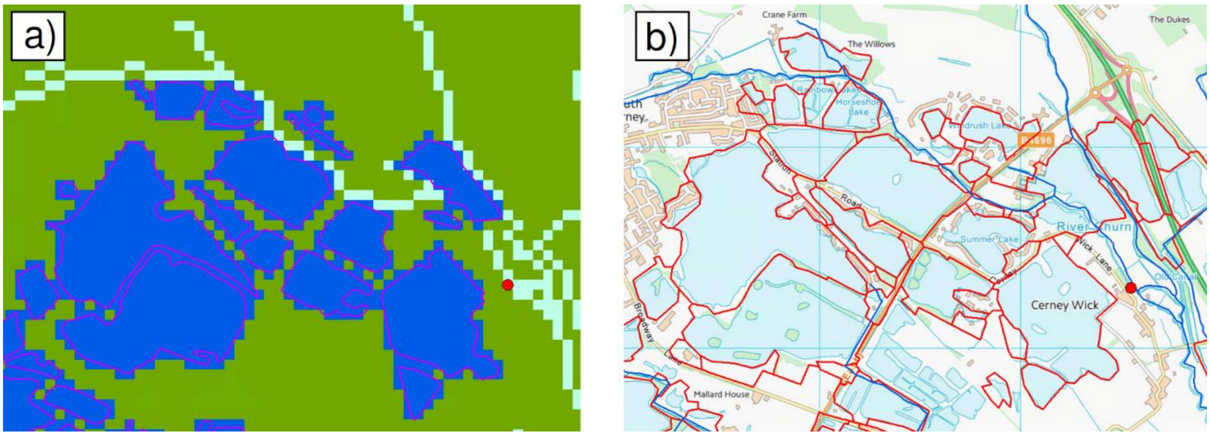

Figure 12 – Representations of Cotswold Water Park: a) IHDTM, where rivers are light blue and water bodies are dark blue; b) LCM2015 (red outlines) and UKCEH 1:50k digital river network (blue lines) over OS Open Raster map data. © Crown copyright 2022 OS.

$FARL_{2015}$ vs *FARL*

Figure 12 plots $FARL_{2015}$ against *FARL* for all NRFA Peak Flow v12.1 catchments. For eight stations, *FARL* values in the Peak Flow dataset do not match those available for the same point on the FEH Web Service, because the *FARL* values in the Peak Flow dataset are relevant to the water bodies that existed when the stations were open and collecting valid peak flow data (i.e. the *FARL* value is matched to the valid peak flow data). For these stations, $FARL_{2015}$ values were also recalculated with the same lakes/reservoirs excluded. Figure 12 shows that $FARL_{2015}$ differs from *FARL* by 0.05 or more for 27 catchments, for one reason in 24 cases and two reasons in 3 cases.

The most common reason for differing $FARL_{2015}$ and *FARL* values is “sand extraction or similar”: many lakes resulting from sand or gravel extraction are present in LCM2015 but not in the surface-type dataset used to create the original *FARL* descriptor. These sand/gravel extraction lakes are most commonly found in the Trent and Nene catchments but may be found elsewhere in smaller quantities. The second most common reason for differences between $FARL_{2015}$ and *FARL* is the absence of major water bodies (Carsington Water, Roadford Lake and Caldecotte Lake) in the dataset used to generate *FARL*. A full list of reasons for all 27 highlighted catchments is given in Table 4. In five of the 27 catchments (33048, 40012, 42007, 44006, 44015), it is unclear which of *FARL* or $FARL_{2015}$ is most accurate. While this does not negate the clear improvements shown in the majority of cases, the potential errors in $FARL_{2015}$ combined with the known errors in *FARL* do demonstrate that any consistently generated, national-scale dataset will contain some errors. Unfortunately, it is not practical to correct missing lakes in $FARL_{2015}$, as it is impractical to review the entire LCM at this level of detail. It should be noted that Table 4 can only show lakes missing from gauged catchments when they are sufficient to cause a difference of more than 0.05 between *FARL* and $FARL_{2015}$, so the missing lakes mentioned there are far from exhaustive at whole UK-scale.

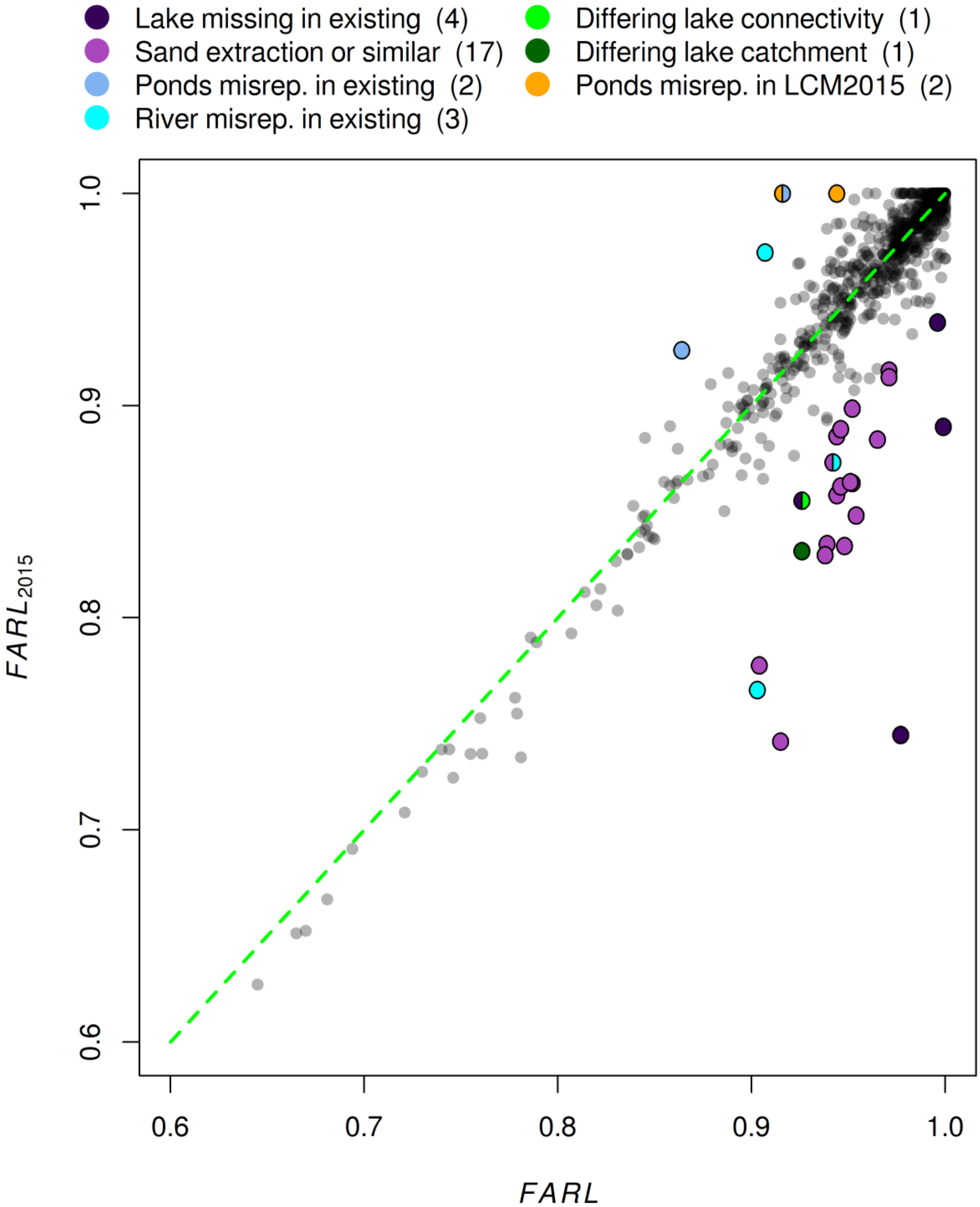

Figure 13 – *FARL*$_{2015}$ vs *FARL* for 917 catchments in NRFA Peak Flow dataset v12.1, with reasons for large differences highlighted.

Table 4 – Reasons for differences above 0.05 between *FARL*$_{2015}$ and *FARL*.

Station	*FARL*	*FARL*$_{2015}$	Reason for *FARL*$_{2015}$ value
27030	0.952	0.8635	Sand/gravel extraction lakes identified in LCM2015.
28005	0.939	0.8347	Sand/gravel extraction lakes identified in LCM2015.
28007	0.954	0.8482	Sand/gravel extraction lakes identified in LCM2015.
28009	0.944	0.8578	Sand/gravel extraction lakes identified in LCM2015.
28015	0.904	0.7775	Sand/gravel extraction lakes identified in LCM2015.
28019	0.948	0.8338	Sand/gravel extraction lakes identified in LCM2015.
28022	0.946	0.8619	Sand/gravel extraction lakes identified in LCM2015.
28058	0.977	0.7447	Carsington Water present in LCM2015.
28095	0.938	0.8295	Sand/gravel extraction lakes identified in LCM2015.
32010	0.915	0.7416	Sand/gravel extraction lakes identified in LCM2015.
33014	0.944	0.8856	Sand/gravel extraction lakes identified in LCM2015.
33015	0.926	0.8552	Caldecotte Lake present in LCM2015. Willen Lake correctly represented as two lakes (not one) by LCM2015.
33048	0.907	0.9722	Two lakes correctly changed from on-line to off-line. One lake missed by LCM2015.
34001	0.971	0.9165	Extra lakes (not necessarily sand/gravel) identified by LCM2015.
38001	0.952	0.8986	Sand/gravel extraction lakes identified in LCM2015.
39001	0.942	0.8732	Reclassification of some water boides from on-line to off-line, and vice versa. Catchment includes Cotswold Water Park.
39006	0.951	0.8639	Sand/gravel extraction lakes identified in LCM2015.
39008	0.946	0.8889	Catchment includes Cotswold Water Park and catchment 39042.
39010	0.903	0.7659	Lake outlets correctly moved from Grand Union Canal to River Colne.
39016	0.965	0.8840	Sand/gravel extraction lakes identified in LCM2015.
39042	0.971	0.9134	Sand/gravel extraction lakes identified in LCM2015. These are near the gauging station, so they have an outsize effect on *FARL*$_{2015}$.
40012	0.926	0.8314	Lake outlet moved from one branch to another. Unknown which is correct.
42007	0.864	0.9261	Watercress beds not identified by LCM2015 as lakes.
44006	0.944	1.0000	Three ponds (< 0.5 ha) missed by LCM2015 partially due to tree growth.
44015	0.916	1.0000	Algae obscures one pond in LCM2015. One real trout pond missed by LCM2015, one erroneous (non) “trout pond” correctly excluded by LCM2015. Change of one other pond from on-line to off-line, unknown if correct.
47006	0.996	0.9392	Roadford Lake present in LCM2015.
47008	0.999	0.8900	Roadford Lake present in LCM2015.

Figure 13 plots the difference between $FARL_{2015}$ and *FARL* for 459 of the 1568 ungauged catchments, the remaining 1109 being excluded for having no on-line water bodies (i.e. $FARL_{2015}$ = *FARL* = 1).
There is a cluster of four catchments in the Nene where $FARL_{2015}$ is from −0.17 to −0.31 less than *FARL*. This has previously been identified as an area where sand/gravel extraction is prevalent. The largest change occurs for the smallest catchment (18 km^2).

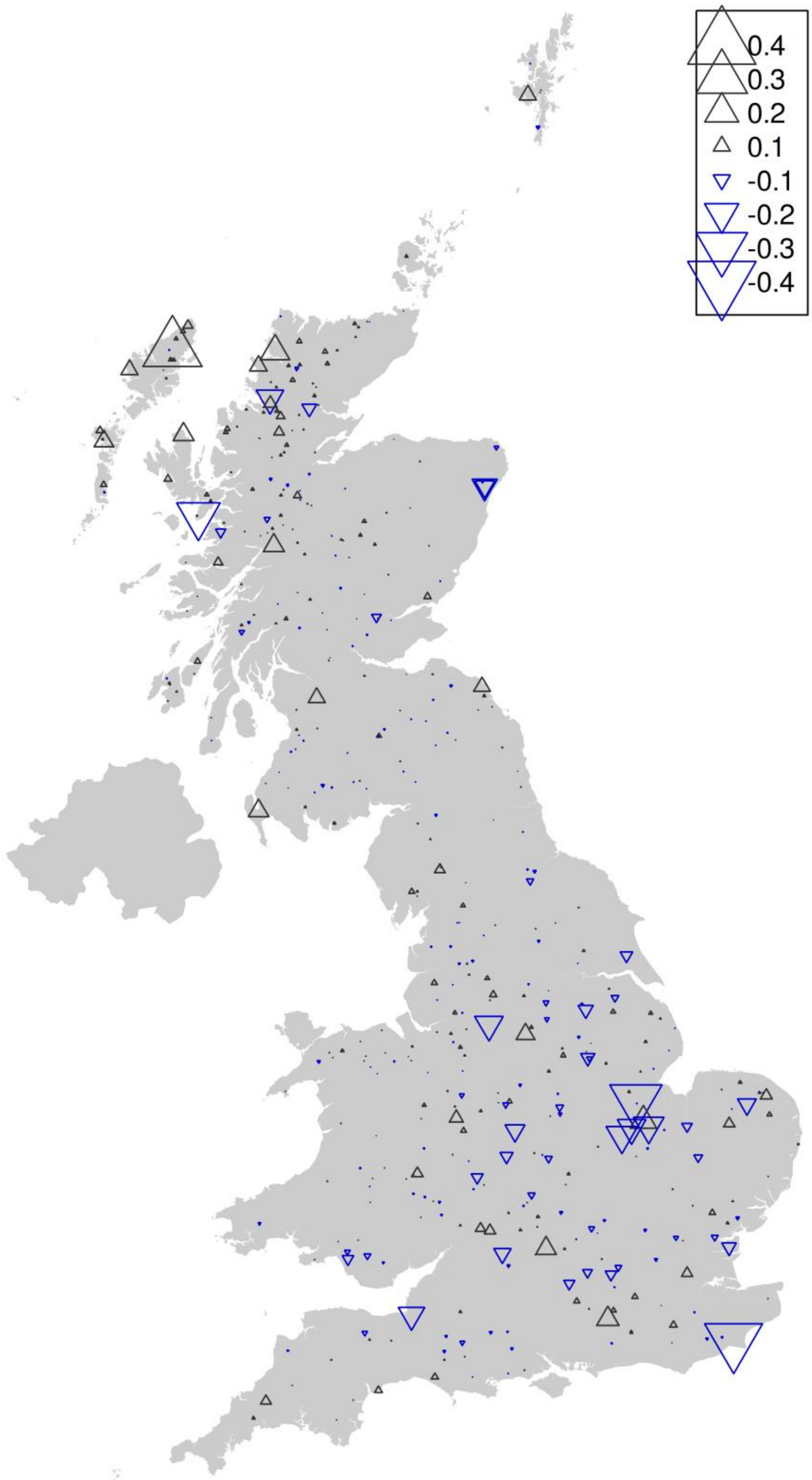

Figure 14 – *FARL*$_{2015}$ value minus *FARL* value for 459 ungauged catchments.

2.4 $BFIHOST_{19SCALED}$

Overview

The new $BFIHOST_{19SCALED}$ descriptor is a slightly modified version of $BFIHOST_{19}$ (Griffin *et al.*, 2019). While catchment-average $BFIHOST_{19}$ is calculated from 30 HOST classes (29 soil classes and surface water), $BFIHOST_{19SCALED}$ is calculated only from the 29 soil classes. Within each catchment that contains surface water, the fractional coverage of surface water is reduced to zero, and the fractional coverages of the 29 soil classes are increased (divided by the total fractional coverage of the 29 soil classes) so that the coverage of the 29 soil classes sums to 1. $BFIHOST_{19SCALED}$ is calculated on these scaled classes using exactly the same method as used to calculate $BFIHOST_{19}$.
Because $BFIHOST_{19SCALED}$ uses the gridded HOST dataset to identify surface water, but $FARL_{2015}$ uses LCM2015, there is not a 100% correspondence between the water bodies removed from $BFIHOST_{19}$ to give $BFIHOST_{19SCALED}$ and the water bodies used to calculate $FARL_{2015}$. As the HOST grids and LCM2015 vector dataset are at very different resolutions, it is not considered practical to use the LCM2015 data to either supplement or replace the representation of lakes in the HOST grids. Furthermore, $FARL_{2015}$ is only affected by on-line water bodies, while the HOST surface water class does not distinguish between on-line and off-line. Nevertheless, $BFIHOST_{19SCALED}$ does overall achieve its aims of improving the variance explained by the new *QMED* equation and reducing the standard errors in the estimates of the coefficients in that equation.

Input data

$BFIHOST_{19SCALED}$ is produced using exactly the same datasets as $BFIHOST_{19}$: 1-km gridded HOST class percentage coverage per grid cell (Boorman *et al.*, 1995) and the IHDTM (Morris & Flavin, 1990; 1994), each of which exists as two grids aligned to different coordinate systems: one for the UK and IoM excluding NI (aligned to the British national grid, OSGB1936), and one for NI (aligned to the Irish national grid, TM75). The HOST class grids are used to provide soil class coverages from which $BFIHOST_{19SCALED}$ is calculated while the IHDTM grids are used to define the river network for catchment-averaging of values.
The calculation of $BFIHOST_{19}$ is detailed by Griffin *et al.* (2019). Calculation of $BFIHOST_{19SCALED}$ only differs from the data and methods published there by the removal of HOST class 30 prior to calculation. Each 1-km grid cell is considered in turn. If HOST class 30 coverage in a cell is greater than zero, then it is set to zero and the coverages of the other 29 HOST classes are increased, maintaining their relationships relative to each other, so that the total coverage of classes 1-29 is 100%. For each catchment along the river network, $BFIHOST_{19SCALED}$ is calculated using the same method, procedure and regression coefficients as $BFIHOST_{19}$, the only difference being the prior removal of HOST class 30 and rescaling of HOST classes 1-29. It must be noted that catchment-average $BFIHOST_{19SCALED}$ (and catchment-average $BFIHOST_{19}$) is not simply a weighted average of $BFIHOST_{19SCALED}$ (or $BFIHOST_{19}$) in each contributing 1-km grid cell. Instead,

catchment-average coverages of each HOST class are found, then *BFIHOST*$_{19SCALED}$ (or *BFIHOST*$_{19}$) is calculated from these.

BFIHOST$_{19SCALED}$ vs *BFIHOST*$_{19}$

As expected, *BFIHOST*$_{19SCALED}$ values are equal to or lower than *BFIHOST*$_{19}$ values (with two exceptions, discussed later). Also, as expected, there is a general but not exact correlation between the value of *FARL*$_{2015}$ and the difference between *BFIHOST*$_{19SCALED}$ and *BFIHOST*$_{19}$ (Figure 14). This is because *FARL*$_{2015}$ depends on the size, location and on-line status of vector water bodies in LCM2015, while the difference between *BFIHOST*$_{19SCALED}$ and *BFIHOST*$_{19}$ depends on the size, but not location or on-line status, of water bodies in the 1-km gridded HOST dataset. For water bodies that do not occupy a full 1 km^2 grid cell, the water is distributed equally across the grid cell, while with LCM2015, it is easy to determine whether the water body is fully in or fully out of the catchment.

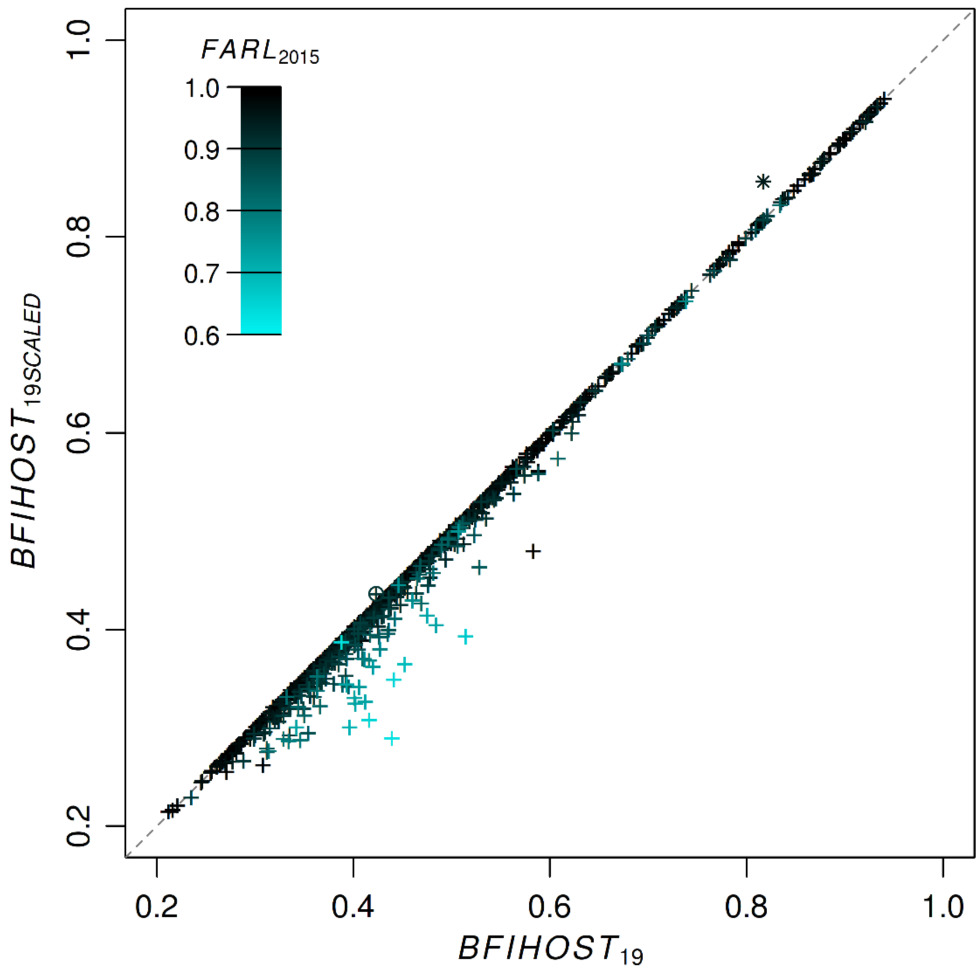

Figure 15 – *BFIHOST*$_{19SCALED}$ vs *BFIHOST*$_{19}$ for 917 NRFA Peak Flow v12.1 catchments, coloured by *FARL*$_{2015}$ value.

$BFIHOST_{19SCALED}$ should always be lower than $BFIHOST_{19}$. However, there are clearly two catchments where this is not the case. The circled catchment, 205034, is regularly excluded from use as a donor due to a serious mismatch between the true size and shape of the catchment (~0.1 km^2), and its IHDTM representation from which FEH catchment descriptors are derived (5.74 km^2), so the values of $BFIHOST_{19SCALED}$ and $BFIHOST_{19}$ for this gauged catchment are unimportant from a practical flood risk perspective. It is noted that a station's suitability for *QMED* estimation is based solely on the quality and uncertainty of the gauged peak flows as assessed by the NRFA, hence not at all on how accurately its upstream catchment is represented digitally. The starred catchment, 39003, has no HOST class 30 coverage, so $BFIHOST_{19SCALED}$ and $BFIHOST_{19}$ should be equal. However, $BFIHOST_{19SCALED}$ was calculated directly from HOST classes summarized to catchment level under a new procedure, not involving $BFIHOST_{19}$ in any way. This procedure made negligible difference in all cases except catchments 39003 and 205034.

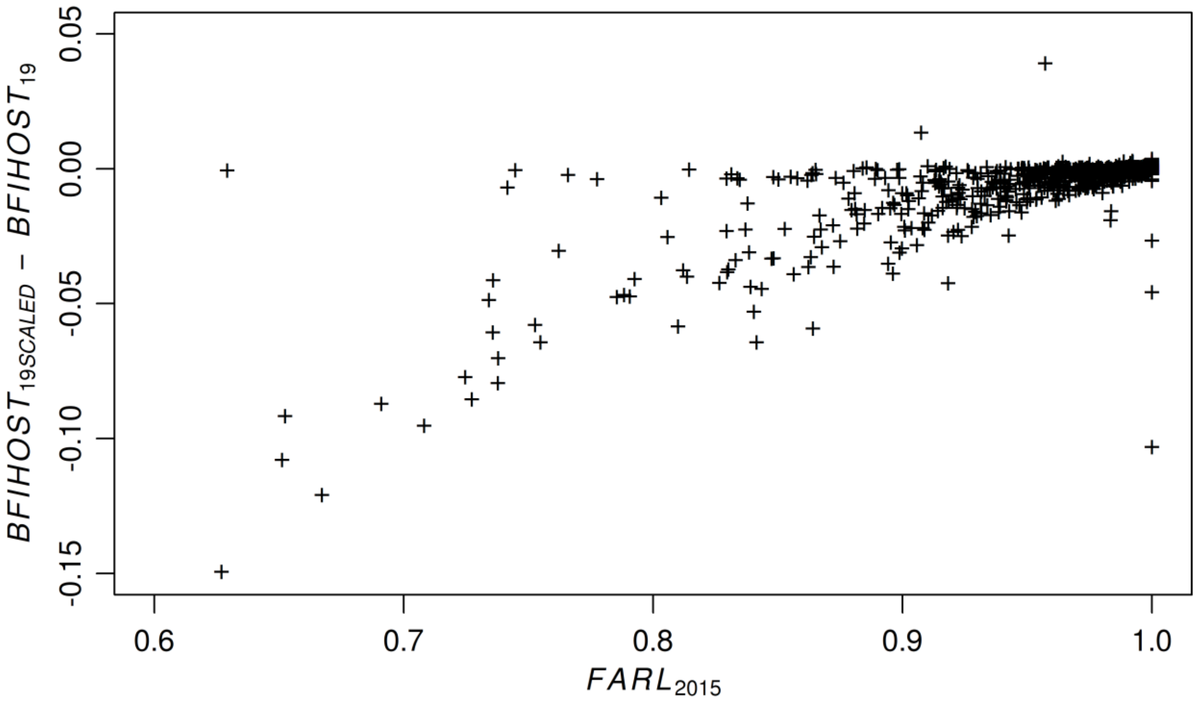

Figure 16 – difference between $BFIHOST_{19SCALED}$ and $BFIHOST_{19}$ vs $FARL_{2015}$ for 917 NRFA Peak Flow v12.1 catchments.

The three "outlier" catchments in Figure 15, where $FARL_{2015}$ is 1 *and* $BFIHOST_{19SCALED}$ is noticeably below $BFIHOST_{19}$ are 45817, 25808 and 31026, from smallest to largest difference. All three of these catchments are 2.30 km^2 or smaller, so will be very sensitive to the exact methods used to extract catchment-average HOST class coverages.
Figure 16 shows the difference between $BFIHOST_{19SCALED}$ and $BFIHOST_{19}$ for 1568 ungauged catchments. The overall distribution of points is similar to Figure 15.

While there appear to be many catchments where $FARL_{2015}$ is 1 and $BFIHOST_{19SCALED}$ is noticeably below $BFIHOST_{19}$, there are in fact only 13 catchments where the difference is more than 0.05. All of these are under 8 km^2, so could be affected significantly by average-sized water bodies. Nine are in the Scottish Highlands, where there are many notable water bodies, while the other four are located near St Mary's Loch, Loch Grannoch, Covenham Reservoir, and a series of lakes near Knaresborough, North Yorkshire.

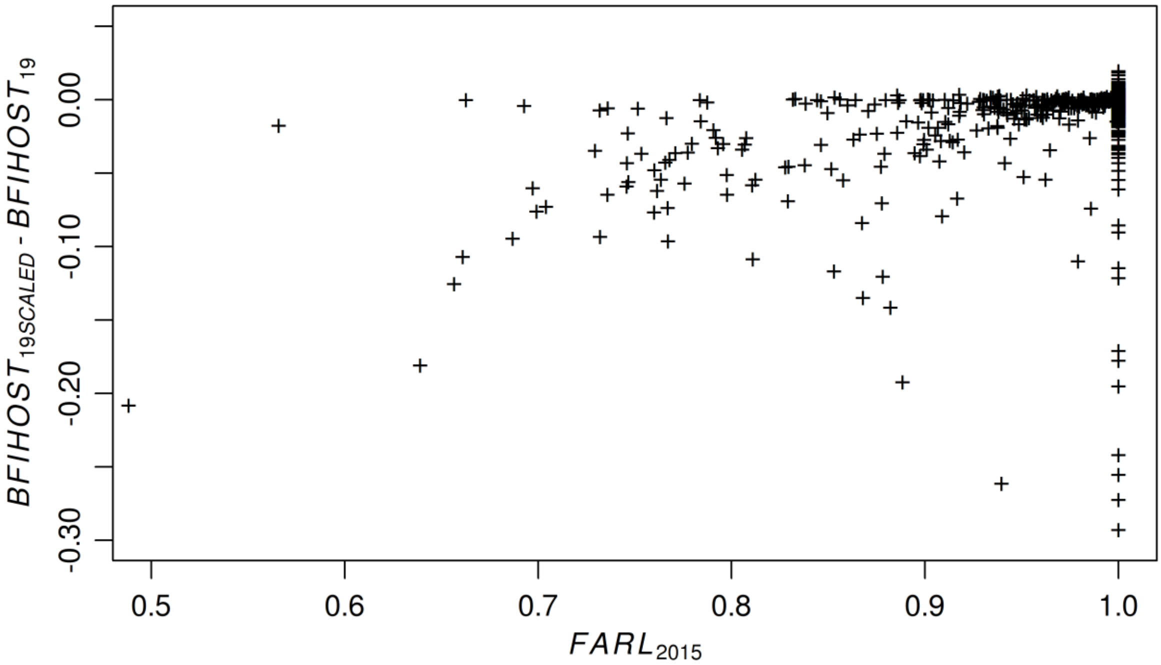

Figure 17 – difference between $BFIHOST_{19SCALED}$ and $BFIHOST_{19}$ vs $FARL_{2015}$ for 1568 ungauged catchments.

For a comparison of $BFIHOST_{19}$ against the original *BFIHOST* descriptor, see Griffin *et al.* (2019).

3. Conclusions

The new catchment descriptors are used to better describe the current state of ungauged catchments with an aim to improve flood frequency estimates for ungauged catchments. As with all FEH methods, expert local knowledge is always crucial, to verify whether national estimates, particularly those based on modelled values like HOST, can be improved by local knowledge. Although not described here, there is uncertainty inherent in all catchment descriptors, and uncertainty is higher where data quality is poor. Differences between different versions of descriptors stem from a combination of model changes, new data, measurement error, modelling error (in the new and old versions) and uncertainty. Unfortunately, a full quantification of uncertainty in catchment descriptors would require a full understanding of uncertainty in the UKCEH Land Cover Map, IHDTM, HOST dataset, Ordnance Survey data, Met Office rainfall products, and all the data underlying these datasets, such as the uncertainty associated with individual rain gauges and soil samples. This is far outside the scope of the catchment descriptor update project.

New FEH *QMED* catchment descriptor equations and pooling-group generation methods will be released in conjunction with Wallingford HydroSolutions' release of WINFAP 6, and reports will be available on UKCEH's FEH webpages.

Acknowledgements

We would like to acknowledge the help of Dan Hollis from the Met Office who helps us with many of our enquiries, as well as our colleagues at the Environment Agency who directed our queries.

References

Bayliss, A. C. (1999) *Catchment descriptors. Volume 5 of the Flood Estimation Handbook*. UK Centre for Ecology & Hydrology, Wallingford.

Bayliss, A. C., Black, K. B., Fava-Verde, A. & Kjeldsen, T. R. (2006). *$URBEXT_{2000}$ – A new FEH catchment descriptor. Calculation, dissemination and application*. Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme. Department for Environment, Food and Rural Affairs, London.

Curley, M., Coonan, B., Ruth, C. E. & Ryan, C. (2023). *Ireland's Climate Averages 1991-2020. Climatological Note No. 22*. Met Éireann, Ireland.

Faulkner, D. (1999) *Rainfall frequency estimation. Volume 2 of the Flood Estimation Handbook*. UK Centre for Ecology & Hydrology, Wallingford.

Griffin, A., Young, A. & Stewart; L. (2019). Revising the BFIHOST catchment descriptor to improve UK flood frequency estimates. *Hydrology Research* 50(6): 1508-1519.

Kjeldsen, T. R., Jones, D. A. & Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme. Department for Environment, Food and Rural Affairs, London.

Met Office, 1993. Standardised Annual Average Rainfall (1961-90) [SAAR 61-90] [Dataset, last accessed 2024-11-07]. https://catalogue.ceh.ac.uk/documents/2a2b1b05-a30f-4e04-9b37-75fa5ef5c26b Met Office, Exeter.

Met Office, Hollis, D., Carlisle, E., Kendon, M., Packman, S., Doherty, A., 2024. HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.3.0.ceda (1836-2023). NERC EDS Centre for Environmental Data Analysis, 18 July 2024. doi:10.5285/b963ead70580451aa7455782224479d5.Moore, R. V., Morris, D. G. & Flavin, R. W. (1994). *Sub-set of UK digital 1:50,000 scale river centreline network*. Institute of Hydrology, Wallingford.

Morris, D.G. & Flavin, R. W. (1990). A digital terrain model for hydrology. *Proc 4th International Symposium on Spatial Data Handling,* 23-27 July, Zürich, 1, 250-262.

Morris, D.G. & Flavin, R. W. (1994). *Sub-set of UK 50m by 50m hydrological digital terrain model grids*. Institute of Hydrology, Wallingford.

Ordnance Survey (2021). *OS Terrain 50* [dataset]. https://www.ordnancesurvey.co.uk/products/os-terrain-50

Otto, F. E. L., van der Wiel, K., van Oldenborgh, G. J., Philip, S., Kew, S. F., Uhe, P. & Cullen, H. (2018). Climate change increases the probability of heavy rains in Northern England/Southern Scotland like those of storm Desmond—a real-time event attribution revisited. *Environmental Research Letters* 13, 024006.

Perry, M., Hollis, D. & Elms, M. (2009). *National Climate Information Centre Climate Memorandum No 24: The Generation of Daily Gridded Datasets of Temperature and Rainfall for the UK*. Met Office, Exeter.

Rowland, C.S.; Morton, R.D.; Carrasco, L.; McShane, G.; O'Neil, A.W.; Wood, C.M. (2017a) *Land Cover Map 2015 (25m raster, GB)* [dataset]. NERC Environmental Information Data Centre. https://doi.org/10.5285/bb15e200-9349-403c-bda9-b430093807c7

Rowland, C.S.; Morton, R.D.; Carrasco, L.; McShane, G.; O'Neil, A.W.; Wood, C.M. (2017b) *Land Cover Map 2015 (25m raster, N. Ireland)* [dataset]. NERC Environmental Information Data Centre. https://doi.org/10.5285/47f053a0-e34f-4534-a843-76f0a0998a2f

Rowland, C.S.; Morton, R.D.; Carrasco, L.; McShane, G.; O'Neil, A.W.; Wood, C.M. (2017c) *Land Cover Map 2015 (vector, GB)* [dataset]. NERC Environmental Information Data Centre. https://doi.org/10.5285/6c6c9203-7333-4d96-88ab-78925e7a4e73

Rowland, C.S.; Morton, R.D.; Carrasco, L.; McShane, G.; O'Neil, A.W.; Wood, C.M. (2017d) *Land Cover Map 2015 (vector, N. Ireland)* [dataset]. NERC Environmental Information Data Centre. https://doi.org/10.5285/60764028-adeb-4316-987a-14b3b21a8f9a

Spackman, E. (1993). Calculation and Mapping of Rainfall Averages for 1961-90. *British Hydrological Society Meeting on Areal Rainfall*, 15 December, University of Salford, Manchester.

Appendix 1: Comparison of *SAAR* versions

This appendix summarises the differences between the current FEH $SAAR_{6190}$ rainfall product and those available from the Met Office through the HadUK dataset ($SAAR_{9120}$). These are also compared with an equivalent 1961-90 estimate derived using HadUK.

FEH $SAAR_{6190}$ (1993)

The standardised average annual rainfall for the period 1961-1990, $SAAR_{6190}$, was generated in 1993, and is the current standard period for catchment average rainfall [Met Office, 1993]. It references "Spackman, E. 1993. Calculation and Mapping of Rainfall Averages for 1961-90, University of Salford, Manchester, 15th Dec.", a talk given to a British Hydrological Society meeting on Areal Rainfall. The current $SAAR_{6190}$ metadata (based on notes from the talk referenced) reads as follows:

> *Rainfall Averages for 1961/90 have been calculated from monthly values assembled from the Met Office Climatological Data Archive for stations open at some time during 1961-90. The monthly totals were compared automatically with nearby stations. Seriously discrepant values were inspected manually were either corrected or discarded. Missing and discarded observations were replaced by estimates calculated by regression against observations from up to 6 nearby long period stations. Average Annual Rainfall values were analysed by determining a fitting equation for each 10km square. The natural logarithm of the rainfall was related to 32 variables which included: relative geographical position (using a quadric representation in National Grid easting and northing), the ratio of land to sea within 3.5km, the local terrain height and a further 21heights at distances of 1km, 3.5km and 12km. The fitting method used an iterative regression on the residuals specially devised to avoid over-fitting. Rainfall values were weighted according to the number of years each station was open during 1961-90 and their distance from the square. The number of station values used for each square depended on the data density and ranged from about 20 within 75km in sparse data areas to about 70 within 15km where data were most plentiful. Final rainfall estimates were obtained bi-linearly interpolating the values calculated from the fitting functions for each of the 4 nearest squares. The resulting Standard Error of these values, worked out from those station averages based on more than 15 years of data, ranges from better than 2% in topographically simple areas to worse than 5% in some upland areas. Monthly values have been obtained by generating the percentage of annual average on a grid with a 5km spacing. These values were obtained as a weighted average of nearby station values where the*

weights depended on the distance of the station and the period for which data were available.

Further information (personal comms.) suggested that “Terrain Heights are averages for 500m x 500m area” so are much smoother than those used by the Met Office (25m native resolution, averaged to 100m resolution by the Met Office).

HadUK $SAAR_{9120}$

The HadUK $SAAR_{9120}$ grid has been generated by the Met Office as part of the HadUK dataset [Met Office, 2024] (FEH analysis is presently using v1.3.0.0).

The methods for deriving monthly grids by the Met Office were originally written in Visual Basic, but have been rewritten twice, first in GIS software and most recently in Python. Due to the characteristics and limitations of each language, the methods were not translated identically during either rewrite, although efforts were made to minimize the differences each time.

There are two approaches available in HadUK grids:

- **An Average-then-Grid approach (AtG, internal to Met Office):**
  - Infill missing months in the station record
  - Calculate the 30-year January average, 30-year February average, etc.,
  - interpolate the twelve grids (using regression and donor gauges) and sum. This approach was also used to generate FEH *SAAR*.
- **A Grid-then-Average approach (GtA, HadUK-Grid on CEDA – used to generate the current version of $SAAR_{9120}$ as of October 2024):**
  - Calculate the total for January 1991, February 1991, etc., to December 2020, just using all available data (stations with a maximum of 2 missing days in the month)
  - Convert values to percentages of the long-term average
  - Interpolate the 360 grids using donor gauges and regression on easting and northing.
  - Sum the grids and divide by 30.

For the regression/infilling procedure, the chosen donors, and their impacts, change over time as records lengthen and stations open or close. The gauge time series are extended to a gridded dataset using regression with easting, northing, elevation and terrain shape, and corrected using inverse distance weightings on the gauges.

The elevation and shape grids used to generate *SAAR* in the 1990s were sourced from Military Survey, but the Met Office now (and for some time) uses Copernicus gridded topography.

Differences between methods

Compared to the 1993 approach ($SAAR_{6190}$), which possibly used a separate model for each 10km square (using stepwise regression), the current Met Office approach uses a more flexible, but single model across the UK. The 1993 method also looks to model “the natural logarithm of rainfall”, whereas the HadUK-Grid methods don’t transform the rainfall at all. The use of natural logarithm in calculations in the older method could result in greater rainfall depths in the wettest parts of the UK, once they are transformed back into millimetre values.

Both the 1993 and HadUK-Grid methods use a similar approach to account for missing data, using regression and six donor gauges, but the older approach seems to use a vastly bigger number of covariates compared to the current approach.

There is some ambiguity in the method of interpolation to a 1km grid: “Final rainfall estimates were obtained bi-linearly interpolating the values calculated from the fitting functions for each of the 4 nearest squares”. As read, this suggest a level of smoothing not seen in the data, so some subtleties have been lost in the summary of the method.

Differences

FEH $SAAR_{6190}$ vs HadUK $SAAR_{6190}$ (AtG)

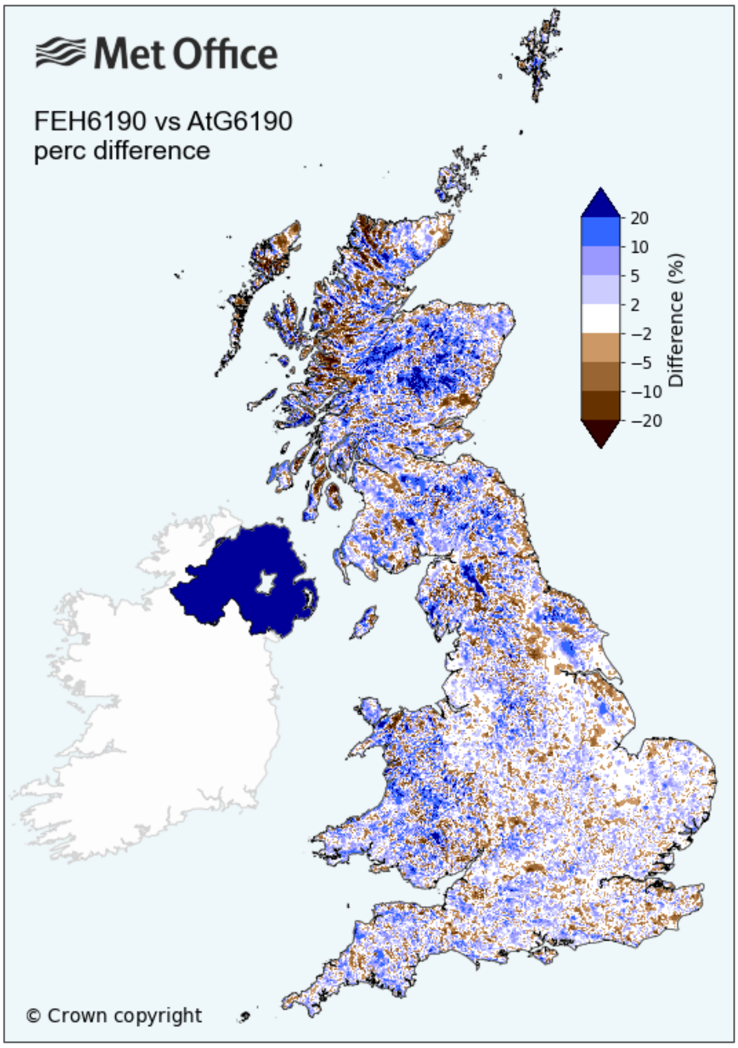

Figure A18 Percentage difference between FEH $SAAR_{6190}$ (shifted) and $SAAR_{6190}$ from the HadUK grid (For GB only)

Figure A18 shows the difference between the original FEH $SAAR_{6190}$ and the HadUK current estimates for the same period using an equivalent method of calculating averages at gauges, then gridding the results (AtG: Average-then-Grid). There is a lot of granularity in the differences across the whole of Great Britain. Some of these big changes are due to the FEH $SAAR_{6190}$ data being much smoother spatially than the HadUK data, so the extremes in the HadUK grid are visible in the comparison.
There is also the issue that the FEH and HadUK grids are offset by 500m, which may show large changes in steep or mountainous areas. However, these large changes may actually result from the alignment of the two datasets (in order to

divide one grid by the other, one must be shifted by 500m). This would not present themselves when calculating catchment averages. Differences in mountainous areas are confounded by differences in smoothing of the topography: 500m in 1993, 100m in HadUK-Grid.
Finally, there is the difference in method, where the 1993 approach for FEH6190 used station weighting based on record length to account for varying record lengths, whereas the HadUK grid uses monthly infilling methods.

HadUK $SAAR_{6190}$ vs HadUK $SAAR_{9120}$ (GtA)

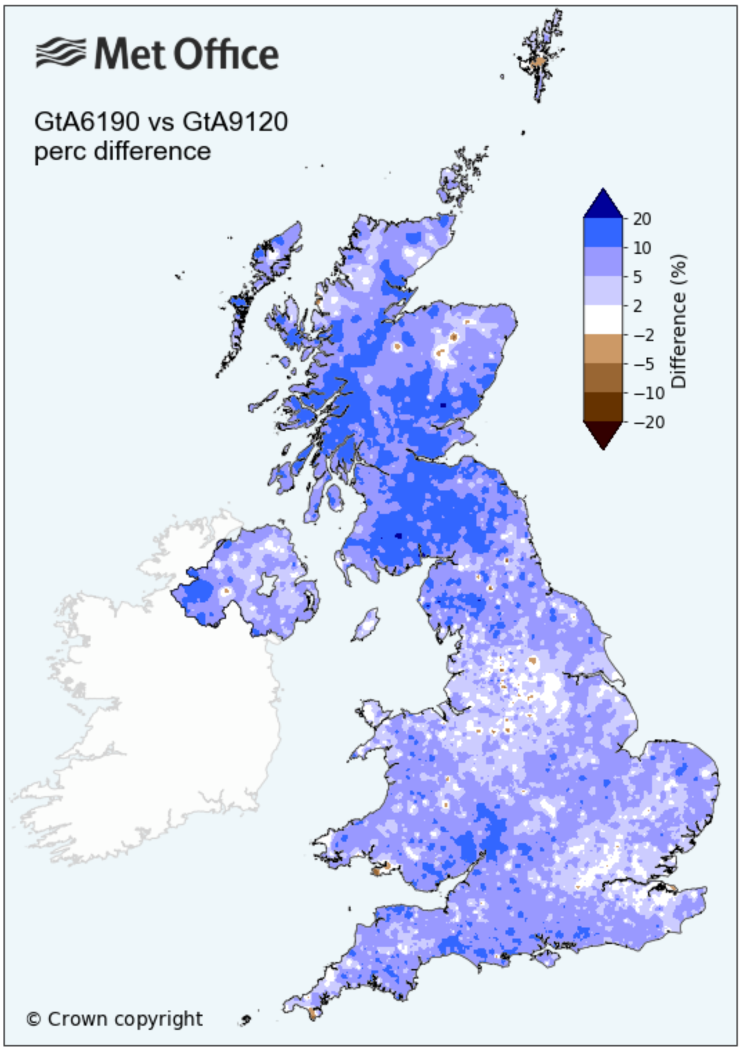

Figure A19 Percentage difference between $SAAR_{6190}$ and $SAAR_{9120}$ using the HadUK grid (using a Grid-then-Average (GtA) method)

Figure A19 shows the differences between the HadUK-Grid 1961-90 and 1991-2020 averages. The “bullseyes” seen in the grid relate to differences in the gauging network over the 60-year period as stations open and close. This is a major issue, since the number of stations being used has almost halved over the 60-year period.

Figure 5 shows the distribution of stations across the UK in January 1961, 1991, and 2021, along with the annual changes in the number of stations used in annual calculations (hence the number of stations in January 1961 is less than the graphed value, which is for the whole of 1961). Alongside the reduction in the number of gauges, there is also a slightly more uniform distribution of gauges, as more of the urban gauges in, for example, Manchester and London, have closed compared to rural areas.

The overall trend suggests the recent 30-year period has been wetter than the 1961-1990 period, which matches up with our knowledge of inter-decadal variability and climate change.

The order of operations undertaken to calculate monthly (and annual) rainfall (AtG vs GtA, as described above) can affect the final gridded rainfall depths. The comparison of the order of operation shows that grid-then-average is always drier than or equal to average-then-grid. Differences can be zero, but can exceed 5% over small upland areas, and 2% over larger upland areas. Hence, this is a small issue compared to network changes and other methodological differences, and is comparable to the reported standard error in the 1993 rainfall depths.

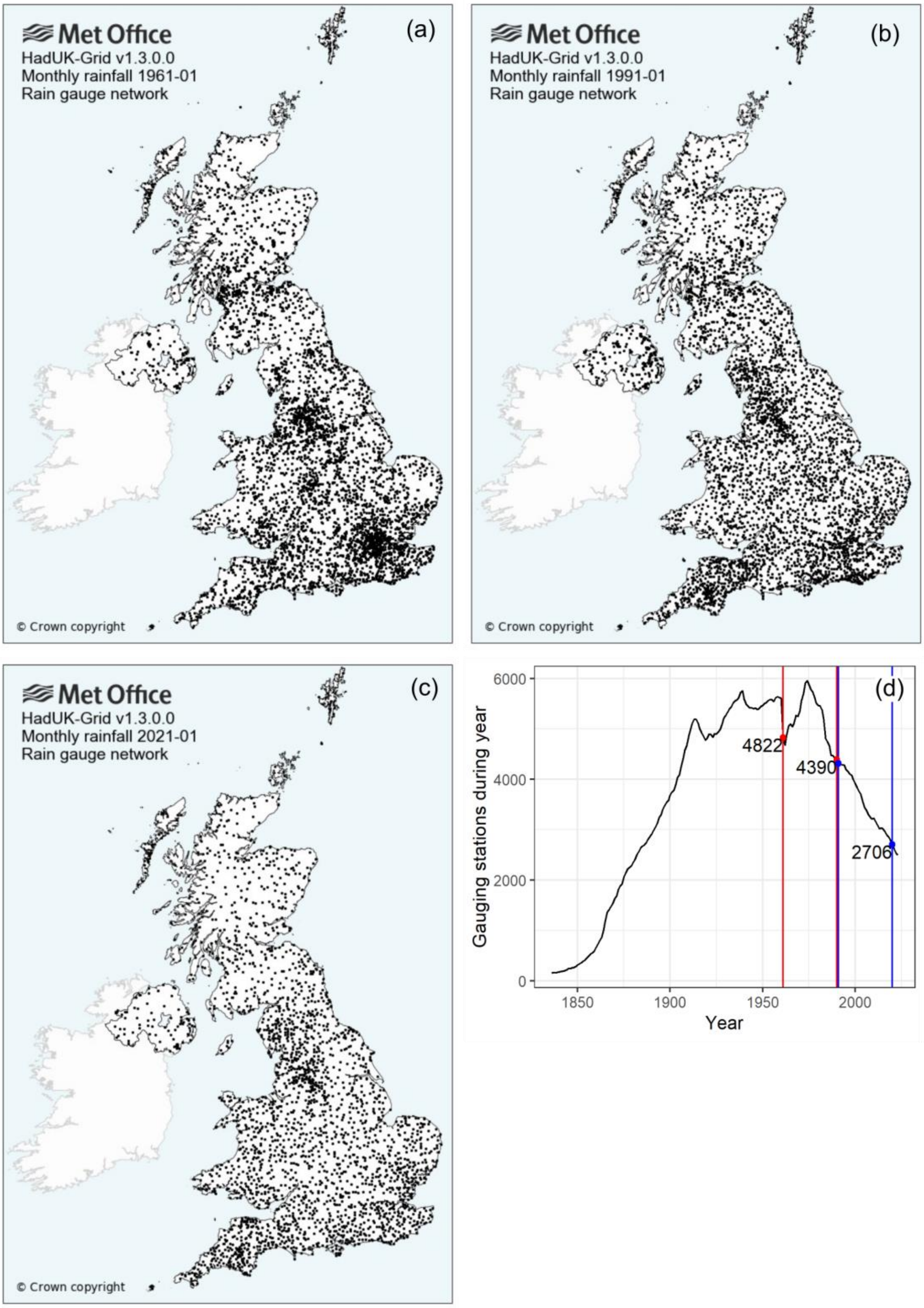

Figure A20 Coverage of stations used in monthly rainfall calculations for HadUK in (a) January 1961, (b) January 1991, (c) January 2021. Annual station counts given in (d), with standard periods 1961-90 and 1991-2020 highlighted in red and blue respectively.

Cumulative Differences

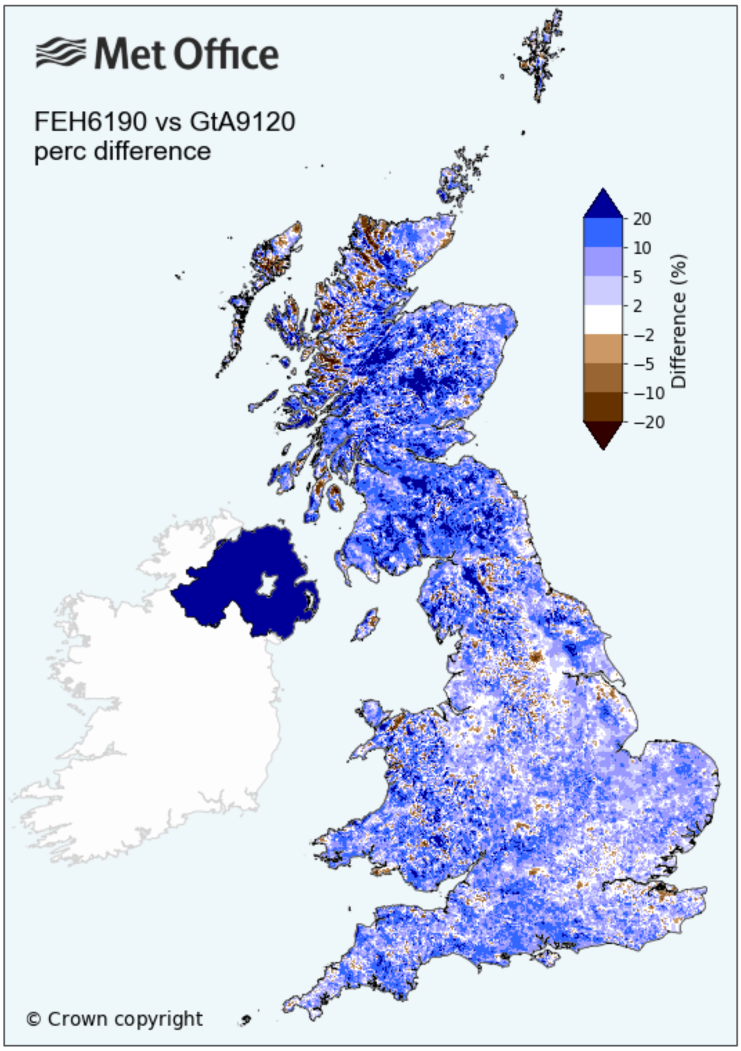

Figure A21 Percentage difference between FEH *SAAR*6190 (shifted) and HadUK *SAAR*9120 using the HadUK grid (using a Grid-then-Average (GtA) method) for GB only.

Figure A21 shows the combined differences between the FEH *SAAR*6190 grid and HadUK *SAAR*9120 grid (v1.3.0.0 as available on CEDA). This combines all the issues discussed above into a single plot, compounding some issues and reducing the apparent impact of others.

Conclusions

The key differences between the two datasets:

- FEH $SAAR_{6190}$ used (in 1993 when developed) more extensive statistical regression to smooth the dataset, beyond what the Met Office considers necessary for monitoring the UK climate.
- FEH $SAAR_{9120}$ is offset by 500m from HadUK-Grid, so some differences observed in the figures are due to misalignment of the datasets.
- There is a large change in the number and location of raingauges over the 60 years of interest, with a shift to fewer, more uniformly distributed raingauges across the UK. For example, 4822 stations were used in 1961 in annual calculations, and 2706 stations used in 2021 in annual calculations.
- The Met Office's methods have been transferred through different programming languages and platforms, which may introduce very small differences, although efforts were made at the time to ensure these were minimized.
- There is also the actual difference in rainfall due to inter-decadal variability and climate change, which is the difference the product aims to capture.

There is no method of "correcting" any current rainfall product from HadUK to match the methods/network without substantial work and resources. Therefore, it is still recommended that HadUK's $SAAR_{9120}$ is the most appropriate rainfall estimator to describe the current condition of UK catchments.

Contact

enquiries@ceh.ac.uk

@UK_CEH

ceh.ac.uk

Bangor

UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road
Bangor
Gwynedd
LL57 2UW

+44 (0)1248 374500

Edinburgh

UK Centre for Ecology & Hydrology
Bush Estate
Penicuik
Midlothian
EH26 0QB

+44 (0)131 4454343

Lancaster

UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue
Bailrigg
Lancaster
LA1 4AP

+44 (0)1524 595800

Wallingford (Headquarters)

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

+44 (0)1491 838800

UK Centre for Ecology & Hydrology