Received: 26 September 2019 | Revised: 6 April 2020 | Accepted: 30 May 2020

DOI: 10.1111/jfr3.12642

ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management | Journal of Flood Risk Management | WILEY

# Estimating design flood runoff volume

**Gianni Vesuviano**[1] | **Elizabeth Stewart**[1] | **Andrew R. Young**[2]

[1]Centre for Ecology and Hydrology, Wallingford, UK

[2]Wallingford HydroSolutions Ltd., Wallingford, UK

**Correspondence**
Gianni Vesuviano, Centre for Ecology and Hydrology, UK.
Email: giaves@ceh.ac.uk

**Abstract**

Standard flood risk estimation methods in the United Kingdom have largely focused on peak flows at ungauged locations. However, the importance of whole-hydrograph and event volume estimation in a design context is increasing with the application of unsteady-state hydraulic models and construction of sustainable drainage systems. Here, we explore the relationship between peak flow estimation accuracy and flood volume estimation accuracy across 780 events in 81 catchments. Runoff hydrographs are modelled using ReFH2, a rainfall-runoff model widely used by practitioners for design flood estimation in the UK. We find that strong performance in peak flow estimation is highly correlated with strong performance in event volume estimation, and that between-event variation in performance is greater than the typical reduction in performance when moving from calibrated to design (regression-based) model parameters. Unfortunately, evaluating model performance in terms of runoff volume is complicated by the fact that measured rainfall hyetographs and runoff hydrographs are themselves estimates that can disagree with each other for legitimate reasons. We demonstrate that it is not always possible, expected or realistic to close the water balance over an event in a topographically defined river catchment. Hence, 'errors' in modelled hydrographs cannot be solely attributed to modelling deficiencies.

**KEYWORDS**

hydrological modelling, hydrology, rainfall-runoff

## 1 | INTRODUCTION

The estimation of flood risk from a hydrological perspective in the United Kingdom has very much focused on the estimation of flood frequency curves, or the relationship between flood peak magnitude and rarity. From the original Flood Studies Report (FSR; NERC, 1975) to the Flood Estimation Handbook (FEH; Institute of Hydrology, 1999) and subsequent updates (Kjeldsen, 2007; Kjeldsen, Jones, & Bayliss, 2008; Kjeldsen, Miller, & Packman, 2013; Wallingford HydroSolutions, 2016), two methods have been provided for flood frequency estimation. The FEH statistical method follows the general index-flood and regionalization principles described by Dalrymple (1960). However, the index flood in the UK is taken as the median annual flood (*QMED*) and the 'region' used to provide additional flood frequency behavioural information is non-geographical and specific to each catchment of interest, consisting only of the gauged catchments judged, through the use of digital catchment descriptors (Bayliss, 1999), to be most hydrologically similar to the catchment of interest, and weighted according to similarity. In contrast, the rainfall-runoff method is a generalised, event-based, lumped rainfall-runoff model of a catchment, used in conjunction with a rainfall depth–duration–frequency (DDF) model and design estimates of initial conditions to estimate a *T*-year flow hydrograph and hence the peak flow corresponding to a specified *T*-year rainfall event. Rainfall-runoff modelling methods and design rainfall DDF models have both been the subject of continuous development. ReFH2 (Kjeldsen et al., 2013; Wallingford HydroSolutions, 2016) is the current version of the Revitalised Flood Hydrograph (ReFH) model (Kjeldsen, 2007) as of 2016, and is widely used for design flood estimation in the UK, largely superseding the FSR/FEH rainfall-runoff model (Houghton-Carr, 1999). The FEH13 DDF rainfall model (Stewart, Vesuviano, Morris, & Prosdocimi, 2014) is currently recommended to provide design rainfall inputs to ReFH2.

The majority of practitioner guidance on the choice of method for a particular problem tends to focus on peak flow estimation (e.g., CNC, 2017; Environment Agency, 2017; SEPA, 2017). Consequently, calibration of ReFH2 and previous UK rainfall-runoff methods (FSR, FEH, and ReFH) has focused on how well they reproduce flood-frequency relationships at gauged sites, using the gauged data alone for shorter return periods, or in a regional analysis with heavy at-site weighting for longer return periods. Neither ReFH2 nor any previous UK rainfall-runoff method has been extensively evaluated in terms of simulated event hydrograph, and therefore simulated runoff volume, the most comprehensive study as of 2019 being an evaluation of the ReFH2 design hydrograph shape against the empirical median hydrograph (Archer, Foster, Faulkner, & Mawdsley, 2000) for 20 small catchments up to 40 $km^2$ (Environment Agency, 2012). Evaluating performance in terms of runoff volume is difficult, as it is often not possible to calculate a closed water balance over an observed event; further rainfall may occur before flows have receded to pre-event levels, and it may not be clear that flows at the start of an event would have followed a pattern of recession in the absence of rainfall. However, with the application of unsteady-state hydraulic models becoming the norm in, for example, reservoir modelling and catchment-scale flood risk modelling, correct estimation of flood hydrograph, and event volume in a design context is becoming increasingly important to flood risk management in the UK. Furthermore, accurate estimation of runoff volume is required to size the storage and/or detention components of any system (or area) for which the outflow rate is less than the peak runoff rate, even if the design criteria focus only on achieving a specified outflow rate. Examples of systems in which temporary storage is normally required to meet design outflow rates include sustainable drainage systems (SuDS), reservoirs with spillways, and flood storage areas.

In this paper, we explore the relationship between the accuracy of peak flow estimation and the accuracy of volume estimation across 81 catchments and 780 observed flood events using the ReFH2 model framework with event-specific initial conditions and, initially, catchment-specific calibrated model parameters. We evaluate how the predictive performance of the modelling framework reduces for both peak flow and volume estimation when using generalised, regression-estimated model parameters, and compare the change in mean model performance with the between-event variation in model performance. Finally, we identify several factors that may lead to poor model performance.

## 2 | DATA

This study was conducted on a dataset of 780 rainfall-runoff events distributed over 81 catchments in England and Wales, all obtained from the Flood Event Database (Bayliss, 2003). This is a subset of the 1,285 events and 101 catchments used in the original development of ReFH that excludes heavily urbanised catchments, catchments located in Scotland (which use Scotland-specific regression equations for parameter estimation) and events from sources other than the Flood Event Database. The 81 study catchments are mapped in Figure 1, where catchment colour represents the number of available events. For each event, catchment-average rainfall hyetographs, gauged runoff hydrographs, and estimates of initial baseflow, $BF_0$, and initial soil moisture, $C_{ini}$, derived from gauged flow records and 1 year of antecedent rainfall and potential evaporation records respectively, were available from the original development of ReFH.

Catchment-average rainfall hyetographs were derived according to Jones (1983), which assumes and requires that any catchment can be approximated as an irregular quadrilateral. Figure 2 compares a selection of catchment boundaries derived from a 50-m DTM, the Integrated Hydrological Digital Terrain Model (IHDTM: CEH, 2014), against their quadrilateral approximations and demonstrates that this assumption is generally appropriate, even for very irregularly shaped catchments.

Two sets of model parameter values were available for each catchment: the event-calibrated values, which vary between catchments, but not events, and minimise error against gauged hydrographs across all events on a

CIWEM Chartered Institution of Water and Environmental Management Journal of Flood Risk Management—WILEY



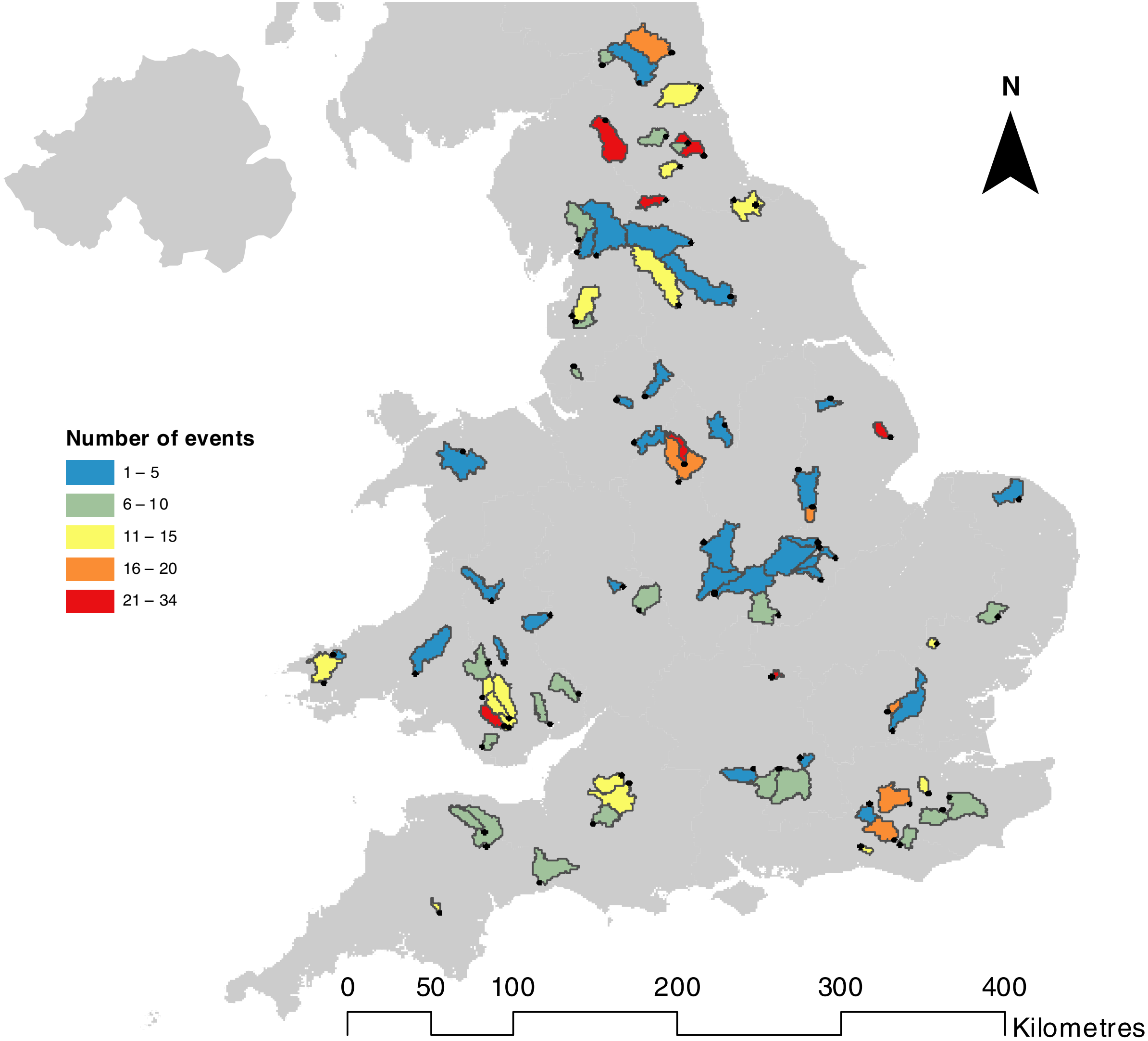


**FIGURE 1** Locations of study catchments and number of events per catchment

single catchment; and the ReFH2 design package values, which are estimated from regression relationships against FEH catchment descriptors. FEH catchment descriptors and event-calibrated model parameter values were available for each catchment, originating from the development of the original version of the ReFH model (ReFH1). Table 1 gives the five-number summary for selected catchment descriptors in this dataset of 81 catchments:

- *AREA*: catchment area (km$^2$)
- *SAAR*:1961–1990 mean annual rainfall (mm)
- *BFIHOST*: estimated baseflow index from soil type (dimensionless)
- *FARL*: flood attenuation from reservoirs and lakes (dimensionless)
- $URBEXT_{2000}$: proportion of catchment urbanised in the year 2000 (dimensionless)
- *DPLBAR*: mean drainage path length (km)
- *DPSBAR*: mean drainage path slope (m/km)
- *PROPWET*: proportion of time that catchment was 'wet' (estimated soil moisture deficit <6 mm) during 1961–1990 (dimensionless).

The presented catchment descriptors are used either to estimate model parameters or initial conditions during ungauged or uncalibrated model runs (*SAAR*, *BFIHOST*, *PROPWET*, *DPLBAR*, and *DPSBAR*), to assess the applicability of using ReFH2 for a particular catchment (*AREA* and *FARL*), or to assess whether 'as rural' modelling can be appropriate ($URBEXT_{2000}$). Wallingford HydroSolutions (2018) shows that the *BFIHOST* values assigned to HOST classes 23 and 25 by Boorman, Hollis, and Lilly (1995) are lower than the baseflow index values derived from gauged flow records in catchments dominated by these soil types, and provides a procedure to adjust the *BFIHOST* estimate to a (typically) more

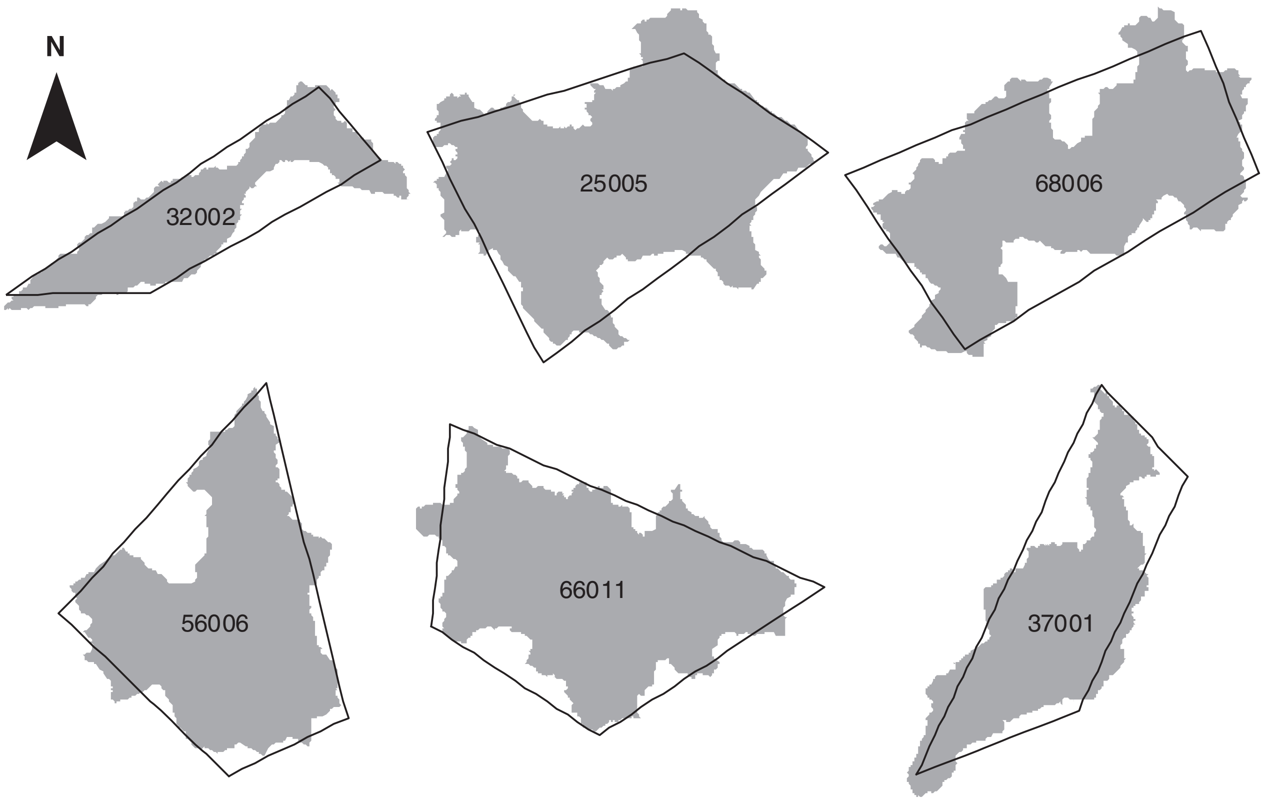


**FIGURE 2** Comparison of IHDTM representations (solid grey) and quadrilateral approximations (black lines) of selected irregular catchments. Each catchment's NRFA station number (nrfa.ceh.ac.uk/data/search) is included inside it

**TABLE 1** Five-number summary of selected catchment descriptors

| Descriptor | Minimum | Lower quartile | Median | Upper quartile | Maximum |
|---|---|---|---|---|---|
| *AREA* | 15.07 | 65.35 | 142.08 | 273.84 | 510.90 |
| *SAAR* * | 577 | 691 | 887 | 1,276 | 2,182 |
| *BFIHOST* * | 0.242 | 0.382 | 0.466 | 0.552 | 0.782 |
| *FARL* | 0.886 | 0.974 | 0.986 | 0.997 | 1.000 |
| $URBEXT_{2000}$ | 0.0000 | 0.0037 | 0.0131 | 0.0437 | 0.2974 |
| *DPLBAR* * | 4.55 | 10.25 | 15.26 | 21.03 | 38.49 |
| *DPSBAR* * | 11.5 | 34.3 | 75.2 | 123.5 | 210.4 |
| *PROPWET* * | 0.25 | 0.32 | 0.39 | 0.53 | 0.71 |

*Note:* Starred catchment descriptors are used in parameter and/or initial condition estimation. *AREA* and *FARL* are used to assess the applicability of ReFH2. $URBEXT_{2000}$ is used to assess the applicability of 'as rural' modelling.

realistic value in any catchment with any level of HOST class 23 or 25 coverage. Catchment-average values of *BFIHOST* were therefore revised for all catchments, the mean increase in *BFIHOST* resulting from this procedure being 0.015 and the largest being 0.098, from 0.238 to 0.336 for Ray at Grendon Underwood (NRFA N$^{o}$. 39017). The *BFIHOST* adjustment procedure in Wallingford HydroSolutions (2018) differs from that described in Griffin et al. (2019b), which had not been finalised when the work described in this paper was performed. It is noted that the catchments span a range of areas, although the maximum catchment area is limited to around 500 km$^2$ due to the increasing implausibility of spatially-uniform rainfall over increasingly larger areas (Kjeldsen, 2007). Similarly, the range of $URBEXT_{2000}$ encompasses the vast majority of gauged UK catchments, although an upper limit of 0.3 was imposed so that all catchments could be modelled 'as rural'—this decision is explained further in 'Methods'. Due to the lack of storage routing in the ReFH2 model structure, no catchments with significant reservoirs or lakes were included. All other catchment descriptors encompass a wide range relative to England and Wales as a whole, except *DPLBAR*, which is strongly correlated with *AREA*. However, values of *DPLBAR* are typical for the range of catchment areas in this dataset.

Further information on UK catchment descriptors is available from Bayliss (1999) and Bayliss, Black, Fava-Verde, and Kjeldsen (2006).

## 3 | METHODS

Rainfall-runoff modelling was conducted using ReFH2 (Kjeldsen et al., 2013; Wallingford HydroSolutions, 2016). ReFH2 is a lumped conceptual model that divides a catchment into three zones: rural, urban pervious and urban impervious (Figure 3, left). As it is a lumped model, spatial uniformity of rainfall is assumed and the spatial arrangement of zones is ignored.

For the rural fraction and urban pervious fraction of the catchment, the PDM loss model (Moore, 2007) with a uniform distribution of soil stores is used to estimate net rainfall, while for the urban impervious fraction of the catchment, a constant percentage runoff is assumed. A 'rural' unit hydrograph (e.g., Chow, Maidment, & Mays, 1988) is used to route only the rural net rainfall, while the urban pervious and urban impervious net rainfalls are added together and routed via a faster 'urban' unit hydrograph. Baseflow is generated by passing recharge through a linear reservoir. Recharge is generated only for the runoff-generating fractions of the rural and urban permeable fractions of the catchment, and the ratio of recharge to runoff is fixed. This approach, described by Appleby (1974) is used as it allows baseflow separation when the rainfall input is unknown. Total runoff is the sum of baseflow, routed rural runoff and routed urban runoff. As a lumped model, information on the spatial arrangement of rural, urban pervious and urban impervious zones in a catchment is not used.

In typical usage, the catchment descriptor $URBEXT_{2000}$ is multiplied by 1.567 to estimate the proportion of total urban area in a catchment (Bayliss et al., 2006). Urban area is by default assumed to be 70% pervious and 30% impervious, a constant percentage

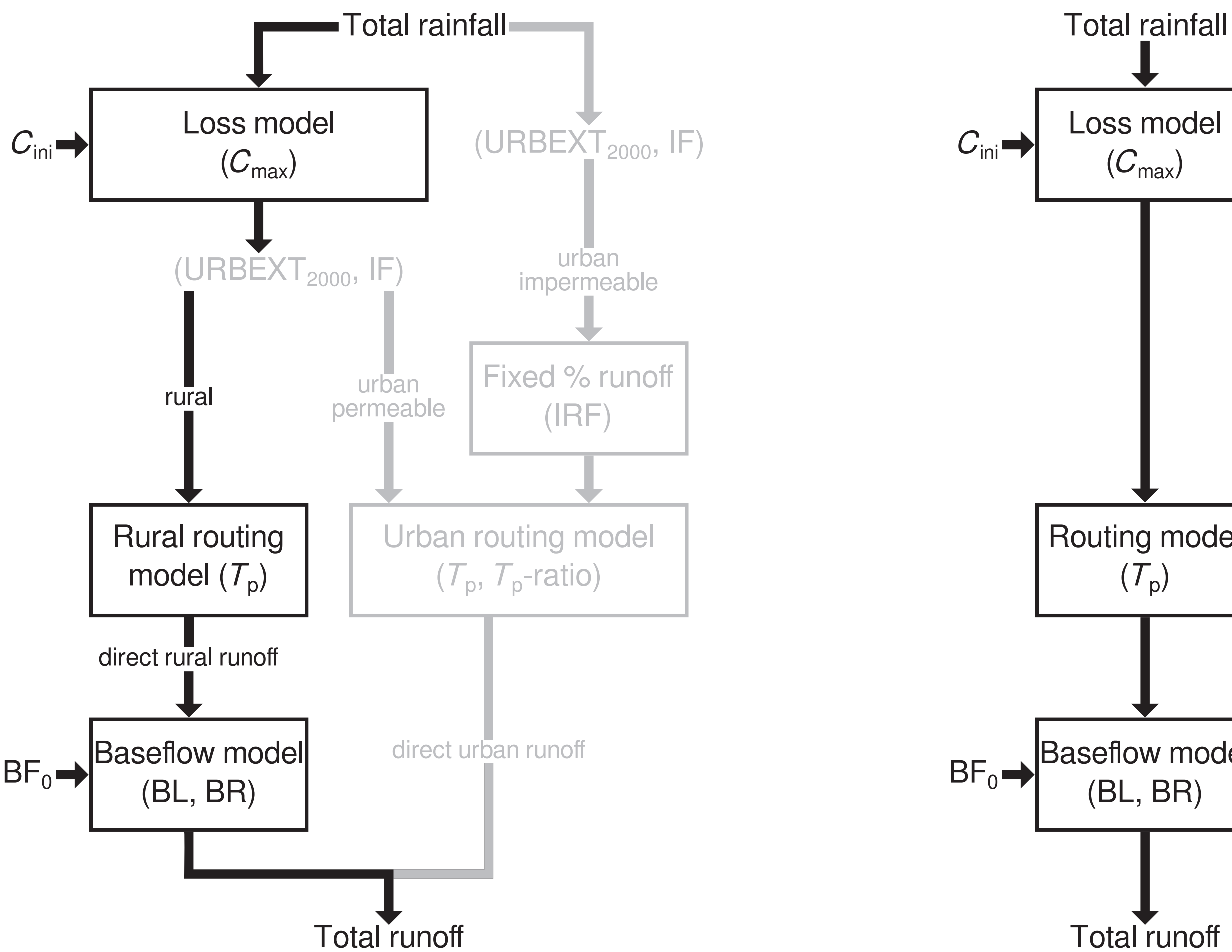


**FIGURE 3** ReFH2 model structure with urban flow paths greyed-out (left); ReFH model structure (right)

runoff of 70% is assumed for the impervious urban area, and the urban unit hydrograph time-to-peak is assumed to be 50% of the rural unit hydrograph time-to-peak (Wallingford HydroSolutions, 2018). However, recent research has shown that the accelerated routing and percentage runoff that should be associated with urban catchments is not consistently apparent until a level of urbanisation represented by an $URBEXT_{2000}$ value of 0.3 is exceeded (Environment Agency, 2012). Hence, it is possible to treat catchments with $URBEXT_{2000}$ <0.3 as essentially rural, with all net rainfall generated by the PDM loss model and routed through one unit hydrograph. This 'as rural' configuration gives ReFH2 the same model structure as the original ReFH model (Kjeldsen, 2007), shown in Figure 3 (right).

ReFH2 in its 'as rural' configuration has four model parameters: baseflow linear reservoir time constant (*BL*, hr), recharge-to-runoff ratio (*BR*, dimensionless), unit hydrograph time-to-peak (*Tp*, hr), and PDM maximum soil moisture store depth ($C_{max}$, mm), and two initial conditions: initial soil moisture ($C_{ini}$, mm) and initial baseflow ($BF_0$, $m^3/s$). In gauged catchments, all four parameters can be calibrated using paired rainfall hyetographs and runoff hydrographs and $BF_0$ can be taken as the flow at the start of the event. $C_{ini}$ can be estimated from antecedent daily mean rainfall and potential evaporation series using DAYMOD, a simple accounting procedure where soil moisture is modelled as a balance between infiltration, soil drainage and potential evaporation. DAYMOD is described fully in Appendix B of Kjeldsen (2007). In ungauged catchments, all parameter and design initial condition values are estimated via regression relationships based on FEH catchment descriptors (Bayliss, 1999). The parameter regressions used in ReFH2 maximise the explained variance in calibrated parameter values at gauged catchments, while the initial condition regression equations are set to minimise the difference between the natural logarithm of the median annual flood from gauged data, ln(*QMED*), and the natural logarithm of the peak flow generated by ReFH2 when applying the 1-in-2 year FEH13 design-duration storm with design initial conditions.

In this study, ReFH2 is used first with full calibration (calibrated parameter values, calibrated $C_{ini}$, observed $BF_0$, as described above for gauged catchments) and again with initial condition calibration (regression-estimated parameter values, calibrated $C_{ini}$, observed $BF_0$). A third case, where all parameter values and all initial conditions are estimated via regression, was not considered, as the regressions for initial conditions are intended to relate the spatially uniform *T*-year design rainfall to the *T*-year flood peak. Design initial conditions are therefore inappropriate for use with observed rainfall events, which do not follow single-peaked design rainfall profiles and are not spatially uniform, hence having variable return periods across the catchment. Furthermore, design initial conditions are intended for use with rarer events than many of the 780 considered here.

Performance is measured in terms of geometric mean bias (*bias*, Equation 1) and factorial standard error (*fse*, Equation 2):

$$bias = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\left(\ln y_{i,\text{modelled}} - \ln y_{i,\text{observed}}\right)\right) \tag{1}$$

$$fse = \exp\left(\sigma\left(\ln y_{\text{modelled}} - \ln y_{\text{observed}}\right)\right) \tag{2}$$

where $n$ is the number of events, $\sigma$ indicates standard deviation, and ln and exp indicate natural logarithm and exponential, respectively. In both equations, $y$ can represent peak flow ($m^3/s$) or total runoff volume ($m^3$), and $n$ can be 1, considering an individual event, 780, considering all events in all catchments together, or another number, for example, when considering all events in one catchment. By default, ReFH2-modelled hydrographs continue until the flow rate has regressed to $BF_0$. In cases where the observed and modelled hydrograph differed in length, the longer hydrograph was trimmed to the length of the shorter hydrograph before comparing runoff volumes.

## 4 | RESULTS AND DISCUSSION

Figure 4 summarises the results of the modelling exercise in terms of estimating peak flow, runoff volume and general hydrograph shape for all events and catchments. The top row of Figure 4a–c shows the results when all four model parameters and both initial conditions are calibrated to or set from gauged records, while the bottom row (Figure 4d–f) shows the results when all four model parameters are estimated via regression equations on catchment descriptors but $C_{ini}$ remains calibrated to gauged antecedent rainfall and potential evaporation, and $BF_0$ remains set from gauged flood event data. The left subplot in each row plots *bias* in peak flow against *bias* in runoff volume for each event, the middle subplot shows *bias* in peak flow and runoff volume as histograms, and the right subplot shows a histogram of time-series correlation between modelled and gauged hydrographs. Note that all axes showing *bias* are logarithmic.

Figure 4 shows that error in peak flow estimation is correlated with error in runoff volume estimation and that both the geometric mean bias and spread of errors associated with runoff volume are smaller than those

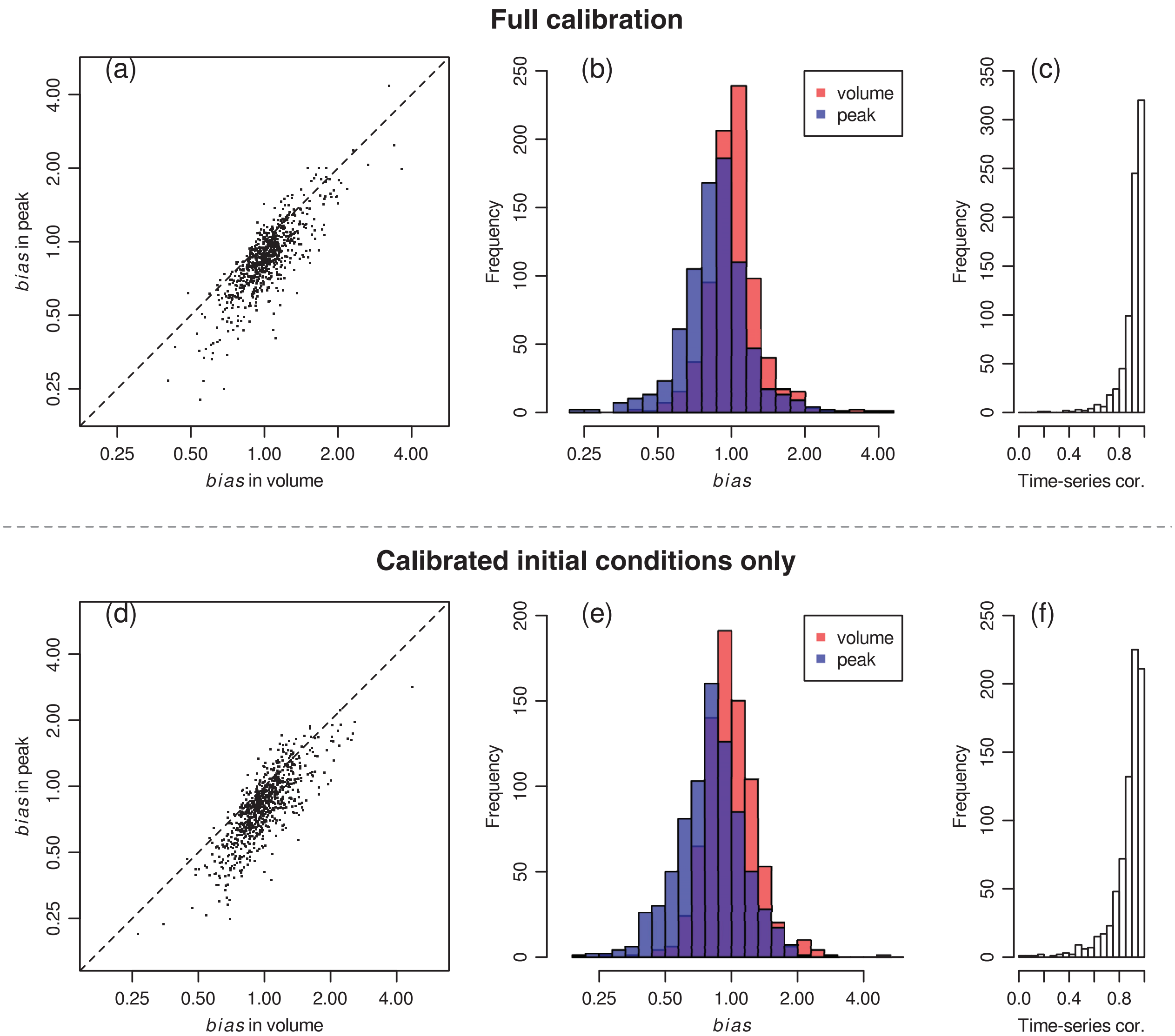


**FIGURE 4** Scatterplot and histogram of *bias* in peak flow and runoff volume, and histogram of time-series correlation between observed and modelled hydrographs for ReFH2 model with full calibration (a–c) and calibration of initial conditions only (d–f)

**TABLE 2** *bias* and *fse* in runoff volume and peak flow across all events for ReFH2 with full calibration and calibrated initial conditions only

| | Full calibration | | Calibrated initial conditions only | |
|---|---|---|---|---|
| | Volume | Peak flow | Volume | Peak flow |
| *bias* | 1.021 | 0.858 | 0.989 | 0.802 |
| *fse* | 1.268 | 1.357 | 1.310 | 1.419 |

associated with peak flow. However, the prevalence of points below/right of the 1:1 line in both scatterplots reveals that the numeric value of the *bias* in runoff volume is generally larger than that in peak flow (for *bias* <1, a larger numeric value indicates a smaller error). Table 2 summarises *bias* and *fse* in runoff volume and peak flow across all events for both levels of calibration. This shows that ReFH2 has a lower *bias* (closer to 1) and smaller *fse* in its estimation of runoff volume than in its estimation of peak flow, regardless of whether the model parameter values are calibrated from data or estimated from regression relationships. It also shows that the decrease in model performance found when moving from calibrated to regression-estimated parameter values is small and that the between-event variation in performance, measured by *fse*, is greater than the typical reduction in performance resulting from a change from calibrated to design parameter values. This demonstrates that ReFH2 is relatively insensitive to variations in the exact parameter values used. All of this is also summarised graphically by comparing Figure 4b,e.

For each event, Figure 5 plots the *bias* in peak runoff or event volume when using calibrated parameter values (on the *y*-axis) with the *bias* in peak runoff or event volume when using design parameter values (on the *x*-axis), using point colour to represent the *bias* in the design parameter value (here simply the design parameter value divided by the calibrated parameter value). *bias* in design parameter value is constant per catchment.

All subplots of Figure 5 show a strong grouping along the 1:1 line, with some scatter. That is, the scale of *bias* in modelled peak flow or runoff volume is generally independent of whether ReFH2 is used with calibrated or design parameter values. In general, *bias* <1 in design parameter values results in greater peak flows and runoff volumes, shown by the prevalence of red points considerably below/right of the 1:1 line. Lower values of $T_p$ and $C_{max}$ conceptually result in a higher-peaked unit hydrograph and smaller soil moisture store, which increase peak flow and runoff volume, respectively. Lower values of *BL* and *BR* correspond to faster-responding baseflow and a lower ratio of recharge to runoff. While the direct effects of *BL* and *BR* on peak flow and runoff volume are less obvious, it is noted that *BL*, *BR* and $C_{max}$ have similar *BFIHOST* coefficients in their respective regressions, while the same is also true of *PROPWET* for *BL* and $C_{max}$. Hence, catchments with *bias* <1 in design *BL* and *BR* may have higher peak flows and runoff volumes simply because they are also often catchments with *bias* <1 in design $C_{max}$.

Conversely, the prevalence of blue points considerably above/left of the 1:1 line shows that *bias* >1 in design parameter values results in smaller peak flows and runoff volumes. However, the scatter of points around the 1:1 line is relatively small compared to the spread along the 1:1 line. This low scatter indicates that errors in peak flow and runoff volume resulting from either calibrated parameter values or parameter values estimated via regression equations are similar for most events. Therefore, the performance of ReFH2 is less sensitive to model parameterization than it is to ReFH2's suitability for modelling a particular event, whether related to the conceptual model structure, the accuracy of gauged flow or rainfall data, or other factors that might influence the difference between modelled and measured flow estimates.

Figure 5 directly illustrates the effects of parameter interaction and equifinality, by showing that design parameter values can give low-*bias* peak flow and runoff volume estimates when they are similar to, far below or far above their corresponding calibrated values. This is shown by noting that on all subplots of Figure 5, points of all colours, representing a wide range of *bias* in the design parameter values, lie near $x = 1$, indicating low *bias* in peak flow or runoff volume when modelling with design parameter values. Although the ReFH2 ‘as rural’

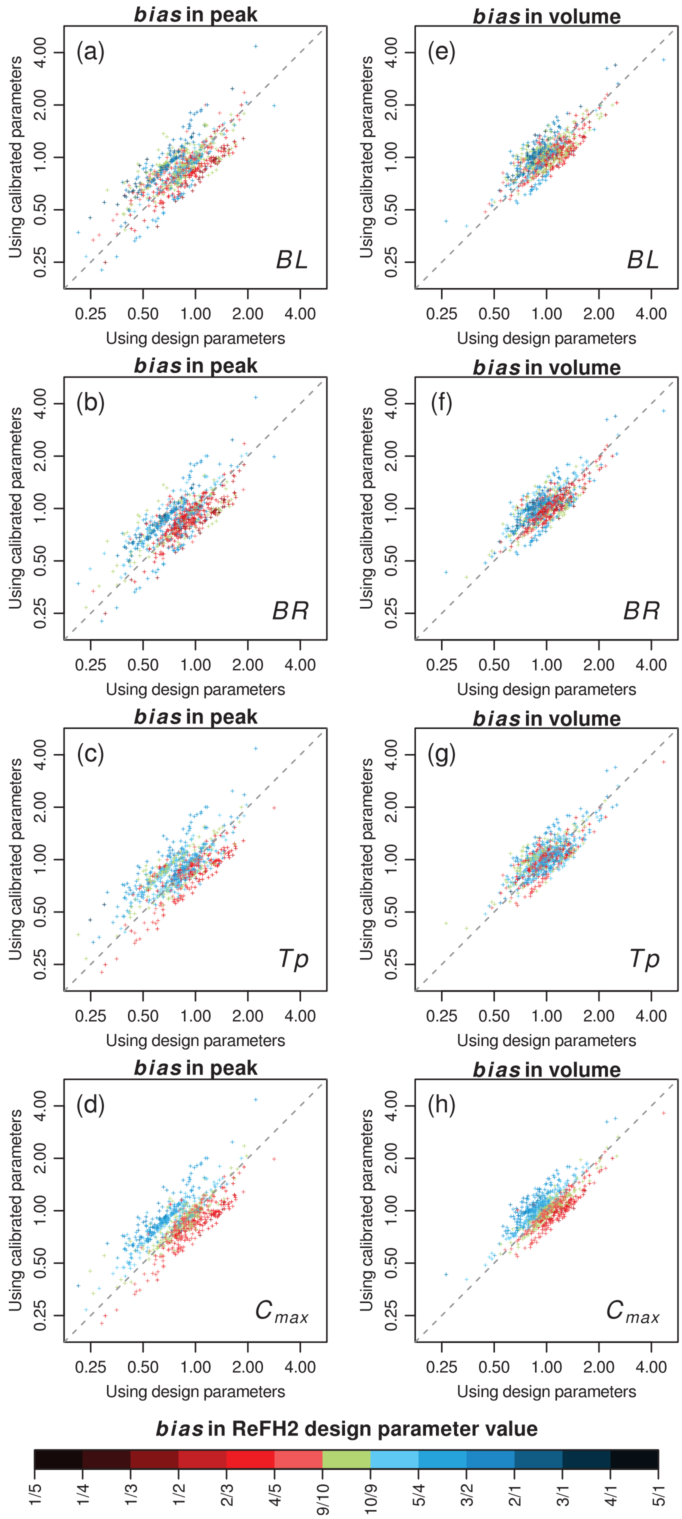


**FIGURE 5** Relationship between *bias* in peak flow or runoff volume and *bias* in ReFH2 parameter value estimate

model has only four parameters and two initial conditions, and is therefore parsimonious compared to many other models, there are multiple ways to achieve two outcomes (a defined peak flow and runoff volume) in any system with more than two independent free variables. Low-*bias* peak and volume estimates with both design and calibrated parameter values are indicated by points near the (1, 1) intersection of both axes. These are more common when the design and calibrated parameter values are similar, as the closeness of the parameterization must result in similar hydrographs. However, parameter interaction can cause hydrographs with similar overall properties to result from a range of parameterizations (equifinality). This is clear from the existence of both red and blue points near to the (1, 1) intersection of both axes in all subplots of Figure 5.

Inspection of individual events where estimated volume is significantly less than gauged volume demonstrates a general difficulty of model calibration and evaluation. In both cases shown in Figure 6, the recorded volume of runoff is greater than the recorded volume of rainfall. In Figure 6a, the general shape of the hydrograph is captured accurately, but the vertical scaling is different, suggesting that the gauged percentage runoff of 117% could result from overestimation at the flow gauge, underestimation at the rain gauges, inappropriate weighting of rain gauges (due to the irregular quadrilateral approximation of the catchment boundary), uncaptured spatial variations in rainfall, or underestimation of contributing area, through neglecting subsurface flows or anthropogenic modification to the catchment (Miller et al., 2014; Vesuviano & Miller, 2019). Figure 6b shows a gauged hydrograph with two equal-magnitude peaks, while ReFH2 estimates a much smaller second peak than first peak. The catchment-average rainfall paired to this event is also double peaked, but the second peak contains considerably less volume of rainfall than the first, despite achieving a similar maximum rate in mm/hour. This could suggest that: the second rainfall peak was not fully captured by all rain gauges contributing to the catchment-average rainfall profile; the second rainfall peak was centred closer to the flow gauging point spatially; or that the calibrated value of $C_{ini}$ was too low (i.e., the soil moisture store was less full in calibration

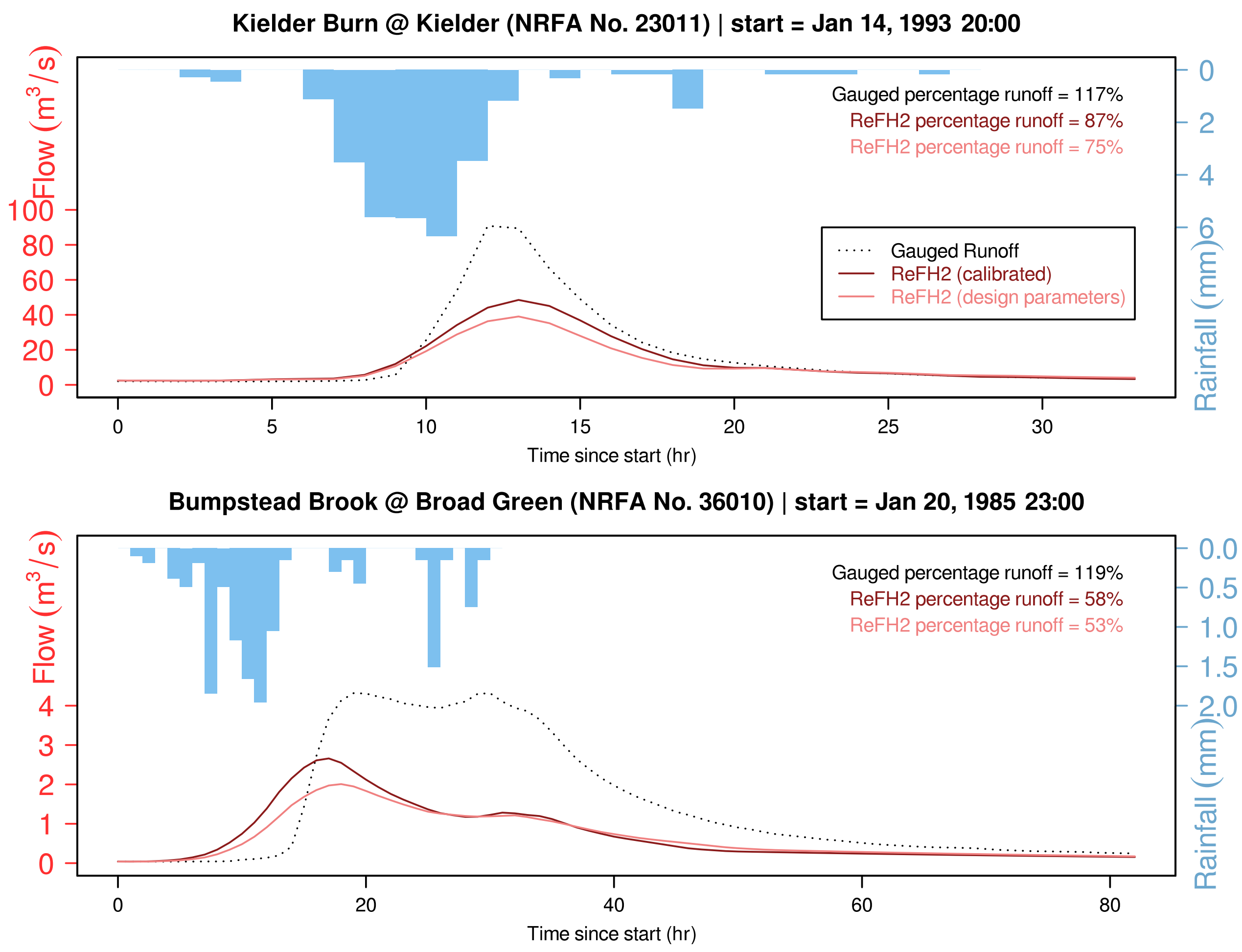


**FIGURE 6** Example hydrographs for two events where modelled runoff is considerably less than gauged runoff

than in reality), leading to more simulated infiltration of the second rainfall peak. The gauged percentage runoff of 119% again appears to violate the water balance, implying systematic errors in estimation of rainfall, runoff or contributing area, or an inaccuracy in representing the initial catchment state through $C_{ini}$. In all cases, the gauged runoff hydrograph is itself an estimate, as is the extrapolation of point rainfall records to a catchment-average hyetograph. Hence, reported 'errors' in model outputs cannot be solely attributed to deficiencies in model structure or poor parameterization.

In addition to the situations described above, total runoff could legitimately exceed total rainfall over the period of an event when there is a significant baseflow component, resulting from the continuing long-term runoff from previous events. Though this is not apparent in the examples presented in Figure 6, it illustrates another way that a total percentage runoff of over 100% may occur and be recorded during an event, even though to do so would fail to close the water balance, appearing to invalidate event-based modelling.

## 5 | IMPLICATIONS FOR DESIGN EVENT MODELLING

Overall, the results show that ReFH2 is consistently able to estimate runoff volumes associated with real rainfall events accurately. However, the vast majority of the 780 flood events in the study data set are not comparable in volume or peak flow to typical design floods. The reasons for this are unavoidable, and include the very low probability of an extreme flood occurring at an active gauging site and the higher potential for damage to monitoring equipment during extreme flows. Consequently, care must be taken when generalising from this study's results to long-return period design event modelling.

One limitation of standard design event rainfall-runoff modelling is that there is only one design rainfall event for each return period, producing one peak flow and one runoff volume. In contrast, a real rainfall event may produce a high peak flow but a small runoff volume, or vice versa. The 2013–2014 UK winter floods exemplify this: flow at Kingston gauging station, near the Thames tidal limit, remained above 250 $m^3/s$ for 76 days, more than doubling the previous record set in 1947 (Muchan, Lewis, Hannaford, & Parry, 2015). However, the peak flow during the same period, 507 $m^3/s$, was only the 12th highest since 1883. In general, a $T$-year peak flow and a $T$-year runoff volume may not coincide, as extreme peak flows are caused by extreme rainfall peaks (shorter-duration accumulations) but extreme flood volumes are often caused by extreme total rainfall depths (longer-duration accumulations). Some design storms, such as those used in ReFH2 and its predecessors, conflate these two characteristics by using fixed rainfall profiles that are independent of return period or rainfall duration. Use of these profiles causes total depth and peak intensity to rise by the same percentage as event duration or return period is increased, and may risk a situation where the most intense part of the design storm has a significantly different rarity than the whole design event. Composite storm profiles (e.g., Keifer & Chu, 1957) are designed so that any subset of the storm centred on the peak has the same return period as the whole storm. This implies a perfect correlation between the occurrences of extreme short-duration and extreme long duration rainfall, even though it is easy to find real counter-examples such as the 2013–2014 UK winter floods mentioned previously.

Multivariate modelling based on observed flood hydrograph volume and observed annual maximum flood peak (Requena, Chebana, & Mediero, 2016) generates a continuum of results, where a range of combinations of peak flow and runoff volume correspond to one return period that reflects the joint occurrence of that peak and volume. However, the relationship between $T$-year peak flow and $T$-year runoff volume is complicated by climate change and the possibility that each may be affected differently. Blöschl et al. (2019) show that the mean annual flood discharge per decade is increasing particularly quickly around the border between England and Scotland; Griffin, Vesuviano, and Stewart (2019) show the same for annual maximum peak flows. In this region, floods normally result from winter rains on catchments with high soil moisture (Bayliss & Jones, 1993) and hence limited capacity to accept rainfall. From this, a flattening of the relationship between percentage runoff and return period is inferred, as available soil storage capacity becomes a smaller fraction of an increasing $T$-year rainfall depth, causing higher percentage runoff for increasingly common rainfall events. Conversely, Gadian et al. (2018) estimate that in summer, total hours of heavy precipitation (>7.6 mm/hour) will increase but individual storms will not increase in volume. Coupled with predicted longer interevent periods and hence typically drier initial soil conditions, it is possible that specified flow peaks will become more common while specified runoff volumes become rarer.

## 6 | CONCLUSIONS

In the UK, practitioner guidance on the choice of flood estimation method tends to focus on peak flow estimation, despite the increasing importance of whole-hydrograph and event volume estimation. A dataset of

780 events recorded in 81 catchments was used to evaluate ReFH2, a rainfall-runoff model widely used in the UK for flood risk analyses, in terms of its ability to model event runoff volume accurately.

It was found that accurate estimation of runoff volume in ReFH2 was strongly correlated with accurate estimation of peak flow, and that over- or under-estimation of event volume was in fact typically smaller than over- or under-estimation of peak flow. Accurate estimation of either peak flow or runoff volume did not depend strongly on accurate calibration of ReFH2's model parameters, corresponding to baseflow lag time, baseflow recharge ratio, unit hydrograph time-to-peak, and maximum soil moisture storage capacity. For individual events, it is possible for estimated parameter values that are far below or above calibrated values to give accurate peak flow and runoff volume estimates. On a case-by-case basis, these occurrences cannot be solely credited to the inaccurate parameterization introducing a modelling error that cancels out the measurement error.

Expressing runoff volumes as a percentage of catchment-average rainfall volume demonstrated that it is not always possible, expected or realistic for a water balance over an event to close. There can be many reasons for this, including but not limited to errors in gauging, uncaptured spatial variations in rainfall, inaccurate estimation of the initial catchment state (e.g., available infiltration capacity), significant or rising baseflow, neglecting subsurface flows, or neglecting anthropogenic flows or modifications to the catchment.

This research demonstrates confidence that the UK's industry-standard rainfall-runoff model can estimate the volume of runoff associated with real rainfall events accurately. However, the majority of observed events are considerably smaller in both peak flow and runoff volume than typical design events of interest, such as the 6-hr, 100-year event widely used in urban drainage design. Further research is required on the accuracy of modelled runoff volumes for long-return period design events, noting that very limited verification event data may be available for this work. More broadly, research is needed to estimate how flood event volume-frequency relationships are related to peak flow-frequency relationships in the UK, how a *T*-year event might be defined as a range of peak flow and event volume combinations, and how these combinations might change under a changing climate.

## ACKNOWLEDGEMENTS

This work was supported by UKCEH National Capability funding. The authors thank James Miller (UKCEH) and two anonymous reviewers for their helpful comments on previous drafts of this manuscript.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are owned by UKCEH (runoff records and catchment descriptors), Wallingford HydroSolutions (calibrated model parameter values), and the Met Office (rainfall records). Restrictions apply to the availability of these data, which were used under license for this study. Data are available from and with the permission of the respective data owners.

## ORCID

*Gianni Vesuviano* https://orcid.org/0000-0003-2157-8875

## REFERENCES

Appleby, F. V. (1974). *Unpublished notes on baseflow modelling*. London: Imperial College.

Archer D, Foster M, Faulkner D and Mawdsley J. 2000. *The synthesis of design flood hydrographs*. Flooding: Risks and reactions, proceedings of CIWEM/CIE Conference (London, October 5, 2000). London: Chartered Institution of Water and Environmental Management; pp. 45–57.

Bayliss, A. (1999). *Catchment descriptors (flood estimation handbook volume 5)*. Wallingford: Institute of Hydrology.

Bayliss, A. C. (2003). *Revitalisation of the FSR/FEH rainfall-runoff method: Flood event database*. Wallingford: Centre for Ecology & Hydrology.

Bayliss, A. C., Black, K. B., Fava-Verde, A., & Kjeldsen, T. R. (2006). *$URBEXT_{2000}$—A new FEH catchment descriptor. Calculation, dissemination and application*. London: Department for Environment, Food and Rural Affairs.

Bayliss, A. C., & Jones, R. C. (1993). *Peaks-over-threshold flood database: Summary statistics and seasonality (IH report 121)*. Wallingford: Institute of Hydrology.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., … Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*, 108–111.

Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). *Hydrology of soil types: A hydrologically-based classification of the soils of the United Kingdom (IH report 126)*. Wallingford: Institute of Hydrology.

CEH. (2014). *Integrated hydrological digital terrain model*. Wallingford: Centre for Ecology & Hydrology Available from https://www.ceh.ac.uk/services/integrated-hydrological-digital-terrain-model

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. Singapore: McGraw-Hill.

CNC. (2017). *Flood estimation: Technical guidance (GN008)*. Cardiff: Cyfoeth Naturiol Cymru.

Dalrymple, T. (1960). *Flood-frequency analyses, manual of hydrology: Part 3. US geological survey water-supply paper 1543-A*. Washington, DC: United States Government Printing Office.

Environment Agency. 2012. *Estimating flood peaks and hydrographs for small catchments (Phase 2). Report SC090031/R0*. Bristol: Environment Agency.

Environment Agency. (2017). *Making better use of local data in flood frequency estimation*. Bristol: Environment Agency.

Gadian, A. M., Blyth, A. M., Bruyere, C. L., Burton, R. R., Done, J. M., Groves, J., … Warner, J. L. (2018). A case study of possible future summer convective precipitation over the UK and Europe from a regional climate projection. *International Journal of Climatology*, *38*, 2314–2324.

Griffin, A., Vesuviano, G., & Stewart, E. (2019). Have trends changed over time? A study of UK peak flow data and sensitivity to observation period. *Natural Hazards and Earth System Sciences*, *19*, 2157–2167.

Griffin, A., Young, A., & Stewart, L. (2019). Revising the BFIHOST catchment descriptor to improve UKflood frequency estimates. *Hydrology Research*, *50*(6), 1508–1519.

Houghton-Carr, H. (1999). *Restatement and application of the flood studies report rainfall-runoff method (flood estimation handbook volume 4)*. Wallingford: Institute of Hydrology.

Institute of Hydrology. (1999). *Flood estimation handbook*. (5 volumes. Wallingford: Institute of Hydrology.

Jones, S. B. (1983). *The estimation of catchment average point rainfall profiles (IH report 87)*. Wallingford: Institute of Hydrology.

Keifer, D. J., & Chu, H. H. (1957). Synthetic storm pattern for drainage design. *ASCE Journal of the Hydraulics Division*, *83.HY4*, 1–25.

Kjeldsen T R. 2007. *Flood estimation handbook supplementary report no. 1. The revitalised FSR/FEH rainfall-runoff method*. Wallingford: Centre for Ecology & Hydrology.

Kjeldsen, T. R., Jones, D. A., & Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation*. Bristol: Environment Agency.

Kjeldsen, T. R., Miller, J. D., & Packman, J. C. (2013). Modelling design flood hydrographs in catchments with mixed urban and rural land cover. *Hydrology Research*, *44*(6), 1040–1057.

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., & Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, *515*, 59–70.

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, *11*(1), 483–499.

Muchan, K., Lewis, M., Hannaford, J., & Parry, S. (2015). The winter storms of 2013/2014 in the UK: Hydrological responses and impacts. *Weather*, *70*(2), 55–61.

NERC. (1975). *Flood Studies Report* (5 volumes. London: Natural Environment Research Council.

Requena, A. I., Chebana, F., & Mediero, L. (2016). A complete procedure for multivariate index-flood model application. *Journal of Hydrology*, *535*, 559–580.

SEPA. (2017). *Flood Modelling guidance for responsible authorities, version 1.1*. Edinburgh: Scottish Environment Protection Agency.

Stewart E, Vesuviano G, Morris D G and Prosdocimi I. 2014. *The new FEH rainfall depth-duration-frequency model: results, comparisons and implications*. 12th British Hydrological Society National Symposium. Birmingham, September 2–4, 2014.

Vesuviano, G., & Miller, J. D. (2019). Design flood estimation and utility of high-resolution calibration data in small, heavily urbanised catchments. *Journal of Flood Risk Management*, *11* (2), e12464.

Wallingford HydroSolutions. (2016). *The revitalised flood hydrograph model ReFH 2.2: Technical guidance*. Wallingford: Wallingford HydroSolutions.

Wallingford HydroSolutions. (2018). *ReFH2 technical note: Applying ReFH2 FEH13 in small clay catchments*. Wallingford: Wallingford HydroSolutions Available from https://www.hydrosolutions.co.uk/app/uploads/2018/06/Applying-ReFH2-FEH13-in-clay-dominated-catchments-v1.pdf

**How to cite this article:** Vesuviano G, Stewart E, Young AR. Estimating design flood runoff volume. *J Flood Risk Management*. 2020;13:e12642. https://doi.org/10.1111/jfr3.12642