



Estimating the UK Index Flood: an Improved Spatial Flooding Analysis

Marinah Muhammad^{1,2} · Zudi Lu¹

Received: 17 January 2019 / Accepted: 5 May 2020 / Published online: 16 June 2020
© The Author(s) 2020

Abstract

Flooding is one of the major natural hazards in the UK. Accurate flood estimation at ungauged catchment is an important component to understand and mitigate flood hazards, but still a difficult issue. This study therefore attempts to explore and improve an index flood estimation model, known as the FEH-QMED model, popular in the UK. It was developed under the assumption that the index flood of QMED, i.e., the median of the set of annual maximum (AMAX) flood data, standing for a flooding level of 2-year return period, can be explained by catchment descriptors. In this study, two fundamentals are empirically explored, including assessing reliability of the nonlinear functional impacts of the catchment descriptors on the logarithmic transformation of QMED, specified by the FEH-QMED model, and the potential to improve the model for more accurate index flood estimation, based on the flooding data of 586 gauged stations across the UK. Through a spatial additive regression analysis, we empirically find that the nonlinear impacts of the catchment descriptors in an updated FEH-QMED model appear reliable. However, spatial correlation tests including Moran's I and Lagrange multiplier tests show that strong spatial dependence exists in the residuals of, but was not fully taken into account by, the QMED type models. We have therefore empirically established new spatial index flood estimation models by proposing spatial autoregressive models to model the impacts of the neighboring sites. Cross-validation assessments demonstrate that the suggested spatial error-based index flood model outperforms the updated FEH-QMED model with a significant improvement, which is robust in the sense of different error measures, say by a reduction of 13.8% of the mean squared error of prediction, for the UK index flood estimation.

Keywords Index flood estimation · Flood catchment descriptors · Flood Estimation Handbook (FEH)-QMED Model · Nonlinear effect of covariates · Spatial neighboring effect · Spatial error model

1 Introduction

Flooding is a major natural hazard in the UK [53], with more than 5 million people living or working in flood prone areas in England and Wales alone [54]. In fact, flooding had caused extensive damage to infrastructure. For example, there was \$3.5 billion paid out by insurance industry due to flood damage between 1998 and 2005 according to [21]. Therefore, flooding is a serious risk of huge economic

loss in the UK, particularly in winter (c.f. [1]). As one of the hydrological phenomena, flooding is a process that is complex and uneasy to understand. Our basic objective, in this study, is therefore to empirically investigate and assess the reliability of, and explore a potential to improve, the existing popular model for the UK index flood estimation.

Due to the emergence of modern data acquisition techniques together with powerful computational tools, voluminous hydrological data have been, and will continue to be, collected. There is an urgent need to explore the unknown and unexpected features in the flooding datasets and assess the reliability of the existing models in the literature. In particular, in hydrology, the issue of an ungauged basin is a challenging problem because drainage basins or waterways, in many parts of the world, are ungauged or poorly gauged, due to the fact that the river flow monitoring stations are mostly located at specific, strategic locations only (c.f. [40]).

✉ Zudi Lu
Z.Lu@soton.ac.uk

¹ Southampton Statistical Sciences Research Institute, and School of Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, UK

² Faculty of Earth Science, Universiti Malaysia Kelantan, Jeli Campus, 17600 Jeli, Kelantan, Malaysia

It is a usual practice that hydrologists have to estimate the design flood using available information from nearby gauged sites and characteristics of locally regional flood catchments when a development project is located at an ungauged basin of a catchment area. This approach is known as a regionalization procedure. It was proposed by [47] for hydro-meteorological modelling with the basic principle of “substitute time for space.” There are several types of regionalization procedures available in the literature (see, for example, [16, 17]). Among them, a popular approach is the index flood procedure introduced by [18], which requires estimation of an index flood. The median value of annual maximum flow series has been chosen as the base for the index flood, that is denoted as QMED. The median means that half of all observed annual maximum flow data at the site of interest are greater than the QMED. In general, a return period, T , is the length of an average time interval between occurrences of two consecutive floods that exceed a threshold of particular magnitude. The probability of (re)occurrence or the so-called exceedance probability with the corresponding magnitude is therefore thought to be $\frac{1}{T}$. This probability is equal to a half for the QMED with $T = 2$, which hence implies that the QMED stands for a flood with a return period of 2 years. The focus in this paper is on estimating this index flood at ungauged basins by using an indirect method, such as regression models, to forge a link between a hydrological parameter of index flood and a set of catchment descriptors, based on the assumption that the flood characteristics can be explained by the catchment descriptors (c.f. [42]).

Different regression models have been developed for estimating the index flood by using different catchment descriptors. A well-established model of this kind in the UK is the median annual maximum flood (QMED) model given in Flood Estimation Handbook (FEH) (c.f. [30, 36]). This model was first published in 1999 and later improved in 2008 by the Centre for Ecology and Hydrology (CEH), and has been widely applied in the UK, e.g., for flood defense planning, flood risk analysis, new development planning and rarity assessment of notable rainfalls or floods. Actually, the FEH has become the industry standard used to estimate local flood risk and develop resilient infrastructure. For details, the reader is referred to [30], [36] and [33]. In the wider literature of hydrological applications, a direct method of using arithmetic median or mean of annual maximum flow data is preferable, in determining the scale values (index flood), to the indirect statistical regression models when flow data of 2–5 years are available, which is however only possible at a gauged catchment area [28, 30]. The literature also shows that statistical indirect models through regression methods, such as the QMED model, are generally more accurate, for prediction of index flood

in ungauged catchment, than conceptual indirect models (see, e.g., [9]), though the former may lack an ability to provide physical interpretation or insight into other inter-related factors in the flood process, causing possibly a large amount of uncertainty. In a broader context, a reliable and valid index flood estimation model is needed and the base for mitigating the hazards of an infrequent extreme event (say 1 in 1000 years event) through flood insurance (c.f. [25]). It can also be used to estimate the scale values that are crucial in determination of frequency and magnitude of the flood peak discharges at a particular ungauged catchment area when designing infrastructure and planning land use [55].

Assessing and improving estimation of index flood at ungauged basins still need a great deal of efforts to dedicate to owing to a large amount of uncertainties associated with index flood estimation procedures (c.f. [34, 50]). In this study, we are focusing to explore two fundamentals on the QMED type model of the FEH, or simply the FEH-QMED model hereafter, in estimating the UK index flood, on the basis of the flooding data of 586 gauged stations around the UK: (i) assessing the reliability of the non-linear functions of different catchment descriptors specified by the updated QMED type model and (ii) exploring the potential to improve the QMED type model. For these purposes, we will choose the FEH-QMED model as a benchmark due to its wide recognition and being taken as basis of most studies on the flood risk estimation and modelling in the UK. The updated flood information in conjunction with the gauged stations and the high quality data on flood peak are available through the National River Flow Archive. The Centre for Ecology and Hydrology (CEH) has as well published the FEH-CD ROM that provides a large number of catchment descriptors, and now the service is made by a website: <https://fehweb.ceh.ac.uk/>. These sources of data make it possible for us to explore our assessment and improvement analyses.

We are making our assessment and potential improvement by proposing appropriate spatial analysis tools to be applied to explore the flooding data set. In order to assess the reliability of the non-linear effects of the catchment descriptors identified by the FEH-QMED model, we will suggest a spatial additive regression analysis (c.f. [39]) to examine the non-linear effects of the catchment descriptor covariates in flood index estimation. This is a kind of semiparametric model [26], by which some unknown or unexpected complex nonlinear features, if existed, would be well identified from the data. Furthermore, in order to check any potential spatial neighboring dependence in the data, a diagnostic tool named Moran's I test (c.f. [45, 48]) will be utilized to examine any significant dependence in the regression residuals. In the literature, although spatial correlation or neighboring impact has been noticed in the updated

FEH procedure (c.f. [34, 35]), it appears to have only been partially considered for improvement by using the QMED model together with a transformation of the data from an appropriately selected (nearby) gauged site in estimating the index flood at an ungauged catchment (c.f. [33]). Differently from these references, in this study, by identification of spatial dependence, we will further suggest a spatial flooding model to explore any potential improvement over the FEH-QMED model by fitting the data with the surrounding neighboring impacts to be more fully considered in our model. Interestingly, we find that the additive analysis confirms the non-linear effects of the catchment descriptors specified by the FEH-QMED model, while the Moran test and a spatial error analysis show that the spatial autocorrelations significantly exist in our flood data. These facts lead us to propose spatial autoregressive index flood estimation models by more fully taking account of the spatial dependence, which we will demonstrate outperform the updated FEH-QMED model of [33] in improvement of prediction, which is robust in the sense of different error measures, say by a reduction of 13.8% of the mean squared error for prediction of the logarithmic QMED.

The structure of the remaining sections is organized as follows. Background information on the source of the UK flood data, the index flood procedure and the equation linking a certain hydrological parameters to a set of catchment descriptors will be briefly introduced in Section 2. Section 3 presents the methodology to be used for our spatial additive and error analyses of the UK flood data set. The detailed analyses and results of the empirical study by our suggested methods, including spatial additive analysis, diagnostic autocorrelation test, spatial index flood model analysis and evaluation, will be conducted and discussed in Section 4. Section 5 concludes with the findings of this study.

2 Data and Index Flood Estimation

We first introduce some background information on the UK flood data source used in this study and on the index flood procedure and the FEH-QMED model as a benchmark needed in the following sections.

2.1 Data Source

We are using the data from the National River Flow Archive (NRFA) to examine the FEH-QMED model for index flood estimation in the UK. Two types of data are used, including the observed flood peak data and the physical catchment descriptors data (see Section 2.3 below for more details). The observed flood peak data are from the HiFlows-UK project, which are available at

<http://nrfa.ceh.ac.uk/data/search>, while the data on physical catchment descriptors are from the FEH CD-ROM 3.0, which now can also be accessed through the website at <https://fehweb.ceh.ac.uk/>. The observations for these two data sets are based on the 602 monitoring stations around the UK that were used in the scientific report of [36], published by the Environment Agency. However, some updates and changes have been made through the HiFlows-UK Notes. The important changes include the list of the stations that have been changed, either removed or added in the HiFlows-UK project for improvement of the dataset. Out of the 602 stations, 6 have been changed to other station numbers with the same station names due to incorrect information, 8 were removed and 2 closed due to data quality. As a result, only 586 rural stations will be considered in this study below. The locations for these 586 rural stations are showed in circle in Fig. 1. This figure has been plotted by using ArcGIS software with the background uploaded from Google Earth map, and then overlaid with the location data of the 586 river flow monitoring stations considered in this study.

2.2 The Flood Regionalization Procedure

If flood records are insufficient, regionalization technique is usually required. In fact, availability of the past flow data has always been problematic to hydrologists in estimating flood at the sites of interest such as where projects are to be designed or developed (see, e.g., [18, 28, 29]). Applying flood frequency analysis without use of regionalization at the sites with insufficient length of flow data records introduces significant uncertainties into the flood estimates. By regionalization technique, hydrologists can utilize the available data and information from nearby gauged stations to estimate the flood at the ungauged sites of interest. By [18], the T -year return flood (say the QMED for $T = 2$) at a given site can be expressed as the product of the index flood (i.e., QMED) and the dimensionless regional growth factor for the return period of T years. According to [28], the issue of estimation of the regional growth factor is well studied in the literature, while the main problem lies in estimation of the index flood. This research is therefore focused on improvement of the existing methodology on estimation of the index flood, denoted by μ_Q below. Note that the index flood μ_Q , or QMED, plays an important part of the FEH statistical method, as shown in [33, 34].

There are various issues in the regionalization procedure with the estimation of the index flood [28]. At the gauged sites, an estimate of μ_Q can be obtained directly by calculating the arithmetic median of the available annual maximum flood flow observations (i.e., QMED). However, at ungauged sites, indirect methods have to be used in estimating μ_Q (or QMED) and there is a large amount of uncertainty associated with the estimate [34]. The most

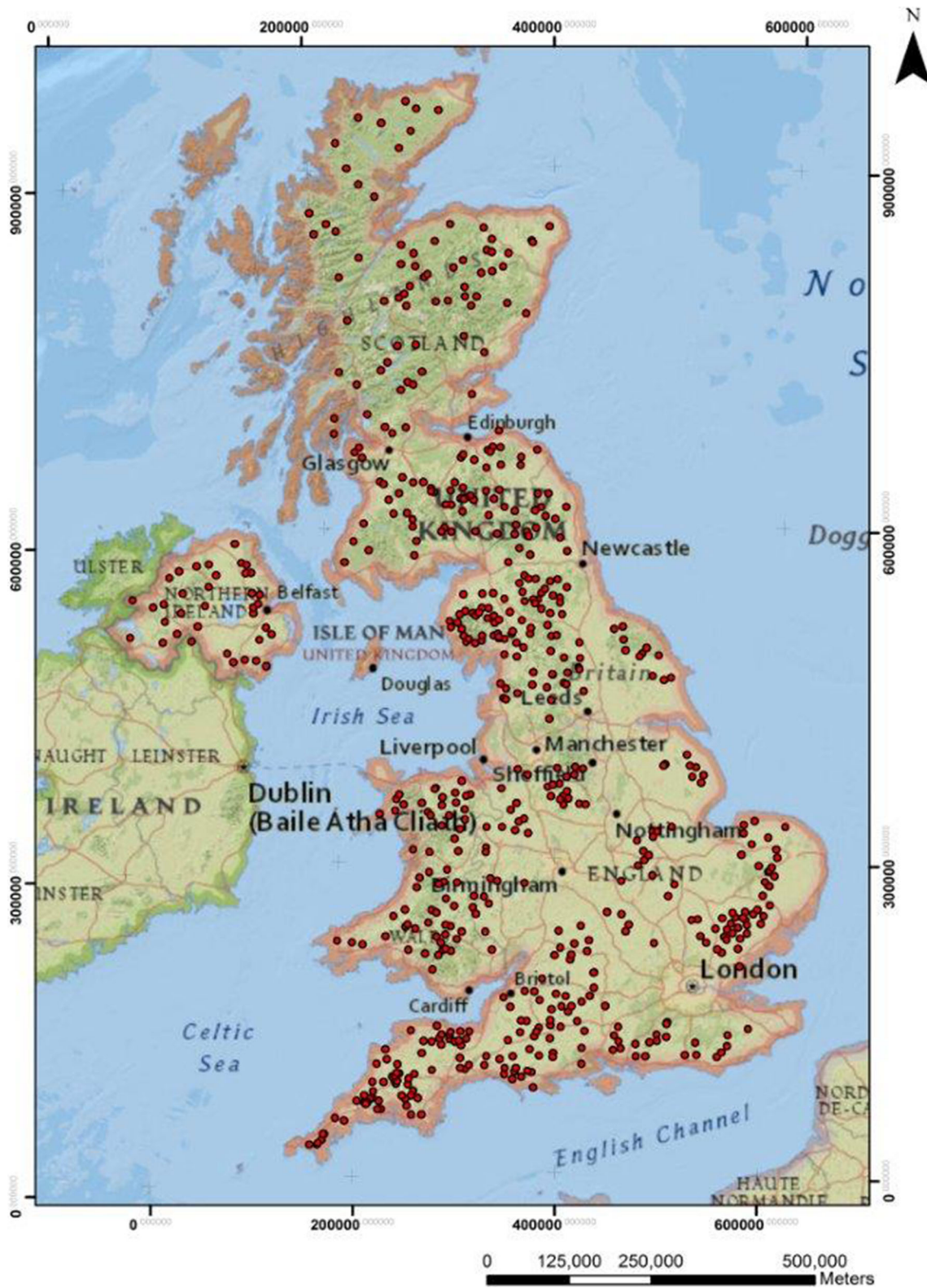


Fig. 1 Locations of the 586 gauging stations on rural catchments, with the background uploaded from Google Earth map

widely used indirect methods are multiregression models, which link μ_Q to a set of appropriate morphological and climatic descriptors of the basins through a power form equation [9, 28]:

$$\mu_Q = ax_1^{\beta_1} x_2^{\beta_2} \dots x_p^{\beta_p}, \tag{1}$$

where x_i is the i^{th} catchment descriptor or some function of it, for $i = 1, \dots, p$, with p the number of the catchment descriptors taken as the explanatory variables for the model, and a and β_i are the model parameters. Note that in the flood regionalization, it is a complex issue deciding an appropriate model structure to link the covariate variables of the chosen catchment descriptors to the dependent variable of river flow. Theoretically, it is impossible to know the “true” model structure (c.f. [55]). Therefore, in practice, hydrologists tend to use some familiar model structure such as the log linear model as indicted in Eq. (1), which is easy to deal with and apply mathematically. By this approach, many index flood estimation models have been developed (c.f. [2, 11, 27, 44, 46, 49]), including those with efforts for improved models [34, 50, 55] (see also [8] for the guidelines on choice of suitable methods and [55] on the potential difficulties for the index flood estimation and modelling). Among, the QMED model published by Flood Estimation Handbook (FEH) [30] is a well-established one for the flood data in the UK.

2.3 The FEH-QMED Model

To introduce the benchmark FEH-QMED model, we need provide some more information on two types of data used in this study.

Typically, there are many flood peaks associated with individual storm events, recorded as continuous time series of streamflow discharges throughout the year. The type of the flood peak data used in this study is the annual maximum (AM) series formed by the maximum discharge in each year, yielding the series $\{w_1 \dots w_n\}$, with w_i for the maximum discharge in the i^{th} year of the n -years record. The AM series can be used to estimate the probability that the maximum flood discharge in a year exceeds a threshold of particular magnitude, say q . This probability is called the annual exceedance probability and can be estimated by $\frac{1}{n} \sum_{j=1}^n I(w_j > q)$, where $I(A)$ is an indicator function with value 1 if A is true and 0 otherwise. The AM series from the 585 stations, with a varying length from 7 to 127 years, will be used to determine the important hydrological parameter in the index flood estimation model, i.e., the median annual maximum flood (QMED) for the index flood.

The CEH (Central for Ecology and Hydrology, formerly known as the Institute of Hydrology, IH) in the UK has also published a CD ROM (c.f. [30]) containing a

modified digital terrain model, called Institute of Hydrology Digital Terrain Model (IHDTM), that has been developed from digitally held rivers and contours taken from the Ordnance Survey 1:50 000 map. The calculation of catchment descriptors in the IHDTM is done through using gridded elevation data. The IHDTM can automatically derive catchment boundary by including the raster data set of drainage path direction based on the steepest route to the neighboring grid nodes in 50 m resolution. This boundary can be applied to compute a large number of catchment descriptors. Nevertheless, the variables that have previously been found to be useful in flood risk study will be included in this study (see Table 1 below for details). The selected characteristic variables are based on [36] who were responsible for improving the original FEH-QMED model, with the updated model, given in Equation (2), simpler but outperforming the original one. We will take this updated model as a benchmark in this study. The observations of all the variables in Table 1 are based on the same 586 stations around the UK.

The updated FEH-QMED model by [36] and [33, 35] uses the median annual maximum flow (QMED) as an index flood for the ungauged catchment flood estimation. The development of this model is based on Eq. (1) that has been linearized by taking logarithm on both sides of (1) for parameter estimation by using generalized least squares or maximum likelihood method for model (2) [19, 51, 52]. The variable selection is carried out by using an exhaustive search to find a relatively small set of variables, with criteria such as the coefficient of determination R^2 , the adjusted R^2 or the predicted error sum of squares applied, providing a good statistical fit to the QMED. This index flood regression model links the QMED with a set of catchment descriptors, namely the catchment area (AREA), the standard average annual rainfall (SAAR), the index of flood attenuation attributable to reservoirs and lakes (FARL) and the based flow index (BFIHOST), explained in Table 1, where the observations for those variables are also based on the same 586 stations around the UK. Equation (2) below is a formula that is widely used in the UK by flood hydrologist to estimate the median flood at an ungauged catchment:

$$\begin{aligned} \ln(QMED) = & 2.1170 + 0.8510 \ln(AREA) \\ & - 1.8734 \left(\frac{1000}{SAAR} \right) + 3.4451 \ln(FARL) \\ & - 3.080BFIHOST^2. \end{aligned} \tag{2}$$

In practice, the quality of flood data has to be ensured before it can be used in any hydrological application analysis. However, some unknown or unexpected features in the flood data cannot be controlled due to its nature. For example, the water level or river flow that represents the flood peak discharge at a particular point in a river

Table 1 Summary of catchment descriptors

Descriptors name	Unit	Range	Note
AREA	km ²	[0; ∞)	Catchment drainage area
SAAR	mm	[0; ∞)	Average annual rainfall
FARL		[0; 1]	Flood attenuation by reservoirs and lakes
BFIHOST		[0; 1]	Base Flow Index (BFI) derived from a spatial dataset of hydrological soil type and land-uses. It is used as a measure of hydrological soil properties

has well-defined spatial effect due to the flow of the river and the lay of the land, while the direction in which the water will flow within a catchment is generally known. Therefore, the local factors influencing the index flood may not be adequately represented in the regression models, and it may be beneficial to adjust the estimates by using local data from the neighboring gauged catchments [33]. The Flood Estimation Handbook (FEH) [30] recommended the use of the data transferring procedure together with the QMED regression model when estimating the index flood at ungauged sites, where one or more suitable gauged sites can be found among the gauged catchments that are geographically close and hydrologically similar. There is no specific guidance as to the limit of the geographical area or necessary degree of hydrological similarity. Nevertheless, studies have shown that geographical proximity gives superior impact in estimating the index flood at ungauged sites in comparison with hydrological similarity [22, 34, 43]. Consecutive studies by [33–35] concluded that geographical proximity representing the spatial impact in the data transferring procedure is very important and useful to improve the regression results.

As [33] has showed, the new FEH-QMED model (2) has been developed by using observations of the spatial data from 602 gauged stations, instead of the 728 stations in the original model, together with the revised data-transfer method that performs better in estimating the index flood at an ungauged catchment. We will use it as a benchmark model as a comparison with our developed models in the following sections.

3 Methodology

It is interestingly noted that the FEH-QMED model (2) is a linear multiple regression model with non-linear effects of the catchment descriptors specified. We are first assessing the reliability of this model in estimating the UK index flood, and then aim to develop an improved model for a better index flood estimation. Here the median annual maximum flood (QMED) in logarithm is considered as the response variable, $Y = \ln(QMED)$, which is a function of the four predictor variables, i.e., the catchment drainage

area (AREA), the average annual rainfall (SAAR), the flood attenuation by reservoirs and lakes (FARL), and the base flow index by hydrology of soil type data (BFIHOST). Two assumptions underlying (2) are as follows: (i) the linearity between $Y = \ln(QMED)$ and the explanatory covariates

$$(X_1, X_2, X_3, X_4) = (\ln(AREA), \frac{1000}{SAAR}, \ln(FARL), BFIHOST^2), \quad (3)$$

and (ii) the independence in the residuals of the multiple regression model (2). These will be assessed by using spatial additive regression analysis to investigate the model structure of the explanatory variables and by spatial error analysis for diagnostic test on any dependence in the residuals, in order to achieve an improved model for index flood estimation. The background information on these methods is provided in Sections 3.1 and 3.2.

3.1 Spatial Additive Analysis

Suppose the response variable, Y , and the $p = 4$ explanatory variables, $X_1 \dots X_p$, given above, are observed with the spatial data, $(Y_i = Y(s_i), X_{i1} = X_1(s_i), \dots, X_{ip} = X_p(s_i))$, $i = 1, 2, \dots, n$, from the $n = 586$ gauged stations around the UK (see Section 2.1), where $s_i = (u_i, v_i)$ stands for the location in (latitude, longitude) of the i th gauged station. In general, the mean of Y_i is a function of X_{i1}, \dots, X_{ip} , say $\mu_i = E(Y_i | X_{i1}, \dots, X_{ip}) = f(X_{i1}, \dots, X_{ip})$. If the FEH model (2) is true, then it is a linear regression

$$\mu_i = \beta_0 + \sum_{j=1}^p \beta_j X_{ij}, \quad (4)$$

where it is often assumed that $e_i = Y_i - \mu_i \sim N(0, \sigma^2)$ for $i = 1, 2, \dots, n$, with e_i 's independent. Equation (4) indicates that the covariates are multiplied by the coefficients, β_1, \dots, β_p for prediction of Y_i .

To explore if Eq. (2) or (4) holds for the given flood data, we are considering a more general model, an additive spatial regression model (c.f. [24, 26, 39]) that is more flexible

than Eq. (4) by replacing the linear predictor with a sum of smoothing functions of the covariates in the form

$$\mu_i = \beta_0 + \sum_{j=1}^p f_j(X_{ij}), \tag{5}$$

where the smooth functions, f_j 's, are unknown non-linear functions of the $p = 4$ covariates (X_1, X_2, X_3, X_4) defined in Eq. (3) in this study, and for identifiability it is often supposed that $E f_j(X_{ij}) = 0$ for $j = 1, \dots, p$. Here although we may assume $e_i = Y_i - \mu_i$ satisfies the conditions as specified under Eq. (4), but it is not necessary for estimation of Eq. (5) as showed in [39] and [26]. By the data $(Y_i, X_{i1} \dots X_{ip}), i = 1, 2, \dots, n$, we can obtain the estimates of the unknown (nonparametric) functions, $\hat{f}_j(X_{ij})$ for $j = 1, \dots, p$, and $\hat{\beta}_0$, by using the so-called backfitting algorithm or other semiparametric smoothing methods; for details, the reader is referred to the related statistical references (c.f. [24, 26, 39]). We did the estimation by the statistical software R with the package ‘‘GAM’’ in the analysis of data in Section 4 below. Clearly, if the resultant estimate $\hat{f}_j(X_{ij})$ is a linear function of X_{ij} for all $j = 1, 2, \dots, p$, then it just demonstrates that Eq. (2) or (4) is true, i.e., the non-linear effects of the catchment descriptors (AREA, SAAR, FARL, BFIHOST) identified in Eq. (3) (c.f. [36]) are well reliable in the FEH-QMED model (2).

3.2 Spatial Dependence and Autoregressive Analysis

After a regression model such as Eq. (4) or (5) is established, we need to check if the obtained model is appropriate by examining the residuals of the regression. In applications, we can define the residuals $e_i, i = 1, 2, \dots, n$, as above, with β_j 's and f_j 's based on their estimated values for calculating the residuals in Eqs. (4) and (5), respectively. Two statistical tests to detect the presence of spatial autocorrelation in the residuals are introduced in this study. Here a spatial weight matrix (c.f. [4]) need to be constructed based on the locations of the $n = 586$ monitoring stations. It is defined as a $n \times n$ matrix, W , for a set of n locations, with elements w_{ij} indicating whether the locations s_i and s_j of the i th and the j th observations are spatially close, where $i \neq j$, and $w_{ii} = 0$ as usual.

Two types of spatial weight matrices are popular, which are contiguity based and distance based, respectively, to be considered in this study. For a contiguity-based matrix, the (i, j) th element is define as

$$w_{ij} = \begin{cases} 1, & \text{if } i \text{ is contiguous to } j, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Here the weight w_{ij} equals one if stations i and j are neighbors in space, and zero otherwise, by the definition

of neighbor based on sharing triangle edges using the Delaunay triangulation concept (c.f. [7, 12, 23, 41]). An advantage with this weight matrix lies in its representation of the proximal action that is an immediate action or neighboring correlation. For another type of spatial weight matrix based on distance, the element is defined as

$$w_{ij} = \begin{cases} 1, & 0 < d_{ij} \leq D, \\ 0, & d_{ij} > D, \end{cases} \tag{7}$$

where d_{ij} is the distance between the locations i and j , with D a defined distance band. Equation (7) indicates that the stations within a defined distance band from each other are categorized as neighbors and have spatial weight equal to one, and zero otherwise. Typically, the upper bound D is taken as the greatest value of the Euclidean distance among the list of the first order nearest neighbors (c.f. [5, 6]). An advantage with this type of weight matrix is for its ability to represent a kind of short-distance action or quasi-proximal action. Both types need to be row-wise standardized so that the weights sum up to one on each row when applied.

We comment that the two types of spatial weight matrices above are much easier to apply than other more complex weight matrices. For example, these two weight matrices only select a relatively small number of the most relevant neighboring gauged sites to be taken into account in modelling, and are hence often more attractive from the application perspectives. For information on other types of more complex weight matrices like using a weight matrix that has values decreasing as distances increase, the reader is referred to the above-mentioned references, which would lead to more neighboring gauged sites to be taken into account in modelling. Whether we should use a more complex spatial weight matrix or not is an interesting model selection issue worth investigation. This is an issue that is beyond the scope of this paper and worth more study, so we will leave it for a future work.

Our first test for spatial dependence in the residuals of a regression model is by Moran's I -statistic [45]. It is the most widely used method in detecting spatial autocorrelation of geo-referenced data (see, e.g., [20, 32, 37]). With a spatial weight matrix W as defined in Eq. (6) or Eq. (7), Moran's I formula for a row-wise standardized weight matrix W reads as

$$I = \frac{e' W e}{e' e}, \tag{8}$$

where $e = (e_1, e_2, \dots, e_n)'$ is an $n \times 1$ vector of the regression residuals. The statistic I asymptotically follows a standard normal distribution under the null hypothesis of no spatial dependence (c.f. [13–15]). In R, this test is implemented using function `lm.morantest`.

The Moran's I test statistics has high power against a range of spatial alternatives, but does not provide much

help in terms of which alternative spatial regression model would be most appropriate. An alternative test for spatial dependence is called Lagrange multiplier (LM) test, which allows a distinction between spatial error models and spatial lag models. Unlike the Moran test, the LM test relies on a structured hypothesis for test of spatial error dependence or spatial lag dependence.

The LM test for spatial error dependence, or the so-called LM error test, is based on the least-squares residuals of spatial error model (11) below involving a spatial weight matrix W , conditional on having a λ parameter not equal to zero in the model. The LM error test statistic takes the form [4]:

$$LM_{err} = \frac{(e'We/s^2)^2}{T} \sim \chi^2(1), \quad (9)$$

with $s^2 = e'e/n$, e denoting a vector of the least-squares residuals and $T = tr[(W + W')W]$. The LM test for spatial lag dependence, or the LM lag test, is based on the residuals from the spatial lag model (12). It can be used to examine whether inclusion of the spatial lag term eliminates spatial dependence in the residuals of the model. This test differs from the two tests outlined above because it allows for the presence of the spatial lag variable in the model. It is conditional on having a ρ parameter not equal to zero in the model, rather than relying on the least-squares residuals. The LM lag test takes the form [4]:

$$LM_{lag} = \frac{(e'Wy/s^2)^2}{nJ} \sim \chi^2(1), \quad (10)$$

where $nJ = T + (WX\beta)'M(WX\beta)/s^2$, $T = tr[(w + w')w]$, $M = I - X(X'X)^{-1}X'$. Both LM tests and their robust forms are implemented in the R function `lm.LMtest`.

As mentioned above, two cases of spatial dependence could arise in Eq. (4) or (5). A spatial error model is appropriate when the interest is in correcting spatial autocorrelation due to use of spatial data, irrespective of whether the model of interest is spatial or not. The spatial error model can be expressed as

$$y = X\beta + e, \quad (11)$$

$$e = \lambda We + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n).$$

Here the $n \times 1$ vector y contains the n observations of the dependent variable Y , and X represents the usual $n \times (1 + p)$ data matrix containing the explanatory variables from model (4), and W is a spatial weight matrix and the parameter λ is the coefficient on the spatially correlated errors. The $(1 + p) \times 1$ vector of parameters, β , reflects the influence of the explanatory variables on variation in the dependent variable Y . We comment that, as often done in practice, assumption of normal distribution for the error vector $\varepsilon \sim N(0, \sigma^2 I_n)$, with I_n an $n \times n$ identity matrix, is applied in modelling here,

which can at least be seen as an approximation, in particular at a stage of investigation with no further information on the distribution. Under such kind of Gaussian assumption (no matter if the true error is Gaussian or not), quasi-maximum likelihood estimation (QMLE) can be applied for estimation of Eq. (11) (see, e.g., [38] for detail of justification on the QMLE in theory). We have also provided a plot of kernel density estimate, together with a normal distribution of the estimated mean and variance, for the data of $Y = \ln(QMED)$, which appears to be normally approximated in distribution as indicated in Fig. 2. Maximum likelihood estimation of the spatial error model [4] can be implemented by the R function `errorsarlm`.

The alternative model is a spatial lag model, which is appropriate when the focus is on the spatial interactions of the dependent variable. Here the existence and strength of spatial interactions are the focus of interest to be assessed. The spatial lag model takes the form:

$$y = \rho Wy + X\beta + \varepsilon, \quad (12)$$

$$\varepsilon \sim N(0, \sigma^2 I_n),$$

where y , X and W are similarly defined as in Eq. (11). The parameter ρ is the coefficient of the spatially lagged dependent variable, Wy , and the parameter β reflects the influence of the explanatory variables on variation in the dependent variable Y . Maximum likelihood method for estimating the parameters of this model [4] can be carried out with the R function `lagsarlm`.

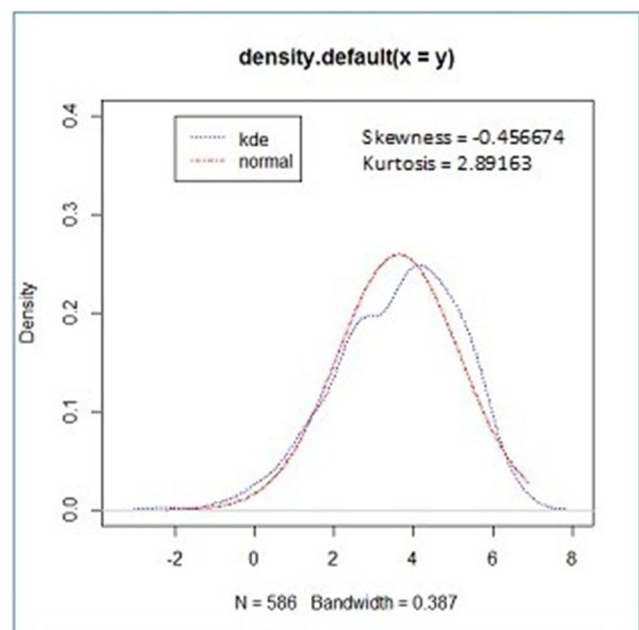


Fig. 2 A plot of kernel density estimate (kde), together with a normal distribution of the estimated mean and variance, for the data of $Y = \ln(QMED)$

4 Empirical Spatial Index Flood Estimation: Results

In this section, we are first conducting empirical spatial additive assessment of the non-linear effects of the covariates specified by the FEH-QMED model given in Eq. (2) and then attempting to develop an improved spatial autoregressive flooding model.

4.1 Empirical Spatial Additive Assessment of the FEH-QMED Model

The analysis starts with fitting the dataset by applying general linear model (4), estimated by using an R function “lm,” with the identified covariates, given in Eq. (3), used in the FEH-QMED model (2). Then we are applying the spatial additive model (5) in Section 3.1, estimated by the “gam” function of the R package “mgcv” [56], to assess if the nonlinear parametric transformations of the catchment descriptors, in Eq. (3), as covariates are reliable in fitting the linear model (4), as given by the FEH-QMED model (2), for the data.

The results from fitting the general linear model (4) and the spatial additive model (5), with $Y = \ln(QMED)$ and the four catchment descriptors in Eq. (3), for the UK flooding data at the 586 gauged stations are shown in Table 2, and the estimated functions of $f_j(\cdot)$'s in Eq. (5) are displayed in Fig. 3. Here the coefficients of the FEH-QMED model given in Eq. (2) (c.f. [36]) are also provided in Table 2 for an easy reference and comparison.

As noticed in Section 3.1, the FEH-QMED model (2) is represented actually in the form of linear regression model (4). The estimated values from the linear model (4) and the FEH-QMED model shown in Table 2 are essentially the same with slight difference owing to using slightly different data sets (c.f. Section 2.1 for detail about the reduced number of the gauged stations considered in this paper). Comparing the linear model with the additive model in Table 2 and Fig. 3, we can see that statistically, both models show the same conclusion with regard to

the significance of the individual covariate effects. The equivalent degrees of freedom (edf, i.e., the number of free parameters for usual parametric models) for all the functions of the four variables are greater than or equal to 1, and the F -statistic values indicate them being significant with their p values less than the significance level of 5%, under the additive model, in Table 2. In terms of model comparison by adjusted R -squares, it is concluded that both the spatial additive and the linear models are very similar with the adjusted R -squared values being 0.938 and 0.935, respectively. Although the adjusted R -square for the additive model (5) is slightly larger, the linear model (4) has better interpretation. Furthermore, for the estimated smoothing functions of the additive model (5) with $p = 4$, plotted in Fig. 3, all the component functions with the four covariates given in Eq. (3) look to be rather linear. This confirms that the specified nonlinear effects of the catchment descriptors in the FEH-QMED model (2) look satisfactorily reliable for the UK flooding data.

4.2 Empirical Tests of Spatial Autocorrelation

We are further investigating whether there is any strong spatial autocorrelation in the residuals of the spatial additive or the linear FEH-QMED model. In fact, if the residuals were a white noise data set, then there should be none of significant autocorrelation when it is seen as a series, which is obviously violated by a simple autocorrelation plot of the residuals as a series for the linear model in Table 2, provided in Fig. 4. We therefore do the spatial tests more formally by using Moran's I and Lagrange multiplier tests as sketched in Section 3.2. This study considers two types of spatial weight matrices that are widely used in estimating the relationship among the spatial observations, that is the contiguity-based and the distance-based matrices. The spatial neighborhoods under contiguity- and distance-based weights are displayed in panels (a) and (b) of Fig. 5, respectively (see Section 3.2 for detail on how to construct these spatial weight matrices with row-wise standardized elements).

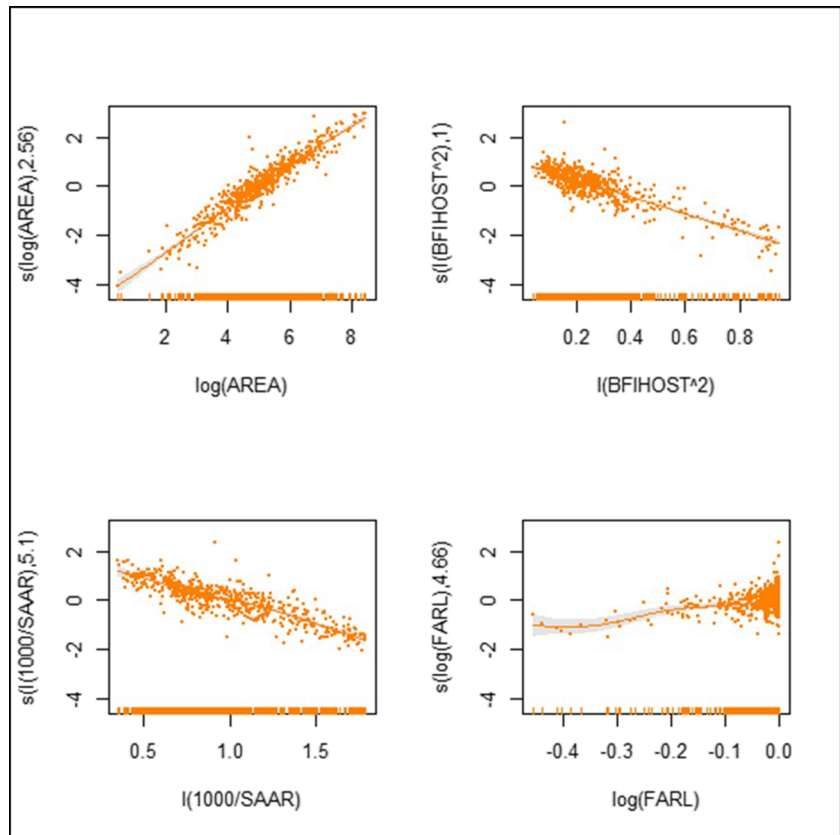
Table 2 Comparison between general linear and additive models

Coefficient	QMED model	Linear model	Additive model		
Intercept	2.1170*	2.1826*	3.6263*		
log(AREA)	0.8510*	0.8527*	2.561	edf	Prob
1000/(SAAR)	- 1.8734*	- 1.8410*	5.098	1218.13	0.000
log(FARL)	3.4451*	3.0335*	4.655	208.99	0.000
(BFIHOST) ²	- 3.0800*	- 3.3479*	1.000	24.13	0.000
Adjusted R -square	0.935	0.935	0.938	1116.56	0.000

edf, effective degree of freedom

*The estimated parameter is significant at 0.01

Fig. 3 Graphical display on effects of four catchment descriptors in additive model



Six tests have been performed to assess the spatial dependence in the residuals. A summary of the six diagnostics tests is reported in Table 3. First, Moran’s *I* score of 0.213 is highly significant, indicating a strong

spatial autocorrelation in the residuals, which are not white noise data. This conclusion is further confirmed by other five tests, including the simple Lagrange multiplier tests for spatial lag dependence (under the spatial lag model (12) and for spatial error dependence (under the spatial error model (11), and the robust Lagrange multiplier tests for spatial lag dependence in the presence of error dependence and for spatial error dependence in the presence of spatial lag dependent variable, and a portmanteau test. From Table 3, it clearly shows that both simple tests of the lag and the error are significant, indicating presence of spatial autocorrelation in the residuals for both types of weight matrices. The robust tests give a better understanding in terms of which alternative spatial dependence would be most appropriate for model fitting, showing that the spatial error model appears more acceptable than the lag model.

Series residuals(lm)

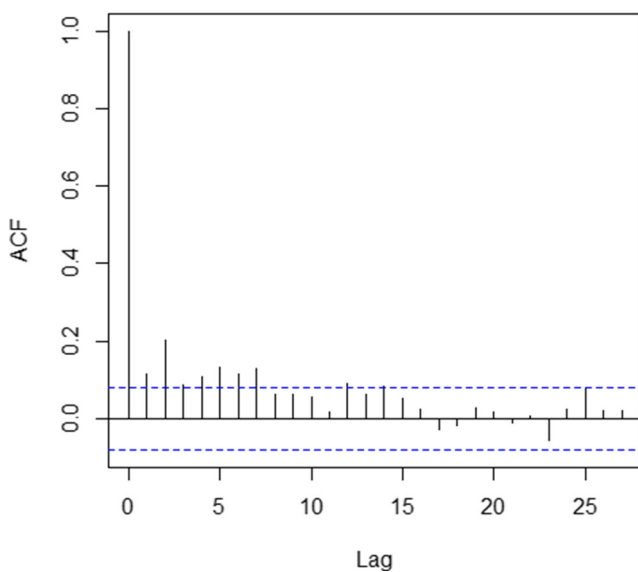
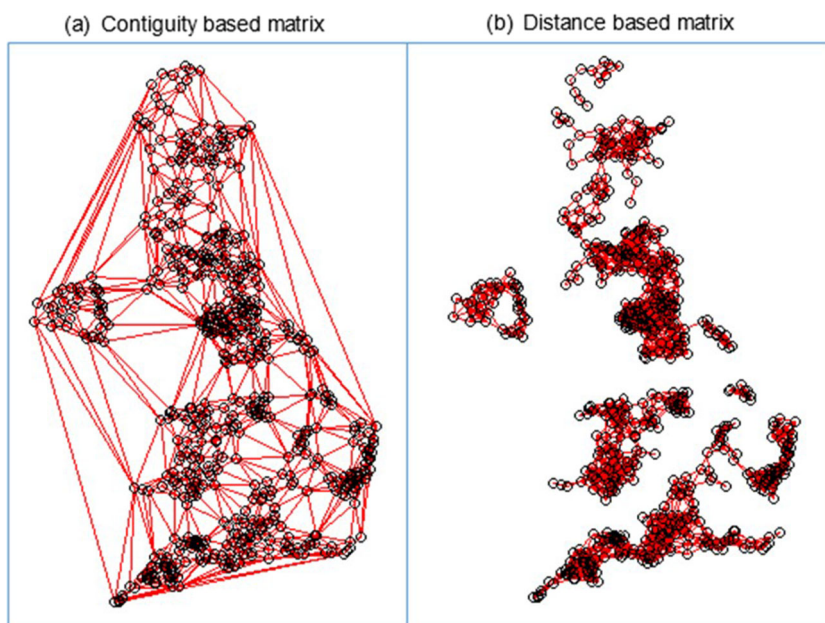


Fig. 4 A simple autocorrelation plot for the residuals seen as a series

4.3 An Improved Index Flood Estimation: Empirical Spatial Autoregressive Flooding Analysis

Based on the features identified in Sections 4.1 and 4.2, we are attempting to find an improved model for the UK index flood estimation. We consider fitting both the spatial error and the spatial lag models introduced in Section 3.2 to the UK flooding data by using both types of spatial weight

Fig. 5 Illustration of spatial relationship in the UK flood peak data



matrices considered above. The results are summarized in Table 4, which gives the estimated values for each of the parameters in all models considered in this analysis, where the value in the round bracket represents the standard error of the estimate. All the estimated regressive coefficients are found to be significant for each model. In addition, the spatial coefficients λ and ρ are significant in both the spatial error and the spatial lag models, given in Eqs. (11) and (12), respectively, for both types of spatial weight matrices. These demonstrate that the spatial relationship in the UK flood peak data can be characterized by both contiguity-based and distance-based weight matrices.

We compare both spatial error and spatial lag models associated with the two types of spatial weight matrices with the benchmark linear FEH-QMED model (Eq. 2). Model evaluation and selection among different fitted models are based on Akaike information criterion (AIC). The AIC is one of the most popular approaches in model selection that compares multiple models by taking into account both fitting accuracy and model parsimony [3, 10]. A preferred

model can be selected via a minimal value of the AIC defined by

$$AIC = 2K - 2\ln(L), \tag{13}$$

where K is the number of estimated unknown parameters in the model and L is the maximum value of the likelihood function for the model. In terms of the AIC value, it is clearly shown in Table 4 that taking account of the spatial dependence in modelling can improve the model performance over the FEH-QMED model. We also see that the spatial error model (11) outperforms the spatial lag model (12), with use of the contiguity weight matrix attaining the lowest AIC value of 519.34, which is significantly lower than the AIC values for other models.

Moran's I test is also examined for the residuals in both the fitted spatial error and spatial lag models to assess whether these spatial flooding models are sufficiently acceptable. The results of the p values (denoted by Prob)

Table 3 Diagnostics for spatial dependence in the regression residuals

Test	Contiguity matrix			Distance matrix		
	DF	Value	Prob	DF	Value	Prob
Moran I (error)	-	0.213	0.0000	-	0.162	0.00000
Lagrange multiplier (lag)	1	20.961	0.0000	1	19.563	0.00001
Lagrange multiplier (error)	1	77.196	0.0000	1	76.611	0.00000
Robust Lagrange multiplier (lag)	1	3.306	0.0690	1	4.672	0.03065
Robust Lagrange multiplier (error)	1	59.541	0.0000	1	64.721	0.00000
Lagrange multiplier (SARMA)	2	80.502	0.0000	2	84.284	0.00000

Table 4 Comparison between general linear and spatial autoregressive models

Coefficients	QMED	Linear	Contiguity matrix		Distance matrix	
			SpErr	SpLag	SpErr	SpLag
Intercept	2.1170	2.1826 (0.0848)	2.1494 (0.0911)	1.6800 (0.1385)	2.1430 (0.1024)	1.6576 (0.1454)
log(AREA)	0.8510	0.8527 (0.0131)	0.8497 (0.0129)	0.8320 (0.0137)	0.8472 (0.0130)	0.8327 (0.0137)
1000/(SAAR)	- 1.8734	- 1.8410 (0.0523)	- 1.8181 (0.0748)	- 1.6554 (0.0660)	- 1.7923 (0.0796)	- 1.6499 (0.0674)
log(FARL)	3.4451	3.0335 (0.2629)	2.9526 (0.2725)	2.8210 (0.2617)	3.1500 (0.2627)	2.9076 (0.2596)
(BFIHOST) ²	- 3.0800	- 3.3479 (0.0969)	- 3.2655 (0.1144)	- 3.1228 (0.1077)	- 3.2959 (0.1169)	- 3.1251 (0.1087)
Lambda			0.4351 (0.0546)		0.5001 (0.0599)	
Rho				0.0975 (0.0220)		0.1007 (0.0232)
AIC		575.90	519.34	558.03	525.78	559.49

The values in the () represent as standard error for each estimated parameter
SpErr, spatial error model; *SpLag*, spatial lag model

given in Table 5 clearly show that spatial autocorrelation is still existent in the spatial lag models but not in the spatial error models. This indicates that allowing the FEH-QMED model error terms modelled by a spatial error model with a contiguity-based weight matrix is acceptable and can significantly improve the model fitting with the spatial effect well characterized. This further confirms that the contiguity-matrix-based spatial error model is more suitable for the UK flooding data than the spatial lag model and other models given in Table 4.

4.4 Cross-validation Assessment of Spatial Index Flood Estimation Models in Prediction

Model evaluation is important in assessing the quality and accuracy of the developed models for practical use in prediction. The key point in assessing the performance of a developed model is to accurately measure the ability of its prediction at an ungauged site by prediction error, $\hat{\epsilon}_j$, say, at site j . Note that the predicted value from a spatial model is not as easy to compute as that from a linear regression

model because it involves spatial neighboring effects in the regression.

This study will suggest the evaluation of a method of computing the predicted values by using leave-one-out cross validation (LOOCV) technique. In this method, for each left-out gauged site j of the $n = 586$ observations of the data set, $(n - 1)$ of them are used for model training and the j th left-out observation is for testing of the prediction at the j th site. By using the training data of $(n - 1)$ observations (except the j th site), we can estimate the model parameters, denoted by $\hat{\beta}$ and $\hat{\lambda}$ or $\hat{\rho}$ in model (11) or (12), which are used to get the predicted value \hat{y}_j that is shown in Table 6, where w_{ik} , $k \neq j$, is the (i, k) -th element of standardized spatial weight matrix, W , i.e., $\sum_{k \neq j} w_{ik} = 1$, for each j . This process is repeated for j from 1 until n .

Hence, the prediction error for each model can be computed as follows

$$\hat{\epsilon}_j = y_j - \hat{y}_j, \quad j = 1, 2, \dots, n, \tag{14}$$

which we will call leave-one-out (LOO) prediction errors, or simply LOO errors. The squared or absolute values

Table 5 Diagnostics for spatial dependence in residuals of spatial autoregressive models

Spatial econometric model	Moran I test			
	Contiguity matrix		Distance matrix	
	Value	Prob	Value	Prob
Spatial lag model	0.1600	0.0000	0.1230	0.0000
Spatial error model	- 0.0114	0.6889	- 0.0029	0.9459

Table 6 Computation of the predicted values

Model	Estimation (using $i = 1, 2, \dots, j - 1, j + 1, \dots, n$)	Prediction ($j = 1, 2, \dots, n$)
Linear	$y_i = x_i \beta + \varepsilon_i$	$\hat{y}_j^{LS} = x_j \hat{\beta}$
Spatial error	$y_i = x_i \beta + \varepsilon_i; \varepsilon_i = \lambda \sum_{k \neq j} w_{ik} \varepsilon_k + e_i$	$\hat{y}_j^{SE} = x_j \hat{\beta} + \hat{\lambda} \sum_{k \neq j} w_{jk} (y_k - x_k \hat{\beta})$
Spatial lag	$y_i = x_i \beta + \rho \sum_{k \neq j} w_{ik} y_k + \varepsilon_i$	$\hat{y}_j^{SL} = x_j \hat{\beta} + \hat{\rho} \sum_{k \neq j} w_{jk} y_k$

of the LOO errors (simply called the squared or absolute errors) are widely applied in evaluating model performance in particular for prediction. Their respective simple arithmetic averages are just the popular LOO cross-validation (LOOCV) mean squared error (*MSE*) and mean absolute error (*MAE*) in machine learning (c.f. [31]). In [33], an error measure equivalent to *MSE*, called root mean squared error (*RMSE*), is applied, which is a square root of *MSE*. These error measures are defined, for example, for prediction of $Y = \ln(QMED)$, by

$$MSE = \frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j^2 = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2, \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n \hat{\varepsilon}_j^2} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}, \tag{16}$$

$$MAE = \frac{1}{n} \sum_{j=1}^n |\hat{\varepsilon}_j| = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|. \tag{17}$$

Clearly, the smaller the MSE, RMSE or MAE, the better the model performs in prediction. but their values are scale dependent. An alternative measure that is based on absolute percentage error, $|(y_j - \hat{y}_j)/y_j|$, is the mean absolute percentage error (*MAPE*),

$$MAPE = \frac{1}{n} \sum_{j=1}^n \frac{|\hat{\varepsilon}_j|}{|y_j|} = \frac{1}{n} \sum_{j=1}^n \frac{|y_j - \hat{y}_j|}{|y_j|}, \tag{18}$$

which is scale free. However, a shortcoming associated with this measure of MAPE is that it is very sensitive to a small number of outlying prediction errors, in particular potentially with small values of $|y_j|$'s. We have also reported this measure of MAPE in model assessment below. We however should take care when explaining this measure for flood prediction, where we are concerned with extreme (large) values of QMED, rather than its small values, which we shall address below.

We are first comparing our proposed spatial flooding models (11) and (12), in terms of the MSE, RMSE and MAE, with the updated FEH-QMED model that is based on a revised data-transfer method of [33] with spatial autocorrelation partially taken account of in prediction. For detail on the updated FEH-QMED model in prediction, the reader is referred to [33] (page 5), where it is recommended that the estimates obtained at ungauged sites by using a regression model should be adjusted by transferring data from one nearby gauged site (if available) as there is a relatively dense gauging network in the UK. Differently from [33], the spatial dependence in our spatial flooding models (11) and (12) is taken into account by the spatial autoregressive models through considering two types of spatial neighboring relationship (i.e., contiguity-based and distance-based spatial weight matrices).

The results for prediction of $Y = \ln(QMED)$ in terms of the MSE, the RMSE and the MAE for different models are summarized in Table 7. Note that the RMSE measure

Table 7 Comparison of model performance for $Y = \ln(QMED)$ by LOOCV method

Model	Spatial weight	Performance measure			
		MSE ¹	RMSE ²	MAE ³	MAPE ⁴
Kjeldson and Jones (2010)	-	0.1586	0.3982	0.2981	0.3274
Spatial error	Contiguity	<i>0.1367</i>	<i>0.3697</i>	<i>0.2701</i>	0.2924
	Distance	0.1388	0.3725	0.2724	0.2994
Spatial lag	Contiguity	0.1509	0.3885	0.2915	0.2905
	Distance	0.1514	0.3891	0.2922	<i>0.2868</i>

¹MSE = mean square error for prediction of $Y = \ln(QMED)$

²RMSE = root mean square error (i.e., \sqrt{MSE}) for prediction of $Y = \ln(QMED)$

³MAE = mean absolute error for prediction of $Y = \ln(QMED)$

⁴MAPE = mean absolute percentage error for prediction of $Y = \ln(QMED)$

Values in italic indicate the smallest value in each of the 4 error measures

Table 8 Comparison of model performance for $QMED = \exp(Y)$ by LOOCV method

Model	Spatial weight	Performance measure			
		MSE ¹	RMSE ²	MAE ³	MAPE ⁴
Kjeldson & Jones (2010)	-	2851	53.39	25.08	0.3074
Spatial error	Contiguity	<i>2281</i>	<i>47.76</i>	<i>21.78</i>	<i>0.2904</i>
	Distance	2369	48.67	22.38	0.2914
Spatial lag	Contiguity	2730	52.25	24.24	0.3125
	Distance	2790	52.82	24.36	0.3129

¹MSE = mean square error for prediction of $QMED$

²RMSE = root mean square error for prediction of $QMED$

³MAE = mean absolute error for prediction of $QMED$

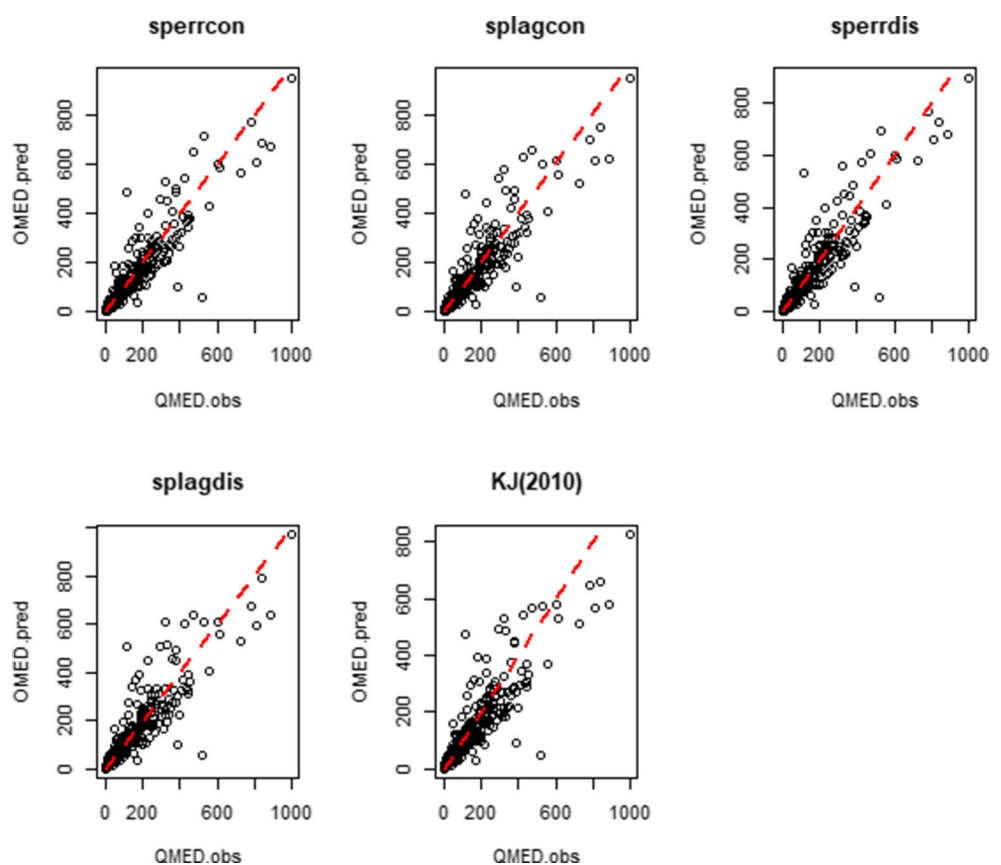
⁴MAPE = mean absolute percentage error for prediction of $QMED$

Values in italic indicate the smallest value in each of the 4 error measures

for prediction of $Y = \ln(QMED)$ was considered by [33]. It is seen from Table 7 that our 4 proposed spatial flooding models, including the spatial error model with contiguity-based weight matrix (sperrcon), the spatial lag model with contiguity-based weight matrix (splagcon), the spatial error model with distance-based weight matrix (sperrdis) and the spatial lag model with distance-based weight matrix (splagdis), all look to outperform the updated FEH-QMED benchmark model of [33] from the prediction perspective in terms of both the MSE, the RMSE and the MAE. Here

for ease of comparisons, the smallest value for each error measure among the 5 models is highlighted in italic in the table. Clearly, the spatial error model with contiguity-based matrix (sperrcon) performs the best in terms of the MSE, the RMSE and the MAE. The percentage of the improvement of this sperrcon flooding model over the updated FEH-QMED model of [33] in reduction of the MSE, the RMSE and the MAE is equal to $(0.1586 - 0.1367)/0.1586 = 13.81\%$, 7.15% and 9.48% , respectively. Therefore, the sperrcon model significantly beats the [33] benchmark model, in

Fig. 6 The scatter plots of the LOOCV based predicted values (y-axis), versus the actual values (x-axis), of the $QMED$ by the 4 models of sperrcon (spatial error model with contiguity weight matrix), splagcon (spatial lag model with contiguity weight matrix), sperrdis (spatial error model with distance weight matrix) and splagdis (spatial lag model with distance weight matrix), and the [33] model. Here the dashed lines stand for the predicted and the observed values being equal in the sub-panels

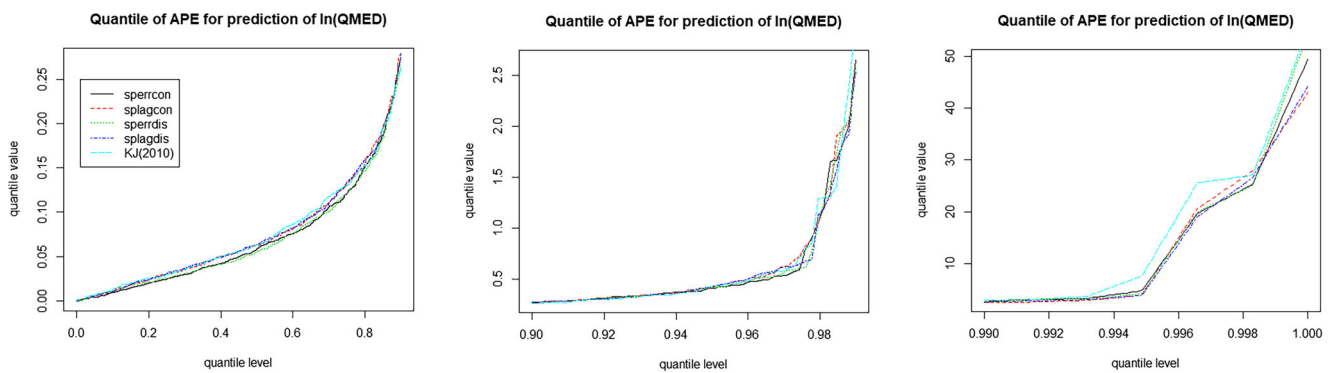


terms of the MSE, RMSE and MAE, for the UK index flood estimation, which is consistent with the model selected by the AIC score in Table 4. As mentioned above, we have also reported, in the 6th column of Table 7, the performance of our 4 proposed spatial flooding models against the model of [33] in terms of the MAPE, defined in Eq. (18). Although in terms of this measure of MAPE it seems to show in Table 7 that the spatial lag models perform better than the spatial error models, but the difference in the MAPE between them is small, and note our sperrcon model still beats the model of [33] in prediction of $Y = \ln(QMED)$ with a reduction of the MAPE by $(0.3274 - 0.2924) / 0.3274 = 10.69\%$. Here we should realize that this MAPE measure may be sensitive to a small number of outlying APE values particularly at the extreme quantile level, say, from 0.998 to 1 (c.f. the quantile curves for the spatial lag models are under those for other models as indicated in the right-hand sub-panel of panel (a) of Fig. 7).

Moreover, notice that it is the $QMED$ that is of a direct interest, rather than $Y = \ln(QMED)$ as an intermediate step, in the practice of index flood estimation. We have

therefore also considered the four error measures of MSE, RMSE, MAE and MAPE for the prediction of the $QMED$ directly, based on $QMED = \exp(Y)$ with Y predicted as shown above, which are reported in Table 8. Here the MSE, RMSE, MAE and MAPE for prediction of the $QMED$ are defined similarly to those for $Y = \ln(QMED)$, with y_j and \hat{y}_j in Eqs. (15)–(18) replaced by $QMED_j = \exp(y_j)$ and $\widehat{QMED}_j = \exp(\hat{y}_j)$. We have also displayed the scatter plots of the LOOCV based predicted value \widehat{QMED} against the observed value $QMED$ for different models in Fig. 6, where the dashed lines stand for the predicted value equal to the observed value. From this figure, the [33] model appears not work well with most of the scattered points under the dashed line, but which model is preferred to in prediction cannot be clearly seen from the scatter plots. Nevertheless, in Table 8, we can clearly see that the spatial lag models perform worse than the other models including the model of [33] in prediction of the $QMED$ in terms of the MAPE. However, in terms of all the four error measures of the MSE, RMSE, MAE and MAPE, the spatial error flooding models, in particular the contiguity weight matrix

(a) Distribution for Errors of Prediction of $\ln(QMED)$



(b) Distribution for Errors of Prediction of $QMED$

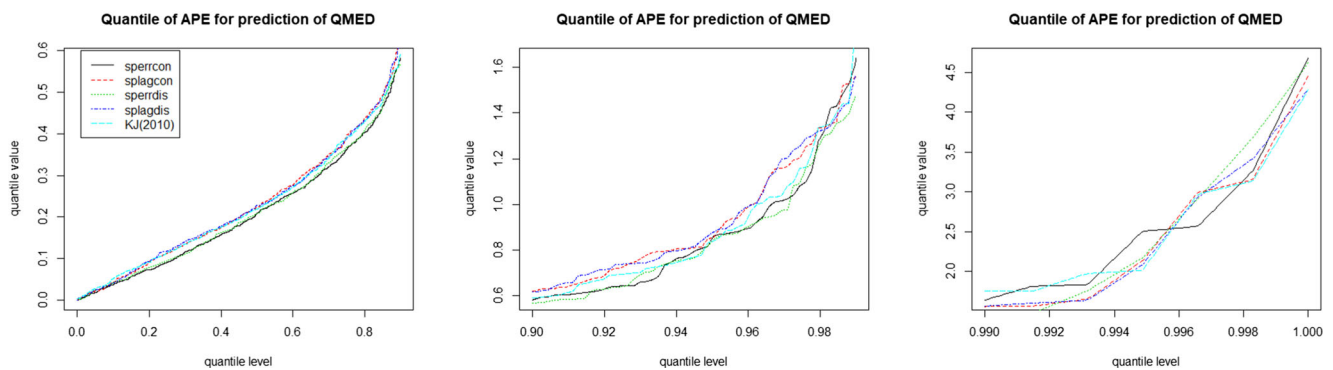


Fig. 7 Plots of quantile value (y-axis) of the absolute percentage error (APE) for the LOOCV based prediction of: **a** $Y = \ln(QMED)$, and **b** $QMED$, against the quantile level (x-axis) for the model of [33] and the 4 models of sperrcon, splagcon, sperrdis and splagdis. For clarity,

the plots are made for the quantile level (x-axis) from 0 to 0.90 (left), from 0.90 to 0.99 (middle) and from 0.99 to 1 (right) in panels **a** and **b**, respectively

based model, all beat the model of [33] significantly in prediction of $QMED$ itself as indicated in Table 8. These results again well demonstrate that our spatial error flooding models (particularly the sperrcon model) outperform the model of [33], which are robust in the sense of different error measures, in estimation of the UK index flood.

We finally provide a comment on the results, particularly on the difference between predictions of $Y = \ln(QMED)$ and $QMED$ in terms of the MAPE. We hence further examine the distributions, in terms of quantiles, of the absolute percentage errors (APEs) for $Y = \ln(QMED)$ and $QMED$ with the different models considered above. Here the quantile value of the APE at the quantile level α , denoted by Q_α , satisfies $P(APE < Q_\alpha) = \alpha$. The plots of the quantile value Q_α of the APE against the quantile level α for different models are displayed in panels (a) and (b) of Fig. 7 for $Y = \ln(QMED)$ and $QMED$, respectively, where for clarity, the quantile levels from 0 to 0.90, from 0.90 to 0.99 and from 0.99 to 1 are partitioned and plotted on the left, the middle and the right panels, respectively. It can be seen in panel (a) of Fig. 7 that the curves of quantile values for both the spatial error and the spatial lag models are basically under that for the model of [33], in particular over the quantile level from 0.99 to 1 (the right-hand panel) with outlying (extremely large) APE values. This explains the performance of the spatial lag models and the spatial error models over the model of [33] in terms of MAPE in Table 7. However, in panel (b) of Fig. 7, the curves of quantile values for the spatial lag models are mostly slightly above that for the model of [33], in particular significantly above for the part of the quantile level from 0.90 and 0.99 (the middle panel). This may also explain why the spatial lag models are slightly worse than the model of [33] in terms of MAPE in Table 8. We can also see in panel (b) that the quantile curves for the spatial error models (sperrcon and sperrdis) are basically under that for the model of [33], so that the spatial error models outperform the [33] model in terms of the MAPE as shown in Table 8. For the performance of the models in terms of other error measures, the results from Tables 7 and 8 can be explained similarly with the spatial error models preferred to the [33] model for the UK index flood estimation by examining the distributions of other error measures, which are omitted to save space.

5 Conclusion

Flood estimation for ungauged catchments is a challenging task. Urgent needs for efficient statistical method to identify and extract the unknown and unexpected complex features from the data sets are obvious, in order to have reliable index flood estimation models. For this purpose, the index flood regression model known as the QMED model in the

Flood Estimation Handbook (FEH), which is the basis of most studies on flood risk estimation and modelling, has been assessed by spatial additive analysis. It is found that the FEH-QMED model (2) appears reliable in identifying the nonlinear effects of the catchment descriptors. Based on this, we have attempted to explore the unexpected features in the UK flood peak data set that went overlooked. We have found that the existing index flood estimation models can be significantly improved by spatial flooding models.

Our empirical findings through the work in the above are summarized as follows:

1. The nonlinear impacts of the explanatory catchment descriptors identified by the FEH-QMED model (2) appear reliable for the UK flooding data through spatial additive assessment.
2. To further improve the FEH-QMED model, we have suggested taking into account the spatial neighboring effects through spatial autoregressive analysis, since the spatial autocorrelation in the residuals of the FEH-QMED type model, as indicated in [35], is significantly strong through Moran's I and Lagrange multiplier tests.
3. The UK flooding data appear to be more suitably modelled by the spatial error type model (11) than by the spatial lag type model (12). Our diagnostic assessment of spatial autoregressive analysis shows that the residuals of the spatial error type models are significantly uncorrelated with p values for contiguity- and distance-based spatial weights being as large as 68.89% and 95.59%, respectively, while the residuals of the spatial lag type models are detected to be significantly correlated yet.
4. We have empirically found that the spatial error type model (11) with contiguity-based spatial weight matrix is a significantly improved new model for the UK index flood estimation in terms of model selection by the AIC value. Moreover, our leave-one-out cross validation assessment shows that our new model outperforms the updated FEH-QMED model based on a revised data-transfer method of [33] in prediction. This conclusion is robust in terms of different error measures, for example, with a reduction of the mean squared prediction error by a percentage of 13.8%.

The improved index flood estimation model for the UK flooding data can be expressed, from Table 4, as

$$\begin{aligned} \ln(QMED_i) &= 2.1494 + 0.8497 \ln(AREA_i) \\ &\quad - 3.2655(BFIHOST_i)^2 \\ &\quad + 2.9526 \ln(FARL_i) \\ &\quad - 1.8181(1000SAAR_i^{-1}) + e_i, \\ e &= 0.4351We + \varepsilon, \end{aligned} \quad (19)$$

where $e = (e_1, e_2, \dots, e_n)'$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ with ε_i being i.i.d. of normal distribution, and W is a $n \times n$ standardized contiguity-based spatial weight matrix based on (6).

It is well known that the FEH-QMED model has become the industry standard used to estimate local flood risk and develop resilient infrastructure. Similarly, it is interesting to investigate the applications of our improved index flood estimation model (19), e.g., for flood defense planning, flood risk analysis, new development planning and rarity assessment of notable rainfalls or floods in the UK, which are left for future research. In addition, potential improvement based on this work may be further considered as discussed in Section 3.2. For example, we may consider model selection with other spatial weight matrices that represent more neighboring sites than the contiguity- or distance-based matrix representation of a short-distance or quasi proximal action in this paper. We may also consider non-Gaussian distribution in model building as shown in Fig. 2. These are beyond the scope of this paper and worth more efforts of investigation for future work.

Acknowledgments The authors are grateful to the editor, an associate editor and three referees for their insightful comments which have greatly improved the presentation of this paper. The first author would like to thank the Ministry of Higher Education of Malaysia and Universiti Malaysia Kelantan for the scholarship offered for this research. We also thank Dr. Thomas Kjeldsen from the Department of Architecture and Civil Engineering at University of Bath for his helps and comments with a lot of related information provided to the first author, which have improved the manuscript. This research was also partially supported by an European Commission's Marie Curie Career Integration Grant, which is acknowledged.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. ABI (2016). New figures reveal scale of insurance response after recent floods. [Retrieved Online; <https://www.abi.org.uk/News/News-releases/2016/01/New-figures-reveal-scale-of-insurance-response-after-recent-floods/>].
2. Acreman, M.C. (1985). Predicting the mean annual flood from basin characteristics in scotland. *Hydrological sciences journal*, 30(1), 37–49.
3. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
4. Anselin, L. (1988). Spatial econometrics: methods and models, volume 4. Springer Science & Business Media.
5. Anselin, L. (2003). Data and spatial weights in spdep, notes and illustrations. Spatial Analysis Laboratory, University of Illinois at Urbana-Champaign. Retrieved online <https://geodacenter.asu.edu/system/files/dataweights.pdf>.
6. Bivand, R. (2016). Creating neighbours. [Retrieved Online; <https://cran.r-project.org/web/packages/spdep/vignettes/nb.pdf>].
7. Bivand, R.S., & Portnov, B.A. (2004). Exploring spatial data analysis techniques using R: the case of observations with no neighbors. In *Advances in Spatial Econometrics*, pages 121–142. Springer.
8. Bocchiola, D., De Michele, C., Rosso, R. (2003). Review of recent advances in index flood estimation. *Hydrology and Earth System Sciences Discussions*, 7(3), 283–296.
9. Brath, A., Castellarin, A., Franchini, M., Galeati, G. (2001). Estimating the index flood using indirect methods. *Hydrological sciences journal*, 46(3), 399–418.
10. Burnham, K.P., & Anderson, D.R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. Springer Science & Business Media.
11. Canuti, P., & Moisello, U. (1982). Relationship between the yearly maxima of peak and daily discharge for some basins in tuscany. *Hydrological Sciences Journal*, 27(2), 111–128.
12. Chen, K., Anthony, S.M., Granick, S. (2014). Extending particle tracking capability with Delaunay triangulation. *Langmuir*, 30(16), 4760–4766.
13. Cliff, A., & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical analysis*, 4(3), 267–284.
14. Cliff, A.D., & Ord, J.K. (1973). Spatial autocorrelation, volume 5. Pion London.
15. Cliff, A.D., & Ord, J.K. (1981). Spatial processes: models & applications, volume 44. Pion London.
16. Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1), 269–290.
17. Cunnane, C. (1989). Statistical distributions for flood frequency analysis. WMO (Series). Secretariat of the World Meteorological Organization.
18. Dalrymple, T. (1960). *Flood-frequency analyses manual of hydrology: part 3*. USGPO: Technical report.
19. Draper, N., & Smith, H. (1981). Applied regression analysis. Number pt. 766 in Applied Regression Analysis. Wiley.
20. Dray, S., Saïd, S., Débias, F. (2008). Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. *Journal of Vegetation Science*, 19(1), 45–56.
21. Ecclestone, P. (2007). Flooding facts. The Telegraph [Online] 15 June. <http://www.telegraph.co.uk/earth/earthnews/3297530/Flooding-facts.html>.
22. Eng, K., Milly, P., Tasker, G.D. (2007). Flood regionalization: a hybrid geographic and predictor-variable region-of-influence regression method. *Journal of Hydrologic Engineering*, 12(6), 585–591.
23. Estivill-Castro, V., & Houle, M.E. (2001). Robust distance-based clustering with applications to spatial data mining. *Algorithmica*, 30(2), 216–242.
24. Friedman, J.H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American statistical Association*, 76(376), 817–823.
25. Gallagher, J. (2014). Learning about an infrequent event: evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics*, 6(3), 206–233.

26. Gao, J., Lu, Z., Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *The Annals of Statistics*, 34(3), 1395–1435.
27. Garde, R., & Kothiyari, U. (1990). Flood estimation in Indian catchments. *Journal of Hydrology*, 113(1), 135–146.
28. Grimaldi, S., Kao, S.-C., Castellarin, A., Papalexiou, S.-M., Viglione, A., Laio, F., Aksoy, H., Gedikli, A. (2011). 2.18 - statistical hydrology. In Wilderer, P. (Ed.) *Treatise on water science, pages 479–517*. Elsevier, Oxford.
29. Hamed, K., & Rao, A.R. (1999). Flood frequency analysis. CRC press.
30. IH (1999). Flood estimation handbook, 5 Volumes. Institute of Hydrology.
31. James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. Springer.
32. Kazembe, L.N., Kleinschmidt, I., Holtz, T.H., Sharp, B.L. (2006). Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *International Journal of Health Geographics*, 5(1), 1.
33. Kjeldsen, T., & Jones, D. (2010). Predicting the index flood in ungauged UK catchments: on the link between data-transfer and spatial model error structure. *Journal of Hydrology*, 387(1), 1–9.
34. Kjeldsen, T.R., & Jones, D. (2007). Estimation of an index flood using data transfer in the UK. *Hydrological sciences journal*, 52(1), 86–98.
35. Kjeldsen, T.R., & Jones, D.A. (2009). An exploratory analysis of error components in hydrological regression modeling. *Water resources research*, 45(2), W02407.
36. Kjeldsen, T.R., Jones, D.A., Bayliss, A.C. (2008). Improving the FEH statistical procedures for flood frequency estimation. Environment Agency.
37. Kosfeld, R., & Dreger, C. (2006). Thresholds for employment and unemployment: a spatial analysis of german regional labour markets, 1992–2000. *Papers in Regional Science*, 85(4), 523–542.
38. Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6), 1899–1925.
39. Lu, Z., Lundervold, A., Tjøstheim, D., Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli*, 13(2), 447–472.
40. Mamun, A.A., Hashim, A., Amir, Z. (2011). Regional statistical models for the estimation of flood peak values at ungauged catchments, Peninsular Malaysia. *Journal of Hydrologic Engineering*, 17(4), 547–553.
41. Maus, A., & Drange, J.M. (2010). All closest neighbors are proper Delaunay edges generalized, and its application to parallel algorithms. Proceedings of Norwegian informatikkonferanse, pages 1–12.
42. Mazvimavi, D., Meijerink, A., Savenije, H., Stein, A. (2005). Prediction of flow characteristics using multiple regression and neural networks: a case study in Zimbabwe. *Physics and Chemistry of the Earth, Parts A/B/C*, 30(11), 639–647.
43. Merz, R., & Blöschl, G. (2005). Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302(1), 283–306.
44. Mimikou, M., & Ordios, J. (1989). Predicting the mean annual flood and flood quantiles for ungauged catchments in Greece. *Hydrological sciences journal*, 34(2), 169–184.
45. Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17–23.
46. NERC (1975). Flood Studies Report, 5 Volumes. Natural Environment Research Council.
47. NRC-US. (1988). *Estimating probabilities of extreme floods: methods and recommended research*. Atlanta: National Academy Press.
48. Paradis, E. (2016). Moran’s autocorrelation coefficient in comparative methods. [Retrieved Online: <https://cran.r-project.org/web/packages/ape/vignettes/MoranI.pdf>].
49. Reimers, W. (1990). Estimating hydrological parameters from basin characteristics for large semiarid catchments. *IAHS Publication*, 191, 187–194.
50. Shu, C., & Burn, D.H. (2004). Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resources Research*, 40(9), W09301.
51. Stedinger, J.R., & Tasker, G.D. (1986). Regional hydrologic analysis, 2, model-error estimators, estimation of sigma and log-pearson type 3 distributions. *Water Resources Research*, 22(10), 1487–1499.
52. Tasker, G.D., & Stedinger, J.R. (1989). An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111(1), 361–375.
53. UKfloods (2012). Introduction to flooding in the UK. <https://sites.google.com/site/ukfloods/>.
54. UK Groundwater Forum (2020). What is groundwater flooding? Online retrieved (on 25/05/2020) from http://www.groundwateruk.org/FAQ_groundwater_flooding.aspx.
55. Wan Jaafar, W., Liu, J., Han, D. (2011). Input variable selection for median flood regionalization. *Water Resources Research*, 47(7), W07503.
56. Wood, S. (2017). Generalized additive models: an introduction with R. Chapman and Hall/CRC, 2 edition.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.